Practical Computation of Matrix Functions

Hans-J. Runckel
Abteilung Mathematik IV
Universität Ulm
7900 Ulm, Germany

and

Uwe Pittelkow
Abteilung Mathematik I
Universität Ulm
7900 Ulm, Germany

Submitted by Richard A. Brualdi

ABSTRACT

Several new representations for an analytic function $f(A)$ of a complex matrix A, and in particular for e^{At} and A^t, are derived, which also are numerically useful in that they avoid the computation of eigenvalues of A.

1. INTRODUCTION

The main purpose of this paper is to derive, in Sections 3 and 4, new representations for an analytic function $f(A)$ of a complex matrix A which are of particular importance for numerical computations (Theorems 2, 3, 6). Section 2 contains new and short proofs of known formulas for $f(A)$ and of related topics in order to give a unified, elementary, and self-contained approach to the subject (avoiding the Jordan canonical form). In Section 4 the new formulas are applied to the solution of linear systems of differential and difference equations and more general operator equations (Theorems 4, 5). In particular, new formulas [(30)–(34)] for e^{At} and A^t are obtained, which are theoretically interesting and numerically useful in that they do not require the computation of eigenvalues of A. The many publications on this subject show that there still is a demand for a theoretically simple and numerically useful method to compute $f(A)$.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/83/010161-18$3.00

2. DEFINITION OF MATRIX FUNCTIONS

Let A be a complex $m \times m$ matrix, whose different eigenvalues are $\lambda_1,\dots,\lambda_k$, and denote by I the $m \times m$ identity matrix. Assume that $c(A)=0$ for $c(\lambda):=\Pi_{\nu=1}^{k}(\lambda-\lambda_\nu)^{n_\nu}$ with $n_1+\cdots+n_k=n$. For instance, $c(\lambda)$ can be the characteristic or the minimal polynomial of A or the polynomial $[c(\lambda)/d(\lambda)]^s=\Pi_{\nu=1}^{k}(\lambda-\lambda_\nu)^s$, where c is the characteristic polynomial of A, d the g.c.d. of c and c', and $s\in\mathbb{N}$ sufficiently large. For $i=1,\dots,k$ we then put $q_i(\lambda):=c(\lambda)/(\lambda-\lambda_i)^{n_i}$, $q(\lambda):=\Sigma_{i=1}^{k}q_i(\lambda)$, $B_i:=q_i(A)$, $B:=q(A)$. All considerations below remain valid if A belongs to an arbitrary topological algebra over $\mathbb{C}$ with identity I such that $c(A)=0$.

LEMMA 1. *B is invertible.*

(See also [12, Lemma 2].)

Proof. $c(A)=0$ and $q(\lambda)-q(\lambda_i)=(\lambda-\lambda_i)r_i(\lambda)$ imply $\Pi_{i=1}^{k}[q(A)-q(\lambda_i)I]^{n_i}=0$, or $p(B)=0$, where $p(\lambda)=\Pi_{i=1}^{k}[\lambda-q(\lambda_i)]^{n_i}$. Since $q(\lambda_i)=q_i(\lambda_i)\neq 0$, $p(0)\neq 0$ follows, and hence $Br(B)=I$ holds with $r(\lambda)=[p(0)-p(\lambda)]/\lambda p(0)$. ■

LEMMA 2. *If $C_i:=B_iB^{-1}=B^{-1}B_i$, $i=1,\dots,k$, then $\Sigma_{i=1}^{k}C_i=I$, $C_iC_j=0$ for $i\neq j$ and $C_i^2=C_i$ for $i,j=1,\dots,k$. If, in addition, $\Sigma_{i=1}^{k}C_i'=I$ holds, where the C_i' are $m\times m$ matrices satisfying $(A-\lambda_iI)^{n_i}C_i'=0$, $i=1,\dots,k$, then $C_i'=C_i$ for $i=1,\dots,k$.*

Proof. Since for $i\neq j$, C_i contains the factor $(A-\lambda_jI)^{n_j}$, it follows that $C_iC_j=0$ and $C_iC_j'=0$. Next, Lemma 1 yields $\Sigma_{i=1}^{k}C_i=I$ and hence $C_i=\Sigma_{j=1}^{k}C_iC_j'=C_iC_i'=\Sigma_{j=1}^{k}C_jC_i'=C_i'$. In particular, $C_i^2=C_i$ follows. ■

As an application of Lemma 2 we obtain

THEOREM 1. *Let*

$$V_i=V_i(A):=\{x\in\mathbb{C}^m:(A-\lambda_iI)^{n_i}x=0\}.$$

Then $\mathbb{C}^m=V_1\oplus\cdots\oplus V_k$. Furthermore the columns of B_j (and C_j) span $V_j(A)$, and the rows of B_j (and C_j) span $V_j(A^T)$; hence rank $B_j=$ rank $C_j=\dim V_j$.

Proof. Let V_j' be the subspace of $\mathbb{C}^m$ spanned by the columns of B_j. Since $(A-\lambda_j I)^{n_j}B_j = c(A) = 0$ holds, $V_j' \subset V_j$ follows. Thus $B = \sum_{j=1}^k B_j$ and Lemma 1 imply $\mathbb{C}^m = V_1' + \cdots + V_k' = V_1 + \cdots + V_k$. Assume that $x_1 + \cdots + x_k = 0$, where $x_j \in V_j$. Since for $i \neq j$, B_i contains the factor $(A - \lambda_j I)^{n_j}$, we have $0 = B_j(x_1 + \cdots + x_k) = B_j x_j = (B_1 + \cdots + B_k)x_j = Bx_j$. Thus, $x_j = 0$, $j = 1,\ldots,k$, and $\mathbb{C}^m = V_1' \oplus \cdots \oplus V_k' = V_1 \oplus \cdots \oplus V_k$. This and $V_j' \subseteq V_j$ imply $V_j' = V_j$ and hence rank B_j = rank $C_j = \dim V_j$.

