Arch. Hist. Exact Sci. (2008) 62:655–717
DOI 10.1007/s00407-008-0026-x

Continued fractions and the origins of the Perron–Frobenius theorem

Thomas Hawkins

Received: 22 October 2007 / Published online: 29 July 2008

Abstract The theory of nonnegative matrices is an example of a theory motivated in its origins and development by purely mathematical concerns that later proved to have a remarkably broad spectrum of applications to such diverse fields as probability theory, numerical analysis, economics, dynamical programming, and demography. At the heart of the theory is what is usually known as the Perron–Frobenius Theorem. It was inspired by a theorem of Oskar Perron on positive matrices, usually called Perron's Theorem. This paper is primarily concerned with the origins of Perron's Theorem in his masterful work on ordinary and generalized continued fractions (1907) and its role in inspiring the remarkable work of Frobenius on nonnegative matrices (1912) that produced, inter alia, the Perron–Frobenius Theorem. The paper is not at all intended exclusively for readers with expertise in the theory of nonnegative matrices. Anyone with a basic grounding in linear algebra should be able to read this article and come away with a good understanding of the Perron–Frobenius Theorem as well as its historical origins. The final section of the paper considers the first major application of the Perron–Frobenius Theorem, namely, to the theory of Markov chains. When he introduced the eponymous chains in 1908, Markov adumbrated several key notions and results of the Perron–Frobenius theory albeit within the much simpler context of stochastic matrices; but it was by means of Frobenius' 1912 paper that the linear algebraic foundations of Markov's theory for nonpositive stochastic matrices were first established by R. Von Mises and V.I. Romanovsky.

Communicated by J.J. Gray.

T. Hawkins (✉)
Department of Mathematics and Statistics, Boston University,
111 Cummington Street, Boston, MA 02215, USA
e-mail: twh@math.bu.edu

Contents

1 Introductory overview

The theory of nonnegative matrices is an example of a substantial component of theoretical linear algebra with a remarkably broad spectrum of applications to such diverse fields as probability theory, numerical analysis, economics, dynamical programming, and demography.[1] At the heart of the theory, sometimes referred to generically as "Perron–Frobenius theory," is what is usually known as the Perron–Frobenius Theorem. This theorem, which is contained in Frobenius' Theorems 4.5–4.8 of Sect. 4, was inspired by a theorem due to Oskar Perron (1880–1975), given below in Sect. 3.4 as Theorem 3.9 and usually called Perron's Theorem.

This essay is concerned with the origins of Perron's Theorem, which he published in a brief paper in 1907, and its role in inspiring the work of Frobenius on nonnegative matrices (1912) that produced, inter alia, Theorems 4.5–4.8.[2] The essay is not at all

[1] For an overview of the theory and some of its applications see [3,56,57,60,64].

[2] The research on which this essay is based was supported by the NSF Science and Technology Studies program under grant SES-0312697. I am indebted to Hans Schneider for his paper [53], which served as the starting point for my own research, and for generously sharing his expertise on the theory of nonnegative matrices with me in many communications. I also wish to thank Wilfried Parys for information on Maurice Potron and for detecting several bibliographical errors and expositional lapses in a draft of this paper.

intended exclusively for readers with expertise in the theory of nonnegative matrices. Anyone with a basic grounding in linear algebra should be able to read this essay and come away with a good understanding of the Perron–Frobenius Theorem as well as its historical origins.

Perron began his studies at the University of Munich and then, as was commonplace, spent semesters at several other universities—Berlin, Tübingen, and Göttingen in his case—but Munich was his mathematical home base.[3] In 1902 he obtained his doctorate there with a dissertation on a problem involving the rotation of a rigid body. He wrote it under the direction Ferdinand Lindemann, who is remembered nowadays for his proof that π is a transcendental number. Also at the University of Munich was Alfred Pringsheim (1850–1941), who was renowned as a brilliant lecturer and conversationalist.[4] Perron's post-doctoral research interests turned in the direction of Pringsheim's current work, which was strongly influenced by Weierstrass' emphasis upon analytical rigor. According to Perron [39, p. 2],

> Although Pringsheim was not directly a student of Weierstrass, in Germany he stands as the most ardent and successful propagandist for function theory in the mold of Weierstrass .... A theme to which he returned time and again with alacrity was the convergence and divergence of infinite processes .... A class of infinite processes that had been created by Euler and later almost entirely neglected was that of continued fractions, which he infused with new life and for which he developed simple yet far-reaching convergence criteria.

As we shall see in Sect. 2, Perron began to contribute to Pringsheim's research program of determining convergence criteria for broad types of continued fractions. One such class was the periodic continued fractions for which Otto Stolz had provided a convergence criterion in 1886. Stolz' criterion was correct but unsatisfactory in several respects and became one of the subjects of Pringsheim's research. Pringsheim's published work was in general meticulous in its rigor but lacking in major original ideas [11, p. 149] and this was true in particular of his work on Stolz' Theorem 2.1. Perron, whose own work adhered to the high standard of rigor set by his mentor, applied that standard albeit in a more original manner by introducing a new and simpler way of characterizing Stolz' criterion, which, he realized, applied as well to the generalized continued fractions introduced by Jacobi in a paper published posthumously in 1868. Although a few mathematicians had dealt with Jacobi's generalized continued fractions, general convergence criteria were completely lacking. Here then was an opportunity to bring Pringsheim's research program to bear upon a vastly more general class of infinite processes. To accomplish this became the goal of Perron's *Habilitationsschrift*, which was published in 1907.

In Sect. 3 I show how Perron's novel approach to Jacobi's generalized continued fractions led him to the properties of nonnegative matrices posited in what is now known as Perron's Theorem (Theorem 3.9) and its historically consequential corollary (Corollary 3.11). The corollary is consequential because it asserts that the remar-

[3] For details on Perron's life and family background see [10].

[4] The information about Pringsheim is based upon Perron's memorial essay [39].

kable properties of positive matrices posited by Perron's Theorem actually hold for all nonnegative matrices A with the property that A^ν has all positive entries for some power ν, thereby suggesting the possibility of a comparable theory for a larger class of nonnegative matrices. As we shall see in Sect. 3.3, in Perron's work on periodic generalized continued fractions with nonnegative coefficients, the matrices associated to the continued fraction algorithm are nonnegative but not in general positive; however, they always possess a νth power that is positive. Thus Perron's Corollary was central to his generalized theory of continued fractions and was the motivation for a slightly weaker version of Perron's Theorem, namely Lemma 3.5, which then led Perron to the full-fledged Perron Theorem and its Corollary. Furthermore, although the nonnegative matrices associated by Perron's method to Jacobi's generalized continued fraction process were generated by the algorithm associated to the process and so not arbitrary nonnegative matrices, the ideas that went into Perron's proof of his convergence theorem (Theorem 3.8) provided him with the outline for the proof of his general theorem and corollary; he discovered that much of what had been established within the context of his version of the Jacobi theory could be readily extricated from that context.

One step in the extrication process, however, proved cumbersome and forced Perron to prove a complicated lemma involving limits (Lemma 3.10), which he himself deemed an unwelcome intrusion into what he regarded as a purely algebraic theorem for which a proof by the customary algebraic means was a desideratum. It was this desideratum that motivated Frobenius' involvement with Perron's work, and in papers of 1908–1909 Frobenius supplied Perron's Theorem, in a slightly generalized form, with a far simpler proof that avoided Lemma 3.10. The further simplifications introduced in his 1909 paper enabled him to prove a sort of converse to Perron's Theorem. Perron's Theorem implied (as Perron realized) that associated to the maximal positive characteristic root ρ_0 posited by his theorem for a positive matrix is a positive characteristic vector.[5] Frobenius showed that no other characteristic root of a positive matrix possesses a nonnegative characteristic vector. This led him to pose the more general problem of determining for a nonnegative matrix which characteristic roots possess nonnegative characteristic vectors. This problem turned out to be consequential since, according to Frobenius, in seeking to solve it he was led to his remarkable theory of nonnegative matrices. In Sect. 4.2 I suggest how the problem may have led to the key notions of his theory.

Nowadays in many applications of the theory of nonnegative matrices the existence of nonnegative characteristic vectors is of direct importance to the application, but neither Frobenius nor Perron indicated any awareness of such applications, which may not have existed at the time.[6] Indeed, the entire history of the Perron–Frobenius Theo-

[5] I prefer to use the term "characteristic vector" rather than "eigenvector".

[6] Two such present-day applications—Markov chains and input–output type economic analysis—existed at the time of Frobenius' work on positive and nonnegative matrices (1908–1912). Markov introduced the eponymous chains in 1908 (see Sect. 5.1) and the Jesuit mathematician Maurice Potron announced an economic theory analogous to input–output analysis in 1911 [40,41] with details given in 1913 [42]. Although Frobenius was apparently unaware of these developments, it is of interest to note that neither Markov nor Potron ascribed an importance to nonnegative characteristic vectors in their respective applications. For further general information about Potron, whose work remained unappreciated until recently, see [1,4,5].

rem from Perron's work to that of Frobenius, provides another significant example of the manner in which mathematics, done for purely theoretical, "non-applied" reasons, turned out to supply an appropriate theoretical foundation for applications outside the realm of pure mathematics.

Sections 2–4 provide a full account of the origins of the Perron–Frobenius Theorem. Nonetheless I would be remiss if I passed over a paper presented by A.A. Markov in December of 1907 (the year Perron published his theorem) and published in 1908. As we shall see in Sect. 5.1, in his paper Markov introduced the theory of what are now rightly called Markov chains. This involved associating to a given chain a matrix P of probabilities, namely what is now called the stochastic matrix of transition probabilities associated to the chain. Such a matrix is of course a special type of nonnegative matrix and it is easily seen from the stochastic nature of P that $\rho_0 = 1$ is a characteristic root and that all other characteristic roots ρ satisfy $|\rho| \leq 1$. In order to push through the probabilistic analysis, however, Markov needed to show that $\rho_0 = 1$ is a simple root, i.e., a root of multiplicity one, and that $|\rho| < 1$ for all other characteristic roots ρ of P. He was apparently unaware of Perron's 1907 paper containing his Theorem 3.9 and Corollary 3.11, which shows that P has the above properties if all its entries are positive or if P is nonnegative with P^ν positive for some power ν. Markov realized that the above properties of the characteristic roots of P do not hold for all nonnegative P, but his efforts to characterize those nonnegative P for which the conclusions do hold were fraught with ambiguities. Nonetheless, I believe it is possible to characterize the P for which his reasoning is valid and to see that, in the light of Frobenius' theory of nonnegative matrices, he had determined a proper subclass of the nonnegative P for which the desired conclusions hold. (A detailed justification of my claims in Sect. 5.1 is given in Appendix 6.3.) Due to the ambiguity of Markov's reasoning and a lack of widespread interest in the theory of Markov chains in the years immediately following 1908, it was not until the 1930s that the linear algebraic aspects of Markov's theory were clarified using Frobenius' theory of nonnegative matrices, as I indicate in Sect. 5.2. The developments sketched in Sect. 5.2 seem to have served as a paradigm for subsequent applications of Perron–Frobenius theory.

Although the general theory that Frobenius built up around Perron's Theorem constituted a remarkable and profound piece of mathematics, it is important to keep in mind that it owed its inspiration to Perron's equally remarkable Theorem, which in turn was a by-product of his highly nontrivial and fertile work on the generalized continued fractions associated to Jacobi's algorithm. It is true that Markov rediscovered Perron's Theorem (but not his Corollary) for the class of positive stochastic matrices, but the stochastic hypothesis is so strong that the existence of a maximal positive root $\rho_0 = 1$ such that $|\rho| \leq \rho_0$ for any other characteristic root ρ is an easy and elementary consequence, whereas this was not the case at the time for arbitrary positive matrices. It was Perron's masterful study of Jacobi's continued fraction algorithm that brought the properties of positive matrices posited in Perron's Theorem to light. Perron was a talented and creative mathematician, whose work has not been accorded the historical attention it deserves. It is my hope that this first step towards a greater and more informed historical appreciation of his mathematical contributions will encourage others to take further steps in this direction.

2 Continued fractions

A continued fraction is a formal expression of the form

$$a_0 + \frac{a_1}{b_1 + \frac{a_2}{b_2 + \cdots}}, \tag{2.1}$$

where the coefficients $a_0, a_1, a_2, \ldots, b_1, b_2, \ldots$ can be any real or complex numbers. As we shall see in Sect. 3.1 such expressions with positive integer coefficients are naturally suggested by the Euclidean algorithm.[7] I will use the following notation, which is due to Pringsheim, to express (2.1) in the typographically simpler form

$$a_0 + \frac{a_1 \,|}{|\, b_1} + \frac{a_2 \,|}{|\, b_2} + \cdots. \tag{2.2}$$

The continued fraction (2.2) is said to converge if the sequence of partial fractions

$$S_n = a_0 + \frac{a_1 \,|}{|\, b_1} + \frac{a_2 \,|}{|\, b_2} + \cdots + \frac{a_n \,|}{|\, b_n} \tag{2.3}$$

has a finite limit S as $n \to \infty$. In this case the continued fraction (2.2) is said to converge to S.

In the sessions of 1898, 1899, and 1900 of the Munich Academy of Sciences, Pringsheim presented three papers on continued fractions that particularly engaged Perron's interest. In his first paper [43] Pringsheim called attention to the need for general convergence criteria for continued fractions. In the second paper [44], he focused on the same problem but for the special case in which the sequences a_n and b_n are positive real numbers. In this case there was a result due to Seidel (1846) and Stern (1848) to the effect that any continued fraction (2.2) with positive coefficients converges if and only if at least one of the two series diverges:

$$\sum_{n=1}^{\infty} \frac{a_1 a_3 \cdots a_{2n-1}}{a_2 a_4 \cdots a_{2n}} b_{2n}, \quad \sum_{n=1}^{\infty} \frac{a_2 a_4 \cdots a_{2n}}{a_1 a_3 \cdots a_{2n-1}} b_{2n+1}. \tag{2.4}$$

Starting from this result, Pringsheim was able to state a simpler necessary condition for convergence: a continued fraction (2.2) with positive coefficients converges provided the series $\sum_{n=1}^{\infty} \sqrt{\frac{b_n b_{n+1}}{a_{n+1}}}$ diverges [44, p. 267]. In his third paper [45] Pringsheim considered the problem of general convergence criteria in another case in which results were already known, namely continued fractions with possibly complex, but periodic, coefficients (defined below).

In a paper presented to the Munich Academy in 1905 [34], Perron turned to the subject of Pringsheim's second paper [44], namely continued fractions with positive

[7] For an overview of the history of continued fractions, see [7].

terms. Pringsheim's result, like that of Stern and Seidel from which it derived, had related the convergence of a continued fraction to that of an infinite series formed from its coefficients. Perron, by contrast, sought to obtain conditions directly on the coefficients a_n, b_n that would guarantee convergence. For example, a special case of one of his results implied that if

$$\limsup_{n\to\infty}\left(\frac{b_n b_{n+1}}{a_{n+1}}+\frac{a_{n+2}b_n b_{n+3}}{a_{n+1}(a_{n+3}+b_{n+2}b_{n+3})}\right)>0$$

then the continued fraction converges [34, p. 321]. As Perron pointed out with examples, his result implied convergence in cases left undecided by Pringsheim's criterion. In another paper of 1905, presented five months later, Perron took up the subject of Pringsheim's third paper [45] on the convergence of continued fractions with periodic coefficients. As we shall see, this work by Perron proved inspirational, giving him the idea for his *Habilitationsschrift* and leading thereby to his theorem on positive matrices.

2.1 Stolz' theorem

The starting point for both Pringsheim's and Perron's work on periodic continued fractions was a theorem due to Otto Stolz (1842–1905). In 1885–86 Stolz, who was a professor at the University of Innsbruck, published a two-volume series of lectures on what he called "general arithmetic from the modern viewpoint." The second volume [58] was devoted to the arithmetic of complex numbers and its final chapter considered the subject of continued fractions. In particular, Stolz considered continued fractions $\frac{a_1|}{|b_1}+\frac{a_2|}{|b_2}+\cdots$ with possibly complex but periodic coefficients [58, p. 299ff]. Here it will suffice to consider purely periodic continued fractions, i.e., those that are periodic from the outset. Thus if the positive integer $\mathbf{k}$ denotes the period, the coefficients of the continued fraction satisfy

$$a_{i\mathbf{k}+j}=a_j,\quad b_{i\mathbf{k}+j}=b_j,\quad j=1,\ldots,\mathbf{k},\ \ i=1,2,3,\ldots. \tag{2.5}$$

If the continued fraction converges to x, i.e., if $x=\lim_{n\to\infty}S_n$ in the notation of (2.3) (with $a_0=0$), then $x=\frac{a_{\mathbf{k}+1}|}{|b_{\mathbf{k}+1}}+\frac{a_{\mathbf{k}+2}|}{|b_{\mathbf{k}+2}}+\cdots$ due to the assumed periodicity, and so

$$x=\frac{a_1|}{|b_1}+\frac{a_2|}{|b_2}+\cdots+\frac{a_{\mathbf{k}}|}{|b_{\mathbf{k}}+x}. \tag{2.6}$$

For any (not necessarily periodic) continued fraction Stolz also denoted the νth partial continued fraction S_ν in the form

$$S_\nu=\frac{A_\nu}{B_\nu}, \tag{2.7}$$

where A_ν and B_ν are the numerator and denominator of S_ν when expressed as a simple fraction. Thus, e.g.,

$$S_3 = \frac{a_1\,|}{|\,b_1} + \frac{a_2\,|}{|\,b_2} + \frac{a_3\,|}{|\,b_3} = \frac{a_1a_3 + a_1b_2b_3}{a_3b_1 + a_2b_3 + b_1b_2b_3},$$

and so $A_3 = a_1a_3 + a_1b_2b_3$ and $B_3 = a_3b_1 + a_2b_3 + b_1b_2b_3$. Stolz established the following recurrence relations satisfied by A_ν and B_ν [58, p.267, eq. (II)].

$$\begin{array}{lll} A_0 = 0, & A_1 = a_1, & A_{\nu+2} = a_{\nu+2}A_\nu + b_{\nu+2}A_{\nu+1} \\ B_0 = 1, & B_1 = b_1, & B_{\nu+2} = a_{\nu+2}B_\nu + b_{\nu+2}B_{\nu+1}. \end{array} \tag{2.8}$$

Using them [58, p. 299], he was able to express (2.6) in the form

$$x = \frac{A_{\mathbf{k}-1}x + A_{\mathbf{k}}}{B_{\mathbf{k}-1}x + B_{\mathbf{k}}}, \tag{2.9}$$

which can be rewritten as a quadratic equation in x:

$$B_{\mathbf{k}-1}x^2 + (B_{\mathbf{k}} - A_{\mathbf{k}-1})\,x - A_{\mathbf{k}} = 0. \tag{2.10}$$

Stolz then proceeded to present the first general convergence theorem for periodic continued fractions [58, pp. 300–302]. It may be summed up as follows.

Theorem 2.1 (Stolz' Theorem) *Let* $\frac{a_1\,|}{|\,b_1} + \frac{a_2\,|}{|\,b_2} + \cdots$ *denote a continued fraction satisfying the periodicity relations* (2.5) *and hence periodic with period* $\mathbf{k}$. *Then the condition*

$$B_{\mathbf{k}-1} \neq 0 \tag{A}$$

is a necessary condition for convergence. To obtain necessary and sufficient conditions two cases must be distinguished. Case I. Suppose the quadratic Eq. (2.10) *has a double root. Then condition* (A) *is also sufficient and the continued fraction converges to this root. Case II. Suppose* (2.10) *has two distinct roots* x_0 *and* x_1. *Then it is also necessary that*

$$|B_{\mathbf{k}} + x_0B_{\mathbf{k}-1}| \neq |B_{\mathbf{k}} + x_1B_{\mathbf{k}-1}|. \tag{B}$$

Assuming (B) *holds, let the notation for* x_0, x_1 *be chosen so that*

$$|B_{\mathbf{k}} + x_0B_{\mathbf{k}-1}| > |B_{\mathbf{k}} + x_1B_{\mathbf{k}-1}|. \tag{B$'$}$$

Then a further necessary condition is

$$A_\nu - x_1B_\nu \neq 0, \quad \nu = 1, 2, \ldots, \mathbf{k} - 2. \tag{C}$$

In Case II, conditions (A), (B$'$) *and* (C) *are necessary and sufficient for convergence and the continued fraction converges to* x_0.

With this theorem Stolz had certainly given a definitive answer to the question of the convergence of periodic continued fractions, but his formulation of his results and the concomitant proofs were neither simple nor insightful. Pringsheim's 1900 paper [45] aimed to improve on this aspect of Stolz' work. The source of the dissatisfaction with Stolz' Theorem involved Case II, the case of distinct roots in the quadratic Eq. (2.10). The roots of this equation are by the usual formula

$$x_i = \left[(A_{\mathbf{k}-1} - B_{\mathbf{k}}) + \varepsilon_i\sqrt{D}\right]/2B_{\mathbf{k}-1}, \quad i = 0, 1, \tag{2.11}$$

where $D = (A_{\mathbf{k}-1} - B_{\mathbf{k}})^2 + 4A_{\mathbf{k}}B_{\mathbf{k}-1}$ is the discriminant, $\sqrt{D}$ is a fixed branch of the square root, and the yet to be determined factors $\varepsilon_i = \pm 1$ are to be chosen so that $\varepsilon_0\varepsilon_1 = -1$, i.e., so that (2.11) gives the two roots. Pringsheim then reworked Case II so that Stolz' condition (B) was replaced by

$$\Re e(S/\sqrt{D}) \neq 0, \quad \text{where } S = A_{\mathbf{k}-1} + B_{\mathbf{k}}. \tag{D}$$

Pringsheim's condition implies that $S \pm \sqrt{D}$ do not have the same absolute value and so the ε_i in (2.11) may be chosen so that $|S - \varepsilon_1\sqrt{D}| < |S + \varepsilon_0\sqrt{D}|$, and x_0 is the value of the continued fraction in Case II. It thus became clearer that in Case II the two distinct roots $x_i = (S + \varepsilon_i\sqrt{D})/2B_{\mathbf{k}-1}$ of the quadratic equation must have different absolute values and the continued fraction converges to the root with the larger absolute value.

2.2 Perron's new approach to Stolz' theorem

Stolz' Condition (B) seemed unenlightening because his lengthy proof lacked any intuitive motivation. It was Perron who discovered the underlying reason why Condition (B) made sense and this enabled him to state Stolz's theorem more clearly and simply than either Stolz or Pringsheim. He showed how to do this in his second paper of 1905 [35]. The discovery on his part was especially important because it revealed to him a new and promising approach to Jacobi's algorithm, an approach which he developed in his *Habilitationsschrift* of 1907 (Sect. 3) and which in turn led him to his theorem on positive matrices, which was also published in 1907 (Sect. 3.4).

Perron observed that the source of Stolz' quadratic Eq. (2.10), namely (2.9), can be expressed in the homogeneous form

$$\rho x = A_{\mathbf{k}-1}x + A_{\mathbf{k}}, \quad \rho = B_{\mathbf{k}-1}x + B_{\mathbf{k}}, \tag{2.12}$$

which I will express in the more suggestive form

$$A\mathbf{v} = \rho\mathbf{v}, \quad A = \begin{pmatrix} A_{\mathbf{k}-1} & A_{\mathbf{k}} \\ B_{\mathbf{k}-1} & B_{\mathbf{k}} \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} x & 1 \end{pmatrix}^t. \tag{2.13}$$

Although Perron did not use any matrix notation, he observed [35, p. 497] that elimination of x in (2.12) shows that ρ is a root of the quadratic equation

$$f(\rho) = \begin{vmatrix} A_{\mathbf{k}-1} - \rho & A_{\mathbf{k}} \\ B_{\mathbf{k}-1} & B_{\mathbf{k}} - \rho \end{vmatrix} = 0, \tag{2.14}$$

which is of course the characteristic equation of A.

If x_0, x_1 denote the roots of Stolz' quadratic Eq. (2.10) as specified by condition (B$'$) of his Theorem 2.1 and if ρ_0, ρ_1 are the corresponding values from (2.12), then we see that $\rho_i = B_{\mathbf{k}-1}x_i + B_{\mathbf{k}}$ for $i = 0, 1$ and Stolz' condition (B$'$) is that $|\rho_0| > |\rho_1|$. Furthermore since $B_{\mathbf{k}-1} \neq 0$ [Condition (A)] it is easily seen that Case II of Stolz' Theorem 2.1 (Stolz' quadratic equation has distinct roots) corresponds to the roots ρ_i of the characteristic Eq. (2.14) being distinct, so that in Case II Stolz' condition (B)/(B$'$) simply says that the characteristic equation must have a root that is strictly greater in absolute value than the other root.

Perron not only reformulated Stolz' condition (B)/(B$'$) in terms of the roots of the characteristic equation, but he gave an entirely different proof of the necessity of the condition $|\rho_0| > |\rho_1|$ in the case of distinct roots. He utilized the recurrence relations (2.8) to deduce from the assumed periodicity that

$$\begin{aligned} A_{i\mathbf{k}+j} &= A_{(i-1)\mathbf{k}+j}A_{\mathbf{k}-1} + B_{(i-1)\mathbf{k}+j}A_j \\ B_{i\mathbf{k}+j} &= A_{(i-1)\mathbf{k}+j}B_{\mathbf{k}-1} + B_{(i-1)\mathbf{k}+j}B_j. \end{aligned} \tag{2.15}$$

These relations formed the starting point of his proof that, of the two roots, one of them, denoted below by ρ_0, is such that

$$L = \lim_{j\to\infty} \left(\frac{\rho_1}{\rho_0}\right)^j \tag{2.16}$$

exists as a finite number [35, Sect. 2], from which it follows immediately that $L = 0$ and $|\rho_0| > |\rho_1|$. Perron was the first to realize the important role of the coefficient matrix A and its characteristic roots in the study of periodic continued fractions. From this new vantage point he also proved that when the characteristic roots ρ of A are equal, the periodic continued fraction always converges [34, Sect. 3].

