On power series of matrices

Kurt Hensel

Journal für die reine und angewandte Mathematik (1826), 155,
pp. 107-110

Among the applications of the system - or matrix - calculus which have been made in the last few decades, the theory of functions or the infinitesimal calculus of matrices, as founded by Volterra and Schlesinger and applied with such great success to linear differential equations, especially by the latter, offers a particularly great interest.

The basis for this investigation is the consideration of a function element or of a power series

$$\mathfrak{B}(T) = a_0 + a_1 T + a_2 T^2 + \cdots$$

whose argument is a system of the nth order

$$T = (t_{i,k})$$

while its coefficients a_i are given scalars.

The main question here is for which values of T $\mathfrak{B}(T)$ converges, that is, in what way the variability of $T = (t_{i,k})$ is constrained by the requirement of convergence. Now this question has found a very nice and simple, but not complete answer by a theorem of E. Weyr (Bulletin des sciences mathématiques sér. 2, vol. XI); but his proof is rather complicated and does not give, as it seems to me, the full insight into the nature of this simple question.

Therefore, I would like to give in these lines a new and complete answer to this question, which follows directly from the results of my treatise "On Fields of Matrices" (this Journal vol. 127, p. 116).

Call two systems T and $T' = P^{-1}TP$ equivalent if each is transformed into the other by a contragradient transformation with an arbitrary multiplier P, then it follows from the well-known and also easily directly provable relation:

$$\mathfrak{B}(T') = \mathfrak{B}(P^{-1}TP) = P^{-1}\mathfrak{B}(T)P,$$

that $\mathfrak{B}(T)$ converges if and only if the same power series is convergent for any system T' equivalent to T.

One can now replace T from the outset by such a simple equivalent T' that for this the question posed here can be found quite trivially.

As is well known, every system of nth order T satisfies a uniquely determined equation of lowest degree, the so-called main equation with scalar coefficients:

$$\psi(t) = t^\nu + b_1 t^{\nu-1} + \cdots + b_\nu = (t-\tau_1)^{\nu_1} \cdots (t-\tau_r)^{\nu_r}, \qquad (1)$$

whose degree $\nu \leq n$ and whose left-hand side, which will be mentioned here only incidentally, is a divisor of the determinant $|t - T|$ belonging to T . Equivalent systems apparently satisfy the same main equation.

The first theorem, which I proved in my work with the simplest means, is now this:

Each system [matrix] T is equivalent to a system T' of the form:

$$T' = \begin{pmatrix} T'_{1,1} & 0 & \cdots & 0 \\ 0 & T'_{2,2} & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & T'_{r,r} \end{pmatrix} \qquad (2)$$

in which each of the diagonal $T'_{i,i}$ is a square system of order n_i, whose main equation:

$$\psi_i(t) = (t-\tau_u)^{\nu_u}$$

consists of only one of the r linear factor powers of $\psi(t)$.

Since now for this equivalent system $T' = (T'_{i,i})$ obviously:

$$\mathfrak{B}(T') = \begin{pmatrix} \mathfrak{B}(T'_{1,1}) & 0 & \cdots & 0 \\ 0 & \mathfrak{B}(T'_{2,2}) & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \mathfrak{B}(T'_{r,r}) \end{pmatrix} \qquad (2a)$$

therefore $\mathfrak{B}(T')$ and thus also $\mathfrak{B}(T)$ converges if and only if it is the same for each of the r partial systems $\mathfrak{B}(T'_{i,i})$, then the question of convergence needs to be examined only for any one of them.

So we can and want to assume from the beginning that already for the original nth order system T the main equation:

$$\psi(t) = (T-\tau)^\nu = 0, \qquad (3)$$

is the power of a single linear factor. Then $(T-\tau)^\nu = 0$ is the lowest power of $T-\tau$ which is zero and the rank numbers of the ν first powers of $T-\tau$:

$$[\,(T-\tau)^0\,] = n,\ [\,T-\tau\,],\ [\,(T-\tau)^2\,],\ \ldots,\ [\,(T-\tau)^{\nu-1}\,]$$

form a non-increasing sequence of positive integers. The same applies, as need only be mentioned here (cf. op. cit. p.144 (4.)), from their ν differences $e_1 \geq e_2 \geq \cdots \geq e_\nu$, where generally:

$$e_i = [\,(T-\tau)^{i-1}\,] - [\,(T-\tau)^i\,], \quad (i = 1, 2, \ldots, \nu) \qquad (4)$$

is defined.

I now divide the system of nth order T into ν^2 partial systems:

$$T = (T_{i,k}), \quad (i, k = 1, 2, \ldots, \nu) \qquad (5)$$

that I combine in turn each $e_1, e_2, \ldots, e_\nu$ rows to a first, second, ..., νth horizontal section each, and proceed in the same way with the columns, so that in general $T_{i,k}$, consists of e_i rows and e_k columns. Then the system $T - \tau$, as shown above, p. 147 (6.), is equivalent to a reduced system:

$$T' - \tau = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \qquad (6)$$

which contains nonzero elements only in the first parallel diagonal right above the diagonal of zero, each time the unit systems corresponding to the partial systems $T_{1,2}$, $T_{2,3}$,.... Then it follows clearly that $(T'-\tau)^2$, $(T'-\tau)^3$, ..., $(T'-\tau)^{\nu-1}$ in sequence only have nonzero entries in the second, third, ..., $(\nu-1)$th upper parallel diagonals and are otherwise all zeros.