The rest follows from $(A^T - \lambda_j I)^{n_j} B_j^T = c(A^T) = 0$. ■

Lemma 2 and Theorem 1 imply

COROLLARY 1. *Though B_j and n_j depend on $c(\lambda)$, C_j and hence $V_j(A)$ and $V_j(A^T)$ are independent of the particular choice of $c(\lambda)$.*

(See also Corollary 3.)

Next, Lemma 2 yields for $1 \leqslant i, j \leqslant k$ and $r, s \geqslant 0$

$$C_{ir}C_{js} = \delta_{ij}C_{i,r+s}, \qquad \text{where} \quad C_{ir} = (A - \lambda_i I)^r C_i. \tag{1}$$

Now for $\nu \geqslant 0$ and $f(\lambda) := \lambda^\nu$, Lemma 2 implies

$$\begin{aligned} f(A) = A^\nu &= \sum_{i=1}^k A^\nu C_i \\ &= \sum_{i=1}^k (A - \lambda_i I + \lambda_i I)^\nu C_i = \sum_{i=1}^k \sum_{s=0}^\nu \binom{\nu}{s} (A - \lambda_i I)^s \lambda_i^{\nu - s} C_i \\ &= \sum_{i=1}^k \sum_{s=0}^\nu \frac{f^{(s)}(\lambda_i)(A - \lambda_i I)^s C_i}{s!} = \sum_{i=1}^k \sum_{s=0}^{n_i - 1} \frac{f^{(s)}(\lambda_i) C_{is}}{s!}, \end{aligned}$$

since $f^{(s)} = 0$ for $s > \nu$ and $C_{is} = (A - \lambda_i I)^s C_i = 0$ for $s \geqslant n_i$ because $c(A) = 0$. Hence this also holds for arbitrary complex polynomials $f(\lambda)$, and we obtain

LEMMA 3. *If for a sequence of complex polynomials $f_\nu(\lambda)$ and for a function $f(\lambda)$, $f^{(s)}(\lambda_i) = \lim_{\nu \to \infty} f_\nu^{(s)}(\lambda_i)$ exist for $0 \leqslant s < n_i$, $1 \leqslant i \leqslant k$, then $f(A) := \lim_{\nu \to \infty} f_\nu(A)$ exists and*

$$f(A) = \sum_{i=1}^k \sum_{s=0}^{n_i - 1} \frac{f^{(s)}(\lambda_i) C_{is}}{s!}. \tag{2}$$

This formula now will be used as the definition of $f(A)$ for every complex function $f(\lambda)$ for which the right side of (2) exists. See also [3], [5], [8], [26], [27], [31], and [10, pp. 104, 107].

One verifies immediately:

$$\text{if} \quad f(A), g(A) \text{ exist,} \quad \text{then}$$

$$(f+g)(A) = f(A) + g(A) \quad \text{and} \quad (fg)(A) = f(A)g(A) \tag{3}$$

(using Leibniz's product differentiation rule);

$$\text{if } f(A) \text{ exists} \quad \text{and} \quad \text{if all } f(\lambda_i) \neq 0, \quad \text{then}$$

$$[f(A)]^{-1} \text{ also exists} \quad \text{and} \tag{4}$$

$$[f(A)]^{-1} = \left(\frac{1}{f}\right)(A) = \sum_{i=1}^{k} \sum_{s=0}^{n_i-1} \left(\frac{1}{f(\lambda)s!}\right)^{(s)}\Bigg|_{\lambda=\lambda_i} C_{is}.$$

If some $C_i = 0$, then $(A - \lambda_i I)^{-1}$ would exist by (4) and (3). Hence all B_i and C_i are $\neq 0$. In particular, there exists r_i with

$$1 \leqslant r_i \leqslant n_i, \; C_{ir_i} = 0, \; C_{i, r_i - 1} \neq 0. \tag{5}$$

This yields the following converse of Lemma 3. If $\lim_{\nu \to \infty} f_\nu(A)$ exists, then $\lim_{\nu \to \infty} f_\nu^{(s)}(\lambda_i)$ exists for $0 \leqslant s < r_i$, $1 \leqslant i \leqslant k$ by multiplying $f_\nu(A)$ by C_{is}, $s = r_i - 1, \ldots, 0$ and using (1).

With $B = q(A)$, (3) and (4) yield $(1/q)(A) = B^{-1}$ and $(f/q)(A) = f(A)B^{-1}$. From this, (2), and $(1/q)^{(s)}(\lambda_i) = (1/q_i)^{(s)}(\lambda_i)$, $0 \leqslant s < n_i$, $1 \leqslant i \leqslant k$, we obtain the following well-known formula (see also [20], [27], and [10, p. 101]):

$$f(A) = \sum_{i=1}^{k} \sum_{s=0}^{n_i-1} \left(\frac{f(\lambda)}{q_i(\lambda)s!}\right)^{(s)}\Bigg|_{\lambda=\lambda_i} (A - \lambda_i I)^s B_i.$$

We observe that

$$p(z) := \sum_{i=1}^{k} \sum_{s=0}^{n_i-1} \left(\frac{f(\lambda)}{q_i(\lambda)s!}\right)^{(s)}\Bigg|_{\lambda=\lambda_i} (z - \lambda_i)^s \prod_{\nu=1, \nu \neq i}^{k} (z - \lambda_\nu)^{n_\nu} \tag{6}$$

is the Lagrange-Sylvester interpolation polynomial (with $p^{(s)}(\lambda_i)=f^{(s)}(\lambda_i)$, $0\leqslant s<n_i$, $1\leqslant i\leqslant k$) and we have $f(A)=p(A)$.

