MARKOV AND THE CREATION OF MARKOV CHAINS.

EUGENE SENETA*

Abstract.
We describe the life, times and legacy of Andrei Andreevich Markov (1856 -1922), and his writings on what became known as Markov chains. One focus is on his first paper [27] of 1906 on this topic, which already contains important contractivity principles embodied in the Markov - Dobrushin coefficient of ergodicity, which in fact makes an explicit appearance in that paper. The contractivity principles are shown directly to underpin a number of results of the later theory. The coefficient is especially useful as a condition number in measuring the effect of perturbation of a stochastic matrix on the stationary distribution (sensitivity analysis). Some recent work in this direction is reviewed from the standpoint of the paper [53], presented at the first of the present series of conferences [63].

Key words. biography, Markov chain, contractivity, stochastic matrix, coefficient of ergodicity, perturbation, condition number, Google matrix, Dobrushin, Besicovitch, Uspensky, Romanovsky.

AMS subject classifications. 01A55, 01A60, 60-03, 60J10, 60E15, 15A51, 15A12

1. Introduction. Andrei Andreevich Markov was born June 14th (June 2nd, old style), 1856, in Ryazan, Imperial Russia, and died on July 20, 1922, in Petrograd, which was – before the Revolution, and is now again – called Sankt Peterburg (St. Petersburg).

In his academic life, totally associated with St. Petersburg University and the Imperial Academy of Science, he excelled in three mathematical areas: the theory of numbers, mathematical analysis, and probability theory. What are now called Markov chains first appear in his work in a paper of 1906 [27], when Markov was 50 years old. It is the 150th anniversary of his birth, and the 100th anniversary of the appearance of this paper that we celebrate at the Markov Anniversary Meeting, Charleston, South Carolina, June 12 - 14, 2006.

Markov's writings on chains occur within his interest in probability theory. On the departure in 1883 of his mentor, Pafnuty Lvovich Chebyshev (1821 - 1894) from the university, Markov took over the teaching of the course on probability theory and continued to teach it yearly, even in his capacity of a Privat-Dozent (lecturer) after his own retirement from the university as Emeritus Professor in 1905.

His papers on Markov chains utilize the theory of determinants (of finite square matrices), and focus heavily on what are in effect finite stochastic matrices. However, explicit formulation and treatment in terms of matrix multiplication, properties of powers of stochastic matrices, and more generally of inhomogeneous products of stochastic matrices, and of associated spectral theory, are somewhat hidden, even though striking results, rediscovered by other authors many years later, follow from ideas in [27]. Our mathematical focus is an exploration of the contractivity ideas of that paper in the context of finite stochastic matrices , and specifically of the structure and usage of the Markov-Dobrushin coefficient of ergodicity.

Markov's motivation in writing the Markov chain papers was to show that the two classical theorems of probability theory, the Weak Law of Large Numbers and the Central Limit Theorem, could be extended to sums of dependent random variables. Thus he worked very much in terms of probabilistic quantities such as moments and

---

*School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia.(eseneta@maths.usyd.edu.au).

expectations, and particularly with positive matrices. The underlying matrix properties of general non-negative stochastic matrices, such as irreducibility, periodicity, stationary (invariant) vector, and asymptotic behavior of powers, which determine the nature of the probabilistic behavior, were not clearly in evidence.

The theory of finite non-negative matrices was beginning to emerge only contemporaneously with Markov's [27], [29] first papers on Markov chains, with the work of Perron [44] and Frobenius [14], [15]. The connection between the two directions, Markov and Perron-Frobenius is probably due to von Mises [40]. The theory of finite Markov chains was then developed from this standpoint in the treatises of Fréchet [13] and Romanovsky [45] on homogeneous finite Markov chains.

In our own times, the heavily influential book on finite homogeneous Markov chains has been that of Kemeny and Snell [23], which, while heavily matrix theoretic in its operations, avoids any mention of spectral theory, and in its discussion of ergodicity is closest in spirit to Markov's original memoir [27].

The issues raised in the preceding paragraphs have been discussed in more detail, especially with respect to coefficients of ergodicity and inhomogeneous products, in the author's paper [48], and the author's book [49]. This book of 1981 has been reissued in paperback form in February, 2006, with an (incomplete) additional bibliography on coefficients of ergodicity. More recently (1996) the paper [57] of the author, written for statisticians, explores some of what might have entered the content of the present paper. However, the technical emphasis there is probabilistic, including Section 5 ("Techniques of Markov's 1906 Paper") and Section 6 ("The Ergodicity Coefficient"). We shall recast some of this into matrix analytic form, and proceed in a generally different direction in our exploration of the consequences of [27].

2. Biographical Notes. In an anniversary paper , it is appropriate to give some details on Markov's life.

Markov's father was Andrei Grigorievich Markov (1823 - ?). In Russian usage the second name is a patronymic. Thus our Markov, baptized in the Russian Orthodox Church as Andrei, became Andrei Andreevich Markov. A.A. Markov's father, on completion of his studies in a theological seminary in 1844, entered the administration of the Ryazan Guberniia (a guberniia (governorship) was an administrative region). He eventually rose to a senior position as counsellor, becoming noted for his directness, honesty, and uncompromising nature, qualities reflected later in his son Andrei Andreevich. His diligence in unmasking financial corruption was not to the taste of his superiors, and he was eventually asked to retire. Consequently he became a para-legal clerk, and found the legal profession much to his taste. He was reputed to be a gambler, and an inveterate card player. Markov family lore has it that he once lost all his family assets to a card-sharp; but the loss was later reinstated.

Andrei Grigorievich married Nadezhda Petrovna Fedorova (1829 - ?) early in 1847. She was the daughter of a guberniia official. They had 6 children: Piotr (1849 - ?), Yevgeniia (1850 - 1920), Pavel (1852, died in childhood), Maria (1854 - 1875), our Andrei (1856 - 1922) and Mikhail (1859 - ?). Andrei Grigorievich was married twice. His second wife, Anna Iosifovna, was also the daughter of an official. They had three children: Vladimir (1871 - 1897), Lidia (1873 - 1942) and Ekaterina (1875 - ?).

Andrei Andreevich's half-brother, Vladimir Andreevich, was on the way to eminence as mathematician at St. Petersburg University in the area of number theory, but died early. He figures significantly in modern Russian mathematical historiography [25]. Number theory was one of the areas in which Andrei Andreevich excelled, and in which he seems to have influenced the young J.V. Uspensky (of whom more

shortly), in a direction akin to Vladimir's.

In the early 1860's Andrei Grigorievich moved with his family from Ryazan to St. Petersburg. He became steward to the estate of Ekaterina Aleksandrovna Valvatieva, a widow, who had two daughters, Maria and Elizaveta. Maria Ivanovna Valvatieva (1860 - 1942) was to become Andrei Andreevich's wife in 1883, when Maria's mother finally assented to the marriage, until then judging Andrei Andreevich's prospects insufficient. At the time of the marriage he was Privat-Dozent at the university, close to defending his Doctoral dissertation, with the prospect of a Professorship.

Andrei Andreevich was a sickly child. In childhood he had a bone disorder; one leg wouldn't straighten at the knee and he had to walk on crutches. He determinedly learned to dispense with the crutches during games by hopping on one leg. In St. Petersburg an eminent surgeon straightened the leg, allowing him to walk normally, although he retained a slight limp all his life. This did not stop him from taking long hikes of which he was fond, sometimes saying "While you can walk, you know you're alive". Thus he was not a "cripple", as oral tradition has it.

His carefree childhood ended in 1866 when, at age 10, he was placed into the 5th St. Petersburg Gimnaziia (High School), which was then on the outskirts of St. Petersburg. (His younger half-brother Vladimir was also a student there.) Markov was not a particularly good student in high school except in mathematics, with a strong interest also in the burning social issues of the time. On occasion he revealed a rather rebellious and uncompromising nature. This was to manifest itself later in numerous clashes with academic colleagues and with the tsarist regime itself. Nevertheless, even in high school he established contact with the St. Petersburg University's mathematics professors A.N. Korkin and E.I. Zolotarev through his precocious, though as it turned out, not new, method of solving differential equations. Finally, on graduation from high school in 1874, he entered in that year the physico-mathematical faculty of St. Petersburg University. In 1877 he received a gold medal and on completion of his studies in 1878 was retained by the university to prepare for a career as an academic. His first mathematical papers appeared in 1879.

His Master's and Doctoral dissertations (defended in 1880 and 1885 respectively) were in number theory. He began lecturing in 1880 as Privat-Dozent, and as already mentioned, in probability theory in 1883.

