ELSEVIER

Systems & Control Letters 41 (2000) 57–66

SYSTEMS & CONTROL LETTERS

www.elsevier.com/locate/sysconle

Controllability of matrix eigenvalue algorithms: the inverse power method ☆

U. Helmke[a,*], P.A. Fuhrmann[b,1]

[a] *Universität Würzburg, Institut für Mathematik, Würzburg, Germany*
[b] *Department of Mathematics, Ben-Gurion University of the Negev, Beer Sheva, Israel*

Received 27 August 1999; received in revised form 1 February 2000; accepted 28 April 2000

Abstract

In this paper we initiate a program to study the controllability properties of matrix eigenvalue algorithms arising in numerical linear algebra. Our focus is on a well-known eigenvalue method, the inverse power iteration defined on projective space. A complete characterization of the reachable sets and their closures is given via cyclic invariant subspaces. Moreover, a necessary and sufficient condition for almost controllability of the inverse power method is derived.

Keywords: Eigenvalue method; Controllability; Invariant subspaces; Projective space

1. Introduction

We consider arbitrary complex matrices $A \in \mathbb{C}^{n\times n}$ with spectrum $\sigma(A) \subset \mathbb{C}$. The *inverse power method* for A then is the discrete dynamical system

$$x_{k+1} = \frac{(A - u_k I)^{-1} x_k}{\|(A - u_k I)^{-1} x_k\|}, \quad k \in \mathbb{N}, \tag{1}$$

defined on the unit sphere

$$S^{2n-1} = \{x \in \mathbb{C}^n \mid \|x\| = 1\}.$$

Here the values of the so-called *origin shifts* u_k can be arbitrary complex numbers such that $A - u_k I$ is invertible. Thus, we assume $u_k \in \mathbb{C} - \sigma(A)$ for all $k \in \mathbb{N}$. For constant shifts this is just the celebrated inverse-power method to compute a dominant eigenvector of A. In this case, the method, despite its overall simplicity and beauty, is however too slowly convergent to be of practical use.

In numerical linear algebra, one is therefore interested in choosing the origin shifts in such a way as to improve convergence speed of the algorithm. Thus, the origin shifts are treated as control variables to modify the system dynamics. A popular choice is the simple feedback strategy

$$u_k = x_k^* A x_k, \quad k \in \mathbb{N},$$

via the Rayleigh quotient. The resulting closed-loop nonlinear dynamical system

$$x_{k+1} = \frac{(A - x_k^* A x_k I)^{-1} x_k}{\|(A - x_k^* A x_k I)^{-1} x_k\|}, \quad k \in \mathbb{N} \tag{2}$$

☆ Partially supported by a grant from the German–Israeli Foundation GIF I-526-034.06/97.

* Correspondence address. Department of Mathematics, University of Wurzburg, Am Hubland, 97074 Wurzburg, Germany. Fax: +49-931-888-4611.

E-mail address: helmke@mathematik.uni-wuerzburg.de (U. Helmke).

[1] Earl Katz Family Chair in Algebraic System Theory.

PII: S0167-6911(00)00039-6

is called the *inverse Rayleigh iteration* on S^{2n-1}, and is frequently considered as a fast numerical algorithm to compute the dominant eigenvector of a matrix. It is well known that for normal matrices the system (2) converges *cubically* fast to the dominant eigenvectors; see, e.g., [6,4]. In general, however, the dynamics of the inverse Rayleigh iteration can be quite complicated. In fact, Batterson and Smilie [1] have shown that there exists an open set of (non-normal) matrices A and an open subset of initial conditions on S^{2n-1} such that (2) does not converge to an eigenvector. Altogether, a complete analysis of the dynamics of (2) seems far from being available.

The question arises whether such complicated dynamics can be avoided using different control strategies. In order to develop a systematic approach to such questions it becomes important to know the controllability properties of numerical eigenvalue algorithms such as (1). In fact, no feedback control strategy for (1) can produce solutions that lie outside of the reachable set of (1). Thus, the reachable sets provide fundamental limitations on the possible convergence behaviour of (1).

It is for such reasons that we became interested in analyzing the reachability properties of numerical eigenvalue methods. In fact, known efficient eigenvalue algorithms such as the QR, LR, or more generally, FG-algorithms, all implement feedback control strategies via suitable origin shifts and controllability of these methods thus becomes an important (unsolved) issue for a theoretical analysis of these algorithms. Although our results have so far not led us to the design of superior numerical algorithms, we believe that it is a fundamental issue to understand the potential limitations of such algorithms in terms of the reachable sets.

We proceed as follows. In Section 2 the inverse-power method is reformulated as a control system defined on projective space. The reachable sets are characterized as the orbits of a subgroup of the centralizer group of A. In Section 3 we focus on the case of cyclic operators. Using tools from the theory of polynomial models (see Ref. [3]), a complete description of the orbits and their orbit closures is obtained. Section 4 treats the general noncyclic case and summarizes the main results of the paper. It is shown that (1) has a dense reachable set if and only if A is cyclic. In contrast, a simple dimension argument shows that the QR-algorithm is not controllable.

2. Power iteration on projective space

For technical reasons it turns out to be more convenient to consider the inverse-power iteration as a dynamical system defined on projective space rather than on a sphere. In this way, we can avoid potential ambiguities due to the phase factors of complex numbers of absolute value one as well as extend the scope by working over an arbitrary field of coefficients. As a reference for algebraic geometry notions such as projective spaces, etc., we refer to Hartshorne [7].

Thus, let F denote an arbitrary field. Given a finite dimensional F-vector space V, the *projective space* $\mathbb{P}(V)$ is defined as the set of all one-dimensional F-linear subspaces of V. More precisely, given any nonzero vector $x \in V$, let

$$X := [x] = \{\lambda x \mid \lambda \in F - \{0\}\}$$

denote the line generated by x in $V - \{0\}$. The projective space of V over F is then defined as

$$\mathbb{P}(V) := \{[x] \mid x \in V - \{0\}\}.$$

In the sequel we will denote points in a vector space by lower-case letters x, y, while the corresponding elements in projective space $\mathbb{P}(V)$ will be denoted by capital letters such as X, Y. If V is finite-dimensional then $\mathbb{P}(V)$ is a projective algebraic variety (or actually the F-rational points of such a variety), endowed with the Zariski topology. If $F = \mathbb{R}, \mathbb{C}$ then the projective spaces $\mathbb{RP}^{n-1} := \mathbb{P}(\mathbb{R}^n)$, $\mathbb{CP}^{n-1} := \mathbb{P}(\mathbb{C}^n)$ are also referred to as $(n-1)$-dimensional real or complex projective space. $\mathbb{CP}^{n-1}$ is a compact complex manifold and $\mathbb{CP}^1 \simeq S^2$ is diffeomorphic to the familiar Riemann sphere.

