Mathematica
Balkanica

New Series Vol. 22, 2008, Fasc. 1-2

On Matrix Functions in Case of Confluent Characteristic Roots

Angel Popov

Presented by St. Nedev

The extension of a given scalar function $f(z)$ of a complex variable to a matrix function $f(A)$ on the set of $n \times n$ complex matrices is often represented as a linear combination $f(A) = \varphi_0 I + \varphi_1 A + \cdots + \varphi_{n-1} A^{n-1}$, where the coefficients $\varphi_k = \varphi_k(\lambda_1, \lambda_2, .., \lambda_n)$ depend on the characteristic roots of the matrix A, and clearly on the function itself. These coefficients can be evaluated by means of the spectral decomposition of $f(A)$, which in case of distinct roots is relatively easy, while it requires more complicated calculations in case of multiple roots. Here we show that the coefficients $\varphi_k(\lambda_1, \lambda_2, .., \lambda_n)$ converge when some subsets of the characteristic roots approach same values (the collection of multiple roots of a matrix B) and the linear combination of the powers of B with the limits obtained presents $f(B)$. Several examples at the end exercise this technique in case of matrices of small size.

Key Words: matrix function, exponential of a matrix, characteristic polynomial, Cayley-Hamilton theorem, spectral decomposition

The set of all $n \times n$ matrices with complex entries will be denoted by $\mathbb{C}_{n\times n}$, and $H_A(\lambda) = \det(\lambda I - A)$ will stay for the characteristic polynomial of the matrix A.

The study of diverse methods to extend a given scalar function to a matrix function, see [5], has been motivated mostly by the significance of the exponential e^A and one-parametric family e^{tA} in several branches of mathematics, see [4], 6]. One of the earliest and quite natural approach was based on the similarity of polynomials and power series. Thus, by means of the Taylor expansion $f(z) = \sum_{k\geq 0} a_k z^k$ of a given analytical function, one defines $f(A) = \sum_{k\geq 0} a_k A^k$ for $A \in \mathbb{C}_{n\times n}$. A translation (if more appropriate) allows to use the Taylor expansion $f(z) = \sum_{k\geq 0} b_k (z-c)^k$, centered at the point c, and then we take $f(A) = \sum_{k\geq 0} b_k (A - cI)^k$.

Convergence criteria, see Ferrar [1] and MacDuffee [2], for matrix power series (first one due to E. Weyr, 1887) exploit the spectrum of the matrix and the radius of convergence of the Taylor expansion. In particular, if the Taylor series has infinite radius of convergence, then $f(A) = \sum_{k\geq 0} a_k A^k$ converges for all matrices $A \in \mathbb{C}_{n\times n}$. We shall narrow our consideration to such kind of analytic functions, which will restrain some tiresome technical details.

A direct corollary from the Cayley-Hamilton theorem, that $H_A(A) = A^n + a_1 A^{n-1} + \cdots + a_n I = 0$, is the fact that all powers $A^k,\ k \geq 0$, can be presented as a linear combination of $I, A, \dots, A^{n-1}$ and the coefficients are polynomials of the characteristic roots $\lambda_1, \lambda_2, ..., \lambda_n$ of the matrix A. Now all powers A^k of $f(A)$ can be represented as linear combinations of $I, A, ..., A^{n-1}$, hence the matrix $f(A)$ will have the form

$$f(A) = \sum_{k=0}^{n-1} \varphi_k A^k,$$

where the coefficients $\varphi_k = \varphi_k(\lambda_1, \lambda_2, .., \lambda_n)$ depend on the characteristic roots of the matrix A, and clearly on the function $f(z)$.

Sylvester, in 1883, considered a direct extension of the Lagrange interpolation defining

$$f(A) = \sum_{j=1}^{n} f(\lambda_j) \prod_{1\leq i\neq j\leq n} \frac{1}{\lambda_j - \lambda_i}(A - \lambda_i I) \tag{1}$$

in case, when the matrix A has distinct characteristic roots $\lambda_1, \lambda_2, .., \lambda_n$.