3. Probability and chain dependence. The stream of Markov's publications in probability was initially motivated by inadequacies in Chebyshev's treatment of the Central Limit Problem in 1887, and begins with a letter to his friend A.V. Vasiliev (1853 - 1929), which Vasiliev published in the *Izvestiia (Bulletin)* of Kazan University's Physico-Mathematical Society. The Weak Law of Large Numbers (WLLN) and the Central Limit Theorem were the focal probabilistic issues of the times.

The paper [27] in which a Markov chain, as a stochastically dependent sequence for which the WLLN holds, first appeared in Markov's writings, was likewise published in the *Izvestiia*. The paper is motivated by the need to construct a counterexample to what Markov interpreted [50], [57] as a claim in 1902 of P.A. Nekrasov (1853 - 1924) that pairwise independence of summands was necessary as well as sufficient for the WLLN to hold. From showing that the WLLN held in the presence of chain dependence, it was an obvious step to investigate that refinement of the WLLN which is the Central Limit Theorem. There followed a stream of papers in this direction [29], [30], [32], now in more high profile journals, but unfortunately omitting any further use of the contractivity methodology of the original 1906 paper. in particular omitting the ergodic coefficient. It should be mentioned that in correspondence [41]

with the St. Petersburg statistician A.A. Chuprov (1874 - 1926), initiated by fiery post-cards in late 1910 by Markov subsequent to Chuprov's mentioning Markov's arch-enemy P.A. Nekrasov in a positive light, Markov was made aware of earlier work on special kinds of Markov chains by Ernst Heinrich Bruns (1848 - 1919), later called "Markov-Bruns Chains" by Romanovsky [45]. Although the publication of Bruns 's book [7] in 1906 is contemporaneous with [27], in the preface Bruns claims that his book arises out of his lectures of the preceding 25 years. Bruns's own methodology, like Markov's, is direct, that is: not matrix-based. Immediately on becoming aware due to Chuprov, of Bruns's work, Markov [34] produced a paper on "Markov-Bruns" chains. This paper together with another stimulated by correspondence with Chuprov were both presented to the Imperial St. Petersburg Academy of Science on the same day (19 January, 1911, o.s.). Markov was not particularly well-read on the relevant probabilistic literature, and indeed appears not to have been conversant with the fact that the Bernoulli-Laplace urn model, of much earlier provenance, could be cast in the form of a homogeneous Markov chain until, apparently, [35]. When these early models of homogeneous Markov chains are cast in transition matrix form, the transition matrices all have zero entries. Markov, as we have mentioned, did not completely resolve the matrix structural issues (reducibility, periodicity) which can arise out of such forms of stochastic matrix. His (probabilistic) methodology was strongly focused on the Method of Moments in the guise of conditional and absolute expectations, and double probability generating functions. These functions are, indeed, closely linked to the determinant [46], [58] and spectral theory of stochastic matrices, and thus necessarily interact with the positioning of zeros in the transition matrix. We omit a more detailed discussion of them from this paper.

4. Markov's academic progress and later years. In 1886 he was appointed Extraordinary (Associate) Professor in the Department of Pure Mathematics, and later in that year elected Adjunct of the Imperial St. Petersburg Academy of Science, at the proposal of Chebyshev. On the 30 January (o.s.), 1890, he was elected Extraordinary (Associate) Academician, in place of V. Ya. Buniakovsky who had died in 1889. Competing for this vacant position was Sofiia Kovalevskaya (Sonia Kowalewski), whose work Markov continued, apparently mistakenly, to attack in characteristically volatile and unreflective fashion even after her death in January, 1891. Markov was promoted to Ordinary (full) Professor in 1893, and elected Ordinary (full) Academician in 1896. His mentor Chebyshev had died in 1894. The year 1903 saw the birth of his son, also Andrei, and thus also Andrei Andreevich Markov (1903 - 1979), who himself was to become an eminent mathematician, and Corresponding Member of the Academy of Sciences of the U.S.S.R. The identical 3-part name with his father has sometimes confused western writers producing photographs.

1900 saw the publication of the first edition of Markov's textbook *Ischislenie Veroiatnostei (The Calculus of Probabilities)*. The second edition appeared in Russian in 1908 and was translated into German by Heinrich Liebmann in 1912 as *Wahrscheinlichkeitsrechnung*, which became well-known in the West. This edition already contains Markov's (Tail) Inequality, a simple and more direct approach to the Bienaymé-Chebyshev Inequality in probability theory. The third edition of 1913, substantially expanded, and with a portrait of Jacob Bernoulli, was timed to appear in the year of the 200th anniversary of Jacob Bernoulli's WLLN. Markov organized a Commemorative Meeting of the Academy of Science in honor of the anniversary. Other speakers were Vasiliev and Chuprov. The fourth edition of 1924 [36] was posthumous, and published in Moscow in the early years of the Soviet era. It is again much expanded

and, of the Russian-language editions, now the most readily available.

The last years of Markov's life coincided with a stormy period of history. The revolution in Russia of February 1917 saw the fall of the monarchy, with the abdication of Tsar Nicholas II and the establishment of the Provisional Government in Russia. The October Revolution (25 October (o.s.); 7th November) of 1917 resulted in Bolshevik seizure of power. The name St. Petersburg, used till 1914 and the beginning of World War I, was changed to Petrograd which was used till 1924, the year of Lenin's death. It then became Leningrad until the demise of the Soviet Union.

The name Imperial Academy of Sciences in 1917 became the Russian Academy of Sciences until 1925, and then the Academy of Sciences of the U.S.S.R. In 1934 the Academy was transferred from Leningrad to Moscow, which had become the capital city after the Bolshevik seizure of power.

5. Markov's legacy. Tributes to Markov soon after his death appeared in the 1923 *Izvestiia* of the Russian Academy of Science, by Ya.V.Uspensky (= J.V. Uspensky) [65] and A.S. Bezikovich (= A.S. Besicovitch) [5]. Besicovitch [6] also wrote a biographical sketch for the posthumous edition of Markov's monograph [36] .

The name of Abram Samoilovitch Besicovitch (1891 - 1970) was to become very well known in the mathematical world. Besicovitch (this was the transliteration which he used from the Russian of his name) graduated in 1912 from the St. Petersburg University, where one of his teachers had been A.A. Markov, and Besicovitch's first paper published in 1915 in the then renamed Petrograd was in probability theory, a new proof of the classical (independent summands) Central Limit Theorem very much in the Chebyshev/Markov tradition. He left Leningrad (as Petrograd had become) illegally in 1924 by crossing the then-nearby border with Finland under cover of darkness. He took up a position at the University of Cambridge, England, in 1927 and was elected Fellow of the Royal Society in 1934, and to the Rouse Ball Chair of Mathematics in 1950, receiving many honors for his mathematical research, which was primarily in real variable analysis and measure theory, specifically almost periodic functions, geometry of plane sets, and Hausdorff measure. His *oeuvre* after his first paper contains very few papers on probabilistic topics, but he did influence a number of mathematicians during his life in England who eventually made very significant contributions to probabilistic topics, including Markov chains. In particular: P.A.P. Moran, I.J. Good and S.J. Taylor. Burkill [8] wrote Besicovitch's biographical essay/obituary.

Yakov Viktorovich Uspensky (1883 - 1947) seems to have been Markov's colleague at St. Petersburg University, and would have taken courses from him, being about 30 years old in 1913, when Markov stopped teaching probability at the University (Besicovitch thus about 22 years old). Uspensky's early work in Russia was in number theory, and gets considerable attention in the chapter on this topic in [25]. There seems to have been influence by the work on number theory of A.A. Markov. Uspensky's Master's thesis in this area was published in St. Petersburg in 1910. According to Markov [3] pp.19–20, Uspensky in May 1913 was Privat-Dozent at St. Petersburg University, and translated the celebrated 4th part of Jacob Bernoulli's *Ars Conjectandi* from Latin into Russian, for the 200th anniversary celebrations of Bernoulli's WLLN. A note in [3], p.73, identifies Uspensky as Academician of the Russian Academy of Science from 1921. His election to the Academy was supported by Markov, V.A. Steklov and A.N. Krylov, who give an account of his publications in the Academy's *Izvestiia*, Ser. 6, 15(1921), pp. 4–5. At this time Uspensky was (full) Professor at the University.

Uspensky's apparently last paper in Russian was published in 1924 in the *Doklady ANSSSR* and is on a probabilistic topic. He appears to have left Leningrad at about this time, and made his way to the United States. His first paper in English, according to Math. Sci. Net., appeared in the *American Mathematical Monthly* in 1927 and was also on a probabilistic topic. A note in [3], p. 167, says he worked in the United States from 1929. In the United States he used the English version James of Yakov (which is, more accurately translated, as Jacob). Although he continued to write in several areas, and gained considerable distinction, it is largely for his book of 1937, *Introduction to Mathematical Probability*, written as Professor of Mathematics at Stanford University, and based on his lectures there, that he is best known. The book [66] discusses only two-state Markov chains within its chapter *Fundamental Limit Theorems*. It is certainly heavily influenced by the work of the St. Petersburg School of Probability, and specifically by Markov, on the Central Limit Problem. Uspensky's book seems to have brought analytical probability, in the St. Petersburg tradition, to the United States, where it remained a primary probabilistic source until the appearance of W. Feller's *An Introduction to Probability Theory and Its Applications* in 1951. Feller's book contains a great deal on Markov chains, specifically the case of a denumerable number of states, for which a matrix/spectral approach is not adequate, and renewal theoretic arguments are employed.

