Probabilistic Approach to Perron Root, the Group Inverse, and Applications

Iddo Ben-Ari[*]
iddo.ben-ari@uconn.edu

Michael Neumann

Department of Mathematics
University of Connecticut
196 Auditorium Rd.
Storrs, CT 06269-3009

Abstract

A probabilistic approach to the study of the Perron root of irreducible nonnegative matrices is presented. Our two main results are a probabilistic representation for the generalized inverse of the generator of a continuous–time finite–state Markov chain and the identification of the Hessian of the Perron root of a nonnegative irreducible matrix, with respect to a certain natural transformation of its entries, as a covariance matrix for additive functionals for a related Markov chain. These provide us with a natural approach – at least from the point of view of the probabilist – to study positivity and convexity properties of perturbations to the Perron eigenvalue. We focus on reestablishing and improving a number of known results in the field.

Keywords: additive functional, central limit theorem, group inverses, Markov chains, nonnegative matrices, Perron roots.

AMS Subject Classification: 15A48, 15A09, 60G44, 60J22, 65C40.

1 Introduction

The Perron root of nonnegative and irreducible matrices is one of the fundamentally important objects of the theory of nonnegative matrices. This is mostly due to the fact that the Perron root determines the leading term in large time asymptotic behavior of linear evolution systems.

In this paper we study the dependence of the Perron root on the entries of the matrix. Our approach is probabilistic and is obtained by relating the matrices in question to a corresponding Markov chain. We focus on aspects where the probabilistic approach is most natural. Our first main result, Theorem 1, and its discrete–time version, Theorem 2, provides a probabilistic formula for the generalized inverse of of the generator of a continuous–time finite–state Markov chain and for the generalized inverse for the transition matrix of a discrete–time chain minus the identity, respectively. In each case, the formula differs from the ones known to us and which are usually used

---

[*] I.B. would like to dedicate this paper to the memory of Michael Neumann, who passed away suddenly on April 21, 2011.

by practitioners, because it contains a term involving a second moment of a random variable. The existence of this term yields almost automatically a number of positivity results and inequalities.

Our second main result, Theorem 4, identifies the Hessian matrix of the Perron root for the generator of a Markov chain, with respect to a certain transformation of the entries of the generator, as a covariance matrix for additive functionals of the process, which is also the covariance matrix in the central limit theorem for the joint distribution of the additive functionals of the chain, as we show in Theorem 5. Since covariance matrices are positive semidefinite, this result leads directly to convexity results. Since the Hessian has a representation in terms of the group inverse of the generator due to Deutsch and Neumann [10, Theorem 3.2] (appearing here as Theorem 3), this result also provides an analytical expression for the covariance in the central limit theorem. The result has applications in probability theory when a tractable analytical expression for the covariance is needed. As a first example, this relation has been recently applied by Ben Ari, Boushaba, Matzavinos, and Roitershtein [2], who studied stochastic model for the motion of bi–peds.

Part of our motivation for examining even known results on the Perron root and vector and on Markov chains is that there is a body of literature, particularly on the latter, which was created by researchers in numerical and applied linear algebra to facilitate computations of quantities such as the stationary probability distribution or the mean first passage matrices, especially for large scale problems. In the course of developing such methods, results have been proven which may have interest to general research on Markov chains, but have been left without a satisfactory probabilistic interpretation of these results. This is an aspect which we have attempted to illuminate in this paper.

Our results and proofs are given in the following sections where, as a general rule, we have left the proofs to the end of each section. In Section 2 we present notation and conventions and recall some basic facts from the theory of continuous–time Markov chains to be used later on. In Section 3 we derive probabilistic formulas for the group inverse. In Section 4 we link the group inverse with perturbations to the Perron root. In the last section, Section 5, we present several applications to the results and methods of the preceding sections.

2 Notation and Conventions

We begin with some general notational conventions to be used throughout the paper. We write $\mathbb{R}_+$ for the set of nonnegative reals and $\mathbb{Z}_+$ for the set of nonnegative integers. We denote the indicator function of a set A by $\mathbf{1}_A$. That is $\mathbf{1}_A(x) = 1$ or $= 0$ according to whether $x \in A$ or not. When $A = \{i\}$, we will simply write $\mathbf{1}_i$ instead of writing $\mathbf{1}_{\{i\}}$, and write $\mathbf{1}$ for the constant function $\mathbf{1}(x) = 1$ for all x. The standard unit vectors in $\mathbb{R}^n$ (or $\mathbb{C}^n$) are denoted by $e_1, \ldots, e_n$. We also denote the standard inner product on $\mathbb{R}^n$ (or $\mathbb{C}^n$) by $\langle \cdot, \cdot \rangle$. We will omit the dependence of both the standard unit vectors and the standard inner product on the dimension of the space. The latter will be always clear from the context. Furthermore, we will identify a linear operator on $\mathbb{R}^n$ (or $\mathbb{C}^n$) with its matrix representation with respect to the standard basis. That is, if A is such an operator, the corresponding matrix $(A_{i,j})_{i,j=1,\ldots,n}$ is given by $A_{i,j} = \langle Ae_j, e_i \rangle$. Conversely, the operator A could be recovered from the matrix A, defining $(A\mathbf{1}_j)(i) = A_{i,j}$. Next we define the notation of an h–transform also known as Doob transform. If T is a linear operator on $\mathbb{C}^n$, and h

is a positive function on $\mathbb{C}^n$, then we let T^h denote the linear operator given by $(T^h)f = \frac{1}{h}T(hf)$. That is T^h is a diagonal change of basis and $(T^h)_{i,j} = \frac{1}{h(i)}T_{i,j}h(j)$.

Throughout the paper we fix a positive integer $\mathcal{N}$. Let $\overline{\mathcal{N}} = \{1, \ldots, \mathcal{N}\}$. Let $\mathcal{E}$ denote the set of directed edges $\overline{\mathcal{N}} \times \overline{\mathcal{N}}$. That is, the elements in $\mathcal{E}$ are of the form (i, j), $i, j \in \overline{\mathcal{N}}$. We write $\mathcal{E}_d$ for the subset of $\mathcal{E}$ consisting of all diagonal elements, that is elements of the form (i, i), $i \in \overline{\mathcal{N}}$. If A is an $\mathcal{N} \times \mathcal{N}$ matrix and $e = (i, j) \in \mathcal{E}$, we write A_e for $A_{i,j}$.

We now present the probabilistic framework for the paper. For more details on the material in this section and on Markov chains in general, we refer the reader to Aldous and Fill [1] and the references cited therein. Below $X := \{X(t) : t \in \mathbb{R}_+\}$ is an irreducible time–continuous Markov chain on the state space $\overline{\mathcal{N}}$. We write P_i for the distribution of X conditioned on $X(0) = i$, for $i \in \overline{\mathcal{N}}$, and denote the corresponding expectation operator by E_i. We recall that the paths of X are piecewise–constant and are right–continuous P_i–almost surely, for all $i \in \overline{\mathcal{N}}$. The probabilistic definition of irreducibility is that for all $i, j \in \overline{\mathcal{N}}$ and $t > 0$, $P_i(X(t) = j) > 0$. The generator of X is a linear operator A on $\mathbb{C}^{\mathcal{N}}$ defined through

$$(Af)(i) = \lim_{t \searrow 0} \frac{1}{t} E_i\big(f(X(t)) - f(X(0))\big) = \frac{d}{dt^+}\, E_i\big(f(X(t))\big)\big|_{t=0},$$

where and in what follows, $\frac{d}{dt^+}$ denotes the right–derivative with respect to the variable t. Thus, A is an $\mathcal{N} \times \mathcal{N}$ matrix whose entries are given by $A_{i,j} = \frac{d}{dt^+}\, P_i(X(t) = j)|_{t=0}$. By the Markov property, we have $\frac{d}{dt^+} P_i(X(t) = j) = AP_i(X(t) = j)$, for all $i, j \in \overline{\mathcal{N}}$ and $t \in \mathbb{R}_+$. Consequently,

$$P_i(X(t) = j) = (e^{At}\mathbf{1}_j)(i). \tag{1}$$

Furthermore, A is an essentially nonnegative matrix, namely, its off–diagonal entries are nonnegative, and it is irreducible. The irreducibility of a matrix A means that for each $i \neq j$, there exists a $k \in \mathbb{N}$ such that $A^k_{i,j} > 0$, and is equivalent to the the irreducibility of X. It is well–known that an essentially nonnegative and irreducible zero row–sum matrix is a generator of an irreducible Markov chain on $\overline{\mathcal{N}}$. For each $i \in \overline{\mathcal{N}}$, set $A_i := \sum_{j \neq i} A_{i,j}$. Of course, $A_i > 0$ by the irreducibility assumption. The times spent during each visit to i form a sequence of independent identically distributed exponential random variables with parameter A_i. A consequence of irreducibility is the existence of a unique invariant probability distribution for X, π. That is π is the unique probability vector in $\mathbb{R}^{\mathcal{N}}$ such that $\sum_i \pi_i P_i(X(t) = j) = \pi_j$, or equivalently $A^t\pi = 0$. For $j \in \overline{\mathcal{N}}$, let

$$\sigma_j \;=\; \inf\big\{t > 0 : X(t^-) \neq j,\ X(t) = j\big\}$$

denote the first entrance time of X to j (or the strictly positive reentrance time when $X(0) = j$). Then the invariant distribution, jump rates, and the expected return times are related by the following formula:

$$\pi_j \;=\; \frac{1}{A_j E_j(\sigma_j)}. \tag{2}$$

Under P_j, the time spent at j up to σ_j is equal to the time of the first jump from j, which, as explained above is exponentially distributed with expectation valued equal to $\frac{1}{A_j}$. Thus, (2) is the statement that π_j is the proportion of the average time spent at j before σ_j.

3 Group Inverse – Probabilistic Representation

Since A is the generator of an irreducible Markov chain, its Perron root, which equals 0, is simple and its nullspace is spanned by $\mathbf{1}$. Let $V_0 = \{v \in \mathbb{C}^{\mathcal{N}} : \langle v, \pi \rangle = 0\}$. Note that $\dim V_0 = \mathcal{N} - 1$. Clearly $AV_0 \subset V_0$, as $A^t \pi = 0$. Since $V_0 \cap \mathrm{Span}\{\mathbf{1}\} = \{0\}$, we conclude that $\mathbb{C}^{\mathcal{N}}$ could be written as the direct sum of two A–invariant subspaces: $\mathbb{C}^{\mathcal{N}} = \mathrm{Span}\{\mathbf{1}\} \oplus V_0$. Clearly, the restriction $(-A)|_{V_0}$ is invertible. Denote its inverse by $A^{\#}$ and extend $A^{\#}$ to $\mathbb{C}^{\mathcal{N}}$ by letting $A^{\#}\mathbf{1} = 0$. The matrix $A^{\#}$ is known in the literature as the group inverse of $-A$, [4], [5, pp. 118], and [6], as it satisfies the matrix equations:

$$A^{\#} \;=\; A^{\#}(-A)A^{\#},\; (-A)A^{\#}(-A) \;=\; (-A),\; \text{and } A^{\#}A \;=\; AA^{\#}.$$

It is important to note that here $A^{\#}$ is taken as the generalized inverse of $-A$ and not of A itself as is customary in much of the literature. Since the spectrum of $A|_{V_0}$ has only elements with strictly negative real part, it follows that

$$A^{\#}|_{V_0} \;=\; \int_0^\infty e^{A|_{V_0} t} dt.$$

Now let $v \to v^\perp$ denote the projection to V_0 along $\mathrm{Span}\{\mathbf{1}\}$ defined through $v^\perp = v - \langle v, \pi \rangle \mathbf{1}$. Since $v = v^\perp + \langle v, \pi \rangle \mathbf{1}$,

$$A^{\#}v \;=\; A^{\#}v^\perp \;=\; \int_0^\infty e^{At}(v - \langle v, \pi \rangle \mathbf{1}) dt \;=\; \int_0^\infty e^{At}\left(v - \langle v, \pi \rangle \mathbf{1}\right) dt.$$

The following proposition then follows directly from this identity applied to $v = \mathbf{1}_j$ and (1).