3. PRACTICAL COMPUTATION OF MATRIX FUNCTIONS

We now want to derive another representation for $p(z)$ in (6) which requires the following notation: Let

$$c(\lambda)=\sum_{r=0}^{n} c_{n-r}\lambda^r,\qquad h_0:=c_0=1,\quad h_r(\lambda)=\lambda h_{r-1}(\lambda)+c_r,\ r=1,\dots,n,$$

$$\text{so that}\qquad h_r(\lambda)=\sum_{s=0}^{r} c_{r-s}\lambda^s\ \ \text{and}\ \ h_n(\lambda)=c(\lambda). \tag{7}$$

Then

$$(\lambda-z)\sum_{r=0}^{n-1} h_{n-r-1}(\lambda)z^r=(\lambda-z)\sum_{r=0}^{n-1}\lambda^r h_{n-r-1}(z)=c(\lambda)-c(z) \tag{8}$$

can be verified immediately. Using (8) we define

$$H(\lambda,z):=\sum_{r=0}^{n-1} h_{n-r-1}(\lambda)z^r\qquad \text{with}\quad H(\lambda,z)=H(z,\lambda). \tag{9}$$

Observe that $H(\lambda,z)$ depends on the particular choice of $c(\lambda)$. If now $c(\lambda)$ is an *arbitrary* complex polynomial satisfying $c(A)=0$, then substituting A for z in (8) and (9) yields

$$(\lambda I-A)H(\lambda,A)=(\lambda I-A)H(A,\lambda)=c(\lambda)I. \tag{10}$$

REMARK 1. If $c(\lambda)=c_A(\lambda):=|\lambda I-A|$, then $c_A(A)=0$ and (10) imply

$$\operatorname{adj}(\lambda I-A)=\sum_{r=0}^{m-1} h_{m-r-1}(\lambda)A^r=\sum_{r=0}^{m-1}\lambda^r h_{m-r-1}(A).$$

REMARK 2. For $c_A(\lambda) \neq 0$, (10) yields

$$\sum_{r=0}^{m-1} \lambda^{m-r-1} \operatorname{trace} h_r(A) = c_A(\lambda) \operatorname{trace}(\lambda I - A)^{-1}. \tag{11}$$

Now $\operatorname{trace} A = -c_A^{(m-1)}(0)/(m-1)!$, and hence from $|tI - (\lambda I - A)^{-1}| = t^m c_A(\lambda - t^{-1})/c_A(\lambda)$ and

$$-\left(\frac{\partial}{\partial t}\right)^{m-1} \frac{t^m c_A(\lambda - t^{-1})}{(m-1)!}\Bigg|_{t=0} = c_A'(\lambda)$$

we obtain $\operatorname{trace}(\lambda I - A)^{-1} = c_A'(\lambda)/c_A(\lambda)$. Then (11) yields

$$\sum_{r=0}^{m-1} \lambda^{m-r-1} \operatorname{trace} h_r(A) = c_A'(\lambda) = \sum_{r=0}^{m-1} (m-r) c_r \lambda^{m-r-1},$$

and by comparing coefficients of λ^{m-r-1} we obtain $\operatorname{trace} h_r(A) = (m-r)c_r$. This and $\operatorname{trace} h_r(A) = \operatorname{trace}(Ah_{r-1}(A)) + mc_r$ from (7) yield the following well-known algorithm (see also [8]): If

$$|\lambda I - A| = \sum_{r=0}^{m} c_{m-r}\lambda^r, \qquad c_0 = 1,$$

then

$$c_r = -\operatorname{trace}\frac{Ah_{r-1}(A)}{r}, \qquad r = 1, \ldots, m.$$

Next, let γ be a suitable simple closed positively oriented curve enclosing $\lambda_1, \ldots, \lambda_k$, or the sum of k simple closed positively oriented sufficiently small circles around $\lambda_1, \ldots, \lambda_k$.

LEMMA 4. *Let $p(\lambda)$ be a complex polynomial of degree $< n$. Then with the notation* (7) *and* (9) *we have*

$$p(z) = \int_\gamma \frac{p(\lambda)H(\lambda, z)}{c(\lambda)2\pi i}\,d\lambda \tag{12}$$

Choosing in particular $p(z) = z^{\nu}$ *yields*

$$\int_{\gamma} \frac{\lambda^{\nu} h_{n-r-1}(\lambda)}{c(\lambda)2\pi i} d\lambda = \delta_{r\nu} \qquad \textit{for} \quad 0 \leqslant \nu, r < n. \tag{13}$$

Proof. (8) implies $1/(\lambda - z) = H(\lambda, z)/[c(\lambda) - c(z)]$. Then by Cauchy's integral formula

$$2\pi i p(z) = \int_{\gamma} \frac{p(\lambda)}{\lambda - z} d\lambda$$

$$= \int_{\gamma} \frac{p(\lambda)H(\lambda, z)}{c(\lambda) - c(z)} d\lambda = \int_{\gamma} \frac{p(\lambda)H(\lambda, z)}{c(\lambda)} d\lambda,$$

since

$$\int_{\gamma} p(\lambda)\{[c(\lambda) - c(z)]^{-1} - c(\lambda)^{-1}\} H(\lambda, z)\, d\lambda$$

$$= c(z) \int_{\gamma} \frac{p(\lambda)H(\lambda, z)}{c(\lambda)[c(\lambda) - c(z)]} d\lambda = 0,$$

because γ can be replaced by an arbitrarily large circle around 0. ∎

We now can prove (see also [2], [7]):

THEOREM 2. *Let* $p(\lambda)$ *be the interpolation polynomial* (6) *of* f. *Then with the notation* (9),

$$f(A) = p(A) = \int_{\gamma} \frac{p(\lambda)H(\lambda, A)}{c(\lambda)2\pi i} d\lambda. \tag{14}$$

If, in particular, f *is analytic at* $\lambda_1, \dots, \lambda_k$, *then*

$$f(A) = \int_{\gamma} \frac{f(\lambda)H(\lambda, A)}{c(\lambda)2\pi i} d\lambda. \tag{15}$$

Proof. (14) follows from $f(A) = p(A)$ in Section 2 and (12). (15) follows from (14), since $[f(\lambda) - p(\lambda)]/c(\lambda)$ has removable singularities at $\lambda_1, \dots, \lambda_k$, and hence $\int_{\gamma} H(\lambda, A)[f(\lambda) - p(\lambda)]/c(\lambda)\, d\lambda = 0$. ∎