We have already mentioned the book of Fréchet [13], which was the first monograph on finite Markov chains from a matrix standpoint. The matrix method for finite Markov chains was subsequently exposited very much from Markov's post-1906 standpoint, in monograph form in Russian, by Romanovsky [45]. It reappeared in English translation by the author of this paper in 1970.

Vsevolod Ivanovich Romanovsky (1879 - 1954), born in the town that became known as Alma Ata, received his secondary eduction in an academic high school ("Reelschule") in Tashkent. He completed his studies at St. Petersburg University in 1906, where he was retained to prepare for an academic career. Then he completed his Master's degree examinations in 1908, at which time he returned to teach mathematics at his old high school. From 1911 to 1915 he was at first Privat-Dozent and then Professor at Warsaw University (at the time part of Poland was still part of the Russian Empire). This university, as a Russian institution, was closed down, and for a year or so from 1915 he worked at Don University at Rostov-on-the-Don, and returned to Tashkent in 1917. From its beginning stages in 1918, till his death he was heavily involved in teaching and research in mathematics, and in administration at what became Tashkent State University (initially called Central Asian University), and with the organization in 1943 and functioning of the Academy of Science of the Uzbek S.S.R. In the early period of his research he worked on differential equations, algebraic questions, and (as expected from his student days at St. Petersburg) on number theory. His later research activities were very largely devoted to the theory and applications of probability theory and mathematical statistics. In spite of his geographical distance from the main academic centers of the Soviet Union, he managed to keep in touch with and publish on statistical topics in the important western European statistical and mathematical journals, such as *Biometrika, Metron, Comptes Rendus Acad. Sci. Paris, Rend. del Circ. Mat. di Palermo,* on issues of mathematics close in spirit to that of the English Biometric School of Karl Pearson, but using the probabilistic methodology of the St. Petersburg School. There was a fundamental paper on finite Markov chains in *Acta Mathematica* in 1936, presumably in imitation of A.A. Markov's French-language publication in this journal in 1910. Romanovsky's

geographical isolation within the Soviet Union seems to have helped him maintain a scientific activity in mathematical statistics and its applications when it was being severely attacked in the (European) Soviet centers. The distance from St. Petersburg-Petrograd-Leningrad on the other hand, would have worked against any personal contact with A.A. Markov in the last decade or so of Markov's life. Romanovsky's most important scientific work was on finite Markov chains (it began in 1928), and on their generalization. His *magnum opus* of 1949 on this topic [45], however, was algebraically intricate, and received little attention in comparison with the theory of denumerable chains developed by Kolmogorov from the 1930's.

In the English-speaking world, finite homogeneous Markov chain theory was reborn with Kemeny and Snell's book, as we have mentioned in our Introduction.

6. Some sources on Markov's life and work. For some time the best source on Markov, and the present author's primary source, on his life, and his publications, in number theory and probability theory, has been [37], a Russian-language book of about 720 pages. The part entitled Probability Theory includes reprinting of 7 of Markov's papers on Markov chains, including [27]. There are several important appendices, in particular an extensive biography by his son, and a survey of Markov senior's writings on number theory and probability by Yu. V. Linnik, N.A. Sapogov and V.N. Timofeev. There are additionally commentaries on the individual papers, the ones on the Markov chain papers are written by Sapogov. His commentary does not encompass all the important ideas in [27] , but makes the important point that the strict positivity of all transition probabilities in Markov's exposition can be relaxed to assuming a strictly positive column, a fact which had already been noticed by S.N. Bernstein (see [4]), and which has played an important part in the theory of ergodicity coefficients. A finite stochastic matrix with a strictly positive column has been called a "Markov" matrix. Also in [37] is a very complete and detailed listing, by year, of all of Markov's publications; of his lithographed course lectures; of literature about Markov; and a name index. The paper published in French in *Acta Mathematica* **33** (1910) 87-104 as "Recherches sur un cas remarquable d'épreuves dépendantes" is not included, but is of course readily available. It is essentially encompassed by the Russian-language articles [29], [30]. The 1908 paper [30] is included in [37].

Recently in a privately printed book [62] some of the contents of [37] have become available in English translation: [28], [31], [33] with Sapogov's commentaries; the sketch by Linnik *et al.*, ; and the biography of his father by A.A. Markov Jr. [38].

The biography by A.A. Markov Jr. is, understandably, written in the Soviet political spirit of the times. For example, P.A. Nekrasov is painted in a very negative way. In the last *glasnost* years of the Soviet Union a more balanced as well as considerably extended biographical study (on the basis of family documents and recollections, and archival documents from a number of archives) was prepared by Grodzensky [16].

Grodzensky's account consists of 5 chapters, contains a number of photographs, and features Markov's activity as a chess player. There are 3 appendices, respectively on number theory, mathematical analysis, and probability theory (this last by B.V. Gnedenko). There is a listing of Markov's publications, which concludes with [37], and a list of 210 references.

In English a recent biographical sketch is given by the author [59]. Markov's motivation for initiating his study of Markov chains, and his interaction with Nekrasov, are encompassed in the author's [50], [57], [60]. Sheynin [61] gives an introductory biographical sketch of Markov, and is largely concerned with describing with the content of Markov's probability monograph [36].

Finally, it is appropriate to mention the study by Basharin, Langville and Nau-
mov [2] prepared for an earlier conference in this series. The present paper (until a
late draft) was written without examining [2], There is, in the event, relatively little
overlap. On the technical side the emphasis in [2] is on the probabilistic aspects, as is
the case in [57], which it cites. The presence of the Markov family photographs in [2]
is very welcome. The reader is encouraged to read both [57], [2] in conjunction with
the present paper.

7. Contractivity principles in Markov's reasoning. In Sections 7.1 to 7.3
we present what may be extracted in essence from Markov [27], specifically its Section
5.

7.1. Markov's Contraction Inequality. Lemma 7.1 below, states what we
call Markov's Contraction Inequality, which is sometimes inappropriately attributed
to Paz [42],

LEMMA 7.1. *If* $\delta = \{\delta_s\}, \mathbf{w} = \{w_s\}$ *, are real-valued column N-vectors, and*
$\delta^T \mathbf{1} = 0$, *then*

$$(7.1) \qquad |\delta^T \mathbf{w}| \leq (\max w_s - \min w_s) \frac{1}{2} \sum_{s=1}^{N} |\delta_s|$$

$$= \max_{h,h'} |w_h - w_{h'}| \frac{1}{2} \sum_{s=1}^{N} |\delta_s|$$

Proof. Let $E = \{s; \delta_s \geq 0\}, \ F = \{s; \delta_s < 0\}$.   Then

$$(7.2) \qquad \sum_{s \in E} \delta_s = - \sum_{s \in F} \delta_s = \frac{1}{2} \sum_{s=1}^{N} |\delta_s|.$$

Also

$$v = \delta^T \mathbf{w} = \sum_{s=1}^{N} \delta_s w_s = \sum_{s \in E} \delta_s w_s + \sum_{s \in F} \delta_s w_s \leq (\max \ w_s) \sum_{s \in E} \delta_s + (\min w_s) \sum_{s \in F} \delta_s$$

$$= (\max w_s - \min w_s) \sum_{s \in E} \delta_s. \qquad\qquad \square$$

7.2. Contractive property of a stochastic matrix. The following lemma
expresses the averaging property of a stochastic matrix.

LEMMA 7.2. *Let* $\mathbf{P} = \{p_{ij}\}, i, j = 1, \cdots, N$ *be a stochastic matrix, so that* $\mathbf{P} \geq$
$\mathbf{0}, \mathbf{P1} = \mathbf{1}$. *Let* $\mathbf{w} = \{w_i\}$ *be a real-valued column N-vector, and put*

$$(7.3) \qquad\qquad\qquad \mathbf{v} = \mathbf{P} \ \mathbf{w}$$

Then, writing $\mathbf{v} = \{v_i\}$,

$$(7.4) \qquad\qquad \max_{h,h'} |v_h - v_{h'}| \leq H \max_{j,j'} |w_j - w_{j'}|$$

where

$$(7.5) \qquad H = \frac{1}{2} \max_{i,j} \sum_{s=1}^{N} |p_{is} - p_{js}|,$$

so $0 \le H \le 1$.