Although Perron seems to have reasoned directly in terms of iteration relations such as (2.15), it should be noted that if we introduce the 2×1 column matrix

$$\mathbf{C}_\nu = \left(A_\nu \ B_\nu\right)^t, \tag{2.17}$$

then the above equations state that $A\mathbf{C}_{(i-1)\mathbf{k}+j} = \mathbf{C}_{i\mathbf{k}+j}$ for any $i \geq 1$. This relationship, and thus (2.15), follows by iteration from the case $i = 1$, i.e.,

$$A\mathbf{C}_j = \mathbf{C}_{\mathbf{k}+j}. \tag{2.18}$$

3 Perron's Habilitationsschrift

Perron pointed out in the beginning of his 1905 paper described above that not only do his new methods "derive the basic formulas of Stolz in a rational manner" but in addition, "As I will show elsewhere, my procedure has the additional advantage that by means of a natural extension of it, the convergence of the general Jacobi continued fraction algorithms can be decided" [35, p. 495]. Perron showed this in his *Habilitationsschrift*, which was published in *Mathematische Annalen* in 1907 [36]. As we shall now see, in this tour de force extension of his work of 1905, he not only established the promised convergence criteria in the far more complicated case of Jacobi's algorithm, but he did it in a manner that led naturally to his theorem on positive matrices.

3.1 Jacobi's continued fraction algorithm

It was well-known that the Euclidean division algorithm, which produces the greatest common divisor of two positive integers x_0, x_1 is related to a continued fraction expansion. To indicate the connection as Perron did [36, p. 2], let $x_0 < x_1$ and write the algorithm as follows:

$$x_1 = ax_0 + r', \quad x_0 = a'r' + r'', \quad r' = a''r'' + r''', \dots. \tag{3.1}$$

Thus $r', r'', r''', \dots$ are the respective remainders and satisfy $x_0 > r' > r'' > r''' > \dots \geq 0$. The general term in (3.1) is

$$r^{(\nu-1)} = a^{(\nu)} r^{(\nu)} + r^{(\nu+1)}.$$

Of course, since x_0, x_1 are positive integers we will eventually get $r^{(\nu+1)} = 0$ and $r^{(\nu)}$ will be the greatest common divisor of x_0, x_1.

If we set

$$\sigma = \frac{x_1}{x_0}, \quad \sigma' = \frac{x_0}{r'}, \quad \sigma'' = \frac{r'}{r''}, \dots, \quad \sigma^{(\nu)} = \frac{r^{(\nu-1)}}{r^{(\nu)}}, \tag{3.2}$$

then from (3.1) we obtain

$$\sigma = x_0/x_1 = a' + r'/x_0 = a' + 1/\sigma', \quad \sigma' = x_0/r' = a' + r''/r' = a' + 1/\sigma'',$$

and so on, so that the algorithm can be written in the form

$$\sigma = a + \frac{1}{\sigma'}, \quad \sigma' = a' + \frac{1}{\sigma''}, \dots, \quad \sigma^{(\nu)} = a^{(\nu)}, \tag{3.3}$$

where the last term lacks $1/\sigma^{(\nu+1)}$ since $1/\sigma^{(\nu+1)} = r^{(\nu+1)}/r^{(\nu)} = 0$. By substituting for $\sigma', \sigma'', \dots, \sigma^{(\nu)}$ in (3.3) we obtain a finite continued fraction expansion for

$\sigma = x_0/x_1$:

$$\sigma = a + \frac{1\,|}{|\,a'} + \frac{1\,|}{|\,a''} + \cdots + \frac{1\,|}{|\,a^{(\nu)}}. \tag{3.4}$$

If x_0 and x_1 are positive real numbers such that $\sigma = x_1/x_0$ is irrational, the algorithm still makes sense but continues indefinitely since $r^{(\nu)} > 0$ for all ν. For this reason, instead of (3.4) we get

$$\sigma = a + \frac{1\,|}{|\,a'} + \frac{1\,|}{|\,a''} + \cdots + \frac{1\,|}{|\,a^{(\nu)} + 1/\sigma^{(\nu+1)}}. \tag{3.5}$$

It turns out that

$$\sigma = a + \frac{1\,|}{|\,a'} + \frac{1\,|}{|\,a''} + \cdots + \frac{1\,|}{|\,a^{(\nu)}} + \cdots. \tag{3.6}$$

Thus $\sigma = \lim_{\nu\to 0} S_\nu$, where $S_\nu = a + \frac{1\,|}{|\,a'} + \frac{1\,|}{|\,a''} + \cdots + \frac{1\,|}{|\,a^{(\nu)}}$. To appreciate the analogy with Perron's extension of these considerations to the context of Jacobi's generalized Euclidean algorithm, it should be noted that: (1) the right-hand side of (3.5) can be expressed as a rational function of $\sigma^{(\nu+1)}$ by undoing the nest of reciprocals; (2) if we set $\sigma^{(\nu+1)} = 0$ in this rational function we get the partial sum S_ν. (2) is easy to see because the final reciprocal in (3.5) is

$$\frac{1}{a^{(\nu)} + 1/\sigma^{(\nu+1)}} = \frac{\sigma^{(\nu+1)}}{\sigma^{(\nu+1)}a^{(\nu)} + 1},$$

which becomes 0 when we set $\sigma^{(\nu+1)} = 0$. Thus with $\sigma^{(\nu+1)} = 0$ (3.5) becomes $a + \frac{1\,|}{|\,a'} + \frac{1\,|}{|\,a''} + \cdots + \frac{1\,|}{|\,a^{(\nu-1)}}$, which is the partial sum S_ν corresponding to the first ν terms of the infinite continued fraction in (3.6)

In the above expressions the coefficients $a^{(\nu)}$ denote positive integers, but as we have seen, by Perron's time mathematicians such as Stolz and Pringsheim were considering the convergence of continued fractions

$$a + \frac{1\,|}{|\,a'} + \frac{1\,|}{|\,a''} + \cdots + \frac{1\,|}{|\,a^{(\nu)}} + \cdots,$$

in which the coefficients $a^{(\nu)}$ can be real or complex numbers; and they had obtained general convergence theorems for continued fractions with positive coefficients and for possibly complex coefficients that are periodic. It was these sort of results that Perron hoped to obtain for the analog of continued fractions associated to the generalized Euclidean algorithm introduced by Jacobi in a paper published posthumously in 1868 [26].

Here is the idea of Jacobi's algorithm.[8] The Euclidean algorithm gives the greatest common divisor of two positive integers. Suppose instead that one wishes an algorithm to determine the greatest common divisor of the three positive integers $\{x_0, x_1, x_2\}$. Proceed as follows. Replace $\{x_0, x_1, x_2\}$ by $\{x_0', x_1', x_2'\}$, where x_0' and x_1' are the respective remainders upon division of x_1 and x_2 by x_0 and $x_2' = x_0$. In the language of *Mathematica*

$$\{x_0, x_1, x_2\} \to \{\text{Mod}[x_1, x_0], \text{Mod}[x_2, x_0], x_0\} \equiv \{x_0', x_1', x_2'\}.$$

Then it can be checked that $\gcd\{x_0, x_1, x_2\} = \gcd\{x_0', x_1', x_2'\}$. Next replace $\{x_0', x_1', x_2'\}$ by $\{x_0'', x_1'', x_2''\}$ following the same algorithm, so that x_0'' and x_1'' are the respective remainders upon division by x_0' and $x_2'' = x_0'$. Then we have $\gcd\{x_0', x_1', x_2'\} = \gcd\{x_0'', x_1'', x_2''\}$, and the idea is to continue this process until the greatest common divisor of $\{x_0^{(\nu)}, x_1^{(\nu)}, x_2^{(\nu)}\}$ is obvious.

For example, suppose that $\{x_0, x_1, x_2\} = \{378, 210, 700\}$. Five iterations of the above-described process yields:

$$\begin{aligned} \{378, 210, 700\} &\to \{210, 322, 378\} \to \{112, 168, 210\} \to \\ \{56, 98, 112\} &\to \quad \{42, 0, 56\} \to \quad \{0, 14, 42\}. \end{aligned}$$

Thus $= \{x_0^{(5)}, x_1^{(5)}, x_2^{(5)}\} = \{0, 14, 42\}$. The next step would call for division by $x_0^{(5)} = 0$, and so the process with 3 numbers stops and continues with the two numbers 14 and 42. Thus $\{0, 14, 42\} \to \{0, 0, 14\}$ and since $\gcd\{0, 0, 14\} = 14$ it follows that $\gcd\{378, 210, 700\} = 14$.

Jacobi considered only three integers as above, but the algorithm can be generalized to $n + 1$ integers $\{x_0, x_1, \ldots, x_n\}$, and it was in this form that Perron considered it. As in the case of $n + 1 = 3$ integers, with $n + 1$ integers the general step in the algorithm is

$$\{x_0^{(\nu)}, \ldots, x_n^{(\nu)}\} \to \{x_0^{(\nu+1)}, \ldots, x_n^{(\nu+1)}\}$$

where $x_0^{(\nu+1)}, \ldots, x_{n-1}^{(\nu+1)}$ are the remainders after division of each of $x_1^{(\nu)}, \ldots, x_n^{(\nu)}$ by $x_0^{(\nu)}$ and $x_n^{(\nu+1)} = x_0^{(\nu)}$. These relations can be expressed in the form

$$x_1^{(\nu)} = a_1^{(\nu)} x_0^{(\nu)} + x_0^{(\nu+1)}, \ldots, x_n^{(\nu)} = a_n^{(\nu)} x_0^{(\nu)} + x_{n-1}^{(\nu+1)}, \; x_0^{(\nu)} = x_n^{(\nu+1)},$$

where the $a_i^{(\nu)}$ are nonnegative integers. Thus the first equation means that $x_0^{(\nu)}$ goes into $x_1^{(\nu)}$ a total of $a_1^{(\nu)}$ times with remainder $x_0^{(\nu+1)}$, and so on; and the last equation means that the last term in the $(\nu + 1)$st $(n + 1)$-tuple is the divisor $x_0^{(\nu)}$, which is the initial term of the νth $n + 1$-tuple.

[8] I have followed Perron's exposition of the algorithm [36, p. 2f].

By analogy with (3.2), set $\sigma_i^{(\nu)} = x_i^{(\nu)}/x_0^{(\nu)}, i = 1, \dots, n$. Then the above equations for the $x_i^{(\nu)}$ can be rewritten in a form analogous to (3.3) above:

$$\begin{aligned} \sigma_1^{(\nu)} &= a_1^{(\nu)} + 1/\sigma_n^{(\nu+1)}, \\ \sigma_2^{(\nu)} &= a_2^{(\nu)} + \sigma_1^{(\nu+1)}/\sigma_n^{(\nu+1)}, \\ \dots \ &\dots \ \dots\dots\dots\dots \\ \sigma_n^{(\nu)} &= a_n^{(\nu)} + \sigma_{n-1}^{(\nu+1)}/\sigma_n^{(\nu+1)}. \end{aligned} \tag{3.7}$$

Using the above equations it is possible, for any fixed $\nu > 0$, to express the initial values $\sigma_i = \sigma_i^{(0)}, i = 1, \dots, n$, as a rational function of $\sigma_1^{(\nu)}, \dots, \sigma_n^{(\nu)}$. Consider, for example, expressing σ_1 as a function of $\sigma_1^{(2)}, \dots, \sigma_n^{(2)}$. From the first equation above with $\nu = 0$ we have $\sigma_1 = \sigma_1^{(0)} = a_1^{(0)} + 1/\sigma_n^{(1)}$ and from the last equation above with $\nu = 1$ we have $\sigma_n^{(1)} = a_n^{(1)} + \sigma_{n-1}^{(2)}/\sigma_n^{(2)}$. Combining these two equations we get

$$\sigma_1 = a_1^{(0)} + \frac{1}{a_n^{(1)} + \frac{\sigma_{n-1}^{(2)}}{\sigma_n^{(2)}}} = \frac{a_1^{(0)}\sigma_{n-1}^{(2)} + (a_1^{(0)}a_n^{(1)} + 1)\sigma_n^{(2)}}{\sigma_{n-1}^{(2)} + a_n^{(1)}\sigma_n^{(2)}},$$

which shows that σ_1 is a rational function of the $\sigma_i^{(2)}$. To indicate that σ_i is a rational function of $\sigma_1^{(\nu)}, \dots, \sigma_n^{(\nu)}$ Perron introduced the notation

$$\sigma_i = \frac{A_i^{(\nu)} + A_i^{(\nu+1)}\sigma_1^{(\nu)} + \dots + A_i^{(\nu+n)}\sigma_n^{(\nu)}}{A_0^{(\nu)} + A_0^{(+1)}\sigma_1^{(\nu)} + \dots + A_0^{(\nu+n)}\sigma_n^{(\nu)}}, \tag{3.8}$$

where $i = 1, \dots, n$. The coefficients $A_i^{(\nu)}$ are polynomials in the $a_i^{(\nu)}$. For example from the above calculation of σ_1 in terms of $\sigma_1^{(2)}, \dots, \sigma_n^{(2)}$ we see that, e.g., $A_1^{(2)} = 0$ because there is no constant term in the numerator, whereas $A_1^{(2+n)} = a_1^{(0)}a_n^{(1)} + 1$.

I pointed out above that the finite approximations to the infinite continued fraction of a positive irrational number σ corresponded to the expression (3.5) for σ considered as a rational function of $\sigma^{(\nu+1)}$ and with $\sigma^{(\nu+1)} = 0$. By analogy, the νth partial fraction approximation to the irrational number σ_i would be what one obtains from (3.8) by setting $\sigma_i^{(\nu)} = 0$ for $i = 1, \dots, n$, to obtain

$$S_i^{(\nu)} = A_i^{(\nu)}/A_0^{(\nu)}, \quad i = 1, \dots, n. \tag{3.9}$$

It is thus the convergence as $\nu \to \infty$ of the partial sums $S_i^{(\nu)}$ corresponding to any real or complex coefficients $a_i^{(\nu)}$ that is the object of study of Perron's Habilitationsschrift.

Before Perron only a few mathematicians had studied Jacobi's algorithm and the generalized continued fractions it produces.[9] Jacobi, Bachmann (1873), and Fürstenau (1874) considered only the case $n+1=3$. Of these only Bachmann's work implied anything about convergence and that only in a very special case. In 1897 Franz Meyer considered the algorithm for any $n \geq 2$ [30] but obtained no convergence theorems. Perron, working within the research program of Pringsheim, was, by contrast, primarily interested in the question of convergence; and he brought to bear on the problem the ideas he had developed in connection with Stolz' Theorem (Sect. 2.2). The analogy between the coefficients A_ν, B_ν of (2.7), which corresponds here to the case $n=1$ (Euclid's algorithm), and the coefficients $A_i^{(\nu)}$, $A_0^{(\nu)}$ gave rise at Perron's hands to the following analog of Stolz' 2-term recursion relation (2.8) [36, p. 6]:

$$A_i^{(\nu)} = \delta_{i\nu},\ 0 \leq i, \nu \leq n, \quad A_i^{(\nu+n+1)} = A_i^{(\nu)} + \sum_{j=1}^{n} a_j^{(\nu)} A_i^{(\nu+j)}, \quad \nu \geq 0. \tag{3.10}$$

As we shall see, this $(n+1)$-term recursion relation, which was not introduced by Meyer, formed the basis for Perron's insight that his approach to Stolz' Theorem could be extended to deal with the convergence of Jacobi's algorithm when the $a_i^{(\nu)}$ are periodic.

To sum up, implicit in Perron's paper is the following definition, which forms the real starting point of his own investigations.

Definition 3.1 Let $a_i^{(\nu)}$, $i \geq 1$, $\nu \geq 0$, denote real or complex numbers. Then the general Jacobi continued fraction algorithm defined by the recursive relations (3.10) is said to converge and have limiting values $(\alpha_1, \ldots, \alpha_n)$ if $\alpha_i = \lim_{\nu\to\infty} A_i^{(\nu)}/A_0^{(\nu)}$ exists as a finite real or complex number for all $i = 1, \ldots, n$.

The analog of continued fractions with positive coefficients are Jacobi algorithms with all coefficients $a_i^{(\nu)} > 0$. Perron's results included the following remarkable convergence theorem, which even allows some coefficients to be zero.[10]

Theorem 3.2 *Suppose that the $a_i^{(\nu)}$ are any nonnegative real numbers with the property that a constant C exists such that for all i and ν*

$$0 < \frac{1}{a_n^{(\nu)}} \leq C \quad \text{and} \quad 0 \leq \frac{a_i^{(\nu)}}{a_n^{(\nu)}} \leq C. \tag{3.11}$$

Then $\alpha_i = \lim_{\nu\to\infty} \frac{A_i^{(\nu)}}{A_0^{(\nu)}}$ exists as a finite number for all $i = 1, \ldots, n$. Moreover, $\alpha_i > 0$ unless for every ν at least one of the $n+1$ numbers $A_i^{(\nu)}, \ldots, A_i^{(\nu+n)}$ vanishes.

The condition (3.11) of course requires that the coefficients $a_i^{(\nu)}$ with $i=n$ be positive. As we shall see in Sect. 3.3, the above theorem implies that any periodic generalized

9 For details and references see the discussion of pp. 5 and 17 of [36].

10 See Satz II [36, p.12], which establishes more than I have stated in Theorem 3.2.

continued fraction with all $a_i^{(\nu)} \geq 0$ and $a_n^{(\nu)} > 0$ is guaranteed to converge because (3.11) is always satisfied.

The analog of periodic continued fractions involves Jacobi algorithms with real or complex coefficients $a_i^{(\nu)}$ that possess the following periodicity property. As in the discussion of Stolz' work, I will limit myself, without any real loss of generality, to the case of purely periodic coefficients $a_i^{(\nu)}$. Thus by analogy with Stolz' (2.5), the Jacobi algorithm defined by (3.10) is purely periodic and of period $\mathbf{k}$ when the coefficients $a_i^{(\nu)}$ satisfy

$$a_i^{(m\mathbf{k}+j)} = a_i^{(j)}, \quad 1 \leq i \leq n, \quad 0 \leq j \leq \mathbf{k} - 1, \tag{3.12}$$

for any positive integer m.

Just as Perron had earlier used Stolz' recurrence relation (2.8) to deduce the fundamental relation (2.15) for purely periodic continued fractions, he now used the recurrence relation (3.10) in conjunction with the periodicity property (3.12) to deduce an analog of (2.15). As in the discussion of (2.15), it will he helpful to introduce matrix notation as in (2.17)–(2.18), even though Perron did not. The analog of the 2×2 matrix A of (2.13) is the $n + 1 \times n + 1$ matrix

$$A = \begin{pmatrix} A_0^{(\mathbf{k})} & \cdots & A_0^{(\mathbf{k}+n)} \\ & \ddots & \\ A_n^{(\mathbf{k})} & \cdots & A_n^{(\mathbf{k}+n)} \end{pmatrix}. \tag{3.13}$$

If we introduce the $n + 1 \times 1$ column matrices

$$\mathbf{C}_\nu = \left(A_0^{(\nu)} \cdots A_n^{(\nu)}\right)^t, \quad \nu \geq 0, \tag{3.14}$$

then A is the matrix with columns $\mathbf{C}_\mathbf{k}, \ldots, \mathbf{C}_{\mathbf{k}+n}$. The analog of the fundamental relation (2.18) for ordinary continued fractions is[11]

$$A\mathbf{C}_\nu = \mathbf{C}_{\mathbf{k}+\nu}. \tag{3.15}$$

As we saw, in seeking necessary conditions for the convergence of a periodic continued fraction, Perron had introduced the 2×2 matrix A of (2.13) in the course of expressing Stolz' formula (2.9) for the limiting value of a convergent $\mathbf{k}$-periodic continued fraction in the homogeneous form (2.12). The reasoning surrounding (2.12)–(2.15) then showed that the limiting value is determined by a characteristic vector associated to an absolutely maximal characteristic root ρ_0 of A. For a convergent Jacobi algorithm (periodic or not) homogeneity is achieved by setting $\alpha_i = \lim_{\nu\to\infty} A_i^{(\nu)}/A_0^{(\nu)}$ equal to x_i/x_0, $x_0 \neq 0$. Thus $(x_0, x_1, \ldots, x_n)$ is a set of homogeneous coordinates for

[11] The relation (3.15) (without the matrix symbolism) is the special case of Perron's formula (9) [36, p.7], that arises when Perron's λ is taken to be the period k of a purely periodic algorithm.

$(\alpha_1, \ldots, \alpha_n)$. Now for any k, whether or not it is a period, it is certainly the case that for a convergent algorithm

$$\alpha_i = \lim_{\nu \to \infty} A_i^{(k+\nu)} / A_0^{(k+\nu)}. \tag{3.16}$$

When k is the period $\mathbf{k}$, the relation $A\mathbf{C}_\nu = \mathbf{C}_{\mathbf{k}+\nu}$ transforms (3.16) into

$$\alpha_i = \lim_{\nu \to \infty} [A\mathbf{C}_\nu]_i / [A\mathbf{C}_\nu]_0,$$

and this limiting value is easily seen to equal $[A\mathbf{x}]_i/[A\mathbf{x}]_0$, where $\mathbf{x} = \left(x_0, \ldots, x_n\right)^t$ and $[A\mathbf{x}]_i$ is the ith coordinate of $A\mathbf{x}$.[12] Thus both $\mathbf{x}$ and $A\mathbf{x}$ give a set of homogeneous coordinates for $(\alpha_1, \ldots, \alpha_n)$, and so $A\mathbf{x} = \rho\mathbf{x}$, which is the analog of (2.13).

Perron had thus discovered that for convergent periodic Jacobi algorithms as well, the limiting values of the algorithm are given by a characteristic vector associated to some (yet to be determined) characteristic root ρ of A. For this reason he termed

$$f(\rho) = |A - \rho I_{n+1}| = 0 \tag{3.17}$$

the characteristic equation of the algorithm [36, p. 30]. Since Perron had earlier proved a result [36, p. 6] that shows that $\det A = (-1)^{n\mathbf{k}}$, it followed that 0 is never a characteristic root of the matrix A associated to a periodic algorithm.

3.2 Convergence theorems for periodic coefficients

As in the far simpler case of ordinary periodic continued fractions, the problem was now to seek necessary conditions for convergence that would prove to be sufficient and to identify the characteristic root ρ that would yield the limiting values of the algorithm. Perron began by considering the relatively simple case of Jacobi algorithms with period $\mathbf{k} = 1$ [36, Sect. 11]. In this case the periodicity relations (3.12) imply that $a_i^{(\nu)} = a_i^{(0)}$ for all i, and so the algorithm is completely determined by the n values $a_i = a_i^{(0)}$. The matrix A of the algorithm as given by (3.13) with $\mathbf{k} = 1$ and so is $A = \left(\mathbf{C}_1 \cdots \mathbf{C}_{1+n}\right)$, i.e.,

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & a_1 \\ 0 & 1 & 0 & \cdots & 0 & a_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & a_n \end{pmatrix}, \tag{3.18}$$

[12] For notational simplicity let $A = (a_{ij})$. Then, letting $\nu \to \infty$,

$$[A\mathbf{C}^{(\nu)}]_i/[A\mathbf{C}^{(\nu)}]_0 = \frac{\sum_{\sigma=0}^n a_{i\sigma} A_\sigma^{(\nu)}/A_0^{(\nu)}}{\sum_{\sigma=0}^n a_{0\sigma} A_\sigma^{(\nu)}/A_0^{(\nu)}} \to \frac{\sum_{\sigma=0}^n a_{i\sigma} x_\sigma/x_0}{\sum_{\sigma=0}^n a_{0\sigma} x_\sigma/x_0} = \frac{[A\mathbf{x}]_i}{[A\mathbf{x}]_0}.$$

and the characteristic equation is

$$f(\rho) = \pm[\rho^{n+1} - a_n\rho^n - \cdots - a_1\rho - 1]. \tag{3.19}$$

To obtain necessary conditions for convergence, Perron used the same approach he had employed in his paper on ordinary periodic continued fractions: deduce necessary properties of the roots ρ of the characteristic equation by considering $\lim_{\nu\to\infty}(\rho_1/\rho_0)$ [see (2.15)–(2.16)].

In the present more general situation, let $\rho_0, \ldots, \rho_d$ denote the distinct characteristic roots and let m_i denote the multiplicity of ρ_i, $i = 0, \ldots, d$. By a line of reasoning that was much more algebraically involved and sophisticated than what had been required to deal with ordinary continued fractions, Perron showed that the assumption of convergence in the case $\mathbf{k} = 1$ implies that a root, which he denoted by ρ_0, exists such that for any other root ρ_i

$$L_i = \lim_{\nu\to\infty} \frac{\rho_i^{-m_i+1}}{\rho_0^{-m_0+1}} \frac{\nu(\nu-1)\cdots(\nu-m_i+2)}{\nu(\nu-1)\cdots(\nu-m_0+2)} \left(\frac{\rho_i}{\rho_0}\right)^{\nu}$$

exists as a finite number [36, p. 35]. For this to be possible either $|\rho_i| < |\rho_0|$ or $|\rho_i| = |\rho_0|$ and $m_i < m_0$. By virtue of this result Perron introduced the following definition.

Definition 3.3 (Regular characteristic polynomial) A polynomial $f(\rho)$ is regular when it has a root ρ_0 with the property that every other root either has smaller absolute value or, when the absolute values are equal, a smaller multiplicity. The root ρ_0 will be called the principal root (*Hauptwurzel*).

Thus a necessary condition for the convergence of a Jacobi algorithm of period $\mathbf{k} = 1$ is the regularity of its characteristic polynomial.

Perron was able to prove the sufficiency of his regularity condition to establish the following theorem.