Applying the residue theorem to (14) and using $p^{(s)}(\lambda_j) = f^{(s)}(\lambda_j)$, $0 \leqslant s < n_j$, $1 \leqslant j \leqslant k$, yields (se [18] for a special case):

COROLLARY 2. *If* $m_j \geqslant n_j$, $1 \leqslant j \leqslant k$, *then*

$$f(A) = \sum_{r=0}^{n-1} A^r \sum_{j=1}^{k} \left(\frac{f(\lambda) h_{n-r-1}(\lambda)(\lambda - \lambda_j)^{m_j}}{c(\lambda)(m_j - 1)!} \right)^{(m_j - 1)} \Bigg|_{\lambda = \lambda_j}$$

$$= \sum_{r=0}^{n-1} h_r(A) \sum_{j=1}^{k} \left(\frac{f(\lambda) \lambda^{n-r-1}(\lambda - \lambda_i)^{m_j}}{c(\lambda)(m_j - 1)!} \right)^{(m_j - 1)} \Bigg|_{\lambda = \lambda_j}$$

provided all required derivatives of f *exist. Since always* $n_j \leqslant n - k + 1$, *one can choose* $m_j = n - k + 1$ *or* $m_j = n$ *for* $1 \leqslant j \leqslant k$.

As a generalization of a result in [6] we then obtain

COROLLARY 3 [see Corollary 1 and (9)].

$$B_j = q_j(A) = \left(\frac{H(\lambda, A)}{(n_j - 1)!} \right)^{(n_j - 1)} \Bigg|_{\lambda = \lambda_j}; \tag{16}$$

if $m_j \geqslant n_j$, *then*

$$C_j = \left(\frac{H(\lambda, A)(\lambda - \lambda_j)^{m_j}}{c(\lambda)(m_j - 1)!} \right)^{(m_j - 1)} \Bigg|_{\lambda = \lambda_j} \quad \text{for } 1 \leqslant j \leqslant k. \tag{17}$$

Observe that C_j is independent of the particular choice of $c(\lambda)$ by Lemma 2.

Proof. (16) follows from Corollary 2 with $f(\lambda) = q_j(\lambda)$ and $m_j = n_j$. Next, $C_j = f_j(A)$, where $f_j(\lambda) = q_j(\lambda)/q(\lambda)$ by Lemma 2. Since $[f_j(\lambda) - 1]/c(\lambda)$ has a removable singularity at λ_j,

$$\left(\frac{(\lambda - \lambda_j)^{m_j} f_j(\lambda)}{c(\lambda)} \right)^{(\nu)} \Bigg|_{\lambda = \lambda_j} = \left(\frac{(\lambda - \lambda_j)^{m_j}}{c(\lambda)} \right)^{(\nu)} \Bigg|_{\lambda = \lambda_j}$$

holds for $0 \leqslant \nu < m_j$. Hence Corollary 2 yields (17). ∎

REMARK 3. If one defines $f(A)=\int_\gamma f(\lambda)(\lambda I-A)^{-1}/2\pi i\,d\lambda$ and $C_j=\int_{\gamma_j}(\lambda I-A)^{-1}/2\pi i\,d\lambda$, where γ_j is a small circle around λ_j, then (10) immediately yields (15) and (17).

REMARK 4. In order to apply Corollary 2, only the different eigenvalues $\lambda_1,\dots,\lambda_k$ of A have to be known, but not their exact multiplicities. Observe that $\lambda_1,\dots,\lambda_k$ are the simple zeros of the kth order polynomial $c^*(\lambda)=c(\lambda)/d(\lambda)$, where $d(\lambda)$ is the g.c.d. of $c(\lambda)$ and $c'(\lambda)$. $c(\lambda)$ may be replaced by $[c^*(\lambda)]^s=\Pi_{\nu=1}^k(\lambda-\lambda_\nu)^s$ for sufficiently large $s\in\mathbb{N}$.

REMARK 5. Assume that $\lambda_1,\dots,\lambda_k$ are known approximately. The advantage of the formulas in Corollary 2 over (2) is that only for the coefficients of $A^0,\dots,A^{n-1}$ or of $h_0(A),\dots,h_{n-1}(A)$ is new computation necessary whenever the accuracy of computation of $\lambda_1,\dots,\lambda_k$ is increased.

REMARK 6. For other definitions of $f(A)$ involving finite differences see [4, 11, 30].

We conclude this section by deriving from (15) new formulas for $f(A)$. (See also [2].)

THEOREM 3. *Let* $1/c(\lambda)=\sum_{\nu=n}^\infty d_\nu\lambda^{-\nu}$ *for* $|\lambda|>M:=\max_{1\leqslant j\leqslant k}\{|\lambda_j|\}$, *and let* $f(\lambda)$ *be analytic in* $|\lambda|<K$ *where* $K>M$. *Then*

$$f(A)=\sum_{\nu=n}^{\infty}d_\nu\left(\frac{f(\lambda)H(\lambda,A)}{(\nu-1)!}\right)^{(\nu-1)}\Bigg|_{\lambda=0}, \tag{18}$$

$$f(A)=\sum_{r=0}^{n-1}A^r\sum_{s=r+1}^{n}c_{n-s}\sum_{\nu=n+r-s}^{\infty}\frac{d_{\nu+s-r}f^{(\nu)}(0)}{\nu!}, \tag{19}$$

$$f(A)=\sum_{r=0}^{n-1}h_r(A)\sum_{\nu=r}^{\infty}\frac{d_{\nu+n-r}f^{(\nu)}(0)}{\nu!}, \tag{20}$$

$$f(A)=\sum_{r=0}^{n-1}A^r\left(\frac{f^{(r)}(0)}{r!}-\sum_{s=0}^{r}c_{n-s}\sum_{\nu=n+r-s}^{\infty}\frac{d_{\nu+s-r}f^{(\nu)}(0)}{\nu!}\right). \tag{21}$$

