ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 29, Number 1, Spring 1999

CONSTRAINED CONVERGENCE

ROBERT E. HARTWIG AND PETER ŠEMRL

ABSTRACT. The convergence of PA^NQ is investigated. The results are then used to obtain information about the convergence of constrained Picard iteration $Y_N = PX_N$, where $X_{N+1} = AX_N + B$. In particular, it is shown that, for given P, A and B there exists an initial condition $X_0 = C$ for which Y_N converges, exactly when $R[\vartheta B] \subseteq R[\vartheta(I-A)]$, where $\vartheta = [P^T, A^TP^T, \dots, (A^{m-1})^TP^T]^T$ and m is the degree of the minimal polynomial of A.

1. Introduction. One of the most basic iterations in matrix theory is the Picard iteration (PI) [**1**]:

$$X_{N+1} = AX_N + B \quad \text{with} \quad X_0 = C, \tag{1.1}$$

where A, B and C are constant complex matrices and A is $n \times n$. In practice, however, it may be that only the constrained matrix $Y_N = PX_N$ is "observable." Since the PI iterations admits the exact solution

$$X_N = \left[\sum_{i=0}^{N-1} A^i\right]B + A^NC, \tag{1.2}$$

we see that

$$Y_N = P\left[\sum_{i=0}^{N-1} A^i\right]B + PA^NC, \tag{1.3}$$

and, hence, that the convergence of both $S_N = P[\sum_{i=0}^{N-1} A^i]B$ as well as that of PA^NC (as $N \to \infty$) ensures that of Y_N. The converse may not be true, however, as seen by taking $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $C = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$ and $P = I$. On the other hand, if we set $U = B - (I - A)C$, then

$$\begin{aligned} P\left[\sum_{i=0}^{N-1} A^i\right]U &= P\left[\sum_{i=0}^{N-1} A^i\right]B - P(I - A^N)C \\ &= PX_N - PC, \end{aligned} \tag{1.4}$$

Received by the editors on May 20, 1997.
AMS (MOS) *Mathematics Subject Classification.* 15A24, 15A09, 56F10.

which shows that $Y_N = PX_N$ converges exactly when $P\left[\sum_{i=0}^{N-1} A^i\right]U$ converges.

In this note we shall therefore examine the convergence of

(i) the constrained powers PA^NQ and

(ii) the constrained sum $S_N = P\left[\sum_{i=0}^{N-1} A^i\right]Q$,

as $N \to \infty$ for general matrices P and Q. The case where P has a left inverse P^- or $PA = AP$ can be investigated entirely by means of polynomial ideals [**1**]. We shall refer to this case as the *one sided case*. Needless to say, the one-sided case could also mean that Q has a right inverse or commutes with A. Unfortunately, however, this theory only partially extends to the case of a two-sided constraint, and linear independence cannot be used. This is not really surprising since we must be able to use simultaneously global (all matrices are of full size) as well as local (P^T and Q are $n \times 1$) convergence, when we are at the entry level.

We are interested in computing expressions of the form $Pf_N(A)Q$ for some polynomial $f_N(\lambda)$. This suggests that we consider the vector space

$$\Gamma = \operatorname{span}\{PA^iQ; i = 0, 1, 2, \dots\}.$$

In general, some of these will vanish or may be linearly dependent. We could construct a basis for Γ simply by starting with the *lowest* nonzero link in the chain $\{PQ, PAQ, PA^2Q, \dots\}$, say $PA^{i_1}Q$, after which we pick $PA^{i_2}Q$ as the next link that is independent of the first one, followed by $PA^{i_3}Q$ as the first link that is independent of the previous two links, and so on. This gives us a basis

$$\mathcal{B}_\Gamma = \{PA^{i_1}Q, \dots, PA^{i_d}Q\}, \quad i_1 < i_2 < \cdots < i_d, \tag{1.5}$$

where $d = \dim(\Gamma)$. Clearly $d \le d_p, d_q$, which are the respective dimensions of the Krylov spaces spanned by

$$\{P, PA, PA^2, \dots\} \quad \text{and} \quad \{Q, AQ, A^2Q, \dots\}.$$

In general the powers of A in a basis for Γ need not be unique nor consecutive, as seen from the case where PAQ is a scalar, i.e., when P^T and Q are $n \times 1$ column vectors. Now, for any polynomial $f(A)$, we have

$$Pf(A)Q = \sum_{k=1}^{d} \alpha_k(f) PA^{i_k} Q, \tag{1.6}$$

and, in particular,

$$PA^N Q = \sum_{k=1}^{d} \alpha_k^{(N)} PA^{i_k} Q. \tag{1.7}$$

But, as always, the problem is to find the coefficients $\alpha_k^{(N)}$, which may be prohibitive. In the one-sided case, the key to finding the coefficients is to use annihilating polynomials and the associated ideal theory. To extend these concepts to the two sided case, however, is a nontrivial task, which we now shall undertake.