    Proof. From (7.3)

$$(7.6) \qquad v_i - v_j = \sum_{s=1}^{N} (p_{is} - p_{js}) w_s,$$

and since $\sum_{s=1}^{N}(p_{is} - p_{js}) = 0$ by stochasticity of $\mathbf{P}$, we may apply (7.1) to (7.6) to obtain

$$|v_i - v_j| \le (\max w_s - \min w_s) \frac{1}{2} \sum_{s=1}^{N} |p_{is} - p_{js}|$$

$$(7.7) \qquad = (\max w_s - \min w_s) H$$

from (7.5), whence (7.4) follows. $\square$

    LEMMA 7.3. *Putting* $\mathbf{P}^{n-1} = \{p_{sr}^{(n-1)}\}, n \ge 1$, *with* $\mathbf{P}^0 = \mathbf{I}$ *(the unit matrix),*

$$(7.8) \qquad \max_{h,h'} |p_{hr}^{(n)} - p_{h'r}^{(n)}| \le H^n, \; n \ge 0.$$

    Proof. Since $\mathbf{P}^n = \mathbf{P}\,\mathbf{P}^{n-1}, n \ge 1$, putting $w_s = p_{sr}^{(n-1)}$ for fixed r , and $s = 1, \cdots, N$, from (7.3) and (7.4):

$$(7.9) \qquad \max_{h,h'} |p_{hr}^{(n)} - p_{h'r}^{(n)}| \le H \max_{j,j'} |p_{jr}^{(n-1)} - p_{j'r}^{(n-1)}|$$

so by iterating (7.9) back, (7.8) obtains. $\square$

    If $\mathbf{P} > \mathbf{0}$ i.e. all entries are positive, as Markov effectively assumes, it is clear that $H < 1$ from the expression (7.5) for H ; and this is also clearly true if $\mathbf{P}$ has a strictly positive column (i.e. is a "Markov" matrix).

    When $H < 1$, (7.9) implies that as $n \to \infty$ all rows of $\mathbf{P}^n$ tend to coincidence (this property was later called "weak ergodicity").

    LEMMA 7.4. *For fixed* r, $\max_h p_{hr}^{(n)}$ *is non-increasing with increasing* n; *and* $\min_h p_{hr}^{(n)}$ *is non-decreasing with increasing* n, *so (since both sequences are bounded) both have limits as* $n \to \infty$. *When* $H < 1$, *all rows of* $\mathbf{P}^n$ *tend to the same limiting probability vector.*

    Proof. Using the notation of Lemma 7.2, since

$$\mathbf{v} = \mathbf{P}\mathbf{w}, \quad v_i = \sum_{j=1}^{N} p_{ij}\, w_j \le (\max w_j) \sum_{j=1}^{N} p_{ij}$$

so max $v_i \leq$ max w_j. Similarly min $v_i \geq$ min w_j.

Putting, for fixed r, $w_s = p_{sr}^{(n-1)}$, $v_i = p_{ir}^{(n)}$ the respective monotonicities follow. Now from (7.7),

$$v_i - v_j \leq (\max w_s - \min w_s)H$$

so the coincidence in the limit as $n - \infty$ of both the maximal and minimal of sequences follows when $H < 1$. □

The property of all rows of $\mathbf{P}^n$ actually tending to the same limiting probability distribution came to be called "strong ergodicity".

Notice that the argument of Lemma 7.4 uses the "backward" form: $\mathbf{P}^n = \mathbf{P}\,\mathbf{P}^{n-1}$; and obtains ergodicity of a finite homogeneous Markov chain, it would seem, at geometric rate of convergence providing $H < 1$, without use of Perron-Frobenius theory of non-negative matrices.

The notation "H" of (7.5) is actually Markov's [27], and the expression (7.5) implicitly appears in this paper. The form of H has been ascribed to Dobrushin [12]. We think [57] it appropriate to call it the Markov-Dobrushin coefficient of ergodicity.

7.3. Attribution. A great deal of theory for stochastic matrices/Markov chains can be developed from the inequality (7.1). It remains true if $\mathbf{w}$ is replaced by an N-vector $\mathbf{z} = \{z_i\}$ each of whose elements may be real or complex, so that

$$(7.10) \qquad |\delta^T \mathbf{z}| \leq \max_{h,h'} |z_h - z_{h'}| \, \frac{1}{2} \sum_{s=1}^{N} |\delta_s|$$

This inequality (7.10) for finite N is due to Alpin and Gabassov [1], where it is proved by induction on N . It follows also from a problem, given without solution in Paz ([42], p.73, Problem 16), and restated in [48], p.583, where (7.10) is derived from it. The inequality (7.10) does not appear in [42]. To rectify the question of attribution further, since (7.1) plays a crucial role in the perturbation (sensitivity) theory of stochastic matrices to be discussed below, and (7.10) plays a crucial role in spectral bounding theory, we restate verbatim Paz's problem:

> **16.** Prove that for any vector ξ such that $\| \xi \| < \infty$ and $\sum \xi_i = 0$ can be expressed in the form $\xi = \sum_{i=1}^{\infty} \zeta_i$ where $\zeta_i = (\zeta_{ij})$ vectors have only two non-zero entries, $\| \zeta_i \| < \infty$, $\sum_j \zeta_{ij} = 0$, and $\| \xi \| = \sum \| \zeta_i \|$.

The norm used in the above is $\| \cdot \|_1$.

The following is, with small notational changes, Lemma 2.4 of [49], p.62. Here $\mathbf{f}_k$ denotes the vector with unity in the kth position, and zeros elsewhere.

LEMMA 7.5. *Suppose $\delta \in \Re^N, N \geq 2, \delta^T \mathbf{1} = 0, \delta \neq \mathbf{0}$. Then for a suitable set $I = I(\delta)$ of ordered pairs of indices (i,j), $i,j = 1, \cdots, N$,*

$$\delta = \sum_{(i,j) \in I} \left(\frac{\eta_{ij}}{2}\right) \gamma(i,j)$$

where $\eta_{i,j} > 0$ and $\sum_{(i,j) \in I} \eta_{i,j} = \| \delta \|_1$, and $\gamma(i,j) = \mathbf{f}_i - \mathbf{f}_j$. □

A proof of this lemma, by induction on N is given on [49], p.63. This can be worked up into a constructive proof. (7.10) (and (7.1)) follow immediately.

Lemma 7.1 clearly remains valid, with essentially the proof given, for real valued vectors $\delta, \mathbf{w}$, of countably infinite length, providing $\| \delta^T \|_1 = \sum |\delta_i| < \infty, \delta^T \mathbf{1} = 0, |w_i| < K < \infty$, in which case

$$(7.11) \qquad\qquad |\delta^T \mathbf{w}| \leq (\sup_s w_s - \inf w_s)\frac{1}{2}\| \delta^T \|_1$$

$$= \sup_{h,h'}|w_h - w_{h'}|\frac{1}{2}\| \delta^T \|_1.$$

We propose that the name Markov's Contraction Inequality be used for both (7.1) and (7.10), although the name Lemma PS, as used in the body of Kirkland, Neumann and Shader [24] is a reasonable compromise. "Paz's Inequality" is not an appropriate name. This is no reflection on A. Paz's excellent [42], Chapter II on finite and countably infinite, homogeneous and inhomogeneous, Markov chains with emphasis on the countably infinite and inhomogeneous, using what we have called the Markov-Dobrushin coefficient (7.5).

For subsequent sections, we need to show the dependence of H on $\mathbf{P}$ explicitly. so we change the notation to more recent usage at (8.1).

8. Some direct consequences of Markov's contractivity principles. With little extra effort, Lemmas 7.1 - 7.4 may be used to obtain direct results, which in qualitative nature are as good as known results using more elaborate (albeit related) superstructure. We give two examples.

8.1. Weak and strong ergodicity of inhomogeneous products. For an $N \times N$ stochastic matrix $\mathbf{P} = \{p_{ij}\}$, write

$$(8.1) \qquad\qquad \tau_1(\mathbf{P}) = \frac{1}{2} \max_{i,j} \sum_{k=1}^{N} |p_{ik} - p_{jk}|,$$

$$(8.2) \qquad\qquad \Delta(\mathbf{P}) = \frac{1}{2} \sum_{k=1}^{N} \max_{i,j} |p_{ik} - p_{jk}|.$$

Notice that

$$\tau_1(\mathbf{P}) \leq \Delta(\mathbf{P}),$$

and that $\Delta(\mathbf{P}) = 0$ (whenever $\tau_1(\mathbf{P}) = 0$, both zero values expressing equality of all rows of $\mathbf{P}$. We have noted that $0 \leq \tau_1(\mathbf{P}) \leq 1$ always, but it is possible for $\Delta(\mathbf{P}) > 1$. For example, if $\mathbf{P} = I$, $\tau_1(\mathbf{P}) = 1$, but $\Delta(\mathbf{P}) = N/2$.

