This is a previous version of the article published in Numerical Algorithms. 2019, 82(2): 397-421. doi:10.1007/s11075-018-0608-x

Optimal rates of linear convergence of the averaged alternating modified reflections method for two subspaces

Francisco J. Aragón Artacho* Rubén Campoy†

November 13, 2018

Abstract

The averaged alternating modified reflections (AAMR) method is a projection algorithm for finding the closest point in the intersection of convex sets to any arbitrary point in a Hilbert space. This method can be seen as an adequate modification of the Douglas–Rachford method that yields a solution to the best approximation problem. In this paper we consider the particular case of two subspaces in a Euclidean space. We obtain the rate of linear convergence of the AAMR method in terms of the Friedrichs angle between the subspaces and the parameters defining the scheme, by studying the linear convergence rates of the powers of matrices. We further optimize the value of these parameters in order to get the minimal convergence rate, which turns out to be better than the one of other projection methods. Finally, we provide some numerical experiments that demonstrate the theoretical results.

Keywords Best approximation problem · Linear convergence · Averaged alternating modified reflections method · Linear subspaces · Friedrichs angle

MSC2010: 65F10 · 65K05 · 65F15 · 15A08 · 47H09 · 90C25 · 41A25

1 Introduction

The *averaged alternating modified reflections (AAMR) algorithm*, introduced in [2], is a projection method for solving best approximation problems in the convex setting. A best approximation problem consists in finding the closest point to any given point in the intersection of a collection of sets. In this work we study problems involving two subspaces U and V in $\mathbb{R}^n$.

Given a point $z \in \mathbb{R}^n$, the corresponding best approximation problem is defined as

$$\text{Find } w \in U \cap V \text{ such that } \|w - z\| = \inf_{x \in U \cap V} \|x - z\|. \tag{1}$$

For any initial point $x_0 \in \mathbb{R}^n$, the AAMR algorithm is iteratively defined by

$$x_{k+1} := (1-\alpha)x_k + \alpha(2\beta P_{V-z} - I)(2\beta P_{U-z} - I)(x_k), \quad k = 0, 1, 2 \ldots. \tag{2}$$

When $\alpha, \beta \in]0,1[$, the generated sequence $\{x_k\}_{k=0}^{\infty}$ converges to a point $x^\star$ such that

$$P_U(x^\star + z) = P_{U\cap V}(z),$$

which solves problem (1). Furthermore, the shadow sequence $\{P_U(x_k + z)\}_{k=0}^{\infty}$ is convergent to the solution $P_{U\cap V}(z)$ even if $\alpha = 1$, see [2, Theorem 4.1]. In fact, when the sets involved are subspaces, we prove that the sequence $\{x_k\}_{k=0}^{\infty}$ is also convergent for $\alpha = 1$, see Corollary 3.1.

*Department of Mathematics, University of Alicante, Spain. e-mail: francisco.aragon@ua.es
†Department of Mathematics, University of Alicante, Spain. e-mail: ruben.campoy@ua.es

Several projection methods have been developed for solving convex feasibility problems in Hilbert spaces, see e.g. [3, 6, 8, 11, 14]. In the case where the sets are subspaces, some of these methods converge to the closest point in the the intersection to the starting point, providing thus a solution of the best approximation problem (1). Among these schemes, probably the two most well-known are the method of alternating projections (AP), which was originally introduced by John von Neumann [24], and the Douglas–Rachford method (DR) [12, 20], which is also referred as averaged alternating reflections. The rate of linear convergence of these methods is known to be the cosine of the Friedrichs angle between the subspaces for DR [4], and the squared cosine of this angle for AP [10]. Several relaxations and generalizations of these methods have been proposed, such as the relaxed and the partial relaxed alternating projections (RAP, PRAP) [1, 7, 23], the generalized alternating projections (GAP) [15, 17], the relaxed averaged alternating reflections (RAAR) [21], and the generalized Douglas–Rachford (GDR) [13], among others. We note that AAMR can also be seen as a modified version of DR, since both methods coincide when $\alpha = \frac{1}{2}$ and $\beta = 1$ in (2).

Thanks to the linearity of the projector operator onto subspaces, projection methods reduce to matrix iterations. Taking advantage of this fact, optimal convergence rates have been obtained in [5] for RAP, PRAP and GDR. By following an analogous matrix analysis, the rate of convergence with optimal parameters for GAP has been recently given in [16].

In the current setting, the rate of convergence of the AAMR algorithm was numerically analyzed in various computational experiments in [2, Section 7]. The goal of this work is to provide the theoretical results that substantiate the behavior of the algorithm that was numerically observed. By following the same approach as in [5], we analyze the linear rate of convergence of the AAMR method by studying the convergence rates of powers of matrices. The rate obtained depends on both the Friedrichs angle and the parameters defining the algorithm. In addition, we also obtain the optimal selection of the parameters according to the Friedrichs angle, so that the rate of convergence is minimized. This rate coincides with the one for GAP, which is the best among the rates of all the projection methods mentioned above. This is not just by chance: the shadow sequences of GAP and AAMR coincide for linear subspaces under some conditions (see Theorem 4.1 and Figure 4).

The remaining of the paper is structured as follows. We present some definitions and preliminary results in Section 2. In Section 3 we collect our main results regarding the rate of convergence of the AAMR method. We compare the rate with optimal parameters of AAMR with the rate of various projection methods in Section 4. In Section 5 we perform two computational experiments that validate the theoretical results obtained. We finish with some conclusions and future work in Section 6.

2 Preliminaries

In this work, our setting is the Euclidean space $\mathbb{R}^n$ with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\|$. For a given set $C \subset \mathbb{R}^n$ we denote by $C^\perp = \{x \in \mathcal{H} : \langle c, x\rangle = 0, \forall c \in C\}$ the *orthogonal complement* of C. Given $x \in \mathbb{R}^n$, a point $p \in C$ is said to be a *best approximation* to x from C if

$$\|p - x\| = d(x, C) := \inf_{c\in C} \|c - x\|.$$

The operator $P_C(x) := \operatorname{argmin}\{\|x - c\|, c \in C\}$ is called the *projector* onto C. When C is closed and convex, P_C is single-valued. In the case when C is a subspace, $P_C(x)$ is sometimes called the *orthogonal projection* of x to C, due to the fact that $x - P_C(x) \in C^\perp$.

Throughout this paper, we assume without lost of generality that U and V are two subspaces of $\mathbb{R}^n$ such that $1 \leq p := \dim U \leq \dim V =: q \leq n - 1$, with $U \neq U \cap V$ and $U \cap V \neq \{0\}$ (otherwise, problem (1) would be trivial). We now recall the concept of principal angles and Friedrichs angle between a pair of subspaces, and a result relating both concepts.

Definition 2.1. *The* principal angles *between* U *and* V *are the angles* $0 \le \theta_1 \le \theta_2 \le \cdots \le \theta_p \le \frac{\pi}{2}$ *whose cosines are recursively defined by*

$$\begin{aligned}\cos\theta_k :=& \langle u_k, v_k\rangle \\ =& \max\left\{\langle u, v\rangle : u \in U, v \in V, \|u\| = \|v\| = 1, \langle u, u_j\rangle = \langle v, v_j\rangle = 0 \text{ for } j = 1,\ldots,k-1\right\},\end{aligned}$$

with $u_0 = v_0 := 0$.

Definition 2.2. *The* Friedrichs angle *between* U *and* V *is the angle in* $\theta_F \in]0, \frac{\pi}{2}]$ *whose cosine is*

$$c_F(U, V) := \sup\left\{\langle u, v\rangle \ : \ u \in U \cap (U \cap V)^\perp, v \in V \cap (U \cap V)^\perp, \|u\| \le 1, \|v\| \le 1\right\}.$$

Fact 2.1. *Let* $\theta_1, \theta_2, \ldots, \theta_p$ *be the principal angles between* U *and* V*, and let* $s := \dim(U \cap V)$. *Then we have* $\theta_k = 0$ *for* $k = 1, \ldots, s$ *and* $\theta_{s+1} = \theta_F > 0$.

Proof. See [5, Proposition 3.3]. □

Remark 2.1. *By our standing assumption that* $U \ne U \cap V$*, we have* $s = \dim(U \cap V) < p$.

The projector operator onto subspaces is known to be a linear mapping. The following result provides a matrix representation of the projectors onto U and V, according to their principal angles. We denote by I_n, 0_n and $0_{m\times n}$, the $n \times n$ identity matrix, the $n \times n$ zero matrix, and the $m \times n$ zero matrix, respectively. For simplicity, we shall omit the subindices when the size can be deduced.

Fact 2.2. *If* $p + q < n$*, we may find an orthogonal matrix* $D \in \mathbb{R}^{n\times n}$ *such that*

$$P_U = D\begin{pmatrix} I_p & 0 & 0 & 0 \\ 0 & 0_p & 0 & 0 \\ 0 & 0 & 0_{q-p} & 0 \\ 0 & 0 & 0 & 0_{n-p-q}\end{pmatrix}D^* \quad \text{and} \quad P_V = D\begin{pmatrix} C^2 & CS & 0 & 0 \\ CS & S^2 & 0 & 0 \\ 0 & 0 & I_{q-p} & 0 \\ 0 & 0 & 0 & 0_{n-p-q}\end{pmatrix}D^*, \qquad (3)$$

where C *and* S *are two* $p \times p$ *diagonal matrices defined by*

$$C := \operatorname{diag}(\cos\theta_1, \ldots, \cos\theta_p) \quad \text{and} \quad S := \operatorname{diag}(\sin\theta_1, \ldots, \sin\theta_p),$$

with $\theta_1, \ldots, \theta_p$ *being the principal angles between* U *and* V.

Proof. See [5, Proposition 3.4]. □

2.1 The averaged alternating modified reflections operator for two subspaces

The AAMR operator was originally introduced for two arbitrary closed and convex sets [2, Definition 3.2]. In this section, we present the scheme in the case of two subspaces, as well as some properties of the operator and its set of fixed points within this context.

Definition 2.3. *Given* $\alpha \in]0, 1]$ *and* $\beta \in]0, 1[$, *the* averaged alternating modified reflections (AAMR) operator *is the mapping* $T_{U,V,\alpha,\beta} : \mathbb{R}^n \mapsto \mathbb{R}^n$ *given by*

$$T_{U,V,\alpha,\beta} := (1-\alpha)I + \alpha(2\beta P_V - I)(2\beta P_U - I).$$

Where there is no ambiguity, we shall abbreviate the notation of the operator $T_{U,V,\alpha,\beta}$ *by* $T_{\alpha,\beta}$.

Fact 2.3. *Let* $\alpha \in]0, 1]$ *and* $\beta \in]0, 1[$. *Then, the AAMR operator* $T_{\alpha,\beta}$ *is nonexpansive.*

Proof. See [2, Proposition 3.3]. □

Proposition 2.1. *Let $\alpha \in]0,1]$ and $\beta \in]0,1[$. Then*

$$\operatorname{Fix} T_{U,V,\alpha,\beta} = U^\perp \cap V^\perp.$$

Proof. Observe that

$$x \in \operatorname{Fix} T_{U,V,\alpha,\beta} \Leftrightarrow P_V(2\beta P_U(x) - x) = P_U(x). \tag{4}$$

Moreover, by [2, Proposition 3.4],

$$P_U(x) = P_{U\cap V}(0), \qquad \text{for all } x \in \operatorname{Fix} T_{U,V,\alpha,\beta}.$$

Therefore, if $x \in \operatorname{Fix} T_{U,V,\alpha,\beta}$, then $0 = P_{U\cap V}(0) = P_U(x)$. Using this equality, together with (4), we deduce that $P_U(x) = P_V(x) = 0$, which implies $x \in U^\perp \cap V^\perp$.

