A probabilistic proof of Perron's theorem

Raphaël Cerf
DMA, École Normale Supérieure

Joseba Dalmau
CMAP, Ecole Polytechnique

January 10, 2018

Abstract

We present an alternative proof of Perron's theorem, which is probabilistic in nature. It rests on the representation of the Perron eigenvector as a functional of the trajectory of an auxiliary Markov chain.

In 1907, Oskar Perron proved the following theorem.

Theorem 1 Let A be a square matrix with posivite entries. Then the matrix A admits a positive eigenvalue λ such that:
i) to λ is associated an eigenvector μ whose components are all positive;
ii) if α is another eigenvalue of A, possibly complex, then $|\alpha| < \lambda$;
iii) any other eigenvector associated to λ is a multiple of μ.

This theorem was subsequently generalized by Frobenius in his work on non–negative matrices in 1912, leading to the so–called Perron–Frobenius theorem [5]. A myriad of mathematical models involve non–negative matrices and their powers, thereby calling for the use of the Perron–Frobenius theorem. Mathematicians have developped generalizations in several directions, notably in infinite dimensions (for infinite matrices [6], for non–negative kernels in arbitrary spaces [1]) and a whole Perron–Frobenius theory has emerged. Hawkins wrote an historical account on the initial developpement of this theory [3]. MacCluer [4] describes several applications of Perron's theorem and review the different proofs that have been found over the years. The original proof of Perron rested on an induction over the size the matrix. A few years later Perron found a proof involving the resolvent of the matrix. A nowadays popular proof, which is found in most textbooks, is due to Wielandt and it rests on a miraculous max–min functional.

We present here an alternative proof of Perron's theorem, which is probabilistic in nature. It rests on an auxiliary Markov chain, and the representation of the Perron eigenvector as a functional of the trajectory of this Markov chain [2]. This formula generalizes the well–known formula for the invariant probability measure of a finite state Markov chain. To ease the exposition, we restrict ourselves to the Perron theorem, and we work with matrices whose entries are all positive. However our proof can be readily extended to primitive matrices, thereby yielding the classical Perron–Frobenius theorem. Our proof might seem lengthy compared to other proofs, yet it is completely self–contained and it requires only classical results of basic algebra and power series.

We introduce next some notation in order to define the auxiliary Markov chain. Let d be a positive integer. Throughout the text, we consider a square matrix $A = (A(i,j))_{1\leq i,j\leq d}$ of size $d \times d$ with positive entries. For $i \in \{\,1,\dots,d\,\}$, we denote by $S(i)$ the sum of the entries on the i–th row of A, i.e.,

$$\forall i \in \{\,1,\dots,d\,\} \qquad S(i) \,=\, \sum_{j=1}^{d} A(i,j)\,,$$

and we create a new matrix $M = (M(i,j))_{1\leq i,j\leq d}$ by setting

$$\forall i,j \in \{\,1,\dots,d\,\} \qquad M(i,j) \,=\, \frac{A(i,j)}{S(i)}\,.$$

Obviously, the sum of each row of M is now equal to one, i.e., M is stochastic, and we think of it as the transition matrix of a Markov chain. So, let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with state space $\{\,1,\dots,d\,\}$ and transition matrix M. Let us fix $i \in \{\,1,\dots,d\,\}$. We denote by E_i the expectation of the Markov chain issued from i and we introduce the time τ_i of the first return of the chain to i, defined by

$$\tau_i \,=\, \inf\,\big\{\, n \geq 1 : X_n = i \,\big\}\,.$$

Finally, we define a function ϕ_i by setting

$$\forall \lambda \geq 0 \qquad \phi_i(\lambda) \,=\, E_i\left(\lambda^{-\tau_i} \prod_{n=0}^{\tau_i - 1} S(X_n)\right).$$

The quantity in the expectation is non–negative, so the function ϕ_i is well defined and it might take infinite values.