Now let $\mathbf{P_1} = \{p_{ij}(1)\}$ and $\mathbf{P_2} = \{p_{ij}(2)\}$ be $N \times N$ stochastic matrices, and put $\mathbf{U} = \{u_{ij}\} = \mathbf{P_2}\,\mathbf{P_1}$. From (7.4)

$$\max_{h,h'} |u_{hj} - u_{h'j}| \leq \tau_1(\mathbf{P_2})\ \max_{i,i'}|p_{ij}(1) - p_{i'j}(1)|$$

for fixed j , so that

$$(8.3) \qquad\qquad \Delta(\mathbf{U}) = \Delta(\ \mathbf{P_2}\,\mathbf{P_1}) \leq \tau_1(\mathbf{P_2})\,\Delta(\mathbf{P_1}).$$

Now put

$$(8.4) \qquad \mathbf{U_{p,r}} = \{u_{ij}^{(p,r)}\} = \mathbf{H_{p+r}} \cdots \mathbf{H_{p+2}} \, \mathbf{H_{p+1}}$$

where $\{\mathbf{H_i}\}$, $\mathbf{i} \geq 1$ are $N \times N$ stochastic matrices.

Since

$$(8.5) \qquad \mathbf{U_{p,r}} = \mathbf{H_{p+r}} \, \mathbf{U_{p,r-1}}$$

as in Lemma 7.4, as $r \to \infty$:

$$(8.6) \qquad \max_h u_{jj}^{(p,r)} \downarrow \overline{u}_j^{(p)}, \ \min_h u_{hj}^{(p,r)} \uparrow \underline{u}_j^{(p)}.$$

for fixed j, p for some limit quantities $\overline{u}_j^{(p)}$, $\underline{u}_j^{(p)}$. Further, from (8.3) and (8.4)

$$(8.7) \qquad \Delta(\mathbf{U_{p,r}}) \leq \tau_1(\mathbf{H_{p+r}}) \, \Delta(\mathbf{U_{p,r-1}})$$

and iterating (8.7)

$$(8.8) \qquad \Delta(\mathbf{U_{p,r}}) \leq \prod_{s=1}^{r} \tau_1(\mathbf{H_{p+s}})(N/2)$$

since $\Delta(\mathbf{I}) = N/2$.

Now, weak ergodicity (for fixed p) is said to obtain if the rows of $\mathbf{U_{p,r}}$ tend to equality as $r \to \infty$; that is, if and only if $\Delta(\mathbf{U_{p,r}}) \to 0$ as $r \to \infty$. From (8.7) and (8.5) we see that weak ergodicity holds for backwards products (8.4) if and only if strong ergodicity (all rows tending to the same probability vector) holds. This result occurs in Chatterjee and Seneta [9] and is discussed in [49], Section 4.6. .

We see that the proof follows very much from the "backward" multiplication structure inherent in Lemma 7.4 .

Further, if we form successive an inhomogeneous matrix products $\mathbf{T_{p,r}}$ stochastic matrices $\{\mathbf{H_i}\}$, $\mathbf{i} \geq 1$, in any order, for fixed $\mathbf{r}$,

$$(8.9) \qquad \Delta(\mathbf{T_{p,r}}) \leq \prod_{\mathbf{s=1}}^{\mathbf{r}} \tau_1(\mathbf{H_{p+s}})(N/2),$$

so a sufficient condition weak ergodicity of the sequence $\mathbf{T_{p,r}}$ is

$$\sum_{\mathbf{s=1}}^{\infty} \{1 - \tau_1(\mathbf{H_{p+s}})\} = \infty,$$

where $\mathbf{H_{p+s}}$, $\mathbf{s} = 1, 2, \cdots, \mathbf{r}$ now simply labels the order of selection of the matrices which go to form $\mathbf{T_{p+r}}$, irrespective of where each new matrix is placed in going from $\mathbf{T_{p,r}}$ to $\mathbf{T_{p,r+1}}$.

Notice that we *have not* used here the submultiplicative property

(8.10) $$\tau_1(\mathbf{P_2}\,\mathbf{P_1}) \leq \tau_1(\mathbf{P_2})\,\tau_1(\mathbf{P_1})$$

for stochastic matrices $\mathbf{P_1}$, $\mathbf{P_2}$. The submultiplicative property is derived, for finite or infinite compatible stochastic matrices, using in effect direct ideas very similar to Markov's contractivity arguments by Isaacson and Madsen [22] as Lemma V.2.3, pp. 143 - 146; and by Iosifescu [21], as Theorem 1.11, pp. 58-59.

We take this opportunity to mention the author's paper [55] which shows that a condition expressed in terms of Birkhoff's coefficient of ergodicity implies a ratio limit property, as well as weak ergodicity, for inhomogeneous products of infinite stochastic matrices. The condition generalizes a classical condition of Kolmogorov. The Birkhoff coefficient $\tau_B(\mathbf{P}) \geq \tau_1(\mathbf{P})$ for a stochastic $\mathbf{P}$ [49], Theorem 3.13.

8.2. Rate of convergence to ergodicity. The Google matrix. Markov's argument embodied in Lemmas 7.3 and 7.4 gives a geometric rate H to equalization of rows as embodied in (7.8), providing $H \equiv \tau_1(\mathbf{P}) < 1$. This is easily extended to the more conventional concept of convergence at geometric rate. For a less direct proof, see the author's [54].

THEOREM 8.1. *Suppose* $\mathbf{P}$ *is* $(N \times N)$ *stochastic, with* $H < 1$, *and suppose* $\pi^T = \{\pi_r\}$ *is the common probability distribution to which each row of* $\mathbf{P}^n$ *converges as* $n \to \infty$. *Then for fixed* r

$$|p_{ir}^{(n)} - \pi_r| \leq H^n, \ n \geq 0.$$

Proof. Since by Lemma 7.4

$$\mathbf{P}^n \to \mathbf{1}\,\pi^T$$

where $\pi \geq \mathbf{0}$, $\pi^T\,\mathbf{1} = 1$, it follows that $\pi^T\,\mathbf{P}^n = \pi^T$, $n \geq 0$, so

$$(\mathbf{I} - \mathbf{1}\,\pi^T)\mathbf{P}^n = \mathbf{P}^n - \mathbf{1}\,\pi^T$$

and

$$(\mathbf{I} - \mathbf{1}\,\pi^T)\mathbf{1} = \mathbf{0}.$$

Now

$$\mathbf{I} - \mathbf{1}\,\pi^T = \{\delta_{ij} - \pi_j\}$$

where δ_{ij} is the Kronecker delta. We see that $\sum_j (\delta_{ij} - \pi_j) = 0$. Hence fixing i, the vector $\delta_i = \{\delta_{ij} - \pi_i\}_{j=1}^N$ satisfies $\delta_i^T\mathbf{1} = 0$.

Thus fixing i, r and using (7.1)

$$
\begin{aligned}
|p_{ir}^{(n)} - \pi_r| &= |\sum_j \{\delta_{ij} - \pi_j\}\, p_{jr}^{(n)}| \\
&\leq \left(\frac{1}{2}\sum_s |\delta_{is} - \pi_s|\right)\left(\max_s p_{sr}^{(n)} - \min_s p_{sr}^{(n)}\right) \\
&= \left(\frac{1}{2}\sum_s |\delta_{is} - \pi_s|\right)\max_{k,t}|p_{kr}^{(n)} - p_{tr}^{(n)}| \\
&\leq \left(\frac{1}{2}\sum_s |\delta_{is} - \pi_s|\right)H^n
\end{aligned}
$$

from Lemma 7.3 ,

$$\leq \frac{1}{2}\left(\sum_s |\delta_{is}| + |\pi_s|\right) H^n$$

$$= \frac{1}{2} 2H^n \leq H^n.$$

This completes the proof. □

The *Google matrix* ([26], p.5) $\mathbf{P}$ is of form

$$\mathbf{P} = \alpha \mathbf{S} + (1 - \alpha)\mathbf{1}\mathbf{v}^T$$

where $0 < \alpha < 1$ and $\mathbf{v}^T > \mathbf{0}^T$ is a probability vector, and both can be arbitrarily chosen. $\mathbf{S}$ is a stochastic matrix. From (8.1)

$$H = \tau_1(\mathbf{P}) = \alpha\tau_1(\mathbf{S}) \leq \alpha$$

since $\tau_1(\mathbf{S}) \leq 1$. By Theorem 8.1, the rate of convergence to the limit distribution vector π^T by the power method is rapid [26], even with a small $(1 - \alpha)$. Using the relations $\pi^T = \pi^T\mathbf{P}$, $\pi^T\mathbf{1} = 1$ it follows immediately that

$$\pi^T = (1 - \alpha)\mathbf{v}^T(\mathbf{I} - \alpha\mathbf{S})^{-1}.$$

9. Measuring sensitivity under perturbation.

9.1. The setting. Norms and bounds. For $\mathbf{x}^T \in \Re^N$, if $\| \cdot \|$ is a vector norm on $\Re^N$, then the corresponding matrix norm for an $(N \times N)$ matrix $\mathbf{B} = \{b_{ij}\}$ is defined by

$$\| \mathbf{B} \| = \sup\{\mathbf{x}^T, \| \mathbf{x}^T \| = 1 : \| \mathbf{x}^T \mathbf{B} \|\}.$$

We focus on the l_p norms on $\Re^N$, where $\| \mathbf{x}^T \|_p = \left(\sum_{i=1}^N |x_i|^p\right)^{1/p}$ in the cases $p = 1, \infty$, where $\| \mathbf{x}^T \|_\infty = \max_i |x_i|$.