To prove the converse implication, pick any $x \in U^\perp \cap V^\perp$. Then we trivially have that $P_V(2\beta P_U(x) - x) = P_U(x)$, and thus $x \in \operatorname{Fix} T_{U,V,\alpha,\beta}$. □

The AAMR scheme is iteratively defined by (2). Using the linearity of the projector operator onto subspaces, we deduce that the iteration takes the form

$$\begin{aligned}
x_{k+1} &= (1-\alpha)x_k + \alpha(2\beta P_{V-z} - I)(2\beta P_{U-z} - I)(x_k) \\
&= (1-\alpha)x_k + \alpha(2\beta P_{V-z} - I)\left(2\beta\left(P_U(x_k+z) - z\right) - x_k\right) \\
&= (1-\alpha)x_k + \alpha\left(2\beta P_{V-z}\left(2\beta\left(P_U(x_k+z) - z\right) - x_k\right) - 2\beta\left(P_U(x_k+z) - z\right) + x_k\right) \\
&= x_k + 2\alpha\beta\left(P_V\left(2\beta\left(P_U(x_k+z) - z\right) - x_k + z\right) - P_U(x_k+z)\right) \\
&= x_k + 2\alpha\beta\left(2\beta P_V P_U(x_k+z) + (1-2\beta)P_V(z) - P_V(x_k) - P_U(x_k+z)\right).
\end{aligned} \tag{5}$$

Fact 2.4. *Let $\alpha \in]0,1]$, $\beta \in]0,1[$ and $z \in \mathbb{R}^n$. Then, one has $\operatorname{Fix} T_{U-z,V-z,\alpha,\beta} \neq \emptyset$ and*

$$\operatorname{Fix} T_{U-z,V-z,\alpha,\beta} = x^* + U^\perp \cap V^\perp, \quad \forall x^* \in \operatorname{Fix} T_{U-z,V-z,\alpha,\beta}.$$

Furthermore, for any $x \in \mathbb{R}^n$,

$$T_{U-z,V-z,\alpha,\beta}(x) = T_{U,V,\alpha,\beta}(x - x^*) + x^*, \quad \forall x^* \in \operatorname{Fix} T_{U-z,V-z,\alpha,\beta}.$$

Proof. Since U and V are subspaces in a finite dimensional space, by [2, Fact 2.11 and Corollary 3.1] we get that $\operatorname{Fix} T_{U-z,V-z,\alpha,\beta} \neq \emptyset$. The remaining assertions are obtained by applying Proposition 2.1 and [2, Proposition 3.6] to $U - q, V - q$ and $-q \in (U-q) \cap (V-q)$, noting that [2, Proposition 3.6] also holds for $\alpha = 1$. □

2.2 Optimal convergence rate of powers of matrices

We denote by $\mathbb{C}^{n\times n}$ ($\mathbb{R}^{n\times n}$), the space of $n \times n$ complex (real) matrices, equipped with the induced matrix norm $\|A\| := \max\{\|Ax\| : x \in \mathbb{C}^n, \|x\| \leq 1\}$. The *kernel* of a matrix $A \in \mathbb{C}^{n\times n}$ is denoted by $\ker A := \{x \in \mathbb{C}^{n\times n} : Ax = 0\}$ and the set of *fixed points* of A is denoted by $\operatorname{Fix} A := \ker(A - I)$. We say A is nonexpansive if $\|Ax - Ay\| \leq \|x - y\|$ for all $x, y \in \mathbb{C}^{n\times n}$.

Definition 2.4. *A matrix $A \in \mathbb{C}^{n\times n}$ is said to be convergent to $A^\infty \in \mathbb{C}^{n\times n}$ if and only if*

$$\lim_{k\mapsto\infty} \|A^k - A^\infty\| = 0.$$

We say A is linearly *convergent to A^∞ with rate $\mu \in [0,1[$ if there exist a positive integer k_0 and some $M > 0$ such that*

$$\|A^k - A^\infty\| \leq M\mu^k, \quad \textit{for all } k \geq k_0.$$

In this case, μ is called a linear convergence rate *of A. When the infimum of all the convergence rates is also a convergence rate, we say this minimum is the* optimal linear convergence rate*.*

For any matrix $A \in \mathbb{C}^{n\times n}$, we denote by $\sigma(A)$ the *spectrum* of A (the set of all eigenvalues). An eigenvalue $\lambda \in \sigma(A)$ is said to be *semisimple* if its algebraic multiplicity coincides with its geometric multiplicity (cf. [22, p. 510]), or, equivalently, if $\ker(A-\lambda I) = \ker\left((A-\lambda I)^2\right)$ (see [5, Fact 2.3]). The *spectral radius* of A is defined by

$$\rho(A) := \max\{|\lambda| : \lambda \in \sigma(A)\},$$

and the second-largest modulus of the eigenvalues of A after 1 is denoted by

$$\gamma(A) := \max\left\{|\lambda| : \lambda \in \{0\} \cup \sigma(A) \setminus \{1\}\right\}.$$

An eigenvalue $\lambda \in \sigma(A)$ with $|\lambda| = \gamma(A)$ is called a *subdominant* eigenvalue.

Fact 2.5. *Let $A \in \mathbb{C}^{n\times n}$. Then A is convergent if and only if one of the following holds:*
(i) $\rho(A) < 1$*;*
(ii) $\rho(A) = 1$ *and* $\lambda = 1$ *is semisimple and is the only eigenvalue on the unit circle.*
When this happens, A is linearly convergent with any rate $\mu \in]\gamma(A), 1[$, and $\gamma(A)$ is the optimal linear convergence rate of A if and only if all the subdominant eigenvalues are semisimple. Furthermore, if A convergent and nonexpansive, then $\lim_{k\to\infty} A^k = P_{\operatorname{Fix} A}$.

Proof. See [22, pp. 617–618, 630] and [5, Theorem 2.12, Theorem 2.15 and Corollary 2.7(ii)]. □

3 Convergence rate analysis

We begin this section with the following theorem that establishes the rate of convergence of the AAMR algorithm in terms of α, β and the Friedrichs angle between the subspaces. We denote the positive part of $x \in \mathbb{R}$ by $x^+ := \max\{0, x\}$.

Theorem 3.1. *Let $\alpha \in]0, 1]$ and $\beta \in]0, 1[$. Then, the AAMR operator*

$$T_{\alpha,\beta} := (1-\alpha)I_n + \alpha(2\beta P_V - I_n)(2\beta P_U - I_n)$$

is linearly convergent to $P_{U^\perp\cap V^\perp}$ with any rate $\mu \in]\gamma(T_{\alpha,\beta}), 1[$, where

$$\gamma(T_{\alpha,\beta}) = \begin{cases} 1-4\alpha\beta(1-\beta), & \text{if } 0 \le c_F < c(\alpha,\beta); \\ \sqrt{4(1-\alpha)\alpha\beta^2 c_F^2 + (1-2\alpha\beta)^2}, & \text{if } c(\alpha,\beta) \le c_F < \widehat{c}_\beta; \\ 1+2\alpha\beta\left(\beta c_F^2 - 1 + c_F\sqrt{\beta^2 c_F^2 - 2\beta + 1}\right), & \text{if } \widehat{c}_\beta \le c_F < 1; \end{cases} \tag{6}$$

with $c_F := \cos\theta_F$ and θ_F being the Friedrichs angle between U and V,

$$\widehat{c}_\beta := \frac{\sqrt{(2\beta-1)^+}}{\beta} \quad \text{and} \quad c(\alpha,\beta) := \begin{cases} \sqrt{\frac{\left((1-4\alpha\beta(1-\beta))^2-(1-2\alpha\beta)^2\right)^+}{4(1-\alpha)\alpha\beta^2}}, & \text{if } \alpha < 1; \\ 0, & \text{if } \alpha = 1. \end{cases} \tag{7}$$

Furthermore, $\gamma(T_{\alpha,\beta})$ is the optimal linear convergence rate if and only if $\beta \neq \frac{1}{1+\sin\theta_F}$ or $\theta_F = \frac{\pi}{2}$.

Proof. To prove the result, we consider two main cases.
Case 1: $p+q < n$. By Fact 2.2, we can find an orthogonal matrix $D \in \mathbb{R}^{n\times n}$ such that (3) holds. After some calculations, we obtain

$$T_{\alpha,\beta} = D \begin{pmatrix} M_{\alpha,\beta} & 0 & 0 \\ 0 & (1-2\alpha\beta)I_{q-p} & 0 \\ 0 & 0 & I_{n-p-q} \end{pmatrix} D^*, \tag{8}$$

where

$$M_{\alpha,\beta} := \begin{pmatrix} 2\alpha\beta(2\beta-1)C^2 + (1-2\alpha\beta)I_p & -2\alpha\beta CS \\ 2\alpha\beta(2\beta-1)CS & 2\alpha\beta C^2 + (1-2\alpha\beta)I_p \end{pmatrix}.$$

Let $s := \dim(U \cap V)$ and let $1 = c_1 = \cdots = c_s > c_{s+1} = c_F \geq c_{s+2} \geq \cdots \geq c_p \geq 0$ be the cosine of the principal angles $0 = \theta_1 = \cdots = \theta_s < \theta_{s+1} = \theta_F \leq \theta_{s+2} \leq \cdots \leq \theta_p \leq \frac{\pi}{2}$ between U and V (see Fact 2.1). By the block determinant formula (see, e.g., [18, (0.8.5.13)]), we deduce after some algebraic manipulation that the spectrum of $T_{\alpha,\beta}$ is given by

$$\sigma(T_{\alpha,\beta}) = \begin{cases} \bigcup_{k=1}^{p} \left\{1 + 2\alpha\beta\left(\beta c_k^2 - 1 \pm c_k\sqrt{\beta^2 c_k^2 - 2\beta + 1}\right)\right\} \cup \{1\}, & \text{if } q = p; \\ \bigcup_{k=1}^{p} \left\{1 + 2\alpha\beta\left(\beta c_k^2 - 1 \pm c_k\sqrt{\beta^2 c_k^2 - 2\beta + 1}\right)\right\} \cup \{1\} \cup \{1-2\alpha\beta\}, & \text{if } q > p. \end{cases}$$

Then $\lambda_{k,r} := 1 + 2\alpha\beta\left(\beta c_k^2 - 1 + (-1)^r c_k\sqrt{\beta^2 c_k^2 - 2\beta + 1}\right)$ are eigenvalues of $T_{\alpha,\beta}$, with $r = 1, 2$ and $k = 1, \ldots, p$. Observe that $\lambda_{k,r} \in \mathbb{R}$ if $c_k \geq \beta^{-1}\sqrt{(2\beta-1)^+} =: \widehat{c}_\beta$, while $\lambda_k^r \in \mathbb{C}$ otherwise. To study the modulus of the eigenvalues $\lambda_{k,r}$, consider the function $f_{\alpha,\beta,r} : [0,1] \to \mathbb{R}$ given by

$$f_{\alpha,\beta,r}(c) := \begin{cases} \left(1 + 2\alpha\beta(\beta c^2 - 1)\right)^2 - 4\alpha^2\beta^2 c^2\left(\beta^2 c^2 - 2\beta + 1\right), & \text{if } c < \widehat{c}_\beta; \\ \left(1 + 2\alpha\beta\left(\beta c^2 - 1 + (-1)^r c\sqrt{\beta^2 c^2 - 2\beta + 1}\right)\right)^2, & \text{if } c \geq \widehat{c}_\beta. \end{cases}$$

Hence, one has $|\lambda_{k,r}|^2 = f_{\alpha,\beta,r}(c_k)$.