Proposition 2 The function ϕ_i is continuous, decreasing on $\mathbb{R}^+$ and

$$\lim_{\lambda\to 0} \phi_i(\lambda) \,=\, +\infty\,, \qquad \lim_{\lambda\to +\infty} \phi_i(\lambda) \,=\, 0\,.$$

Proof. In fact, the function ϕ_i can be written as a power series in the variable $1/\lambda$, as follows:

$$
\begin{aligned}
\phi_i(\lambda) \;=\; \sum_{k=1}^{\infty} \frac{1}{\lambda^k} E_i\Bigg(\Big(1_{\{\tau_i=k\}} \prod_{n=0}^{k-1} S(X_n)\Big)\Bigg) \;=\; \\
\sum_{k=1}^{\infty} \frac{1}{\lambda^k} \sum_{i_1,\dots,i_{k-1}\neq i} S(i)S(i_1)\cdots S(i_{k-1}) P\big(X_1 = i_1, \dots, X_{k-1} = i_{k-1}, X_k = i\big) \\
= \sum_{k=1}^{\infty} \frac{1}{\lambda^k} \sum_{i_1,\dots,i_{k-1}\neq i} S(i)M(i,i_1)\cdots S(i_{k-1})M(i_{k-1},i) \\
= \sum_{k=1}^{\infty} \frac{1}{\lambda^k} \sum_{i_1,\dots,i_{k-1}\neq i} A(i,i_1)\cdots A(i_{k-1},i)\,.
\end{aligned}
$$

Since A has positive entries, the series contains non vanishing terms, and this implies that ϕ_i is decreasing and tends to ∞ as λ goes to 0. Let R be the radius of the convergence circle of this series. From classical results on powers series, we know that $\phi_i(\lambda)$ is continuous for $\lambda > R$. To prove that ϕ_i is continuous, we have to show that $\phi_i(R) = +\infty$. Let B be the matrix obtained from A by removing the i–th row and the i–th column and let $\gamma_1, \dots, \gamma_{d-1}$ be its eigenvalues (possibly complex), arranged so that $|\gamma_1| \geq \cdots \geq |\gamma_d|$. Let m (respectively M) be the minimum (respectively the maximum) of the entries of A. For any $k \geq 1$, we have

$$
\begin{aligned}
\sum_{i_1,\dots,i_{k-1}\neq i} A(i,i_1)\cdots A(i_{k-1},i) \;\geq\; \frac{m^2}{M} \sum_{i_1,\dots,i_{k-1}\neq i} A(i_1,i_2)\cdots A(i_{k-1},i_1) \\
= \frac{m^2}{M}\mathrm{trace}(B^k) \;=\; \frac{m^2}{M}\Big(\gamma_1^k + \cdots + \gamma_{d-1}^k\Big)\,.
\end{aligned}
$$

Although the eigenvalues $\gamma_1, \dots, \gamma_{d-1}$ might be complex numbers, the trace of B^k is a positive real number. We can also a prove a similar inequality in the reverse direction, and we conclude that the power series defining ϕ_i converges if and only if the series

$$
\sum_{k=1}^{\infty} \frac{1}{\lambda^k}\Big(\gamma_1^k + \cdots + \gamma_{d-1}^k\Big)
$$

converges. This is certainly the case if $|\lambda| > |\gamma_1|$, therefore $R \leq |\gamma_1|$. Let us define, for $n \geq 1$,

$$
S_n(\lambda) \;=\; \sum_{k=1}^{n} \frac{1}{\lambda^k}\Big(\gamma_1^k + \cdots + \gamma_{d-1}^k\Big)\,.
$$

We shall rely on the following result on geometric series.

Lemma 3 Let z be a complex number such that $|z| \leq 1$. Then

$$\lim_{n\to\infty} \frac{1}{n}\big(z+\cdots+z^n\big) \;=\; \begin{cases} \;0 & \text{if} \quad z\neq 1\,,\\ \;1 & \text{if} \quad z=1\,.\end{cases}$$

Proof. For $z=1$, the result is obvious. For $z\neq 1$, we compute

$$\frac{1}{n}\big(z+\cdots+z^n\big) \;=\; \frac{z-z^{n+1}}{n(1-z)}\,,$$

and we observe that this quantity goes to 0 when n goes to ∞. □

Lemma 3 implies that, for λ a complex number such that $|\lambda| = |\gamma_1|$,

$$\lim_{n\to\infty} \frac{1}{n} S_n(\lambda) \;=\; \mathrm{card}\,\big\{\, j : 1\leq j\leq d,\ \lambda=\gamma_j\,\big\}\,.$$