Theorem 3.4 *A purely 1-periodic Jacobi algorithm* (3.10) *defined by any complex numbers* $a_1^{(0)}, \ldots, a_n^{(0)}$ *converges if and only if its characteristic polynomial is regular. In the case of convergence, the homogeneous coordinates* $\mathbf{x} = \begin{pmatrix} x_0 & \cdots & x_n \end{pmatrix}^t$ *of the limiting values* $(\alpha_1, \ldots, \alpha_n)$ *are given by any solution* $\mathbf{x} \neq \mathbf{0}$ *of* $A\mathbf{x} = \rho_0\mathbf{x}$*,* ρ_0 *being the principal root, since this system of equations implies* $x_0 \neq 0$ *and that the ratios* $\alpha_i = x_i/x_0$ *are uniquely determined.*

Perron wrote down the explicit formulas giving $\alpha_i = x_i/x_0$ in terms of ρ_0 and the coefficients $a_i = a_i^{(0)}$ [36, p. 38], formulas that show that $x_0 \neq 0$.[13]

[13] The simple nature of A as given by (3.18) implies that for any characteristic root ρ $A\mathbf{x} = \rho\mathbf{x}$ reduces to $\rho x_0 = x_n$, and $\rho x_i = x_{i-1} + a_i x_n$ for $1 \leq i \leq n$. From these relations and the fact that $\rho \neq 0$—det $A = \pm 1$ for the matrix of any $\mathbf{k}$-periodic algorithm as Perron showed—it follows that $x_0 \neq 0$ and that the relations may be solved for the ratios x_i/x_0 in terms of ρ and $a_1, \ldots, a_n$. Thus any matrix of the form (3.18) will have just one linearly independent characteristic vector for each characteristic root ρ, no matter what its multiplicity.

As an illustration of Theorem 3.4, consider the 1-periodic algorithm defined for $n = 4$ by $\{a_1^{(0)}, \ldots, a_4^{(0)}\} = \{-1, -2, 2, 1\}$. The matrix A associated to this algorithm by (3.18) and its characteristic polynomial are easily seen to be

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \quad f(\rho) = \det(A - \rho I) = -(\rho - 1)^3(\rho + 1)^2.$$

Thus the characteristic polynomial is regular with $\rho_0 = 1$, since $\rho_0 = 1$ has a greater multiplicity than $\rho_1 = -1$. Up to scalar multiples, the sole characteristic vector for $\rho_0 = 1$ is $\mathbf{x} = \begin{pmatrix} 1 & 0 & -2 & 0 & 1 \end{pmatrix}^t$, and so by Theorem 3.4 $\alpha_i = \lim_{\nu\to\infty} A_i^{(\nu)}/A_0^{(0)} = x_i/x_0 = 0/1, -2/1, 0/1, 1/1 = 0, -2, 0, 1$.

As Perron observed, "The investigation of convergence takes a far more difficult form when the algorithm has period $k > 1$" [36, p. 39]. Nonetheless, he did establish sufficient conditions for convergence in the general complex $\mathbf{k}$-periodic case [36, Satz VIIIa, Satz VIIIb, p. 44]. Both theorems posit the regularity of the characteristic polynomial of A as one of the conditions, and both establish that the limiting values of the algorithm are given by a specific characteristic vector for the principal root ρ_0, namely $\mathbf{x} =$ row λ of the matrix of cofactors of $\rho_0 I - A$, where by hypothesis the λth row is not identically zero.

For what is to follow, it will be helpful to recall why the above vector $\mathbf{x}$ is a characteristic vector for ρ_0 and to introduce a concomitant notation that will be used extensively in what is to follow. If M is any square matrix, let Adj M denote the associated *adjunct matrix*, i.e., the matrix with (i, j) entry equal to the (j, i) cofactor of det M. In other words Adj M is the transpose of the matrix of cofactors of A. A fundamental identity of the theory of determinants going back to Lagrange, Gauss, and Cauchy was that $M \operatorname{Adj} M = (\det M)I$. With $M = \rho_0 I - A$ we have $(\rho_0 I - A) \operatorname{Adj} (\rho_0 I - A) = 0$. Since the columns of $(\rho_0 I - A) \operatorname{Adj} (\rho_0 I - A)$ are $(\rho_0 I - A)\mathbf{c}_i$, $i = 1, \ldots, n$, where $\mathbf{c}_i$ denotes the ith column of $\operatorname{Adj} (\rho_0 I - A)$, it follows that whenever $\mathbf{c}_i \neq \mathbf{0}$, it is a characteristic vector of A for ρ_0. This was a well-known elementary fact to mathematicians such as Perron and Frobenius.

3.3 Perron's lemma

For periodic algorithms of any period $\mathbf{k}$, Perron realized that his convergence Theorem 3.2 for algorithms with coefficients $a_i^{(\nu)}$ satisfying

$$a_i^{(\nu)} \geq 0, \quad a_n^{(\nu)} > 0, \quad i = 1, \ldots n-1, \quad \nu = 0, 1, 2, \ldots \tag{3.20}$$

provided another set of sufficient conditions for convergence. That is, when an algorithm satisfies the nonnegativity conditions of (3.20) and has in addition periodic coefficients $a_i^{(\nu)}$ as per (3.12), then the quotients $1/a_n^{(\nu)}$ and $a_i^{(\nu)}/a_n^{(\nu)}$ can assume only a finite number of values, which means that these expressions are bounded with respect to i and ν and so satisfy the convergence conditions (3.11) of his Theorem 3.2. Thus any **k**-periodic algorithm that is nonnegative in the sense of (3.20) necessarily converges, and the main problem for such algorithms becomes the determination of the limiting values $(\alpha_1, \ldots, \alpha_n)$ of the partial sums $(S_1^{(\nu)}, \ldots, S_n^{(\nu)}) = (A_1^{(\nu)}/A_0^{(\nu)}, \ldots, A_n^{(\nu)}/A_0^{(\nu)})$. As we saw following (3.16), Perron knew that in the homogeneous form $(x_0, x_1, \ldots, x_n)$, $\alpha_i = x_i/x_0$, the limiting values define a characteristic vector for some characteristic root ρ' of the matrix A of (3.13). It would have been natural to ask whether the characteristic polynomial is regular so that the relevant characteristic root would most likely be the principal root ρ_0, as it was in the above-described results on 1-periodic algorithms (Theorem 3.4). Perron discovered that, thanks to the nonnegativity conditions (3.20) this was indeed the case. How he arrived at this conclusion needs to be considered, for the reasoning employed eventually led him to Theorem 3.9 (Perron's Theorem).

By virtue of the nonnegativity conditions (3.20) Perron realized that the $(n+1)$-term recursion relation (3.10) implies that for all $\nu \geq 2n$ one has $A_i^{(\nu)} > 0$ for all $i = 0, 1, \ldots, n$, which means $\mathbf{C}_\nu = \left(A_0^{(\nu)}, \ldots, A_n^{(\nu)}\right)^t$ has positive components for all $\nu \geq 2n$. The fundamental relation $A\mathbf{C}_\nu = \mathbf{C}_{\mathbf{k}+\nu}$ of (3.15) then shows that since $A = \left(\mathbf{C}^{\mathbf{k})}, \ldots, \mathbf{C}^{\mathbf{k}+n)}\right)$ we have

$$
\begin{aligned}
A^2 &= A\left(\mathbf{C}^{(\mathbf{k})}, \ldots, \mathbf{C}^{(\mathbf{k}+n)}\right)\\
&= \left(A\mathbf{C}^{(\mathbf{k})}, \ldots, A\mathbf{C}^{(\mathbf{k}+n)}\right)\\
&= \left(\mathbf{C}^{(2\mathbf{k})}, \ldots, \mathbf{C}^{(2\mathbf{k}+n)}\right),
\end{aligned}
$$

and thus in general that

$$
A^\nu = \left(\mathbf{C}^{(\nu\mathbf{k})}, \ldots, \mathbf{C}^{(\nu\mathbf{k}+n)}\right). \tag{3.21}
$$

Hence for $\nu \geq 2n/\mathbf{k}$ we have $A^\nu > 0$.[14] Since the characteristic roots of the matrix A^ν are the νth powers of the characteristic roots of A with the same characteristic vectors, information about the characteristic roots of A can be obtained from information about the characteristic roots of $B = A^\nu$, as we shall see below.

The positivity of $B = A^\nu$ apparently induced Perron to ask whether certain properties of the characteristic roots of B germane to the regularity of its characteristic equation might hold simply because $B > 0$, i.e., independently of the fact that $B = A^\nu$ with A the matrix (3.13) associated to the periodic algorithm. He was able to prove the following *Hilfsatz* [36, p. 47].

Lemma 3.5 (Perron's Lemma) *Let $B = (b_{ij})$ be any $n \times n$ matrix with all coefficients $b_{ij} > 0$. Then B has at least one positive root. The greatest positive root ρ_0 has multiplicity one, and all the cofactors of $\rho_0 I - B$ are positive.*

The positivity of the cofactors of $\rho_0 I - B$ means that Adj $(\rho_0 I - B) > 0$ and so (as noted at the end of the previous section) its columns are positive characteristic vectors for ρ_0.

[14] The notations $M > 0$ and $M \geq 0$ for matrices M with all positive, respectively, all nonnegative, coefficients were introduced by Frobenius in 1909 [14].

Perron's Lemma represents the first part of Perron's Theorem (Theorem 3.9 below), the second part being that $|\rho| < \rho_0$ for all other roots ρ of B. The truth of Perron's Lemma is easy to verify for any 2×2 matrix $B > 0$: the characteristic roots of B are given by the familiar quadratic formula, from which it is easily seen that the roots are real and distinct and that the larger is positive; the positivity of the 1×1 cofactors of $\rho_0 I - B$ also follows readily from the quadratic formula for ρ_0. Such considerations probably suggested to Perron the possibility that Lemma 3.5 might be true for $n \times n$ matrices; but for them the above sort of straightforward verification is not feasible since, among other things, there is no formula for the roots of the characteristic polynomial. Not surprisingly, Perron sought to establish the general validity of the lemma by induction on n. His proof was entirely correct, although not explained well. An exposition of this historic proof is given Appendix 6.1.

I now consider how Perron utilized Lemma 3.5 to show how to determine the limiting values $x_0, x_1, \ldots, x_n$ for a **k**-periodic algorithm satisfying the nonnegativity conditions (3.20). These deliberations on his part are historically important because they revealed to him the possibility of extending Lemma 3.5 to include the assertion that $|\rho| < \rho_0$ holds for all roots $\rho \neq \rho_0$ of B.

The nonnegativity conditions (3.20) imply, as we saw, that the matrix A associated to the algorithm by (3.13) satisfies $A^\nu > 0$ for all $\nu \geq 2n/\mathbf{k}$. As the discussion leading to (3.17) indicates, since the algorithm converges Perron knew that its limiting values $x_i/x_0 = \lim_{\nu\to\infty}(A_i^{(\nu)}/A_0^{(\nu)})$, $i = 1, \ldots, n$, have the property that $\mathbf{x} = \left(x_0, \ldots, x_n\right)^t$ satisfies $A\mathbf{x} = \rho'\mathbf{x}$ for some characteristic root ρ' of A. All that remained was to determine which root, and this is where Lemma 3.5 proved useful. Since $A^\nu > 0$ for all $\nu > 2n/\mathbf{k}$, Lemma 3.5 as applied to $B = A^\nu > 0$ says that B has a largest positive root σ_0 of multiplicity one. Now it was well known that if $\rho_0, \rho_1, \ldots, \rho_n$ are the characteristic roots of A, each root being listed as often as its multiplicity, then $\rho_0^\nu, \rho_1^\nu, \ldots, \rho_n^\nu$ is a similar listing of the characteristic roots of A^ν. Suppose by Lemma 3.5 that $\sigma_0 = \rho_0^\nu$. Then it follows that (1) $\rho_0 > 0$, that (2) ρ_0 is the largest positive root of A, and that (3) ρ_0 is simple.[15] Having established these properties of ρ_0, Perron then used determinant–theoretic relations to show that the positivity of all cofactors of $\sigma_0 I - A^\nu$ (a consequence of Lemma 3.5 applied to $B = A^\nu > 0$) implied the same for the cofactors of $\rho_0 I - A$ [36, pp. 49–50].

Before proceeding to Perron's proof that ρ_0 is in fact the root ρ' whose characteristic vectors $\mathbf{x} = \left(x_0\ x_1, \ldots, x_n\right)^t$ yield the limiting values of the algorithm, it will be helpful to point out a fact realized by Perron, namely that the above reasoning establishes the following corollary to Lemma 3.5:

Corollary 3.6 (Corollary to Perron's Lemma) *If A is any nonnegative matrix such that $A^\nu > 0$ for some integer $\nu > 0$, then A has a at least one positive root. The greatest positive root ρ_0 is simple, and* Adj $(\rho_0 I - A) > 0$.

[15] Perron gave only a proof of (1), from which (2) and (3) follow easily. His proof of (1), however, takes for granted that for every $\nu \geq 2n/\mathbf{k}$ the largest positive root of A^ν is the νth power of the *same* root ρ_0 of A [36, p. 49]. (1) can be proved without such an assumption as follows. Let $\nu \geq 2n/\mathbf{k}$ be fixed. By Lemma 3.5 there is an $\mathbf{x} > \mathbf{0}$ such that $A^\nu\mathbf{x} = \sigma_0\mathbf{x} = \rho_0^\nu\mathbf{x}$. There is also a $\mathbf{y} \neq \mathbf{0}$ such that $A\mathbf{y} = \rho_0\mathbf{y}$, from which $A^\nu\mathbf{y} = \rho_0^\nu\mathbf{y}$ follows. Since $\sigma_0 = \rho_0^\nu$ is simple, $\mathbf{y}$ is just a multiple of $\mathbf{x}$ and so $A\mathbf{x} = \rho_0\mathbf{x}$, which implies that $\rho_0 > 0$ because $A \geq 0$ and $\mathbf{x} > \mathbf{0}$.

To show that $\rho' = \rho_0$ Perron used a relation he had derived in his general study of $\mathbf{k}$-periodic algorithms [36, (38), p. 41]: If the algorithm converges to $\mathbf{x} = \big(x_0\ x_1, \ldots, x_n\big)^t$ so that $A\mathbf{x} = \rho'\mathbf{x}$ for some root ρ' of the characteristic equation, then for any $j = 0, 1, \ldots, n$,

$$\rho' = \lim_{\nu\to\infty} A_i^{((\nu+1)\mathbf{k}+j)}/A_i^{(\nu\mathbf{k}+j)}, \quad i = 0, 1, \ldots, n. \tag{3.22}$$

The ratio in (3.22) is the ratio of the (i, j) entries of $A^{\nu+1}$ and A^{ν}. From (3.22) and the fact that $A_i^{(\nu k+j)} > 0$ for ν sufficiently large, it follows that[16]

$$\rho' = \lim_{\nu\to\infty} \frac{\sum_{i=0}^{n} A_i^{((\nu+1)k+i)}}{\sum_{i=0}^{n} A_i^{(\nu k+i)}}. \tag{3.23}$$

In view of the formula (3.21) for A^{ν}, (3.22) states in more familiar notation that ρ' is the limit of $\operatorname{tr}(A^{\nu+1})/\operatorname{tr}(A^{\nu})$. Since traces are sums of characteristic roots, Perron could write (3.23) as

$$\rho' = \lim_{\nu\to\infty} \frac{\sum_{i=0}^{n} \rho_i^{\nu+1}}{\sum_{i=0}^{n} \rho_i^{\nu}}, \tag{3.24}$$

where the summation is over all roots of A, counted according to multiplicity. The only roots that count in the limit are those with absolute value $M = \max_i |\rho_i|$. Perron denoted the possible roots with absolute value M by: M with multiplicity $r \geq 0$, $-M$ with multiplicity $r_0 \geq 0$ and $Me^{\pm\varphi_i}$, $i = 1, \ldots, m$, each with multiplicity $r_i \geq 0$. Thus a possible root of absolute value M is actual only if its multiplicity is positive. He then expressed the right-hand side of (3.24) in a form that showed that the limit could exist if and only if $r_i = 0$ for $i = 0, 1, \ldots, m$ [36, pp. 50–51], which implied that $r > 0$, i.e., that M is a root and that $\rho' = M$. It then follows immediately that $\rho_0 = M = \rho'$. It also followed, as Perron noted, that the characteristic equation of A is regular in the sense of Definition 3.3, the principal root being the maximal positive root ρ_0. This is because the above reasoning showed that $|\rho_i| < M = \rho_0$ for any root $\rho_i \neq \rho_0$.

As the above outline of Perron's reasoning suggests, the *only* part of it that used the fact that A is the matrix (3.13) associated to a periodic algorithm satisfying the nonnegativity conditions (3.20) ensuring convergence was the reasoning leading to

[16] If we write $A^{\nu} = (a_{ij}^{(\nu)})$ so that $a_{ij}^{(\nu)} = A_i^{(\mathbf{k}\nu+j)}$, then (3.22) implies that $a_{ii}^{(\nu+1)}/a_{ii}^{(\nu)} = \rho' + \varepsilon_i^{(\nu)}$, where $\lim_{\nu\to\infty} \varepsilon_i^{(\nu)} = 0$ for all i. Thus $a_{ii}^{(\nu+1)} = (\rho' + \varepsilon_i^{(\nu)})a_{ii}^{(\nu)}$, and so

$$q_\nu = \frac{\sum_{i=0}^{n} a_{ii}^{(\nu+1)}}{\sum_{i=0}^{n} a_{ii}^{(\nu)}} = \rho' + \frac{\sum_{i=0}^{n} \varepsilon_i^{(\nu)} a_{ii}^{(\nu)}}{\sum_{i=0}^{n} a_{ii}^{(\nu)}}.$$

If $\underline{\varepsilon}^{(\nu)} = \min_i \varepsilon_i^{(\nu)}$ and $\overline{\varepsilon}^{(\nu)} = \max_i \varepsilon_i^{(\nu)}$, then since $a_{ii}^{(\nu)} > 0$ for all sufficiently large ν it follows that $\rho' + \underline{\varepsilon}^{(\nu)} \leq q_\nu \leq \rho' + \overline{\varepsilon}^{(\nu)}$ for all sufficiently large ν.

the existence of a characteristic root ρ' satisfying (3.22). In effect he had proved, as he realized, the following extension of his Corollary 3.6.

Proposition 3.7 *Suppose $A = (a_{ij})$ is any nonnegative matrix that has the following properties: (1) $A^\nu > 0$ for some $\nu > 0$; (2) there is a characteristic root ρ' such that*

$$\lim_{\nu\to\infty} a_{ij}^{(\nu+1)}/a_{ij}^{(\nu)} = \rho' \quad \text{for all } i, j,$$

where $a_{ij}^{(\nu)}$ denotes the (i, j) entry of A^ν. Then there is a root $\rho_0 > 0$ of multiplicity 1 such that $|\rho| < \rho_0$ for all other characteristic roots ρ. Furthermore, the cofactors of $\rho_0 I - A$ are all positive so that any column of Adj $(\rho_0 I - A)$ *provides an $\mathbf{x} > 0$ for which $A\mathbf{x} = \rho_0\mathbf{x}$.*

As we have seen, for the special matrices A corresponding to periodic, nonnegative algorithms properties (1) and (2) are satisfied, and so Perron obtained from the reasoning behind Proposition 3.7 the following result [36, p.51].

Theorem 3.8 (Perron's Satz IX) *A purely periodic algorithm of any period $\mathbf{k}$ that is nonnegative in the sense of* (3.20) *converges. Furthermore, the characteristic equation of the associated matrix A of* (3.13) *has a simple root $\rho_0 > 0$ such that $\rho_0 > |\rho|$ for all other characteristic roots ρ of A. The limiting values $x_0, x_1, \ldots, x_n$ of the algorithm are obtained by solving $A\mathbf{x} = \rho_0\mathbf{x}$ or by computing the positive cofactors of any row of the cofactor matrix of $\rho_0 I - A$.*

It should be noted that the matrix $A = \left(\mathbf{C}^{(\mathbf{k})}, \ldots, \mathbf{C}^{(\mathbf{k}+n)}\right)$ associated to an algorithm of period $\mathbf{k} < n$ will include among its columns [by (3.10)] columns of the identity matrix I_{n+1}. Hence A will contain zero coefficients so that $A \not> 0$. For example, the algorithm with period $\mathbf{k} = 2$ and $n = 3$ defined by $a_i^{(\nu)} = 0$ for $i = 1, 2$ and $a_3^{(0)} = 1$, $a_3^{(1)} = 2$, has associated matrix

$$A = \begin{pmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 \end{pmatrix},$$

which contains many zero coefficients although of course $A \geq 0$. Thus it does not satisfy the conditions of Perron's Lemma 3.5; and yet as the proof of Satz IX showed, because $A^\nu > 0$ for $\nu \geq 2n/\mathbf{k} = 3$, its characteristic roots possess the remarkable properties posited in that theorem.

Perron realized that his above-sketched proof of Satz IX depended very little upon the fact that A is the matrix associated to a nonnegative periodic algorithm. As the above proof-sketch indicates, he had used that fact to conclude that: (1) $A^\nu > 0$ for all sufficiently large ν; (2) there is a characteristic root ρ' of A for which (3.22) holds. Conclusion (2) was a consequence of the fact that the algorithm converged. From this it followed that if $(x_0, x_1, \ldots, x_n)$ are the limiting values of the algorithm expressed in homogeneous form, then $A\mathbf{x} = \rho'\mathbf{x}$ for some characteristic root ρ'; and it is precisely this root ρ' for which (3.22) was established. The rest of the proof did not depend on A being an algorithm matrix and so in effect implied what I stated above as Proposition 3.7. With these considerations in mind, Perron observed after stating Satz IX that [36, pp. 51–52]:

> It would be of major importance to prove the regularity of the characteristic equation in a purely algebraic way *without advance knowledge of the convergence of the algorithm*. Then the convergence could be deduced at once from Satz VIIIa, and the above result would again be obtained. This idea can actually be carried out. That is, the power sums of the characteristic equation . . .
>
> $$\sum \rho^{\nu} = A_0^{(\nu k)} + A_1^{(\nu k)} + \cdots + A_n^{(\nu k)} \; [= \mathrm{tr}\; A^{\nu}]$$
>
> are thus all positive, and from this it can be concluded that no root of greater absolute value than ρ_0 could exist. Then some inequalities about the growth of the power sums with increasing exponents follow from which, in view of the simplicity of the greatest positive root ρ_0, it can also be concluded that no other root with absolute value ρ_0 is at hand. And yet to precisely carry out the above sketched line of thought would require various lengthy supporting considerations and so I believed it preferable to give the above presentation.

Satz VIIIa, to which Perron referred in the above quotation, was one of his theorems giving sufficient conditions for the convergence of a (real or complex) **k**-periodic algorithm [36, p. 44]. One condition was the regularity of the characteristic polynomial (Definition 3.3), which in view of Corollary 3.6 to Perron's Lemma meant that all characteristic roots $\rho \neq \rho_0$ of A satisfy $|\rho| < \rho_0$. The other was a condition on the cofactors of $\rho_0 I - A$ that is satisfied because Corollary 3.6 guarantees that the cofactors of $\rho_0 I - A$ are all positive. Thus *if* the regularity of the characteristic equation of A could be established independently of convergence, i.e., independently of the assumption that the algorithm satisfies the nonnegativity conditions (3.20) that guarantee convergence, convergence would be guaranteed instead by Satz VIIa. In view of Proposition 3.7 such an independent proof could be given for any nonnegative matrix A satisfying the first property—$A^{\nu} > 0$ for some $\nu \geq 1$—provided property (2) could be established for any such A. As the above quotation indicates, Perron believed he could carry out such a proof but that it would be far more complicated than the proof of Satz IX.

3.4 Perron's theorem

Within 6 months of submitting his *Habilitationsschrift* to the *Annalen*, Perron published a paper confirming his claims. The paper was entitled "Towards the theory of matrices" [37], and in it Perron proposed to show how many of the proof ideas he had developed in his *Habilitationsschrift* could be used to give new and simple proofs of known results about matrices and their characteristic equations as well as to establish some new ones. Among the new ones was the theorem alluded to in his *Habilitationsschrift* that any positive matrix has a regular characteristic equation [36, Sect. 5], i.e., what is now usually called Perron's Theorem:

Theorem 3.9 (Perron's Theorem) *Let A be any square matrix such that $A > 0$. Then A has a characteristic root $\rho_0 > 0$ of multiplicity one such that $\rho_0 > |\rho|$ for all other characteristic roots ρ of A. Moreover, all the cofactors of $\rho_0 I - A$ are positive.* (*Hence* $\mathbf{x} > 0$ *exists such that* $A\mathbf{x} = \rho_0\mathbf{x}$.)

Perron's proof was dictated by the reasoning behind Satz IX, which implied Proposition 3.7. The assumption that $A > 0$ in Theorem 3.9 is a special case of property (1) of Proposition 3.7 and so he now proved that property (2) holds for any positive matrix. Theorem 3.9 then follows by the same reasoning as that behind Proposition 3.7, i.e., by the reasoning implicit in his proof of Satz IX, as Perron himself explained so as not to repeat it [37, p. 261].

The proof that property (2) of Proposition 3.7 holds was achieved by means of the following lemma [37, pp. 259–261].

Lemma 3.10 (Perron's Limit Lemma) *Let $A = (a_{ij})$ be $n \times n$ with $a_{ij} > 0$ for all i and j. If A^ν is denoted by $(a_{ij}^{(\nu)})$ then:* (i) $\lim_{\nu\to\infty} a_{ij}^{(\nu)}/a_{nj}^{(\nu)}$ *exists as a finite number that is independent of j. Denote it by x_i/x_n.* (ii) $\rho' = \lim_{\nu\to\infty} a_{ij}^{(\nu+1)}/a_{ij}^{(\nu)}$ *exists as a finite positive number that is independent of i and j.* (iii) *If $\mathbf{x} = \left(x_1, \ldots, x_n\right)^t$ with the x_i as in* (i)*, then $A\mathbf{x} = \rho'\mathbf{x}$. Hence ρ' is a positive characteristic root of A.*

Perron's proof of Lemma 3.10 seems to represent the working out of what he had outlined in his remark to Satz IX quoted above. The proof of part (i) is indicative of his method of proof. For a fixed value of i, let $\beta^{(\nu)}$ and $B^{(\nu)}$ denote, respectively, the minimum and the maximum of the n positive numbers $a_{ij}^{(\nu)}/a_{nj}^{(\nu)}$, $j = 1, \ldots, n$, so that $\beta^{(\nu)} \le B^{(\nu)}$ for all ν. Perron showed that the sequence $\beta^{(\nu)}$ is increasing, and that $B^{(\nu)}$ is decreasing. From "a known theorem" it then followed that $\beta = \lim_{\nu\to\infty} \beta^{(\nu)}$ and $B = \lim_{\nu\to\infty} B^{(\nu)}$ exist as finite numbers with $\beta \le B$ [36, p. 259].[17] Then a fairly complicated "ε–δ" type argument was give to show that $\beta = B$, which implies (i) of Lemma 3.10.