Let us analyze some properties of the function $f_{\alpha,\beta,r}$. When $\widehat{c}_\beta > 0$, observe that $f_{\alpha,\beta,r}$ is continuous at $\widehat{c}_\beta$, since

$$\lim_{c \to \widehat{c}_\beta^-} f_{\alpha,\beta,r}(c) = \lim_{c \to \widehat{c}_\beta^+} f_{\alpha,\beta,r}(c) = \left(1 - 2\alpha(1-\beta)\right)^2.$$

Define the auxiliary function $g_{\beta,r}(c) := \beta c^2 - 1 + (-1)^r c\sqrt{\beta^2 c^2 - 2\beta + 1}$ for $c \geq \widehat{c}_\beta$. Then,

$$f_{\alpha,\beta,r}(c) = \begin{cases} 4(1-\alpha)\alpha\beta^2 c^2 + (1-2\alpha\beta)^2, & \text{if } c < \widehat{c}_\beta; \\ \left(1 + 2\alpha\beta g_{\beta,r}(c)\right)^2, & \text{if } c \geq \widehat{c}_\beta. \end{cases}$$

The derivative of $f_{\alpha,\beta,r}$ is given for $c \neq \widehat{c}_\beta$ by

$$f'_{\alpha,\beta,r}(c) = \begin{cases} 8(1-\alpha)\alpha\beta^2 c, & \text{if } c < \widehat{c}_\beta; \\ 4\alpha\beta\left(1 + 2\alpha\beta g_{\beta,r}(c)\right)(-1)^r \dfrac{\left(\sqrt{\beta^2c^2-2\beta+1} + (-1)^r\beta c\right)^2}{\sqrt{\beta^2c^2-2\beta+1}}, & \text{if } c > \widehat{c}_\beta. \end{cases}$$

Further, we claim that $1 + 2\alpha\beta g_{\beta,2}(c) > 1 + 2\alpha\beta g_{\beta,1}(c) \geq 0$ for all $c > \widehat{c}_\beta$. Indeed, since

$$(2\beta^2c^2 - 2\beta + 1)^2 = 4\beta^2c^2(\beta^2c^2 - 2\beta + 1) + (2\beta-1)^2 \geq (2\beta c)^2(\beta^2c^2 - 2\beta + 1),$$

we deduce, after taking square roots and reordering, that

$$-1 \leq 2\beta\left(\beta c^2 - 1 - c\sqrt{\beta^2c^2 - 2\beta + 1}\right) = 2\beta g_{\beta,1}(c) < 2\beta g_{\beta,2}(c),$$

from where the assertion easily follows.

All the above properties of the function $f_{\alpha,\beta,r}$ can be summarized as follows:

- For all $0 \leq c < d \leq \widehat{c}_\beta$,

$$F^0_{\alpha,\beta} := (1-2\alpha\beta)^2 \leq f_{\alpha,\beta,r}(c) \leq f_{\alpha,\beta,r}(d) \leq \left(1 - 2\alpha(1-\beta)\right)^2.$$

- For all $\widehat{c}_\beta \leq c < d \leq 1$,

$$F^1_{\alpha,\beta} := (1-4\alpha\beta(1-\beta))^2 \leq f_{\alpha,\beta,1}(d) \leq f_{\alpha,\beta,1}(c)$$
$$\leq (1-2\alpha(1-\beta))^2 \leq f_{\alpha,\beta,2}(c) < f_{\alpha,\beta,2}(d) \leq 1.$$

In view of Fact 2.5, we have to show that the eigenvalue $\lambda = 1$ is semisimple and the only eigenvalue in the unit circle. According to the monotonicity properties of $f_{\alpha,\beta,r}$, we have that $|\lambda_{k,r}| \leq 1$ for all $k = 1,\dots,p$ and $r = 1,2$. Further,

$$|\lambda_{k,r}| = 1 \Leftrightarrow k \in \{1,2,\dots,s\} \text{ and } r = 2,$$

in which case $\lambda_{k,2} = 1$. Thus, we have shown that $\rho(T_{\alpha,\beta}) = 1$ and $\lambda = 1$ is the only eigenvalue in the unit circle.

Let us see now that $\lambda = 1$ is semisimple. First observe that, for any $\lambda \in \mathbb{C}$, given the block diagonal structure of $T_{\alpha,\beta}$, one has

$$\ker(T_{\alpha,\beta} - \lambda I) = \ker\left((T_{\alpha,\beta} - \lambda I)^2\right) \iff \ker(M_{\alpha,\beta} - \lambda I) = \ker\left((M_{\alpha,\beta} - \lambda I)^2\right).$$

Then, we can compute

$$M_{\alpha,\beta} - I = 2\alpha\beta \begin{pmatrix} (2\beta-1)C^2 - I_p & -CS \\ (2\beta-1)CS & -S^2 \end{pmatrix}.$$

Observe that the matrices C and S can be decomposed as

$$C = \begin{pmatrix} I_s & 0 \\ 0 & \widetilde{C} \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} 0_s & 0 \\ 0 & \widetilde{S} \end{pmatrix}, \tag{9}$$

where both $\widetilde{C}$ and $\widetilde{S}$ are diagonal matrices and $\widetilde{S}$ has strictly positive entries. Hence,

$$M_{\alpha,\beta} - I = 2\alpha\beta \begin{pmatrix} -2(1-\beta)I_s & 0 & 0 & 0 \\ 0 & (2\beta-1)\widetilde{C}^2 - I_{p-s} & 0 & -\widetilde{C}\widetilde{S} \\ 0 & 0 & 0 & 0 \\ 0 & (2\beta-1)\widetilde{C}\widetilde{S} & 0 & -\widetilde{S}^2 \end{pmatrix},$$

and one has that $\ker(M_{\alpha,\beta} - I) = \ker\left((M_{\alpha,\beta} - I)^2\right)$ if and only if $\ker(M_0) = \ker\left(M_0^2\right)$, where

$$M_0 := \begin{pmatrix} (2\beta-1)\widetilde{C}^2 - I_{p-s} & -\widetilde{C}\widetilde{S} \\ (2\beta-1)\widetilde{C}\widetilde{S} & -\widetilde{S}^2 \end{pmatrix}.$$

Since $\det(M_0) = \det\left(\widetilde{S}^2\right) \neq 0$ (again, by the block determinant formula), we conclude that $\lambda = 1$ is a semisimple eigenvalue. Then, since $T_{\alpha,\beta}$ is nonexpansive by Fact 2.3, we have by Fact 2.5 that $T_{\alpha,\beta}$ is linearly convergent to $P_{\operatorname{Fix} T_{\alpha,\beta}}$ with any rate $\mu \in]\gamma(T_{\alpha,\beta}), 1[$, and $\operatorname{Fix} T_{\alpha,\beta} = U^\perp \cap V^\perp$ by Proposition 2.1.

Furthermore, we can also deduce from the monotonicity properties of $f_{\alpha,\beta,r}$ that the subdominant eigenvalues of $T_{\alpha,\beta}$ are determined by

$$\gamma(T_{\alpha,\beta}) = \max\{|\lambda_{s+1,2}|, |\lambda_{1,1}|\}$$
$$= \max\left\{\left|1 + 2\alpha\beta\left(\beta c_F^2 - 1 + c_F\sqrt{\beta^2 c_F^2 - 2\beta + 1}\right)\right|, 1 - 4\alpha\beta(1-\beta)\right\}.$$

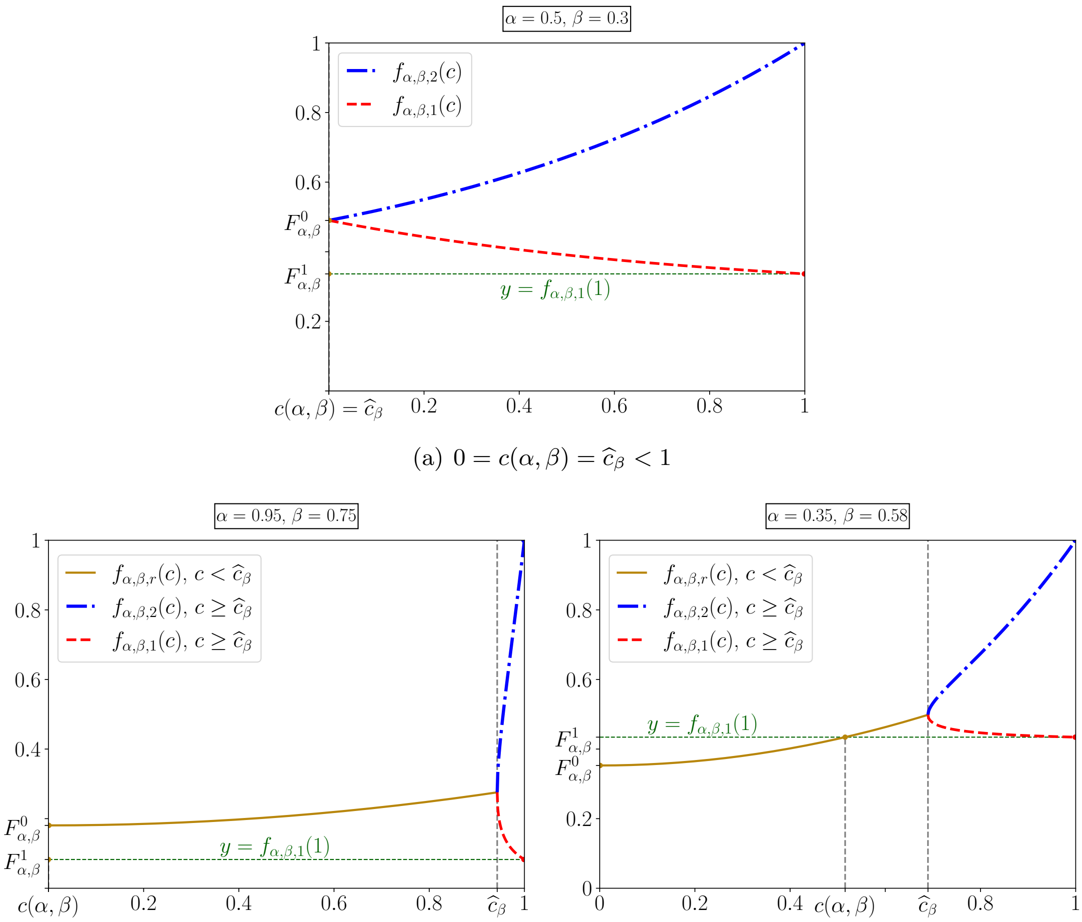
To prove (6), let us compute the value of $\gamma(T_{\alpha,\beta})$. If $c_F > \widehat{c}_\beta$, then $|\lambda_{1,1}| < |\lambda_{s+1,2}|$. Otherwise,

$$|\lambda_{s+1,2}| \le |\lambda_{1,1}| \Leftrightarrow f_{\alpha,\beta,2}(c_F) \le f_{\alpha,\beta,1}(1) \Leftrightarrow 4(1-\alpha)\alpha\beta^2 c_F^2 \le F^1_{\alpha,\beta} - F^0_{\alpha,\beta}.$$

Consequently, if we define

$$c(\alpha,\beta) := \begin{cases} \sqrt{\frac{F^1_{\alpha,\beta}-F^0_{\alpha,\beta}}{4(1-\alpha)\alpha\beta^2}}, & \text{if } F^1_{\alpha,\beta} > F^0_{\alpha,\beta}; \\ 0, & \text{otherwise;} \end{cases}$$

which is equivalent to the expression in (7), we obtain (6). Three possible scenarios for $f_{\alpha,\beta,r}$ and the constants $c(\alpha,\beta)$ and $\widehat{c}_\beta$ depending on the values of α and β are shown in Figure 1.