Suppose that the minimal polynomial of A has the form

$$\psi_A(\lambda) = \prod_{i=1}^{s} (\lambda - \lambda_i)^{m_i} \quad \text{with degree } m. \tag{1.8}$$

The ideal that we shall need is the vector space

$$W = \{f(\lambda); PA^i f(A)Q = 0 \text{ for } i = 0, 1, \dots, m-1\}. \tag{1.9a}$$

It is easily seen to be an ideal if we rewrite W as

$$W = \{f(\lambda); Pg(A)f(A)Q = 0 \text{ for all } g(\lambda) \in \mathbf{C}[\lambda]\}. \tag{1.9b}$$

We shall further need the following characterization which holds for each fixed value of scalar α:

$$W = \{f(\lambda); P(A - \alpha I)^i f(A)Q = 0 \text{ for } i = 0, 1, \dots, m-1\}, \tag{1.9c}$$

which is an immediate consequence of the binomial theorem.

We note that $W \subseteq K = \{f(\lambda); Pf(A)Q = 0\}$, which is a vector space but will not be an ideal except in very special cases such as where P (Q) has a left (right) inverse or commutes with A.

Since $\mathbf{C}[\lambda]$ is a principal ideal domain, we know there exists a monic generator for W, say $\phi(\lambda) = \phi_{PQ}(\lambda)$ which necessarily must divide $\psi_A(\lambda)$. Apart from the case where $\phi(\lambda) = 1$, we can without loss of generality assume that ϕ has the form

$$\phi(\lambda) = \prod_{i=1}^{t} (\lambda - \lambda_i)^{p_i} \quad \text{with degree } l, \tag{1.10}$$

where $t \le s$ and $1 \le p_i \le m_i$. We may think of ϕ as the "minimal annihilating polynomial" relative to the two spaces, $R(Q)$ and $RS(P)$. The case where $\phi = 1$ can only trivially happen in the one sided case. For notational convenience we shall also exclude it in the discussion that follows. The final results, however, will also hold in this special case. We now stress that

$$\begin{aligned} f(\lambda) \in W \quad \Longleftrightarrow \quad \phi \mid f \quad \Longleftrightarrow \quad & PA^i f(A) Q = 0 \\ & \text{for } i = 0, 1, \dots, m-1, \end{aligned} \tag{1.11}$$

while
(1.12)

$$\begin{aligned} h(\lambda) \notin W \quad \Longleftrightarrow \quad \phi \nmid h \quad & \Longleftrightarrow \quad PA^i h(A) Q \neq 0 \quad \text{for some } i \\ & \Longleftrightarrow \quad P(A - \alpha I)^j h(A) Q \neq 0 \quad \text{for some } j. \end{aligned}$$

For convenience, we define the *effective* spectrum of A, relative to P and Q, by

$$\tilde{\sigma} = \sigma_{PQ}(A) = \{\lambda_k ; (\lambda - \lambda_k) \mid \phi(\lambda)\} \tag{1.13}$$

and call the numbers p_i the P-Q *effective* indices of A.

Needless to say, we can also speak of the *effective* spectral radius

$$\tilde{\rho} = \rho_{PQ} = \max\{|\lambda_k| ; \lambda_k \in \tilde{\sigma}\}. \tag{1.14}$$

We may now capitalize on our new ideal. Indeed, if $PA^i f(A) Q = PA^i g(A) Q$ for all i, then $PA^i[f(A) - g(A)]Q = 0$ for all i and $\phi|(f-g)$. This can again be characterized via differentiation.

Let us next introduce the spectral components, which are nothing but a convenient way of writing down the powers of the Jordan form.

2. The spectral components. A convenient way of handling convergence is by using the spectral theorem [**2**]:

$$f(A) = \sum_{k=1}^{s} \sum_{j=0}^{m_k - 1} f^{(j)}(\lambda_k) Z_k^j, \tag{2.1}$$

where the spectral components $\{Z_k^j; k = 1, \ldots, s, j = 0, \ldots, m_k - 1\}$ form a canonical basis for $\langle I, A, A^2, \ldots \rangle$, the span of the powers of A.

Let us next consider the constrained components PZ_k^jQ, $k = 1, \ldots, s$, $j = 0, \ldots, m_k - 1$. We shall first show that for ineffective eigenvalues the constrained components drop out. And as such we can speak of the effective spectral components.

Lemma 2.1. *If $\lambda_r \notin \tilde{\sigma}$, then $PA^iZ_r^jQ = 0$ for all $i, j = 0, 1, \ldots$.*

Proof. Recall that, if $\psi_A(\lambda) = (\lambda - \lambda_k)^{m_k} \cdot \Psi_k(\lambda)$, then $\gcd(\Psi_1(\lambda), \ldots, \Psi_s(\lambda)) = 1$ and hence that there exist $g_k(\lambda)$ such that $1 = g_1(\lambda)\Psi_1(\lambda) + \cdots + g_s(\lambda)\Psi_s(\lambda)$. It is well known that $Z_k^0 = h_k(A)$ where $h_i(\lambda) = g_i(\lambda)\Psi_i(\lambda)$. Now, for $\lambda_r \notin \tilde{\sigma}$, i.e., for $r > t$, we see that $\phi \mid \Psi_r(\lambda)$ and thus $\phi \mid (\lambda - \lambda_r)^j h_r(\lambda)$. This means that $PA^iZ_r^jQ = 0$ for all i. In particular, $PZ_r^jQ = 0$. □

We next turn to the effective spectral components.