Proposition 1.

$$A^{\#}_{i,j} \;=\; \int_0^\infty \left(P_i(X(t) = j) - \pi_j\right) dt. \tag{3}$$

This proposition is known, see Aldous and Fill [1]. This expression does not seem to provide an easy way of studying the analytic properties of $A^{\#}$. We comment that the expression in (3) has the following discrete–time analog, and for more details we refer the reader to Meyer [27] and Kemeny and Snell [20]. Let P denote the transition matrix of an irreducible and aperiodic Markov chain and let π denote its invariant probability distribution. Recall that the fundamental matrix Z may be defined as

$$Z_{i,j} \;=\; \delta_{i,j} + \sum_{k=1}^{\infty} \left(P^k_{i,j} - \pi_j \right).$$

Then $A^{\#}$, the generalized inverse of $-A$, where $A = P - I$ is related to the fundamental matrix by the relation:

$$A^{\#}_{i,j} \;=\; Z_{i,j} - \pi_j.$$

Below we present two alternatives to Proposition 1 of probabilistic nature. The first should be viewed as a generalization of the relation (2), because it describes $A^{\#}$ in terms of proportions of times spent at sites. The second representation for $A^{\#}$ is in terms of expectations of first and second moments of hitting times. The existence of second moments plays a key role in proving some of the results that will follow and is in our opinion the main contribution of this formula.

To present the first formula, we need some notation. Let U be a random variable distributed according to π. There is no loss of generality assuming that X and U are defined on the same probability space and are independent. Set $T = \sigma_U$, that is T is the first time a site, randomly chosen according to π, is hit. Let

$$N_{i,j}(t) \;=\; \begin{cases} \int_0^t \mathbf{1}_i(X(s))ds, & i = j, \\ \#\{s > 0 : X(s^-) = i,\; X(s) = j\}, & i \neq j. \end{cases} \tag{4}$$

That is, $N_{i,i}(t)$ is the occupation time of X at i up to time t and $N_{i,j}(t)$, $i \neq j$, is the number of jumps from from i to j, up to time t. We are ready to state the result:

Proposition 2.

$$A^{\#}_{i,j} \;=\; E_i(N_{j,j}(T)) - \pi_j E_i(T) \;=\; E_i(T)\left(\frac{E_i(N_{j,j}(T))}{E_i(T)} - \pi_j\right).$$

The derivation of this result from Proposition 1 is intuitively clear. The choice of the random time T guarantees that $X(T)$ is distributed according to π. Therefore by the strong Markov property, $X(T+s)$ is distributed according to π, for all $s \in \mathbb{R}_+$. Formally, one observes that the integrand in (3) vanishes for $t \geq T$, which allows to replace the time integral over the infinite interval $[0,\infty)$ with an integral over a finite interval $[0,T]$, leading to the formula in Proposition 2. We now present a second probabilistic representation for $A^{\#}$:

Theorem 1. *For all* $i, j \in \overline{\mathcal{N}}$,

$$A^{\#}_{i,j} \;=\; \frac{1}{A_j}\left(\frac{E_j(\sigma_j^2)}{2(E_j(\sigma_j))^2} - \frac{1}{A_j E_j(\sigma_j)} - \mathbf{1}_{i\neq j}\frac{E_i(\sigma_j)}{E_j(\sigma_j)}\right).$$

Since $(A^{\#})^t\pi = 0$, it follows from the theorem that for all $j \in \overline{\mathcal{N}}$

$$\frac{E_j(\sigma_j^2)}{2A_j(E_j(\sigma_j))^2} - \frac{1}{A_j^2 E_j(\sigma_j)} \;=\; \sum_{k\neq j}\frac{\pi_k E_k(\sigma_j)}{A_j E_j(\sigma_j)} \;=\; \pi_j \sum_{k\neq j}\pi_k E_k(\sigma_j).$$

Combining this with (2) allows to eliminate the second moment term in Theorem 1 to obtain:

Corollary 1. $A^{\#}_{i,j} \;=\; \pi_j\left(\sum_{k\neq j}\pi_k E_k(\sigma_j) - \mathbf{1}_{i\neq j}E_i(\sigma_j)\right).$

The discrete analog to this result is [27, Equation 3.6]. However, we already mentioned above, we will show that the second moment term is useful.

Since much of the study in the literature is on discrete–time Markov chains, obtaining a discrete–time version of Proposition 2 and Theorem 1 may be of interest. Since we focus on the consequences of the existence of the second moment term, we will restrict the discussion here to the discrete–time version of Theorem 1.

Let P be the transition matrix of an irreducible Markov chain $Y := \{Y_n : n \in \mathbb{Z}_+\}$ on $\overline{\mathcal{N}}$. Then $A = P - I$ is the generator of an irreducible continuous–time Markov chain X. Note that X and Y

share the same invariant probability distribution. Furthermore, $A_j = 1 - P_{j,j} \in (0,1]$. For $j \in \overline{\mathcal{N}}$, define the discrete entrance time to j by letting

$$\tau_j \ = \ \min\{n \in \mathbb{Z}_+ \backslash \{0\} : Y_n = j\}.$$

Note that if $Y_0 = j$, then τ_j is the time until first return to j and is therefore not less than 1. We recall that $\pi_j = \frac{1}{E_j(\tau_j)}$, a fact that will frequently be used in the sequel.

Theorem 2. *Let $A = P - I$. Then*

$$A^{\#}_{i,j} \ = \ \frac{1}{2}\left(\frac{E_j(\tau_j^2)}{(E_j(\tau_j))^2} - \frac{1}{E_j(\tau_j)}\right) - \mathbf{1}_{i \neq j} E_i(\tau_j)\pi_j. \tag{5}$$

As an immediate corollary to Theorem 2 we obtain the discrete analog of Corollary 1 by replacing all occurrences of σ_j with τ_j. The proof is identical to the proof in the continuous setting.

We now briefly discuss some estimates on the terms appearing in (5). We begin with the expression $E_i(\tau_j)\pi_j$ appearing in the off–diagonal entry $A^{\#}_{i,j}$. Conditioning on the first step, we obtain that $E_j(\tau_j) = 1 + \sum_{l \neq j} P_{j,l} E_l(\tau_j)$. On rewriting this as

$$\frac{E_j(\tau_j) - 1}{1 - P_{j,j}} \ = \ \sum_{l \neq j} \frac{P_{j,l}}{1 - P_{j,j}} E_l(\tau_j),$$

and multiplying both sides by π_j, we arrive at the inequalities:

$$\min_{l \neq j} E_l(\tau_j)\pi_j \ \leq \ \frac{1 - \pi_j}{1 - P_{j,j}} \ \leq \ \max_{l \neq j} E_l(\tau_j)\pi_j.$$

Note that $1 - \pi_j \leq \frac{1-\pi_j}{1-P_{j,j}}$ and therefore $\max_j(1 - \pi_j) \leq \max_j \max_{l \neq j} E_l(\tau_j)\pi_j$. But the left–hand side is equal to $1 - \min_j \pi_j \geq 1 - \frac{1}{\mathcal{N}}$, so $1 - \frac{1}{\mathcal{N}} \leq \max_j \max_{l \neq j} E_l(\tau_j)\pi_j$. This inequality and the characterization of the chains attaining it were obtained by Kirkland in [23].

We turn to a discussion on the diagonal terms $A^{\#}_{j,j}$. Our estimates rely on the existence of the second moment term in (5). We have that

$$\frac{E_j(\tau_j^2)}{(E_j(\tau_j))^2} - \frac{1}{E_j(\tau_j)} \ = \ E_j\left(\frac{\tau_j}{E_j(\tau_j)} - 1\right)^2 + 1 - \frac{1}{E_j(\tau_j)}$$

The first term on the right–hand side is positive and is bounded below by $E_j\left(\mathbf{1}_1(\tau_j)\left(\frac{\tau_j}{E_j\tau_j} - 1\right)^2\right)$, with equality holding if and only if $P_{j,j} = 1$, or $P_{j,j} = 0$ and τ_j is P_j–almost surely a constant. The former case violates irreducibility and will be discarded. We then have that

$$\frac{E_j(\tau_j^2)}{(E_j(\tau_j))^2} - \frac{1}{E_j(\tau_j)} \ \geq \ P_{j,j}(1 - \pi_j)^2 + (1 - \pi_j) \ = \ (1 - \pi_j)\left(1 + P_{j,j}(1 - \pi_j)\right).$$

Substituting this into (5), we arrive to the following corollary:

Corollary 2.

$$A^{\#}_{j,j} \;\geq\; \frac{1}{2}(1-\pi_j)\,(1+P_{j,j}(1-\pi_j))\,.$$

Furthermore, an equality holds if and only if τ_j *is a constant* P_j*–almost surely. In this case* $P_{i,i}=0$*, for all* $i \in \overline{\mathcal{N}}$.

The last claim in the corollary follows from the simple fact that due to the irreducibility condition, $P_j(\sigma_i < \sigma_j) > 0$ for all $i \neq j$. Therefore if $P_{i,i} > 0$ for some i, we can force the chain to pass through i and stay at i for an arbitrary number of steps before returning to j. Thus τ_j attains infinitely many values, each with positive P_j–probability.

We also comment that Corollary 6 in Section 4, which provides a lower bound on the second–order derivatives of the Perron eigenvalue, is proved exactly in the same manner.

Recall that

$$E_j(\tau_j^2) \;=\; \int_0^\infty P_j(\tau_j > k)2kdk \;=\; \sum_{k=0}^{\infty} P_j(\tau_j > k)(2k+1) \;=\; E_j(\tau_j) + 2\sum_{k=0}^{\infty} P_j(\tau_j > k)k.$$

Now let $m_j := \min_{i\neq j}(1-P_{ij}) = 1 - \max_{i\neq j} P_{i,j}$. Then, for all $l \geq 1$,

$$P_j(\tau_j > l) \;\geq\; [1 - P_j(\tau_j = 1)]m_j^{l-1},$$

and so

$$\sum_{k=0}^{\infty} P_j(\tau_j > k)k \;\geq\; \frac{1-P_{j,j}}{(1-m_j)^2} \;=\; \frac{1-P_{j,j}}{\max_{i\neq j}(P_{i,j})^2}.$$

Summarizing, we have proved that:

Corollary 3.

$$A^{\#}_{j,j} \;\geq\; (1-P_{j,j})\frac{\pi_j^2}{\max_{i\neq j}(P_{i,j})^2}.$$

It is tempting to compare the lower estimates for $A^{\#}_{j,j}$ given in Corollary 2 with the one in Corollary 3. In many numerical examples the values that one obtains for Corollary 2 are higher than the corresponding ones for Corollary 3, but it is by no means universally so. As an example we give the transition matrix:

$$T \;=\; \begin{pmatrix} 0.1667 & 0.4692 & 0.1838 & 0.1803 \\ 0.365 & 0.1355 & 0.2542 & 0.2453 \\ 0.2724 & 0.3802 & 0.0360 & 0.3114 \\ 0.3638 & 0.5382 & 0.0364 & 0.0616 \end{pmatrix}.$$

For $A = I - T$, $(A^{\#}_{1,1}, A^{\#}_{2,2}, A^{\#}_{3,3}, A^{\#}_{4,4}) = (0.8333, 0.8645, 0.9640, 0.9384)$. The corresponding lower estimates the diagonal entries of $A^{\#}$ produced via Corollary 2 are given by: $(0.3956, 0.3526, 0.4348, 0.4198)$, while the corresponding lower estimates for the diagonal entries of $A^{\#}$ produced via Corollary 3 are given by $(0.2459, 0.3692, 0.0810, 0.1294)$.