(a) $0 = c(\alpha,\beta) = \widehat{c}_\beta < 1$

(b) $0 = c(\alpha,\beta) < \widehat{c}_\beta < 1$

(c) $0 < c(\alpha,\beta) < \widehat{c}_\beta < 1$

Figure 1: The three possible scenarios for the function $f_{\alpha,\beta,r}(c)$

To conclude the proof, let us see that the subdominant eigenvalues are semisimple if and only if $\beta^2 c_F^2 - 2\beta + 1 \ne 0$ or $c_F = 0$. The candidate eigenvalues to be subdominant are $\lambda_{1,1}$ and $\lambda_{s+1,2}$, possibly simultaneously.

Consider first the case where $\lambda_{1,1} = 1 - 4\alpha\beta(1-\beta)$ is subdominant, and compute

$$M_{\alpha,\beta} - \lambda_{1,1} I = 2\alpha\beta \begin{pmatrix} -(2\beta-1)S^2 & -CS \\ (2\beta-1)CS & C^2 - (2\beta-1)I_p \end{pmatrix}.$$

Using the decomposition of C and S given in (9), we get

$$M_{\alpha,\beta} - \lambda_{1,1} I = 2\alpha\beta \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -(2\beta-1)\widetilde{S}^2 & 0 & -\widetilde{C}\widetilde{S} \\ 0 & 0 & 2(1-\beta)I_s & 0 \\ 0 & (2\beta-1)\widetilde{C}\widetilde{S} & 0 & \widetilde{C}^2 - (2\beta-1)I_{p-s} \end{pmatrix},$$

and one has that $\ker(M_{\alpha,\beta} - \lambda_{1,1} I) = \ker\left((M_{\alpha,\beta} - \lambda_{1,1} I)^2\right)$ if and only if $\ker(M_1) = \ker\left(M_1^2\right)$, where

$$M_1 := \begin{pmatrix} -(2\beta-1)\widetilde{S}^2 & -\widetilde{C}\widetilde{S} \\ (2\beta-1)\widetilde{C}\widetilde{S} & \widetilde{C}^2 - (2\beta-1)I_{p-s} \end{pmatrix}.$$

Since we are assuming that $\lambda_{1,1}$ is subdominant, it necessarily holds that $\frac{1}{2} < \beta < 1$. Hence,

$$\det(M_1) = \det\left((2\beta-1)^2 \widetilde{S}^2\right) \neq 0,$$

and one trivially has that $\ker(M_1) = \ker\left(M_1^2\right)$, which proves that $\lambda_{1,1}$ is semisimple.

Consider now the case where $\lambda_{s+1,2} = 1 + 2\alpha\beta\left(\beta c_F^2 - 1 + c_F\sqrt{\beta^2 c_F^2 - 2\beta + 1}\right)$ is a subdominant eigenvalue. Denote by $\Delta_F := \sqrt{\beta^2 c_F^2 - 2\beta + 1}$ and compute

$$M_{\alpha,\beta} - \lambda_{s+1,2} I = 2\alpha\beta \begin{pmatrix} (2\beta-1)C^2 - c_F(\beta c_F + \Delta_F)I_p & -CS \\ (2\beta-1)CS & C^2 - c_F(\beta c_F + \Delta_F)I_p \end{pmatrix}.$$

Let $k \in \{1, \ldots, p-s\}$ be such that $c_F = c_{s+1} = c_{s+2} = \cdots = c_{s+k} > c_{s+k+1}$. Then

$$C = \begin{pmatrix} I_s & 0 & 0 \\ 0 & c_F I_k & 0 \\ 0 & 0 & \widetilde{C} \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} 0_s & 0 & 0 \\ 0 & s_F I_k & 0 \\ 0 & 0 & \widetilde{S} \end{pmatrix},$$

where both $\widetilde{C}$ and $\widetilde{S}$ are diagonal matrices and $\widetilde{C}$ has entries strictly smaller than c_F. Hence, one has

$$M_F := M_{\alpha,\beta} - \lambda_{s+1,2} I = 2\alpha\beta \begin{pmatrix} m_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & m_2 & 0 & 0 & m_{25} & 0 \\ 0 & 0 & m_3 & 0 & 0 & m_{36} \\ 0 & 0 & 0 & m_4 & 0 & 0 \\ 0 & m_{52} & 0 & 0 & m_5 & 0 \\ 0 & 0 & m_{63} & 0 & 0 & m_6 \end{pmatrix},$$

where $m_1 := (2\beta - 1 - c_F(\beta c_F + \Delta_F))I_s$, $m_2 := -c_F(\Delta_F + (1-\beta)c_F)I_k$, $m_3 := (2\beta-1)\widetilde{C}^2 - c_F(\beta c_F + \Delta_F)I_{p-k-s}$, $m_4 := (1 - c_F(\beta c_F + \Delta_F))I_s$, $m_5 := -c_F(\Delta_F - (1-\beta)c_F)I_k$, $m_6 := \widetilde{C}^2 - c_F(\beta c_F + \Delta_F)I_{p-k-s}$, $m_{25} := -c_F s_F I_k$, $m_{36} := -\widetilde{C}\widetilde{S}$, $m_{52} := (2\beta-1)c_F s_F I_k$ and $m_{63} := (2\beta-1)\widetilde{C}\widetilde{S}$. Thus, if we denote by $M_{\{2,5\}} := \begin{pmatrix} m_2 & m_{25} \\ m_{52} & m_5 \end{pmatrix}$ and by $M_{\{3,6\}} := \begin{pmatrix} m_3 & m_{36} \\ m_{63} & m_6 \end{pmatrix}$, we get that

$$\ker(M_F) = \ker\left(M_F^2\right) \iff \ker\left(M_{\{2,5\}}\right) = \ker\left(M_{\{2,5\}}^2\right) \text{ and } \ker\left(M_{\{3,6\}}\right) = \ker\left(M_{\{3,6\}}^2\right).$$

On the one hand, by the block determinant formula we have that

$$\begin{aligned}\det\left(M_{\{3,6\}}\right) &= \det(m_3 m_6 - m_{63}m_{36})\\ &= \det\left((2\beta-1)\widetilde{C}^4 + c_F^2(\beta c_F+\Delta_F)^2 I_{p-k-s} - 2\beta c_F(\beta c_F+\Delta_F)\widetilde{C}^2 + (2\beta-1)\widetilde{C}^2\widetilde{S}^2\right)\\ &= \det\left((2\beta-1-2\beta c_F(\beta c_F+\Delta_F))\widetilde{C}^2 + c_F^2(\beta c_F+\Delta_F)^2 I_{p-k-s}\right)\\ &= \det\left((-(\beta^2 c_F^2-2\beta+1)-\beta^2 c_F^2 - 2\beta c_F\Delta_F)\widetilde{C}^2 + c_F^2(\beta c_F+\Delta_F)^2 I_{p-k-s}\right)\\ &= \det\left(-\left(\beta c_F+\Delta_F\right)^2\left(\widetilde{C}^2 - c_F^2 I_{p-s-k}\right)\right).\end{aligned}$$

Observe that $\beta c_F + \Delta_F = 0$ if and only if $\beta = \frac{1}{2}$ and $c_F = 0$, in which case $M_{\{3,6\}} = 0_{2p-k-s}$. If $\beta c_F + \Delta_F \neq 0$, then $\det\left(M_{\{3,6\}}\right) \neq 0$. Thus, in either case, we get $\ker\left(M_{\{3,6\}}\right) = \ker\left(M_{\{3,6\}}^2\right)$.

On the other hand, we can rewrite

$$M_{\{2,5\}} = -c_F\begin{pmatrix}(\Delta_F+(1-\beta)c_F)I_k & s_F I_k\\ -(2\beta-1)s_F I_k & (\Delta_F-(1-\beta)c_F)I_k\end{pmatrix},$$

and one has

$$M_{\{2,5\}}^2 = c_F^2\begin{pmatrix}\left((\Delta_F+(1-\beta)c_F)^2-(2\beta-1)s_F^2\right)I_k & 2\Delta_F s_F I_k\\ -2\Delta_F(2\beta-1)s_F I_k & \left((\Delta_F-(1-\beta)c_F)^2-(2\beta-1)s_F^2\right)I_k\end{pmatrix}.$$

Observing that

$$\begin{aligned}(\Delta_F-(1-\beta)c_F)^2-(2\beta-1)s_F^2 &= 2\Delta_F\left(\Delta_F-(1-\beta)c_F\right),\\ (\Delta_F+(1-\beta)c_F)^2-(2\beta-1)s_F^2 &= 2\Delta_F\left(\Delta_F+(1-\beta)c_F\right),\end{aligned}$$

we deduce that $M_{\{2,5\}}^2 = -2\Delta_F c_F M_{\{2,5\}}$. If $c_F = 0$, then $M_{\{2,5\}} = 0_{2k}$. Therefore, $\ker\left(M_{\{2,5\}}\right) = \ker\left(M_{\{2,5\}}^2\right)$ if and only if $\Delta_F \neq 0$ or $c_F = 0$.

Summarizing the discussion above, we have shown that

$$\ker\left(M_F\right) = \ker\left(M_F^2\right) \iff \Delta_F \neq 0 \text{ or } c_F = 0.$$

Finally, observe that $\Delta_F = 0$ if and only if $c_F = \widehat{c}_\beta$, in which case $\lambda_{s+1,1}$ is a subdominant eigenvalue, and this proves the last assertion in the statement.

Case 2: $p+q \geq n$. We can take some $k \geq 1$ such that $n' := n+k > p+q$, and consider $U' := U \times \{0_{k\times 1}\} \subset \mathbb{R}^{n'}$, $V' := V \times \{0_{k\times 1}\} \subset \mathbb{R}^{n'}$, and $T'_{\alpha,\beta} := T_{U',V',\alpha,\beta} = (1-\alpha)I + \alpha(2\beta P_{V'} - I)(2\beta P_{U'} - I)$. Since $P_{U'} = \begin{pmatrix}P_U & 0\\ 0 & 0_k\end{pmatrix}$ and $P_{V'} = \begin{pmatrix}P_V & 0\\ 0 & 0_k\end{pmatrix}$, it holds that

$$T'_{\alpha,\beta} = \begin{pmatrix}T_{\alpha,\beta} & 0\\ 0 & I_k\end{pmatrix}.$$

Therefore, $\sigma(T_{\alpha,\beta}) \cup \{1\} = \sigma(T'_{\alpha,\beta})$ and $\gamma(T_{\alpha,\beta}) = \gamma(T'_{\alpha,\beta})$. Note that the principal angles between U' and V' are the same that the ones between U and V. Hence, the result follows from applying Case 1 to $T'_{\alpha,\beta}$. □

Remark 3.1. *For simplicity, we have assumed that* $\dim U = p \leq q = \dim V$. *If this is not the case and* $q < p$, *observe that one has to exchange the matrix decomposition of* P_U *and* P_V *given in* (3). *In this case, one can check that the matrix* $T_{\alpha,\beta}$ *obtained corresponds to the transpose of the one given in* (8). *Hence, the spectrum of* $T_{\alpha,\beta}$ *remains the same and thus all the results in Theorem 3.1 also hold.*

Remark 3.2. *The expression in* (6) *corroborates what it was numerically observed in [2]: there are values of α and β for which the rate of convergence of AAMR does not depend on the value of the Friedrichs angle for all angles larger than* $\arccos c(\alpha,\beta)$.