Let $A \in F^{n\times n}$ and $u \in F$ not an eigenvalue of A. Then $A - uI : F^n \to F^n$ is an invertible linear map and thus defines an algebraic map

$$A - uI : \mathbb{P}(F^n) \to \mathbb{P}(F^n),$$

$$X \mapsto (A - uI)X$$

and similarly for $(A - uI)^{-1}$. Thus, the inverse power method

$$X_{k+1} = (A - u_k I)^{-1} X_k, \quad k \in \mathbb{N}, \tag{3}$$

with shifts $u_k \in F - \sigma(A)$, $X_k \in \mathbb{P}(F^n)$, defines a nonlinear control system on projective space $\mathbb{P}(F^n)$.

Let

$$\mathscr{R}_A(X_0) := \left\{ \prod_{i=0}^{k} (A - u_i I)^{-1} X_0 \mid u_i \in F - \sigma(A),\ k \in \mathbb{N}_0 \right\} \tag{4}$$

denote the *reachable set* of (3), starting from X_0. Thus $\mathscr{R}_A(X_0)$ is the set of all states obtained from X_0 after a finite number of iterations in (3). We say that (3) is *almost controllable* if the system has a reachable set that is dense in $\mathbb{P}(F^n)$.

Our first result identifies $\mathscr{R}_A(X_0)$ with the orbit of an algebraic group action on $\mathbb{P}(F^n)$. In the sequel we will completely analyze the structure of these orbits. Given an arbitrary polynomial $p \in F[z]$, the matrix $p(A) \in F^{n\times n}$ is invertible if and only if $p(z)$ is coprime with the characteristic polynomial $q(z) = \det(zI - A)$ of A. Given two polynomials $p, q \in F[z]$, let $p \wedge q$ and $p \vee q$ denote the greatest common divisor and the least common multiple of the polynomials p and q, respectively. Thus, p and q are coprime if and only if $p \wedge q = 1$. With this notation, consider

$$\begin{aligned} \Gamma_A &:= F[A]^{\times} := F[A] \cap GL(n,F) \\ &= \{ p(A) \mid p \in F[z] \text{ and } p \wedge q = 1 \}. \end{aligned} \tag{5}$$

Lemma 2.1. *For $A \in F^{n\times n}$ let $m_A \in F[z]$ denote the minimal polynomial of A and $\Gamma_A := \{p(A) \mid p \in F[z], p$ coprime with $\det(zI - A)\}$. Then Γ_A is a closed algebraic subgroup of $GL(n,F)$ of dimension $\dim_F \Gamma_A = \deg m_A$.*

Proof. Obviously, Γ_A is multiplicatively closed. Let $p \in F[z]$ be coprime with q. By the Euclidean algorithm, there exist polynomial $a, b \in F[z]$ with $\deg a < n$ which solve the Bezout equation

$$a(z)p(z) + b(z)q(z) = 1.$$

By the Cayley–Hamilton theorem we have $a(A)p(A) = I$ and therefore $p(A)^{-1} = a(A) \in \Gamma_A$. Thus, Γ_A is a group. Again, using the Cayley–Hamilton theorem, we see that every element of Γ_A has a unique representation as $p(A)$ with a p coprime with q and $\deg p < \deg m_A$. From the resultant test for coprimeness we see that

$$\Omega_q = \{ p \in F[z] \mid p \wedge q = 1,\ \deg p < \deg m_A \}$$

defines a Zariski open subset of F^d, $d = \deg m_A$, and indeed a complement of a hypersurface. Therefore, Ω_q and hence Γ_A define an affine algebraic variety of dimension $d = \deg m_A$. □

Remark. If A is cyclic then Γ_A coincides with the centralizer subgroup $C_A := \{S \in GL(n,F) \mid SA = AS\}$, but in general Γ_A is a proper subgroup of C_A.

Over $\mathbb{C}$ an equivalent description of Γ_A is

$$\Gamma_A = \left\{ \exp\left(\sum_{i=0}^{n-1} \alpha_i A^i \right) \middle| \alpha_i \in \mathbb{C} \right\}.$$

Consider now the algebraic group action

$$\begin{aligned} \alpha : \Gamma_A \times \mathbb{P}(F^n) &\to \mathbb{P}(F^n) \\ (p(A), X) &\mapsto p(A)X. \end{aligned} \tag{6}$$

Let $\simeq$ denote the associated equivalence relation on projective space, i.e. for lines $X, Y \in \mathbb{P}(F^n)$ we have $X \simeq Y$ if and only if $Y = p(A)X$ for some $p(A) \in \Gamma_A$. The equivalence classes of $\simeq$ are then the *orbits*

$$\Gamma_A \cdot X = \{ p(A)X \mid p(A) \in \Gamma_A \} \subset \mathbb{P}(F^n). \tag{7}$$

The next result shows that the reachable sets for (1) coincide with the orbits of Γ_A. Recall that a subset of $\mathbb{P}(F^n)$ is called quasi-projective if it is an open subset of a projective variety.