Immediately after stating Theorem 3.9 Perron added two historically consequential comments [36, p. 262]:

> Although this is a purely algebraic theorem, nevertheless I have not succeeded in proving it with the customary tools of algebra. The theorem remains valid, by the way, when the a_{ik} are only partly positive but the rest are zero *provided only that a certain power of the matrix A exists for which none of the entries are zero*.

The second sentence, of course, reflects Perron's realization that the reasoning in his *Habilitationsschrift* implied Proposition 3.7. Combined with Lemma 3.10, that proposition implies the result noted by Perron. In other words, Perron had also established the following corollary to Theorem 3.9:

Corollary 3.11 (Perron's Corollary) *Let $A \ge 0$ be such that $A^\nu > 0$ for some power $\nu \ge 1$. Then the conclusions of Theorem 3.9 still hold.*

[17] The "known theorem" was presumably that increasing (respectively, decreasing) sequences of real numbers that are bounded above (respectively, below) converge.

Although Perron simply mentioned Corollary 3.11 in passing, his remark called attention to the fact that there is a substantial class of nonnegative matrices for which the remarkable conclusions of Perron's Theorem 3.9 remain valid. As we shall see, Frobenius pursued this implication to a definitive conclusion in 1912 and in the process created his remarkable theory of nonnegative matrices. It was Perron's first sentence above, however, that initially drew Frobenius' interest to the theory of positive and nonnegative matrices.

The first sentence reflects Perron's dissatisfaction with his proof of Theorem 3.9. What he seems to have meant was that his proof depended upon the limit considerations of Lemma 3.10 and its proof, and so required more than the "customary tools of algebra," such as the theory of determinants. This is how Frobenius interpreted Perron's remark [13, p. 404], and so he took up the challenge implicit in it: to give a determinant-based proof of Perron's Theorem that avoided limit considerations.[18]

4 Frobenius' theory of nonnegative matrices

Although Frobenius no doubt regularly scanned the pages of *Mathematische Annalen*, which had become the journal of the rival Göttingen school of mathematics of Klein, Hilbert, and Minkowski, it seems unlikely to me that he would have paid any attention to Perron's *Habilitationsschrift* when it appeared in the first issue of 1907, due to its subject matter. But when Perron's 16-page paper "Towards the theory of matrices" [37] appeared in a subsequent issue that year, it is not surprising that Frobenius, an expert on the theory of matrices, read it and responded to the challenge set forth by Perron of providing a proof of his Theorem 3.9 that would avoid his Limit Lemma 3.10. Thus while Perron was writing up a detailed study of the convergence of nonperiodic Jacobi algorithms, which appeared in 1908 [38], Frobenius set himself the task of a proof of Perron's Theorem that avoided his Limit Lemma. He succeeded, and his results were published in 1908 [13].

4.1 Frobenius' papers of 1908 and 1909

Frobenius proved a slightly stronger version of Perron's Theorem 3.9 [13, Sect. 1], which may be stated as follows.

Theorem 4.1 (Frobenius' version of Perron's Theorem) *Let $A > 0$ be $n \times n$. Then the following hold. (I) A has a positive characteristic root and hence a maximal positive root ρ_0. Furthermore, ρ_0 is simple, and* $\mathrm{Adj}\,(\rho I - A) > 0$ *for all $\rho \geq \rho_0$. (II) If ρ' is any other characteristic root of A, then $|\rho'| < \rho_0$.*

[18] It does not seem that Perron was seeking a purely algebraic proof of his theorem in the sense of a proof that was completely free of propositions from analysis. For example, he never expressed a similar dissatisfaction with his proof of his seminal Lemma 3.5 despite the fact that it repeatedly invokes basic theorems from analysis such as the intermediate value theorem for continuous functions. (See Appendix 6.1.) The intermediate value theorem was also invoked in Frobenius' "limit-free" proof of Perron's Theorem. It should also be noted that although Frobenius took up the challenge of a determinant-based proof of Perron's Theorem, as a student of Weierstrass he was not adverse to employing results from complex analysis, notably Laurent expansions, in his proofs of theorems about matrices [24, pp. 29ff., 52ff.].

Part I of Frobenius' version represents a slightly improved version of Perron's Lemma 3.5, the improvement being that Adj $(\rho I - A)$ is not only positive for $\rho = \rho_0$ but also for all $\rho > \rho_0$. His proof of Part I, like Perron's of his Lemma 3.5, was by induction on n, but it was far shorter and simpler. This was due in part to the fact that Frobenius used the induction hypothesis that Adj $(\rho I - A) > 0$ for all $\rho \geq \rho_0$ in his proof.[19] This induction hypothesis enabled Frobenius to prove quickly via cofactor expansions that A has positive roots and so a maximal positive root ρ_0. It also enabled him to give a quick proof, by further, more subtle, cofactor expansions, that Adj $(\rho - A) > 0$ for $\rho \geq \rho_0$. The simplicity of ρ_0 then followed from an identity not used by Perron, namely

$$\varphi'(\rho) = \sum_{\alpha=1}^{n} \varphi_{\alpha\alpha}(\rho), \tag{4.1}$$

where $\varphi_{\alpha\alpha}(\rho)$ denotes the αth *principal minor determinant of* $\rho I - A$, i.e., the $n-1 \times n-1$ minor determinant (and cofactor, since $\alpha + \alpha = 2\alpha$ is even) obtained from $\rho I - A$ by deleting row α and column α. Equation (4.1) showed that $\varphi'(\rho_0) > 0$ (and hence that ρ_0 is simple) because $\varphi_{\alpha\alpha}(\rho_0)$ is the (α, α) entry of the positive matrix Adj $(\rho_0 I - A)$.

It was Perron's proof of Part II of Theorem 4.1 that had required his Limit Lemma 3.10. Frobenius' proof of Part II required less than a full page [13, p. 406] and avoided Perron's Limit Lemma 3.10. Nonetheless it was somewhat contrived. As we shall see, he soon discovered a proof that was even briefer and yet straightforward, a proof that seems to have provided the fillip for a new and remarkably rewarding line of research that ultimately led to his masterful paper of 1912 on nonnegative matrices.

Besides supplying a far simpler proof of the "noteworthy properties" [13, p. 404] of positive matrices that Perron had called to attention, Frobenius also considered briefly what could be said when A is simply assumed to be nonnegative:

> If the matrix A is only assumed to have elements $a_{\alpha\beta} \geq 0$, then by means of the above methods of proof and continuity considerations it is easy to determine the modifications under which the above theorems remain valid. The greatest root ... $[\rho_0]$ ... is real and ... ≥ 0. It can be a multiple root, but only when all the principal determinants ... $[\varphi_{\alpha\alpha}(\rho_0)]$ ... vanish [13, pp. 408–409].

For later reference, I will sum up the above quotation as

Proposition 4.2 *If $A \geq 0$ there is a nonnegative root ρ_0, which is greatest in the sense that $|\rho'| \leq \rho_0$ for all characteristic roots ρ' of A. It is no longer the case that ρ_0 is necessarily simple, but in order for it to be a multiple root it is necessary that $\varphi_{\alpha\alpha}(\rho_0) = 0$ for all α, where $\varphi_{\alpha\alpha}(\rho)$ is defined following* (4.1).

As Frobenius said, Proposition 4.2 was an easy consequence Perron's Theorem, obtained as a limiting case of that theorem. Thus, e.g., in the limit Adj $(\rho_0 I - A) > 0$ for

[19] In proving Lemma 3.5 Perron could have used the weaker induction hypothesis Adj $(\rho_0 I - A) > 0$ but did not, thereby unnecessarily complicating his proof; see Appendix 6.1.

$A > 0$ becomes $\mathrm{Adj}\,(\rho_0 I - A) \geq 0$ for $A \geq 0$. In particular, the (α, α) entry of $\mathrm{Adj}\,(\rho_0 I - A)$, namely $\varphi_{\alpha\alpha}(\rho_0)$, is nonnegative; and since by (4.1) $\varphi'(\rho_0)$ is the sum of all the $\varphi_{\alpha\alpha}(\rho_0)$, it follows that $\varphi'(\rho_0) \geq 0$ with $\varphi'(\rho_0) = 0$ only if all $\varphi_{\alpha\alpha}(\rho_0)$ vanish, thereby giving the above necessary condition for ρ_0 to be a multiple root. Just how much of Perron's Theorem is lost in the limiting case is illustrated by the Jordan–Weierstrass canonical form matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

which has ρ_0 being zero (rather than positive) and with multiplicity three (rather than being simple).

In 1909 Frobenius published a sequel to his paper on positive matrices, which seems to have been inspired by his discovery that certain properties of characteristic roots could be established quickly and simply by means of what would now be called inner product considerations. This discovery had been made already in the 1860s by Clebsch and Christoffel within the context of hermitian symmetric matrices, but Frobenius was apparently unaware of their use of the technique.[20] Having now discovered it, he showed how it could be used to give a very short and simple proof of Part II of Theorem 4.1 [14, pp. 411-412]. He also used the same technique to prove the following proposition, which he regarded as a converse to Perron's Theorem. It seems to be the first sign of interest in a line of investigation that was eventually to lead him to a series of remarkable results about nonnegative matrices.

The proposition in question is the following.

Proposition 4.3 *Let $A > 0$. Then if $\mathbf{y}$ is a nonnegative characteristic vector for some root ρ' of A, it must be that $\rho' = \rho_0$ and hence that $\mathbf{y} > \mathbf{0}$.*

With the use of modern inner (or dot) product notation $\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \cdots + x_n y_n$, Frobenius' proof goes like this [14, p. 410]. First of all since $A > 0$ and $\mathbf{y} \geq \mathbf{0}$, $A\mathbf{y} \neq \mathbf{0}$, and so the equation $A\mathbf{y} = \rho'\mathbf{y}$ implies that $\rho' > 0$. Now let $\mathbf{x}$ be the positive characteristic vector for ρ_0 that exists by virtue of Perron's Theorem 3.9 applied to the tranposed matrix A^t, so that $A^t\mathbf{x} = \rho_0\mathbf{x}$. Then by hypothesis

$$\rho'(\mathbf{x} \cdot \mathbf{y}) = (\mathbf{x} \cdot \rho'\mathbf{y}) = (\mathbf{x} \cdot A\mathbf{y}) = (A^t\mathbf{x} \cdot \mathbf{y}) = (\rho_0\mathbf{x} \cdot \mathbf{y}) = \rho_0(\mathbf{x} \cdot \mathbf{y}).$$

Since $\mathbf{x} > \mathbf{0}$ and $\mathbf{0} \neq \mathbf{y} \geq \mathbf{0}$ means that $\mathbf{x} \cdot \mathbf{y} > 0$, it follows by canceling $\mathbf{x} \cdot \mathbf{y}$ in the above equation that $\rho' = \rho_0$.

Frobenius' Proposition 4.3 shows that for positive matrices the only nonnegative characteristic vectors $\mathbf{y}$ that exist are those associated to the maximal positive root ρ_0; and because ρ_0 is a simple root $\mathbf{y}$ must actually be positive. From this point of view Proposition 4.3 suggests the following more general problem:

[20] On the technique as used by Clebsch and Christoffel and Frobenius' failure to use it earlier, see [24, p. 29].

Problem 4.4 *Given a nonnegative matrix A, determine the characteristic roots of A for which nonnegative characteristic vectors exist.*

Nowadays, in many applications of the theory of nonnegative matrices, the existence of positive or nonnegative characteristic vectors is of great importance; but, as noted in Sect. 1, Frobenius' work predated those applications. It was evidently as a problem of purely mathematical interest that Frobenius eventually decided that Problem 4.4 was worth investigating.

4.2 Frobenius' 1912 paper on nonnegative matrices

According to Frobenius it was Problem 4.4 that led him to his remarkable discoveries about nonnegative matrices. His paper of 1912 presenting these discoveries began as follows [15, p. 546].

> In my works ..., [13,14], ... I developed the properties of positive matrices and extended them with the necessary modifications to nonnegative [matrices]. The latter, however, require a far more in-depth investigation, to which I was led by the problem treated in §11.

The problem of §11 is Problem 4.4 above. Let us now consider how that problem may have led him to his discoveries.

For $A > 0$ Frobenius' Proposition 4.3 solves Problem 4.4 by showing that the maximal root ρ_0 is the only such root. However, if A and A' are both positive matrices with respective maximal roots ρ_0 and ρ_0', then it is easily seen that the nonnegative matrix $B = \begin{pmatrix} A & 0 \\ 0 & A' \end{pmatrix}$ has nonnegative characteristic vectors for both ρ_0 and ρ_0'. That is, if $\mathbf{x}, \mathbf{x}'$ denote positive characteristic vectors for A, ρ_0 and A', ρ_0', respectively, it is easily seen by block multiplication that $\mathbf{y} = \begin{pmatrix} \mathbf{x} \\ \mathbf{0} \end{pmatrix}$ and and $\mathbf{y}' = \begin{pmatrix} \mathbf{0} \\ \mathbf{x}' \end{pmatrix}$ are nonnegative characteristic vectors of B for ρ_0 and ρ_0', respectively. Frobenius, who was an expert on the application of matrix algebra to linear algebraic problems, had utilized the symbolical algebra of block partitioned matrices on many occasions, especially in his work on principal transformations of theta functions [24, Sect. 3].

The above observations about the nonnegative matrix B indicate that, more generally, Frobenius' Problem 4.4 is trivial to solve for any nonnegative matrix in block diagonal form

$$B = \begin{pmatrix} R_{11} & 0 & \cdots & 0 \\ 0 & R_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & R_{kk} \end{pmatrix}, \quad R_{11} > 0, \ldots, R_{kk} > 0. \tag{4.2}$$

If $\rho_0^{(1)}, \ldots \rho_0^{(k)}$ are the maximal roots of $R_{11} \ldots, R_{kk}$, then, as in the case of two diagonal blocks, B has a nonnegative characteristic vector $\mathbf{y}^{(i)}$ for $\rho_0^{(i)}$, $i = 1, \ldots, m$. More generally, let P_σ denote the $n \times n$ permutation matrix obtained from the identity matrix I_n by permuting its rows according to the permutation $\sigma \in S_n$, and consider the similar matrix $A = P_\sigma B P_\sigma^{-1} = P_\sigma B P_\sigma^t$. (Since P_σ is an orthogonal matrix, $P_\sigma^{-1} = P_\sigma^t$.) Then A is nonnegative because the similarity transformation $B \to (P_\sigma B)P_\sigma^t$ involves first permuting the rows of B by σ and then permuting the columns of the resulting matrix, viz $P_\sigma B$, by σ. Furthermore, $\mathbf{z}^{(i)} = P_\sigma \mathbf{y}^{(i)}$ is also nonnegative and is easily seen to be a characteristic vector of A for $\rho_0^{(i)}$. Thus Frobenius' Problem 4.4 is solved for any nonnegative matrix *permutationally similar* (in the sense described above) to a matrix in the block form (4.2).

Frobenius' Problem would be rather trivial if every nonnegative matrix were permutationally similar to a matrix in the block diagonal form (4.2), but this is not the case. For example, $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ is not permutationally similar to a matrix in the form (4.2), since the transposition $\sigma = (12)$ is the sole nontrivial permutation of two objects and $P_\sigma A P_\sigma^t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. On the other hand, as Frobenius realized [15, p. 555], any nonnegative matrix is permutationally similar to a matrix in a lower triangular block form

$$\begin{pmatrix} R_{11} & 0 & 0 & \cdots & 0 \\ R_{21} & R_{22} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ R_{m1} & R_{m2} & R_{m3} & \cdots & R_{mm} \end{pmatrix}, \tag{4.3}$$

where now each diagonal $k_i \times k_i$ block R_{ii} is nonnegative—but not necessarily positive—and, assuming no further reduction is possible, the diagonal blocks R_{ii} have the property that they are not permutationally similar to a matrix in the block form

$$\begin{pmatrix} P & 0 \\ Q & R \end{pmatrix}, \tag{4.4}$$

since if, e.g., $P_\sigma R_{11} P_\sigma^t = \begin{pmatrix} P & 0 \\ Q & R \end{pmatrix}$ then the similarity transformation generated by

$$P_{\sigma^*} = \begin{pmatrix} P_\sigma & 0 & \cdots & 0 \\ 0 & I_{k_2} & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & I_{k_m} \end{pmatrix}$$

would make (4.3) permutationally similar to a more refined lower triangular block form with the block R_{11} replaced by (4.4).

Thus Frobenius' Problem 4.4 requires dealing with an "irreducible" lower triangular block form (4.3) rather than the diagonal block form (4.2). This means it is necessary to know, first of all, to what extent the nonnegative matrices R_{ii} that occur on the diagonal possess the properties of positive matrices set forth in Perron's Theorem 3.9. Frobenius called such nonnegative matrices "indecomposable" (*unzerlegbar*). Nowadays they are said to be *irreducible*, and to avoid confusion I will use the current terminology.[21] Thus a nonnegative matrix A is irreducible if it is not permutationally similar to a matrix of the form (4.4). Nonnegative matrices that are permutationally similar to the block form (4.4) he referred to as "decomposable"; I will use the current term *reducible*.

It should be noted that if a nonnegative matrix A is reducible, so that $A = P_\sigma \begin{pmatrix} P & 0 \\ Q & R \end{pmatrix} P_\sigma^{-1}$, then $A^\nu = P_\sigma \begin{pmatrix} P^\nu & 0 \\ Q^* & R^\nu \end{pmatrix} P_\sigma^{-1}$ can never be positive. Recall that Perron's Corollary 3.11 shows that any nonnegative matrix such that $A^\nu > 0$ for some power ν possess all the properties of a positive matrix posited in Perron's Theorem 3.9. The above considerations show that the class of irreducible matrices includes all nonnegative matrices satisfying Perron's condition $A^\nu > 0$. This fact may have raised the hope in Frobenius' mind that the larger class of irreducible matrices might share some of the remarkable properties of those satisfying Perron's condition; if so, the solution of Frobenius' Problem 4.4 would be greatly advanced. The first task then would be to investigate the extent to which irreducible matrices satisfy the conclusions of Perron's Theorem.

The above characterization of the concept of a reducible matrix—and hence also an irreducible one—is the characterization that Frobenius used in his reasoning and is, as I have suggested, probably the form in which he was led to it by Problem 4.4. It is, however, possible to formulate the concept in a form directly related to the coefficient array of a reducible matrix. That is, an $n \times n$ matrix $A \geq 0$ is reducible if and only if there exist $p > 0$ rows of A and $q = n - p > 0$ *complementary* columns of A such that there are zeros at all the intersections of these rows and columns. This was Frobenius' official definition of reducibility [15, p. 548]. For example, if

$$A = \begin{pmatrix} a_{11} & 0 & a_{13} & a_{14} & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & 0 & a_{33} & a_{34} & 0 \\ a_{41} & 0 & a_{43} & a_{44} & 0 \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix},$$

where the a_{ij} are positive, then rows 1, 3, and 4 and complementary columns 2 and 5 have zeros at their intersections (so $p = 3$ and $q = 2$). To see that A is reducible in the original sense consider the transposition $\sigma = (2, 4)$ of the columns of A; it puts the two columns with the "intersection zeros" at the far right, and σ applied to the rows of the resulting matrix puts the intersection zeros in the upper right-hand corner,

[21] There is, of course, a certain degree of analogy between Frobenius' definition of an indecomposable nonnegative matrix and his definition of an irreducible matrix representation of a finite group; I suspect Frobenius, who was aware of the analogy, choose his terminology to avoid confusion of the two notions.

i.e.,

$$P_\sigma A P_\sigma^t = \begin{pmatrix} a_{11} & a_{14} & a_{13} & 0 & 0 \\ a_{41} & a_{44} & a_{43} & 0 & 0 \\ a_{31} & a_{34} & a_{33} & 0 & 0 \\ a_{21} & a_{24} & a_{23} & a_{22} & a_{25} \\ a_{51} & a_{54} & a_{53} & a_{52} & a_{55} \end{pmatrix}, \quad P_\sigma = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

so that $P_\sigma A P_\sigma^t$ is in the form (4.4) and A is reducible in the first-mentioned sense.

To avoid possible confusion in what is to follow, it should be noted that in order for a matrix $A \geq 0$ to be reducible it must be at least 2×2. Thus, although Frobenius never mentioned it, it follows that every 1×1 $A \geq 0$ is irreducible, including $A = (0)$. The fact that $A = (0)$ is irreducible is relevant to the solution of Frobenius' Problem 4.4 because it means that some of the irreducible blocks R_{ii} in (4.3) can be (0). (See the discussion following (4.10) below.) Many of Frobenius' theorems about irreducible matrices, however, do not hold for $A = (0)$ and so in discussing them $A = (0)$ will be excluded by stiputating that the nonnegative matrices A under consideration do not include the zero matrix.

Nowadays, graph-theoretic notions are used with profit in the theory of nonnegative matrices. For example, for an $n \times n$ matrix $A = (a_{ij}) \geq 0$, the directed graph $G(A)$ of A, is defined as follows: $G(A)$ has vertices $1, \ldots, n$ and a directed edge $i \to j$ exists when $a_{ij} > 0$. Then it turns out that A is irreducible in Frobenius' sense precisely when $G(A)$ is connected in the following sense: either $G(A)$ has one vertex (so A is 1×1) or if $G(A)$ has at least two vertices, then for any two vertices $i \neq j$ there is a directed path from i to j.

Judging by the contents of Frobenius 1912 paper, it seems that in exploring the properties of irreducible matrices he focused upon the question of whether the maximal root ρ_0 of Proposition 4.2 is necessarily simple. He observed that if A is reducible and so permutationally similar to $\begin{pmatrix} P & 0 \\ Q & R \end{pmatrix}$, then

$$0 = \varphi(\rho_0) = \det(\rho_0 I - A) = \det(\rho_0 I - P) \cdot \det(\rho_0 I - R),$$

and so one of the principal minor determinants,[22] $\det(\rho_0 I - P)$ or $\det(\rho_0 I - R)$, must vanish. By means of determinant-theoretic considerations combined with the matrix algebra of block-partitioned matrices, he was able prove the converse: If some principal minor determinant of $\rho_0 I - A$ vanishes, then A must be reducible. In this way he proved that a nonnegative $A \neq 0$ is reducible if and only if some principal minor of $\rho_0 I - A$ vanishes. Stated another way, his result was that a nonnegative $A \neq 0$ is irreducible if and only if none of the principal minors of $\rho_0 I - A$ vanish. This meant in particular that when $A \neq 0$ is irreducible none of the degree $n - 1$ principal minors $\varphi_{\alpha\alpha}(\rho_0)$ vanish, i.e., all are positive, and so by the identity (4.1) he had used in his 1908 paper $\varphi'(\rho_0) = \sum_{\alpha=1}^{n} \varphi_{\alpha\alpha}(\rho_0) > 0$, which means that ρ_0 is simple. The positivity of

[22] A principal minor determinant of $\rho_0 I - A$ of degree $n - k$ is one obtained by deleting the same k rows and columns of $\rho_0 I - A$, e.g., by deleting the first k rows and the first k columns.

the diagonal elements $\varphi_{\alpha\alpha}(\rho_0)$ of Adj $(\rho_0 I - A)$ together with Adj $(\rho_0 I - A) \geq 0$ then implied, by a determinant identity, that Adj $(\rho_0 I - A) > 0$.[23] Finally, the fact that all roots ρ' of any $A \geq 0$ satisfy $|\rho'| \leq \rho_0$ (Proposition 4.2) shows that $\rho_0 = 0$ is only possible for an irreducible A when $A = (0)$. In this way Frobenius obtained his first substantial result on nonnegative matrices:

Theorem 4.5 (Irreducible Matrix Theorem) *If $A \neq 0$ is an irreducible matrix, then ρ_0 is simple and positive and* Adj $(\rho_0 I - A) > 0$. *Hence there is an* $\mathbf{x} > \mathbf{0}$ *such that* $A\mathbf{x} = \rho_0\mathbf{x}$. *All other characteristic roots ρ' satisfy $|\rho'| \leq \rho_0$.*[24]

Frobenius was eventually able to strengthen the above theorem by showing that Adj $(\rho I - A) > 0$ for all $\rho \geq \rho_0$ [15, p. 552], as in his version of Perron's Theorem (Theorem 4.1).

Frobenius' Theorem 4.5 showed that nonzero irreducible matrices possessed almost all the properties of the matrices covered by Perron's Corollary 3.11, namely nonnegative matrices satisfying Perron's condition that $A^\nu > 0$ for some power ν. The sole difference was that the strict inequality $|\rho'| < \rho_0$ of Perron's Theorem 3.9 and Corollary 3.11 is replaced by $|\rho'| \leq \rho_0$. Frobenius introduced the irreducible matrices

$$A = \begin{pmatrix} 0 & a_{12} & 0 & \cdots & 0 \\ 0 & 0 & a_{23} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & a_{n-1n} \\ a_{n1} & 0 & 0 & \cdots & 0 \end{pmatrix}, \tag{4.5}$$

where $b = a_{21}a_{23}, \ldots, a_{n-1n}a_{n1} \neq 0$, to show that $|\rho'| \leq \rho_0$ is best possible [15, p.559].[25] It is easily seen that the characteristic polynomial of A is $\varphi(\rho) = \rho^n - b$, and so the characteristic roots are $\rho_k = \sqrt[n]{|b|} \cdot \epsilon^k$, where $\epsilon = e^{(2\pi i)/n}$ and $k = 0, 1, \ldots, n-1$. Thus $\rho_0 = \sqrt[n]{|b|}$, and all roots satisfy $|\rho_k| = \rho_0$.