We now look for the values of the parameters α and β and V, in order to that minimize the rate of convergence of the AAMR method obtained in Theorem 3.1.

Theorem 3.2. *The infimum of the linear convergence rates of the AAMR operator $T_{\alpha,\beta}$ attains its smallest value at $\alpha^\star = 1$ and $\beta^\star = \frac{1}{1+\sin\theta_F}$, where θ_F is the Friedrichs angle between U and V; i.e., it holds*

$$\frac{1-\sin\theta_F}{1+\sin\theta_F} = \gamma\left(T_{1,\beta^\star}\right) \leq \gamma\left(T_{\alpha,\beta}\right) \quad \text{for all } (\alpha,\beta)\in\,]0,1]\times\,]0,1[\,.$$

Furthermore, $\gamma\left(T_{1,\beta^\star}\right)$ is an optimal linear convergence rate if and only if $\theta_F = \frac{\pi}{2}$.

Proof. Let us look for the values of parameters α and β that minimize the rate $\gamma(T_{\alpha,\beta})$ given by (6). Define the sets $D :=]0,1]\times]0,1[$,

$$\begin{aligned}
D_1 &:= \left\{(\alpha,\beta)\in D : \beta < \frac{1}{1+s_F}\right\},\\
D_2 &:= \left\{(\alpha,\beta)\in D : \frac{1}{1+s_F} \leq \beta \leq \frac{1}{1+s_F^2} \text{ or } \alpha \geq \frac{1-\beta(1+s_F^2)}{\beta\left(4(1-\beta)^2-s_F^2\right)}, \beta > \frac{1}{1+s_F^2}\right\},\\
D_3 &:= \left\{(\alpha,\beta)\in D : \alpha < \frac{1-\beta(1+s_F^2)}{\beta\left(4(1-\beta)^2-s_F^2\right)}, \beta > \frac{1}{1+s_F^2}\right\},
\end{aligned}$$

and the functions

$$\begin{aligned}
\Gamma_1(\alpha,\beta) &:= 1+2\alpha\beta\left(\beta c_F^2 - 1 + c_F\sqrt{\beta^2c_F^2-2\beta+1}\right), \quad \text{for } (\alpha,\beta)\in D_1,\\
\Gamma_2(\alpha,\beta) &:= \sqrt{4(1-\alpha)\alpha\beta^2c_F^2+(1-2\alpha\beta)^2}, \quad \text{for } (\alpha,\beta)\in D,\\
\Gamma_3(\alpha,\beta) &:= 1-4\alpha\beta(1-\beta), \quad \text{for } (\alpha,\beta)\in D,
\end{aligned}$$

having $D = D_1\cup D_2\cup D_3$. Hence, we can define the convergence rate in terms of the parameters α and β through the function

$$\Gamma(\alpha,\beta) := \gamma(T_{\alpha,\beta}) = \begin{cases} \Gamma_1(\alpha,\beta), & \text{if } (\alpha,\beta)\in D_1,\\ \Gamma_2(\alpha,\beta), & \text{if } (\alpha,\beta)\in D_2,\\ \Gamma_3(\alpha,\beta), & \text{if } (\alpha,\beta)\in D_3,\end{cases}$$

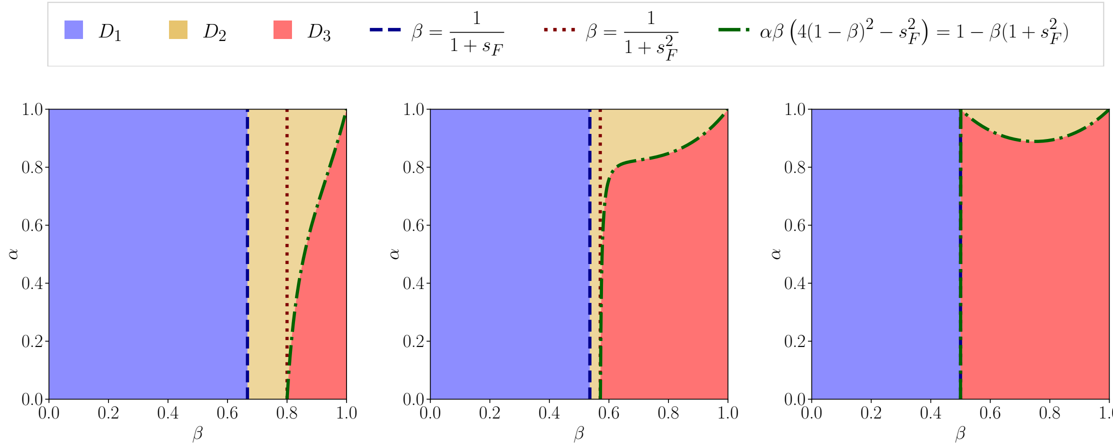
see Figure 2.

The function Γ is piecewise defined, continuous and differentiable on the interior of each of the three regions D_1, D_2 and D_3, but is not differentiable on the boundaries. Let us analyze the three problems of minimizing the function Γ over the closure of each of the three pieces. The gradient of the functions Γ_1, Γ_2 and Γ_3 are given by

$$\nabla\Gamma_1(\alpha,\beta) = \begin{pmatrix} 2\beta\left(\beta c_F^2-1+c_F\sqrt{\beta^2c_F^2-2\beta+1}\right)\\ 2\alpha\left(\beta c_F^2-1+c_F\sqrt{\beta^2c_F^2-2\beta+1}\right)\left(\frac{\beta c_F+\sqrt{\beta^2c_F^2-2\beta+1}}{\sqrt{\beta^2c_F^2-2\beta+1}}\right)\end{pmatrix},$$

$$\nabla\Gamma_2(\alpha,\beta) = \frac{1}{\sqrt{4(1-\alpha)\alpha\beta^2c_F^2+(1-2\alpha\beta)^2}}\begin{pmatrix}2\beta\left(\beta c_F^2-1+2\alpha\beta(1-c_F^2)\right)\\ 2\alpha\left(2\beta c_F^2-1+2\alpha\beta(1-c_F^2)\right)\end{pmatrix},$$

$$\nabla\Gamma_3(\alpha,\beta) = \begin{pmatrix}-4\beta\left(1-\beta\right)\\ -4\alpha\left(1-2\beta\right)\end{pmatrix}.$$



(a) Friedrichs angle: $\frac{\pi}{6}$ radians (b) Friedrichs angle: $\frac{\pi}{3}$ radians (c) Friedrichs angle: $\frac{\pi}{2}$ radians

Figure 2: Piecewise domain of the function $\Gamma(\alpha, \beta)$ for three different values of the Friedrichs angle

To minimize Γ over $\overline{D_1}$, we assert that

$$\frac{\partial \Gamma_1}{\partial \alpha}(\alpha, \beta) < 0, \frac{\partial \Gamma_1}{\partial \beta}(\alpha, \beta) < 0, \quad \text{for all } (\alpha, \beta) \in D_1. \tag{10}$$

Indeed, on the one hand, since $c_F^2 < 1$, then

$$c_F^2\left(\beta^2 c_F^2 - 2\beta + 1\right) = \beta^2 c_F^4 - 2\beta c_F^2 + c_F^2 < \beta^2 c_F^4 - 2\beta c_F^2 + 1 = \left(1 - \beta c_F^2\right)^2.$$

Thus, taking square roots and reordering, we get that $\beta c_F^2 - 1 + c_F\sqrt{\beta^2 c_F^2 - 2\beta + 1} < 0$. On the other hand, one has $\beta c_F + \sqrt{\beta^2 c_F^2 - 2\beta + 1} \geq 0$, with equality if and only if $c_F = 0$ and $\beta = \frac{1}{2}$. In this case, the point has the form $\left(\alpha, \frac{1}{2}\right) \notin D_1$ for $\alpha \in]0, 1]$. We have therefore shown that (10) holds, and thus the unique minimum of Γ over $\overline{D_1}$ is attained at $\left(1, \frac{1}{1+s_F}\right)$.

Let us consider now the problem of minimizing Γ over $\overline{D_2}$. To address this problem, we consider two cases. Suppose first that $c_F = 0$ and observe that

$$\Gamma_2(\alpha, \beta) = \sqrt{(1 - 2\alpha\beta)^2} \geq 0, \quad \text{for all } (\alpha, \beta) \in \overline{D_2},$$

having $\Gamma_2(\alpha, \beta) = 0$ if and only if $2\alpha\beta = 1$. Since $\left(1, \frac{1}{2}\right)$ is the only point in $\overline{D_2}$ satisfying this equation, we deduce that it is the unique minimum.

Suppose now that $c_F > 0$. In this case, we claim that Γ_2 attains its minimimum over the region $\overline{D_2 \cup D_3} = [0, 1] \times \left[\frac{1}{1+s_F}, 1\right]$ at the point $\left(1, \frac{1}{1+s_F}\right) \in D_2$, and so does Γ over $\overline{D_2}$. Indeed, observe that Γ_2 is smooth on the interior of the set $D_2 \cup D_3$. Moreover, $\nabla\Gamma_2$ only vanishes at $(0, 0)$. Therefore, the minimum has to be attained at some point in the boundary. Note that, for all $\beta \in \left[\frac{1}{1+s_F}, 1\right]$, the following holds:

(i) $\Gamma_2(0, \beta) = 1$.
(ii) $\Gamma_2(1, \beta) = 2\beta - 1$, which attains its minimum at $\beta = \frac{1}{1+s_F}$.
(iii) The function $\alpha \mapsto \Gamma_2(\alpha, \beta)$ is the square root of a positive non-degenerated convex parabola,

$$\alpha \mapsto \Gamma_2(\alpha, \beta) = \sqrt{4\beta^2 s_F^2 \alpha^2 - 4\beta(1 - \beta c_F^2)\alpha + 1};$$

which attains its minimum at $\alpha^\star(\beta) := \frac{1 - \beta c_F^2}{2\beta s_F^2}$. Since $\alpha^\star\left(\frac{1}{1+s_F}\right) = \frac{1+s_F}{2s_F} \geq 1$, we have

$$\Gamma_2\left(1, \frac{1}{1+s_F}\right) < \Gamma_2\left(\alpha, \frac{1}{1+s_F}\right), \quad \text{for all } \alpha \in [0, 1].$$

On the other hand, $\alpha^\star(1) = \frac{1}{2}$, which implies

$$\Gamma_2\left(\frac{1}{2}, 1\right) < \Gamma_2(\alpha, 1), \qquad \text{for all } \alpha \in [0,1].$$

Then, noting that

$$\Gamma_2\left(1, \frac{1}{1+s_F}\right) = \frac{1-s_F}{1+s_F} < c_F = \Gamma_2\left(\frac{1}{2}, 1\right),$$

we have shown by (i)–(iii) that Γ_2 attains its minimum over $\overline{D_2 \cup D_3}$ at $\left(1, \frac{1}{1+s_F}\right) \in D_2$, as claimed.