This implies in particular that $\big|S_n(\gamma_1)\big|$ goes to ∞ with n. Observing that $\big|S_n(\gamma_1)\big| \leq S_n(|\gamma_1|)$, we conclude that

$$\phi_i(|\gamma_1|) \;=\; \lim_{n\to\infty} S_n(|\gamma_1|) \;=\; +\infty\,.$$

Therefore $R=|\gamma_1|$ and moreover $\phi_i(R)=+\infty$. □

Proposition 2 implies that ϕ_i is one to one from $]R,+\infty[$ onto $]0,+\infty[$, thus there exists a unique positive real number λ_i such that $\phi_i(\lambda_i)=1$. The next result is the key to our proof of the Perron–Frobenius theorem. We define a vector μ_i by setting

$$\forall j\in\{\,1,\dots,d\,\} \qquad \mu_i(j) \;=\; E_i\Bigg(\sum_{n=0}^{\tau_i-1}\Big(1_{\{X_n=j\}}\lambda_i^{-n}\prod_{k=0}^{n-1}S(X_k)\Big)\Bigg)\,.$$

Theorem 4 The value λ_i is an eigenvalue of A and the vector μ_i is an associated left eigenvector whose components are all positive and finite.

Proof. Let us note $E_i,\tau_i,\lambda_i,\mu_i$ simply by E,τ,λ,μ. Let us compute

$$\begin{aligned}\sum_{j=1}^{d}\mu(j)A(j,k) \;&=\; \sum_{j=1}^{d}\mu(j)S(j)M(j,k)\\ &= \sum_{j=1}^{d}\sum_{n\geq 0}E\Bigg(1_{\{\tau>n\}}\lambda^{-n}\Big(\prod_{t=0}^{n-1}S(X_t)\Big)1_{\{X_n=j\}}f(j)M(j,k)\Bigg)\\ &= \sum_{j=1}^{d}\sum_{n\geq 0}E\Bigg(1_{\{\tau>n\}}\lambda^{-n}\Big(\prod_{t=0}^{n}S(X_t)\Big)1_{\{X_n=j\}}1_{\{X_{n+1}=k\}}\Bigg)\end{aligned}$$

$$
\begin{aligned}
&= E\left(\sum_{n=0}^{\tau-1} 1_{\{X_{n+1}=k\}}\lambda^{-n}\Big(\prod_{t=0}^{n} S(X_t)\Big)\right)\\
&= \lambda\, E\left(\sum_{n=1}^{\tau} 1_{\{X_n=k\}}\lambda^{-n}\Big(\prod_{t=0}^{n-1} S(X_t)\Big)\right).
\end{aligned}
$$

Suppose that $k \neq i$. Then the term in the last sum vanishes for $n = 0$ or $n = \tau$, and we obtain

$$
\sum_{j=1}^{d} \mu(j)A(j,k) \;=\; \lambda\mu(k)\,.
$$

For $k = i$, we obtain, noticing that $\mu(i) = 1$,

$$
\sum_{j=1}^{d} \mu(j)A(j,i) \;=\; \lambda\, E\left(\lambda^{-\tau}\prod_{t=0}^{\tau-1} S(X_t)\right) \;=\; \lambda\,\phi_i(\lambda) \;=\; \lambda\,\mu(i)\,.
$$

Thus we have proved that $\mu A = \lambda\mu$. Since $\mu(i) = 1$, these equations imply that $\mu(1),\dots,\mu(d)$ are all positive and finite. $\square$

Proposition 5 Let α be an eigenvalue of A, possibly complex, and let ν be an associated left eigenvector. Let $i \in \{\,1,\dots,d\,\}$ be such that $\nu(i) \neq 0$. Either ν and μ_i are proportional (in which case $\alpha = \lambda_i$) or $|\alpha| < \lambda_i$.