Then

$$(9.1) \qquad \| \mathbf{B} \|_1 = \max_i \sum_{j=1}^N |b_{ij}|, \ \| \mathbf{B} \|_\infty = \max_j \sum_{i=1}^N |b_{ij}|.$$

Dobrushin [12] showed that for an $(N \times N)$ stochastic matrix $\mathbf{P}$,

$$(9.2) \qquad \tau_1(\mathbf{P}) = \sup\{\delta^T, \| \delta^T \|_1 = 1, \delta^T\mathbf{1} = 0 : \| \delta^T\mathbf{P} \|_1\}$$

in the context of a Central Limit Theorem for non-homogeneous Markov chains (in-homogeneous products of stochastic matrices). The Central Limit Theorem direction was Markov's main concern, but, as we have noted, he never seems to have used the coefficient $\tau_1(\mathbf{P})$ in this setting. The submultiplicative property (8.10) follows trivially from (9.2).

For any $(N \times N)$ matrix $\mathbf{B} = \{b_{ij}\}$ we may define a Markov-Dobrushin-type coefficient of ergodicity more generally by

(9.3) $\tau_1(\mathbf{B}) = \sup\{\delta^T, \| \delta^T \|_1 = 1, \ \delta^T \mathbf{1} = 0 : \ \| \delta^T \mathbf{B} \|_1\}$

whence [51]

(9.4) $$\tau_1(\mathbf{B}) = \frac{1}{2} \max_{i,j} \sum_{s=1}^{N} |b_{is} - b_{js}|.$$

Suppose $\mathbf{P} = \{p_{ij}\}$ is an $N \times N$ stochastic matrix containing a single irreducible set of indices, so that there is a unique stationary distribution vector $\pi^T = \{\pi_i\}$, $[\pi^T(\mathbf{I} - \mathbf{P}) = \mathbf{0}^T, \pi^T \mathbf{1} = 1]$. Let $\overline{\mathbf{P}}$ be any other $(N \times N)$ stochastic matrix with this structure (the irreducible sets need not coincide), and $\overline{\pi}^T = \{\overline{\pi}_i\}$ its unique stationary distribution vector. Under the assumption on $\mathbf{P}$ the corresponding *fundamental matrix* [23] $\mathbf{Z}$ exists, where $\mathbf{Z} = (I - \mathbf{P} + \mathbf{1}\pi^T)^{-1}$. Set $\mathbf{E} = \{e_{ij}\} = \overline{\mathbf{P}} - \mathbf{P}$. Suppose that there exists an $N \times N$ matrix $\mathbf{C} = \{c_{ij}\}$ such that:

(9.5) $$\overline{\pi}^T - \pi^T = \overline{\pi}^T \mathbf{E}\, \mathbf{C}$$

THEOREM 9.1. *Under our prior conditions on P and $\overline{P}$, and assuming (9.5) holds:*

(9.6) $$\| \overline{\pi}^T - \pi^T \|_1 \le \tau_1(\mathbf{C}) \| \mathbf{E} \|_1$$

(9.7) $$\| \overline{\pi}^T - \pi^T \|_\infty \le \mathcal{T}(\mathbf{C}) \| \mathbf{E} \|_1$$

where

(9.8) $$\mathcal{T}(\mathbf{C}) = \frac{1}{2} \max_j \left(\max_{k,k'} |c_{kj} - c_{k'j}| \right).$$

Proof. (9.6) follows by imitating the last part of the proof of [53] Theorem 2, using (9.2).

For the jth column vector of $\mathbf{C}$ write $\mathbf{C}_{\cdot j}$, and for the kth row of $\mathbf{E}$ write $\mathbf{E}_{k\cdot}^T$.

From (9.5)

$$\overline{\pi}_j - \pi_j = \sum_k \overline{\pi}_k (\mathbf{EC})_{kj}$$

so

$$|\overline{\pi}_j - \pi_j| \le \max_k |(\mathbf{EC})_{kj}|$$
$$= \max_k |\mathbf{E}_{k\cdot}^T \, \mathbf{C}_{\cdot j}|$$

so by Markov's Contraction Inequality (7.1) since $\mathbf{E}_k^T \mathbf{1} = 0$,

$$\leq \max_k \left(\max_{h,h'} |c_{hj} - c_{h'j}| \left(\frac{1}{2} \sum_s |e_{ks}| \right) \right)$$

$$= \frac{1}{2} \left(\max_{h,h'} |c_{hj} - c_{h'j}| \left(\max_k \sum_s |e_{ks}| \right) \right)$$

so

(9.9) $$|\overline{\pi}_j - \pi_j| \leq \frac{1}{2} \left(\max_{h,h'} |c_{hj} - c_{h'j}| \right) \| \mathbf{E} \|_1$$

where

$$\| \mathbf{E} \|_1 = \max_k \sum_s |e_{ks}|.$$

Thus from (9.9)

(9.10) $$\max_j |\pi_j - \pi_j| = \| \overline{\pi}^T - \pi^T \|_\infty \leq \mathcal{T}(\mathbf{C}) \| E \|_1.$$

This completes the argument. □

The result (9.6) was obtained by the author [53] Theorem 2, in the case $\mathbf{C} = \mathbf{C}(\mathbf{u}, \mathbf{v})$ where:

(9.11) $$\begin{aligned} \mathbf{C}(\mathbf{u}, \mathbf{v}) &= (\mathbf{I} - \mathbf{P} + \mathbf{1}\,\mathbf{u}^T)^{-1} - \mathbf{1}\mathbf{v}^T \\ &= (\mathbf{Z}^{-1} + \mathbf{1}(\mathbf{u} - \pi)^T)^{-1} - \mathbf{1}\mathbf{v}^T \\ &= \mathbf{Z} - \frac{\mathbf{1}(\mathbf{u} - \pi)^T \mathbf{Z}}{\mathbf{u}^T \mathbf{1}} - \mathbf{1}\,\mathbf{v}^T \end{aligned}$$

since $\mathbf{Z}\mathbf{1} = \mathbf{1}$, for any (real) $\mathbf{v}$, and any (real $\mathbf{u}$) such that $\mathbf{u}^T \mathbf{1} \neq 0$, using Bartlett's Identity. Notice that $\mathbf{C}(\pi, \pi) = \mathbf{A}^\sharp = \mathbf{Z} - \mathbf{1}\pi^T$, the *group generalized inverse* [39] $\mathbf{A}^\sharp$ of $\mathbf{A} = \mathbf{I} - \mathbf{P}$; while $\mathbf{C}(\pi, \mathbf{0}) = \mathbf{Z}$.

In fact it is shown in [53] that $\tau_1(\mathbf{C}(\mathbf{u}, \mathbf{v})) = \tau_1(\mathbf{A}^\sharp) = \tau_1(\mathbf{Z})$.

The steps in the proof of (9.7) are due to Kirkland, Neumann and Shader [24], Theorem 2.2, in the case $\mathbf{C} = \mathbf{A}^\sharp$, The point of imitating the steps here is primarily to show that, in the guise of "Lemma PS", Markov's Contraction Inequality is the central ingredient.

Cho and Meyer [11] have shown that the bound on the right of (9.7) in a different guise also occurs in 1984 in Haviv and Van der Heyden [17], and later, in Cho and Meyer [10]. Haviv and Van der Heyden [17] also used Lemma 7.1, and Hunter [20], in the proof of his Theorem 3.2, ascribes the result to these authors.