Finally, observe that if $(\alpha, \beta) \in D_3$, it holds that $\frac{1}{2} < \beta < 1$. Then

$$\frac{\partial \Gamma_3}{\partial \alpha}(\alpha, \beta) < 0, \frac{\partial \Gamma_3}{\partial \beta}(\alpha, \beta) > 0, \quad \text{for all } (\alpha, \beta) \in D_3.$$

Thus, there exists some point $(\alpha_3^\star, \beta_3^\star) \in \overline{D_3}$ with $\alpha_3^\star \beta_3^\star \left(4(1-\beta_3^\star)^2 - s_F^2\right) = 1 - \beta_3^\star(1+s_F^2)$ such that

$$\Gamma_3(\alpha_3^\star, \beta_3^\star) < \Gamma_3(\alpha, \beta), \qquad \text{for all } (\alpha, \beta) \in D_3.$$

Note that $(\alpha_3^\star, \beta_3^\star)$ lies on the boundary curve between D_2 and D_3. Since Γ is continuous on D, it holds that $\Gamma_2(\alpha_3^\star, \beta_3^\star) = \Gamma_3(\alpha_3^\star, \beta_3^\star)$ and hence,

$$\Gamma\left(1, \frac{1}{1+s_F}\right) \leq \Gamma(\alpha_3^\star, \beta_3^\star) < \Gamma(\alpha, \beta), \text{ for all } (\alpha, \beta) \in D_3.$$

Hence, all the reasoning above proves that

$$\operatorname*{argmin}_{(\alpha,\beta)\in D} \Gamma(\alpha, \beta) = \left(1, \frac{1}{1+s_F}\right),$$

with $\Gamma\left(1, \frac{1}{1+s_F}\right) = \frac{1-s_F}{1+s_F}$. Finally, by the last assertion in Theorem 3.1, $\gamma\left(T_{1,\frac{1}{1+s_F}}\right)$ is an optimal linear convergence rate if and only if $c_F = 0$, as claimed. ☐

Corollary 3.1. *Let $\alpha \in]0,1]$ and $\beta \in]0,1[$. Given $z \in \mathbb{R}^n$, choose any $x_0 \in \mathbb{R}^n$ and consider the sequence generated, for $k = 0, 1, 2, \ldots$, by*

$$x_{k+1} = T_{U-z,V-z,\alpha,\beta}(x_k) = (1-\alpha)x_k + \alpha(2\beta P_{V-z} - I)(2\beta P_{U-z} - I)(x_k).$$

Let $\gamma(T_{\alpha,\beta})$ be given by (6). Then, for every $\mu \in]\gamma(T_{\alpha,\beta}), 1[$, the sequence $(x_k)_{k\geq 0}$ is R-linearly convergent to $P_{\operatorname{Fix} T_{U-z,V-z,\alpha,\beta}}(x_0)$ and the shadow sequence $(P_U(z+x_k))_{k\geq 0}$ is R-linearly convergent to $P_{U\cap V}(z)$, both with rate μ, in the sense that there exists a positive integer k_0 such that

$$\|P_U(z+x_k) - P_{U\cap V}(z)\| \leq \left\|x_k - P_{\operatorname{Fix} T_{U-z,V-z,\alpha,\beta}}(x_0)\right\| \leq \mu^k, \quad \textit{for all } k \geq k_0. \tag{11}$$

Proof. According to Fact 2.4 we have that $\operatorname{Fix} T_{U-z,V-z,\alpha,\beta} \neq \emptyset$ and

$$x_{k+1} = T_{U-z,V-z,\alpha,\beta}(x_k) = T_{U,V,\alpha,\beta}(x_k - x^*) + x^*,$$

for $x^* := P_{\operatorname{Fix} T_{U-z,V-z,\alpha,\beta}}(x_0)$. Hence, one has

$$\|x_k - x^*\| = \|T_{U,V,\alpha,\beta}(x_{k-1} - x^*)\| = \cdots = \|T^k_{U,V,\alpha,\beta}(x_0 - x^*)\|.$$

Again by Fact 2.4, one has

$$\operatorname{Fix} T_{U-z,V-z,\alpha,\beta} = x^* + U^\perp \cap V^\perp,$$

and by the translation formula for projections (c.f. [11, 2.7(ii)]),

$$x^* = P_{\operatorname{Fix} T_{U-z,V-z,\alpha,\beta}}(x_0) = P_{x^*+U^\perp\cap V^\perp}(x_0) = P_{U^\perp\cap V^\perp}(x_0 - x^*) + x^*,$$

which implies $P_{U^\perp\cap V^\perp}(x_0 - x^*) = 0$, and therefore,

$$\|x_k - x^*\| = \left\|T^k_{U,V,\alpha,\beta}(x_0 - x^*)\right\| = \left\|\left(T^k_{U,V,\alpha,\beta} - P_{U^\perp\cap V^\perp}\right)(x_0 - x^*)\right\|.$$

Let $\nu \in]\gamma(T_{\alpha,\beta}), \mu[$. Since $\nu > \gamma(T_{\alpha,\beta})$, by Theorem 3.1, there exists a positive integer k_1 and some $M > 0$ such that

$$\left\|T^k_{U,V,\alpha,\beta} - P_{U^\perp\cap V^\perp}\right\| \leq M\nu^k, \quad \text{for all } k \geq k_1.$$

Let $k_0 \geq k_1$ be a positive integer such that

$$\left(\frac{\mu}{\nu}\right)^k \geq M\|x_0 - x^*\|, \quad \text{for all } k \geq k_0.$$

Then, we deduce that

$$\|x_k - x^*\| \leq \left\|T^k_{U,V,\alpha,\beta} - P_{U^\perp\cap V^\perp}\right\| \|x_0 - x^*\| \leq M\nu^k\|x_0 - x^*\| \leq \mu^k,$$

for all $k \geq k_0$, which proves the second inequality in (11).

By [2, Proposition 3.4] and the translation formula for projections (c.f. [11, 2.7(ii)]), we can deduce that

$$P_U\left(z + P_{\operatorname{Fix} T_{U-z,V-z,\alpha,\beta}}(x_0)\right) = P_{U\cap V}(z).$$

Thus, the first inequality in (11) is a consequence of this, and the linearity and nonexpansiveness of P_U. Indeed

$$\begin{aligned}\|P_U(z + x_k) - P_{U\cap V}(z)\| &= \|P_U(z + x_k) - P_U(z + P_{\operatorname{Fix} T_{U-z,V-z,\alpha,\beta}}(x_0))\| \\ &= \|P_U(x_k - P_{\operatorname{Fix} T_{U-z,V-z,\alpha,\beta}}(x_0))\| \\ &\leq \|x_k - P_{\operatorname{Fix} T_{U-z,V-z,\alpha,\beta}}(x_0)\|,\end{aligned}$$

which completes the proof. □

4 Comparison with other projection methods

In this section, we compare the rate of AAMR with optimal parameters obtained in Section 3 with the rates of various projection methods analyzed in [5, 16]. We summarize the key features of these schemes in Table 1, where we recall the operator defining the iteration of each method, as well as the optimal parameters and rates of convergence when these schemes are applied to linear subspaces. Note that all these rates only depend on the Friedrichs angle θ_F between the subspaces.

On the one hand, we observe that the rates for AAMR and GAP coincide. Moreover, their optimal parameters are closely related, in the sense that

$$\alpha^\star_{AAMR} = \alpha^\star_{GAP} \quad \text{and} \quad \alpha^\star_{1,GAP} = \alpha^\star_{2,GAP} = 2\beta^\star_{AAMR}.$$

We explain this behavior in Section 4.1, where under some conditions, we show that the shadow sequences of GAP and AAMR coincide for linear subspaces (Theorem 4.1). On the other hand, we note that the rate for AAMR/GAP is considerably smaller than the one of other methods, see Figure 3. We numerically demonstrate this with a computational experiment in Section 5.

Method	Optimal parameter(s)	Rate
Alternating Projections $AP = P_V P_U$	–	$\cos^2\theta_F$
Relaxed Alternating Projections $RAP = (1-\alpha)I + \alpha P_V P_U$	$\alpha^\star = \frac{2}{1+\sin^2\theta_F}$	$\frac{1-\sin^2\theta_F}{1+\sin^2\theta_F}$
Generalized Alternating Projections $GAP = (1-\alpha)I + \alpha\left(\alpha_1 P_V + (1-\alpha_1)I\right)\left(\alpha_2 P_U + (1-\alpha_2)I\right)$	$\alpha^\star = 1$ $\alpha_1^\star = \alpha_2^\star = \frac{2}{1+\sin\theta_F}$	$\frac{1-\sin\theta_F}{1+\sin\theta_F}$
Douglas–Rachford $DR = (1-\alpha)I + \alpha(2P_V - I)(2P_U - I)$	$\alpha^\star = \frac{1}{2}$	$\cos\theta_F$
Averaged Alternating Modified Reflections $AAMR = (1-\alpha)I + \alpha(2\beta P_V - I)(2\beta P_U - I)$	$\alpha^\star = 1,\ \beta^\star = \frac{1}{1+\sin\theta_F}$	$\frac{1-\sin\theta_F}{1+\sin\theta_F}$

Table 1: Rates of convergence with optimal parameters of AP, RAP, GAP, DR and AAMR when they are applied to two subspaces

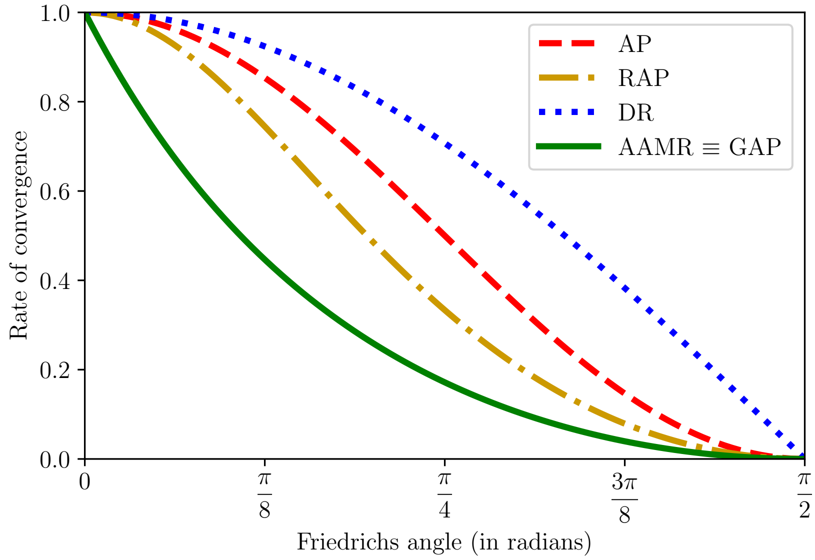


Figure 3: Comparison of the rates of linear convergence with optimal parameters of AP, RAP, DR, AAMR and GAP

4.1 Relationship between AAMR and GAP for subspaces

Fält and Giselsson have recently obtained in [16] the rate of convergence with optimal parameters for the generalized alternating projections (GAP) method for two subspaces. This iterative scheme is defined by

$$z_{k+1} = (1-\alpha)z_k + \alpha\left(\alpha_1 P_V + (1-\alpha_1)I\right)\left(\alpha_2 P_U + (1-\alpha_2)I\right)(z_k), \tag{12}$$

where $\alpha \in]0,1]$ and $\alpha_1, \alpha_2 \in]0,2]$.

The next result shows that, for subspaces, the shadow sequences of GAP and AAMR coincide when $\alpha_1 = \alpha_2 = 2\beta$ and the starting point of AAMR is chosen as $x_0 = 0$; see Figure 4 for a simple example in $\mathbb{R}^2$. This is not the case for general convex sets, as shown in Figure 5: GAP gives a point in the intersection of the sets, while AAMR solves the best approximation problem (1). Figures 4 and 5 were created with `Cinderella` [9].