Proof. Let α, ν and i be as in the statement of the proposition. We suppose that $\alpha \neq 0$, otherwise there is nothing to prove. Let ν be an associated left eigenvector. We have

$$
\forall k \in \{\,1,\dots,d\,\} \qquad \nu(k) \;=\; \frac{1}{\alpha}\sum_{j=1}^{d} \nu(j)A(j,k)\,.
$$

Let us focus on the equation for $k = i$. We divide by $\nu(i)$ (which is assumed to be non zero) and we isolate the term $j = i$ in the sum to obtain

$$
1 \;=\; \frac{1}{\alpha}A(i,i) + \frac{1}{\alpha}\sum_{j\neq i}\frac{\nu(j)}{\nu(i)}A(j,i)\,.
$$

We expand $\nu(j)$ in the above equation as a sum, and we get

$$
1 \;=\; \frac{1}{\alpha}A(i,i) + \frac{1}{\alpha^2}\sum_{j\neq i}\sum_{j'}\frac{\nu(j')}{\nu(i)}A(j',j)A(j,i)
$$

$$= \frac{1}{\alpha}A(i,i) + \frac{1}{\alpha^2}\sum_{j\neq i} A(i,j)A(j,i) + \frac{1}{\alpha^2}\sum_{j\neq i}\sum_{j'\neq i}\frac{\nu(j')}{\nu(i)}A(j',j)A(j,i)\,.$$

Iterating n times this procedure, we get

$$1 \;=\; \frac{1}{\alpha}A(i,i) + \cdots + \frac{1}{\alpha^{n+1}}\sum_{i_1,\dots,i_n\neq i} A(i,i_1)A(i_1,i_2)\cdots A(i_n,i)$$
$$+ \frac{1}{\alpha^{n+1}}\sum_{i_0,i_1,\dots,i_n\neq i}\frac{\nu(i_0)}{\nu(i)}A(i_0,i_1)A(i_1,i_2)\cdots A(i_n,i)\,.$$

If $\phi_i(|\alpha|) = +\infty$, then it follows from proposition 2 and the defintion of λ_i that $|\alpha| < \lambda_i$ and we are done. From now onwards, we suppose that $\phi_i(|\alpha|) < +\infty$. In the proof of proposition 2, we worked out a power series expansion of ϕ_i. The convergence of this series at $|\alpha|$ implies in particular that the general term of this series goes to 0, hence

$$\lim_{n\to\infty}\frac{1}{\alpha^{n+1}}\sum_{i_1,\dots,i_n\neq i} A(i,i_1)A(i_1,i_2)\cdots A(i_n,i) \;=\; 0\,.$$

Let m (respectively M) be the minimum (respectively the maximum) of the entries of A. For any $i_0 \neq i$, we have

$$\sum_{i_1,\dots,i_n\neq i} A(i_0,i_1)A(i_1,i_2)\cdots A(i_n,i) \;\leq$$
$$\frac{M}{m}\sum_{i_1,\dots,i_n\neq i} A(i,i_1)A(i_1,i_2)\cdots A(i_n,i)\,.$$

It follows that, for any $n \geq 1$,

$$\frac{1}{|\alpha|^{n+1}}\sum_{i_0,i_1,\dots,i_n\neq i}\frac{\nu(i_0)}{\nu(i)}A(i_0,i_1)A(i_1,i_2)\cdots A(i_n,i) \;\leq$$
$$\frac{Md\max_{1\leq j\leq d}|\nu(j)|}{m|\nu(i)|}\frac{1}{|\alpha|^{n+1}}\sum_{i_1,\dots,i_n\neq i} A(i,i_1)A(i_1,i_2)\cdots A(i_n,i)$$

and we conclude from the previous inequality that

$$\lim_{n\to\infty}\frac{1}{\alpha^{n+1}}\sum_{i_0,i_1,\dots,i_n\neq i}\frac{\nu(i_0)}{\nu(i)}A(i_0,i_1)A(i_1,i_2)\cdots A(i_n,i) \;=\; 0\,.$$

We send now n to ∞ in the identity and we get

$$1 \;=\; \frac{1}{\alpha}A(i,i) + \sum_{n=1}^{+\infty}\frac{1}{\alpha^{n+1}}\sum_{i_1,\dots,i_n\neq i} A(i,i_1)A(i_1,i_2)\cdots A(i_n,i)\,.$$

