Bernoulli **23**(3), 2017, 2058–2082
DOI: 10.3150/16-BEJ804

Representations for the decay parameter of Markov chains

JINWEN CHEN[1,*], SIQI JIAN[2,†] and HAITAO LI[1,**]

[1]*Department of Mathematics, Tsinghua University, China.*
E-mail: [*]*jchen@math.tsinghua.edu.cn;* [**]*liht0618@163.com*
[2]*School of Statistics, Capital University of Economics and Business, Beijing China.*
E-mail: jiansiqifrank@163.com

In this paper, we give variational representations for decay parameters of Markov chains. In continuous-time cases, the representation involves Donsker–Varadhan's famous I-functional, from which some dual representations are given, which are expected to he useful in estimating the lower and upper bounds of the decay parameter. As a consequence, dual representations for decay parameters of discrete time Markov chains are derived. For continuous-time chains with finite states, we also give another form of dual formulas, which can be regarded as a version of the one for the Perron–Frobenius eigenvalue, with nonnegative matrices replaced by Q-matrices of the chains. Connections with quasi-stationarity and quasi-ergodicity of absorbing Markov chains are discussed. An interpretation for the corresponding variational solutions is given.

Keywords: decay parameter; Markov chain; quasi-ergodicity; quasi-stationarity

1. Introduction

Let $P(t) = \{P_{ij}(t), i, j \in E, t \geq 0\}$ be a standard transition functions on a countable state space E, that is, it has the following properties:

(i) $P_{ij}(t) \geq 0$ and $\lim_{t\to 0} P_{ij}(t) = \delta_{ij} = P_{ij}(0)$ $\forall i, j \in E$.
(ii) $\sum_{j\in E} P_{ij}(t) \leq 1$ for all $t \geq 0$, $i \in E$.
(iii) (*Chapman–Kolmogorov equation, or the semigroup property*)

$$P_{ij}(t+s) = \sum_{k\in E} P_{ik}(s)P_{kj}(t) \qquad \text{for all } s, t \geq 0 \text{ and } i, j \in E.$$

$P(t)$ is said to be conservative or honest if $\sum_{j\in E} P_{ij}(t) = 1$ for all $i \in E$ and all $t \geq 0$. In this case, it is well known that there is a Markov chain $X = \{X_t, t \geq 0\}$ on E, with transition function $P(t)$.

If $P(t)$ is non-conservative, we add an extra state, say 0, to E, and extend $P(t)$ to $\tilde{E} = E \cup \{0\}$ so that the extension, still denoted as $P(t)$, is conservative on $\tilde{E}$. Then we have a Markov chain $X = \{X_t, t \geq 0\}$ on $\tilde{E}$ with transition function $P(t)$, and 0 is an absorbing state of X.

[†]Corresponding author.

Remark 1.1. (1) In this paper, we specify E to be $\{1, 2, \dots, N\}$ with $N \le \infty$;

(2) It is well known that if $P(t)$ is a standard transition function on E, then for each $i \in E$,

$$P_{ii}(t) > 0, \qquad \forall t \ge 0,$$

and for any $i, j \in E$ with $i \neq j$, either:

(2a) $P_{ij}(t) = 0\ \forall t > 0$, or
(2b) $P_{ij}(t) > 0\ \forall t > 0$,

cf. [7], Theorem 1.3, see also [1]. In this paper we assume (2b) for every pair $i \neq j$ in E (in this sense, we may call $P(t)$, or equivalently the chain X, irreducible on E).