Lemma 2.2. *If $\lambda_k \in \tilde{\sigma}$, then $PA^iZ_k^{p_k}Q = 0 \neq PZ_k^{p_k - 1}Q$ for all i.*

Proof. Again we claim that $\phi | (\lambda - \lambda_k)^{p_k} \cdot h_k(\lambda)$, since $(\lambda - \lambda_r)^{p_r} | (\lambda - \lambda_r)^{m_r} | \Psi_k$ for $r \neq k$. Since $Z_k^{p_k} = (A - \lambda_k I)^{p_k} \cdot Z_k^0 / p_k!$ we see that $PA^iZ_k^{p_k}Q = 0$ for all i. Next observe that $\phi \nmid (\lambda - \lambda_k)^{p_k - 1} \cdot h_k(\lambda)$, and hence $(\lambda - \lambda_k)^{p_k - 1} \cdot h_k(\lambda) \notin W$. This means that, for some i, $PA^i(A - \lambda_k I)^{p_k - 1} \cdot Z_k^0 Q \neq 0$, or equivalently that some $P(A - \lambda_k I)^i (A - \lambda_k I)^{p_k - 1} \cdot Z_k^0 Q \neq 0$ for some i. But, by the first part, we know that the terms with $i = 1, 2, \ldots,$ all vanish. Consequently, the only term that can be nonzero is $P(A - \lambda_k I)^{P_k - 1} \cdot Z_k^0 Q = (p_k - 1)! \cdot PZ_k^{p_k - 1}Q \neq 0$. We shall refer to these nonzero components as the "maximal links" corresponding to λ_k.

Since the spectral components $\{Z_k^j; k = 1, \ldots, s, j = 0, \ldots, m_k - 1\}$ form a basis for $\langle I, A, A^2, \ldots \rangle$, we know that the constrained components $\{PZ_k^jQ\}$ span Γ. Now, based on the above two lemmas, we see that in this list only effective components survive and that, on the other hand, we only know that inside the λ_k-block $\{PZ_k^0Q, PZ_k^1Q, \ldots, PZ_k^{p_k-1}Q\}$ of effective constrained components corresponding to λ_k, only the highest term, $PZ_k^{p_k-1}Q$, is guaranteed to be nonzero. Some of the lower terms may vanish! Moreover, the nonzero terms may be linearly dependent. This is in stark contrast to the one sided case where this cannot happen. Indeed, in the scalar case, where both P^T and Q are single columns, each block is a string of scalars, which must be dependent if there is a second nonzero term besides the last term. On account of this, it seems difficult to find $\dim \Gamma$ using the spectral components.

Example. If $A = \begin{bmatrix} 6 & 2 & 2 \\ -2 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, then $\psi = (\lambda - 4)^2(\lambda - 2)$ and, with $\lambda_1 = 4$ and $\lambda_2 = 2$, the spectral components are

$$Z_1^0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

$$Z_1^1 = \begin{bmatrix} 2 & 2 & 2 \\ -2 & -2 & -2 \\ 0 & 0 & 0 \end{bmatrix},$$

and

$$Z_2^0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Now, if $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $Q = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, then $PZ_1^0Q = 0$, yet $PZ_1^1Q = \begin{bmatrix} 0 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq 0$ and $PZ_2^0Q = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \neq 0$. The maximal links are independent here.

We do know that the set $\{PZ_k^jQ\}$ contains a basis for Γ, of the form

$$\{PZ_1^{i_1}Q, PZ_1^{i_2}Q, \ldots, PZ_1^{i_r}Q | PZ_2^{j_1}Q, PZ_2^{j_2}Q, \ldots, PZ_2^{j_u}Q | \ldots | PZ_t^{k_1}Q, \ldots\}, \tag{2.2}$$

which only contains effective components. On account of the scalar case, where P^T and Q are single columns, we observe that

(i) a basis need not contain the highest links $PZ_k^{p_k-1}Q \neq 0$.

(ii) Different bases can use links from different blocks, and the "holes" need not match.

The conclusion is that there is no point in expanding $Pf(A)Q$ in terms of this basis, since we would not know how to characterize the coefficients.

In any case we may expand a given function $f_N(\lambda)$ again as

$$Pf_N(A)Q = \sum_{k=1}^{s} \sum_{j=0}^{m_k-1} f_N^j(\lambda_k) \cdot (PZ_k^jQ), \tag{2.3}$$

which, when dropping ineffective eigenvalues, reduces to

$$Pf_N(A)Q = \sum_{k=1}^{t} \sum_{j=0}^{p_k-1} f_N^j(\lambda_k) \cdot (PZ_k^jQ). \tag{2.4}$$

In particular,

$$PA^NQ = \sum_{k=1}^{t} \sum_{j=0}^{p_k-1} \mathbf{D}^j[\lambda^N]\Big|_{\lambda=\lambda_k} \cdot (PZ_k^jQ), \tag{2.5}$$

and

$$P\left[\sum_{i=0}^{N-1} A^i\right]Q = \sum_{i=0}^{N-1} \sum_{k=1}^{t} \sum_{j=0}^{p_k-1} \mathbf{D}^j[\lambda^i]\Big|_{\lambda=\lambda_k} \cdot (PZ_k^jQ). \tag{2.6}$$

If the constrained components PZ_k^jQ were linearly independent, then the convergence of $Pf_N(A)Q$ immediately reduces to a study of the coefficients $f_N^j(\lambda_k)$. In our case, however, this cannot be done.