Theorem 2.1. *Let F be an algebraically closed field. Then the reachable set $\mathscr{R}_A(X_0)$ of each $X_0 \in \mathbb{P}(F^n)$ is a nonsingular quasiprojective subvariety of $\mathbb{P}(F^n)$ and*

$$\mathscr{R}_A(X_0) = \Gamma_A \cdot X_0, \quad \text{for all } X_0 \in \mathbb{P}(F^n). \tag{8}$$

Proof. By the fundamental theorem of algebra, every monic polynomial $p \in F[z]$ splits into linear factors as

$$p(z) = \prod_{i=1}^{m} (z - u_i), \quad u_i \in F,$$

where p is a coprime with a polynomial q if and only if $q(u_i) \neq 0$ for all i. If $X \in \mathscr{R}_A(X_0)$, then

$$X = \prod_{i=1}^{m} (A - u_i I)^{-1} X_0$$

and $u_i \in F - \sigma(A)$. Thus, $X = p(A)^{-1}X_0$ with $p(A) = \prod_{i=1}^{m}(A - u_i I) \in \Gamma_A$ and therefore $X \in \Gamma_A \cdot X_0$. Conversely, any $X \in \Gamma_A \cdot X_0$ is of the form $X = p(A)^{-1}X_0$ with $p(A) \in \Gamma_A$. Thus, $X = \prod_{i=1}^{m}(A - u_i I)^{-1} X_0 \in \mathscr{R}_A(X_0)$, with $p(z) = \prod_{i=1}^{m}(z - u_i)$. □

Remark. If A has all its eigenvalues in F then we still have

$$\mathscr{R}_A(X_0) = \Gamma_A \cdot X_0, \quad \text{for all } X_0 \in \mathbb{P}(F^n).$$

If F is a *real closed field*, e.g. $F=\mathbb{R}$ the field of real numbers, then all irreducible polynomials have either degree 1 or 2. Thus, in this case, by allowing quadratic shift transformations

$$X_{k+1} = p_k(A)^{-1}X_k$$

with p_k monic and irreducible over F, then $\mathscr{R}_A(X_0) = \Gamma_A \cdot X_0$, for all $X_0 \in \mathbb{P}(F^n)$. However, by using only linear shifts of the form $(A-u_kI)^{-1}$ the reachable set is in general only a strict subset of $\Gamma_A \cdot X_0$.

Remark. For $F=\mathbb{C}$, let $\mathscr{R}_A(x_0)$ denote the reachable set of $x_0 \in S^{2n-1}$ for the power iteration on the sphere S^{2n-1} and let

$$\Gamma_A \cdot x_0 = \left\{ \frac{p(A)x_0}{\|p(A)x_0\|} \;\middle|\; p(A)\in\Gamma_A \right\}$$

denote the Γ_A-orbit on S^{2n-1}. Then the sets differ by phase factors, i.e. $\mathscr{R}_A(x_0) \subset \Gamma_A \cdot x_0$ and

$$\Gamma_A \cdot x_0 = \{\lambda x \mid x \in \mathscr{R}_A(x_0),\ \lambda\in\mathbb{C},\ |\lambda|=1\}.$$

3. The cyclic case

In this section, we study the controllability properties of the inverse power method for the case the matrix A is cyclic. In this case, all the information about A, up to similarity, is contained in the characteristic polynomial of A. This allows us to apply polynomial methods to the problem. In particular, we will use the theory of polynomial models developed by the second author (see [2,3]).

We begin by introducing polynomial models for the case of scalar polynomials. We assume an arbitrary field F and consider a monic polynomial $q \in F[z]$. We denote by π_q the operator of taking the remainder modulo q. By $X_q = \operatorname{Im}\pi_q$ we denote the linear space of remainders modulo q, i.e.

$$X_q = \{p \in F[z] \mid \deg p < \deg q\}.$$

We define an $F[z]$-module structure on X_q by

$$p \cdot f = \pi_q p f, \quad f \in X_q.$$

We define the shift map S_q acting in X_q by $S_q f = \pi_q z f$. Clearly we have the isomorphism $X_q \simeq F[z]/qF[z]$. This is an isomorphism both as linear spaces as well as modules over the ring of polynomials $F[z]$.

The shift operators are important as each cyclic operator is isomorphic to a uniquely determined shift.

Theorem 3.1. *Let $A: F^n \to F^n$ be a cyclic transformation with b a cyclic vector for A. Let q be the characteristic polynomial of A and let the map $\phi: X_q \to F^n$ be defined by $\phi(p) = p(A)b$. Then ϕ is an isomorphism and the following diagram is commutative*:

$$\begin{array}{ccc} X_q & \xrightarrow{\ \phi\ } & F^n \\ {\scriptstyle S_q}\big\downarrow & & \big\downarrow{\scriptstyle A} \\ X_q & \xrightarrow{\ \phi\ } & F^n \end{array}$$

The S_q-invariant subspaces of X_q, which are just the submodules of X_q, are closely related to factorizations of q. In fact, we have the following.

Proposition 3.1. *Given a monic polynomial $q \in F[z]$.*

1. *A subspace V of X_q is a submodule if and only if it has a representation $V = q_1X_{q_2}$ for some factorization $q = q_1q_2$.*
2. *Let $q = q_1q_2 = p_1p_2$ be two factorizations. Then we have the inclusion*

   $$q_1X_{q_2} \subset p_1X_{p_2} \tag{9}$$

   if and only if $p_1 \mid q_1$, or equivalently $q_2 \mid p_2$.
3. *Given factorizations $q = p_iq_i$, $i = 1,\dots,s$, then $\bigcap_{i=1}^{s} p_iX_{q_i} = pX_q$ with p the l.c.m of the p_i and q the g.c.d. of the q_i.*
4. *Given factorizations $q = p_iq_i$, $i = 1,\dots,s$ then $\sum_{i=1}^{s} p_iX_{q_i} = pX_q$ with q the l.c.m of the q_i and p the g.c.d. of the p_i.*

In the study of polynomial models, coprimeness of polynomials is of central importance. It relates to the characterization of cyclic vectors of the shift, to direct sum of submodules decomposition of a polynomial model and finally to invertibility of polynomials in the shift.

Thus, we can state

Theorem 3.2. *Let q be monic.*

1. *Let $f \in X_q$. Then f is a cyclic vector for S_q if and only if $f \wedge q = 1$.*
2. *Let $p \in F[z]$. Then $p(S_q)$ is invertible if and only if $p \wedge q = 1$.*
3. *Let $q = q_1q_2$. Then $X_q = q_2X_{q_1} \oplus q_1X_{q_2}$ if and only if $q_1 \wedge q_2 = 1$.*

So far we have used only the module structure of the polynomial model X_q. We find it appropriate to quote, adapting the notation, a remark by Kalman et al. [5]: "It is embarassing to have to observe that the state space X_q has even more structure than claimed. It is even a ring with multiplication

$$p_1 \cdot p_2 = \pi_q(p_1 p_2), \tag{10}$$

where $p_1 p_2$ is the ordinary product in $F[z]$. No system-theoretic interpretation of (10) is known at present, etc.". Actually, the natural structure to consider for our purposes is the F-algebra structure of X_q.