Under the above assumptions, a fundamental result proved in [17] is that there exists the limit

$$-\lambda_1 = \lim_{t\to\infty} \frac{1}{t} \log P_{ij}(t), \tag{1.1}$$

which is independent of i and j. If the chain X is transient or null recurrent on E, then for each pair $i, j \in E$ with $i \neq j$,

$$P_{ij}(t) \le C_{ij} e^{-\lambda_1 t} \qquad \forall t \ge 0$$

for some constant C_{ij}. In each of these cases, λ_1 characterizes the exponential decay rate of the process, thus is called the decay parameter of $P(t)$, or equivalently of the chain X. It is of the same significance as the exponential ergodic rate for an ergodic chain. If the chain X is ergodic with the unique stationary distribution π, denote by λ_1^* the largest constant for which

$$\| P_{i\cdot}(t) - \pi \|_{\mathrm{TV}} \le C_i e^{-\lambda_1^* t} \qquad \forall t > 0$$

for each $i \in E$ with some constant C_i, where $\|\mu\|_{\mathrm{TV}}$ denote the total variation of the signed measure μ. λ_1^* is shown to be closely related to the spectral gap of the corresponding generator of the chain, and has been well studied. To be more precise, let Q be the matrix of transition rates of the chain (called the Q-matrix, see [1], page 64, for its definition and for more details). If the chain is conservative, then 0 is a trivial eigenvalue of Q. Let σ^* be the spectral gap of Q. It is proved that if the chain is reversible, then

$$\lambda_1^* \ge \sigma^*,$$

with equality holds in certain specific cases, see [4], Chapter 9, [5], Chapter 8, [6,9] and the references cited therein. To see the significance of λ_1 in the non-conservative case, we first focus on the case of finite E. If we denote by ρ the Perron–Frobenius eigenvalue of the nonnegative matrix e^Q, then it is easy to check that $\lambda_1 = -\log \rho$. Furthermore, since Q is a *Metzler–Leontief* matrix (cf. [24], page 45), it is shown that λ_1 is the principal eigenvalue of $-Q$: λ_1 is an eigenvalue of $-Q$, and any other eigenvalue λ satisfies that $\operatorname{Re}\lambda > \lambda_1$. The well-known Perron–Frobenius theorem gives further significant results:

(i) For the Perron–Frobenius eigenvalue ρ of a nonnegative (finite) matrix P, there exist the strictly positive left and right eigenvalues $\mathbf{u}'$ and $\mathbf{v}$ which are unique up to constant multiples.

(ii) If P is primitive, then

$$\lim_{n\to\infty}\rho^{-n}P^n = \mathbf{vu}' \tag{1.2}$$

see [24], Chapter 1. As for the case of infinite E, see Theorem 2.3 in the Section 2.1. It is also noted that λ_1 is closely linked to the principal eigenvalue of the generator of the process. To describe it, define

$$\lambda^* = -\lim_{t\to\infty}\frac{\log\|P(t)\|_u}{t}, \tag{1.3}$$

where $\|P(t)\|_u$ is the norm of the operator $P(t)$ on $C_b(E)$. Then the spectrum of $-Q$ is contained in $\{\lambda : \mathrm{Re}\,\lambda \geq \lambda^*\}$. From the definition of λ_1 it can be easily seen that

$$\lambda_1 \geq \lambda^*. \tag{1.4}$$

In many typical cases, λ^* is really in the spectrum. In such cases, λ^* is regarded as the principal eigenvalue of $-Q$. The simplest but nontrivial case is the one where E is finite, as we have just discussed. In this case, the equality holds in (1.4). Another typical case is that when E is an open set with compact closure in some connected manifold, and $P(t)$ is the semigroup associated with a certain elliptic second order differential operator L on E with Dirichlet boundary conditions. In this case, it is shown in [14] that λ^* is in the spectrum of $-L$ and represents the decay rate of $P(t)$. Besides, it is also known that λ_1 is relevant to existence of a quasi-stationary distribution (*QSD*) of the process (cf. [27] and [20]). Thus computation or estimate of λ_1 is of great interests. A classical result is for the Perron–Frobenius eigenvalue ρ of a nonnegative matrix $A = (a_{i,j})$, which provides two variational formulas for ρ in dual form:

$$\rho = \sup_{u>0}\inf_i \frac{\sum_j a_{i,j}u_j}{x_i} = \inf_{u>0}\sup_i \frac{\sum_j a_{i,j}u_j}{u_i} \tag{1.5}$$