In fact, if one denotes by $p_j(\lambda),\ j = 1, 2, ..., n$ the polynomials

$$p_j(\lambda) = \prod_{1\leq i\neq j\leq n} \frac{\lambda - \lambda_i}{\lambda_j - \lambda_i}, \tag{2}$$

we identify the spectral decomposition of the matrix A in terms of orthogonal idempotents $P_j = p_j(A),\ j = 1, 2, ..., n$, i.e.

$$A = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_n P_n,$$
$$P_i^2 = P_i, \quad \text{and} \quad P_i P_j = 0 \quad \text{for} \quad 1 \leq i \neq j \leq n.$$

Thus, for all $k \geq 0$, one has $A^k = \lambda_1^k P_1 + \lambda_2^k P_2 + \cdots + \lambda_n^k P_n$, and the substitution $A = \sum\limits_{j=1}^{n} \lambda_j P_j$ in the power series gives

$$f(A) = \sum_{j=1}^{n} f(\lambda_j) P_j,$$

as it was defined in (1).

Now, having $p_j(\lambda) = \sum\limits_{k=0}^{n-1} p_{jk}\lambda^k$, we can evaluate the coefficients φ_k explicitly:

$$\varphi_k(\lambda_1, \lambda_2, .., \lambda_n) = \sum_{j=1}^{n} f(\lambda_j) p_{jk}. \tag{3}$$

Note that the Lagrange interpolation $L(\lambda) = \sum_{j=1}^{n} f(\lambda_j)p_j(\lambda)$ provides the unique polynomial of degree $\leq n-1$ which assumes the set of prescribed values at a collection of (distinct) selected points, and the matrices $P_j(\lambda), j = 1, 2, ..., n$ do not depend on the function $f(z)$.

The case of multiple rots is more complicated. Buchheim, in 1886, generalized the Sylvester's approach exploiting the Hermit's (generalized Lagrange) interpolation that defines uniquely the polynomial of degree $\leq n-1$ which along with its derivatives assumes prescribed values at a collection of (distinct) points. A complete picture of the current frame for this approach can be found in 3], Chapter 5, where the spectral decomposition and related topics are discussed in full details.

Here we shall see how one can sidestep the technique of spectral decomposition in case of multiple roots by examining the coefficients $\varphi_k(\lambda_1, \lambda_2, ..., \lambda_n)$ in case of distinct roots (3), and then evaluating some limits to construct the function in case of confluent roots.

Our main result here is the following

Theorem 1. *Let $f(z)$ be an analytic function, and let $B \in \mathbb{C}_{n\times n}$ be a matrix with distinct characteristic roots $\mu_1, \mu_2, ..., \mu_r$ of multiplicities $m_1, m_2, ..., m_r$, respectively, $m_1 + \cdots + m_r = n$. Assume A is a matrix with distinct characteristic roots $\lambda_1, ..., \lambda_{m_1}$ (in a neighborhood of μ_1), $\lambda_{m_1+1}, ..., \lambda_{m_1+m_2}$ (in a neighborhood of μ_2), ..., and $\lambda_{n-m_r+1}, ..., \lambda_n$ (in a neighborhood of μ_r).*

Then the coefficients $\varphi_k(\lambda_1, \lambda_2, ..., \lambda_n)$ of $f(A)$ in (3) converge when $\lambda_1, ..., \lambda_{m_1}$ approach μ_1, $\lambda_{m_1+1}, ..., \lambda_{m_1+m_2}$ approach μ_2, ..., and $\lambda_{n-m_r+1}, ..., \lambda_n$ approach μ_r, and if $\lim \varphi_k = \psi_k(\mu_1, \mu_2, ..., \mu_r)$, then $f(B) = \sum_{k=0}^{n-1} \psi_k B^k$.

First, we shall see that $f(A)$ is continuous function of the matrix variable. Convergency in $\mathbb{C}_{n\times n}$ is a natural extension of the entry-wise convergency of the series of (i, j) entries for each pair $1 \leq i, j \leq n$. In terms of any matrix norm $\|A\|$ a matrix series $A_1, A_2, ... \to 0$ if and only if the numerical series $\|A_1\|, \|A_2\|, ... \to 0$. Here we shall use one easy and quite popular matrix norm defined by $\|A\| = \max_{1\leq i,j\leq n} |a_{ij}|$. It is easy to check, see [4, Chapter 4], that the norm introduced satisfies the following properties

$$\|A\| \geq 0, \text{ and } \|A\| = 0 \text{ if and only if } A = 0,$$

$$\|cA\| = |c|\|A\|, c \in \mathbb{C}, \text{ and } \|A + B\| \leq \|A\| + \|B\|,$$

$$\|AB\| \leq n\|A\|\|B\|, \text{ hence } \|A_1A_2...A_k\| \leq n^{k-1}\|A_1\|\|A_2\|...\|A_k\| \text{ for } k \geq 2.$$

