In other words, formula (6.23) implies that $-\boldsymbol{\varepsilon}^{(p)}$ is the residue of the vector function $\lambda^{p-1}\boldsymbol{w}(\lambda)$ at infinity.

Next, according to inequality (6.21), all the eigenvalues of the operator $\boldsymbol{B}$ belong to the disk of radius $\rho < 1$ centered at the origin on the complex plane: $|\lambda_j| \le \rho < 1,\ j = 1, 2, \dots, n$. Then the integrand in the second integral of formula (6.23) is an analytic vector function of λ outside of this disk, i.e., for $|\lambda| > \rho$, because the operator $(\boldsymbol{B} - \lambda\boldsymbol{I})^{-1}$ exists (i.e., is bounded) for all $\lambda :\ |\lambda| > \rho$. This function is the analytic continuation of the function $\lambda^{p-1}\boldsymbol{w}(\lambda)$, where $\boldsymbol{w}(\lambda)$ is originally defined by the series (6.22) that can only be proven to converge outside of a larger disk $|\lambda| \le \|\boldsymbol{B}\| + \eta$. Consequently, the contour of integration in (6.23) can be altered, and instead of $r \ge \|\boldsymbol{B}\| + \eta$ one can take $r = \rho + \zeta$, where $\zeta > 0$ is arbitrary, without changing the value of the integral. Therefore, the error can be estimated as follows:

$$\begin{aligned} \|\boldsymbol{\varepsilon}^{(p)}\| &= \frac{1}{2\pi}\left\| \int\limits_{|\lambda|=\rho+\zeta} \lambda^p (\boldsymbol{B} - \lambda\boldsymbol{I})^{-1}\boldsymbol{\varepsilon}^{(0)} d\lambda \right\| \\ &\le (\rho+\zeta)^{p+1} \max_{|\lambda|=\rho+\zeta} \|(\boldsymbol{B} - \lambda\boldsymbol{I})^{-1}\| \|\boldsymbol{\varepsilon}^{(0)}\|. \end{aligned} \tag{6.24}$$

In formula (6.24), let us take $\zeta > 0$ sufficiently small so that $\rho + \zeta < 1$. Then, the right-hand side of inequality (6.24) vanishes as p increases, which implies the convergence: $\|\boldsymbol{\varepsilon}^{(p)}\| \longrightarrow 0$ when $p \longrightarrow \infty$. This completes the proof of sufficiency.

To prove the necessity, suppose that inequality (6.21) does not hold, i.e., that for some λ_k we have $|\lambda_k| \ge 1$. At the same time, contrary to the conclusion of the theorem, let us assume that the convergence still takes place for any choice of $\boldsymbol{x}^{(0)}$: $\boldsymbol{x}^{(p)} \longrightarrow \boldsymbol{x}$ as $p \longrightarrow \infty$. Then we can choose $\boldsymbol{x}^{(0)}$ so that $\boldsymbol{\varepsilon}^{(0)} = \boldsymbol{x} - \boldsymbol{x}^{(0)} = \boldsymbol{e}_k$, where $\boldsymbol{e}_k$ is the eigenvector of the operator $\boldsymbol{B}$ that corresponds to the eigenvalue λ_k. In this case, $\boldsymbol{\varepsilon}^{(p)} = \boldsymbol{B}^p\boldsymbol{\varepsilon}^{(0)} = \boldsymbol{B}^p\boldsymbol{e}_k = \lambda_k^p\boldsymbol{e}_k$. As $|\lambda_k| \ge 1$, the sequence $\lambda_k^p\boldsymbol{e}_k$ does not converge to $\boldsymbol{0}$ when p increases. The contradiction proves the necessity. ∎

REMARK 6.2 Let us make an interesting and important observation of a situation that we encounter here for the first time. The problem of computing the limit $\boldsymbol{x} = \lim\limits_{p\to\infty} \boldsymbol{x}^{(p)}$ is ultimately well conditioned, because the result $\boldsymbol{x}$ does not depend on the initial data at all, i.e., it does not depend on the initial guess $\boldsymbol{x}^{(0)}$. Yet the algorithm for computing the sequence $\boldsymbol{x}^{(p)}$ that converges according to Theorem 6.2 may still appear computationally unstable. The instability may take place if along with the inequality $\max_j |\lambda_j| = \rho < 1$ we have $\|\boldsymbol{B}\| > 1$. This situation is typical for non-self-adjoint (or non-normal) matrices $\boldsymbol{B}$ (opposite of Theorem 5.2).

Indeed, if $\|\boldsymbol{B}\| < 1$, then the norm of the error $\|\boldsymbol{\varepsilon}^{(p)}\| = \|\boldsymbol{B}^p\boldsymbol{\varepsilon}^{(0)}\|$ decreases monotonically, this is the result of Theorem 6.1. Otherwise, if $\|\boldsymbol{B}\| > 1$, then for some $\boldsymbol{\varepsilon}^{(0)}$ the norm $\|\boldsymbol{\varepsilon}^{(p)}\|$ will initially grow, and only then decrease. The