Let

$$\Gamma_q := \{p \in X_q \mid p \text{ is invertible in } X_q\}$$

denote the *group of units* of X_q. We will study some of the properties of this F-algebra and focus especially on the group of units Γ_q in X_q. We have an easy characterization of the units in X_q.

Lemma 3.1. *Let $p \in X_q$. Then $p \in \Gamma_q$ if and only if p and q are coprime, i.e.*

$$\Gamma_q = \{p \in X_q \mid p \wedge q = 1\}. \tag{11}$$

Proof. Assume p and q are coprime, then there exist a solution to the Bezout equation $ap + bq = 1$, or $a \cdot p = 1$. Clearly $a = p^{-1}$. Conversely, assume $a \cdot p = 1$. Then there exists a polynomial b for which $ap + bq = 1$ which implies the coprimeness of p and q. □

The following result characterizes the orbits of the Γ_q-action via factors of q.

Theorem 3.3. *Let $x \in X_q$.*

1. *Every orbit $\Gamma_q \cdot x$ is of the form*

   $$\Gamma_q \cdot x = \Gamma_q \cdot t, \tag{12}$$

   where $t = x \wedge q$ is the greatest common factor. In particular, there are only finitely many distinct orbits. They are given as $\Gamma_q \cdot t$, where $t \in X_q$ is an arbitrary monic factor of q.
2. *We have, for $x_1, x_2 \in X_q$, $\Gamma_q \cdot x_1 = \Gamma_q \cdot x_2$ if and only if $x_1 \wedge q = x_2 \wedge q$.*
3. *There is a bijective correspondence between orbits of Γ_q and monic factors of q.*

Proof.

1. Let $x = ut$ with $t = x \wedge q$. Thus, u is a coprime with q, i.e. $u \in \Gamma_q$, and therefore

   $$\Gamma_q \cdot x = (\Gamma_q \cdot u) \cdot t = \Gamma_q \cdot t.$$

2. Clearly, from (1),

   $$\Gamma_q \cdot x_1 = \Gamma_q \cdot x_2 \quad \Leftrightarrow \quad \Gamma_q \cdot (x_1 \wedge q) = \Gamma_q \cdot (x_2 \wedge q),$$

   which shows that

   $$x_1 \wedge q = x_2 \wedge q \Rightarrow \Gamma_q \cdot x_1 = \Gamma_q \cdot x_2.$$

   Conversely, let t_1, t_2 be two arbitrary factors of q with $\Gamma_q \cdot t_1 = \Gamma_q \cdot t_2$. Then $t_1 = p \cdot t_2$ for some $p \in \Gamma_q$. We have $p \cdot t_2 = pt_2 + \alpha q$, for some $\alpha \in F[z]$, and therefore $t_1 = pt_2 + qr_1$. Since t_1 divides q it must divide pt_2. Since $p \wedge q = 1$ we have $t_1 \wedge p = 1$. Thus, t_1 divides t_2. By symmetry, t_2 divides t_1. As both are monic, we have the equality $t_1 = t_2$.
3. This is a reflection of the unique factorization theorem. Clearly, every factor t of q determines an orbit $\Gamma_q \cdot t$ and, by (1) and (2) every orbit has such a unique representation. □

In the sequel we shall denote, for $M \subset X_q$, by $\bar{M}$ the Zariski closure of M in X_q. We need the following elementary lemma.

Lemma 3.2. *Let $A : F^n \to F^m$ be a surjective linear map. Then the image $A(\Omega)$ of any Zariski open subset $\Omega \subset F^n$ is Zariski open in F^m.*

Proof. Let $B : F^m \to F^n$ be a linear right inverse of A, i.e. $AB = I$. Then $A(\Omega) = B^{-1}(\Omega)$. Since the pre-image of a Zariski open subset under a polynomial map is Zariski open, we are done. □

Theorem 3.4. *Let $x \in X_q$. Then*

$$\overline{\Gamma_q \cdot x} = \{p \cdot x \mid p \in X_q\} = X_q \cdot x. \tag{13}$$

Proof. The set of coefficient vectors of polynomials p coprime with q defines a Zariski open subset of F^n, characterized by the nonvanishing of the resultant of p and q.

Consider the linear map

$$\psi : F^n \to X_q,$$

$$p \mapsto p \cdot x,$$

whose image is the linear space

$$\psi(F^n) = \{p \cdot x \mid p \in X_q\}.$$

We have to show that the image $\psi(\Omega) = \Gamma_q \cdot x$ of the Zariski open subset $\Omega := \{p \in F^n \mid p \wedge q = 1\}$ is again a Zariski open subset of $\psi(F^n)$. The result then follows as any nonempty Zariski open subset of an affine (or irreducible) space is dense. But this immediately follows from the previous lemma. □

Theorem 3.5. *Let t_1, t_2 be factors of q. Then*

1. *We have*

$$\Gamma_q \cdot t_1 \cap \overline{\Gamma_q \cdot t_2} \neq \emptyset \;\Leftrightarrow\; \Gamma_q \cdot t_1 \subset \overline{\Gamma_q \cdot t_2} \;\Leftrightarrow\; \overline{\Gamma_q \cdot t_1} \subset \overline{\Gamma_q \cdot t_2} \;\Leftrightarrow\; t_2 \mid t_1. \quad (14)$$

2. *We have*

$$\overline{\Gamma_q \cdot t} = \bigcup_{t|s|q} \Gamma_q \cdot s. \quad (15)$$

Proof.

1. For the first equivalence the reverse direction is trivial. Assume $p \cdot t_1 \in \overline{\Gamma_q \cdot t_2}$ for $p \in \Gamma_q$. By continuity of multiplication in X_q with respect to the Zariski topology this implies $\Gamma_q \cdot t_1 \subset \overline{\Gamma_q \cdot t_2}$. The second equivalence is trivial. For the third note that

$$\Gamma_q \cdot t_1 \subset \overline{\Gamma_q \cdot t_2} \;\Leftrightarrow\; t_1 \in \{p \cdot t_2 \mid p \in X_q\}.$$

Thus, suppose $t_1 = p \cdot t_2$ for some $p \in X_q$. Then $t_1 = p \cdot t_2 = pt_2 + rq$ for some $r \in F[z]$. Thus, $pt_2 = t_1 - rq$ and therefore t_2 divides $t_1 - rq$. Since t_2 divides q this implies that t_2 divides t_1.