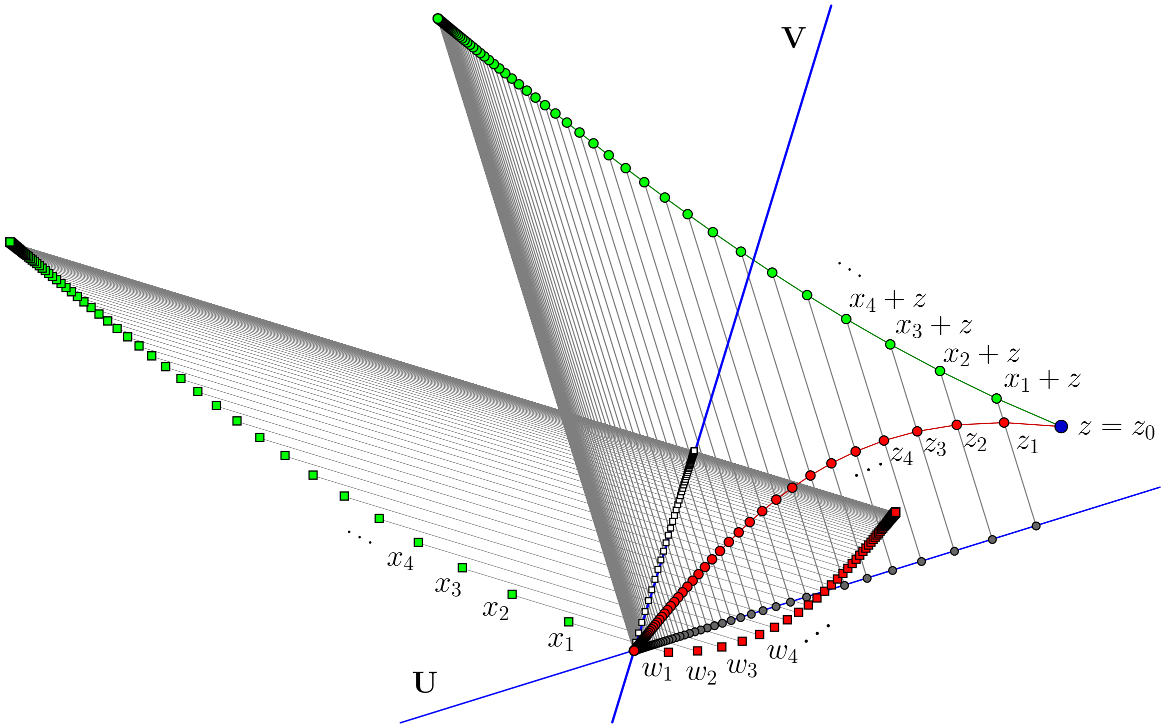


Figure 4: Graphical representation of Theorem 4.1 for two lines in $\mathbb{R}^2$. The sequence $\{x_k + z\}_{k=0}^{\infty}$ is generated by AAMR with $x_0 = 0$, while the sequence $\{z_k\}_{k=0}^{\infty}$ is generated by GAP. We also represent $w_k := (2\beta - 1)(z_k - z)$

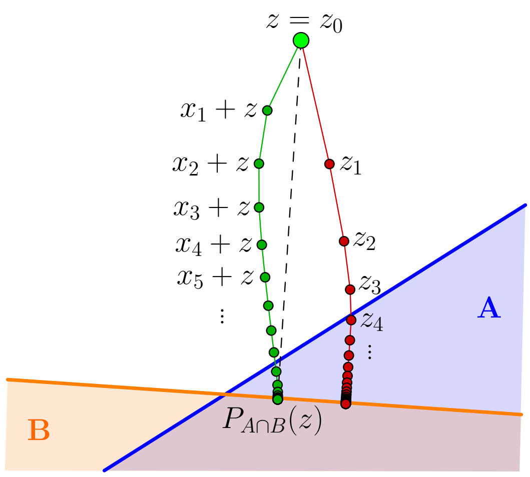


Figure 5: The sequence $\{x_k + z\}_{k=0}^{\infty}$, with $x_0 = 0$, generated by AAMR converges to $P_{A\cap B}(z)$ and thus solves the best approximation problem, while the sequence $\{z_k\}_{k=0}^{\infty}$, with $z_0 = z$, generated by GAP only converges to some point in $A \cap B$

Theorem 4.1. *Given $z \in \mathbb{R}^n$, set $x_0 = 0$ and consider the AAMR sequence $\{x_k\}_{k=0}^{\infty}$ generated by (2) with parameters $\alpha \in]0, 1]$ and $\beta \in]0, 1[$. Let $\{z_k\}_{k=0}^{\infty}$ be the sequence generated by GAP (12) with parameters α and $\alpha_1 = \alpha_2 = 2\beta$ and starting point $z_0 = z$. Then, one has*

$$P_U(x_k + z) = P_U(z_k) \quad \textit{and} \quad P_V(x_k) = (2\beta - 1)P_V(z_k - z), \tag{13}$$

for all $k = 0, 1, 2, \ldots$.

Proof. To simplify the notation, let $\eta := 2\beta$. We shall prove (13) by induction. Since both equalities clearly hold for $k = 0$, we can assume that they are valid for some $k \geq 0$. By (5), the sequence generated by the AAMR scheme satisfies

$$\begin{aligned}P_U(x_{k+1}) &= P_U(x_k) + \alpha\eta\big(\eta P_U P_V P_U(x_k+z) \\&\quad + (1-\eta)P_U P_V(z) - P_U P_V(x_k) - P_U(x_k+z)\big)\\&= \left(\alpha\eta^2 P_U P_V P_U - \alpha\eta P_U P_V + (1-\alpha\eta)P_U\right)(x_k)\\&\quad + \alpha\left(\eta^2 P_U P_V P_U + \eta(1-\eta)P_U P_V - \eta P_U\right)(z),\end{aligned} \tag{14}$$

and,

$$\begin{aligned}P_V(x_{k+1}) &= P_V(x_k) + \alpha\eta\left((\eta-1)P_V P_U(x_k+z) + P_V((1-\eta)z - x_k)\right)\\&= \alpha\eta(\eta-1)P_V P_U(x_k+z) + P_V\left(\alpha\eta(1-\eta)z + (1-\alpha\eta)x_k\right).\end{aligned} \tag{15}$$

Thanks to the linearity of the projectors onto subspaces and using $\alpha_1 = \alpha_2 = \eta$, the GAP iteration (12) takes the form

$$z_{k+1} = (1-\alpha\eta(2-\eta))\, z_k + \alpha\left(\eta^2 P_V P_U z_k + \eta(1-\eta)P_V z_k + \eta(1-\eta)P_U z_k\right);$$

and thus this scheme verifies

$$P_U(z_{k+1}) = \left(\alpha\eta^2 P_U P_V P_U + \alpha\eta(1-\eta)P_U P_V + (1-\alpha\eta)P_U\right)(z_k), \tag{16}$$

and

$$P_V(z_{k+1}) = (1-\alpha\eta)P_V(z_k) + \alpha\eta P_V P_U(z_k). \tag{17}$$

Then, by (15), the induction hypothesis (13) and (17), we obtain

$$\begin{aligned}P_V(x_{k+1}) &= \alpha\eta(\eta-1)P_V P_U(z_k) + (1-\alpha\eta)P_V(x_k) + P_V(\alpha\eta(1-\eta)z)\\&= \alpha\eta(\eta-1)P_V P_U(z_k) + (1-\alpha\eta)(\eta-1)P_V(z_k - z) + P_V(\alpha\eta(1-\eta)z)\\&= (\eta-1)\left(\alpha\eta P_V P_U(z_k) + (1-\alpha\eta)P_V(z_k)\right) + (1-\eta)P_V(z)\\&= (\eta-1)P_V(z_{k+1} - z),\end{aligned}$$

which proves the second equation in (13) for $k+1$. Finally, by (16), (13) and (14), we have that

$$\begin{aligned}P_U(z_{k+1}) &= \alpha\eta^2 P_U P_V P_U(x_k+z) + \alpha\eta P_U P_V(-x_k + (1-\eta)z) + (1-\alpha\eta)P_U(x_k+z)\\&= \left(\alpha\eta^2 P_U P_V P_U - \alpha\eta P_U P_V + (1-\alpha\eta)P_U\right)(x_k)\\&\quad + \alpha\left(\eta^2 P_U P_V P_U + \eta(1-\eta)P_U P_V - \eta P_U\right)(z) + P_U(z)\\&= P_U(x_{k+1} + z),\end{aligned}$$

which proves the first equation in (13) for $k+1$ and completes the proof. □

5 Computational experiments

In this section we demonstrate the theoretical results obtained in the previous sections with two different numerical experiments. In both experiments we consider randomly generated subspaces U and V in $\mathbb{R}^{50}$ with $U \cap V \neq \{0\}$. We have implemented all the algorithms in *Python 2.7* and the figures were drawn with *Matplotlib* [19].

The purpose of our first computational experiment is to exhibit the piecewise expression of the convergence rate $\gamma(T_{\alpha,\beta})$ given in Theorem 3.1. To this aim, we generated 500 pairs of random subspaces. For each pair of subspaces, we chose 10 random starting points with $\|x_0\| = 1$. Then, for each of these instances, we ran the AAMR method with $\alpha = 0.8$ and $\beta \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$. The algorithm was stopped when the shadow sequence satisfies

$$\|P_U(x_n + x_0) - P_{U\cap V}(x_0)\| < \epsilon := 10^{-8}, \tag{18}$$

where $(x_n)_{n=0}^{+\infty}$ is the sequence iteratively defined by (2) with $z = x_0$. According to Corollary 3.1, for any $\mu \in]\gamma(T_{\alpha,\beta}), 1[$, the left-hand side of (18) is bounded by μ^n for n big enough. Therefore, an estimate of the maximum number of iterations is given by

$$\frac{\log \epsilon}{\log \gamma(T_{\alpha,\beta})}. \tag{19}$$

The results are shown in Figure 6, where the points represent the number of iterations required by AAMR to satisfy (18), and the lines correspond to the estimated upper bounds given by (19). We clearly observe that the algorithm behaves in accordance with the theoretical rates. We emphasize the fact that (19) is expected to be a good upper bound on the number of iterations only when this number is *sufficiently* large. We can indeed find a few instances in the plot, especially those which require a small number of iterations, exceeding its estimated upper bound.