Let us now tackle this problem in its simplest form by first considering the bilinear forms $\underline{x}^T A^N \underline{y}$ and $\sum_{i=0}^{N-1} \underline{x}^T A^i \underline{y}$, as $N \to \infty$.

The method that we shall use is a variation of that used in interpolation and coding [**4**].

3. A special case. It is clear that PA^NQ converges if and only if $\underline{p}_i^T A^N \underline{q}_j$ converges for all i and j, where $\underline{p}_i^T$ and $\underline{q}_j$ are the ith row and jth column in P and Q, respectively. As such, let us first focus on the bilinear form $\underline{x}^T A^N \underline{y}$. From the spectral theorem, we know that

$$\underline{x}^T A^N \underline{y} = \sum_{k=1}^{s} \sum_{j=0}^{m_k-1} \mathbf{D}^j[\lambda^N]\Big|_{\lambda=\lambda_k} (\underline{x}^T Z_k^j \underline{y}), \tag{3.1}$$

which we may write as

$$\underline{x}^T A^N \underline{y} = \sum_{k=1}^{s} \sum_{j=0}^{m_k-1} \binom{N}{j} \lambda_k^{N-j} \cdot \beta_k^j, \tag{3.2}$$

where $\beta_k^j = (j!\underline{x}^T Z_k^j \underline{y})$. It is clear from this that the zero eigenvalue, say λ_2, does not contribute to this sum, as long as $N \geq m_2$. Therefore, in what follows, we shall assume that the matrix is invertible.

To avoid the dependency of β_k^j on $\underline{x}$ and $\underline{y}$, we shall keep our summation running from $k = 1, \ldots, s$ and $j = 0, \ldots, m_{k-1}$, even though some of the terms will vanish.

Now $\underline{x}^T A^N \underline{y}$ converges as $N \to \infty$ if and only if $\{\underline{x}^T A^N \underline{y}, \underline{x}^T A^{N+1} \underline{y}, \ldots, \underline{x}^T A^{N+m-1} \underline{y}\}$ all converge as $N \to \infty$. Before we write this in matrix form, let us first define the matrices

$$\underline{b}_k = [\beta_k^0, \beta_k^1, \ldots, \beta_k^{m_k-1}]^T = 1, \ldots, s.$$

We also need the $m \times m_k$ matrix

$$M_k^{(N)} = M(\lambda_k, N, m_k) = \begin{bmatrix} R(\lambda_k, N) \\ R(\lambda_k, N+1) \\ \vdots \\ R(\lambda_k, N+m-1) \end{bmatrix}_{(m \times m_k)}, \tag{3.3}$$

where

(3.4)

$$R(\lambda_k, N) = \left[\lambda_k^N \cdot \binom{N}{0}, \lambda_k^{(N-1)} \cdot \binom{N}{1}, \ldots, \lambda_k^{(N-m_k+1)} \cdot \binom{N}{m_k-1} \right].$$

Further, set

$$\begin{aligned} M^{(N)} &= [M(\lambda_1, N, m_1), M(\lambda_2, N, m_2); \dots , M(\lambda_s, N, m_s)] \\ &= [M_1^{(N)}, \dots , M_s^{(N)}]. \end{aligned}$$

Using these matrices, we may stack the terms $\underline{x}^T A^{N+i} \underline{y}$ as

$$\begin{bmatrix} \underline{x}^T A^N \underline{y} \\ \underline{x}^T A^{N+1} \underline{y} \\ \vdots \\ \underline{x}^T A^{N+m-1} \underline{y} \end{bmatrix} = M^{(N)} \begin{bmatrix} \underline{b}_1 \\ \underline{b}_2 \\ \vdots \\ \underline{b}_s \end{bmatrix} = M^{(N)} \underline{b}. \tag{3.5}$$

Thus $\underline{x}^T A^N \underline{y}$ converges if and only if $M^{(N)} \underline{b}$ converges. The aim of what follows is to show that we may uncouple the eigenvalue dependency and, as such, solve the convergence problem.

Our first step is to simplify $M(\lambda_k, N, m_k)$ and to factor out the powers of λ_k. This gives

$$M(\lambda_k, N, m_k) = F(k, m) \cdot B(m, m_k, N) \cdot (\lambda_k^N \cdot F_k^{-1}), \tag{3.6}$$

where

$$F(k, N) = \operatorname{diag}(1, \lambda_k, \dots , \lambda_k^{N-1}), \quad F_k = F(k, m_k) \tag{3.7}$$

and

$$B_k = B(m, m_k, N) = \begin{bmatrix} R(1, N) \\ R(1, N+1) \\ \vdots \\ R(1, N+m-1) \end{bmatrix}_{(m \times m_k)}. \tag{3.8}$$