These considerations led Frobenius to define an irreducible matrix $A \neq 0$ to be *primitive* if $|\rho'| < \rho_0$ for every characteristic root $\rho' \neq \rho_0$. In other words, those irreducible matrices that possess *all* the properties posited by Perron's Theorem 3.9 are called primitive. The remaining irreducible matrices he termed *imprimitive*. Thus the matrices in (4.5) are imprimitive, whereas any $A \geq 0$ satisfying Perron's condition that $A^\nu > 0$ for some ν is primitive by Perron's Corollary 3.11. The obvious question

[23] What Frobenius used, without any explanation, was the fact that if B is any matrix and Adj $(B) = (\beta_{ij})$, then det $B = 0$ implies $\beta_{ij}\beta_{ji} = \beta_{ii}\beta_{jj}$ for all $i \neq j$. Since when $B = \rho_0 I - A$, β_{ii} and β_{jj} are positive (being principal minors), it follows from $\beta_{ij}\beta_{ji} = \beta_{ii}\beta_{jj}$ that β_{ij} and β_{ji} are not just nonnegative but positive. The identity $\beta_{ij}\beta_{ji} = \beta_{ii}\beta_{jj}$ follows from a very special case of a well-known identity due to Jacobi [2, p. 50]. It also follows readily from the more basic identity $B \cdot$ Adj $(B) = (\det B)I$, which when det $B = 0$ implies rank Adj $(B) \leq 1$, and so all 2×2 minors of Adj (B) must vanish. (Viewed in modern terms, $B \cdot$ Adj $(B) = 0$ means that the range of Adj (B) is contained in the null space of B. When 0 is a simple root of B this means rank Adj $B \leq 1$.)

[24] Theorem 4.5 is not stated by Frobenius as a formal theorem, although it is alluded to in his prefatory remarks. The proof is given on pp. 549–550 of [15].

[25] In terms of the graph-theoretic characterization of irreducibility given above in a footnote to Frobenius' official definition, $G(A)$ is a directed n-cycle and so connected.

is: are there any irreducible matrices that are primitive besides those satisfying Perron's condition? Frobenius showed that the answer is "no" [15, p. 553]:

Theorem 4.6 (Primitive Matrix Theorem) *An irreducible matrix A is primitive if and only if $A^{\nu} > 0$ for some power ν.*

In order to establish Theorem 4.6 it is necessary to prove that a primitive matrix has a power that is positive, since the converse is obvious (as noted at the beginning of this section). It is interesting to see where Frobenius got the idea for his proof. As we saw, in his 1908 paper responding to Perron's call for a more satisfactory proof of Theorem 3.9, Frobenius had done just that by giving a simple determinant-based proof that avoided Perron's Limit Lemma 3.10. But Frobenius did not stop there. As a mathematician he was characteristically thorough and delighted in exploring mathematical relations from every conceivable angle within the framework of his chosen approach to the subject. Thus even though Perron's Limit Lemma was no longer needed to establish Perron's Theorem, Frobenius could not refrain from considering the possibility of a simpler proof. It was based upon the following result [13, p. 408]: if a nonnegative $A \neq 0$ has a simple root ρ_0, which strictly dominates in absolute value all other characteristic roots, then for any (i, j) entry

$$\lim_{k\to\infty} \frac{[A^k]_{ij}}{\rho_0^k} = \frac{[\mathrm{Adj}\,(\rho_0 I - A)]_{ij}}{\varphi'(\rho_0)}, \quad \varphi(\rho) = \det(\rho I - A).$$

By virtue of this identity from his 1908 paper, Frobenius saw how to prove the Primitive Matrix Theorem. That is, the identity applies when A is primitive because by definition ρ_0 has the requisite dominance property. Also, by the Irreducible Matrix Theorem 4.5, $\mathrm{Adj}\,(\rho_0 I - A) > 0$, which implies both numerator and denominator in the above limit are positive, since by (4.1) $\varphi'(\rho_0)$ is the sum of the diagonal terms $\varphi_{\alpha\alpha}(\rho_0)$ of $\mathrm{Adj}\,(\rho_0 I - A) > 0$. Because the limit is positive, it follows that for all sufficiently large values of k the expressions $[A_{ij}^k]/\rho_0^k$ will be positive for all (i, j). Since $\rho_0 > 0$, it follows that $A^k > 0$ for all sufficiently large k. This then establishes the Primitive Matrix Theorem 4.6.

Frobenius also obtained a simple sufficient condition for primitivity as a byproduct of his investigation of the properties of imprimitive matrices (see Appendix 6.2 for the simple proof):

Theorem 4.7 (Trace Theorem) *If A is irreducible and* $\mathrm{tr}\, A > 0$*, then A is primitive. Hence all imprimitive A have* $\mathrm{tr}\, A = 0$ *and so all diagonal entries must be zero.*

Frobenius' Primitive Matrix Theorem showed that Perron's condition $A^{\nu} > 0$ exactly characterized those irreducible matrices satisfying all the properties posited in Perron's Theorem, i.e., the matrices A Frobenius called primitive. The imprimitive matrices could now be seen as the class of nonnegative irreducible matrices that lay outside the province of Perron's investigations. Frobenius' exploration of their properties yielded his most profound results on nonnegative matrices.

Frobenius' discoveries regarding imprimitive matrices are summarized in Theorem 4.8 below. A broad outline of the main ideas by means of which he established the theorem are given in Appendix 6.2. In stating it I will use the notation $A \sim_\sigma B$ to mean that A is permutationally similar to B by means of the permutation matrix P_σ, so that $P_\sigma A P_\sigma^{-1} = P_\sigma A P_\sigma^t = B$.

Theorem 4.8 (Imprimitive Matrix Theorem) *Let $A \neq 0$ be an $n \times n$ imprimitive matrix, and let k denote the number of characteristic roots of A with absolute value equal to ρ_0. Then:* (1)

$$A^k \sim_\sigma \begin{pmatrix} R_{11} & 0 & \cdots & 0 \\ 0 & R_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & R_{kk} \end{pmatrix}, \tag{4.6}$$

where each square block R_{ii} is primitive; (2) *If the characteristic polynomial of A is expressed in the notation*

$$\varphi(\rho) = \det(\rho I - A) = \rho^n + a_1\rho^{n1} + a_2\rho^{n_2} + \cdots, \quad a_i \neq 0, \tag{4.7}$$

then k is the greatest common divisor of the differences $n - n_1, n_1 - n_2, \ldots$; (3) *If ρ' is any characteristic root of A, then so is $\epsilon\rho'$, where $\epsilon = e^{2\pi i/k}$;* (4) *The k roots with absolute value ρ_0 are given by $\rho_i = \epsilon^i \rho_0$, $i = 0, \ldots, k-1$, and they are all simple;* (5) *If the characteristic polynomial of A is expressed in the notation*

$$\varphi(\rho) = \rho^n + b_1\rho^{n-k} + b_2\rho^{n-2k} + \cdots + b_m\rho^{n-mk}, \tag{4.8}$$

where $b_m \neq 0$ but $b_i = 0$ for some $i < m$ is possible, and if

$$\psi(\rho) = \rho^m + b_1\rho^{m-1} + b_2\rho^{m-2} + \cdots + b_m, \tag{4.9}$$

then $\psi(\rho)$ has a simple positive root that is larger than the absolute value of any other root.

A few comments about this remarkable theorem are in order. First of all, the integer k, which figures so prominently in the theorem, is nowadays usually called the *index of imprimitivity of A*. The definition of k makes sense for $k = 1$ as well and simply defines a primitive matrix. Part (2) gives an easy way to determine k if the characteristic polynomial of A is known. Stated geometrically, part (3) says that the set of characteristic roots of A is invariant under rotations by $2\pi/k$ radians; and (4) says that the k roots of absolute value ρ_0 form a regular k-gon inscribed in the circle $|z| = \rho_0$, with one vertex at $z = \rho_0$. Although Frobenius certainly recognized these simple geometrical consequences of his results, he did not mention them. What fascinated him was the more algebraic part (5), which "shows most palpably the minor modification by means of which the properties of positive matrices are transferred to imprimitive ones, while at the same time they remain entirely unchanged in their validity for primitive ones" [15, p. 558]. Let me explain. In (5) m is the integer such that $n - mk = \ell$, where $\ell \geq 0$ is the lowest power of ρ actually occurring in $\varphi(\rho)$. Thus $\rho^\ell\psi(\rho^k) = \varphi(\rho)$. From this and (4) it follows readily that the roots of ψ are kth powers of roots of φ and that, in particular, ρ_0^k is a positive root of ψ that is greater than the absolute values of all other roots. Of course, when $k = 1$, i.e., when A is primitive, $\rho^\ell \psi(\rho) = \varphi(\rho)$ and so both ψ and φ share the same special properties with respect to ρ_0 that are true for the characteristic polynomials of positive matrices as per Perron's Theorem.

We have now seen how the problem Frobenius posed to himself—that of determining the characteristic roots of a nonnegative matrix A that possess nonnegative characteristic vectors (Problem 4.4)—may have led him, via the lower triangular block forms (4.3), into his penetrating study of irreducible matrices, the distinction between primitive and imprimitive matrices being motivated by Perron's Theorem 3.9 and its Corollary 3.11. Most of Frobenius' paper [15] is concerned with the theory of irreducible matrices. Having worked out that theory, he then turned, in the penultimate section of his paper [15, Sect. 11], to Problem 4.4. I will conclude the discussion of Frobenius' paper by indicating his solution.

Recall that Problem 4.4 seems to have been motivated by Proposition 4.3 from his 1909 paper: When $A > 0$, the maximal positive root ρ_0 is the sole characteristic root of A with a nonnegative characteristic vector. This solves Problem 4.4 for positive matrices. Once Frobenius had established the Irreducible Matrix Theorem (Theorem 4.5), the same inner product argument used to prove Proposition 4.3, yields an analogous solution to Problem 4.4 for irreducible A: If $A \neq 0$ is irreducible, then the only characteristic root possessing a nonnegative characteristic vector is ρ_0 [15, pp. 554–555].

Suppose now that A is reducible. Then, as already indicated in (4.3), permutations exist so that A is permutationally similar to a matrix of the form

$$\begin{pmatrix} R_{11} & 0 & 0 & \cdots & 0 \\ R_{21} & R_{22} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ R_{m1} & R_{m2} & R_{m3} & \cdots & R_{mm} \end{pmatrix}, \tag{4.10}$$

where the diagonal blocks R_{jj} are irreducible. Nowadays in the theory of nonnegative matrices (4.10) is called a Frobenius normal form for A. Although the irreducible diagonal blocks R_{jj} in (4.10) are uniquely determined by A up to permutational similarity, their ordering on the diagonal depends in general on the chosen normal form (4.10). For example, if A is permutationally similar to the normal form T_1 it is also permutationally similar to the normal form T_2, where

$$T_1 = \begin{pmatrix} R_{11} & 0 & 0 \\ R_{21} & R_{22} & 0 \\ R_{31} & 0 & R_{33} \end{pmatrix} \quad \text{and} \quad T_2 = \begin{pmatrix} R_{11} & 0 & 0 \\ R_{31} & R_{33} & 0 \\ R_{21} & 0 & R_{22} \end{pmatrix},$$

since the block-transposition $\tau = (2, 3)$ applied to the rows and columns of T_1 results in T_2.

With this in mind, suppose that (4.10) is some normal form for A, and let $\rho_0^{(j)}$ denote the maximal root of the irreducible block R_{jj}, $j = 1, \ldots, m$, in the ordering associated to the chosen normal form (4.10). Using the above-mentioned solution to

Problem 4.4 for irreducible matrices, Frobenius showed via block multiplication that if ρ' is a characteristic root of a reducible A, then ρ' can have a nonnegative characteristic vector only if ρ' is one of above the maximal roots $\rho_0^{(j)}$, $j = 1, \ldots, m$. Now assume that ρ' is one of the $\rho_0^{(j)}$ and consider when it possesses a nonnegative characteristic vector. Since it is possible that $\rho' = \rho_0^{(j)}$ for several values of j, let λ denote the *largest* of all indices j, in the ordering associated to the given normal form, for which $\rho_0^{(j)} = \rho'$. Frobenius showed that if $\rho' = \rho_0^{(\lambda)}$ is strictly greater than the maximal roots of all blocks further down the diagonal, i.e., if

$$\rho_0^{(\lambda)} > \rho_0^{(\lambda+i)} \quad \text{for all } i = 1, \ldots, m - \lambda, \tag{4.11}$$

then $\rho' = \rho_0^{(\lambda)}$ has a nonnegative characteristic vector. His solution to Problem 4.4 then followed by establishing the converse, so as to prove

Theorem 4.9 *If $A \neq 0$ is nonnegative, then a root ρ' of A has a nonnegative characteristic vector if and only if A has a normal form* (4.10) *such that $\rho' = \rho_0^{(\lambda)}$ (with index λ as defined above) for which* (4.11) *holds.*

As an illustration of this theorem consider the following matrix [54, p. 168]:

$$A = \begin{pmatrix} 5 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 1 & 1 & 1 & 4 & 0 \\ 0 & 0 & 1 & 0 & 6 \end{pmatrix}.$$

This is in a normal form (4.10) with the five diagonal entries representing the five irreducible blocks R_{jj} of A. Thus $\rho' = 5, 0, 5, 4, 6$ are all possible candidates for having a nonnegative characteristic vector, although the normal form defining A guarantees this, by virtue of (4.11), only in the case of $\rho' = 6$. Whether other roots have nonnegative characteristic vectors depends on whether other normal forms for A exist with a different ordering of the diagonal blocks so that condition (4.11) applies to characteristic roots $\rho' \neq 6$. It turns out that

$$A \sim_\tau \begin{pmatrix} 5 & 0 & 0 & 0 & 0 \\ 1 & 6 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 4 \end{pmatrix}, \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 1 & 2 & 4 \end{pmatrix}. \tag{4.12}$$

It is clear from (4.11) applied to the normal form in (4.12) that not only $\rho' = 6$ but also $\rho' = 5$ and $\rho' = 4$ have nonnegative characteristic vectors. Whether or not there is a nonnegative characteristic vector for $\rho' = 0$ or another linearly independent one for the double root $\rho' = 5$ depends upon what further normal forms (4.10) are permutationally similar to A.

The above example illustrates that Theorem 4.9, as a solution to Problem 4.4, is not entirely satisfying because it depends upon knowing all possible normal forms for A. Frobenius' proof of Theorem 4.9 actually involved ideas that later proved to be the key to a definitive solution [54, pp. 162, 168]. His own proof methods, however, lacked the graph-theoretic viewpoint that brings the underlying ideas to fruition as in [54, pp. 163–169]. For example, by means of graph theoretic notions based on (4.11) it follows that $\rho' = 4, 6, 5$ each have one independent nonnegative characteristic vector, whereas $\rho' = 0$ has none. The complete solution to Problem 4.4 for the A of Theorem 4.9 can be read off from the information implicit in a graph associated to A (the reduced graph of A). Graph theory, however, was in its infancy in 1912, and Frobenius was not impressed by what could be achieved by applying the theory to linear algebra [53, p. 143].

Frobenius' theory of nonnegative matrices as presented in his extraordinary paper of 1912 [15] was his last major contribution to mathematics. Five years later, at the age of 67, he died of a heart condition that had plagued him for many years.

5 Markov chains 1908–1936

We saw in Sect. 3 that Jacobi's generalization of the Euclidean algorithm had led Perron, in his further generalization of it, to introduce a nonnegative matrix associated to any such algorithm that is periodic. Furthermore the existence of a characteristic root ρ_0, which possesses certain dominance properties relative to the other characteristic roots was of interest to his primary concern: the convergence of a periodic algorithm and the calculation of its limiting values. It is rather remarkable that at roughly the same time as Perron's work considerations derived from an entirely difference source, namely the theory of probability, led A.A. Markov (1856–1922) to a type of probabilistic model to which is associated a (stochastic) nonnegative matrix and that, furthermore, the existence of a characteristic root (namely $\rho_0 = 1$) with dominance properties relative to the other characteristic roots was critical in order to carry out his primary objective, namely the analytical calculation of the associated probabilistic functions so as to show that certain laws of large numbers that had been established by Chebyshev for independent sequences apply as well to many cases of dependent sequences.

Markov's paper was presented to the Academy of Sciences in St. Petersburg on 5 December 1907 and published in its proceedings in 1908 [28]. A German translation was appended to the German edition of his lectures on the theory of probability in 1912 [29]—the same year that Frobenius published his remarkable results on nonnegative matrices. There is no evidence that Frobenius was aware of Markov's paper. Indeed, as we have seen in Sect. 4, Frobenius' theory of nonnegative matrices was inspired by the results of Perron, and, as we shall now see, by creating his theory, Frobenius unwittingly resolved all the linear algebraic problems Markov had posed for stochastic matrices—but did not completely resolve. Markov's work is nonetheless historically relevant to the present essay for two reasons: (1) Within the more restricted context of the nonnegative matrices Markov considered, i.e., stochastic matrices, he seems to have anticipated a key concept of Frobenius' theory, namely the concept of an irreducible matrix, as well as a few of Frobenius' results; and so we need to consider to what extent his work did anticipate Frobenius. (2) As we shall see in Sect. 5.2, it was not until the 1930s that an interest in Markov chains became widespread and it was then by means of Frobenius' theory that it was developed rigorously and in complete generality for nonnegative, rather than just positive, stochastic matrices. In this way, the theory of Markov chains became one of the earliest developed applications of the Perron–Frobenius theory and seems to have served to call general attention among mathematicians and mathematically inclined scientists to the existence and utility of the theory.

5.1 Markov's paper of 1908

In his paper Markov considered a sequence of numbers

$$x_1, x_2, \ldots, x_k, x_{k+1}, \ldots \tag{5.1}$$

Initially he assumed that each x_k can have three values $\alpha = -1$, $\beta = 0$, and $\gamma = +1$, and then he generalized to the case in which each x_k can take any finite number n of distinct values $\alpha, \beta, \ldots, \mu, \ldots \nu, \ldots$. In the general case he introduced the probability $p_{\mu\nu}$ that (for any k) if $x_k = \mu$, then $x_{k+1} = \nu$. Thus for any μ we must have

$$p_{\mu,\alpha} + p_{\mu,\beta} + \cdots = 1 \tag{5.2}$$

We see that with these assumptions (5.1) defines what is now called an n-state Markov chain with transition probability matrix $P = (p_{\mu\nu})$, and in fact Markov himself occasionally referred to (5.1) as a chain.[26]

Markov wished to calculate the probability distribution of the sum $x_1 + x_2 + \cdots + x_n$ for increasingly large values of n and in this connection the properties of the characteristic roots of P or, equivalently, its transpose $A = P^t$, were critical to performing these calculations. Nowadays A is usually termed a stochastic matrix, i.e., a nonnegative matrix with columns adding to 1, but, especially in older literature, a nonnegative matrix with row sums adding to 1 (e.g., P) is called stochastic [16, v. 2, p. 83]. Markov vacillated between the systems P and $A = P^t$ in his paper. To avoid confusion I will refer to *column stochastic* and *row stochastic* matrices. Incidentally, Markov made no use of matrix notation in his paper, just as Perron had used none in his *Habilitationsschrift* the preceding year. He expressed all his reasoning in terms of determinants and systems of linear equations, written out without any abbreviated matrix notation.

Markov began by treating in considerable detail the case of three numerical states. Then he turned to case of any finite number n of numerical states $\alpha, \beta, \gamma, \ldots$. The

[26] See e.g., pp. 569, 571 and 576 of the English translation by Petelin. In what is to follow, all page references will he to this translation (cited in the bibliographic reference for Markov's 1908 paper [28]).

mathematical analysis was essentially the same in the more general case, but that analysis (for n states) depended upon linear algebraic assumptions that became more difficult to establish in the n-state case. These assumptions involved the characteristic roots of P or, equivalently, those of $A = P^t$. It is to easy to realize, as Markov did, that $y = 1$ is a characteristic root. For example, if $\mathbf{x} = \begin{pmatrix}1 & 1 & \cdots & 1\end{pmatrix}^t$, then $P\mathbf{x} = \mathbf{x}$ follows from the row-stochastic nature of P. It is also easy to see that because P is row-stochastic, every root $y \neq 1$ of $\varphi(y)$ satisfies $|y| \leq 1$.[27] The attendant mathematical analysis, however, required assuming that (1) $y = 1$ be a simple root and that (2) $|y| < 1$ for all roots $y \neq 1$.

Markov set forth the following conditions to guarantee the validity of (1) and (2) [28, p. 571]:

> Before proceeding to further conclusions, it is necessary to note that [I] we are investigating only chains $x_1, x_2, \ldots, x_n, \ldots$, where the appearance of some of the numbers $\alpha, \beta, \gamma, \ldots$ does not preclude the appearance of others. This important condition could be expressed by means of determinants in the following form: [ID] The determinant
>
> $$\begin{vmatrix} u, & p_{\beta,\alpha}, & p_{\gamma,\alpha}, & \cdots \\ p_{\alpha,\beta}, & v, & p_{\gamma,\beta}, & \cdots \\ p_{\alpha,\gamma}, & p_{\beta,\gamma}, & w & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{vmatrix}$$
>
> with arbitrary elements
>
> $$u, v, w, \ldots$$
>
> does not reduce to a product of several determinants of the same type.
> This condition is not enough, however, for our purposes and thus [IID] we must assume that the determinant we have chosen does not reduce to the product of several determinants with
>
> $$u = p_{\alpha,\alpha}, \quad v = p_{\beta,\beta}, \quad w = p_{\gamma,\gamma}, \ldots.$$

The bracketed roman numerals [I], [ID], and [IID] have been added in order to facilitate discussing the conditions Markov assumed to hold for the matrix P. The "D" in ID and IID is to indicate that these conditions are formulated in terms of determinants. In what follows I will refer to them as Markov's Conditions I, ID, and IID, respectively.

Although lacking in mathematical precision, Markov's Condition I is usually interpreted, and rightly so, as an equivalent form of Frobenius' notion of irreducibility [53, pp. 146–147]. The strongest support for this interpretation comes from Markov's proof

[27] If $\mathbf{x}$ is a characteristic vector for y and $m = \max_i |x_i|$, let i_0 be such that $|x_{i_0}| = m$. Then the i_0th equation of $yI = P\mathbf{x}$ is $yx_{i_0} = \sum_{j=1}^n p_{i_0 j} x_j$. Taking absolute values and using the triangle inequality implies $|y|m \leq \sum_{j=1}^n p_{i_0 j}|x_j| \leq \left(\sum_{j=1}^n p_{i_0 j}\right)m = 1 \cdot m$, whence $|y| \leq 1$. As we shall see in Appendix 6.3, Markov sought variations on this line of reasoning that would prove $|y| < 1$ for all roots $y \neq 1$ for P satisfying certain conditions.

of assumption (2) above, namely that $|y| < 1$ for all roots $y \neq 1$. As I indicate in Appendix 6.3.2, in dealing with the first of two cases, Markov appears to have tacitly assumed in his proof that P satisfies the following condition:

Condition I*. *If C, D is any partition of $\{1, \dots, n\}$ into nonempty sets, then there is an $\mu \in C$ and a $\nu \in D$ for which $p_{\mu\nu} \neq 0$.*

This condition is easily seen to be equivalent to P being irreducible in Frobenius' sense. For if Condition I* fails to hold, then a partition C, D exists such that $p_{ij} = 0$ for all $i \in C$ and all $j \in D$. This means P is reducible in Frobenius' sense: there are zeros at the intersections of the $p = |C|$ rows of P and the $q = |D|$ complementary columns. The failure of Condition I* also means that Markov's Condition I fails to hold. To see this, assume without loss of generality that the n state values $\alpha, \beta, \dots$ are $1, \dots, n$. Then if, e.g., x_1 has a value $i \in C$ then so must x_2 since $p_{ij} = 0$ for all $(i, j) \in C \times D$. Thus x_k will never take a value in D for any k, contrary to Markov's Condition I.

It is thus fairly certain that Markov's probabilistically motivated Condition I was interpreted by him to mean Condition I*, which is equivalent to assuming the irreducibility of P. Further support for identifying Markov's Condition I with irreducibility in the guise of Condition I* comes from his Condition ID, which is easily seen to imply irreducibility. That is, if P is reducible, so that a permutation $\sigma \in S_n$ exists for which $P_\sigma P P_\sigma^{-1} = \begin{pmatrix} L & 0 \\ M & N \end{pmatrix}$, then if $P(u_1, \dots, u_n)$ denotes P with its diagonal elements $p_{11}, \dots, p_{nn}$ replaced by variables $u_1, \dots, u_n$, we have with $\tau = \sigma^{-1}$

$$P_\sigma P(u_1, \dots, u_n) P_\sigma^{-1} = \begin{pmatrix} L(u_{\tau(1)}, \dots, u_{\tau(k)}) & 0 \\ M & N(u_{\tau(k+1)}, \dots, u_{\tau(n)}) \end{pmatrix},$$

where L is $k \times k$. Thus

$$\det[P(u_1, \dots, u_n)] = \det[L(u_{\tau(1)}, \dots, u_{\tau(k)})] \cdot \det[N(u_{\tau(k+1)}, \dots, u_{\tau(n)})]$$

is "a product of several determinants of the same type," contrary to Markov's Condition ID.[28]

Evidently Markov had anticipated Frobenius' key notion of an irreducible matrix, albeit restricted to the context of stochastic matrices. Furthermore, in 1911, and thus also before Frobenius, Maurice Potron, a mathematical economist familiar with the work of Perron and Frobenius published during 1907–1909, introduced the equivalent of the notions of reducible and irreducible nonnegative matrices.[29] Here we have yet another example of an instance of multiple discovery involving Frobenius. Other instances include the theory of finite group characters and representations (T. Molien,

[28] Markov's Condition ID is given in terms of $A(u_1, \dots, u_n)$, $A = P^t$, but of course $A(u_1, \dots, u_n)$ and $P(u_1, \dots, u_n)$ have the same determinant.