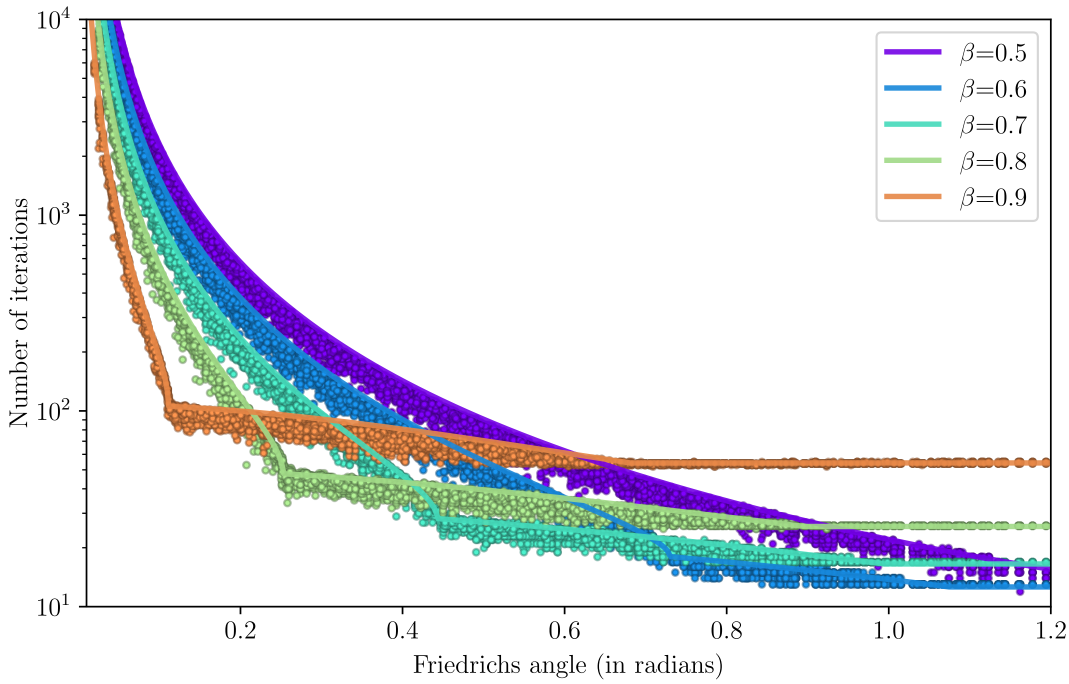


Figure 6: Number of iterations required to converge for the AAMR algorithm with $\alpha = 0.8$ and five different values of the parameter β, with respect to the Friedrichs angle. The lines correspond to the approximate upper bounds given by (19) and the theoretical rates (6)

In our second experiment we compare the performance of AP, RAP, DR and AAMR, when their parameters are selected to be optimal (see Table 1). For 100 pairs of subspaces, we generated 50 random starting points with $\|x_0\| = 1$. For a fair comparison, we monitored the shadow sequence for all the algorithms. We also used the stopping criterion (18), with $\epsilon = 10^{-8}$. The results of this experiment are summarized in Figure 7, where we show in three different graphics the median, the difference between the maximum and the median, and the coefficient of variation of the number of iterations needed to converge for each pair of subspaces. As expected, since the rate of convergence of AAMR is the smallest amongst all the compared methods (see Table 1 and Fig. 3), this algorithm is clearly the fastest, particularly for small angles. Moreover, we can observe that AAMR is one of the most robust methods (together with RAP), which makes the median to be a good representative of the rate of convergence.

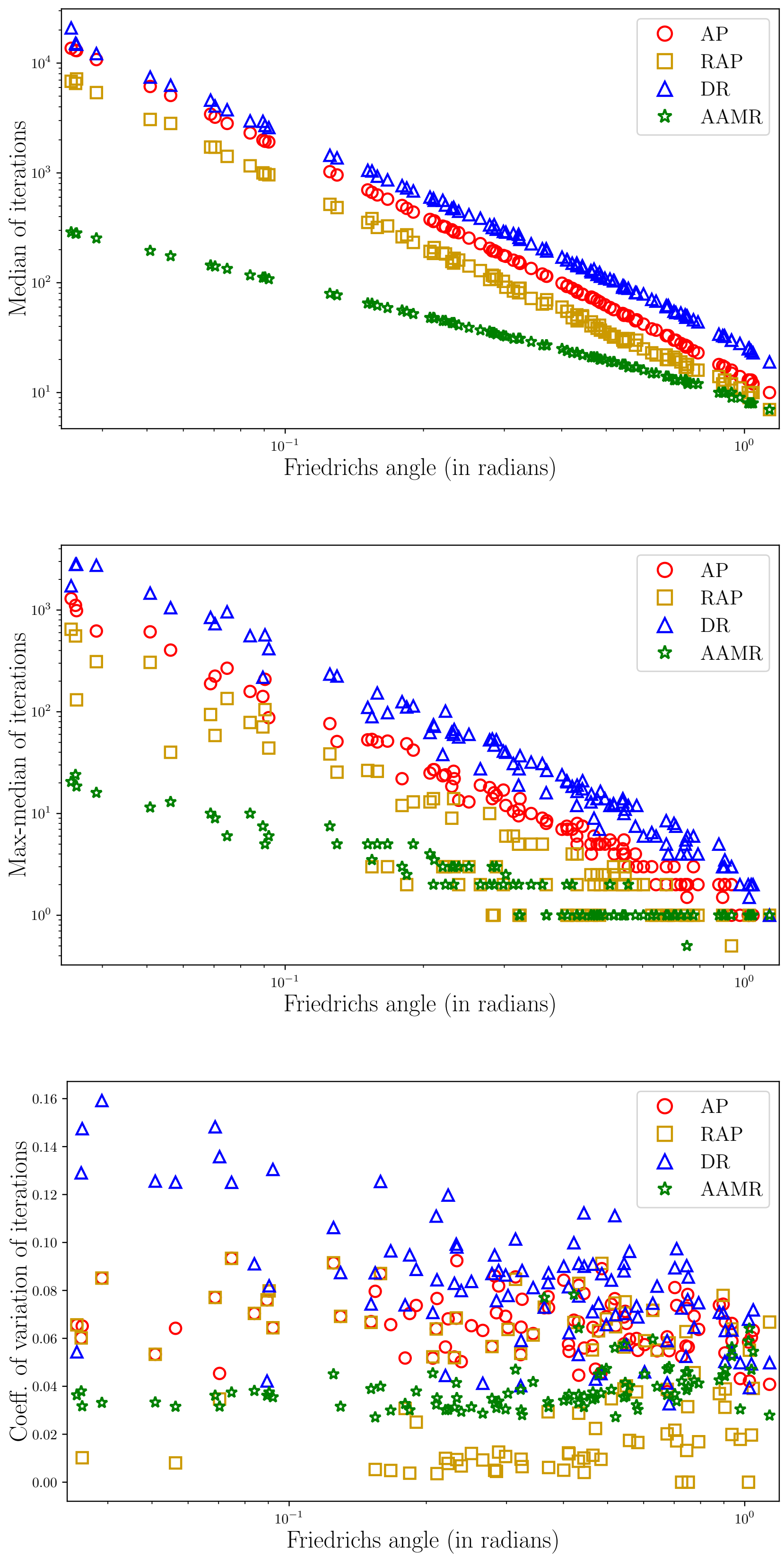


Figure 7: Median, difference between the maximum and the median, and coefficient of variation of the required number of iterations with respect to the Friedrichs angle of AP, RAP, DR and AAMR for their respective optimal parameters

6 Conclusions

We have computed the rate of linear convergence of the *averaged alternating modified reflections (AAMR) method*, which was originally introduced in [2], for the case of two subspaces in a Euclidean space. We have additionally found the optimal selection of the parameters defining the scheme that minimizes this rate, in terms of the Friedrichs angle. The rate with optimal parameters coincides with the one of the generalized alternating projections (GAP) method, which is the best among all known rates of projection methods. This coincidence motivated us to study the relationship between AAMR and GAP when they are applied to subspaces. We have discovered that, under some conditions, their associated shadow sequences overlap, which explains the coincidence of the rates. This is not the case for general convex sets.

The developed theoretical results validate the conclusions drawn in the numerical analysis of the convergence rate developed in [2, Section 7]. The sharpness of these theoretical results were additionally demonstrated in this work with two computational experiments.

The analysis in this work was done for the case of linear subspaces in a finite-dimensional space. It would be interesting to investigate in future research whether the results can be extended to infinite-dimensional spaces; or even more, to study the rate of convergence of the method when it is applied to two arbitrary convex sets.

Acknowledgements The authors would like to thank one of the anonymous referees of [2] for suggesting us this interesting project.

This work was partially supported by Ministerio de Economía, Industria y Competitividad (MINECO) of Spain and European Regional Development Fund (ERDF), grant MTM2014-59179-C2-1-P. F.J. Aragón Artacho was supported by the Ramón y Cajal program by MINECO and ERDF (RYC-2013-13327) and R. Campoy was supported by MINECO and European Social Fund (BES-2015-073360) under the program "Ayudas para contratos predoctorales para la formación de doctores 2015".

References

[1] Agmon, S.: The relaxation method for linear inequalities. Canad. J. Math. 6(3), 382–392 (1954)

[2] Aragón Artacho, F.J., Campoy, R.: A new projection method for finding the closest point in the intersection of convex sets. Comput. Optim. Appl., 1–34 (2017). DOI: 10.1007/s10589-017-9942-5

[3] Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces. Springer, New York (2011)

[4] Bauschke, H.H., Cruz, J.B., Nghia, T.T., Phan, H.M., Wang, X.: The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle. J. Approx. Theory 185, 63–79 (2014)

[5] Bauschke, H.H., Cruz, J.B., Nghia, T.T., Phan, H.M., Wang, X.: Optimal rates of linear convergence of relaxed alternating projections and generalized Douglas–Rachford methods for two subspaces. Numer. Algor. 73(1), 33–76 (2016)

[6] Cegielski, A.: Iterative methods for fixed point problems in Hilbert spaces. Lecture Notes in Mathematics, 2057. Springer, Heidelberg (2012)

[7] Cegielski, A., Suchocka, A.: Relaxed alternating projection methods. SIAM J. Optim. 19(3), 1093–1106 (2008)

[8] Censor, Y., Cegielski, A.: Projection methods: an annotated bibliography of books and reviews. Optimization 64(11), 2343–2358 (2015)

[9] `Cinderella` software. http://www.cinderella.de

[10] Deutsch, F.: The angle between subspaces of a Hilbert space. In: Singh S.P. (eds) Approximation Theory, Wavelets and Applications, pp. 107–130. Springer, Dordrecht (1995)

[11] Deutsch, F.: Best approximation in inner product spaces. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 7. Springer-Verlag, New York (2001)

[12] Douglas, J., Rachford, H. H.: On the numerical solution of heat conduction problems in two and three space variables. Trans. Amer. Math. Soc. 82, 421–439 (1956)

[13] Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. 55(1), 293–318 (1992)

[14] Escalante, R., Raydan, M.: Alternating projection methods. Fundamentals of Algorithms, 8. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2011)

[15] Fält, M., Giselsson, P.: Line search for generalized alternating projections. Preprint (2016). arXiv: 1609.05920

[16] Fält, M., Giselsson, P.: Optimal convergence rates for generalized alternating projections. Preprint (2017). arXiv: 1703.10547

[17] Gubin, L.G., Polyak, B.T., Raik, E.V.: The method of projections for finding the common point of convex sets. USSR Comp. Math. Math. Phys. 7(6), 1–24 (1967)

[18] Horn, R.A., Johnson, C.R.: Matrix analysis. Second edition. Cambridge University Press, Cambridge (2013)

[19] Hunter, J.D.: Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9(3), 90–95 (2007)

[20] Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. 16(6), 964–979 (1979)

[21] Luke, D.R.: Finding best approximation pairs relative to a convex and a prox-regular set in a Hilbert space. SIAM J. Optim. 19(2), 714–739 (2008)

[22] Meyer, C.D.: Matrix analysis and applied linear algebra. SIAM, Philadelphia (2000)

[23] Motzkin, T.S., Schoenberg, I.J.: The relaxation method for linear inequalities. Canad. J. Math. 6(3), 393–404 (1954)

[24] Von Neumann, J.: Functional operators II: The geometry of orthogonal spaces. Princeton University Press (1950). (Reprint of mimeographed lecture notes first distributed in 1933.)