We now recall the combinatorial identity

$$\begin{aligned} \binom{N+k}{r} &= \binom{k}{0}\binom{N}{r} + \binom{k}{1}\binom{N}{r-1} + \binom{k}{2}\binom{N}{r-2} + \cdots \\ &\quad + \begin{cases} \binom{k}{r}\binom{N}{0} & \text{if } k \geq r, \\ \binom{k}{k}\binom{N}{r-k} & \text{if } k \leq r, \end{cases} \end{aligned} \tag{3.9}$$

obtained by drawing r objects from a set of $N+k$ elements. These identities may be written in matrix form as

$$(3.10)\qquad \binom{N+k}{r} = \left[\binom{k}{0},\binom{k}{1},\dots,\binom{k}{k},0,\dots,0\right]\begin{bmatrix}\binom{N}{r}\\ \binom{N}{r-1}\\ \vdots\\ \binom{N}{0}\\ 0\\ \vdots\\ 0\end{bmatrix},$$

which then precisely gives the LU factorization of the $(N+1)\times(N+1)$ matrix $B(N+1,N+1,N)$. Indeed,

$$(3.11)\qquad \begin{bmatrix}\binom{N}{0} & \binom{N}{1} & \cdots & \binom{N}{N}\\ \binom{N+1}{0} & \binom{N+1}{1} & & \cdot\\ \binom{N+2}{0} & \binom{N+2}{1} & \cdots & \cdot\\ \vdots & & & \vdots\\ \binom{2N}{0} & \cdots & \cdots & \binom{2N}{N}\end{bmatrix}$$

$$= \begin{bmatrix}\binom{0}{0} & & & \\ \binom{1}{0} & \binom{1}{1} & & O\\ \binom{2}{0} & \binom{2}{1} & \binom{2}{2} & \\ \vdots & & \ddots & \\ \binom{N}{0} & \binom{N}{1} & \cdots & \binom{N}{N}\end{bmatrix}\begin{bmatrix}\binom{N}{0} & \binom{N}{1} & \binom{N}{2} & \cdots & \binom{N}{N}\\ 0 & \binom{N}{0} & \binom{N}{1} & \cdots & \\ & & \binom{N}{0} & \cdots & \\ O & & & \ddots & \vdots\\ & & & & \binom{N}{0}\end{bmatrix}.$$

From (3.11), we immediately obtain, for $N \geq m$ as a submatrix product, the nonsquare LU factorization of $B(m,m_k,N)$ as $L_k T_k^{(N)}$, in which B_k and L_k are $m\times r$ and $T_k^{(N)}$ is $r\times r$, where $m_k = r$.

Moreover, L_k is independent of N. Consequently,

$$(3.12)\quad \begin{bmatrix} \binom{N}{0} & \binom{N}{1} & \cdots & \binom{N}{r-1} \\ \binom{N+1}{0} & \binom{N+1}{1} & \cdot & \\ \binom{N+2}{0} & \binom{N+2}{1} & \cdot & \\ \vdots & & & \vdots \\ \binom{N+m-1}{0} & \cdots & \cdots & \binom{N+m-1}{r-1} \end{bmatrix}$$

$$= \begin{bmatrix} \binom{0}{0} & & & & \\ \binom{1}{0} & \binom{1}{1} & & O & \\ \binom{2}{0} & \binom{2}{1} & \binom{2}{2} & & \\ \vdots & & & \ddots & \binom{r-1}{r-1} \\ \vdots & & & & \vdots \\ \binom{m-1}{0} & \binom{m-1}{1} & \cdots & \cdots & \binom{m-1}{r-1} \end{bmatrix}$$

$$\cdot \begin{bmatrix} \binom{N}{0} & \binom{N}{1} & \binom{N}{2} & \cdots & \binom{N}{r-1} \\ 0 & \binom{N}{0} & \binom{N}{1} & \cdot & \cdot \\ & & \binom{N}{0} & \cdots & \cdot \\ & O & & \ddots & \\ & & & & \binom{N}{0} \end{bmatrix},$$

and we arrive at $M^{(N)}\underline{b} = [F(1,m)B_1, F(2,m)B_2, \dots, F(s,m)B_s] \cdot \operatorname{diag}(\lambda_1^N F_1^{-1}, \dots, \lambda_s^N F_s^{-1})\underline{b} = [F(1,m)L_1, F(2,m)L_2, \dots, F(s,m)L_s] \cdot \operatorname{diag}(\lambda_1^N T_1 F_1^{-1}, \dots, \lambda_s^N T_s F_s^{-1})\underline{b}$. The key observation now is that $F(k,m)L_k = \Omega_k F_k$, where, for $i \geq j$,

$$(3.13)\qquad (\Omega_k)_{ij} = \binom{i}{j}\lambda_k^{i-j}, \quad i = 0, \dots, m-1,\ j = 0, \dots, m_k - 1,$$

and zero otherwise. We then have

$$M^{(N)}\underline{b} = [\Omega_1, \dots, \Omega_s] \cdot \operatorname{diag}(F_1, \dots, F_s) \times \operatorname{diag}(\lambda_1^N T_1 F_1^{-1}, \dots, \lambda_s^N T_s F_s^{-1})\underline{b}.$$

The matrix $[\Omega_1, \dots, \Omega_s]$ is (a multiple of) the confluent Vandermonde matrix (or Wronskian) [**3**] and as such is constant and invertible. Likewise, the matrix $\operatorname{diag}(F_1, \dots, F_s)$ is independent of N and invertible since the zero eigenvalue had been excluded earlier.