Next we discuss the Kemeny constant. Let

$$\tilde{K} \;:= \operatorname{Tr}(A^{\#}) \;=\; \frac{1}{2}\sum_j \left(\frac{E_j(\tau_j^2)}{(E_j(\tau_j))^2}\right) - \frac{1}{2}.$$

Combining the fact that $A^{\#}\mathbf{1} = 0$ and (5) we observe that for all $i \in \overline{\mathcal{N}}$

$$0 \;=\; (A^{\#}\mathbf{1})_i \;=\; \sum_j A^{\#}_{i,j} \;=\; \operatorname{Tr}(A^{\#}) - \sum_{j\neq i} E_i(\tau_j)\pi_j.$$

Thus $\tilde{K} = \sum_{j\neq i} \pi_j E_i(\tau_j)$. This was observed by Kemeny and Snell [20], and, more recently, discussed in Doyle [11], Hunter [17] and [18] and Levene and Loizou [26]. The constant $K = \tilde{K} + 1$ is known in the literature as the Kemeny constant for P or as the expected time for mixing. Note that $\tilde{K}$ is equal to the discrete analog of $E_i(T)$, where T is the random variable defined above (4), explaining the terminology. By Corollary 2 and the discussion preceding it, we now obtain that:

Corollary 4. $K \geq \frac{\mathcal{N}+1}{2}$, *with equality holding if and only if* τ_j *is a constant with respect to* P_j, *for all* j. *In this case* $\pi_j = \frac{1}{\mathcal{N}}$ *and* $A^{\#}_{j,j} = \frac{1}{2}(1 - \frac{1}{\mathcal{N}})$.

This result is known. See for example Hunter [17].

We close this section with some formulas relating the discrete–time chain whose transition matrix is P and the continuous–time process whose generator is $A = P - I$. Matching the formula in Theorem 2 with that in Theorem 1 leads to the following:

Corollary 5.

1. $E_i(\sigma_j) \;=\; E_i(\tau_j) \times \begin{cases} 1, & i \neq j, \\ \dfrac{1}{1 - P_{j,j}}, & i = j. \end{cases}$

2. $$(1 - P_{j,j})E_j(\sigma_j^2) \;=\; E_j(\tau_j^2) + \frac{1 + P_{j,j}}{1 - P_{j,j}}\frac{1}{\pi_j}.$$

Proofs

We begin with the proof of Proposition 2.

Proof of Proposition 2. There is no loss of generality assuming that T is a stopping time for X because it is a stopping time for the process $\{(X(t), U) : t \in \mathbb{R}_+\}$. Observe that:

$$\int_0^\infty (P_i(X(t) = j) - \pi_j)dt \;=\; \lim_{M\to\infty} \int_0^M (P_i(X(t) = j) - \pi_j)dt \;=\; \lim_{M\to\infty} \int_0^M E_i\left(\mathbf{1}_j(X(t)) - \pi_j\right) dt.$$

However,

$$\int_0^M E_i\left(\mathbf{1}_j(X(t)) - \pi_j\right) dt \;=\; E_i\left(\int_0^{M\wedge T} \left(\mathbf{1}_j(X(t)) - \pi_j\right) dt\right) + E_i\left(\int_T^M \mathbf{1}_{\{T<M\}}\left(\mathbf{1}_j(X(t)) - \pi_j\right) dt\right). \tag{6}$$

Since T is a stopping time, the strong Markov property and the fact that $X(T)$ is distributed according to π imply that the second term on the right–hand side is equal to

$$\int_0^M \left(\int_u^M \left(P_\pi(X(t-u)=j) - \pi_j \right) dt \right) dP_i(T \leq u),$$

where $P_\pi := \sum_k \pi_k P_k$. Since π is the invariant distribution for X, $P_\pi(X(s)=j) = \pi_j$, for all $s \in \mathbb{R}_+$, and so the inner sum is equal to 0. In particular the second term on the right–hand side of (6) vanishes. As for the first term on the right–hand side of (6), we observe that it is equal to the expected value of a random variable, the time integral, which is dominated by the random variable $2T$. Since $E_i(2T) < \infty$, the result follows from the dominated convergence theorem by letting $M \to \infty$. $\square$

Next, we prove Theorem 1.

Proof of Theorem 1. Let $G(z) = (zI - A)^{-1}$. The operator G is known as the Green's function or the z–potential of A. It is known that G is the analytical continuation of $G(z) = \int_0^\infty e^{-t(zI-A)} dt$, for $\Re(z) > 0$. Probabilistically this can be expressed as

$$G_{i,j}(z) \;=\; E_i \left(\int_0^\infty \left(e^{-zt} \mathbf{1}_j(X(t)) \right) dt \right), \quad \Re(z) > 0.$$

By the strong Markov property, for $i \neq j$, we have that:

$$G_{i,j}(z) \;=\; E_i \left(e^{-z\sigma_j} \right) G_{j,j}(z),$$

while

$$G_{j,j} \;=\; E_j \left(\int_0^J e^{-zt} dt \right) + E_j \left(e^{-z\sigma_j} \right) G_{j,j}(z),$$

where J is the time X jumps from j and is exponentially distributed with parameter A_j. Then $\int_0^J e^{-zt} dt = \frac{1-e^{-zJ}}{z}$, and the expectation of this random variable is

$$\frac{1 - \frac{A_j}{z+A_j}}{z} \;=\; \frac{1}{z+A_j},$$

which will hereby be denoted by $K_j(z)$. Let $H_j(z) = \frac{1-E_j e^{-z\sigma_j}}{z}$. We then have:

$$G_{i,j}(z) \;=\; \frac{K_j(z)}{zH_j(z)} \cdot \begin{cases} 1, & i = j, \\ \\ E_i \left(e^{-z\sigma_j} \right), & i \neq j. \end{cases} \tag{7}$$

Note that $E_j \left(e^{-z\sigma_j} \right)$ and $\frac{K_j(z)}{H_j(z)}$ are analytic at 0 and so do not vanish there. The Laurent series of G at $z = 0$ can be written as

$$G(z) \;=\; \frac{T_{-1}}{z} + T_0 + zT_1 + \dots. \tag{8}$$

This leads to several observations. First, one has $T_{-1} = \lim_{z\to 0} zG(z)$. By (7),

$$(T_{-1})_{i,j} \;=\; \frac{K_j(0)}{H_j(0)} \;=\; \frac{1}{A_j E_j(\sigma_j)} \;=\; \pi_j.$$

Equivalently, T_{-1} is the projection $T_{-1}v = \langle v, \pi\rangle \mathbf{1}$. Next, note that V_0 is $zG(z)$–invariant. However, these operators are analytic at 0, while $\frac{d^k}{dz^k}(zG(z))|_{z=0} = k!T_k$. Hence V_0 is T_k–invariant. We also have that $zG(z)\mathbf{1} = \mathbf{1}$ and therefore $T_k\mathbf{1} = 0$. As a conclusion, for $v \in V_0$, $G(z)v = T_0v + zT_1v + \ldots$, and by taking $z \to 0$, it follows that $T_0 = A^{\#}$. Observe that $H_j(z) = E_j\sigma_j - \frac{z}{2}E_j(\sigma_j^2) + O(z^2)$, as $z \to 0$. Therefore,

$$\begin{aligned}(T_0)_{j,j} \;&=\; (zG_{j,j}(z))'|_{z=0} \;=\; -\frac{H_j'(0)}{H_j(0)^2}K_j(0) + \frac{K_j'(0)}{H_j(0)} \\ &=\; \frac{E_j(\sigma_j^2)}{2A_j(E_j\sigma_j)^2} - \frac{1}{A_j^2E_j(\sigma_j)}.\end{aligned}$$

Finally,

$$(T_0)_{i,j} = \frac{d}{dz}\left[\frac{K_j(z)}{H_j(z)}E_j\left(e^{-z\sigma_j}\right)\right]\Big|_{z=0} \;=\; (T_0)_{j,j} - (T_{-1})_{i,j}E_i(\sigma_j) \;=\; (T_0)_{j,j} - \frac{E_i(\sigma_j)}{A_jE_j(\sigma_j)}.$$

□

We close with the proof of Theorem 2, an adaptation of the proof of Theorem 1.

Proof of Theorem 2. Let $G(z) = (zI - P)^{-1} = z^{-1}(I - z^{-1}P)^{-1}$. In particular, G is the analytic continuation of the von–Neumann series $z^{-1}\sum_{k=0}^{\infty} z^{-k}P^k$. Since $P_{i,j}^k = P_i(X(k) = j)$, we have that:

$$G_{i,j}(z) \;=\; z^{-1}\sum_{k=0}^{\infty} z^{-k}P_i(X(k) = j). \tag{9}$$

Now let $0 \le \eta_0 < \eta_1 < \ldots$ be the times X visits j. Then $X(k) = j$ if and only if $k = \eta_l$, for some l. In particular,

$$G_{i,j}(z) \;=\; z^{-1}\sum_{l=0}^{\infty}\sum_{k=0}^{\infty} z^{-k}P_i(\eta_l = k) \;=\; z^{-1}\sum_{l=0}^{\infty} E_i(z^{-\sigma_l}).$$

Note that although η_0 may be equal to 0 (which is equivalent to $i = j$). However, for all $l \in \mathbb{Z}_+$, the distribution of $\eta_{l+1} - \eta_l$ under P_i coincides with that of τ_j under P_j. Therefore, applying the strong Markov property successively one arrives at the formula:

$$G_{i,j}(z) \;=\; \frac{E_i\left(z^{-\eta_0}\right)}{z(1 - E_j\left(z^{-\tau_j}\right))}.$$