Proof. Let γ be a sufficiently large circle with center at 0. Then (15) implies $f(A)=\sum_{\nu=n}^\infty d_\nu\int_\gamma f(\lambda)H(\lambda,A)\lambda^{-\nu}/2\pi i\,d\lambda$, which is (18) by Cauchy's integral formula. Observing (7) and (9), we again obtain from (15)

$$f(A)=\sum_{r=0}^{n-1}A^r\sum_{s=0}^{n-r-1}c_{n-r-1-s}\sum_{\nu=n}^{\infty}d_\nu\int_\gamma\frac{f(\lambda)\lambda^{s-\nu}}{2\pi i}d\lambda,$$

which implies (19). (20) is obtained similarly from $f(A)=\sum_{r=0}^{n-1}h_r(A)\int_\gamma f(\lambda)\lambda^{n-r-1}/c(\lambda)2\pi i\,d\lambda$. Finally, using $g_r(\lambda):=\sum_{s=0}^{r}c_{n-s}\lambda^s=c(\lambda)-\lambda^{r+1}h_{n-r-1}(\lambda)$, we obtain

$$\int_\gamma\frac{f(\lambda)h_{n-r-1}(\lambda)}{c(\lambda)}d\lambda=\int_\gamma f(\lambda)\lambda^{-r-1}d\lambda-\int_\gamma\frac{f(\lambda)g_r(\lambda)\lambda^{-r-1}}{c(\lambda)}d\lambda.$$

Substituting this in (15) yields (21).

REMARK 7. The obvious advantage of the formulas (18)–(21) over the earlier ones is that the eigenvalues $\lambda_1,\ldots,\lambda_k$ of A need not be known. In $1/c(\lambda)=\sum_{\nu=n}^{\infty}d_\nu\lambda^{-\nu}$, we have $d_\nu=\int_\gamma\lambda^{\nu-1}/c(\lambda)2\pi i\,d\lambda$ and hence $d_n=1$ (since $c_0=1$), and for $\nu>n$,

$$\sum_{s=0}^{n}c_sd_{\nu-s}=\int_\gamma\frac{\sum_{s=0}^{n}c_s\lambda^{\nu-s-1}}{c(\lambda)2\pi i}d\lambda=\int_\gamma\frac{\lambda^{\nu-n-1}}{2\pi i}d\lambda=0.$$

Therefore d_ν can be computed recursively by

$$d_\nu=-\sum_{s=1}^{n}c_sd_{\nu-s}\qquad\text{for}\quad\nu>n\tag{22}$$

using $d_\nu=0$ for $\nu<n$ and $d_n=1$. From this the following upper bounds for d_ν are obtained by induction on ν. Put $a:=\sum_{s=1}^{n}|c_s|$. If $a\geqslant1$, then $|d_\nu|\leqslant a^{\nu-n}$, $\nu\geqslant n$. If $a<1$, then $|d_{n\nu+\rho}|\leqslant a^\nu<1$ for $1\leqslant\rho\leqslant n$, $\nu\geqslant0$. For a connection of the d_ν with the Lucas polynomials see [2]. By induction on ν one can easily prove that

$$d_\nu=\sum_{\substack{\nu_1+2\nu_2+\cdots+n\nu_n=\nu-n,\\ \nu_i\geqslant0}}\frac{(-1)^{\nu_1+\cdots+\nu_n}c_1^{\nu_1}\cdots c_n^{\nu_n}(\nu_1+\cdots+\nu_n)!}{\nu_1!\cdots\nu_n!}$$

holds for $\nu\geqslant n$.

4. APPLICATIONS

Let $T\colon F \to F$ be a linear operator, where F is a vector space of complex functions $y(t)$, $t \in R$ (a region in $\mathbb{R}$ or $\mathbb{C}$). We define

$$T\left(\sum_{r=0}^{N} y_r(t)A_r\right) := \sum_{r=0}^{N} (Ty_r(t))A_r$$

for $N \in \mathbb{N}$, $y_r(t) \in F$, and arbitrary complex $m \times m$ matrices A_r, $0 \leqslant r \leqslant N$. T may be D or E, where $Dy(t) = y'(t)$, $Ey(t) = y(t+1)$. More generally, T may be a polynomial in D and E.

Next, let L be an open set in $\mathbb{R}$ or $\mathbb{C}$ containing $\lambda_1,\dots,\lambda_k$, and let $y(\lambda,t)$, $f_{i_0\cdots i_u}(\lambda,t)$, $0 \leqslant i_0,\dots,i_u \leqslant N$ be complex functions of $\lambda \in L$, $t \in R$ such that $y(A,t)$ and $f_{i_0\cdots i_u}(A,t)$ exist according to (2) or (14). We assume that $(\partial/\partial\lambda)^s(T^r y(\lambda,t))$ and $(\partial/\partial\lambda)^s f_{i_0\cdots i_u}(\lambda,t)$ exist and that $(\partial/\partial\lambda)^s y(\lambda,t) \in F$ for $0 \leqslant r \leqslant u$, $0 \leqslant s < \mu := \max\{n_1,\dots,n_k\}$, $\lambda \in L$, $t \in R$.

As an illustration how to apply matrix functions we prove (see also [29])

Theorem 4. *Assume that $y(\lambda,t)$ satisfies*

$$\sum_{i_0,\dots,i_u=0}^{N} f_{i_0\cdots i_u}(\lambda,t)[y(\lambda,t)]^{i_0}[Ty(\lambda,t)]^{i_1}\cdots[T^u y(\lambda,t)]^{i_u} = 0 \tag{23}$$

and

$$T^r\left(\left(\frac{\partial}{\partial\lambda}\right)^s y(\lambda,t)\right) = \left(\frac{\partial}{\partial\lambda}\right)^s (T^r y(\lambda,t)) \qquad \text{for} \quad \lambda \in L, \quad t \in R, \tag{24}$$

$1 \leqslant r \leqslant u$ and $0 \leqslant s < \mu$. Then $y(A,t)$ satisfies

$$T^r(y(A,t)) = (T^r y(\lambda,t))|_{\lambda=A} \tag{25}$$

and

$$\sum_{i_0,\dots,i_u=0}^{N} f_{i_0,\dots,i_u}(A,t)[y(A,t)]^{i_0}\cdots[T^u y(A,t)]^{i_u} = 0 \qquad \text{for} \quad t \in R. \tag{26}$$

If, in addition, $T^r y(\lambda,t_0) = \alpha_r$ for $0 \leqslant r \leqslant v$, $\lambda \in L$, and $t_0 \in R$, then $T^r y(A,t_0) = \alpha_r I$, $0 \leqslant r \leqslant v$.