(cf. [24] or Appendix D in [22]). An extension to the case of an infinite E will be given in Section 4, see (4.7). A number of papers had worked on representation for λ_1 in continuous time cases, some powerful formulas were obtained in certain concrete cases. For example, for a birth–death process with birth rates $\{b_n\}$ and death rates $\{d_n\}$, it was proved in [26] that

$$\lambda_1 = \sup_{u\in\mathcal{U}}\inf_n\left\{b_n + d_n - \frac{b_{n-1}d_n}{u_n} - u_{n+1}\right\}, \tag{1.6}$$

where $\mathcal{U} = \{u = (u_j, j \geq 1) : u_j > 0\,\forall j \geq 1\}$ is the set of all strictly positive functions on E. This variational representation can provide good lower bounds for λ_1. [8,9] and [6] provide some other representations and powerful approximation procedures for λ_1.

The aim of the present paper is to provide representation for λ_1 for more general chains. The formula we are going to derive is motivated by the fundamental works of [13] and [14]. It is proved in [13] that if L is an elliptic second order differential operator L on a compact metric space E, and $P(t)$ is the associated semigroup, then for any continuous function V on E,

$$-\lambda_V = \sup_{\mu \in M_1(E)} \left[\int V \, d\mu - I(\mu) \right], \tag{1.7}$$

where λ_V is the principal eigenvalue of $L + V$, $M_1(E)$ is the space of probability measures on E, and I is defined by

$$I(\mu) = -\inf_{u \in \mathcal{D}(L)^+} \int \frac{Lu}{u} \, d\mu, \tag{1.8}$$

where $\mathcal{D}(L)$ is the domain of L and $\mathcal{D}(L)^+ = \{u \in \mathcal{D}(L) : \inf_x u(x) > 0\}$. A similar formula is proved in [14] when E is an open set with compact closure in R^d (or in a more general connected manifold without boundary), and L is an elliptic second order differential operator on E with Dirichlet boundary conditions. The significance of (1.7) is that it generalizes the following classical variational formula for λ_V when L is self-adjoint w.r.t. a measure ν on E:

$$-\lambda_V = \sup_{f \in L^2(\nu), \|f\|_2 = 1} \left[\int V f^2 \, d\nu + (Lf, f) \right]. \tag{1.9}$$

The formula (1.7) involves the functional I which is typically used as the large deviation rate function for the family $\{L_t, t \geq 0\}$ of empirical measures of a Markov process $\{X_t, t \geq 0\}$ on an arbitrary topological space E:

$$L_t = \frac{1}{t} \int_0^t \delta_{X_s} \, ds,$$

where δ_x is the Dirac measure centered at x. A large deviation principle for $\{L_t\}$ with the rate function I states that

$$-\inf_{\mu \in A^\circ} I(\mu) \leq \liminf_{t \to \infty} \frac{1}{t} \log P_\pi(L_t \in A) \leq \limsup_{t \to \infty} \frac{1}{t} \log P_\pi(L_t \in A) \leq -\inf_{\mu \in \bar{A}} I(\mu)$$

for every subset A of $M_1(E)$ which is Borel measurable with respect to some preassembled topology on $M_1(E)$, A° and $\bar{A}$ are the interior and closure of A respectively, π is the initial distribution of the process. The above inequalities can be roughly expressed as