Hence, we have uncoupled the convergence of $M^{(N)}\underline{b}$ into the convergence of each of the matrix sequences:

$$G_k^{(N)} = T_k(\lambda_k^N F_k^{-1})\underline{b}_k, \quad k = 1, \dots, s. \tag{3.14}$$

Setting $\lambda_k = \alpha$, $m_k = r$ and $\underline{b}_k = [b_0, b_1, \dots, b_{r-1}]^T$, we have

$$G_k^{(N)} = \begin{bmatrix} \binom{N}{0} & \binom{N}{1} & \binom{N}{2} & \cdots & \binom{N}{r-1} \\ 0 & \binom{N}{0} & \binom{N}{1} & \cdot & \cdot \\ & & \binom{N}{0} & \cdots & \cdot \\ & O & & \ddots & \vdots \\ & & & & \binom{N}{0} \end{bmatrix} \cdot \begin{bmatrix} \alpha^N & & & & \\ 0 & \alpha^{N-1} & & & O \\ & & \alpha^{N-2} & & \\ & & & \ddots & \\ & & & & \alpha^{N-r+1} \end{bmatrix} \begin{bmatrix} b_0 \\ \vdots \\ b_{r-1} \end{bmatrix}. \tag{3.15}$$

Let us now analyze the various cases that can occur. From Lemma (2.1) and (2.2), we recall that $\underline{b}_k \neq \underline{0}$, $k = 1, \dots, s$, exactly when λ_k is an effective eigenvalue with respect to $\underline{x}$ and $\underline{y}$, and that in particular $\beta_k^{p_k-1} \neq 0$ where $p_k = p(\underline{x}, \underline{y})$. As such, we focus on an effective eigenvalue λ_k with effective index p_k. There are two cases that can occur.

Case 1. $p_k = 1$. (λ_k has effective index 1.) In this case $b_0 = \beta_k^0 \neq 0$, and from (3.15), we see that $(G_k^{(N)})_{11} = \lambda_k^N b_0$. We may now conclude that

(i) $(G_k^N)_{11}$ converges if and only if either $|\lambda_k| < 1$ or $\lambda_k = 1$.

(ii) On the other hand, $(G_k^{(N)})_{11} = \lambda_k^N b_0$ converges to zero if and only if $|\lambda_k| < 1$.

Case 2. $p_k > 1$. The p_kth entry in $G_k^{(N)}$ is $\lambda_k^{N-p_k+1} \cdot \beta_k^{p_k-1}$. Now, since $\beta_k^{p_k-1} \neq 0$, we see that if this converges then either $|\lambda_k| < 1$ or $\lambda_k = 1$. But, if $|\lambda_k| < 1$, then all the remaining terms automatically converge to zero. On the other hand, if $\lambda_k = 1$, then the convergence of the $(p_k - 1)$st entry in $G_k^{(N)}$ shows that

$$\beta_k^{p_k-2} + N\beta_k^{p_k-1} \quad \text{converges with} \quad \beta_k^{p_k-1} \neq 0.$$

This is impossible, and hence this case cannot occur. We may conclude that, in Case 2,

$$G_k^{(N)} \quad \text{converges} \quad \Longleftrightarrow \quad |\lambda_k| < 1 \quad \Longleftrightarrow \quad \text{it converges to zero.}$$

Combining the two cases, we see that for each effective λ_k:

(a) $G_k^{(N)}$ converges if and only if either $|\lambda_k| < 1$ or $\lambda_k = 1$ with $p_k = 1$.

(b) $G_k^{(N)}$ converges to zero if and only if $|\lambda_k| < 1$.

We stress again that we did not make use of the possible linear independence of the nonzero effective constrained components. For the one sided case, the independence of the $Z_k^j Q$ is easily established and allows us to jump from matrix to scalar convergence. For the two sided case no such short cut seems to be available.

Because PA^NQ converges (to zero) exactly when each $\underline{p}_i^T A^N \underline{q}_j$ converges (to zero) we are in a position to apply the above for each row $\underline{p}_i^T$ and each column $\underline{q}_j$.

4. The general case. For convenience, let us start with the zero convergence case first.

Theorem 1. *The following are equivalent*:

(i) $PA^NQ \to 0$,

(ii) $\rho_{PQ}(A) < 1$,

(iii) *the roots of* $\phi_{PQ}(\lambda)$ *lie inside the unit circle*,

(iv) $|\lambda_k| \geq 1 \Rightarrow P(A^i Z_k^0)Q = 0$ *for all* $i = 0, 1, \ldots$,

(v) $P\left[\sum_{i=0}^{N-1} A^i\right] Q$ *converges to a limit* L, *in which case the limit equals*

$$L = P(I - A)^D Q. \tag{4.1}$$

Proof. The equivalence of (i), (ii) and (iii) follows directly from (3.16).

(iii)⇒(iv). If $|\lambda_k| \geq 1$, then λ_k is not an effective eigenvalue and thus, by Lemma 1, part (iv) follows.

(iv)⇒(iii). If (iv) holds, and recalling that $Z_k^0 = h_k(A)$, we see from (1.9a) that $|\lambda_k| \geq 1 \Rightarrow \phi | h_k(\lambda)$. But $(\lambda - \lambda_k) \nmid h_k(\lambda)$. This means that all roots of ϕ must be inside the unit circle.

(v)⇒(i). Since PA^NQ is the Nth term in a convergent series, it must tend to zero.