On the other hand, since G is analytic with the exception of a finite number of poles, including a simple pole at 1, we obtain the Laurent expansion

$$G(z) \;=\; \frac{T_{-1}}{z-1} + T_0 + (z-1)T_1 + \ldots.$$

Here, T_{-1} is a projection given by $T_{-1}v = \langle v, \pi\rangle \mathbf{1}$. For $j \in \mathbb{Z}_+$, V_0 is T_j–invariant and $\mathrm{Span}\{\mathbf{1}\} \subset \ker(T_j)$. Therefore by letting $z \to 1$, it follows that $A^{\#} = T_0$. Equivalently, $A^{\#}_{i,j} = \frac{d}{dz}((z-1)G_{ij}(z))|_{z=1}$. Now

$$(z-1)G_{i,j}(z) \;=\; \frac{k(z)}{h(z)},$$

where

$$k(z) \;=\; E_i\left(z^{-\eta_j}\right) \quad \text{and} \quad h(z) \;=\; \frac{z\left(1 - E_j\left(z^{-\tau_j}\right)\right)}{z-1},$$

and therefore

$$A^{\#}_{i,j} \;=\; \frac{k'(0)}{h(0)} - \frac{k(0)h'(0)}{h^2(0)}. \tag{10}$$

To compute $A^{\#}_{i,j}$ explicitly, recall that by Taylor series expansion,

$$E_i\left(u^{\tau_j}\right) \;=\; 1 + (u-1)E_i(\tau_j) + \frac{1}{2}(u-1)^2 E_i(\tau_j(\tau_j - 1)) + o\left((u-1)^2\right). \tag{11}$$

Letting $u = z^{-1}$ and $i = j$, we obtain that

$$1 - E_j\left(z^{-\tau_j}\right) \;=\; -\left(z^{-1}-1\right)E_j\left(\tau_j\right) - \frac{1}{2}\left(z^{-1}-1\right)^2 E_j\left(\tau_j(\tau_j-1)\right) + o(z^{-1}-1\big).$$

Thus

$$h(z) \;=\; E_j\left(\tau_j\right) + \frac{z^{-1}-1}{2}\left(E_j\left(\tau_j\right)\right)^2 E_j\left(\tau_j(\tau_j-1)\right) \;=\; E_j\left(\tau_j\right) - (z-1)\frac{E_j(\tau_j(\tau_j-1))}{2} + o(z-1).$$

In particular, $h(0) = E_j\left(\tau_j\right)$ and $h'(0) = -\frac{1}{2}E_j\left(\tau_j(\tau_j-1)\right)$. Now $k(0) = 1$ and when $i = j$, $k(z) = 1$ for all z because $\eta_0 = 0$, so $k'(0) = 0$. However, when $i \neq j$, $k(z) = E_i\left(z^{-\tau_j}\right)$ and so by (11), $k'(0) = -E_i\left(\tau_j\right)$. Substituting these into (10), we obtain that:

$$A^{\#}_{i,j} \;=\; \frac{E_j\left(\tau_j(\tau_j-1)\right)}{2(E_j\left(\tau_j\right))^2} - \mathbf{1}_{i\neq j}\frac{E_i\left(\tau_j\right)}{E_j\left(\tau_j\right)}.$$

Thus the formula follows by algebraic manipulation along with the fact that $\pi_j = (E_j(\tau_j))^{-1}$. $\square$

4 Perturbations to Perron root

In this section we obtain a probabilistic interpretation of the first and second–order partial derivatives of the Perron root of A with respect to its entries, or more precisely, some transformation of the entries, to be described below. The main result, Theorem 4 identifies the Hessian matrix of the Perron eigenvalue as a covariance matrix of some family of random variables. Since covariance matrices are always positive semidefinite, this observation allows to derive some positivity and convexity results, which will be presented in the next section. We also prove a central limit theorem for the joint distribution of all additive functionals of X, and show that the covariance matrix for the limiting multi–variate normal distribution coincides with the above Hessian (or covariance matrix).

We begin with a well–known result on the first and second–order derivatives of the Perron root of A with respect to its entries. It serves as the bridge between probabilistic notions of the invariant distribution and the covariance matrix, and the analytic notions of the first and second–order derivatives of an analytic function, which are at the center of this section. The existence of the derivatives of all orders of the Perron root with respect to the matrix entries at A follows from the fact that the root is a simple eigenvalue of A and hence it is an analytic function in the entries of the matrix, see Wilkinson [32, pp.66–67]. Let $\lambda = \lambda(A)$ denote the Perron eigenvalue of A and denote by $\lambda_{(i,j)}$ the derivative of λ with respect to $A_{i,j}$ and $\lambda_{(k,l),(i,j)}$ the second order derivative with respect to $A_{i,j}$ and $A_{k,l}$. We are ready to state the result:

Theorem 3 (Deutsch and Neumann [10, Theorem 3.2])**.**

$$\lambda_{(i,j)} \;=\; \pi_i, \quad \lambda_{(i,j),(k,l)} \;=\; \pi_i A^{\#}_{j,k} + \pi_k A^{\#}_{l,i}.$$

The formula for the first derivative of the Perron root predates reference [10]. It can be found in Stewart [31, p.305, Exer. 1] and in many other places.

Combining Theorem 2 and Theorem 3 we have the following corollary:

Corollary 6. *Under the conditions of Theorem* 2 *and Theorem* 3 *we have that for all* $i, j \in \overline{\mathcal{N}}$*:*

$$\lambda_{(i,j),(j,i)} \;\geq\; \frac{1}{2}\left[(\pi_i + \pi_j - 2\pi_i\pi_j) + \pi_i P_{j,j}(1-\pi_j)^2 + \pi_j P_{i,i}(1-\pi_i)^2\right].$$

In particular,

$$\lambda_{(i,i),(i,i)} \;\geq\; \pi_i(1-\pi_i)[1 + P_{i,i}(1-\pi_i)].$$

Recall the definition of $\mathcal{E}$ and $\mathcal{E}_d$ from the first paragraph of Section 2. As will soon to become evident, the variables $\{\alpha_e : e \in \mathcal{E}\}$ are more natural than the actual entries of A. We consider the entries of A as variables and introduce a new set of variables $\{\alpha_e : e \in \mathcal{E}\}$ which we define as follows:

$$\alpha_e = \begin{cases} A_e, & e \in \mathcal{E}_d \text{ or } A_e = 0, \\ \ln A_e, & \text{otherwise.} \end{cases}$$

When $e \notin \mathcal{E}_d$ and $A_e = 0$, the edge e is non–existent from the point of view of the Markov chain X. Because of this reason, we will disregard these variables, and the dependence of the Perron eigenvalue on these will not be studied here.

We let U be the $\mathcal{N} \times \mathcal{N}$ matrix, whose all diagonal entries are all 1's and whose off–diagonal entries coincide with those of A. Note that $U_{i,j} = \frac{dA_{i,j}}{d\alpha_{i,j}}$. We will use U to simplify some of the formulas below.

Let $H = \{H_{e,e'} : e \in \mathcal{E}\}$ denote the Hessian matrix of λ defined through

$$H_{e,e'} = \frac{\partial^2 \lambda}{\partial \alpha_e \partial \alpha_{e'}}.$$

Theorem 3 now implies that:

Corollary 7.

1. $\dfrac{\partial \lambda}{\partial \alpha_{i,j}} \;=\; \pi_i U_{i,j}.$

2.

$$H_{e,e'} \;=\; \lambda_{e,e'} U_e U_{e'} + \begin{cases} 0, & e \neq e', \\ \\ \pi_i U_{i,j}, & e = e' = (i,j). \end{cases}$$

We now state the main result of this section, a probabilistic interpretation of H. We first need some notation. For $v \in \overline{\mathcal{N}}$, let $R_v = \frac{\sigma_v}{E_v(\sigma_v)}$ and recall the definition of the matrix–valued process N in (4). We have that:

Theorem 4. *Let $v \in \overline{\mathcal{N}}$. Then:*

$$H_{e,e'} \;=\; E_v\left[\left(\frac{N_e(\sigma_v) - R_v E_v(N_e(\sigma_v))}{\sqrt{E_v(\sigma_v)}}\right)\left(\frac{N_{e'}(\sigma_v) - R_v E_v(N_{e'}(\sigma_v))}{\sqrt{E_v(\sigma_v)}}\right)\right].$$

That is, H is the covariance matrix of the centered random variables

$$\left\{\frac{N_e(\sigma_v) - R_v E_v(N_e(\sigma_v))}{\sqrt{E_v(\sigma_v)}} : e \in \mathcal{E}\right\}$$

with respect to the measure P_v.

We comment that the main ingredient in the proof of Theorem 4 is a martingale stopping technique. To the best of our knowledge, this technique has been first used by Ney and Nummelin in [29] for studying Markov additive processes. More recently, and closer in spirit to our work, O'Cinneide [30] has applied the technique to prove a large number of results, mostly inequalities for Perron eigenvalues, including three of Cohen's [8] conjectures. Since our work is concerned with derivatives of the Perron root, we need to differentiate under the expectation sign. This requires an elaboration of the martingale stopping technique.

We further comment that by its very definition, a covariance matrix is positive semidefinite, a feature which is crucial for the convexity results in the next section.

Next we prove a central limit theorem for N or, more precisely, a limit theorem for the moment generating function of some properly centered and scaled version of N. This observation allows us to extend the probabilistic interpretation of H. Let $Z = \{Z_{i,j}(t) : t \in \mathbb{R}_+ \backslash \{0\}\}$ denote the $\mathcal{N} \times \mathcal{N}$ matrix–valued process Z defined through

$$Z_{i,j}(t) \;=\; \frac{N_{i,j}(t) - \pi_i U_{i,j} t}{\sqrt{t}}. \tag{12}$$

We define the dot product of two $\mathcal{N} \times \mathcal{N}$ matrices θ and Z, $\theta \cdot Z$, by letting $\theta \cdot Z = \sum_{i,j=1}^{\mathcal{N}} \theta_{i,j} Z_{i,j}$. We have:

Theorem 5. *There exists an $\epsilon > 0$ such that for all complex–valued matrices θ, with $\|\theta\|_\infty < \epsilon$,*

$$\lim_{t \to \infty} E_i\left(\exp(\theta \cdot Z(t))\right) \;=\; e^{\frac{1}{2}(H\theta)\cdot\theta}, \quad i \in \overline{\mathcal{N}}.$$

Theorem 5 has the immediate corollary:

Corollary 8. *The distribution and all moments of Z converge as $t \to \infty$ to those of a centered normal vector with covariance matrix H. In particular, for all $k \in \overline{\mathcal{N}}$ and $m \in \mathbb{Z}_+$,*

$$\lim_{t\to\infty} E_k(Z_{i,j}(t)^m) \ = \ \mathbf{1}_{2\mathbb{Z}_+}(m) \cdot (H_{(i,j),(i,j)})^{m/2}(m-1)!!, \tag{13}$$

where $(m-1)!!$ is the product of all odd positive integers less than or equal to $(m-1)$, defined as 1 when $m = 0, 1$. Furthermore by Theorem 4 and Corollary 7,

$$H_{(i,j),(i,j)} \ = \ (2A^{\#}_{j,i}U_{i,j} + 1)\pi_i U_{i,j}.$$

We note that the last part of the corollary is due to the fact that by Theorem 5, the moment generating function of $Z_{i,j}(t)$ converges to that of a centered normal random variable with variance $H_{(i,j),(i,j)}$. The m–th moment of such a random variable is equal the expression on the right–hand side in (13).