Conversely, let $t_1 = at_2$ for some $a \in F[z]$. Write $a = \alpha q + p$ for $p \in F[z]$, $\deg p < \deg q$. Then

$$t_1 = at_2 = \alpha t_2 q + t_2 p = p \cdot t_2,$$

with $p \in X_q$.

2. We observe

$$x \in \overline{\Gamma_q \cdot t} \Leftrightarrow \Gamma_q \cdot x \subset \overline{\Gamma_q \cdot t} \Leftrightarrow \Gamma_q \cdot (x \wedge t) \subset \overline{\Gamma_q \cdot t} \Leftrightarrow t \mid (x \wedge q) \mid q \Leftrightarrow x \in \bigcup_{t|s|q} \Gamma_q \cdot s. \qquad \square$$

In Theorem 3.4 the polynomials $p \in X_q$ that occur in a representation of an element $a \in X_q \cdot x$ as $a = p \cdot x$ are by no means uniquely determined. The next result describes a unique representation of elements in $X_q \cdot x$.

Theorem 3.6. *Let $t = x \wedge q$ be the greatest common divisor of the polynomials x and q, $q = st$. Then*

1. *Any element in $\overline{\Gamma_q \cdot x} = X_q \cdot t$ has a unique description as $p \cdot t$ with $\deg p < \deg s$. In particular, we have*

$$\overline{\Gamma_q \cdot x} = X_s \cdot t = \{p \cdot t \mid \deg p < \deg s\}.$$

2. *We have*

$$\Gamma_q \cdot x = \Gamma_q \cdot t = \Gamma_s \cdot t = t\Gamma_s \cdot 1.$$

More explicitly,

$$\Gamma_q \cdot t = \{p \cdot t \mid \deg p < \deg s,\ p \wedge s = 1\}.$$

Proof.

1. From Theorem 3.4 we have

$$\overline{\Gamma_q \cdot x} = \overline{\Gamma_q \cdot t} = \{p \cdot t \mid p \in X_q\}.$$

Writing

$$p = as + r, \quad \deg r < \deg s = \deg q - \deg t,$$

we obtain

$$p \cdot t = as \cdot t + r \cdot t = a \cdot q + r \cdot t = r \cdot t,$$

and thus $X_q \cdot t = X_s \cdot t$.

For the uniqueness assume $p_1 \cdot t = p_2 \cdot t$ for polynomials p_1, p_2 with $\deg p_i < \deg s$, $i = 1, 2$. Then $(p_1 - p_2) \cdot t = 0$ in X_q, i.e. there exists $r \in F[z]$ such that $0 = (p_1 - p_2)t + rq$. Thus, $q = ts$ divides $(p_1 - p_2)t$, which is impossible as $\deg(p_1 - p_2) < \deg s$.

2. Obviously, $\Gamma \cdot x = \Gamma \cdot t$. Write any $p \in \Gamma_q$ as $p = as + r$ with $\deg r < \deg s$. Then $p \cdot t = p \cdot r$. Since p is coprime with q it is in particular coprime with s. Thus, $r = p - as$ must be coprime with s and therefore we obtain the inclusion

$$\{p \cdot t \mid p \in \Gamma_q\} \subset \{r \cdot t \mid r \in \Gamma_s\}.$$

To show the reverse inclusion, let $r \in \Gamma_s$ be given, $\deg r < \deg s$. Let s' be such that $u = s \wedge t$, $s = s'u$. s' clearly exists. Then $s' \wedge t = 1$. Multiplying the Bezout equation by $1 - r$ we obtain polynomials a, b with $as' + bt + r = 1$. Consider the polynomial $f = as' + r$. Since $r \wedge s = 1$ and $s' \mid s$, we have $f \wedge s' = 1$. Moreover, $f + bt = 1$ and so f must be coprime with t. Thus, f is a coprime with s' and t and therefore it is coprime with $q = st$.

Choose polynomials α, p with $f = \alpha q + p$, $\deg p < \deg q$. Then p is coprime with q and we have

$$\begin{aligned} r \cdot t &= (f - as) \cdot t = f \cdot t - a \cdot q \\ &= f \cdot t = \alpha q \cdot t + p \cdot t \\ &= p \cdot t. \end{aligned}$$

Hence,

$$\{r \cdot t \mid r \in \Gamma_s\} \subset \{p \cdot t \mid p \in \Gamma_q\}. \qquad \square$$

We are now able to give a detailed analysis of the structure of the orbits.

Corollary 3.1. *Let $q = q_1^{\nu_1} \cdots q_r^{\nu_r}$ be a primary decomposition, $\nu_i \geqslant 1$. Then*

1. *The dimension of each orbit is given by*
$$\dim \Gamma_q \cdot x = \dim \overline{\Gamma_q \cdot x} = \deg q - \deg(x \wedge q).$$
2. *$\Gamma_q \cdot 1$ is the unique Zariski open and dense orbit in X_q.*
3. *There are exactly r closed orbits in X_q and they are given by*
$$\Gamma_q \cdot (q_1^{\nu_1 - 1} \cdots q_r^{\nu_r}), \ldots, \Gamma_q \cdot (q_1^{\nu_1} \cdots q_r^{\nu_r - 1}).$$
4. *We have*
$$\dim \Gamma_q \cdot x = \dim X_q \quad \Leftrightarrow \quad x \wedge q = 1, \tag{16}$$
and
$$\dim \Gamma_q \cdot x = 1 \quad \Leftrightarrow \quad \deg(x \wedge q) = 1. \tag{17}$$
The latter is achievable if and only if q has a root in F.

Proof.