Continuity of $f(A)$ follows directly from the following

Lemma 1. *Let $f(z)$ be analytic in the complex plane. Then for every matrix $A \in \mathbb{C}_{n\times n}$ there exists a positive constant $c = c(A)$, such that for every matrix Q with $\|Q\| = \varepsilon \leq \|A\|$ the inequality $\|f(A + Q) - f(A)\| \leq \varepsilon.c$ holds.*

P r o o f. The power series $f(z) = \sum_{k\geq 0} a_k z^k$ converges for all $z \in \mathbb{C}$ and we calculate $f(A+Q)-f(A) = \sum_{k\geq 1} a_k C_k$, where the matrix $C_k = (A+Q)^k - A^k$ is a sum of $2^k - 1$ terms of a kind

$$W_{p,q} = A^{p_1}Q^{q_1}A^{p_2}Q^{q_2} \; ...$$

with $p_1 + p_2 + \cdots = p \geq 0$, $q_1 + q_2 + \cdots = q \geq 1$, and $p + q = k$.

Now, for $a = n\|A\|$, we have

$$\|W_{p,q}\| \leq n^{k-1}\|A\|^p\|Q\|^q \leq n^{k-1}\|A\|^{k-1}\|Q\| \leq \varepsilon a^{k-1},$$

hence $\|B_k\| \leq 2\varepsilon(2a)^{k-1}$. Thus, $\|f(A + Q) - f(A)\| \leq 2\varepsilon \sum_{k\geq 1} a_k(2a)^{k-1}$, the last power series converges, and one can easily identify the constant c needed. ■

Next, we shall scrutinize the coefficients $\varphi_k(\lambda_1, \lambda_2, ..., \lambda_n)$ in (3). It is clear that $\varphi_k(\lambda)$ have the form $\varphi_k(\lambda) = \frac{1}{\Delta}\phi_k(\lambda)$, where $\Delta = \prod_{1\leq i\neq j\leq n} (\lambda_i - \lambda_j)$, and $\phi_k(\lambda) = \phi_k(\lambda_1, \lambda_2, ..., \lambda_n)$ is a polynomial of $\lambda_1, \lambda_2, ..., \lambda_n$ and $f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)$. Thus, when $\lambda_1, \lambda_2, ..., \lambda_n$ approach distinct values $\mu_1, \mu_2, ..., \mu_n$ each coefficient $\varphi_k(\lambda)$ converges to $\varphi_k(\mu)$.

In case of confluent roots it is more suitable to work with the Newton's interpolation

$$N(\lambda) = d_0 + d_1(\lambda-\lambda_1) + d_2(\lambda-\lambda_1)(\lambda-\lambda_2) + \cdots + d_{n-1}(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n),$$

where the coefficients (divided differences) d_j, $j = 0, 1, ..., n - 1$ are defined recursively by

$$d_0 = f_1 = f(\lambda_1), \quad d_1 = [f_1, f_2] = \frac{f(\lambda_1) - f(\lambda_2)}{\lambda_1 - \lambda_2}, \quad \text{and}$$

$$d_j = [f_1, f_2, \ldots, f_{j+1}] = \frac{[f_1, \ldots, f_j] - [f_2, \ldots, f_{j+1}]}{\lambda_1 - \lambda_{j+1}}, \text{ for } j \geq 2.$$

The polynomial $N(\lambda)$, of degree $\leq n-1$, takes values $f_1 = f(\lambda_1), \ldots, f_n = f(\lambda_n)$ at (distinct) points $\lambda_1, \ldots, \lambda_n$, hence it coincides with the Lagrange interpolation. Thus, the coefficients $\varphi_k(\lambda_1, \lambda_2, ..., \lambda_n)$ will be polynomials in $d_0, d_1, \ldots, d_{n-1}$ and it is enough to see that the divided differences $d_0, d_1, \ldots, d_{n-1}$ converge when $\lambda_1, \lambda_2, \ldots, \lambda_n$ approach some particular (confluent) values.