We would like to comment in regard to Theorem 5 and Corollary 8, that central limit theorems for additive functionals of Markov chains have been studied by many authors through different methods and are valid in much more general setting than presented here. For references representing the various approaches see Meyn and Tweedie [28] who utilize the regenerative structure, Dehling, Denker and Phillip [9] who utilize mixing arguments, and Gordin and Lifsic [16] and Kipnis and Varadhan [22] who utilize Martingale methods. Here we make three contributions. The first, and perhaps the most important, is an analytic identification of the covariance through the generalized inverse. The second is that in the case of a finite-state Markov chain, our proof provides a concise spectral-theoretic explanation to the central limit phenomenon. We are not the first to prove central limit by expansion of the Perron root. Lalley [25] has applied this idea to prove a central limit theorem for additive functionals of one dimensional Gibbs states. Finally, additive functionals appearing in the literature are usually restricted to linear combinations of occupation times, $\{N_e(t) : e \in \mathcal{E}_d\}$, a proper subclass of the additive functionals considered here, linear combinations of $\{N_e(t) : e \in \mathcal{E}\}$. The analysis for the former class corresponds to diagonal perturbations to A, while the analysis of the latter class considers perturbations to all entries of A (except for the vanishing off–diagonal entries).

Proofs

Proof of Theorem 4. We need some notation. We denote the entrywise product or Hadamard product of two $\mathcal{N} \times \mathcal{N}$ matrices A and B by $A \odot B$, defined as $A \odot B$ is the $\mathcal{N} \times \mathcal{N}$ matrix with $(A \odot B)_e = A_e B_e$. We also let $\dot{\exp}(B)$ denote the entrywise exponential of B, that is the $\mathcal{N} \times \mathcal{N}$ matrix given by $(\dot{\exp}(B))_e = \exp(B_e)$. Finally, let $\overline{B}$ denote the $\mathcal{N} \times \mathcal{N}$ matrix whose diagonal entries are all 0's and whose off–diagonal entries coincide with those of B.

Let η be a matrix of real numbers and define a family of linear operators $\mathcal{T} = \{T_t : t \geq 0\}$ by letting

$$(T_t f)(i) \ = \ E_i\big(e^{\eta\cdot N(t)} f(X(t))\big). \tag{14}$$

It follows from the Markov property that $\mathcal{T}$ is a semigroup. We now find its generator. Recalling that the probability of no jumps in $[0, \Delta t]$ is $1 - A_i \Delta t(1 + o(1))$, the probability of one jump to $j \neq i$

is $\Delta t(1 + o(1))A_{i,j}$, and that the distribution of number of jumps in $[0, \Delta t]$ has finite exponential moments in some neighborhood of the origin, it follows that

$$\frac{d}{dt^+}(T_t\mathbf{1}_j)(i)|_{t=0} \ = \ \begin{cases} A_{i,j}e^{\eta_{i,j}}, & j \neq i, \\ -A_i + \eta_{i,i}, & j = i. \end{cases}$$

That is, the generator of $\mathcal{T}$ is the matrix $A_\eta = A \odot e\dot{x}p(\overline{\eta}) + \eta - \overline{\eta}$. Consequently,

$$(T_t f)(i) \ = \ (e^{A_\eta t} f)(i). \tag{15}$$

Since we are interested in differentiating $\lambda(A)$ with respect to the variables $\{\alpha_e : e \in \mathcal{E}\}$ we observe that the transformation $\alpha_e \to \alpha_e + \Delta\alpha_e$ corresponds to $A_e \to A_e e^{\Delta\alpha_e}$, for all off–diagonal e, and $A_e \to A_e + \Delta\alpha_e$, for $e \in \mathcal{E}_d$. Or, in matrix notation, $\alpha \to \alpha + \Delta\alpha$ corresponds to $A \to A \odot \exp(\overline{\Delta\alpha}) + \Delta\alpha - \overline{\Delta\alpha} = A_{\Delta\alpha}$. We write $\lambda(\eta)$ for the Perron root of A_η. Therefore differentiation of $\lambda(A)$ corresponds to differentiation of $\lambda(\eta)$ at $\eta = 0$.

To study the derivatives of $\lambda(\eta)$ we employ a martingale argument. We refer the reader unfamiliar with martingales to either Williams [33], Durrett [12], or any graduate–level textbook in probability theory. Let $\{\mathcal{F}_t : t \in \mathbb{R}_+\}$ denote the filtration associated with X. That is $\mathcal{F}_t$ is generated by sets of the form $\{X(s) \in A : s \leq t, A \subset \overline{\mathcal{N}}\}$. Let

$$M_\eta(t) \ = \ \frac{1}{\varphi_\eta(X(0))}\exp(\eta \cdot N(t) - \lambda(\eta)t)\varphi_\eta(X(t)).$$

We will show that $\{(M_\eta(t), \mathcal{F}_t) : t \in \mathbb{R}_+\}$ is a martingale. Observe that by the Markov property,

$$E_i[M_\eta(t+s)|\mathcal{F}_t] \ = \ \frac{1}{\varphi_\eta(X(0))}\exp(\eta \cdot N(t) - \lambda(\eta)t)E_{X(t)}\big(\exp(\eta \cdot N(s) - \lambda(\eta)s)\varphi_\eta(X(s))\big).$$

Consider now the expectation on the right–hand side. We have to show it is equal to $\varphi_\eta(X(t))$. This expectation could be rewritten as $e^{-\lambda(\eta)s}(T_s\varphi_\eta)(X(t))$ which by (15) is equal to $\varphi_\eta(X(t))$.

We wish to stop this martingale at σ_i. Unfortunately, the above martingale is not uniformly integrable and therefore Doob's optional stopping theorem is not applicable here. Indeed, in Theorem 5 below, we prove that $\frac{\eta \cdot N(t) - \lambda(A)t}{\sqrt{t}}$ converges in law as $t \to \infty$ to a normal distribution. In particular, $\eta \cdot N(t) - \lambda(\eta)t$ attains values of order $\sqrt{t}$ with non-vanishing probability, which clearly violates uniform integrability. In order to stop the martingale, we apply a different approach, a change of measure.

Consider the family of linear operators $S = \{S_t : t \in \mathbb{R}_+\}$ defined by letting

$$(S_t f)(v) \ = \ e^{-\lambda(\eta)t}\frac{1}{\varphi_\eta(v)}(T_t(f\varphi_\eta))(v).$$

The generator of this semigroup, B, is equal to the h–transform $(A_\eta - \lambda(\eta)I)^{\varphi_\alpha}$. The h–transform of an operator was defined in the first paragraph of Section 2. It follows from this formula that B is a generator of an irreducible Markov chain. Let $\tilde{P}_v$ denote the distribution of this Markov chain starting from $v \in \overline{\mathcal{N}}$. Then since $(S_t f)(v) = E_v\left(M_\eta(t)f(X(t))\right)$, it follows that $\tilde{P}_v|_{\mathcal{F}_t}$ is absolutely continuous with respect to $P_v|_{\mathcal{F}_t}$ and the corresponding Radon-Nikodym derivative is $M_\eta(t)$. Let $\tilde{E}_v$ denote the expectation with respect to $\tilde{P}_v$. Note that irreducibility of X with respect to P_v implies irreducibility with respect to $\tilde{P}_v$, as

$$\tilde{P}_v(X(t)=j) \;=\; E_v\big(M_\eta(t)\mathbf{1}_j(X(t))\big) > 0.$$

In particular, $\sigma_v < \infty$, $\tilde{P}_v$-almost surely. Now by the strong Markov property,

$$\begin{aligned}\tilde{P}_v(\sigma_v \le t) \;&=\; E_v\left(M_\eta(t)\mathbf{1}_{\{\sigma_v \le t\}}\right)\\ &=\; E_v\left(M_\eta(\sigma_v)\mathbf{1}_{\{\sigma_v \le t\}} g(t-\sigma_v)\right),\end{aligned}$$

where $g(s) = E_v M_\eta(s) \equiv 1$. Therefore on letting $t \to \infty$, it follows from the dominated convergence theorem that

$$1 = \tilde{P}_v(\sigma_v < \infty) = E_v\big(M_\eta(\sigma_v)\big). \tag{16}$$

Thus, we were able to stop the martingale at σ_v while preserving the expectation value.

The next step is to differentiate (16) under the expectation sign. To justify this, we observe that $M_\eta(\sigma_v) = \exp(\eta \cdot N(\sigma_v) - \lambda(\eta)\sigma_v)$, and there is no longer any reference to φ_η. Since throughout the discussion v is to remain fixed, in the sequel we write M_η for $M_\eta(\sigma_v)$. Suppose now that δ is a $\mathcal{N} \times \mathcal{N}$ matrix, then

$$M_{\eta+\delta} \;=\; M_\eta \times \exp\left(\delta \cdot N(\sigma_v) - \left(\lambda(\eta+\delta) - \lambda(\eta)\right)\sigma_v\right).$$

Thus,

$$E_v(M_{\eta+\delta}) \;=\; \tilde{E}_v(\exp(\delta \cdot N(\sigma_v) - (\lambda(\eta+\delta) - \lambda(\eta))\sigma_v).$$

Since for all $e \in \mathcal{E}$, $N_e(\sigma_v)$ possesses finite exponential moments under $\tilde{P}_v$ in some neighborhood of the origin and since $\lambda(\cdot)$ is analytic, it follows that there exists some $\epsilon > 0$ such that

$$\tilde{E}_v\big(\exp(\epsilon \sum_{e\in\mathcal{E}} N_e(\sigma_v))\big) < \infty.$$

This, combined with the fact that $\exp\left(\delta \cdot N(\sigma_v) - \left(\lambda(\eta+\delta) - \lambda(\eta)\right)\sigma_v\right)$ is analytic in δ in some neighborhood of the 0–matrix, guarantee that one can differentiate (16) under the expectation to all orders.

It remains to differentiate (16) to obtain expression for the partial derivatives of λ with respect to the entries of η. These derivatives should not be confused with the derivatives of λ with respect to the entries of A. Differentiating once we obtain that:

$$0 \;=\; \frac{\partial}{\partial \eta_e} E_v(M_\eta) \;=\; E_v\left(\left(N_e(\sigma_v) - \frac{\partial \lambda(\eta)}{\partial \eta_e}\sigma_v\right) M_\eta(\sigma_v)\right).$$

It follows that

$$\frac{\partial \lambda}{\partial \eta_e}(0) \;=\; \frac{E_v(N_e(\sigma_v)))}{E_v(\sigma_v)}, \tag{17}$$

which gives a probabilistic proof to Corollary 7-(1). We are interested in the second-order derivatives and therefore we differentiate once again, this time with respect to $\eta_{e'}$ to obtain

$$\begin{aligned} 0 &= \frac{\partial^2}{\partial\eta_{e'}\partial\eta_e}E_v(M_\eta) \\ &= -E_v\left(\frac{\partial^2\lambda(\eta)}{\partial\eta_{e'}\partial\eta_e}\sigma_v M_\eta\right) \\ &+ E_v\left((N_e(\sigma_v)-\frac{\partial\lambda(\eta)}{\partial\eta_e}\sigma_v)(N_{e'}(\sigma_v)-\frac{\partial\lambda(\eta)}{\partial\eta_{e'}}\sigma_v)M_\eta\right). \end{aligned}$$

Thus, in view of (17) and when $\eta = 0$, we obtain that

$$\frac{\partial^2\lambda}{\partial\eta_{e'}\partial\eta_e}(0)E_v(\sigma_v) \;=\; E_v\big((N_e(\sigma_v)-R_vE_vN_e(\sigma_v))\,(N_{e'}(\sigma_v)-R_vE_vN_{e'}(\sigma_v))\big).$$

This concludes the proof.

$\square$

Next we prove the central limit theorem, Theorem 5.

Proof of Theorem 5. We adopt notation from the proof of Theorem 4. Let ψ_η denote the eigenvector of A_η corresponding to $\lambda(\eta)$, normalized to satisfy $\langle\psi_\eta,\varphi_\eta\rangle = 1$.