Proof. (24) and (2) imply (25). (25) and (3) yield

$$\sum f_{i_0 \cdots i_u}(A,t)[y(A,t)]^{i_0} \cdots [T^u y(A,t)]^{i_u}$$

$$= \sum f_{i_0 \cdots i_u}(A,t)\left[\, y(\lambda,t)|_{\lambda=A}\right]^{i_0} \cdots \left[\,(T^u y(\lambda,t))|_{\lambda=A}\right]^{i_u}$$

$$= \left(\sum f_{i_0 \cdots i_u}(\lambda,t) y(\lambda,t)^{i_0} \cdots [T^u y(\lambda,t)]^{i_u}\right)\Bigg|_{\lambda=A}$$

which is $=0$ by (23). This proves (26), and $T^r y(A,t_0)=\alpha_r I$ follows from (25), (2) and Lemma 2. ∎

THEOREM 5. *Assume that* (24) *and hence* (25) *hold. Assume furthermore that* $y(\lambda,t)$ *satisfies*

$$Ty(\lambda,t)=\lambda y(\lambda,t),\ y(\lambda,t_0)=1 \qquad \text{for} \quad \lambda \in L, \quad t,t_0 \in R. \tag{27}$$

Then

$$Ty(A,t)=Ay(A,t) \ \text{ and } \ y(A,t_0)=I \qquad \text{for} \quad t \in R.$$

In addition

$$y(A,t)=\sum_{r=0}^{n-1} \varphi_r(t) A^r,$$

where for $0 \leqslant r < n$

$$\varphi_r(t)=\sum_{j=1}^{k} \left(\frac{\partial}{\partial\lambda}\right)^{n_j-1} \left(\frac{y(\lambda,t) h_{n-r-1}(\lambda)(\lambda-\lambda_j)^{n_j}}{c(\lambda)(n_j-1)!}\right)\Bigg|_{\lambda=\lambda_j} \tag{28}$$

and

$$c(T)\varphi_r(t)=0, \quad T^\nu \varphi_r(t_0)=\delta_{r\nu} \qquad \text{for} \quad t \in R, \quad 0 \leqslant r,\nu < n. \tag{29}$$

Observe that also (14), (15), Cor. 2, (19) and (21) can be used to evaluate $\varphi_r(t)$.

Proof. The first part of Theorem 5 is a special case of Theorem 4. (28) follows from Corollary 2. In order to prove (29) we observe that (24) and (27) imply

$$T\left(\left(\frac{\partial}{\partial\lambda}\right)^s y(\lambda,t)\right)=\left(\frac{\partial}{\partial\lambda}\right)^s (Ty(\lambda,t))=\left(\frac{\partial}{\partial\lambda}\right)^s(\lambda y(\lambda,t))$$

$$=\lambda\left(\frac{\partial}{\partial\lambda}\right)^s y(\lambda,t)+s\left(\frac{\partial}{\partial\lambda}\right)^{s-1} y(\lambda,t)$$

or

$$(T-\lambda)\left(\frac{\partial}{\partial\lambda}\right)^s y(\lambda,t)=s\left(\frac{\partial}{\partial\lambda}\right)^{s-1} y(\lambda,t)$$

or

$$(T-\lambda)^{s+1}\left(\frac{\partial}{\partial\lambda}\right)^s y(\lambda,t)=s(T-\lambda)^s\left(\frac{\partial}{\partial\lambda}\right)^{s-1} y(\lambda,t)$$

$$=s!(T-\lambda)y(\lambda,t)=0 \qquad \text{for} \quad \lambda\in L, \quad t\in R.$$

Hence

$$(T-\lambda_j)^{n_j}\left[\left(\frac{\partial}{\partial\lambda}\right)^s y(\lambda,t)\bigg|_{\lambda=\lambda_j}\right]=0, \qquad 0\leqslant s<n_j, \quad 1\leqslant j\leqslant k, \quad t\in R,$$

and therefore $c(T)\varphi_r(t)=0$, $t\in R$. Finally, (24), (27) and (13) imply

$$T^\nu\varphi_r(t_0)=\sum_{j=1}^{k}\left(\frac{\lambda^\nu h_{n-r-1}(\lambda)}{c(\lambda)(n_j-1)!}\right)^{(n_j-1)}\Bigg|_{\lambda=\lambda_j}$$

$$=\int_\gamma\frac{\lambda^\nu h_{n-r-1}(\lambda)}{c(\lambda)2\pi i}d\lambda=\delta_{r\nu} \qquad \text{for} \quad 0\leqslant r,\nu<n. \qquad \blacksquare$$

REMARK 8. In order to find n solutions $y=\varphi_r(t)$, $0\leqslant r<n$, of $c(T)y=0$ with $T^\nu\varphi_r(t_0)=\delta_{r\nu}$, $0\leqslant r,\nu<n$, one only has to determine a solution $y=y(\lambda,t)$ of $Ty=\lambda y$ with $y(\lambda,t_0)=1$ for $\lambda\in L$ and evaluate $\varphi_r(t)$ according to (28), provided (24) holds.

REMARK 9. The equation $(T^p + A_1T^{p-1} + \cdots + A_pT^0)X(t) = 0$ with $\nu \times \nu$ matrices $A_1, \ldots, A_p$ can be transformed into $TY = AY$ with

$$A = \begin{pmatrix} 0 & I & \cdots & 0 \\ & \ddots & \ddots & \\ 0 & \cdots & 0 & I \\ -A_p & \cdots & -A_2 & -A_1 \end{pmatrix}.$$

The eigenvalues of A are the roots of $|\lambda^p I + \lambda^{p-1}A_1 + \cdots + A_p| = 0$.