$$P(L_t \in d\mu) \sim e^{-tI(\mu)}. \tag{1.10}$$

Thus $I(\mu)$ gives the exponential decay rate of the probability that L_t being close to an anomalous state μ. Establish of a large deviation principle for $\{L_t\}$ and construction of the rate function involves the computation of λ_V (which is just λ_1 when E is finite and $V = 0$ as we have pointed out), see Section 7 in [25] and Section 3.1, Section 6.5 in [10]. For explicit description of a large deviation principle, its more general theory and applications, see [10–12,25] and the recent book [22]. Note that the principal eigenvalue λ_V in both [13] and [14] is defined in the same way as the one for λ^* by (1.3), with $P(t)$ replaced by the Feynman–Kac semigroup $P^V(t)$ determined by $L + V$. This definition indicates some connection between λ_V and large deviation behavior of the associate Markov process. [13] and [14] reveal such a connection in terms of the functional I. The proofs of the results are purely analytical, no relying on any large deviation result. In [4], we investigated this problem for Markov chains with countably infinite states from the point of view of large deviations. We proved the large deviation principle for the empirical processes of such a chain under some standard tightness condition, and derived some relationships among λ_1, λ_V, some other parameters and quasi-stationary distributions. In the present paper, we are going to handle this problem without the aid of large deviation results, providing variational formulas for λ_1 similar to that for λ_V (which is λ^* when $V = 0$) given in [13] and [14]. We note that λ_1 may differ from λ^* in general. The approach we will adopt is different from that of [13] and [14]: besides a standard truncation argument, we will use suitably tilted transition functions to obtain some linear representations for the perturbed I-functional, see Theorem 3.4. The main variational formula we will prove is that

$$\lambda_1 = \inf_{\mu \in M_1(E)} I(\mu). \tag{1.11}$$

Similar formula will also be given when the chain is perturbed by some function V, see Theorem 3.8. This variational representation has at least the following features:

(1) If the chain is reversible, then (1.11) admits a more explicit form in terms of the matrix $Q = (q_{ij})$ of transition rates:

$$\lambda_1 = \frac{1}{2} \inf_{\mu \in M_1(E)} \sum_{ij} [\sqrt{\mu_i q_{ij}} - \sqrt{\mu_j q_{ji}}]^2 \tag{1.12}$$

(see Section 2.2), which corresponds to (1.9) for a self-adjoint L. For example, for a birth–death chain with birth rates $\{b_n\}$ and death rates $\{d_n\}$,

$$\lambda_1 = \frac{1}{2} \inf_{\mu \in M_1(E)} \sum_i [\mu_i b_i + \mu_{i+1} d_{i+1} - 2\sqrt{\mu_i \mu_{i+1} b_i d_{i+1}}]^2; \tag{1.13}$$

(2) (1.11) applies to general chains, which will lead to the following “dual” max-min and min-max variational formulas:

$$-\lambda_1 = \sup_{\mu \in M_1^q(E)} \inf_{u \in \mathcal{U}} \int \frac{Qu}{u} \, d\mu = \inf_{u \in \mathcal{U}} \sup_{\mu \in M_1^q(E)} \int \frac{Qu}{u} \, d\mu, \tag{1.14}$$

where $\mathcal{U}$ is as in (1.6),

$$(Qu)_i = \sum_j q_{ij} u_j = \sum_{j \neq i} q_{ij} (u_j - u_i),$$

with $q_i = \sum_{j \neq i} q_{ij}$ and

$$M_1^q(E) = \left\{ \mu \in M_1(E) : \mu(q) \triangleq \sum_i q_i \mu_i < \infty \right\},$$

see Corollary 3.3. The dual form is different from that in (1.5) which is not a minimax form. When E is finite, a version of (1.5) in terms of the Q-matrix is given in Section 4, see (4.10). To see the significance of (1.14), we note that for nonnegative infinite matrices, dual formulas for the Perron–Frobenius eigenvalue ρ of exactly the same form as (1.5) were unknown. A version was known as in (4.4), for which two sets of functions $\mathcal{U}$ and $\mathcal{U}_\alpha$ are involved. Applying (1.14), we give in Section 4 a version of it for ρ. It is also worth noting that slightly different from (1.8) for defining the functional I, the matrix Q in stead of the generator L is used in the integral in (1.14), with $\mathcal{D}(L)^+$ replaced by $\mathcal{U}$. This makes the dual formula more applicable in deriving both lower and upper bounds of λ_1. As a consequence, it follows that