Define $P_\eta f = \langle f,\psi_\eta\rangle\varphi_\eta$. Then P_η is a projection whose image and kernel are A–invariant. But then, by the Jordan canonical form of A_η and equations (14) and (15) we have that

$$T_t \;=\; e^{A_\eta t} \;=\; e^{\lambda(\eta)t}P_\eta + e^{\gamma_\eta t}\Gamma(\eta,t), \tag{18}$$

where

$$\gamma_\eta \;=\; \max\Re\left(\{\lambda \text{ eigenvalue of } A_\eta : \lambda \;\neq\; \lambda(\eta)\}\right) \;<\; \lambda(\eta),$$

and where $\Gamma(\eta,t)$ is a linear operator satisfying $\Gamma(\eta,t)P_\eta = P_\eta\Gamma(\eta,t) = 0$. We note that for all η in some open neighborhood of 0 and for all $t \geq 0$, $\Gamma(\eta,t)$ is uniformly bounded and that $\limsup_{\eta\to 0}\gamma_\eta < 0$.

Below we write $\lambda'(0)$ for the $\mathcal{N}\times\mathcal{N}$ matrix given by $(\lambda'(0))_e = \frac{\partial\lambda}{\partial\eta_e}(0)$. Fix $k\in\overline{\mathcal{N}}$ and let

$$\Lambda(\theta,t) \;=\; E_k(e^{\theta\cdot Z(t)}) \;=\; e^{-\sqrt{t}\theta\cdot\lambda'(0)}E_k\left(e^{\frac{\theta}{\sqrt{t}}\cdot N(t)}\right).$$

Note that the last equality follows from the definition of Z in (12). On putting $\eta = \frac{1}{\sqrt{t}}\theta$, we obtain that

$$\Lambda(\theta,t) \;=\; e^{-\sqrt{t}\theta\cdot\lambda'(0)}(T_t\mathbf{1})(k) = \left(\underbrace{e^{-\sqrt{t}\theta\cdot\lambda'(0)}e^{\lambda(\frac{1}{\sqrt{t}}\theta)}P_{\frac{1}{\sqrt{t}}\theta}}_{(I)} + \underbrace{e^{-\sqrt{t}\theta\cdot\lambda'(0)}e^{\gamma_{\frac{1}{\sqrt{t}}\theta}t}\Gamma(\frac{1}{\sqrt{t}}\theta,t)}_{(II)}\right)\mathbf{1}(k),$$

where the first equality is due to the definition of T_t, (15), and the second is due to (18). We observe that as $t \to \infty$, $(II) \to 0$ at an exponential rate because Γ is uniformly bounded and $\lim_{t\to\infty} \gamma_{\frac{1}{\sqrt{t}}\theta} < 0$. Therefore it remains to estimate the asymptotic behavior of (I). By Taylor's expansion and analyticity of λ in some neighborhood of the origin, we have that

$$\lambda(\frac{1}{\sqrt{t}}\theta)t \;=\; \sqrt{t}\theta \cdot \lambda'(0) + \frac{1}{2}(H\theta)\cdot\theta + O\left(t^{-1/2}\right).$$

In addition, φ_η and ψ_η are analytic in some neighborhood of the origin. This implies $P_{\frac{1}{\sqrt{t}}\theta} \to \langle f, \pi\rangle \mathbf{1}$ as $t \to \infty$. Thus,

$$\lim_{t\to\infty} \Lambda(\theta, t) \;=\; e^{\frac{1}{2}(H\theta)\cdot\theta},$$

and the proof is complete. □

5 Applications to Positive Matrices

In this section we show some applications of the results and techniques presented in the preceding sections, focusing mostly on convexity results related to the Perron root.

Let B be an irreducible nonnegative matrix. Let $\lambda(B)$ denote its Perron eigenvalue and let ψ_B and φ_B denote, respectively, its corresponding positive left and right eigenvectors, normalized so that $\langle \psi_B, \varphi_B\rangle = 1$. In analogy to Section 4, we introduce a new set of variables $\{\beta_e : e \in \mathcal{E}\}$ which we define as follows:

$$\beta_e \;=\; \begin{cases} B_e, & e \in \mathcal{E}_d \text{ or } B_e = 0, \\ \ln B_e & \text{otherwise.} \end{cases}$$

The variables corresponding to vanishing off–diagonal entries will play no role in the following discussion.

We need to transform B into a generator of an irreducible Markov chain. Recall the definition of the h-transform T^h of a linear operator T, given in the first paragraph of Section 2. Of course, the spectrum of T^h coincides with that of T, because T^h and T are similar matrices. Now let $A = (B - \lambda(B)I)^{\varphi_B}$. It follows that A is the generator of some irreducible Markov chain on X. Furthermore, $\psi_B\varphi_B$ is a left eigenvector for A corresponding to the eigenvalue 0. Thus, π the invariant probability distribution is given by $\pi = \psi_B\varphi_B$.

Throughout this discussion, A and B above will be fixed as "reference" matrices and we will consider small perturbations to them obtained as follows. Suppose now that $\tilde{B}$ is obtained from B by changing some of its entries, including the vanishing off-diagonal entries. This perturbation induces a perturbation to A which we denote by $\tilde{A}$, given by $\tilde{A} = A + (\tilde{B} - B)^{\varphi_B}$. We note that

$$\tilde{A} \;=\; (B - \lambda(B)I + \tilde{B} - B)^{\varphi_B} \;=\; (\tilde{B} - \lambda(B)I)^{\varphi_B}.$$

Hence $\tilde{A}_{i,j} = \frac{1}{\varphi_B(i)}(\tilde{B}_{i,j} - \lambda(B)\delta_{i,j})\varphi_B(j)$. In particular, $\frac{dA_{i,j}}{dB_{i,j}} = \frac{\varphi_B(j)}{\varphi_B(i)}$. This impies $\frac{d\alpha_{i,j}}{d\beta_{i,j}} = 1$. We also observe that $\lambda(\tilde{A}) = \lambda(\tilde{B} - \lambda(B)I) = \lambda(\tilde{B}) - \lambda(B)$. Finally, a straightforward computation shows that $(B - \lambda(B)I)^{\#} = (A^{\#})^{1/\varphi_B}$ or, equivalently, that $A^{\#} = ((B - \lambda(B)I)^{\#})^{\varphi_B}$. These observations have the following theorems as immediate corollaries to Theorem 3 and Theorem 4, respectively.

Theorem 6.

1. $\dfrac{\partial\lambda(B)}{\partial B_{i,j}} \;=\; \dfrac{\varphi_B(j)}{\varphi_B(i)}\pi_i \;=\; \varphi_B(j)\psi_B(i)$.

2.

$$\begin{aligned}\frac{\partial^2\lambda(B)}{\partial B_{k,l}\partial B_{i,j}} \;&=\; \frac{\varphi_B(j)}{\varphi_B(i)}\frac{\varphi_B(l)}{\varphi_B(k)}\left(\pi_i A^{\#}_{j,k} + \pi_k A^{\#}_{li}\right) \\ &=\; \psi_B(i)\varphi_B(l)(B-\lambda(B)I)^{\#}_{j,k} + \psi_B(k)\varphi_B(j)(B-\lambda(B)I)^{\#}_{l,i}.\end{aligned}$$

Theorem 7.

1. $\dfrac{\partial\lambda(B)}{\partial\beta_{i,j}} \;=\; \dfrac{\partial\lambda(B)}{\partial\alpha_{i,j}} \;=\; \pi_i \;=\; \psi_B(i)\varphi_B(i)$.

2. $\dfrac{\partial^2\lambda(B)}{\partial\beta_{e'}\partial\beta_e} \;=\; H_{e,e'}$, *where H is the covariance matrix of Theorem 4. In particular, $\lambda(B)$ is a convex function of the variables $\{\beta_e : e \in \mathcal{E}\}$.*

The convexity of $\lambda(B)$ with respect to the diagonal entries of B was studied by several authors, see Cohen [7], Elsner [13], Friedland [15], and Deutsch and Neumann [10], and the references therein. Kingman [21] has shown that if all entries of B are log–convex functions of a single variable t, then $\lambda(B)$ is a convex function of t. The following Corollary to Theorem 7–(2) is a slight improvement of Kingman's result:

Corollary 9. *Let $I \subset \mathbb{R}$ be a nonempty interval interval. Suppose that for every $t \in I$, there exists an essentially nonnegative irreducible matrix $B(t)$, where the functions $B_e(\cdot)$, $e \in \mathcal{E}_d$, are convex and the functions $B_e(\cdot)$, $e \notin \mathcal{E}_d$, are* log*–convex. Then the function $t \to \lambda(B(t))$ is convex on I.*

To see why the corollary holds, for a given B, let β denote the $\mathcal{N} \times \mathcal{N}$ matrix whose e'th entry is β_e, and define $\tilde\lambda(\beta)$ as $\lambda(B)$. By Theorem 4, $\tilde\lambda$ is an increasing and convex function of the entries of β. Therefore for any $\rho \in [0,1]$ and $s,t \in I$, we have that

$$\begin{aligned}\lambda(B(\rho s + (1-\rho)t)) \;&=\; \tilde\lambda(\beta(\rho s + (1-\rho)t)) \;\le\; \tilde\lambda(\rho\beta(s) + (1-\rho)\beta(t)) \\ &\le\; \rho\tilde\lambda(\beta(s)) + (1-\rho)\tilde\lambda(\beta(t)) \;=\; \rho\lambda(B(s)) + (1-\rho)\tilde\lambda(B(t)).\end{aligned}$$

This shows that $t \to \lambda(B(t))$ is convex.

Another convexity–related result we prove is the following, due to Kirkland, Neumann, Ormes, and Xu [24].

Theorem 8 (Kirkland, Neumann, Ormes and Xu [24])**.** *For all $i,j \in \overline{\mathcal{N}}$ such that $B_{i,j} > 0$, we have that $\frac{\partial^2 \ln\lambda(B)}{\partial(\ln B_{i,j})^2} \ge 0$. An equality holds if and only if all of the following hold:*

1. $i \neq j$

2. $B_{i,j}\varphi_B(j) = \lambda(B)\varphi_B(i)$

3. *All sequences of states* $i_1, \ldots, i_l$ *satisfying* $i_1 = j$, $i_k \neq i$, *for* $k < l$, $i_l = i$, *and* $B_{i_k,i_{k+1}} > 0$, *for all* $k = 1, \ldots, l-1$ *are of the same length.*

We comment that our proof to Theorem 8 is essentially a probabilistic interpretation of the ideas originally used to prove the theorem in [24]. This illustrates the usefulness of the second moment term in Theorem 2. The key to proof of the theorem in [24] is [24, Lemma 2.1], which is an inequality between two series. Although not immediately clear, it could be shown that this inequality reduces to the statement that the variance of some random variable is nonnegative, something which is always true. In our approach, we use Theorem 2 to express the second derivative of the logarithm of the Perron root as the sum of the variance of a random variable and a nonnegative term, see (21).