EXAMPLE 1. For $T = D$ (27) has the solution $y(\lambda, t) = e^{\lambda(t-t_0)}$ for $\lambda, t, t_0 \in \mathbb{C}$. Hence by Theorem 5, $Y' = AY$ has the solution $Y(t) = y(A, t) = e^{A(t-t_0)}$, which can be computed for $t, t_0 \in \mathbb{C}$ according to (2), (15), (20):

$$\begin{aligned} e^{A(t-t_0)} &= \sum_{i=1}^{k} \sum_{s=0}^{n_i-1} \frac{(t-t_0)^s e^{\lambda_i(t-t_0)} C_{is}}{s!} \\ &= \int_\gamma \frac{e^{\lambda(t-t_0)} H(\lambda, A)}{c(\lambda) 2\pi i} d\lambda \\ &= \sum_{r=0}^{n-1} h_r(A) \sum_{\nu=r}^{\infty} \frac{d_{\nu+n+r}(t-t_0)^\nu}{\nu!}, \end{aligned}$$

and according to (15), (19), (21), (28) we obtain

$$e^{A(t-t_0)} = \sum_{r=0}^{n-1} \varphi_r(t) A^r$$

with

$$\begin{aligned} \varphi_r(t) &= \int_\gamma \frac{e^{\lambda(t-t_0)} h_{n-r-1}(\lambda)}{c(\lambda) 2\pi i} d\lambda \\ &= \sum_{s=r+1}^{n} c_{n-s} \sum_{\nu=n+r-s}^{\infty} \frac{d_{\nu+s-r}(t-t_0)^\nu}{\nu!} \\ &= \frac{(t-t_0)^r}{r!} - \sum_{s=0}^{r} c_{n-s} \sum_{\nu=n+r-s}^{\infty} \frac{d_{\nu+s-r}(t-t_0)^\nu}{\nu!}, \qquad 0 \leqslant r < n, \end{aligned} \tag{30}$$

Observe that the d_ν are given by (22) and that $c(D)\varphi_r(t)=0$, $D^\nu\varphi_r(t_0)=\delta_{r\nu}$, $0\leqslant r,\nu<n$. Since $e^{\lambda_j(t-t_0)}\neq 0$, $(e^{A(t-t_0)})^{-1}$ exists by (4) for all $t,t_0\in\mathbb{C}$. The first representation of $e^{A(t-t_0)}$ shows that $e^{A(t-t_0)}$ tends to zero (remains bounded) as $t\to+\infty$ if all $\operatorname{Re}\lambda_j<0$ (if, for each j, $\operatorname{Re}\lambda_j<0$ or $\operatorname{Re}\lambda_j=0$ and $C_{js}=0$ for $s>0$). The converse follows by multiplying $e^{A(t-t_0)}$ by C_{js}, $s=r_j-1,\dots,0$, $1\leqslant j\leqslant k$, and using (5), (1).

See also [1], [5], [9], [12]–[16], [18], [19], [21]–[23], [31], [10, pp. 116–129], and [15, pp. 36–38].

EXAMPLE 2. For $T=E$, (27) has the solution $y(\lambda,t)=\lambda^{t-t_0}$ for $\lambda,t,t_0\in\mathbb{C}$, $\lambda\neq 0$. Hence by Theorem 5 the system of difference equations $EY=AY$ has the solution $Y(t)=y(A,t)=A^{t-t_0}$, which can be computed for $t,t_0\in\mathbb{C}$ according to (2), (15) provided all $\lambda_i\neq 0$:

$$A^{t-t_0}=\sum_{i=1}^{k}\sum_{s=0}^{n_i-1}\binom{t-t_0}{s}\lambda_i^{t-t_0-s}C_{is}=\int_\gamma\frac{\lambda^{t-t_0}H(\lambda,A)}{c(\lambda)2\pi i}\,d\lambda,$$

where γ does not wind around 0.

If, in particular, $k=t-t_0$ and $k+1\in\mathbb{N}$, then also $\lambda_j=0$ is allowed and (20) yields

$$A^k=\sum_{r=0}^{n-1}h_{n-r-1}(A)d_{k+1+r}\qquad\text{with } d_\nu \text{ from (22).}\tag{31}$$

Next, (28), (19), and (21) yield for $k+1\in\mathbb{N}$

$$A^k=\sum_{r=0}^{n-1}\psi_r(k)A^r,\tag{32}$$

where

$$\psi_r(k)=\sum_{s=r+1}^{n}c_{n-s}d_{k+s-r}$$

$$=\delta_{r,k}-\sum_{s=0}^{r}c_{n-s}d_{k+s-r}$$

satisfies

$$c(E)\psi_r(k)=0, \quad E^\nu\psi_r(0)=\delta_{r\nu}, \qquad 0\leqslant r,\nu<n.$$

As in Example 1, one can easily see that A^k tends to zero (remains bounded) as $k\to+\infty$ iff all $|\lambda_j|<1$ (iff for each j $|\lambda_j|<1$ or $|\lambda_j|=1$ and $C_{js}=0$ for $s>0$).

REMARK 10. In order to determine for $k+1\in\mathbb{N}$ the fundamental system $\psi_r(k)$, $0\leqslant r<n$, of $c(E)y=0$ with $E^\nu\psi_r(0)=\delta_{r\nu}$, $0\leqslant r,\nu<n$, one has to compute the particular solution d_ν (22) of the inhomogeneous equation $\sum_{s=0}^n c_{n-s}d_{\nu-n+s}=\delta_{n\nu}$ with $d_\nu=0$ for $\nu<n$ and $d_n=1$. Then $\psi_r(k)$ is given by (32).