Let us assume that $\lambda_1, ..., \lambda_{m_1}$ approach some value μ_1, $\lambda_{m_1+1}, ..., \lambda_{m_1+m_2}$ approach another one μ_2, ..., etc. The convergency of each coefficient $d_j(\lambda)$ is easily granted by the specific structure of divided differences, e.g. when $\lambda_s, \lambda_{s+1}, ..., \lambda_t$ approach same value μ, there is no effect on each divided difference $[f_p, f_{p+1}, \ldots, f_q]$ with

$$\{s, s+1, \ldots, t\} \cap \{p, p+1, \ldots, q\} = \varnothing.$$

Otherwise we can present it as a linear combination of terms like

$$[f_{s-1}, f_s], \ [f_t, f_{t+1}], \ \text{and } [f_a, f_{a+1}, \ldots, f_b], \ \text{with } s \leq a < b \leq t.$$

Obviously, the first two have limits $[f(\lambda_{s-1}), f(\mu)]$, and $[f(\mu), f(\lambda_{t+1})]$, while the third kind converges to a constant multiple of the derivative $f^{b-a}(\mu)$.

Now, having in mind the considerations above, the proof of the theorem follows directly form the following

Lemma 2. *Let $B \in \mathbb{C}_{n\times n}$ be a given matrix. Then for every $\varepsilon > 0$ there exists a matrix Q with $\|Q\| < \varepsilon$ such that the matrix $A = B + Q$ has n distinct characteristic roots.*

P r o o f. Assume B has (not necessarily distinct) characteristic roots $\mu_1, \mu_2, ..., \mu_n$ recognizing their multiplicities. Let $J = T^{-1}BT$ be the Jordan canonical form with a transition matrix T, and let $c = \|T^{-1}\|.\|T\|$. It is clear that there exists a diagonal matrix $D = diag(\delta_1, \delta_2, \ldots, \delta_n)$ with $\|D\| < \varepsilon/cn^2$, such that the characteristic roots $\lambda_1 = \mu_1 + \delta_1, \lambda_2 = \mu_2 + \delta_2, \ldots, \lambda_n = \mu_n + \delta_n$ of the matrix $J + D$ are all distinct. Then the matrix $A = T^{-1}(J + D)T = B + Q$ clearly satisfies the requirements: it has the same (distinct) characteristic roots as the matrix $J + D$, and $\|Q\| = \|T^{-1}DT\| \leq n^2\|T^{-1}\|\|D\|\|T\| < \varepsilon$. ∎

Next, we want to show how the technique discussed works in case of matrices of small order.

The case of 2×2 matrices is trivial.

Assume A is a matrix with distinct characteristic roots λ_1, λ_2. The polynomials in (2) are $p_1(\lambda) = \frac{1}{\lambda_1 - \lambda_2}(\lambda - \lambda_2)$, and $p_2(\lambda) = \frac{1}{\lambda_2 - \lambda_1}(\lambda - \lambda_1)$, and we have

$$L(\lambda) = f(\lambda_1)p_1(\lambda) + f(\lambda_2)p_2(\lambda) = \frac{\lambda_1 f(\lambda_2) - \lambda_2 f(\lambda_1)}{\lambda_1 - \lambda_2} + \frac{f(\lambda_1) - f(\lambda_2)}{\lambda_1 - \lambda_2}\lambda.$$

So, $f(A) = \frac{1}{\lambda_1-\lambda_2}[\lambda_1 f(\lambda_2) - \lambda_2 f(\lambda_1)]I + \frac{1}{\lambda_1-\lambda_2}[f(\lambda_1) - f(\lambda_2)]A$, and after evaluation of the limits of the coefficients, when $\lambda_1, \lambda_2 \to \lambda_0$, one get

$$f(B) = [f(\lambda_0) - \lambda_0 f'(\lambda_0)]I + f'(\lambda_0)B \tag{4}$$

for any matrix B with double characteristic root λ_0.

For example, the matrix $B = \begin{pmatrix} 6 & 2 \\ -8 & -2 \end{pmatrix}$ has double root $\lambda_1 = \lambda_2 = 2$ and one calculates

$$e^{Bt} = (e^{2t} - 2te^{2t})I + te^{2t}B = e^{2t}\begin{pmatrix} 1+4t & 2t \\ -8t & 1-4t \end{pmatrix}.$$

Similarly, one calculates

$$\ln B = [\ln 2 - 1]I - \frac{1}{2}B = \begin{pmatrix} 2+\ln 2 & 1 \\ -4 & -2+\ln 2 \end{pmatrix},$$

and

$$\sqrt{B} = [\sqrt{2} - \frac{1}{\sqrt{2}}]I + \frac{1}{2\sqrt{2}}B = \frac{1}{\sqrt{2}}\begin{pmatrix} 4 & 1 \\ -4 & 0 \end{pmatrix}.$$

For matrices of larger size ($n = 3, 4$) one can derive formulas similar to (4), but here we prefer to display some particular examples.