[29] Potron spoke of "partially reduced" matrices [40, p. 1130] by which he meant the equivalent of reducible matrices. His strongest results about solutions $\mathbf{x} \geq \mathbf{0}$, $\mathbf{y} \geq \mathbf{0}$ to $(sI - A)\mathbf{x} = B\mathbf{y}$ with A, B nonnegative and $s \geq \rho_0$ were for A that are not partially reduced. For bibliographic information on Poltron, see footnote 6 in Sect. 1.

W. Burnside, H. Maschke [20]), matrix algebra (Cayley, Laguerre [21]), the bilinear covariant or derivative of a 1-form (Darboux [23]), and the normal form for a matrix over $\mathbb{Z}$ (H.J.S. Smith [22, Sects. 5–6]). In all the above-mentioned instances Frobenius went further in developing the theory, and in most cases with far greater rigor, than any of his fellow discoverers. This is true in particular regarding Markov and the theory of irreducible matrices, as we shall see.

Let us now consider Markov's Condition IID in the above quotation. The substitutions $u = p_{\alpha,\alpha}, v = p_{\beta,\beta}, w = p_{\gamma,\gamma}, \cdots$ would seem to bring us back again to P and its determinant. Markov seems to be asserting that if the nonzero coefficients p_{ij} of P are regarded as variables, then $\det P$ does not factor into a product of several determinants. As we shall see below with the Example (5.4), this condition (so interpreted) is not implied by Condition I. I can find no place in Markov's paper where Condition IID is used directly, i.e., in its determinant formulation. However, as I suggest in Appendix 6.3.2, in proving assumption (2) ($|y| < 1$ for $y \neq 1$) in the second case of his proof, Markov seems to have tacitly assumed that P satisfies the following condition.

Condition II*. For any partition of $\{1, \ldots, n\}$ into nonempty sets E, F there is no corresponding partition of $\{1, \ldots, n\}$ into nonempty sets G, H such that $p_{ij} = 0$ for all $(i, j) \in (G \times F) \cup (H \times E)$.

It is not difficult to see that Condition IID implies Condition II* or, equivalently, that if Condition II* fails to hold then the same is true of Condition IID, so that $\det P$, with the nonzero entries of P regarded as variables, factors into two determinants of like nature.

To see this, suppose Condition II* fails to hold so that partitions E, F and G, H exist and $p_{ij} = 0$ for all $(i, j) \in (G \times F) \cup (H \times E)$. Let the number of integers in E, F, G, H be denoted, respectively, by e, f, g, h. Then there are permutations $\sigma \in S_n$ that map G onto $\{1, \ldots, g\}$ and H onto $\{g + 1, \ldots, n\}$. Likewise there are permutations τ that map E onto $\{1, \ldots, e\}$ and F onto $\{e + 1, \ldots, n\}$. If P_σ and P_τ denote the corresponding permutation matrices, then $P_\sigma P P_\tau^t$ has the block form

$$P_\sigma P P_\tau^t = \begin{pmatrix} L_{g\times e} & 0_{g\times f} \\ 0_{h\times e} & M_{h\times f} \end{pmatrix} \stackrel{\text{def}}{=} Q, \tag{5.3}$$

where the subscripts indicate the dimensions of the respective blocks and $0_{g\times f}$ and $0_{h\times e}$ denote blocks of zeros.[30]

Now since (according to my interpretation) in Condition IID Markov was thinking of the nonzero entries of P as variables it seems likely he assumed that $\det P$ is not identically zero. Indeed, in his detailed analysis of the three-state case he expressly ruled out the case in which $\det P = 0$ [28, p. 555–556]. Since permutation matrices have determinant ± 1, $\det P = \pm \det Q$. Thus $\det P \neq 0$ means $\det Q \neq 0$, and then rank considerations show that for this to be the case it is necessary that $e = g$ and

30 Note that in general $\sigma \neq \tau$ and so P and $P_\sigma P P_\tau^t$ need not be permutationally similar in the sense introduced in discussing Frobenius' 1912 paper.

$f = h$.[31] Thus the diagonal blocks in (5.3) are square, and (5.3) yields $\det P = \pm \det Q = \pm(\det L_{e\times e})(\det M_{f\times f})$. Moreover, for $n > 2$ there is enough leeway in the choice of σ or τ to make $\det(P_\sigma P_\tau) = +1$ so that $\det P = \det Q = (\det L_{e\times e})(\det M_{f\times f})$. In other words, $\det P$ as a function of the nonzero p_{ij} (regarded as variables) factors into the product of two similar determinants of orders e and f. This is precisely what happens when Markov's Condition IID, as interpreted by me, fails to hold. This means that Condition IID implies Condition II*. In sum: Markov's proof that all roots $y \neq 1$ of P satisfy $|y| < 1$ is valid for all P satisfying Conditions I*–II*, and his Conditions I (or ID) and IID imply, respectively, Conditions I* and II*.

Markov's proof of assumption (1) ($y = 1$ is a simple root) is also discussed in Appendix 6.3.1. As I explain there, the proof is valid when $P > 0$—the hypothesis under which Markov initially carried out his analysis—but in the more general case $P \geq 0$ his very sketchy proof does not seem viable. Frobenius' Theorem 4.5 of 1912, of course, implies that any probability transition matrix P that satisfies Markov's Condition I* (irreducibility) has $y = 1$ as a simple root. And Frobenius' Theorem 4.6 implies any P satisfying Markov's Condition I* will also satisfy his assumption (2) ($|y| < 1$) if and only if Perron's condition $P^\nu > 0$ is satisfied. Markov's own proof of assumption (2) depending, as it seems, on Conditions I*–II*, does not yield as general a result. For example, if

$$P = \begin{pmatrix} 0 & p_{12} & p_{13} & 0 \\ p_{21} & 0 & 0 & p_{24} \\ 0 & p_{32} & p_{33} & 0 \\ p_{41} & 0 & 0 & p_{44} \end{pmatrix} \tag{5.4}$$

denotes any row-stochastic matrix with all entries of the form p_{ij} positive, then it is irreducible, i.e., it satisfies Markov's Condition I*.[32] However, it does not satisfy Condition II* by virtue of the partitions $E = \{2, 3\}$, $F = \{1, 4\}$ and $G = \{1, 3\}$, $H = \{2, 4\}$; all coefficients of P with indices (i, j) in $G \times F$ or in $H \times E$ are zero. As we have seen, this means P also fails to satisfy Condition IID.[33] Indeed, based on the considerations leading to (5.3) we may take for the row permutation $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}$ and for the column permutation $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}$ to obtain $\det P_\sigma = \det P_\tau = 1$ and

$$P_\sigma P P_\tau = \begin{pmatrix} p_{32} & p_{33} & 0 & 0 \\ p_{12} & p_{13} & 0 & 0 \\ 0 & 0 & p_{21} & p_{24} \\ 0 & 0 & p_{41} & p_{44} \end{pmatrix} \text{ so } \det P = \begin{vmatrix} p_{32} & p_{33} \\ p_{12} & p_{13} \end{vmatrix} \cdot \begin{vmatrix} p_{21} & p_{24} \\ p_{41} & p_{44} \end{vmatrix}.$$

[31] Consider the first g rows of the matrix in (5.3). They all have the same f coordinates equal to zero and so can be thought of as g vectors in $\mathbb{R}^e$. Thus to be linearly independent it must be that $g \leq e$ or, equivalently $h = n - g \geq n - e = f$. Now consider the last h rows of the matrix in (5.3). They can likewise be thought of as h vectors in $\mathbb{R}^f$. Thus for these rows to be linearly independent it is necessary that $h \leq f$. Combining the two inequalities gives $h = f$ and therefore $g = e$.

[32] The directed graph $G(P)$ contains the 4-cycle $1 \to 3 \to 2 \to 4 \to 1$ and so is connected.

[33] Hence this example shows that Condition I does not imply Condition IID.

Nonetheless P is primitive by virtue of Frobenius' Trace Theorem 4.7, since tr $P = p_{33} + p_{44} > 0$. Thus $|y| < 1$ for all roots $y \neq 1$ even though P fails to satisfy Condition IID. It would seem Markov realized his Condition IID was overly strong for in a footnote to its statement he remarked that "Our conclusions may be extended to many cases we have excluded" [28, p. 571n].[34]

Markov had originally obtained his results for row-stochastic matrices P assuming $P > 0$ and then sought to extend them to $P \geq 0$ [28, p. 574n]. His proof that $|y| < 1$ for all characteristic roots $y \neq 1$ certainly applies when $P > 0$, since Conditions I* and II* are satisfied. And his proof that 1 is a simple root then follows (as he realized) from a lemma due to Minkowski (Appendix 6.3.1). Thus within the more limited context of positive row-stochastic matrices he had independently discovered some of Perron's results about positive matrices. However, as Schneider has pointed out [53, p. 147], there is no mention of Perron's result that Adj $(I - P) > 0$ and its consequence that $y = 1$ has a positive characteristic vector for $A = P^t$—a consequence that has since become of considerable importance to the theory of Markov chains. Likewise, Perron's Corollary 3.11 implied that row-stochastic $P \geq 0$ possessing a positive power have characteristic roots with the properties Markov wished to establish, but Markov gave no indication in his paper that he realized this. And of course, also missing is the deeper insight, implied by Frobenius' Primitive Matrix Theorem 4.6, that a row-stochastic $P \geq 0$ satisfying Markov's Condition I* (irreducibility) possesses characteristic roots with the properties Markov needed if and only if it has a positive power.

Although within the context of stochastic matrices $P \geq 0$ Markov seems to have anticipated Frobenius' notion of irreducibility, it was obscured by his emphasis upon the determinant-based Condition ID in lieu of a precise and explicit mathematical formulation of irreducibility. Although his main result that irreducible row-stochastic matrices satisfying Condition II* are primitive was correct, it failed to characterize primitive row-stochastic matrices. Also his proof of the simplicity of the root $y = 1$ for P satisfying Condition I* was not at all rigorous when $P \not> 0$ due to its dependence upon an unproven generalized version of Minkowski's lemma. By contrast, Frobenius' 1912 paper [15] was based on careful definitions, and by means of clear and rigorous proofs he obtained definitive results on irreducible matrices that went far beyond anything found in the paper of Markov, who, after all, was primarily interested in the probabilistic aspects of his chains, which mainly involved him with analytical derivations.

That same probabilistic focus naturally limited Markov's attention to the more amenable class of stochastic matrices, whereas the work of Perron and Frobenius revealed, in retrospect, that the theorems discovered by Markov were more generally true and were but a part of the rich theory of nonnegative matrices. Even the fact that any nonnegative matrix A possesses a root $\rho_0 \geq 0$ with the property that $|\rho| \leq \rho_0$ for all other roots ρ (the limiting case of Perron's Theorem) was an unexpected result that came out of Perron's penetrating study of Jacobi's algorithm, whereas the same result is trivial when A is stochastic (as noted above). And, of course, from Perron's penetrating

[34] Cf. the discussion of the possible meaning of Markov's Condition IID in [53, pp. 146–147].

study came his even more surprising result that when $A^\nu > 0$ the above inequalities become strict and ρ_0 is simple and positive. As we have seen, it was these remarkable discoveries by Perron that engaged Frobenius' interest in the theory of nonnegative matrices and ultimately led to his masterly paper of 1912. I will now briefly consider how the theory developed in Frobenius' paper was applied to give a clear and rigorous treatment of Markov's theory of chains for the case $P \not> 0$.

5.2 The Perron–Frobenius theorem and Markov chains

Frobenius had concluded his paper of 1912 with one application, which was to the theory of determinants [15, Sect. 14]. If $X = (x_{ij})$ is a matrix of n^2 independent variables x_{ij} then it was well-known that det X is an irreducible polynomial in these n^2 variables. From his theory of nonnegative matrices, he now deduced that if $\widetilde{X}$ is the matrix obtained from X by setting some $x_{ij} = 0$, then if $\widetilde{X}$ is irreducible as a nonnegative matrix in the obvious sense, the polynomial det $\widetilde{X}$ is still irreducible. Thus although Frobenius had written a definitive work on irreducible nonnegative matrices, Markov's theory being unfamiliar, the sole known application was to the theory of determinants. Frobenius' paper thus represented a definitive study of a type of matrix that was not at the time seen to be relevant to many applications or related to the main topics of the linear algebra of the time. For example, the new generation of texts on the theory of matrices that appeared in the early 1930s by Schreier and Sperner [55], Turnbull and Aitkin [59], and Wedderburn [63] make no mention of the Perron–Frobenius theory, being devoted to the main topics in linear algebra, such as canonical matrix forms, properties of symmetric, orthogonal, hermitian and unitary matrices and their applications to quadratic and bilinear forms.

Even though Markov's paper was translated into German and appended to the 1912 German translation of his book on the theory of probability [29], it is uncertain how widely read it was. Apparently, those who did discuss Markov chains in the period 1912–1930 limited their attention to the case $P > 0$ [53, p. 147], perhaps because, as we have seen, when $P > 0$ Markov's proofs are correct and comprehensible. In the late 1920s there was a renewed interest in Markov chains on the part of a large number of mathematicians, who became more or less simultaneously interested in the subject. Some of them, including J. Hadamard and M. Fréchet apparently reinvented aspects of the theory without knowing of Markov's pioneering work [19, p. 2083, 2083n.3]. In the early 1930s, in the midst of the revival of interest in Markov chains, two applied mathematicians, R. von Mises and V.I. Romanovsky, independently applied Frobenius' theory of nonnegative matrices in order to deal with chains corresponding to nonnegative P that need not be positive.

5.2.1 R. von Mises

In 1920 Richard von Mises (1883—1953), became the first director of the newly formed Institute for Applied Mathematics at the University of Berlin.[35] The arrival

[35] See in this connection [6, pp. 148–153].

of von Mises in fact coincided with a period of renewed vitality and ascendency for mathematics at Berlin, and von Mises, with his dynamic personality, was a key player in this revival. In 1921 he became the founder and editor of a journal devoted to applied mathematics, *Zeitschrift für angewandte Mathematik und Mechanik*. In the first issue he wrote an introductory essay [61] in which he made the point that the line between pure and applied mathematics is constantly shifting with time as mathematical theories find applications [61, p. 456]. Such an area of pure mathematics was constituted by the Perron–Frobenius theory of nonnegative matrices. As we saw, it was the pure mathematics of ordinary and generalized continued fractions that motivated Perron's work, which Frobenius further developed solely by virtue of its interesting algebraic content. Von Mises sought to apply this theory to a problem at the foundations of statistical mechanics.

This occurred in his 1931 book *The Calculus of Probabilities and its Application to Statistics and Theoretical Physics* [62], which formed part of his lectures on applied mathematics. The application to theoretical physics, which constituted the fourth and final section of his book, had to do with the statistical mechanics of gases that had been developed by Maxwell and Boltzmann in the 19th century, with alternative statistical models arising in the 20th century from the work of Planck, Bose, Einstein and Fermi. All of these physical theories shared a common assumption. Stated in the neutral language of the theory of probability, the assumption was the following. Suppose there are k "states" $S_1, \ldots, S_k$ that a certain "object" can be in. Let p_i denote the probability that the object is in state S_i. Then the assumption is that all states are equally likely, i.e., that $p_i = 1/k$ for $i = 1, \ldots, k$. In Boltzmann's theory, the states represented small cells of equal volume in the 3-dimensional momentum space of an ideal gas molecule (the object) [62, p. 432]. In Planck's quantum theory the states represented k energy levels $0, h\nu, 2h\nu, \ldots, (k-1)h\nu$ that the ideal gas molecule (the object) may have, where h denotes Planck's constant [62, pp. 439–440]. In the Bose–Einstein–Fermi theory [62, pp.446–449] the states are the occupancy numbers $0, 1, 2, \ldots, k-1$ for a cell of volume h^3 and fixed energy in the six-dimensional phase space of an ideal gas molecule. The object in this case is such a cell.

Such a priori assignations of probabilities were anathema to von Mises' approach to probability theory, according to which probabilities were relative frequencies obtained from a repeated experiment, where the experiment could be an empirical one or an Einsteinian "thought experiment." Von Mises believed that he could describe a thought experiment that would provide a sound probabilistic basis for the above assumptions as follows [62, p.532ff.]. Imagine k urns $U_1, \ldots, U_k$. Each urn contains k lots, which are numbered from 1 to k. From an arbitrarily chosen urn U_{x_0} a lot is drawn. Let x_1 denote the number of the drawn lot. Proceed to urn U_{x_1} and draw a lot. Let x_2 denote the number drawn. Then proceed to urn U_{x_2} and draw a lot, and so on. Then a sequence $x_0, x_1, x_2, \ldots$ is generated, where each x_i is an integer between 1 and k. This is, of course, an example of what is now called a k-state Markov chain, and von Mises was aware that the mathematics of his thought experiment was "closely connected" to "the problem of Markov chains" [62, p. 562]. He saw in this model a way to justify the a priori "equal probability" assumption underlying the various statistical gas models.

I will use symbolical matrix and vector notation in describing von Mises' work, even though he himself used none. Thus let the components of $\mathbf{v}^{(0)} = \left(p_1^{(0)} \cdots p_k^{(0)}\right)^t$ denote the initial probabilities of being in states $S_1, \ldots, S_k$, respectively. And let $P = (p_{ij})$ denote the matrix of transition probabilities and $A = P^t$ its transpose. Von Mises' own notation was chosen so that the coefficients a_{ij} of A defined his transition probabilities, i.e., he defined a_{ij} as the probability of moving from state j to state i [62, p. 533]. Thus A alone is considered by von Mises, and A is column-stochastic. The central question for him was the question of when a limiting probability distribution $\mathbf{v}_\infty = \lim_{n\to\infty} A^n \mathbf{v}_0$ exists that is independent of the initial probabilities.

Von Mises was familiar with Frobenius' three papers on positive and nonnegative matrices, about which he may have learned from Frobenius' former star student Issai Schur, who was also a professor at Berlin. Thus in a footnote [62, p. 536n] von Mises wrote:

> A large part of the propositions that will be derived here and in Sects. 4 and 5 are closely related to the algebraic theory of matrices with nonnegative elements that was developed in three works ... [13–15] ... by *G. Frobenius*. A part of the results of course follow only from the special property of our matrices that the column sums have the value 1.

von Mises utilized Frobenius' notions of reducible and irreducible matrices, as well as the related notion of complete reducibility[36] in his work [62, pp. 534–536]; and he was clearly guided by Frobenius' results, especially those in his paper of 1912. However, von Mises couched everything in probabilistic terms and notation and presented his own proofs rather than appealing to or reproducing Frobenius' own more general proofs.

To facilitate easy comparison and comprehension, in describing von Mises' application of Frobenius' theory, I will use the more familiar notation of Frobenius. I will also limit my discussion to the case of irreducible stochastic matrices A, even though von Mises presented a number of more general results as well. In terms of application to the above-described assumption common to all the statistical theories of gases, von Mises' principal theorem was the following [62, p.548]:

Theorem 5.1 *If A is* (a) *irreducible,* (b) *has* $\operatorname{tr} A > 0$*, and* (c) *is symmetric, then for any* $\mathbf{v}_0 \neq 0$,

$$\mathbf{v}_\infty = \lim_{n\to\infty} A^n \mathbf{v}_0 = \left(\tfrac{1}{k} \cdots \tfrac{1}{k}\right)^t . \tag{5.5}$$

Although von Mises gave his own proof, his Theorem 5.1 is an easy consequence of Frobenius' theorems. For example, the assumptions that A is irreducible with $\operatorname{tr} A > 0$ means that A is primitive by Frobenius' Trace Theorem 4.7. From this it follows readily

[36] Complete reducibility is defined below in Appendix 6.2, Theorem 6.5.

that $\mathbf{v}_\infty = \lim_{n\to\infty} A^n \mathbf{v}_0$ exists and is a positive probability vector for $\rho_0 = 1$. The symmetry of A then implies that $\mathbf{v}_\infty = \left(\frac{1}{k} \cdots \frac{1}{k}\right)^t$.[37]

Referring to the considerations culminating in Theorem 5.1, von Mises, in keeping with his frequentist approach to probability theory, declared that

> Our ... deductions are not based on an assumption about probabilities of fixed individual states and also not on the ergodic hypothesis,[38] but rather exclusively on the assumptions a) to c), which concern the transition probabilities and of which only the last is quantitatively decisive. *It is not the assumption that certain states are equally likely, which is hardly physically meaningful, but rather* [*the assumption*] *that between these states symmetrical ... transition probabilities exist, that forms the proper foundation for the kinetic theory of gas and similar physical-statistical theories*. [62, p. 555]

Although von Mises' work on the probabilistic foundations of statistical mechanics was not explicitly about Markov chains, it was known to those working in this area.[39] Incidentally, one of von Mises' students, Lothar Collatz, applied some of Frobenius' results to a problem in numerical analysis [8], thereby suggesting a vast new area for application that proved quite fertile, as can be seen from Varga's 1962 book, *matrix iterative analysis* [60].

5.2.2 V.I. Romanovsky

V.I. Romanovsky (1879–1954) was born in Vernyi (now Alma-Ata) in Kazakestan and by 1918 had returned to nearby Tashkent in Uzbekistan as professor of probability and mathematical statistics. During 1900–1908, he had been a student and then docent at the University of Saint Petersburg. In 1904 he completed his doctoral dissertation under the direction of Markov at the University, where Markov had been a professor since 1886.[40]

In 1929 Romanovsky published a paper (in French) in the proceedings of the Academy of Sciences of the U.S.S.R. "On Markoff chains" [46]. After giving the basic definitions, he explained that "We call the series of such trials discrete Markoff chains because this eminent geometer was the first to consider them. Here we will expound some new results concerning the general case, which was not considered by Markoff" [46, p. 203]. By the "general case" he meant the generic case in which all the

---

[37] Symmetry means that $A = A^t = P$ and so $\mathbf{v}_\infty$ is a characteristic probability vector for P, which means it is a multiple of $\mathbf{e} = \left(1 \ \cdots \ 1\right)^t$, since $P\mathbf{e} = \mathbf{e}$ is an immediate consequence of the row-stochastic nature of P.

[38] For the statement of this hypothesis see [62, pp. 521–522]. Von Mises joined the ranks of those who criticized invoking it in conjunction with Boltzmann's theory and devoted many pages to critiquing it [62, pp. 526–532].

[39] See, e.g., the paper by Hadamard and Fréchet [19, p.2083], where von Mises' work is called to attention and praised. Hadamard and Fréchet also state (on p. 2083) that von Mises (among others mentioned) did his work without knowledge of Markov's paper [28]. Although the basis for this statement is uncertain, it seems to be based on their belief that Markov's work was only available in Russian, whereas, as noted earlier, a German translation was available since 1912 in the German edition of Markov's book [29].

[40] For further information on Romanovski, see [9].

characteristic roots of $A = P^t$ are distinct, where (as in the above discussion of Markov's work) $P = (p_{ij})$ denotes the matrix of transition probabilities of an n-state chain. Of course one of these roots is $\rho_0 = 1$. In addition to assuming no multiple roots, Romanovsky also assumed that $\rho = -1$ was not a root. For $k = 0, 1, 2, \ldots$ he considered the probabilities $q_i^{(k)}$ of being in the ith state after k iterations of the process. Although he did not use any matrix notation—working with systems of linear equations as had Markov–Romanovsky realized the equivalent of $\mathbf{v}^{(k+1)} = A\mathbf{v}^{(k)}$, where $A = P^t$ and for any k $\mathbf{v}^{(k)} = \left(q_1^{(k)} \cdots q_n^{(k)}\right)^t$. He also realized the immediate implication that $\mathbf{v}^{(k)} = A^k\mathbf{v}^{(0)}$. However, he erroneously assumed that since $\rho = -1$ was excluded as a characteristic root, all roots $\rho \neq 1$ satisfy $|\rho| < 1$ so that $\mathbf{v}^{(\infty)} = \lim_{k\to\infty} A^k\mathbf{v}^{(0)}$ exists [46, p. 204].

At this point in time Romanovsky was not familiar with Frobenius' paper of 1912, which makes it clear that it is only for primitive matrices A (with or without multiple roots) that the above reasoning is valid. In particular, Frobenius' example (4.5) in the special case

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{5.6}$$

is a stochastic matrix with index of imprimitivity $k = 3$ satisfying all Romanovsky's explicit assumptions but having the three cube roots of unity as characteristic roots so that $|\rho| = 1$ for all roots ρ and $\mathbf{v}^{(\infty)} = \lim_{k\to\infty} A^k\mathbf{v}^{(0)}$ does not exist. Ignorant of Frobenius' work, Romanovsky repeated his error in two notes in the *Comptes rendus* of the French Academy of Sciences in 1930 [47,48]. A Czech mathematician, J. Kaucký, spotted the error, and in a 1930 note in the *Comptes rendus* [27] he gave as a counter example the matrix in (5.6), albeit without mentioning Frobenius. Kaucký concluded by pointing out that "the classical theory of A.A. Markoff" shows that $\mathbf{v}^{(\infty)} = \lim_{k\to\infty} A^k\mathbf{v}^{(0)}$ exists when $P = (p_{ij}) > 0$. His remark reflects the fact that in the initial phase of the reawakened interest in Markov chains, Markov's "classical" theory was restricted to the case $P > 0$, perhaps because Markov's efforts to extend the theory to some $P \geq 0$ were, as we saw, unclear and partly untenable.

At the session of 19 January 1931 of the Académie des Sciences Romanovsky responded to Kaucký's criticism with a note "On the zeros of stochastic matrices" [49].[41] As the title suggests, here he attempted a more careful examination of the possibilities for the characteristic roots of certain stochastic matrices. Some of his propositions (I and II) are valid for any row-stochastic matrix P and not just for those satisfying the additional condition—no zero columns—imposed by him; one of them (Prop. III about a characteristic vector $\mathbf{x}$ for $\rho_0 = 1$) is ambiguously stated and, depending on the interpretation, either contains an unnecessary hypothesis (if $\mathbf{x} \geq \mathbf{0}$

41 Judging by his Remark [49, p. 267], Romanovsky was the first to use the term "stochastic matrix." For him it meant (i) $P \geq 0$ (ii) with row sums equaling 1 and (iii) no zero column. Nowadays condition (iii) is not usually included in the definition of a stochastic matrix, and I have not included this condition in my references to stochastic matrices.

is asserted) or is false (if $\mathbf{x} > \mathbf{0}$ is asserted). The next two (IV–V) are incorrect.[42] Anyone well-versed in the results of Frobenius' 1912 paper would have realized these defects.[43] Romanovsky had clearly not yet studied Frobenius' paper and was probably not yet aware of its existence.