9.2. Measuring sensitivity. The relative effect on π^T of the perturbation $\mathbf{E}$ to $\mathbf{P}$ is measured in a natural way by the quantity

(9.12) $$\frac{\| \overline{\pi}^T - \pi^T \| / \| \pi^T \|}{\| \mathbf{E} \| / \| \mathbf{P} \|}.$$

From (9.6), using $\| \cdot \|_1$, and taking $\mathbf{C} = \mathbf{A}^\sharp$ in (9.5), we see since $\| \mathbf{P} \|_1 = 1$ that (9.12) $\leq \tau_1(\mathbf{A}^\sharp)$, so $\tau_1(\mathbf{A}^\sharp)$ is a natural condition number to measure the relative sensitivity of π^T to perturbation of $\mathbf{P}$.

The foundation paper from which sensitivity theory developed is [47].

Cho and Meyer [11] survey various condition numbers $\kappa_l, l = 1, \cdots, 8$ which have occurred in the literature which satisfy

$$\| \bar{\pi}^T - \pi^T \|_p \leq \kappa_l \| \mathbf{E} \|_q$$

where $(p, q) = (1, 1)$ or $(\infty, 1)$ depending on l. In this sense, in particular from their Section 4 for $(p, q) = (1, 1)$, $\kappa_6 = \tau_1(\mathbf{A}^\sharp)$, and for $(p, q) = (1\infty, 1)$, $\kappa_3 = \kappa_8 = \mathcal{T}(\mathbf{A}^\sharp)$ are condition numbers.

However, one might argue that, inasmuch as a condition number should bound (9.12), the same norm should be used for numerator and denominator of the left-hand side of (9.12). This is not the case in expressing (9.7) in form (9.12).

In their Remark 4.1, Cho and Meyer [11], p.148, point out, in order to obtain a fair comparison between the bounding tightness of $\kappa_6 = \tau_1(\mathbf{A}^\sharp)$ and $\kappa_3 \equiv \kappa_8 = \mathcal{T}(\mathbf{A}^\sharp)$ that $(\bar{\pi}^T - \pi^T)\mathbf{1} = 0$, so $\| \bar{\pi}^T - \pi^T \|_\infty \leq (1/2)\| \bar{\pi}^T - \pi \|_1$. Hence

$$\kappa_3 = \mathcal{T}(\mathbf{A}^\sharp) \leq \frac{1}{2}\tau_1(\mathbf{A}^\sharp) = \frac{1}{2}\kappa_6,$$

from which they conclude that κ_3 is the tighter condition number.

However, one might argue that from (9.9) which underlies (9.7), that

$$\sum_j |\bar{\pi}_j - \pi_j| \leq \frac{1}{2}\sum_j \left(\max_{h,h'} |c_{hj} - c_{h'j}| \right) \max_k \| \mathbf{E}_k^T \|_1,$$

and since from (9.4)

$$\tau_1(\mathbf{C}) \leq \Delta(\mathbf{C}) \equiv \frac{1}{2}\sum_j \left(\max_{h,h'} |c_{hj} - c_{h'j}| \right),$$

defining $\Delta(\mathbf{C})$ in analogy to (8.2), it follows that the consequent bound on $\| \bar{\pi}^T - \pi \|_1$ is not as tight as when using $\tau_1(\mathbf{A}^\sharp)$.

In their role as condition numbers, $\tau_1(\mathbf{A}^\sharp)$ and $\mathcal{T}(\mathbf{A}^\sharp)$ are not really directly comparable in regard to size, since different versions of the norm $\| \bar{\pi}^T - \pi \|$ are involved.

9.3. Recent related results. The discussion of Sections 9.1 and 9.2 has revolved around (9.5). Hunter [19] Theorems 2.1 - 2.2 has obtained this equality by using the general form $\mathbf{G}$ of the g-inverse of $\mathbf{I} - \mathbf{P}$:

$$(9.13) \qquad \mathbf{G} = [\mathbf{I} - \mathbf{P} + \mathbf{t}\mathbf{u}^T]^{-1} - \mathbf{1}\mathbf{v}^T + \mathbf{g}\pi^T$$

for any real $\mathbf{t}, \mathbf{v}, \mathbf{u}, \mathbf{g}$ satisfying $\pi^T\mathbf{t} \neq 0$, $\mathbf{u}^T\mathbf{1} \neq 0$, by showing

$$(9.14) \qquad \bar{\pi}^T - \pi^T = \bar{\pi}^T\mathbf{E}\mathbf{G}(\mathbf{I} - \mathbf{1}\pi^T)$$

and that $\mathbf{E}\mathbf{C}(\mathbf{u}, \mathbf{v})$ is a special case of $\mathbf{E}\mathbf{G}(\mathbf{I} - \mathbf{1}\pi^T)$.

In fact, we can see immediately from (9.13) that $\mathbf{C}(\mathbf{u}, \mathbf{v})$ itself is a special case of the g-inverse G .

Equation (9.14) leads to a bound of general appearance :

$$(9.15) \qquad \parallel \overline{\pi}^T - \pi^T \parallel_1 \leq \tau_1(\mathbf{G}(\mathbf{I} - \mathbf{1}\pi^T)) \parallel \mathbf{E} \parallel_1.$$

It would seem plausible that $\tau_1(\mathbf{G}(\mathbf{I} - \mathbf{1}\pi^T))$ may give a tighter bound, for some parameter vectors, than $\tau_1(\mathbf{A}^\sharp) = \tau_1(\mathbf{Z})$, but the author has shown that in fact all these values are the same, namely $\tau_1(\mathbf{A}^\sharp)$.

Hunter [20], Corollary 5.1.1, derives the bound

$$(9.16) \qquad \parallel \overline{\pi}^T - \overline{\pi}^T \parallel_1 \leq \operatorname{tr}(\mathbf{A}^\sharp)\parallel \mathbf{E} \parallel_1.$$

This bound is not as strict as (9.6), with $\mathbf{C} = \tau_1(\mathbf{A}^\sharp)$, since [56], p.165, (10), states that

$$(9.17) \qquad \tau_1(\mathbf{A}^\sharp) \leq \operatorname{tr}(\mathbf{A}^\sharp).$$

Work on these issues by J. Hunter and the author is in progress. For the time being, the bound [53]:

$$(9.18) \qquad \parallel \overline{\pi}^T - \pi^T \parallel_1 \leq \tau_1(\mathbf{A}^\sharp)\parallel \mathbf{E} \parallel_1$$

remains sharp for the norm used.

REFERENCES

[1] YU. A. ALPIN AND N.Z. GABASSOV, *A remark on the problem of localization of the eigenvalues of real matrices (in Russian)*, Izv. Vyssh. Uchebn. Zaved. Matematika, 11(1976), pp. 98–100.

[2] G.P. BASHARIN, A.N. LANGVILLE AND V.A. NAUMOV, *The life and work of A.A. Markov*, Linear Algebra Appl., 386(2004), pp. 3–26.[This was in the Special Issue, pp. 1- 408, on the *Conference on the Numerical Solution of Markov Chains*, 2003.]

[3] J. BERNOULLI, *O Zakone Bol'shikh Chisel (On the Law of Large Numbers)*. Nauka, Moscow, 1986.

[4] S. N. BERNSTEIN, *Sobranie Sochinenii: Tom IV [Collected Works, Vol. IV], Teoriia Veroiatnostei i Matematicheskaia Statistika [1911–1946]*, Nauka, Moscow, 1964.

[5] A.S. BEZIKOVICH [BESICOVITCH], *A. A. Markov's contributions to probability theory (in Russian)*, Izvestiia Rossiiskoi Akademii Nauk, 17(1923), pp. 45–52.

[6] A.S. BEZIKOVICH [BESICOVITCH], *Biograficheskii Ocherk. [Biographical Sketch]*, [36], pp.III – XIV.

[7] H. BRUNS, *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*, Teubner, Leipzig, 1906. [See esp. Achtzehnte Vorlesung].

[8] J.C. BURKILL, *Abram Samoilovitch Besicovitch. 1891-1970*, Biographical Memoirs of Fellows of the Royal Society, 17(1971), pp. 1–16.

[9] S. CHATTERJEE AND E. SENETA, *Towards consensus: some convergence theorems on repeated averaging*, J. Appl. Prob., 14(1977), pp. 89–97.

[10] G.E. CHO AND C.D. MEYER, *Markov chain sensitivity measured by mean first passage time*, Linear Algebra Appl., 316(2000), pp. 21–28.

[11] G.E. CHO AND C.D. MEYER, *Comparison of perturbation bounds for the stationary distribution of a Markov chain*, Linear Algebra Appl., 335(2001), pp. 137–150.

[12] R.L. DOBRUSHIN, *Central limit theorem for non-stationary Markov chains, I and II*, Theory Prob. Appl., 1(1956), pp. 65–80, 329–83 [English translation].

[13] M. Fréchet, *Recherches théoriques modernes sur le calcul des probabilités. Secnd livre. Méthode des fonctions arbitraires. Théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles.* Gauthier-Villars, Paris, 1938.