The characteristic roots of $B = \begin{pmatrix} 1 & -3 & 4 \\ 4 & -7 & 8 \\ 6 & -7 & 7 \end{pmatrix}$ are $\lambda_1 = 3, \lambda_2 = \lambda_3 = -1$.

For $\varepsilon > 0$ we examine the Lagrange interpolation at $\lambda_1 = 3, \lambda_2 = -1 + \varepsilon, \lambda_3 = -1$, and construct the polynomials

$$p_1 = \frac{(\lambda+1-\varepsilon)(\lambda+1)}{4(4-\varepsilon)},\ p_2 = \frac{(\lambda-3)(\lambda+1)}{(-4+\varepsilon)\varepsilon},\ p_3 = \frac{(\lambda-3)(\lambda+1-\varepsilon)}{(-4)(-\varepsilon)},$$

Now, for e^{Bt} we shall examine

$$L(\lambda) = e^{3t}p_1 + e^{(-1+\varepsilon)t}p_2 + e^{-t}p_3$$

$$= e^{3t}p_1 - e^{-t}\frac{\lambda-3}{4(4-\varepsilon)}[(4\frac{e^{\varepsilon t}-1}{\varepsilon} + 1)(\lambda+1) + 4 - \varepsilon],$$

which converges (when $\varepsilon \to 0$) to

$$\frac{1}{16}e^{3t}(\lambda+1)^2 - \frac{1}{16}e^{-t}[(4t+1)(\lambda+1) + 4].$$

Hence,

$$e^{Bt} = \frac{1}{16}e^{3t}(B+I)^2 - \frac{1}{16}e^{-t}[(4t+1)(B+I)+4I]$$

$$= e^{3t}\begin{pmatrix}1 & -1 & 1\\ 2 & -2 & 2\\ 2 & -2 & 2\end{pmatrix} + e^{-t}\begin{pmatrix}-2t-1 & t+1 & -1\\ -4t-2 & 2t+2 & -2\\ -2t-2 & t+2 & -2\end{pmatrix}$$

For $\sqrt{B}$ we have to examine

$$L(\lambda) = \sqrt{3}\, p_1 + \sqrt{-1+\varepsilon}\, p_2 + \sqrt{-1}\, p_3$$

$$= \sqrt{3}p_1 - i\frac{\lambda-3}{4(4-\varepsilon)}[(4\frac{\sqrt{1-\varepsilon}-1}{\varepsilon}+1)(\lambda+1)+4-\varepsilon],$$

which converges (when $\varepsilon \to 0$) to

$$\frac{\sqrt{3}}{16}(\lambda+1)^2 + \frac{i}{16}(\lambda-3)^2,$$

and then

$$\sqrt{B} = \frac{\sqrt{3}}{16}(B+I)^2 + \frac{i}{16}(B-3I)^2$$

$$= \frac{\sqrt{3}}{2}\begin{pmatrix}2 & -1 & 2\\ 4 & -4 & 4\\ 4 & -4 & 4\end{pmatrix} + \frac{i}{2}\begin{pmatrix}2 & -1 & -4\\ 0 & 4 & -4\\ -2 & 3 & -4\end{pmatrix}.$$

The calculations for matrices of larger size and roots of higher multiplicity are more complicated, and may be executed with the help of L'Hospital rule. Sometimes, (when the degree of the minimum polynomial of the matrix is less than the degree of the characteristic polynomial), one can simplify a lot of calculations, based on some specific matrix relations.