We now assume that already $T - \tau$ for the given system T has this reduced form. Then, for the power series to be investigated we get for $\mathfrak{B}(T)$ the equation (7)

$$\begin{aligned} \mathfrak{B}(T) &= \mathfrak{B}(\tau + (T-\tau)) \\ &= a_0 + a_1(\tau + (T-\tau) + a_2(\tau + (T-\tau))^2 + \cdots, \end{aligned}$$

and since the scalar τ is interchangeable with any other, since furthermore $(T-\tau)^\nu$ and all higher powers of $T - \tau$ are zero, and since finally the ν first powers of $T - \tau$, which are multiplied only by scalars, have all elements standing at different places, which cannot cancel each other, then one obtains for $\mathfrak{B}(T)$ the following simple representation (7a):

$$\mathfrak{B}(T) = \mathfrak{B}(\tau) + \mathfrak{B}'(\tau)(T-\tau) + \frac{\mathfrak{B}''(\tau)}{2!}(\tau)(T-\tau)^2 + \cdots + \frac{\mathfrak{B}^{(\nu-1)}(\tau)}{(\nu-1)!}(\tau)(T-\tau)^{\nu-1}.$$

Then, $\mathfrak{B}(T)$ converges if and only if the ν first derivatives

$$\mathfrak{B}(t),\ \mathfrak{B}'(t), \ldots, \mathfrak{B}^{(\nu-1)}(t)$$

for the ν-fold root $t = \tau$ of the main equation for T converge. Thus $\mathfrak{B}(T)$ diverges if τ lies outside the circle of convergence $\mathfrak{R}_\rho$, of $\mathfrak{B}(t)$, and $\mathfrak{B}(T)$ converges if τ is inside it, because then not only $\mathfrak{B}(t)$ but also all its derivatives converge for $t = \tau$. If, on the other hand, τ lies on the periphery of the circle of convergence, then there is convergence for $t = T$ if and only if the first ν derivatives of $\mathfrak{B}(t)$ are convergent for this ν-fold root τ of the main equation. But if in this case only the last of them $\mathfrak{B}^{(\nu-1)}(\tau))$ converges, the same applies to all earlier derivatives according to a known proposition (cf. e.g. Pringsheim, Functional Theory, p. 321). This condition is necessary in this case and sufficient for the convergence of $\mathfrak{B}(T)$.

Since in the most general case that the main equation for T has the form (6.):

$$\psi(t) = \prod(t - \bar{\tau})^{\bar{\nu}},$$

for each of the r partial systems $T'_{i,i}$ resulting there in (2.) the same applies, so the following very general theorem finally results:

A power series $\mathfrak{B}(t)$ diverges for $t = T$, if only one of the r-roots of the main equation $\psi(t) = 0$ for T lies outside the convergence circle $\mathfrak{R}_\rho$, it always converges when all r-roots are inside $\mathfrak{R}_\rho$. For every $\bar{\nu}$ finite that lies on the periphery of $\mathfrak{R}_\rho$, $\mathfrak{B}^{(\bar{\nu}-1)}(\bar{\tau})$ must be convergent if $\bar{\tau}$ is an $\bar{\nu}$-fold root of the main equation of T.

E. Weyr established and proved only the first part of this general theorem in the above-mentioned treatise.

Obviously the theorem can also be expressed in the following form:

The power series $\mathfrak{B}(t)$ converges for $t = T$ if and only if for each root $\bar{\tau}$ of the characteristic equation for T, $\mathfrak{B}^{(\bar{\nu}-1)}(\bar{\tau})$ converges when $\bar{\tau}$ is an $\bar{\nu}$-fold root of the same.

All whole transcendent functions like e^t, $\sin t$, $\cos t$,... whose functional elements $\mathfrak{B}(t)$ converge in the whole complex plane, are therefore also defined for each argument $T = (t_{i,k})$. But since

$$\ln(1+t) = \mathfrak{B}(t) = t - \frac{t^2}{2} + \frac{t^3}{3} - \cdots$$

converges for every t in and on the unit circle with the exception of $t = -1$, but otherwise diverges, then the function $\ln(1+T)$ defined by the series

$$\mathfrak{B}(T) = T - \frac{T^2}{2} + \frac{T^3}{3} - \cdots$$

also converges for all the systems T, whose roots $\bar{\tau}$ lie in or on the unit circle, but are not equal to -1. Since at last the power series

$$\mathfrak{B}(t) = \sum_1^\infty \frac{t^{\nu-1}}{\nu}$$

has the unit circle as convergence region, and on its periphery is absolutely convergent, and since

$$\mathfrak{B}'(t) = \sum_{1}^{\infty} \frac{t^\nu}{\nu^2}$$

diverges on that circle only for $t = 1$, while

$$\mathfrak{B}''(t) = \sum \frac{\nu - 1}{\nu} t^{\nu - 2}$$

diverges on the same circle throughout, then $\mathfrak{B}(T)$ converges if and only if for all roots $\bar{\tau}$ of the main equation of T $|\bar{\tau}| \leq 1$, with the restriction that $\bar{\tau} = 1$ may be at most a single, all other $\bar{\tau}$ with $|\bar{\tau}| = 1$ be at most double roots of the same.