1. Let $t = x \wedge q$, $q = st$. Then
$$\begin{aligned}\dim \Gamma_q \cdot x &= \overline{\dim \Gamma_q \cdot x} \\ &= \dim \{ p \cdot (x \wedge q) \mid \deg p \\ &\qquad < \deg q - \deg(x \wedge q)\} \\ &= \dim X_s = \deg q - \deg(x \wedge q).\end{aligned}$$
2. By 1, $\overline{\Gamma_q \cdot x} = X_q$ implies that deg $(x \wedge q) = 0$, i.e. $x \wedge q = 1$. Conversely, if $x \wedge q = 1$ then $\Gamma_q \cdot x = \Gamma_q \cdot 1 = \Gamma_q$ is dense in X_q.
3. For a factor t of q
$$\overline{\Gamma_q \cdot t} = \bigcup_{t|s|q} \Gamma_q \cdot s = \Gamma_q \cdot t$$
if and only if there is no nontrivial factor s of q divisible by t, $s \neq t$. From this the claim follows.
4. This is an immediate consequence of the dimension formula for orbits. □

Corollary 3.2. *There is a bijective correspondence between*

1. *Orbits $\Gamma_q \cdot x$, $x \in X_q$,*
2. *closure of orbits $\overline{\Gamma_q \cdot x}$, $x \in X_q$,*
3. *S_q-invariant subspaces* of X_q,
4. *monic polynomial factors of q.*

Proof. The equivalence (1) ⇔ (2) ⇔ (4) is obvious.

Now, given any factor t of $q = st$, the subspace $V = tX_s$ is shift invariant and conversely, any shift invariant subspace has such a representation. This proves (4) ⇔ (3). □

In particular, we note that the closures $\overline{\Gamma_q \cdot x}$ of Γ_q-orbits coincide with the shift invariant subspaces via
$$\overline{\Gamma_q \cdot x} = tX_s,$$
for any factorization $q = ts$.

The identification of closures of orbits with shift invariant subspaces allows us to transport the arithmetic of shift invariant subspaces to the closures of orbits.

Corollary 3.3. *Let $s \wedge t$ and $s \vee t$ denote the greatest common divisor and the least-common multiple of the polynomials s and t, respectively. For factors s, t of q we have*
$$\overline{\Gamma_q \cdot s} \cap \overline{\Gamma_q \cdot t} = \overline{\Gamma_q \cdot (s \vee t)}, \tag{18}$$
and
$$\overline{\Gamma_q \cdot s} + \overline{\Gamma_q \cdot t} = \overline{\Gamma_q \cdot (s \wedge t)}. \tag{19}$$

Recall that given a factorization $q = q_1 \cdots q_r$ with pairwise coprime polynomials q_i, there is a direct sum decomposition of X_q into $F[z]$-submodules as
$$X_q = \pi_1 X_{q_1} \oplus \cdots \oplus \pi_r X_{q_r},$$
with $\pi_i = \prod_{j \neq i} q_j$. Although X_q and X_{q_i} have algebra structures, the above decomposition is not compatible with these structures. To see that this decomposition can nevertheless be interpreted in terms of algebras, we need to introduce a rather special algebra structure on products of polynomial model algebras.

Definition 3.1. Let $q_1, \ldots, q_r$ be pairwise coprime polynomials and $\pi_i = \prod_{j \neq i} q_j$. The *twisted algebra structure* on $X_{q_1} \times \cdots \times X_{q_r}$ is defined by
$$(p_1, \ldots, p_r) \circ (p_1', \ldots, p_r') := (\pi_1 p_1 p_1', \ldots, \pi_r p_r p_r'). \tag{20}$$

Note that π_i is invertible in X_{q_i} for all i. It is easily seen that this defines a commutative algebra structure on $X_{q_1} \times \cdots \times X_{q_r}$. The unit element is $(\pi_1^{-1}, \ldots, \pi_r^{-1})$ and an element $(p_1, \ldots, p_r) \in X_{q_1} \times \cdots \times X_{q_r}$ is invertible with respect to the twisted algebra structure if and only if for all $i = 1, \ldots, r$, p_i is invertible in X_{q_i}. Moreover, we have
$$(p_1, \ldots, p_r)^{-1} = (\pi_1^{-2} p_1^{-1}, \ldots, \pi_r^{-2} p_r^{-1}).$$
Thus, the group of units of the twisted algebra $X_{q_1} \times \cdots \times X_{q_r}$ coincides with the twisted products of groups of units $\Gamma_{q_1} \times \cdots \times \Gamma_{q_r}$.

Theorem 3.7. *Let $q = q_1 \cdots q_r$ be a factorization with pairwise coprime polynomials q_i and let $\pi_i = \prod_{j \neq i} q_j$, $i = 1,\dots,r$. Then, with respect to the twisted algebra structure on $X_{q_1} \times \cdots \times X_{q_r}$, the linear map*

$$\phi : X_{q_1} \times \cdots \times X_{q_r} \to X_q$$

$$\phi(p_1,\dots,p_r) = \sum_{i=1}^{r} \pi_i p_i, \tag{21}$$

is an algebra isomorphism. It induces an isomorphism of groups of units

$$\phi : \Gamma_{q_1} \times \cdots \times \Gamma_{q_r} \to \Gamma_q. \tag{22}$$

Proof. It is well known that $\phi : X_{q_1} \times \cdots \times X_{q_r} \to X_q$ is an $F[z]$-module isomorphism (see Fuhrmann [3]). Thus, it remains to show that ϕ is an algebra homomorphism. We have

$$\begin{aligned} &\phi(p_1,\dots,p_r)\phi(p'_1,\dots,p'_r) \\ &\quad= \left(\sum_{i=1}^{r} \pi_i p_i\right)\left(\sum_{j=1}^{r} \pi_j p'_j\right) \\ &\quad= \sum_{i,j=1}^{r} \pi_i\pi_j p_i p'_j. \end{aligned}$$

Since q divides $\pi_i\pi_j$ for $i \neq j$, we have $\pi_i\pi_j = 0$ in X_q for $i \neq j$. Thus, we obtain

$$\phi(p_1,\dots,p_r)\phi(p'_1,\dots,p'_r) = \sum_{i=1}^{r} \pi_i^2 p_i^2.$$

But

$$\begin{aligned} &\phi((p_1,\dots,p_r) \circ (p'_1,\dots,p'_r)) \\ &\quad= \phi(\pi_1 p_1 p'_1,\dots,\pi_r p_r p'_r) \\ &\quad= \sum_{i=1}^{r} \pi_i(\pi_i p_i p'_i) = \sum_{i=1}^{r} \pi_i^2 p_i p'_i \\ &\quad= \phi(p_1,\dots,p_r)\phi(p'_1,\dots,p'_r). \end{aligned}$$