We further remark that if all the entries of B are strictly positive, then the matrix whose entries are $\frac{\partial \ln \lambda(B)}{\partial \ln B_{i,j}}$, $i,j \in \overline{\mathcal{N}}$ is known in the literature as the elasticity matrix associated to B. It provides a normalized measurement of the change of λ induced by a change in the entries. Indeed, by the chain rule

$$\frac{\partial \ln \lambda(B)}{\partial \ln B_{i,j}} \;=\; \frac{\partial \lambda(B)}{\partial B_{i,j}} \frac{\partial \ln \lambda(B)}{\partial \lambda(B)} \frac{\partial B_{i,j}}{\partial \ln B_{i,j}} \;=\; \frac{1}{\lambda(B)} \frac{\partial \lambda(B)}{\partial B_{i,j}} B_{i,j},$$

and by Theorem 3–(1),

$$\frac{\partial \ln \lambda(B)}{\partial \ln B_{i,j}} \;=\; \frac{1}{\lambda(B)} \varphi_B(j)\psi_B(i)B_{i,j} \geq 0.$$

Thus summing over all i,j, we obtain

$$\sum_{i,j} \frac{\partial \ln \lambda(B)}{\partial \ln B_{i,j}} \;=\; \frac{1}{\lambda(B)} \langle \psi_B, B\varphi_B \rangle \;=\; 1.$$

Next we prove a result on the relation between the eigenvectors of B and its "topological" structure, originally proved by Elsner and Johnson [14], where it was proved using matrix–theoretic methods. Although our proof contains ingredients present in the original proof, mostly the idea of cycle decomposition, our approach presents the probabilist's view on this problem and illustrates the usefulness of probabilistic techniques in this and similar contexts.

To begin, let E be a $\{0,1\}$–valued matrix defined by letting $E_e = 1$ if and only if $B_e > 0$, $e \in \mathcal{E}$. That is, E is the indicator function of the existing edges. We now have:

Theorem 9 (Elsner and Johnshon [14]). *For all* $\epsilon \geq 0$, $\frac{d}{d\epsilon}\lambda(B + \epsilon E) \geq 1$ *or, equivalently, for all* $\epsilon \geq 0$, $\lambda(B(\epsilon)) \geq \lambda(B) + \epsilon$.

Note that it follows from Theorem 6–(1) that $\frac{d}{d\epsilon}\lambda(B + \epsilon E)|_{\epsilon=0} = \langle \psi_B, E\varphi_B \rangle$. Therefore we obtain the following energy-like estimate:

Corollary 10. *Let* E *be an irreducible* $(0,1)$*–matrix, and let* B *be any nonnegative matrix satisfying that* $B_{i,j} > 0$ *if and only if* $E_{i,j} = 1$. *If* ψ_B *and* φ_B *are respectively, positive left and right eigenvectors of* B *corresponding to the Perron root* $\lambda(B)$, *then* $\langle \psi_B, E\varphi_B \rangle \geq \langle \psi_B, \varphi_B \rangle$.

Proofs

We begin with a lemma used in the proof of Theorem 8. We then prove the lemma, and continue proving the theorem from the lemma. We close the section with the proof of Theorem 9.

Lemma 1. *Let P be the transition matrix of an irreducible discrete–time Markov chain. Then for all $i, j \in \overline{\mathcal{N}}$, $\frac{\partial^2 \ln \lambda}{\partial (\ln P_{i,j})^2} \geq 0$. Given a pair $i, j \in \overline{\mathcal{N}}$, then an equality holds if and only if all of the following conditions hold:*

1. $i \neq j$.

2. $P_{i,j} = 1$.

3. *All paths from j to i are of the same length (equal to $E_i(\tau_i) - 1$).*

Proof of Lemma 1. In [24] it was shown that

$$\frac{\partial^2 (\ln \lambda(P))}{\partial (\ln P_{i,j})^2} \;=\; \pi_i P_{i,j} \left(2 P_{i,j} A^{\#}_{j,i} + 1 - \pi_i P_{i,j} \right). \tag{19}$$

This is obtained from Theorem 3 (where $A = P - I$) and the chain rule. Recall now from Theorem 2 that

$$2A^{\#}_{j,i} \;=\; \pi_i^2 E_i(\tau_i^2) - \pi_i - 2\mathbf{1}_{i \neq j} E_j(\tau_i) \pi_i. \tag{20}$$

Conditioning on the first step of the discrete-time chain Y, it follows from the Markov property that

$$E_i(\tau_i^2) \;=\; E_i\left(\sum_{l \neq i} \mathbf{1}_l(Y_1) E_l\big((\tau_i + 1)^2\big) \right) + P_{i,i} \;=\; P_{i,j} E_j\big((\tau_i + 1)^2\big) + \underbrace{\sum_{l \neq i,j} P_{i,l} E_l((\tau_i + 1)^2) + P_{i,i}}_{= \Delta_{i,j}}.$$

Note that $\Delta_{i,j} \geq 0$ with equality holding if and only if $P_{i,j} = 1$. Plugging this into (20) we have

$$2A^{\#}_{j,i} = \pi_j^2 \left(P_{i,j} E_j((\tau_j + 1)^2) + \Delta_{i,j} \right) - \pi_j - 2\pi_i \mathbf{1}_{i \neq j} E_j(\tau_j)$$

Hence,

$$2P_{i,j} A^{\#}_{j,i} + 1 - \pi_i P_{i,j} = \pi_j^2 P_{i,j} \left(P_{i,j} E_j((\tau_j + 1)^2) + \Delta_{i,j} \right) - \pi_i P_{i,j} - 2\mathbf{1}_{i \neq j} \pi_i P_{i,j} E_j(\tau_j) + 1 - \pi_i P_{i,j}.$$

Multiplying both sides by $\pi_i P_{i,j}$ we arrive to

$$\begin{aligned}
\frac{\partial^2 \lambda(P)}{\partial (\ln P_{i,j})^2} \;&=\; \pi_i P_{i,j} \left(\pi_i^2 P_{i,j}^2 E_j\big((\tau_i + 1)^2\big) + \pi_i^2 P_{i,j} \Delta_{i,j} - \pi_i P_{i,j} - 2\mathbf{1}_{j \neq i} \pi_i P_{i,j} E_j\left(\tau_i\right) + 1 - \pi_i P_{i,j} \right) \\
&=\; \pi_i P_{i,j} E_j\Big(\big(\pi_i P_{i,j} (\tau_i + 1) - 1 \big)^2 \Big) + (\pi_i P_{i,j})^2 \big(\pi_i \Delta_{i,j} + 2\mathbf{1}_{i = j} E_j(\tau_i) \big).
\end{aligned} \tag{21}$$

The right–hand side is clearly nonnegative. Since our discussion is restricted to pairs i, j for which $P_{i,j} > 0$, the expression on the right–hand side is equal to 0 if and only if all of the following hold:

1. $i \neq j$

2. $\pi_i P_{i,j}(\tau_i + 1) = 1$, P_i–almost surely.

3. $\Delta_{i,j} = 0$, namely $P_{i,j} = 1$.

This completes the proof. $\square$

We now prove Theorem 8.

Proof of Theorem 8. Similarly to the discussion preceding Theorem 7, we let $P = \frac{1}{\lambda(B)} B^{\varphi_B}$. Then P is the transition function of an irreducible Markov chain. A perturbation $\tilde{B}$ to B induces a perturbation $\tilde{P}$ to P defined through $\tilde{P} = \frac{1}{\lambda(B)}(B + \tilde{B} - B)^{\varphi_B}$. Therefore $\frac{dP_{i,j}}{dB_{i,j}} = \frac{1}{\lambda(B)} \frac{\varphi_B(j)}{\varphi_B(i)}$, or on viewing $\tilde{B}$ as a perturbation induced by a perturbation to P, $\frac{dB_{i,j}}{dP_{i,j}} = \lambda(B) \frac{\varphi_B(i)}{\varphi_B(j)}$. Now note that $\lambda(\tilde{P}) = \frac{1}{\lambda(B)} \lambda(\tilde{B})$. Thus by the chain rule, for all pairs i, j such that $B_{i,j} > 0$, we have that $\frac{\partial^2 \ln \lambda(B)}{\partial (\ln B_{i,j})^2} = \lambda(B) \frac{\partial^2 \ln \lambda(P)}{\partial (\ln P_{i,j})^2}$. From Lemma 1, the right–hand side is nonnegative and is strictly positive if and only if $\imath \neq j$, $P_{i,j} = 1$ and all paths from j to i are of the same length. Restating the latter two conditions in terms of B completes the proof of the theorem. $\square$

We now prove Theorem 9.

Proof of Theorem 9. Let $B(\epsilon) = B + \epsilon E$. Note that it is enough to prove that $\lambda'(B(0)) \geq 1$, because the derivative of $\lambda(B(\cdot))$ at $\epsilon = \epsilon_0 > 0$ is obtained by replacing B with $B(\epsilon_0)$ and differentiating $\lambda(B(\epsilon_0 + \epsilon))$ at $\epsilon = 0$. We will work with a discrete-time Markov chain, whose transition function we define as follows. For $i \in \overline{\mathcal{N}}$, let $d_i = \sum_j E_{i,j}$. Let $P = \mathrm{diag}(d_1^{-1}, \ldots, d_n^{-1})E$. Then P is the transition matrix of an irreducible Markov chain, $Y = \{Y_n : n \in \mathbb{Z}_+\}$. The matrix P will remains fixed throughout the discussion below.

Let C be any nonnegative matrix which is zero exactly where E (or B) is. Then for all $e \in \mathcal{E}$,

$$C_e \; = \; P_e \frac{C_e}{P_e}.$$

Or, on letting $N_{i,j}(n)$ denote the number of time Y has jumped from i to j up to time n, one can reformulate the relation above between C and P through

$$C_{i,j} = E_i \left(\left(\prod_{e \in \mathcal{E}} \left(\frac{C_e}{P_e} \right)^{N_e(1)} \right) \mathbf{1}_j(Y_1) \right).$$

The Markov property then implies that for all $n \in \mathbb{N}$,

$$C^n_{i,j} = E_i \left(\prod_{e \in \mathcal{E}} \left(\frac{C_e}{P_e} \right)^{N_e(n)} \mathbf{1}_j(Y_n) \right).$$

We now fix some $i \in \overline{\mathcal{N}}$. Let

$$Z_n(C) = \left(\prod_{e \in \mathcal{E}} \left(\frac{C_e}{P_e} \right)^{N_e(n)} \right) \mathbf{1}_i(Y_n).$$

Then $C^n_{i,i} = E_i(Z_n(C))$. Next, note that $\ln \lambda(C) = \lim_{n\to\infty} \frac{1}{n} \ln C^n_{i,i}$. Therefore,

$$\begin{aligned}\ln \lambda(B(\epsilon)) - \ln \lambda(B) &= \lim_{n\to\infty} \frac{1}{n} \big(\ln E_i\big(Z_n(B(\epsilon))\big) - \ln E_i\big(Z_n(B)\big) \big) \\ &= \lim_{n\to\infty} \frac{1}{n} \ln \frac{E_i\big(Z_n(B(\epsilon))\big)}{E_i\big(Z_n(B)\big)}. \end{aligned} \tag{22}$$

Now for any two integrable positive random variables U and V, defined on the same probability space, we have that $E(VE(U) - UE(V)) = 0$. Therefore with positive probability there exists a realization of the vector (U, V) for which $VE(U) \geq UE(V)$. By the assumption on positivity of U and V, this inequality is equivalent to $E(U)/E(V) \geq U/V$, or $\ln \frac{E(U)}{E(V)} \geq \ln \frac{U}{V}$. Taking $U = Z_n(B(\epsilon))$ and $V = Z_n(B)$, we see that there exists a realization of Y satisfying

$$\ln \frac{E_i Z_n(B(\epsilon))}{E_i Z_n(B)} \;\geq\; \ln \frac{Z_n(B(\epsilon))}{Z_n(B)}. \tag{23}$$

Since we assumed that $Y_n = i$, this realization is a loop (starting and ending at the same point) and therefore it can be decomposed into k loops with no self intersection, $\Gamma_1, \dots, \Gamma_k$. More precisely, there are k increasing subsequences $I_1, \dots, I_k$ of $\{0, 1, \dots, n\}$, where $I_l = (n^l_1, \dots, n^l_{|I(l)|})$ and the following conditions holds:

1. For all l, $Y_{n^l_1} = Y_{n^l_{|I(l)|}}$, while the elements $Y_{n^l_1}, \dots, Y_{n^l_{|I_l|-1}}$ are all distinct.

2. For all l and $j \in \{1, \dots, |I_l| - 1\}$, we have that $P_{Y_{n^l_j}, Y_{n^l_{j+1}}} > 0$.

3. Each $s \in \{0, \dots, n\}$ belongs to either one or two subsequences and if it belongs to two sequences I_l and $I_{l'}$, $l < l'$, then $n^{l'}_1 = s = n^l_j$ for some $j \in \{1, \dots, |I_l| - 1\}$ and $n^l_{j+1} = \inf\{k > n^l_j : Y_k = Y_{n^l_j}\} + 1$.