[14] G. Frobenius, *Über Matrizen aus positiven Elementen*, S.-B. Preuss. Akad. Wiss.,(Berlin), (1908), pp. 471–476; (1909), pp. 514–518.

[15] G. Frobenius, *Über Matrizen aus nicht negativen Elementen*, S.-B. Preuss. Akad. Wiss.,(Berlin), (1912), pp. 456–477.

[16] S. Ia. Grodzensky, *Andrei Andreevich Markov 1856–1922.* Nauka, Moscow, 1987.

[17] M. Haviv and L. Van der Heyden, *Perturbation bounds for the stationary probabilities of a finite Markov chain*, Adv. Appl. Prob., 16(1984), pp. 804–818.

[18] C.C. Heyde and E. Seneta, eds., *Statisticians of the Centuries*, Springer, New York, 2001.

[19] J.J. Hunter, *Stationary distributions and mean first passage times of perturbed Markov chains*, Linear Algebra Appl., 410(2005), pp. 217–243.

[20] J.J. Hunter, *Mixing times with applications to perturbed Markov chains*, LInear Algebra Appl., (2006) to appear.

[21] M. Iosifescu, *Finite Markov Processes and Their Applications*, Wiley, Chichester; Editura Tehnica, Bucharest, 1980.

[22] D.L. Isaacson and R.W. Madsen, *Markov Chains*, Wiley, New York, 1976.

[23] J.G. Kemeny and J.L. Snell, *Finite Markov Chain*, Van Nostrand, Princeton N.J., 1960.

[24] S.J. Kirkland. M. Neumann and B.L. Shader, *Applications of Paz's inequality to perturbation bounds for Markov chains*, Linear Algebra Appl., 268(1998), pp. 183–196.

[25] A.N. Kolmogorov and A. P. Yushkevich, eds., *Mathematics in the 19th Century; Mathematical Logic, Algebra, Number Theory, Probability Theory*, Birkhäuser, Basel, Boston, 1992.

[26] A.N. Langville and C.D. Meyer, *The use of linear algebra by web search engines*, IMAGE, 33(2004), pp. 2–5.

[27] A.A. Markov, *Extension of the law of large numbers to dependent quantities (in Russian)*, Izvestiia Fiz.-Matem. Obsch. Kazan Univ., (2nd Ser.), 15(1906), pp. 135–156 [Also [37], pp. 339–61].

[28] A.A. Markov, *The extension of the law of large numbers onto quantities depending on each other*, [62], pp. 143–158. Translation into English of [27].

[29] A.A. Markov, *Investigation of a notable instance of dependent trials (in Russian)*, Izvestiia Akad. Nauk (St. Petersburg), (6th Ser.), 1(3)(1907), pp.61-80. [Also in [37], pp. 363–397.]

[30] A.A. Markov, *Extension of limit theorems of the calculus of probabilities to sums of quantities associated into a chain(in Russian)*, Zapiski Akad. Nauk (St. Petersburg), Fiz.-Mat. Otd., (7th Ser.). 22(9)(1908), 29 pp. [Also [37], pp. 363–397.]

[31] A.A. Markov, *The extension of the limit theorems of the calculus of probability onto a sum of magnitudes connected into a chain*, [62], pp. 159–178. Translation into English of [30].

[32] A.A. Markov, *Investigation of the general case of trials associated into a chain [in Russian]*, Zapiski. Akad. Nauk (St. Petersburg) Fiz.-Matem. Otd., (8th Ser.) 25(3)(1910), 33 pp. [Also in [37]. pp. 465–507.]

[33] A.A. Markov, *An investigation of the general case of trials connected into a chain*, [62], pp. 181–203. Translation into English of [32] .

[34] A.A. Markov, *On dependent trials not forming a true chain (in Russian)*, Izvestiia Akad. Nauk (St. Petersburg), (6th Ser.), 5(2)(1911), pp.113—126. [Also in [37], pp. 399–416.]

[35] A.A. Markov, *On a problem of Laplace (in Russian)*, Izvestiia Akad. Nauk (St. Petersburg), (6th Ser.), 9(2)(1915), pp. 87–104. [Also in [37], pp. 549–571.]

[36] A.A. Markov, *Ischislenie Veroiatnostei (Calculus of Probabilities)*, Fourth (posthumous) edn., Gosudarstvennoie Izdatelstvo, Moscow, 1924.

[37] A.A. Markov, *Izbrannie Trudy (Selected Works)*, A.N.S.S.S.R., Leningrad, 1951.

[38] A.A. Markov [Jr.], *The Biography of A. A. Markov.* In [62], pp. 242-256. Translation into English of [37], pp. 599–613.

[39] C.D. Meyer, *The role of the generalized inverse in the theory of finite Markov chains*, SIAM Rev., 17(1975), pp. 443–464.

[40] R. von Mises, *Wahscheinlichkeitsrechnung*, Fr. Deuticke, Leipzig and Vienna,1931.

[41] Kh.O. Ondar (Ed.), *The Correspondence Between A.A. Markov and A.A. Chuprov on the Theory of Probability and Mathematical Statistics*, [Translated by C. & M. Stein], Springer, New York. 1981.

[42] A. Paz, *Introduction to Probabilistic Automata*, Academic Press, New York,1971.

[43] U. PAŬN, *A class of ergodicity coefficients, and applications*, Math. Reports (Bucharest), 4(2002), pp. 225–232.

[44] O. PERRON, *Zur Theorie der Matrices*, Math. Ann., 64(1907), pp. 248–263.

[45] V.I. ROMANOVSKY, *Diskretnie Tsepi Markova*, Gostekhizdat, Moscow, 1949. [English translation by E. SENETA, *Discrete Markov Chains*, Wolters-Noordhoff, Groningen, 1970].

[46] H. SCHNEIDER, *The concepts of irreducuibility and full indecomposability of a matrix in the works of Frobenius, König and Markov*, Linear Algebra Appl., 18(1977), pp. 139–162.

[47] P.J. SCHWEITZER, *Perturbation theory and finite Markov chains*, J. Appl. Prob., 5(1968), pp. 401–413.

[48] E. SENETA, *Coefficients of ergodicity: structure and applications.* Adv. Appl. Prob., 11(1979), pp.576–590.

[49] E. SENETA, *Non-negative Matrices and Markov Chains*, Second edn., Springer, New York, 1981, 2006(revised printing).

[50] E. SENETA, *The central limit problem and linear least squares in pre-revolutionary Russia: The background*, Mathematical Scientist, 9(1984), pp. 37–77.

[51] E. SENETA, *Explicit forms for ergodicity coefficients and spectrum localization*, Linear Algebra. Appl., 60(1984), pp. 187–197.

[52] E. SENETA, *Perturbation of the stationary distribution measured by ergodicity coefficients*, Adv. Appl. Prob., 20(1988), pp. 228–230.

[53] E. SENETA, *Sensitivity analysis, ergodicity coefficients, and rank-one updates for finite Markov chains.* In [63], pp. 121–129.

[54] E. SENETA, *Applications of ergodicity coefficients to homogeneous Markov chains*, Contemporary Mathematics, 149(1993), pp. 189–199.

[55] E. SENETA, *Ergodicity for products of infinite stochastic matrices*, J. Theoret. Probab., 6(1993), pp. 345–362.

[56] E. SENETA, *Sensitivity of finite Markov chains under perturbation,* Statist. Probab. Lett., 17(1993), pp. 163 – 168.

[57] E. SENETA, *Markov and the birth of chain dependence theory*, International Statistical Review, 64(1996), pp. 255–263.

[58] E. SENETA, *Complementation in stochastic matrices and the GTH algorithm*, SIAM J. Matrix Anal. Appl., 19(1998), pp. 556–563.

[59] E. SENETA, *Andrei Andreevich Markov*, In [18], pp.243–247.

[60] E. SENETA, *Statistical regularity and free will: L.A.J. Quetelet and P.A. Nekrasov.* International Statistical Review, 71(2003), pp. 319–334.

[61] O.B.SHEYNIN, *A.A. Markov's work on probability.* Archive Hist. Exact Sci., 39(1989), pp. 337–77.

[62] O.B.SHEYNIN, ED., *Probability and Statistics. Russian Papers.*, Selected and Translated by Oscar Sheynin. NG Verlag, Berlin, 2004.

[63] W.J. STEWART, ED., *Numerical Solution of Markov Chains*, Marcel Dekker, New York, 1991.

[64] C.P. TAN. *Coefficients of ergodicity with respect to vector norms.* J. Appl. Prob., 20(1983), pp. 277–287.

[65] J.V. USPENSKY, *A sketch of the scientific work of A. A. Markov (in Russian)*, Izvestiia Rossiskoi Akademii Nauk, 17(1923), pp. 19–34.

[66] J.V. USPENSKY, *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937.