By 1933 Romanovsky had become familiar with Frobenius' paper, for in that year he published a paper in the *Bulletin de la Société mathématique de France* entitled "A theorem on the zeros of nonnegative matrices" [50], which began by noting that the zeros of such matrices "are profoundly studied by G. Frobenius" in [15]. The theorem of the title was a corrected generalization of his faulty Proposition VI in the note of 19 January 1931. Three years later, in 1936, Romanovsky published a lengthy memoir in *Acta Mathematica* entitled "Investigations on Markoff chains" [51], and by that time he had evidently digested all three of Frobenius' papers on positive and nonnegative matrices [13–15]. Citing these three papers, he wrote in his introductory remarks

> Since the theory of stochastic matrices and their zeros plays a fundamental role in the theory of Markoff chains and is intimately connected to the theory of nonnegative matrices developed by G. Frobenius, I will begin my memoir with an exposition of the results of G. Frobenius ....

Romanovsky devoted 33 of the 105 pages of his memoir to Frobenius' theory and its application to stochastic matrices, thereby exposing his readers to all of Frobenius' significant results and making clear their relevance to the theory of stochastic matrices and Markov chains. In 1945 he incorporated his exposition of Frobenius' theory into a book on discrete markov chains (in Russian). Citing Romanovsky's book and several of his earlier papers, Felix Gantmacher devoted a chapter to the Perron–Frobenius theory of nonnegative matrices in his book (in Russian) on the theory of matrices, which appeared in 1953.[44] Gantmacher's book represented the first genuinely comprehensive treatise on matrix theory and has since become a classic. It was translated into German in 1958 and into English in 1959 and is still in print as [16]. An English translation of Romanovsky's book was published in 1970 by E. Seneta [52].

---

[42] In the 16 January 1933 session of the Académie, Émile Ostenc gave simple counter examples to IV–VI [32]. He made no reference to Frobenius' 1912 paper [15].

[43] The most interesting and historically significant part of Romanovsky's paper is the concluding paragraphs where he responded to Kaucký's criticism by attempting to characterize those P which admit all primitive kth roots of unity, $k \geq 3$, as characteristic roots. These paragraphs are of interest because they involved what turns out to be an alternative characterization of the degree of imprimitivity k of an irreducible matrix, a characterization that has a graph-theoretic interpretation (A is cyclical of index k). Romanovsky himself made no reference to the theory of graphs and it is doubtful he was thinking in such terms, since his ideas were motivated by the well-known determinant theoretic formula for the coefficients of the characteristic polynomial $\varphi(r) = |rI - A|$, as is evident from his subsequent, more detailed papers [50, p. 215] and [51, p. 163].

[44] In 1937 Gantmacher and Krein [17] had already used Perron's Lemma 3.5 as proved by Frobenius in 1908 to develop their theory of strictly positive (respectively nonnegative) matrices—$n \times n$ matrices such that all $k \times k$ minors are positive (respectively, nonnegative) for all $k = 1, \ldots, n$. Such matrices arise in the mechanical analysis of small oscillations. See [18] for a comprehensive account.

6 Appendix

6.1 Perron's proof of Lemma 3.5

Perron's Lemma 3.5 as stated and proved by him [36, pp. 47–49] runs as follows.

Lemma 6.1 *If $A > 0$ then A has at least one positive characteristic root. The largest positive root ρ_0 is simple, and the cofactors of $\rho_0 I - A$ are all positive.*

The proof was by induction on the dimension of A, but Perron did not include in his induction hypotheses the positivity of the cofactors, i.e., that $\mathrm{Adj}\,(\rho_0 I - A) > 0$. Perhaps this was because he began the induction with 1×1 matrices, for which $\mathrm{Adj}\,(\rho_0 I - A)$ does not exist. Inclusion of $\mathrm{Adj}\,(\rho_0 I - A) > 0$ among the induction hypotheses would have made his proof a bit shorter and simpler, as noted below.

Let $A = (a_{ij}) > 0$, where $i, j = 0, 1, \ldots, n$. The lemma is obviously true for $n = 0$ (1×1 matrices), and so assume it is true for all $n \times n$ matrices and consider the above $n + 1 \times n + 1$ matrix A. Set $\varphi_{n+1}(\rho) = \det\,(\rho I - A)$. To utilize the inductive hypothesis Perron considered the $n \times n$ matrix B obtained from A by deleting its first row and column. Thus $B = (a_{ij})$ with $i, j = 1, \ldots, n$. Let $\varphi_n(\rho) = \det\,(\rho I - B)$. Then by the induction hypothesis φ_n has a maximal positive root σ_0, which is simple. Note that since $\varphi_n(\rho) = \rho^n + \cdots$, $\lim_{\rho\to+\infty} \varphi_n(\rho) = +\infty$. Thus

$$\varphi_n(\rho) > 0 \quad \text{for all } \rho > \sigma_0. \tag{6.1}$$

In order to relate φ_{n+1} and φ_n Perron presumably turned to the well known Laplace expansions of determinants by cofactors.[45] Applied to the determinant $\varphi_{n+1}(\rho) = \det\,(\rho I - A)$ with $\rho > \sigma_0$ the Laplace expansion by cofactors across row 0 gives

$$\varphi_{n+1}(\rho) = (\rho - a_{00})\varphi_n(\rho) - \sum_{i=1}^{n} a_{0i}[\mathrm{Adj}\,(\rho I - A)]_{i0}, \tag{6.2}$$

since the $(0, i)$ cofactor is the $(i, 0)$ coefficient of $\mathrm{Adj}\,(\rho I - A)$. Now each coefficient $\mathrm{Adj}\,(\rho I - A)_{i0}$ being the $(0, i)$ cofactor of $\det(\rho I - A)$, is $\pm$ the $n \times n$ determinant obtained by deletion of row 0 and column i of $\rho I - A$. If that determinant is in turn given a Laplace expansion down the first column of the undeleted part of $\rho I - A$, the corresponding cofactors are coefficients of $\mathrm{Adj}\,(\rho I - B)$. In this manner one obtains

$$[\mathrm{Adj}\,(\rho I - A)]_{i0} = \sum_{j=1}^{n} a_{j0}[\mathrm{Adj}\,(\rho I - B)]_{ij}, \quad i \geq 1. \tag{6.3}$$

Substituting (6.3) into (6.2), Perron obtained the following relations, here summarized in matrix notation:

$$\varphi_{n+1}(\rho) = (\rho - a_{00})\varphi_n(\rho) - \mathbf{r}\,\mathrm{Adj}\,(\rho I - B)\mathbf{c}, \tag{6.4}$$

[45] Perron simply stated the equivalent of (6.4) without any derivation.

where $\mathbf{r} = \left(a_{01} \cdots a_{0n}\right) > \mathbf{0}$ is a row matrix and $\mathbf{c} = \left(a_{10} \cdots a_{n0}\right)^t > \mathbf{0}$ is a column matrix. They correspond, respectively, to the first row and first column of A with the a_{00} coefficient deleted.

Perron rewrote (6.4) in the form

$$\frac{\varphi_{n+1}(\rho)}{\varphi_n(\rho)} = \rho - a_{00} - \mathbf{r}\mathbf{d}(\rho), \quad \mathbf{d}(\rho) = \frac{\mathrm{Adj}\,(\rho I - B)\mathbf{c}}{\varphi_n(\rho)}. \tag{6.5}$$

His first goal was to prove that $\lim_{\rho \to \sigma_0^+} \mathbf{d}(\rho)_i = +\infty$ for at least one value of i, for then (6.5) and $\mathbf{r} > \mathbf{0}$ implies that $\lim_{\rho \to \sigma_0^+} \varphi_{n+1}(\rho)/\varphi_n(\rho) = -\infty$. Since by (6.1) $\varphi_n(\rho) > 0$ for $\rho > \sigma_0$, it follows that $\varphi_{n+1}(\rho) < 0$ for values of ρ slightly greater than σ_0. But since $\varphi_{n+1}(\rho) = \rho^{n+1} + \cdots$, $\lim_{\rho \to +\infty} \varphi_{n+1}(\rho) = +\infty$. Hence $\varphi_{n+1}(\rho)$ takes both negative and positive values on (σ_0, ∞) and the mean value theorem then implies that $\varphi_{n+1}(\rho) = 0$ for at least one value of $\rho \in (\sigma_0, \infty)$. In other words, A has a positive characteristic root and so a maximal positive root $\rho_0 > \sigma_0$.

To establish the above limiting behavior of $\mathbf{d}(\rho)$ as $\rho \to \sigma_0^+$, Perron proceeded as follows. From the formula for $\mathbf{d}(\rho)$ it follows that its components are rational functions of the form

$$(\mathbf{c}_i \rho^{n-1} + \text{lower powers of } \rho)/\varphi_n(\rho),$$

which means that $\mathbf{d}(\rho)$ has an expansion of the form

$$\mathbf{d}(\rho) = \frac{\mathbf{c}}{\rho} + \text{higher powers of} \frac{1}{\rho} \tag{6.6}$$

valid for all sufficiently large ρ. It should also be noted that, in more modern notation, $\mathbf{d}(\rho) = (\rho I - B)^{-1}\mathbf{c}$, a relation which Perron expressed in terms of the equations defined by

$$\rho \mathbf{d}(\rho) = \mathbf{c} + B\mathbf{d}(\rho). \tag{6.7}$$

It follows immediately from (6.6) that $\mathbf{d}(\rho) > \mathbf{0}$ for all sufficiently large ρ, and from this Perron concluded that

$$\mathbf{d}(\rho) > \mathbf{0} \quad \text{for all} \quad \rho > \sigma_0. \tag{6.8}$$

For if (6.8) failed to hold then (by continuity considerations) there would be a $\rho_1 > \sigma_0$ for which $\mathbf{d}(\rho) \geq \mathbf{0}$ but $\mathbf{d}(\rho) \not> \mathbf{0}$, i.e., $\mathbf{d}(\rho_1)_i = 0$ for at least one i. However, since by (6.7) $\rho_1 \mathbf{d}(\rho_1) = \mathbf{c} + B\mathbf{d}(\rho_1)$, the ith equation would be $0 = \rho_1 \mathbf{d}(\rho_1)_i = \mathbf{c}_i + B\mathbf{d}(\rho_1)$, which is impossible since $\mathbf{c} > \mathbf{0}$, $B > 0$ and $\mathbf{d}(\rho_1) \geq 0$.

Thus (6.8) is proved, and so for $\rho > \sigma_0$ we have $\mathbf{d}(\rho) = \mathrm{Adj}\,(\rho I - B)\mathbf{c}/\varphi_n(\rho) > \mathbf{0}$ and we also have $\varphi_n(\rho) > 0$ by (6.1). Thus $\mathrm{Adj}\,(\rho I - B)\mathbf{c} > \mathbf{0}$ for all $\rho > \sigma_0$. *If* Perron had included $\mathrm{Adj}\,(\rho_0 I - A) > 0$ in his induction hypotheses, he could have used $\mathrm{Adj}\,(\sigma_0 I - B) > 0$ to conclude that $\mathrm{Adj}\,(\sigma_0 I - B)\mathbf{c} > 0$ and thus that $\lim_{\rho \to \sigma_0^+} \mathbf{d}(\rho)_i = +\infty$ for *all* i since $\varphi_n(\sigma_0) = 0$. Instead, he had to give a further argument to show that at least one of the components $\mathbf{d}_i(\rho)$ has a pole at $\rho = \sigma_0$ [36, p. 48].

To show that ρ_0 is a simple root, Perron observed that if

$$\frac{d}{d\rho}\left[\frac{\varphi_{n+1}(\rho)}{\varphi_n(\rho)}\right]_{\rho_0} = \frac{\varphi'_{n+1}(\rho_0)}{\varphi_n(\rho_0)} > 0. \tag{6.9}$$

then $\varphi'_{n+1}(\rho_0) > 0$ since $\varphi_n(\rho_0) > 0$ by (6.1); and so ρ_0 would be simple.

To establish that the derivative in (6.9) is positive, Perron differentiated the fundamental relation (6.5) to obtain

$$\frac{d}{d\rho}\left[\frac{\varphi_{n+1}(\rho)}{\varphi_n(\rho)}\right]_{\rho_0} = 1 - \mathbf{r}\cdot\mathbf{d}'(\rho_0).$$

Since the row vector $\mathbf{r} > \mathbf{0}$, (6.9) will certainly follow if $\mathbf{d}'(\rho_0) < \mathbf{0}$. To confirm this Perron differentiated (6.6) to obtain

$$\mathbf{d}'(\rho) = -\frac{\mathbf{c}}{\rho^2} + \text{ higher powers of } \frac{1}{\rho}.$$

Thus for all sufficiently large values of $\rho > \rho_0$, $\mathbf{d}'(\rho) < \mathbf{0}$. Another continuity argument then shows that $\mathbf{d}'(\rho_0) < \mathbf{0}$,[46] thereby establishing the simplicity of ρ_0.

To show that Adj $(\rho_0 I - A) > 0$, Perron observed that the definition of $\mathbf{d}(\rho)$ in (6.5) implies that Adj $(\rho_0 I - B)\mathbf{c} > \mathbf{0}$ since $\rho_0 > \sigma_0$ implies by (6.8) that $\mathbf{d}(\rho_0) > \mathbf{0}$ and (6.1) implies that $\varphi_n(\rho_0) > 0$. Now the cofactor expansion (6.3) above, expressed in matrix form and with $\rho = \rho_0$, asserts that $[\text{Adj }(\rho_0 I - A)]_{i0} = [\text{Adj }(\rho_0 I - B)\mathbf{c}]_i$ and so is positive for $i = 1, \ldots, n$. Since $[\text{Adj }(\rho_0 I - A)]_{00} = \varphi_n(\rho_0) > 0$ as well, the first column of Adj $(\rho_0 I - A)$ has all positive entries. This conclusion was reached by using the induction hypothesis on the matrix B, whose characteristic equation occurs in the Laplace expansion along row $i = 0$ given in (6.2). If instead the Laplace expansion were done along row $i \neq 0$, the induction hypothesis applied in the same manner to the matrix B' obtained from A by deleting row $i \neq 0$ and column 0, would lead to formulas showing that the ith column of Adj $(\rho_0 I - A)$ has all positive entries.

6.2 An outline of Frobenius' proof of Theorem 4.8

A key to the further investigation of the primitive–imprimitive distinction for Frobenius derived from a line of thought that he had used in the past, starting with his 1878 paper on matrix algebra [12, p. 358ff.]. It provides another illustration, beyond those given in [24], of the manner in which matrix algebra was an agent of mathematical discovery

46 If $\mathbf{d}'(\rho_0) < \mathbf{0}$ failed to hold then there would be a value $\rho_1 \geq \rho_0$ such that $\mathbf{d}'(\rho) < \mathbf{0}$ for all $\rho > \rho_1$ but $\mathbf{d}'(\rho_1) \leq \mathbf{0}$, i.e., $[\mathbf{d}'(\rho_1)]_i = 0$ for some i. Differentiation of (6.7) gives $\rho\mathbf{d}'(\rho) = -\mathbf{d}(\rho) + B\mathbf{d}'(\rho)$. For $\rho = \rho_1$ the ith equation in this system is $0 = [\mathbf{d}'(\rho_1)]_i = -[\mathbf{d}(\rho_1)]_i + [B\mathbf{d}'(\rho_1)]_i$. Since (as shown above) $\mathbf{d}(\rho) > 0$ for all $\rho > \sigma_0$ and so for $\rho = \rho_1$, and since $B > 0$ and $\mathbf{d}'(\rho_1) \leq 0$, the right-hand side of the ith equation cannot add up to zero, contrary to assumption.

for Frobenius. Let $\varphi(\rho) = |\rho I - A|$, and set $\varphi(s,t) = (\varphi(t) - \varphi(s))/(t-s)$. Then by the Cayley–Hamilton Theorem, which Frobenius had independently discovered—and was the first to prove—in 1878 [24, Sect. 2.1], we have $\varphi(A) = 0$ and so for s not a characteristic root $\varphi(s, A) = \varphi(s)(sI - A)^{-1} = \text{Adj}\,(sI - A)$. The expansion of $\varphi(t)$ in powers of $t - s$ shows that

$$\varphi(s,t) = \frac{\varphi(t) - \varphi(s)}{t-s} = \varphi'(s) + \frac{1}{2}\varphi''(s)(t-s) + \cdots + \frac{1}{n!}\varphi^{(n)}(s)(t-s)^{n-1}.$$

By setting $t = A$ in the above equation, Frobenius deduced that

$$\text{Adj}\,(sI - A) = \varphi(s, A) = \varphi'(s)I + \frac{1}{2}\varphi''(s)(A - sI) + \cdots + \frac{1}{n!}\varphi^{(n)}(s)(A - sI)^{n-1}$$

even when s is a characteristic root. This implies the following lemma.

Lemma 6.2 *For any $n \times n$ matrix A and any s,* Adj $(sI - A)$ *is a linear combination of $A^0, A, \ldots, A^{n-1}$.*

From this lemma and the Irreducible Matrix Theorem 4.5 Frobenius readily deduced the following key lemma [15, p. 551].

Lemma 6.3 *If $A \neq 0$ is irreducible, then for any fixed pair of indices (i, j) the n coefficients $[A^m]_{ij}$, $m = 0, \ldots, n-1$, cannot all vanish.*

The proof is as follows. By Lemma 6.2, Adj $[\rho_0 I - A]_{ij}$ is a linear combination of the nonnegative numbers $[A^m]_{ij}$, $m = 0, \ldots, n-1$ and by the Irreducible Matrix Theorem we know that Adj $[\rho_0 I - A]_{ij} > 0$, which means that the coefficients $[A^m]_{ij}$, $m = 0, \ldots, n-1$, cannot all vanish.

A first consequence of Lemma 6.3 is an easy-to-apply sufficient condition for an irreducible matrix to be primitive [15, p. 553], namely the Trace Theorem 4.7 stated already in Sect. 4.2: *If A is irreducible and if* tr $A > 0$*, then A is primitive. Hence all imprimitive A have* tr $A = 0$ *and so all diagonal entries must be zero*. The proof is quite simple. Suppose tr $A > 0$ and that, e.g., $a_{11} > 0$. Then $[A^m]_{11} > 0$ since it is a sum of nonnegative terms one of which is $a_{11}^m > 0$. Now by Lemma 6.3 above for any i there is an $l < n$ for which $[A^l]_{i1} > 0$. Similarly $m < n$ exists with $[A^m]_{1j} > 0$. Since $[A^{l+m}]_{ij}$ contains the term $[A^l]_{i1}[A^m]_{1j}$, it is positive. In other words, $A^{l+m} > 0$, and so A is primitive by the Primitive Matrix Theorem 4.6.

The next theorem formed Frobenius' entrée into a deeper understanding of imprimitive matrices [15, p.554].

Theorem 6.4 *If $A \neq 0$ is any nonzero $n \times n$ matrix such that $A, A^2, \ldots, A^n$ are all irreducible, then A is primitive. Hence if $A \neq 0$ is imprimitive, there is an integer m, $1 < m \leq n$, such that A^m is reducible.*

Again the proof is easy, given what has gone before. Suppose that A and all its powers up to A^n are irreducible. Then since in particular A is irreducible, the Irreducible Matrix Theorem 4.5 shows that $B = \text{Adj}\,(\rho_0 I - A) > 0$. This implies that $BA > 0$ as well, since $[BA]_{ij} = 0$ would hold only if the jth column of A is 0, but then A would be reducible, contrary to assumption. Since by Lemma 6.2 B is a linear combination of A^m, $m = 0, \ldots, n-1$, $BA > 0$ is a linear combination of A^m, $m = 1, \ldots, n$, which means that not all of the n quantities $[A^m]_{11}$, $m = 1, \ldots, n$, can vanish. Thus $\operatorname{tr} A^{m_0} > 0$ for one of these values of m. Since A^{m_0} is irreducible, the Trace Theorem implies it is primitive and so $\rho_0^{m_0}$ strictly dominates the absolute values of all other characteristic roots $(\rho')^{m_0}$. Then ρ_0 strictly dominates the absolute values of all other roots ρ' of A and so A is primitive. This establishes the first statement in the theorem, and the second then follows immediately.

With Theorem 6.4 in mind, Frobenius obtained the following result [15, pp. 554–556].

Theorem 6.5 *If $A \neq 0$ is irreducible but A^m is reducible for some $m > 1$, then A^m is completely reducible, in the sense that A^m is permutationally similar to a block diagonal matrix in which the diagonal blocks are all irreducible.*

The starting point of the proof was again the Irreducible Matrix Theorem 4.5, specifically the fact that $\operatorname{Adj}(\rho_0 I - A) > 0$. As we have seen, this means that both the equations $A\mathbf{x} = \rho_0\mathbf{x}$ and $A^t\mathbf{y} = \rho_0\mathbf{y}$ have positive solutions obtained by using a column, respectively, row of $\operatorname{Adj}(\rho_0 I - A) > 0$. Frobenius also realized that a result from his 1909 paper, namely Proposition 4.3, remains valid for irreducible A (by the same line of reasoning): *the only nonnegative characteristic vector of A is (up to a positive multiple) the positive characteristic vector $\mathbf{x}$ corresponding to the maximal root ρ_0.*

Now since A^m is reducible and hence permutationally similar to a matrix in lower triangular block form, we can assume without loss of generality that A^m itself is in the block form

$$A^m = \begin{pmatrix} R_{11} & 0 & 0 & \cdots & 0 \\ R_{21} & R_{22} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ R_{\ell 1} & R_{\ell 2} & R_{\ell 3} & \cdots & R_{\ell\ell} \end{pmatrix},$$

where the diagonal blocks R_{ii} are irreducible. Using the existence of $\mathbf{x}$ and $\mathbf{y}$ and the italicized fact given above together with block multiplication, Frobenius deduced that all the nondiagonal blocks R_{ij}, $i \neq j$, must vanish, implying that A^m is indeed completely reducible. Furthermore the reasoning showed that each irreducible block R_{ii} has ρ_0^m as its maximal root.

Although Frobenius' step-by-step arguments leading up to Theorems 6.4–6.5 were simple and straightforward, piecing them together as he did so as to achieve these theorems was an act of brilliance. Even more brilliant was the way he was able to use these theorems to arrive at his remarkable Imprimitive Matrix Theorem 4.8. To do so required reasoning of a more complex nature, and for this reason the remainder of the outline of Frobenius' proof of Theorem 4.8 will be less complete than what has preceded.

Although Frobenius' proof of Theorem 4.8 is correct, it was not presented with his customary lucidity, possibly due to the more complicated nature of the reasoning. The following lemma, which was not formally stated by Frobenius, represents the guiding idea of the entire proof (see [15, p. 557]).

Lemma 6.6 *Let $A \neq 0$ be imprimitive and let ρ_i, $i = 0, \ldots, k-1$, denote the $k > 1$ characteristic roots of absolute value ρ_0. Then for any positive integer m, A^m is completely reducible into primitive blocks R_{ii} if and only if all k roots ρ_i satisfy $\rho_i^m = \rho_0^m$. In that case the number of diagonal blocks R_{ii} is precisely k.*

Lemma 6.6 leaves it unclear whether or not integers m actually exist for which A^m is completely reducible into primitive parts, but this follows readily. That is, if A is imprimitive, then we know by Theorem 6.4 that there is a power m_0, $1 < m_0 \leq n$, such that A^{m_0} reduces and so reduces completely by Theorem 6.5. The reasoning behind Lemma 6.6 implies that either all the irreducible blocks R_{ii} of A^{m_0} are primitive or all are imprimitive. In the latter case, we know a power $R_{ii}^{m_i}$ exists that is completely reducible. Thus if $M = \prod_{i=1}^{\ell} m_i$, A^M completely reduces into a greater number of irreducible blocks than in A^{m_0}. If these blocks are all imprimitive we can repeat the above reasoning to get an even larger power of A that reduces into a yet larger number of irreducible parts. Since the total number of irreducible parts cannot exceed the dimension n of A, it follows that this process must come to a stop, i.e., there will be a power $m \leq n$ such that A^m is completely reducible into primitive parts.

Let h denote the smallest power for which A^h is completely reducible into primitive parts. Then by Lemma 6.6 h is the smallest power such that all the k roots ρ_i, $i = 0, \ldots, k$, satisfy $\rho_i^h = \rho_0^h$, i.e., such that all k quotients ρ_i/ρ_0 are hth roots of unity. In particular, it follows that $k \leq h$.

Frobenius then considered the characteristic equation of A:

$$\varphi(\rho) = |\rho I - A| = \rho^n + c_1\rho^{n-1} + \cdots + c_m\rho^{n-m} + \cdots + c_n.$$

Consider the coefficient c_m. If m is not divisible by h, then $m = ph + q$ where p, q are nonnegative integers and $1 \leq q < h$. Thus for every quotient ρ_i/ρ_0 we have $(\rho_i/\rho_0)^m = (\rho_i/\rho_0)^{hp}(\rho_i/\rho_0)^q = (\rho_i/\rho_0)^q$. Hence if all (ρ_i/ρ_0) were mth roots of unity, they would all be qth roots of unity, which is impossible since $q < h$. This means (by Lemma 6.6) that A^m is not completely reducible into primitive parts, i.e., either A^m is imprimitive or is completely reducible into irreducible blocks R_{jj} that are all imprimitive. Thus in either case the Trace Theorem 4.7 implies $\operatorname{tr} A^m = 0$, or, equivalently that the sum of the mth powers of all the roots of $\varphi(\rho)$ vanishes. From Newton's identities Frobenius then deduced by induction that $c_m = 0$ for m not divisible by h [15, p. 557].