(iv)⇒(v). The easiest way to do this is to make use of the following identity from [**1**]:

Lemma 3. *For any matrix A with $\lambda_1 = 1$ and index m_1, we have*

$$\sum_{i=0}^{N-1} A^i = (I - A)^D (I - A^N) + \sum_{j=0}^{m_1 - 1} \binom{N}{j+1} Z_1^j, \tag{4.2}$$

where $(\cdot)^D$ is the Drazin inverse of $(\cdot)$.

It follows at once that

$$P\left[\sum_{i=0}^{N-1} A^i\right] Q = P(I - A)^D (I - A^N) Q + \sum_{j=0}^{P_1 - 1} \binom{N}{j+1} P Z_1^j Q, \tag{4.3}$$

where $p_1 \geq 0$ is the effective P-Q index of $\lambda_1 = 1$. Hence, when $\lambda_1 = 1$ is not an effective eigenvalue, then the last summation in (4.3) vanishes. Now, by the equivalence of (i) and (iv), we also know that $PA^NQ \to 0$, and because $(I - A)^D$ is a polynomial in A, we may conclude that $P[\sum^{N-1} A^i]Q \to P(I - A)^D Q$.

It should be remarked here that, when $\phi_{PQ}(\lambda) = 1$, then $PA^NQ = 0$ for all $N \in \mathbf{N}$, in which case condition (iii) is vacuously true.

Next, let us return to the convergence of the Picard iteration (1.3).

Theorem 2. *The following are equivalent.*

(i) *The Picard iteration (1.3) converges to limit L'.*

(ii) *$P[\sum_{i=0}^{N-1} A^i]U$ converges.*

(iii) $PA^N U \to 0$.

(iv) $\rho_{PU}(A) < 1$.

(v) *The roots of* $\phi_{PU}(\lambda)$ *lie inside the unit circle.*

(vi) $|\lambda_k| \geq 1 \Rightarrow P(A^i Z_k^0)U = 0$ *for all* $i = 0, 1, \ldots,$ *in which case the limit equals*

$$L' = P(I - A)^D B + P Z_1^0 C. \tag{4.4}$$

Proof. The equivalence of (i) and (ii) was observed just after (1.4). The equivalence of (ii)–(vi) follows from Theorem 1 with Q replaced by U. To obtain the actual limit L', we apply (4.1) to U. This gives

$$Y_N = PX_N = P\left[\sum_{i=0}^{N-1} A^i\right]U + PC \longrightarrow P(I-A)^D[B-(I-A)C]+PC,$$

from which (4.4) follows. □

As a special case we set $B = 0$ and $U = (A - I)C$ in Theorem 2 and (1.3), which will finally give us the constrained power convergence.

Corollary 1. *The following are equivalent.*

(i) $PA^N C$ *converges to a limit* L''.

(ii) $P[\sum_{i=0}^{N-1} A^i(A - I)]C$ *converges.*

(iii) $PA^N(A - I)C \to 0$.

(iv) $\rho_{P(A-I)C}(A) < 1$.

(v) $(\lambda - \lambda_k)|\phi_{PC}(\lambda) \Rightarrow |\lambda| < 1$ *or* $\lambda_k = 1$ *with* $p_k = 1$.

(vi) $|\lambda_k| \geq 1, \Rightarrow P[A^i(A - I)Z_k^0]C = 0$ *for all* $i = 0, 1, \ldots,$ *in which case the limit equals*

$$L'' = P Z_1^0 C. \tag{4.5}$$

Proof. We only need to show (v). From (vi), the fact that $Z_k^0 = h_k(A)$ and (1.9a), we know that $|\lambda_k| \geq 1 \Rightarrow \phi_{PC}|(\lambda - 1)h_k(\lambda)$. But, since

$h_k(\lambda_k) \neq 0$, we see that none of the λ_k outside or on the unit circle, except 1, can be roots of ϕ_{PC}. For $\lambda_1 = 1$, we can have at most an effective index of one. Alternatively, we could use the convergence result of Section 3 directly.

We note that in (vi) we may separate out the eigenvalue 1. Indeed, if $|\lambda_k| \geq 1, \neq 1$, then it is not effective and $\phi | h_k(\lambda)$. In other words, $P[A^i Z_k^0]C = 0$ for all $i = 0, 1, \ldots$. For $\lambda_1 = 1$, the condition $P[A^i(A - I)Z_1^0]C = 0$ for all $i = 0, 1, \ldots$, does not seem to allow simplification.

Let us now use the above results to address the related problem, in which we are given P, A and B and wish to find out when there exists an initial matrix C for which the Picard iteration converges.

5. Existence of suitable initial conditions.

Suppose we are given P, A and B and wish to find out if we can and how we should pick our starting C, in order to ensure convergence. This we now address.