The characteristic roots of the matrix $B = \begin{pmatrix}2 & 1 & 1 & 1\\ -1 & 0 & -1 & -1\\ 6 & 1 & -1 & 1\\ -6 & -1 & 4 & 0\end{pmatrix}$ are $\lambda_1 = -4$, and a triple root $\lambda_2 = \lambda_3 = \lambda_4 = 1$. We shall examine the Lagrange interpolation at $\lambda_1 = -4,\ \lambda_2 = 1-\varepsilon,\ \lambda_3 = 1,\ \lambda_4 = 1+\varepsilon$, and construct the polynomials

$$p_1(\lambda) = \frac{[(\lambda-1)^2-\varepsilon^2](\lambda-1)}{(-5+\varepsilon)(-5)(-5-\varepsilon)},\quad p_2(\lambda) = \frac{(\lambda+4)(\lambda-1)(\lambda-1-\varepsilon)}{(5-\varepsilon)(-\varepsilon)(-2\varepsilon)},$$

$$p_3(\lambda) = \frac{(\lambda+4)(\lambda-1+\varepsilon)(\lambda-1-\varepsilon)}{(5)(\varepsilon)(-\varepsilon)}, \quad p_4(\lambda) = \frac{(\lambda+4)(\lambda-1+\varepsilon)(\lambda-1)}{(5+\varepsilon)(2\varepsilon)(\varepsilon)}.$$

For e^{Bt} we have to examine

$$L(\lambda) = e^{-4t}p_1 + e^{(1-\varepsilon)t}p_2 + e^t p_3 + e^{(1+\varepsilon)t}p_4 = e^{-4t}p_1(\lambda) + e^t\frac{\lambda+4}{5(25-\varepsilon^2)}R,$$

where $R = R_1 + R_2$ with

$$R_1 = \frac{1}{2\varepsilon^2}\{2\varepsilon^2(25-\varepsilon^2) - 5\varepsilon^2(e^{t\varepsilon}+e^{-t\varepsilon})(\lambda+1) + 25\varepsilon(e^{t\varepsilon}-e^{-t\varepsilon})(\lambda+1)\},$$

$$R_2 = \frac{1}{2\varepsilon^2}\{5(5+\varepsilon)e^{-t\varepsilon} - 2(25-\varepsilon^2) + 5(5-\varepsilon)e^{t\varepsilon}\}(\lambda-1)^2.$$

One can evaluate $lim_{\varepsilon\to 0}R$ applying twice L'Hospital rule, but here we shall shorten that procedure because of some specific relations that our matrix B satisfies. Direct calculations show that

$$(B-I)^2 = 25P,\ (B-I)^3 = -125P,\ (B-I)(B+4I) = 5Q,\ (B-I)^2(B+4I) = 0,$$

where

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \end{pmatrix}, \quad Q = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{pmatrix}.$$

Now, all terms in R_2 have $(\lambda-1)^2$ and it will vanish after multiplying by $\lambda+4$ and replacing our matrix B. Thus, it is enough to evaluate $\lim_{\varepsilon\to 0} R_1$ which converges to $5[(5t-1)(\lambda-1)+5]$.

So, with the matrices P and Q introduced above we have

$$e^{Bt} = e^{-4t}P + \frac{1}{5}(5t-1)e^tQ + \frac{1}{5}e^t(B+4I) = e^{-4t}P + e^t(I-P+tQ).$$

At the end, we shall briefly discuss the range of reliability of this technique. As we see in the examples above, the validity of this approach is beyond the class of power series with infinite radius of convergence. A natural frame for a reasonable use of the result of Theorem 1 is typical for all the methods in this field, [5], namely the function $f(z)$ should be well defined on the spectrum of the matrix B, i.e. for every characteristic root λ_0 of multiplicity m, the function $f(z)$ and the derivatives $f'(z), f''(z), ..., f^{(m-1)}(z)$ should be defined at (a neighborhood) of λ_0.

References

[1] W.L. Ferrar, *Finite Matrices,* Clarendon Press, 1951.
[2] C.C. MacDuffee, *Theory of Matrices,* Chelsea, New York, 1956.
[3] P. Lankaster, *Theory of Matrices,* Academic Pres, New York-London, 1969, (Russian translation 1978).
[4] M. Artin, *Algebra,* Prentice Hall, New Jersey 1991.
[5] R.F. Rinehart, The equivalence of definitions of a matri function, *Amer. Math. Monthly*, **62**, 6, 1955, 395-414.
[6] C. Moler, C. Van Loan, Nineteen dubious was to compute the exponential of a matrix, twenty five yeatrs later, *SIAM Review,* **45**, 1, 2003, 3-49.

American University in Bulgaria,
2700 Blagoevgrad, BULGARIA
E-mail: apopov@aubg.bg

Received 12.07.2007