Let us explain these conditions. The first means that the path Γ_l, defined as the sequence $(Y_{n^l_1}, Y_{n^l_2}, \dots, Y_{n^l_{|I_l|}})$, is a loop having no sub–loops. The second means that this loop consists of allowed transitions, or that it has positive probability. The third explains the rule governing the allocation of times $\{0, \dots, n\}$ between two loops and how we break a loop that contains sub–loops into loops that do not, by simply removing the "nested" sub–loops.

Let $N^l_{i,j}$ denote the number of times Γ_l has i followed by j, or simply put, the number of transitions from i to j. Of course $N^l_{i,j}$ is either 0 or 1, because Γ_l does not contain sub–loops. Let C^l be $\mathcal{N} \times \mathcal{N}$ matrix defined through $C^l_e = N^l_e C_e$, and let $\rho_{C^l} = \prod_{e\in\mathcal{E}} N^l_e C_e$, the product of the non–zero elements in C^l. Observe that since Γ_l is a loop, $\lambda(C^l) = \rho_{C^l}^{1/|I_l|}$.

Clearly, $Z_n(C) = \prod_{l=1}^k \rho_{C^l}$. Hence taking $C = B(\epsilon)$ once and then taking $C = B$, we find that

$$\frac{Z_n(B(\epsilon))}{Z_n(B)} \;=\; \prod_{l=1}^k \left[\frac{\rho_{B(\epsilon)^l}}{\rho_{B^l}}\right] \;=\; \prod_{l=1}^k \left[\frac{\lambda(B(\epsilon)^l)}{\lambda(B^l)}\right]^{|I_l|}.$$

Now

$$\left[\frac{\lambda(B(\epsilon)^l)}{\lambda(B^l)}\right]^{|I_l|} = \prod_{\{e\in\mathcal{E}: B^l_e > 0\}} \frac{B_e + \epsilon}{B_e}.$$

Let $\eta \in (0,1)$ be arbitrary. Then for ϵ sufficiently small, $\frac{B_e+\epsilon}{B_e} = 1 + \frac{\epsilon}{B_e} \geq e^{(1-\eta)\frac{f\epsilon}{B_e}}$ for all $e \in \mathcal{E}$. This implies

$$\ln \frac{Z_n(B(\epsilon))}{Z_n(B)} \geq (1-\eta)\sum_{l=1}^{k} \sum_{\{e:\in\mathcal{E}:B_e^l>0\}} \frac{\epsilon}{B_e} \;=\; (1-\eta)\epsilon \sum_{l=1}^{k} |I_l| \frac{1}{|I_l|} \sum_{\{e:\in\mathcal{E}:B_e^l>0\}} \frac{1}{B_e}.$$

As $\sum_{l=1}^{k} |I_l| \geq n$, there exists now some l_0 such that

$$\frac{1}{n} \ln \frac{Z_n(B(\epsilon))}{Z_n(B)} \;\geq\; \frac{(1-\eta)}{|I_{l_0}|} \sum_{\{e:\in\mathcal{E}:B_e^{l_0}>0\}} \frac{\epsilon}{B_e} \;\geq\; (1-\eta)\epsilon \left(\prod_{\{e:\in\mathcal{E}:B_e^{l_0}>0\}} \frac{1}{B_e^{l_0}} \right)^{1/|I_{l_0}|} \;=\; \frac{(1-\eta)\epsilon}{\lambda(B^{l_0})},$$

where we have used the arithmetic–geometric mean inequality to obtain the second inequality. But since $B_e^{l_0} \leq B_e$ for all $e \in \mathcal{E}$, we have $\lambda(B^{l_0}) \leq \lambda(B)$. Thus,

$$\frac{1}{n} \left(\ln Z_n(B(\epsilon)) - \ln Z_n(B) \right) \;\geq\; \frac{(1-\eta)\epsilon}{\lambda(B)}.$$

From (22) and (23) we then obtain

$$\frac{\ln \lambda(B(\epsilon)) - \ln(\lambda(B))}{\epsilon} \;\geq\; \frac{1-\eta}{\lambda(B)}.$$

Taking $\epsilon \to 0$ and then $\eta \to 0$ shows that $(\ln \lambda(B(\epsilon)))'|_{\epsilon=0} \geq \frac{1}{\lambda(B)}$, or, equivalently, $\lambda'(B(\epsilon)) \geq 1$, completing the proof. ☐

References

[1] David Aldous and James A. Fill. Reversible Markov Chains and Random Walks on Graphs. preliminary draft, available on
`http://www.stat.berkeley.edu/~aldous/RWG/book.html`.

[2] Iddo Ben-Ari, Khalid Boushaba, Anastasios Matzovinos, and Alexander Roitershtein Stochastic Analysis of the Motion of DNA Nanomechanical Bipeds In press, *Bulletin of Mathematical Biology*
`http://dx.doi.org/10.1007/s11538-010-9600-x`

[3] I. Ben-Ari, O. Kurth, and M. Neumann. In press, *Linear Algebra and its Applications*
`http://dx.doi.org/10.1016/j.laa.2011.04.044`

[4] A. Ben–Israel and T.N.E. Greville, Generalized Inverses: Theory and Applications, 2nd ed., *Springer*, New York, 2003.

[5] Abraham Berman and Robert J. Plemmons. Nonnegative Matrices in the Mathematical Sciences, volume 9 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[6] S. L. Campbell and C. D. Meyer. Generalized Inverses of Linear Transformations. *Dover Publications*, New York, 1991.

[7] Joel E. Cohen. Random evolutions and the spectral radius of a nonnegative matrix. *Math. Proc. Cambridge Philos. Soc.*, 86(2):345–350, 1979.

[8] Joel E. Cohen Eigenvalue inequalities for random evolutions: origins and open problems (Inequalities in statistics and probability (Lincoln, Neb., 1982) *IMS Lecture Notes Monogr. Ser.*, 5:41–53, 1984.

[9] Herold Dehling, Manfred Denker, and Walter Philipp. Central limit theorems for mixing sequences of random variables under minimal conditions. *Ann. Probab.*, 14(4):1359–1370, 1986.

[10] Emeric Deutsch and Michael Neumann. On the first and second order derivatives of the Perron vector. *Linear Algebra Appl.*, 71:57–76, 1985.

[11] Peter Doyle. The Kemeny constant of a Markov chain. preprint, available on `http://math.dartmouth.edu/~doyle/docs/kc/kc.pdf`.

[12] Richard Durrett. Probability: Theory and Examples. *Duxbury Press*, Belmont, CA, second edition, 1996.

[13] Ludwig Elsner. On convexity properties of the spectral radius of nonnegative matrices. *Linear Algebra Appl.*, 61:31–35, 1984.

[14] Ludwig Elsner and Charles R. Johnson. Nonnegative matrices, zero patterns, and spectral inequalities. In *Proceedings of the Fourth Haifa Matrix Theory Conference* (*Haifa,* 1988), volume 120, pages 225–236, 1989.

[15] Shmuel Friedland. Convex spectral functions, *Linear and Multilinear Algebra*, 9(4):299–316, 1980/81.

[16] M. I. Gordin and B. A. Lifšic. Central limit theorem for stationary Markov processes. *Dokl. Akad. Nauk SSSR*, 239(4):766–767, 1978.

[17] Jeffrey J. Hunter. Mixing times with applications to perturbed Markov chains. *Linear Algebra Appl.*, 417(1):108–123, 2006.

[18] Jeffrey J. Hunter. Coupling and mixing times in a Markov chain. *Linear Algebra Appl.*, 430(10):2607–2621, 2009.

[19] I. Karatzas and S. E. Shreve. Brownian Motion and Stochastic Calculus. Graduate Text in Mathematics, Vol. 113. *Springer–Verlag*, New Your, 1991.

[20] John G. Kemeny and J. Laurie Snell. Finite Markov Chains. *Springer-Verlag*, New York, 1976. Reprinting of the 1960 original, Undergraduate Texts in Mathematics.

[21] J. F. C. Kingman. A convexity property of positive matrices. *Quart. J. Math. Oxford Ser. (2)*, 12:283–284, 1961.

[22] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1):1–19, 1986.

[23] S. Kirkland. On a question concerning condition numbers for Markov chains. *SIAM J. Matrix Anal. Appl.*, 23(4):1109–1119 (electronic), 2002.

[24] S. J. Kirkland, M. Neumann, N. Ormes, and J. Xu. On the elasticity of the Perron root of a nonnegative matrix. *SIAM J. Matrix Anal. Appl.*, 24(2):454–464 (electronic), 2002.

[25] S. Lalley. Ruelle's Perron-Frobenius theorem and the central limit theorem for additive functionals of one-dimensional Gibbs states. Adaptive statistical procedures and related topics (Upton, N.Y., 1985). *IMS Lecture Notes Monogr. Ser.*, 8 (1986), 428–446.

[26] M. Levene and G. Loizou. Kemeny's constant and the random surfer. *Amer. Math. Monthly*, 109:741–745, 2002.

[27] Carl D. Meyer. The role of the group generalized inverse in the theory of finite Markov chains. *SIAM Rev.*, 17:443–464, 1975.

[28] Sean Meyn and Richard L. Tweedie. Markov Chains and Stochastic Stability. *Cambridge University Press*, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.

[29] P. Ney and E. Nummelin. Markov additive processes I. Eigenvalue properties and limit theorems. *Ann. Probab.*, 15(2):561–592, 1987.

[30] C. O'Cinneide. Markov additive processes and Perron-Frobenius eigenvalue inequalities. em Ann. Probab., 28(3):1230–1258, 2000.

[31] G. W. Stewart. Introduction to Matrix Computations. *Academic Press*, New York, 1973.

[32] J. H. Wilkinson. The Algebraic Eigenvalue Problem. *Oxford Univ. Press*, London, 1965.

[33] David Williams. Probability with Martingales. *Cambridge Mathematical Textbooks. Cambridge University Press*, Cambridge, 1991.