Theorem 3. *Suppose P, A and B are given. There exists an initial condition $X_0 = C$, for which the Picard iteration* (1.3) *converges, if and only if*

$$R[\vartheta B] \subseteq R[\vartheta(I - A)], \tag{5.1}$$

where $\vartheta = \begin{bmatrix} P \\ PA \\ \vdots \\ PA^{m-1} \end{bmatrix}$ *and m is the degree of the minimal polynomial of A.*

Proof. Suppose first that such a C exists. By Theorem 2, we know that the roots of the relative minimal polynomial $\phi_{PU}(\lambda)$ all lie inside the unit circle. This means that $(\lambda - 1)$ is coprime to $\phi_{PU}(\lambda)$. Hence, by Euclid's algorithm there exist $r(\lambda)$ and $s(\lambda)$ such that

$$\phi_{PU}(\lambda) \cdot r(\lambda) + (1 - \lambda)s(\lambda) = 1. \tag{5.2}$$

Consequently, $\lambda^i \phi_{PU}(\lambda) \cdot r(\lambda) + \lambda^i(1 - \lambda)s(\lambda) = \lambda^i$ for all $i = 0, 1, \ldots$. Replacing λ by A and pre-multiplying by P followed by post-multiplying by U, we arrive at

$$PA^i \phi_{PU}(A) \cdot r(A)U + PA^i(I - A)s(A)U = PA^iU, \tag{5.3}$$

in which the first term vanishes by the definition of $\phi_{PU}(\lambda)$.

Replacing U by $B-(I-A)C$, we then see that

$$PA^i(I-A)[s(A)U+C]=PA^iB. \tag{5.4}$$

Stacking these identities shows that $\vartheta(I-A)F=\vartheta B$, where $F=s(A)U+C$ is a fixed matrix, thus ensuring the necessity.

Conversely, if $R[\vartheta B]\subseteq R[\vartheta(I-A)]$, then there exists an F such that $\vartheta(I-A)F=\vartheta B$, which in turn implies that $PA^i(I-A)F=PA^iB$ and hence that $PA^i[B-(I-A)F]=0$. If we select $C=F$, then the Picard iteration now reduces to

$$P\left[\sum_{i=0}^{N-1}A^i\right][B-(I-A)F]=\sum_{i=0}^{N-1}PA^i[B-(I-A)F]=0,$$

which converges.

Having established existence, let us conclude by examining the set of all such initial conditions C for which the PI converges. In fact, we have

Proposition 1. *Given P, A and B. The set of all C such that the Picard iteration converges is given by*

$$\{C; C=F+D, \text{ where } \vartheta(I-A)F=\vartheta B, \text{ and } PA^ND \text{ converges}\}.$$

Proof. If $C=F+D$, with $\vartheta(I-A)F=\vartheta B$ and PA^ND convergent, then $PA^iU=PA^i[B-(I-A)(F+D)]=[PA^iB-PA^i(I-A)F]-PA^i(I-A)D$. Summing this yields

$$\sum_{i=0}^{N-1}PA^iU=-\sum_{i=0}^{N-1}PA^i(I-A)D=-P(I-A^N)D,$$

which converges. Conversely, let $C-F=D$. Then $\sum_{i=0}^{N-1}PA^i[B-(I-A)C]=\sum_{i=0}^{N-1}PA^i(A-I)(C-F)=\sum_{i=0}^{N-1}PA^i(A-I)D=P(A^N-I)D$ converges. Thus, PA^ND converges as desired. □

Remarks. (i) We have seen that $\rho_{PQ_1} < 1$ and $\rho_{PQ_2} < 1$ imply $\rho_{P(Q_1+Q_2)} < 1$. Without Theorem 1, this is not easily seen.

(ii) The matrix T of (3.12) is a triangular Toeplitz matrix $[a_{j-1}]$ with $a_i = \binom{N}{i}$. Its inverse T^{-1} is of the same type as $[b_{j-1}]$, where $b_i = (-1)^i\binom{N-1-i}{i}$. This may be seen by inverting $\rho(\lambda) = (1+\lambda)^N$.

(iii) We may use Corollary 1 to characterize the set $\tau = \{Q; PA^N \times Q \text{ converges}\}$. Indeed, this shows that

$$R(Q) \subseteq N[\vartheta(A-I)Z_k^0] \quad \text{for all } |\lambda_k| \geq 1.$$

When $P = I$ this reduces to $R(Q) \subseteq E_1 \oplus \sum_{|\lambda_i|<1}^{\oplus} W_i$, where E_1 is the span of all the eigenvectors corresponding to $\lambda_1 = 1$ and W_i is the generalized eigenspace corresponding to λ_i. However, it seems to be difficult to characterize the nullspace of $N[\vartheta(A-I)Z_k^0]$ in general.

(iv) We have seen that $\underline{x}^T A^N \underline{y} = 0$ for all N exactly when $\phi_{xy} = 1$. It would be of interest to find suitable range/rowspace conditions for this to hold.

We close with some open questions:

(i) Can we use the above splitting to find $\dim(W)$ or $\dim(K)$?

(ii) When exactly does the two sided case reduce to the one sided case? That is, when does PA^NQ convergent imply A^NQ or PA^N convergent?

(iii) How are the spaces W and Γ related?

REFERENCES

1. X. Chen and R.E. Hartwig, *The Picard iteration revisited*, submitted.

2. F.R. Gantmacher, *The theory of matrices*, Vol. 1, Chelsea, New York, 1960.

3. R.E. Hartwig, *Applications of the Wronskian and Gram matrices*, Linear Algebra Appl. **43** (1982), 229–241.

4. V. Mani and R.E. Hartwig, *Blackbox interpolation*, SIAM J. Matrix Anal. **16** (1995), 21–28.

Mathematics Department, North Carolina State University, Raleigh, NC 27695
E-mail address: hartwig@math.ncsu.edu

Department of Mathematics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia
E-mail address: peter.semrl@fmf.uni-lj.si