Using (32) we obtain

THEOREM 6. *$f(A)=\sum_{\nu=0}^\infty A^\nu f^{(\nu)}(0)/\nu!$ exists iff*

$$f(A)=\sum_{r=0}^{n-1}A^r\left(\frac{f^{(r)}(0)}{r!}+\sum_{\nu=n}^{\infty}\frac{\psi_r(\nu)f^{(\nu)}(0)}{\nu!}\right)$$

exists and (see (28)) for $0\leqslant r<n$

$$\varphi_r(t)=\left(\frac{\partial}{\partial\lambda}\right)^r y(\lambda,t)\bigg|_{\lambda=0}+\sum_{\nu=n}^{\infty}\psi_r(\nu)\left(\frac{\partial}{\partial\lambda}\right)^\nu\left(\frac{y(\lambda,t)}{\nu!}\right)\bigg|_{\lambda=0},$$

provided the right side exists.

In particular, always (see (30)))

$$e^{A(t-t_0)}=\sum_{r=0}^{n-1}A^r\left(\frac{(t-t_0)^r}{r!}+\sum_{\nu=n}^{\infty}\frac{\psi_r(\nu)(t-t_0)^\nu}{\nu!}\right).$$

Observe that Remark 7 yields $|\psi_r(\nu)|\leqslant a^{\nu+1-n}$ for $\nu\geqslant n$ if $a\geqslant 1$ and $|\psi_r(n\nu+\rho)|\leqslant a^\nu<1$ for $1\leqslant\rho\leqslant n$, $\nu\geqslant 1$ if $a<1$, $0\leqslant r<n$.

See also [2, 17, 19, 24, 25, 28].

REFERENCES

1 O. Borůvka, Remark on the use of Weyr's theory of matrices for the integration of linear differential equations with constant coefficients, *Časopis Pěst. Mat.* 79:151–155 (1954).

2 M. Bruschi and P. E. Ricci, An explicit formula for $f(A)$ and the generating functions of the generalized Lucas polynomials, *SIAM J. Math. Anal.* 13:162–165 (1982).
3 H. T. Chieh, Evaluations of matrix functions by real similarity transformation, *J. Franklin Inst.* 295:69–79 (1973).
4 C. Davis, Explicit functional calculus, *Linear Algebra Appl.* 6:193–199 (1973).
5 S. Deards, On the evaluation of e^{At}, *Matrix Tensor Quart.* 23:141–142 (1973).
6 D. Ž. Djoković, Eigenvectors obtained from the adjoint matrix, *Aequationes Math.* 2:94–97 (1968).
7 L. Fantappiè, Sulle funzioni di una matrice, *An. Acad. Brasil. Ciênc.* 26:25–33 (1954).
8 J. S. Frame, Matrix functions and applications, II, IV, *IEEE Spectrum* 1(4):102–108, (6):123–131 (1964).
9 E. P. Fulmer, Computation of the matrix exponential, *Amer. Math. Monthly* 82:156–159 (1975).
10 F. R. Gantmacher, *The Theory of Matrices I*, Chelsea, New York, 1977.
11 W. Jurkat and A. Peyerimhoff, Über Äquivalenzprobleme und andere limitierungs-theoretische Fragen bei Halbgruppen positiver Matrizen, *Math. Ann.* 159:234–251 (1965).
12 R. B. Kirchner, An explicit formula for e^{At}, *Amer. Math. Monthly* 74:1200–1204 (1967).
13 C. Kluczny, A certain form of the matrix e^{At}, *Zeszyty Nauk. Politech. Śląsk. Mat.-Fiz.* 25:3–10 (1974).
14 I. I. Kolodner, On exp(tA) with A satisfying a polynomial, *J. Math. Anal. Appl.* 52(3):514–524 (1975).
15 G. Kowalewski, *Einführung in die Theorie der kontinuierlichen Gruppen*, Akademische Verlagsgesellschaft, Leipzig, 1931.
16 M. Kumorovitz, Une solution du systéme linéare homogène d'équations différentielles du premier ordre à coefficients constants, *Ann. Soc. Polon. Math.* 23:190–200 (1950).
17 J. L. Lavoie, The m-th power of an $n \times n$ matrix and the Bell polynomials, *SIAM J. Appl. Math.* 29(3):511–514 (1975).
18 Y. Lehrer, On functions of matrices, *Rend. Circ. Mat. Palermo* (2) 6:103–108 (1957).
19 B. Z. Linfield, On the explicit solution of simultaneous linear difference equations with constant coefficients, *Amer. Math. Monthly* 47:552–554 (1940).
20 B. Lis, Particular spectral theory in finite-dimensional spaces, *Comment. Math. Prace Mat.* 18:51–61 (1974).
21 C. B. Mohler, Difficulties on computing the exponential of a matrix, in *2nd USA-Japan Computer Conference Proceedings* (Tokyo, 1975), AFIPS Press, Montvale, N.J., 1975, pp. 79–82.
22 J. Parizet, Détermination de l'exponentielle et recherche du "logarithme" d'une elément d'une algèbre de Banach unitaire engendrant une sous-algébre de dimension finie, *C.R. Acad. Sci. Paris Sér. A–B* 273:A971–A974 (1971).
23 E. J. Putzer, Avoiding the Jordan canonical form in the discussion of linear systems with constant coefficients, *Amer. Math. Monthly* 73:2–7 (1966).

24 M. A. Rashid, Powers of a matrix, Z. *Angew Math. Mech.* 55(5):271–272 (1975).
25 P. E. Ricci, Sulle potenze di una matrice, *Rend. Mat.* (6) 9(1):179–194 (1976).
26 H. Richter, Über Matrixfunktionen, *Math. Ann.* 122:16–34 (1950).
27 R. F. Rinehart, The equivalence of definitions of a matric function, *Amer. Math. Monthly* 62:395–414 (1955).
28 G. Roy, Puissances d'une matrice de polynôme minimal connu, *Rev. Roumaine Math. Pures Appl.* 20(10):1211–1213 (1975).
29 H.-J. Runckel and U. Pittelkow, Matrix functions and two-sided linear operator equations, *Arch. Math.*, to appear.
30 J. D. Stafney, Functions of a matrix and their norms, *Linear Algebra Appl.* 20(1):87–94 (1978).
31 M. N. S. Swamy, On a formula for evaluating e^{At} when the eigenvalues of A are not necessarily distinct, *Matrix Tensor Quart*, 23:67–72 (1972).

Received 3 August 1980; revised 28 June 1982