This proves the claim. Any algebraic isomorphism induces an isomorphism between the groups of units, so the result follows. □

Corollary 3.4. *Let $q = q_1^{v_1} \cdots q_r^{v_r}$ be a primary decomposition. Then, with $\pi_i = \prod_{j \neq i} q_j$, we have an isomorphism of groups of units*

$$\phi : \Gamma_{q_1^{v_1}} \times \cdots \times \Gamma_{q_r^{v_r}} \to \Gamma_q,$$

$$\phi(p_1,\dots,p_r) = \sum_{i=1}^{r} \pi_i p_i.$$

Remark. If q is a prime in $F[z]$, then X_q is a finite-field extension of F and $\Gamma_q = X_q - \{0\}$. More generally, X_{q^v} is an algebra extension of F of dimension $v \deg q$. In fact

$$X_{q^v} = X_q \otimes_F X_{z^v}.$$

Moreover, every element of X_{q^v} can be written as

$$p = p_0 + p_1 q + \cdots + p_{v-1} q^{v-1}, \quad p_i \in X_q,$$

and

$$p \in \Gamma_{q^v} \quad \Leftrightarrow \quad p_0 \in \Gamma_q \quad \Leftrightarrow \quad p_0 \neq 0.$$

There is also a canonical matrix representation of elements of X_{q^v} via Toeplitz matrices. For any $p = \sum_{i=0}^{v-1} p_i q^i$ set

$$T(p) := \begin{pmatrix} p_0 & & & \\ p_1 & \cdot & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ p_{v-1} & & p_1 & p_0 \end{pmatrix}.$$

A simple computation shows that

$$T(pp') = T(p)T(p')$$

with multiplication of elements p_i, p'_j understood to take place in X_q. The above Toeplitz matrices form an algebra over X_q and this matrix algebra is isomorphic to X_{q^v}.

As another consequence of the previous theorem we deduce.

Corollary 3.5. *Let $q = q_1 \cdots q_r$ be a factorization with pairwise coprime polynomials q_i, and let $t_1,\dots,t_r$ be factors of $q_1,\dots,q_r$, respectively. Then the isomorphism of algebras*

$$\phi : X_{q_1} \times \cdots \times X_{q_r} \to X_q$$

$$\phi(p_1,\dots,p_r) = \sum_{i=1}^{r} \pi_i p_i,$$

restricts to an algebraic isomorphism of orbits

$$\phi : \Gamma_{q_1}(t_1) \times \cdots \times \Gamma_{q_r}(t_r) \to \Gamma_q(t_1 \cdots t_r).$$

Proof. Let A and B be algebras with A', B' the groups of units. For the A', B' actions on A and B, respectively we have

$$\phi(g\cdot x)=\phi(g)\phi(x),$$

for any $g\in A'$, $x\in A$. Thus, ϕ maps A'-orbits bijectively onto B'-orbits. The result follows. □

In particular, we see for the Zariski closures of $\Gamma_{q_i}(t_i)$ the isomorphism

$$\overline{\Gamma_q(t_1\cdots t_r)}\simeq\overline{\Gamma_{q_1}(t_1)}\times\cdots\times\overline{\Gamma_{q_r}(t_r)}.$$

If $F=\mathbb{C}$, it is easily seen that the Zariski closure of an orbit coincides with the usual topological closure. For later reference we state now the result over $\mathbb{C}$.

Corollary 3.6. *Let $F=\mathbb{C}$ and let $q=ts$. Then with respect to the usual euclidean topology we have*

1. *Each orbit $\Gamma_q\cdot t$ is a complex submanifold of $X_q\simeq\mathbb{C}^n$ of complex dimension* $\dim\Gamma_q\cdot t=\deg q-\deg t$.
2. *The closure of each orbit $\Gamma_q\cdot t$ is $\overline{\Gamma_q\cdot t}=tX_s$ which is a shift-invariant subspace of X_q.*
3. *For any factorization $q=q_1\cdots q_r$ by pairwise coprime polynomials and factors t_i of q_i, $i=1,\dots,r$, we have the homeomorphisms*

$$\Gamma_q(t_1\cdots t_r)\simeq\Gamma_{q_1}(t_1)\times\cdots\times\Gamma_{q_r}(t_r),$$

$$\overline{\Gamma_q(t_1\cdots t_r)}\simeq\overline{\Gamma_{q_1}(t_1)}\times\cdots\times\overline{\Gamma_{q_r}(t_r)}.$$

4. The general case

We return now to our investigation of the reachable sets for the inverse power iterations. So far, we have focussed on cyclic matrices. Before proving our main results we therefore extend first the analysis to arbitrary matrices.

Let $A\in F^{n\times n}$ denote an arbitrary matrix and $x_0\in F^n$. Then

$$\langle A\,|\,x_0\rangle:=\operatorname{span}_F\{A^ix_0\mid i\in\mathbb{N}\}$$

is the smallest A-invariant subspace of F^n containing x_0. Certainly, x_0 is a cyclic vector for the restriction $\bar{A}:=A|_{\langle A|x_0\rangle}$ of A to $\langle A|x_0\rangle$. Thus,

$$\bar{A}:\langle A\,|\,x_0\rangle\to\langle A\,|\,x_0\rangle$$

is cyclic with cyclic vector x_0. The proof of the next lemma is trivial and therefore omitted.

Lemma 4.1. *Let $A\in F^{n\times n}$, $x_0\in F^n$ and $X_0:=[x_0]\in\mathbb{P}(F^n)$. Then $\mathscr{R}_A(X_0)=\mathscr{R}_{\bar{A}}(X_0)$ and $\Gamma_A\cdot X_0=\Gamma_{\bar{A}}\cdot X_0$. In particular, for an algebraically closed field F, we have*

$$\mathscr{R}_A(X_0)=\Gamma_{\bar{A}}\cdot X_0,$$

for all $X_0\in\mathbb{P}(F^n)$.

Thus, for F algebraically closed, the reachable sets $\mathscr{R}_A(X_0)$ are orbits of the cyclic operator $\bar{A}$ and we can apply the results of the previous section.