The fact that $c_m = 0$ whenever m is not divisible by h implies first of all that $h \leq n$. For if $h > n$ then *all* coefficients c_m of φ vanish and $\varphi(\rho) = \rho^n$, which is impossible since $\rho_0 > 0$ is a root. Thus $h \leq n$ and

$$\varphi(\rho) = \rho^n + a_1\rho^{n-m_1h} + a_2\rho^{n-m_2h} + \cdots, \tag{6.10}$$

where $a_i \neq 0$ for all i and $m_1 < m_2 < \cdots$. From this special form for $\varphi(\rho)$ it follows that if ϵ is any hth root of unity, then

$$\varphi(\epsilon\rho) = \epsilon^n\varphi(\rho), \quad \varphi'(\epsilon\rho) = \epsilon^{n-1}\varphi'(\rho).$$

These relations show that if ρ' is any root of φ then so is $\epsilon\rho'$, and if ρ is a simple root (so $\varphi'(\rho) \neq 0$) then so is $\epsilon\rho$. It thus follows that if $\epsilon = e^{2\pi i/h}$ (a primitive hth root of unity) then the $h \geq k$ roots $\epsilon^i \rho_0$, $i = 0, \ldots, h-1$, all have absolute value 1, which means that $h = k$ and the much-discussed special roots ρ_i, $i = 0, \ldots, k-1$, are precisely the roots $\epsilon^i \rho_0$, $i = 0, \ldots, k-1$, and they are all simple.

From the above proof-sketch, with h everwhere now replaced by k, Parts (1)–(4) of the Imprimitive Matrix Theorem 4.8 follow. Part (5) then follows readily, as indicated following the statement of the theorem in Sect. 4.2.

6.3 Markov's proofs

6.3.1 Simplicity of $y = 1$

Markov based his proof in the $n \times n$ case on the fact (already used in dealing with the $n = 3$ case) that the derivative of $\varphi(y) = \det(yI - A)$, $A = P^t$, is the sum of its principal minors of degree $n - 1$, an identity also used by Frobenius that same year [13, p. 405]; see (4.1). Thus

$$\varphi'(y) = \sum_{\ell=1}^{n} \varphi_{\ell\ell}(y), \tag{6.11}$$

where $\varphi_{\ell\ell}(y)$ is the principal minor determinant of $\varphi(y) = \det(yI - A)$ obtained by deleting row and column ℓ. It will helpful to express (6.11) with $y = 1$ as Markov did. Thus $\varphi'(1)$ is equal to

$$\begin{vmatrix} 1 - p_{\beta\beta} & -p_{\gamma\beta} & \cdots \\ -p_{\beta\gamma} & 1 - p_{\gamma\gamma} & \cdots \\ \cdots & \cdots & \cdots \end{vmatrix} + \begin{vmatrix} 1 - p_{\alpha\alpha} & -p_{\gamma\alpha} & \cdots \\ -p_{\alpha\gamma} & 1 - p_{\gamma\gamma} & \cdots \\ \cdots & \cdots & \cdots \end{vmatrix} + \cdots. \tag{6.12}$$

All of the terms in this sum are of the form $\det M$ where $M = (m_{ij})$ is a matrix with the following properties:

$$m_{ij} \leq 0 \quad \text{for all } i \neq j \quad (\text{Markov's equation } (*)), \tag{6.13}$$

$$\sum_{i=1}^{n} m_{ij} \geq 0 \quad \text{for all } j \quad (\text{Markov's equation } (**)). \tag{6.14}$$

Thus (6.13) states that M is such that the off-diagonal entries are nonpositive, and (6.14) states that the diagonal entry m_{jj} is sufficiently positive that the sum of all the entries in column j is nonnegative. To see that this is true of the matrices in (6.12), note that, since P is row-stochastic, the sum of the first column of the first matrix is

$$1 - \sum_{\mu \neq \alpha} p_{\beta\mu} = p_{\beta\alpha} \geq 0.$$

In general the sum of the λth column is $p_{\lambda\alpha} \geq 0$. Likewise the column sums in the other matrices are other coefficients of P and so are all nonnegative. The only coefficients of P that do not occur among these column sums are the n probabilities $p_{\mu\mu}$. The desired conclusion then is that $\det M > 0$, or at least that $\det M \geq 0$ with $\det M = 0$ only in extraordinary circumstances that would insure that at least one of the determinants in (6.12) would be positive so that $\varphi'(1) > 0$.

Markov realized that in 1900 Minkowski had proved a "similar proposition" [28, p. 572n] in a 1900 paper on units in an algebraic number field. Minkowski's result was the following lemma, which he stated and proved by induction at the beginning of his brief paper [31, p. 316].

Lemma 6.7 (Minkowski) *Let $M = (m_{ij})$ be such that* (6.13) *and* (6.14) *hold in the stronger sense of strict inequality. Then* $\det M > 0$.

Minkowski's Lemma thus shows via (6.11) that $\varphi'(1) > 0$ when all $p_{\mu\nu}$ are strictly positive, and this was, as noted above, Markov's initial working hypothesis. Realizing the hypothesis that $P > 0$ was unnecessarily restrictive, however, he replaced the positivity of P by his Conditions ID–IID of Sect. 5.1 and posited the following generalization of Minkowski's Lemma: under the weaker conditions (6.13)–(6.14) $\det M$ "cannot be a negative quantity and might be zero only in the extreme cases when all the inequalities (**) are equalities, or when it becomes a product of several determinants of the same type and among these there is a determinant for which all the inequalities (**) are turned into equalities" [28, p. 572]. He only hinted at the possibility of a proof by induction that would be based upon the fact that the principal minors of M are "similar determinants of lower order" [28, p. 573]; but, as Schneider has observed [53, p. 146], such an approach is of questionable viability since the principal minors of an irreducible M need not be irreducible.[47] In any case Markov certainly did not provide a viable proof of his generalization of Minkowski's Lemma and so did not provide a proof that $y = 1$ is simple except in the case $P > 0$ (to which Minkowski's Lemma applies).

6.3.2 All roots $y \neq 1$ have $|y| < 1$

Let us now consider Markov's proof that $|y| < 1$ for all characteristic roots $y \neq 1$ of P. Given a root $y \neq 1$, it follows that

$$P\mathbf{x} = y\mathbf{x}, \quad \mathbf{x} = \left(\alpha' \ \beta' \ \gamma' \ \cdots\right)^t \neq \mathbf{0}.$$

[47] For example, if $M = I - P$, where $P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$, then P and M are irreducible (as can be seen readily from their graphs) but the matrix of the $(3, 3)$ principal minor is reducible. It would seem that Ostrowski was misleadingly generous when in 1937 he wrote that Markov "had showed that" if M is irreducible and satisfies (*) and (**), then $\det M \geq 0$ with $\det M = 0$ if and only if all the expressions in (**) are zero [33, p. 73].

Since the rows of P sum to 1, $P\mathbf{x}_0 = \mathbf{x}_0$ for $\mathbf{x}_0 = \begin{pmatrix}1 & 1 & 1 & \cdots\end{pmatrix}^t$. This means that not all the coefficients of $\mathbf{x}$ can be equal, for, if they were, then $\mathbf{x} = \alpha'\mathbf{x}_0$, which means $\mathbf{x}$ is a characteristic vector for 1 rather than y. Markov proceeded to distinguish two cases: either (i) not all components of $\mathbf{x}$ have the same absolute value or (ii) all components of $\mathbf{x}$ have the same absolute value.

Regarding case (i) he reasoned as follows [28, p. 574]:

> there must be at least one equation [in the system $y\mathbf{x} = P\mathbf{x}$] where the multiplier y on the left-hand side could be any of the numbers $\alpha', \beta', \gamma', \ldots$ with largest absolute value, while the right-hand side has coefficients different from 0 and contains numbers $\alpha', \beta', \gamma', \cdots$, less [in absolute value] than the largest in absolute value. From this it follows that $|y| < 1$, because the sum of the coefficients of $\alpha', \beta', \gamma', \cdots$ in the right-hand side of any equation is equal to 1.

Here is what I believe Markov is saying. It will be helpful to denote the characteristic vector $\mathbf{x}$ by $\mathbf{x} = \begin{pmatrix}x_1 & \cdots & x_n\end{pmatrix}^t$ rather than use Markov's notation. Set $m = \max_i |x_i|$. Then there are, say, p values of i, $1 \le p < n$, such that $|x_i| = m$, whereas $|x_i| < m$ for the other $n - p$ values of i. He then assumed (in the above quotation) that *for at least one of these values of i, say $i = i_0$, $p_{i_0,j} \neq 0$ for some j such that $|x_j| < m$.* Assuming this for the moment, it then does indeed follow from consideration of the i_0th equation that

$$|y|m = |yx_{i_0}| = |\sum_{j=1}^{n} p_{i_0 j}x_j| \le \sum_{j=1}^{n} p_{i_0 j}|x_j| < \Big(\sum_{j=1}^{n} p_{i_0 j}\Big)m = 1\cdot m,$$

where the strict inequality follows because $|x_j| < m$ for some j with $p_{i_0 j} \neq 0$. Hence $|y| < 1$.

Why does Markov's italicized assumption have to hold? Let $C = \{i : |x_i| = m\}$ and $D = \{i : |x_i| < m\}$. Thus C consists of the p values of i described above and D is the complementary set of i so that C and D partition $\{1, \ldots, n\}$ into nonempty sets. Here the partition C, D is determined by the chosen characteristic vector, but in order that it apply to any characteristic vector of any probability matrix P the italicized assumption must hold for any partition C, D of $\{1, \ldots, n\}$ into nonempty sets. That this condition is equivalent to the irreducibility of P can be seen as follows. Suppose it fails to hold, then a partition C, D exists such that $p_{ij} = 0$ for every $(i, j) \in C \times D$. That is, there is a 0 at the intersections of the $p = |C|$ rows corresponding to $i \in C$ with the $q = n - p$ complementary columns corresponding to $j \in D$. This is Frobenius' criterion for reducibility (Sect. 4.2). Thus in his proof for case (i) Markov seems to have tacitly assumed that P satisfies the following condition, which is equivalent to irreducibility in Frobenius' sense.

Condition I*. *If C, D is any partition of $\{1, \ldots, n\}$ into nonempty sets, then there is an $i \in C$ and a $j \in D$ for which $p_{ij} \neq 0$.*

As already noted in Sect. 5.1, Condition I* seems to be a purely mathematical version of Markov's more probabilistically oriented Condition I and seems to have motivated his determinant-based Condition ID.

Let us now consider case (ii): $A\mathbf{x} = y\mathbf{x}$, $y \neq 1$, and all components of $\mathbf{x} = (\alpha' \; \beta' \; \gamma' \cdots)^t$ have the same absolute value, although not all can have the same complex argument. Markov's reasoning for case (ii) proceeded as follows [28, pp. 574–575]:

> Let us assume next that all the numbers $\alpha', \beta', \gamma', \cdots$ have the same absolute value. We know, however, that they are not all equal to one another. Thus they can be divided into two groups: One group of numbers equals α' and the other is unequal to α', being different in argument. On the other hand, having in mind the basic conditions, the set of sums
>
> $$\begin{array}{l} p_{\alpha,\alpha}\alpha' \;+\; p_{\alpha,\beta}\beta' \;+\; p_{\alpha,\gamma}\gamma' \;+\; \cdots, \\ p_{\beta,\alpha}\alpha' \;+\; p_{\beta,\beta}\beta' \;+\; p_{\beta,\gamma}\gamma' \;+\; \cdots, \\ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots\cdots . \end{array}$$
>
> cannot be divided into two sets so that from all the numbers $\alpha', \beta', \gamma', \cdots$ the sums of the first set would contain, with coefficients different from zero, only terms equal to α', the sums of the second set, only the ones not equal to α'. Consequently one of these sums certainly contains, with coefficients different from zero, both numbers equal to α' and numbers not equal to α'; thus the absolute value of the sum equals the product of the absolute values of y and α' and must be smaller than the absolute value of α' because for numbers with distinct arguments the absolute value of their sum is smaller than the sum of their absolute values and not equal to it.

Here is my interpretation of the above reasoning. Again I use the preferable notation $\mathbf{x} = (x_1 \cdots x_n)^t$, rather than $\mathbf{x} = (\alpha' \; \beta' \cdots)^t$. Let E consist of all indices i such that $x_i = x_1$ and F the complementary set of indices so that $x_i \neq x_1$ for all $i \in F$. Keep in mind that all components have the same absolute value: $|x_i| = m$ for all i. Markov assumed (in the above quotation) that

> *(II*) it is not possible to determine a partition of* $\{1, \ldots, n\}$ *(representing row sums 1 through n) into nonempty sets* G, H *with the following property: For each* $i \in G$ $p_{ij} = 0$ *for all* $j \in F$*; for each* $i \in H$ $p_{i,j} = 0$ *for all* $j \in E$.

By virtue of (II*) he concluded, quite rightly, that there exists a row index i_0 (corresponding to the i_0th sum above) with the property that column indices j and k exist such that $x_j = x_1$ (i.e., $j \in E$) and $x_k \neq x_1$ (i.e., $k \in F$) and both $p_{i_0,j} \neq 0$ and $p_{i_0k} \neq 0$.

Granted this conclusion, it does indeed follow that $|y| < 1$ because $p_{i_0j}x_j$ $(j \in E)$ and $p_{i_0k}x_k$ $(k \in F)$ are nonzero complex numbers with different arguments. Since the triangle inequality is only an equality for complex numbers with the same argument, from the i_0th row of the system $y\mathbf{x} = P\mathbf{x}$ we obtain by taking absolute values

$$|y| \cdot m = |yx_{i_0}| = |\sum_{\ell=1}^{n} p_{i_0\ell}x_\ell| < \sum_{\ell=1}^{n} |p_{i_0\ell}x_\ell| = \Big(\sum_{\ell=1}^{n} p_{i_0\ell}\Big)m = m,$$

whence $|y| < 1$.

As in the discussion of case (i), if Markov's assumption (II*) is freed from its dependence on the characteristic vector **x**, then it takes the following form.

Condition II*. For any partition of $\{1, \ldots, n\}$ into nonempty sets E, F there is no corresponding partition of $\{1, \ldots, n\}$ into nonempty sets G, H such that $p_{ij} = 0$ for all $(i, j) \in (G \times F) \cup (H \times E)$.

Markov's proof that $|y| < 1$ in case (ii) is clearly valid for any P satisfying Condition II*.

References

1. G. Abraham-Frois and E. Lendjel, editors. *Les Oeuvres Economiques de l'Abbé Potron*. L'Harmattan, Paris, 2004.
2. R. Baltzer. *Theorie und Anwendungen der Determinanten*. S. Hirzel, Leipzig, 3rd edition, 1870.
3. A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, New York, 1979.
4. C. Bidard, G. Erreygers, and W. Parys. Review of [1]. *European J. of the History of Economic Thought*, 13:163–167, 2006.
5. C. Bidard, G. Erreygers, and W. Parys. Our daily bread: Maurice Potron, from Catholicism to mathematical economics. *European J. of the History of Economic Thought*, 16(1), March 2009. Forthcoming.
6. K.-R. Biermann. *Die Mathematik und ihre Dozenten an der Berliner Universität 1810–1920*. Akademie-Verlag, Berlin, 1973.
7. C. Brezinski. *History of Continued Fractions and Padé Approximants*. Springer-Verlag, Berlin, 1991.
8. L. Collatz. Einschliessungssatz für die charakteristischen Zahlen von Matrizen. *Mathematische Zeitschrift*, 48:221–226, 1942.
9. Sh. K. Formanov and R. Mukhamedkhanova. On the origin and development of research in probability theory and mathematical statistics in Uzbekistan up to the middle of the twentieth century (in Russian). *Uzbek. Mat. Zh.*, 4:64–71, 2004.
10. E. Frank. Oskar Perron (1880–1975). *Journal of Number Theory*, 14:281–291, 1982.
11. H. Freudenthal. Pringsheim, Alfred. In *Dictionary of Scientific Biography*, volume 11, pages 148–149. Charles Scribner's Sons, New York, 1975.
12. G. Frobenius. Über lineare Substitutionen und bilineare Formen. *Jl. für die reine u. angew. Math.*, 84:1–63, 1877. Reprinted in *Abhandlungen* 1, 343-405.
13. G. Frobenius. Über Matrizen aus positiven Elementen. *Sitzungsberichte der Akademie der Wiss. zu Berlin*, pages 471–476, 1908. Reprinted in *Abhandlungen* 3, 404–409.
14. G. Frobenius. Über Matrizen aus positiven Elementen II. *Sitzungsberichte der Akademie der Wiss. zu Berlin*, pages 514–518, 1909. Reprinted in *Abhandlungen* 3, 410–414.
15. G. Frobenius. Über Matrizen aus nicht negativen Elementen. *Sitzungsberichte der Akademie der Wiss. zu Berlin*, pages 456–477, 1912. Reprinted in *Abhandlungen* 3, 546–567.
16. F. Gantmacher. *Matrix Theory*. AMS Chelsea Publishing, 2000. This work, published in two volumes, is an English translation of Gantmacher's *Teoriya Matrits* (Moscow, 1953). It first appeared in 1959.
17. F. Gantmacher and M. Krein. Sur les matrices complètement non négative et oscillatoires. *Compositio mathematica*, 4:445–476, 1937.
18. F. Gantmacher and M. Krein. *Ozillationsmatrizen, Ozillationskerne und klein Schwingungen mechanischer Systeme*. Akademie–Verlag, Berlin, 1960. Originally published in Russian (first edn. 1941). German version of the second edition edited by Alfred Stöhr.
19. J. Hadamard and M. Fréchet. Sur les probabilités discontinues des événements en châine. *Zeitschrift für angwandte Mathematik und Mechanik*, 13:92–97, 1933. Reprinted in J. Hadamard *Oeuvres* 4, 2083–2088.
20. T. Hawkins. Hypercomplex numbers, Lie groups and the creation of group representation theory. *Arch. Hist. Exact Sci.*, 8:243–287, 1972.
21. T. Hawkins. Another look at Cayley and the theory of matrices. *Archives internationales d'histoire des sciences*, 26:82–112, 1977.

22. T. Hawkins. Weierstrass and the theory of matrices. *Archive for History of Exact Sciences*, 17:119–163, 1977.
23. T. Hawkins. Frobenius, Cartan, and the Problem of Pfaff. *Archive for History of Exact Sciences*, 59:381–436, 2005.
24. T. Hawkins. Frobenius and the symbolical algebra of matrices. *Archive for History of Exact Sciences*, 62:23–57, 2008.
25. R. A. Howard, editor. *Dynamic Probabilistic Systems. Vol I: Markov Models*. Wiley, New York, 1971.
26. C. G. J. Jacobi. Allgemeine Theorie der kettenbruchähnlichen Algorithmen, in welchen jede Zahl aus drei vorhergehenden gebildet wird. *Jl. für die reine u. angew. Math.*, 69:29–64, 1868. Published posthumously by E. Heine. Reprinted in *Werke* 6, 385–426.
27. J. Kaucký. Remarques à la note de M. V. Romanovsky. *Comptes Rendus, Acad. Sci. Paris*, 191:919–921, 1930.
28. A. A. Markov. Rasprostranenie predel'nykh teorem ischisleniya veroyatnostei na summu velichin svyazannykh v tsep'. *Zap. (Mem.) Imp. Akad Nauk St. Peterb. Fiz.–Mat. Ser. 8, No. 3*, 1908. German translation by H. Liebmann on pp. 272–98 of [29]. English translation by G. Petelin on pp. 552–75 of [25]
29. A. A. Markov. *Warscheinlichkeitsrechnung*. B. G. Teubner, Leipzig, 1912.
30. F. Meyer. Über kettenbruchähnliche Algorithmen. In F. Rudio, editor, *Verhandlungen des ersern internationalen Mathematiker-Krongresses in Zürich vom 9. bis 11. August 1897*, pages 168–181. Teubner, 1898.
31. H. Minkowski. Zur Theorie der Einheiten in den algebraischen Zahlkörpern. *Göttinger Nachrichten, Math. -physik. Kl.*, pages 90–93, 1900. Reprinted in *Abhandlungen* 1, 316–319.
32. E. Ostenc. Sur les zéros des matrices stochastiques. *Comptes Rendus, Acad. Sci. Paris*, 196:150–151, 1933.
33. A. Ostrowski. Über die Determinanten mit überweigender Hauptdiagonale. *Comment. Math. Helv.*, 10:69–96, 1937. Reprinted in *Papers* 1, 31–58.
34. O. Perron. Note über die Konvergenz von Kettenbrüchen mit positiven Gliedern. *Sitzungsberichte der mathematisch–physikalischen Klasse der k. b. Akademie der Wissenschaften zu München 1905*, 35:315–322, 1906.
35. O. Perron. Über die Konvergenz periodischer Kettenbrüche. *Sitzungsberichte der mathematisch–physikalischen Klasse der k. b. Akademie der Wissenschaften zu München 1905*, 35:495–503, 1906.
36. O. Perron. Grundlagen für eine Theorie des Jacobischen Kettenbruchalgorithmus. *Math. Ann.*, 64:1–76, 1907.
37. O. Perron. Zur Theorie der Matrices. *Math. Ann.*, 64:248–263, 1907.
38. O. Perron. Über die Konvergenz der Jacobi-Kettenalgorithmen mit komplexen Elementen. *Sitzungsberichte der mathematisch–physikalischen Klasse der k. b. Akademie der Wissenschaften zu München 1908*, pages 401–481, 1908. Submitted at the 7 December 1907 session.
39. O. Perron. Alfred Pringsheim. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 56:1–6, 1953.
40. M. Potron. Quelques propriétés des substitutiones linéaires à coefficients ≥ 0 et leur application aux problèmes de la production et des salaires. *Comptes Rendus, Acad. Sci. Paris*, 153:1129–1132, 1911.
41. M. Potron. Application aux problèmes de la "production suffisante" et du "salaire vitale" de quelques propriétés des substitutions linéaires à coefficientes ≥ 0. *Comptes Rendus, Acad. Sci. Paris*, 153:1458–1459, 1911.
42. M. Potron. Quleques propriétés des substitutiones linéaires à coefficients ≥ 0 et leur application aux problèmes de la production et des salaires. *Annales scientifiques École Normale Sup. Paris*, (3) 30: 53–76, 1913.
43. A. Pringsheim. Ueber die Convergenz unendlicher Kettenbrüche. *Sitzungsberichte der mathematisch–physikalischen Klasse der k. b. Akademie der Wissenschaften zu München 1898*, pages 295–324, 1899.
44. A. Pringsheim. Ueber ein Convergenz-Kriterium für Kettenbrüche mit positiven Gliedern. *Sitzungsberichte der mathematisch–physikalischen Klasse der k. b. Akademie der Wissenschaften zu München 1899*, pages 261–268, 1900.
45. A. Pringsheim. Ueber die Convergenz periodischer Kettenbrüche. *Sitzungsberichte der mathematisch–physikalischen Klasse der k. b. Akademie der Wissenschaften zu München 1900*, pages 463–488, 1901.
46. V. Romanovsky. Sur les chaînes de Markoff. *Doklady Akademii nauk SSSR A*, pages 203–208, 1929.
47. V. Romanovsky. Sur les chaînes discrètes de Markoff. *Comptes Rendus, Acad. Sci. Paris*, 191:450–452, 1930.
48. V. Romanovsky. Sur les chaînes biconnexes continues de Markoff. *Comptes Rendus, Acad. Sci. Paris*, 191:695–697, 1930.

49. V. Romanovsky. Sur les zéros de matrices stocastiques. *Comptes Rendus, Acad. Sci. Paris*, pages 266–269, 1931.
50. V. Romanovsky. Un théorème sur les zéros des matrices non négatives. *Bulletin de la Société mathématique de France*, 61:213–219, 1933.
51. V. Romanovsky. Recherches sur les chaînes de Markoff. *Acta mathematica*, 66:147–251, 1936.
52. V. Romanovsky. *Discrete Markov Chains*. Wolters–Nordhoff, Groningen, 1970. Translated form the Russian edition (1945) by E. Seneta.
53. H. Schneider. The concept of irreducibility and full decomposibility of a matrix in the works of Frobenius, König and Markov. *Journal of Linear Algebra and its Applications*, 1977.
54. H. Schneider. The influence of the marked reduced graph of a nonnegative matrix on the Jordan form and on related properties: A survey. *Journal of Linear Algebra and its Applications*, 84:161–189, 1986.
55. O. Schreier and E. Sperner. *Vorlesungen über Matrizen*. B. G. Teubner, Leipzig, 1932.
56. E. Seneta. *Non-negative Matrices. An Introduction to Theory and Applications*. Wiley, New York, 1973.
57. E. Seneta. *Non-negative Matrices and Markov Chains*. Springer–Verlag, 1981.
58. O. Stolz. *Vorlesungen über allgemeine Arithmetik, nach den neueren Ansichten. Zweiter Theil: Arithmetik der complexen Zahlen mit geometrischen Anwendungen*. Teubner, Leipzig, 1886.
59. H. W. Turnbull and A. C. Aitken. *An Introduction to the Theory of Canonical Matrices*. Blackie and Son, London & Glasgow, 1932.
60. R. Varga. *Matrix Iterative Analysis*. Prentice–Hall, Englewood Cliffs, N. J., 1962.
61. R. von Mises. Über die Aufgaben und Ziele der angewandten Mathematik. *Zeitschrift für angewandte Mathematik und Mechanik*, 1:1–15, 1921.
62. R. von Mises. *Vorlesungen aus dem Gebiete der angewandten Mathematik. Band I: Warscheinlichkeitsrechnung und ihrer Anwendung in der Statistik und theoretischen Physik*. F. Deuticke, Leipzig, 1931.
63. J. H. M. Wedderburn. *Lectures on Matrices*. American Mathematical Society, New York, 1934.
64. W. H. M. Zijm. *Nonnegative Matrices in Dynamic Programming*. Mathematisch Centrum, Amsterdam, 1983.