Recall that α might be complex. Taking the modulus, we conclude that $\phi_i(|\alpha|) \geq 1$, and since ϕ_i is decreasing, then $|\alpha| \leq \lambda_i$. It remains to examine the case $|\alpha| = \lambda_i$. We suppose that the eigenvector ν associated to α is normalized so that $\nu(i) = 1$. We denote by $|\nu|$ the vector whose coordinates are the modulus of the coordinates of ν, i.e., $|\nu|(j) = |\nu(j)|$ for $1 \leq j \leq d$. Since $\nu A = \alpha\nu$ and the entries of A are positive, then

$$\forall k \in \{\, 1, \dots, d \,\} \qquad |\nu|(k) \;\leq\; \frac{1}{\lambda} \sum_{j=1}^{d} |\nu|(j) A(j,k)\,.$$

Starting from this inequality, we proceed as previously, that is, we isolate the term corresponding to $j = i$ in the sum, we bound from above the term $|\nu(j)|$ for $j \neq i$, we iterate the procedure n times. We check that the ultimate term goes to 0 when we send n to ∞, as we get the inequality

$$\forall k \in \{\, 1, \dots, d \,\} \qquad |\nu|(k) \;\leq\; \mu_i(k)\,.$$

For $k \in \{\, 1, \dots, d \,\}$, we have

$$\lambda |\nu|(k) \;=\; \Big| \sum_{j=1}^{d} \nu(j) A(j,k) \Big| \;\leq\; \sum_{j=1}^{d} |\nu|(j) A(j,k)\,.$$

It follows that

$$\sum_{k=1}^{d} \big(\mu_i(k) - |\nu|(k)\big) A(k,i) \;\leq\; \lambda \left(\mu(i) - \nu(i)\right) \;=\; 0\,.$$

This equation implies that $\mu_i = |\nu|$ and that all the intermediate inequalities were in fact equalities. Since all the entries of A are positive and $\nu(i) = 1$, then necessarily all the components of ν are non–negative real numbers and $\nu = \mu_i$ and $\alpha = \lambda$. $\square$

The λ_i's are positive eigenvalues of A, the eigenvectors μ_i have positive coordinates, thus proposition 5 readily implies the following result.

Corollary 6 The values $\lambda_1, \dots, \lambda_d$ are all equal. Their common value λ is a simple eigenvalue of A. The eigenvectors $\mu_1, \dots, \mu_d$ are proportional.

Finally, we normalize these eigenvectors by imposing that the sum of the components is equal to 1, thereby getting a probability distribution.

Corollary 7 The left Perron–Frobenius eigenvector μ of A is given by

$$\forall i \in \{\, 1, \dots, d \,\} \qquad \mu(i) \;=\; \frac{1}{E_i\left(\displaystyle\sum_{n=0}^{\tau_i - 1} \left(\lambda^{-n} \prod_{t=0}^{n-1} S(X_t) \right) \right)}\,.$$

This formula (already proved in [2]) is a generalization of the classical formula for the invariant probability measure of a Markov chain. Indeed, in the particular case where A is stochastic, S is constant equal to 1, λ is also equal to 1, the formula of the corollary becomes the well–known formula

$$\forall i \in \{\,1,\dots,d\,\} \qquad \mu(i) \,=\, \frac{1}{E_i(\tau_i)}\,.$$

References

[1] Krishna B. Athreya and Peter Ney. A renewal approach to the Perron-Frobenius theory of nonnegative kernels on general state spaces. *Math. Z.*, 179(4):507–529, 1982.

[2] Raphaël Cerf and Joseba Dalmau. A markov chain representation of the normalized perronfrobenius eigenvector. *Electron. Commun. Probab.*, 22:6 pp., 2017.

[3] Thomas Hawkins. Continued fractions and the origins of the Perron-Frobenius theorem. *Arch. Hist. Exact Sci.*, 62(6):655–717, 2008.

[4] C. R. MacCluer. The many proofs and applications of Perron's theorem. *SIAM Rev.*, 42(3):487–498, 2000.

[5] E. Seneta. *Non-negative matrices and Markov chains*. Springer Series in Statistics. Springer, New York, 2006. Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544].

[6] D. Vere-Jones. Ergodic properties of nonnegative matrices. I. *Pacific J. Math.*, 22:361–386, 1967.