Theorem 4.1. *Let F be an algebraically closed field, $A\in F^{n\times n}$, $x_0\in F^n$ and $X_0=[x_0]\in\mathbb{P}(F^n)$.*

1. *Every reachable set $\mathscr{R}_A(X_0)$ is a quasiprojective subvariety of $\mathbb{P}(F^n)$ of dimension*

$$\dim\mathscr{R}_A(X_0)=\dim\langle A\,|\,x_0\rangle-1.$$

2. *For the Zariski closure $\overline{\mathscr{R}_A(X_0)}$ we have*

$$\overline{\mathscr{R}_A(X_0)}=\mathbb{P}(\langle A|x_0\rangle).$$

3. *For $X_0,X_1\in\mathbb{P}(F^n)$ we have*

$$\mathscr{R}_A(X_0)\cap\overline{\mathscr{R}_A(X_1)}\neq\emptyset\;\Leftrightarrow\;\mathscr{R}_A(X_0)\subset\overline{\mathscr{R}_A(X_1)}\;\Leftrightarrow\;\langle A\,|\,x_0\rangle\subset\langle A|x_1\rangle.$$

Proof. By the above lemma we have $\mathscr{R}_A(X_0)=\mathscr{R}_{\bar{A}}(X_0)$ and therefore it suffices to prove the result for cyclic linear operators A, and hence for the shift operator S_q on X_q. Passing from a vector space V to the projective space $\mathbb{P}(V)$ reduces the dimension by 1. The result follows immediately from the corresponding result in Section 3, i.e. from Theorem 3.5. □

Corollary 4.1. *The inverse power method* (3) *on $\mathbb{P}(F^n)$ is almost controllable if and only if A is cyclic.*

The next result is an immediate consequence of Corollary 3.2.

Theorem 4.2. *Let F be an algebraically closed field, $A\in F^{n\times n}$, $x_0\in F^n$ and $X_0=[x_0]\in\mathbb{P}(F^n)$. There is a bijective correspondence between*

1. *closures $\overline{\mathscr{R}_A(X_0)}$ of reachable sets,*
2. *cyclic A-invariant subspaces of F^n.*

In the cyclic case we can be a bit more specific.

Theorem 4.3. *Let F be an algebraically closed field, $A\in F^{n\times n}$, $x_0\in F^n$ and $X_0=[x_0]\in\mathbb{P}(F^n)$. There is a bijective correspondence between*

1. *closures $\overline{\mathscr{R}_A(X_0)}$ of reachable sets,*

2. *A-invariant subspaces of* F^n,
3. *factors of the characteristic polynomial of* A.

Proof. Immediate consequence of Corollary 3.2. □

Corollary 4.2. *Let* $A \in F^{n\times n}$ *be cyclic*, F *an algebraically closed field, and let* $q = q_1^{\nu_1} \cdots q_r^{\nu_r}$ *be the primary decomposition of the characteristic polynomial of* A. *Then*,
1. *There are only a finite number of reachable sets of* (3).
2. *There is a unique reachable set that is open and dense in* $\mathbb{P}(F^n)$.
3. *There are exactly r Zariski closed reachable sets, and these are the points in* $\mathbb{P}(F^n)$ *defined by the one-dimensional invariant subspaces* V_j *such that* $A \mid V_j$ *has characteristic polynomial* q/q_j.

There is an elegant way to characterize the lattice of A-invariant subspaces and hence the closure relations of reachable sets of (3) by the combinatorics of faces of the standard simplex. To this end let $F = \mathbb{C}$ and assume that $A \in \mathbb{C}^{n\times n}$ has n distinct eigenvalues. Without loss of generality we assume $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, with $\lambda_i \neq \lambda_j$. Let

$$\Delta_n := \left\{ (t_1, \ldots, t_n) \in \mathbb{R}^n | t_i \geqslant 0,\ \sum_{i=1}^n t_i = 1 \right\}$$

denote the standard $(n-1)$-dimensional simplex in $\mathbb{R}^n$. Consider the map

$$\mu : \mathbb{P}(\mathbb{C}^n) \to \Delta_n$$

expressed in homogeneous coordinates by

$$\mu([x_1 : \cdots : x_n]) := (t_1, \ldots, t_n),$$

$$t_j := \frac{|x_j|^2}{|x_1|^2 + \cdots + |x_n|^2}, \quad j = 1, \ldots, n.$$

Then μ is a surjective map that maps each invariant subspace of A onto a subsimplex of Δ_n. Thus, the k-dimensional invariant subspaces of A correspond bijectively via μ to the k-dimensional subsimplices of Δ_n, $k = 0, \ldots, n$. In particular, the unique open and dense reachable set of (3) is mapped onto the interior of Δ_n, while the one-dimensional eigenspaces of A correspond to the n vertices of Δ_n. We have

$$\overline{\mathcal{R}_A(X_0)} \subset \overline{\mathcal{R}_A(X_1)} \quad \Leftrightarrow \quad \mu(\overline{\mathcal{R}_A(X_0)}) \subset \mu(\overline{\mathcal{R}_A(X_1)})$$

so that the correspondence preserves the inclusion order on reachable sets and subsimplices, respectively.

Example. For $n = 3, 4$ there are exactly 7, 15 invariant subspaces of A that correspond bijectively to the 7, 15 subsimplices of Δ_3, Δ_4, respectively.

References

[1] S. Batterson, J. Smilie, The dynamics of Rayleigh quotient iteration, SIAM J. Numer. Anal. 26 (1989) 624–636.
[2] P.A. Fuhrmann, Algebraic system theory: An analyst's point of view, J. Franklin Inst. 301 (1976) 521–540.
[3] P.A. Fuhrmann, A Polynomial Approach to Linear Algebra, Springer, New York, 1996.
[4] G.H. Golub, C.F. Van Loan, Matrix computations, 2nd Edition, The John Hopkins University Press, Baltimore, 1989.
[5] R.E. Kalman, P.L. Falb, M.A. Arbib, Topics in Mathematical System Theory, McGraw-Hill, New York, 1969.
[6] B.N. Parlett, The Rayleigh quotient iteration and some generalizations for nonnormal matrices, Math. Comp. 28 (1974) 679–693.
[7] R. Hartshorne, Algebraic Geometry, Springer, New York, 1977.