$$-\lambda_1 = \inf_{u\in\mathcal{U}} \sup_i \frac{(Qu)_i}{u_i} = \inf_{u\in\mathcal{U}} \sup_i \left[\sum_{j\neq i} q_{ij}\frac{u_j}{u_i} - q_i\right] \tag{1.15}$$

which can be regarded as a generalization of (1.6) to general chains. When applied to a birth–death chain with birth rates $\{b_n\}$ and death rates $\{d_n\}$, it follows that

$$\lambda_1 = \sup_{u\in\mathcal{U}} \inf_n \left\{b_n + d_n - b_n\frac{u_{n+1}}{u_n} - d_n\frac{u_{n-1}}{u_n}\right\} \tag{1.16}$$

which is another version of (1.6);

(3) The functional I has been found to be important in analyzing long-term behaviors of a Markov process. It connects many important concepts. Besides its close relationship with the principal eigenvalue of the generator as described above, (1.10) also indicates its connection with stationary distributions and ergodicity of the Markov process. Indeed, it is known that $I(\mu) = 0$ iff μ is stationary for the process (cf. [25], Corollary 7.26). Thus, if the stationary distribution is unique and the large deviation (1.10) is verified, then L_t will converge to this unique stationary, see [11–13] and [14]. Equation (1.11) allows one to explore the power of I to derive consequences concerning decay rates and quasi-ergodic behaviors of the (sub-)Markov processes under consideration. For example, we give in Section 3 necessary and sufficient condition for $I(m) = \inf_{\mu\in M_1(E)} I(\mu) = \lambda_1$ for some (and then unique) $m \in M_1(E)$. These conditions concern the λ_1-*positivity* of $P(t)$ and the λ_1 invariant measures and vectors, and such a minimizer m is a certain quasi-stationary distribution of the process, which differs from the classical ones that have been extensively studied for long time as surveyed in [20]. We will call such distribution m a quasi-ergodic distribution, since we will show in Sections 3 and 4 that it is closely related to the quasi-ergodic behavior of the time average of the process, more precisely, m is the quasi-limit of $\{L_t, t \geq 0\}$, see Theorem 3.6 and (3.15). We also prove that if $I(\nu) = 0$ for some $\nu \in M_1(E)$, then the process is conservative.

We point out that variational formulas like the ones studied in this paper can be found in many other settings. Besides the classical ones given in (1.5) for the Perron–Frobenius eigenvalue of a nonnegative matrix, [14] also derived a dual form (a kind different from (1.5)) for the principal eigenvalue λ_V described previously in this section. Similar variational formulas appear in even more complicated situations. For example, in the context of particles in a random medium, [18] and [19] derived a variational formula for the effective Hamiltonian used in describing the solution to certain HJB equations. For random walks in a random environment, [21] and [23] provide variational formulas for the limiting quenched free energy. In [16], variational formulas are derived for the limiting time constant of last-passage percolation and for the limiting free energy of directed polymers both evolving in certain random media. When the configuration space is finite, [16] explained the connection of the variational formulas with the Perron–Frobenius theorem in more details.

The next section contains some preliminaries for the decay parameter λ_1 and the functional I. The derivation of the variational formulas in terms of the functional I are given in Section 3. Existence and an interpretation of the corresponding variational solution are also provided. In Section 4 we study the connections of the problems considered in Section 3 with related ones for discrete time chains in two directions. In one direction, we apply the results obtained in Section 3 to give dual variational formulas for the Perron–Frobenius type eigenvalues of nonnegative infinite matrices (see (4.7)). The dual form is different from the classical one in the Perron–Frobenius theory (see (1.5)). In the other direction, for finite Markov chains in continuous-time, we also give a second form of dual formulas for λ_1, see (4.10), which can be regarded as a version of (1.5) for the Perron–Frobenius eigenvalue, with nonnegative matrices replaced by Q-matrices of the chains.

2. Preliminaries

2.1. The decay parameter

Recall that we are considering an irreducible (sub-)Markov transition function $P(t)$ on a countable state space E which we specify to be $\{1, 2, \dots, N\}$ with $N \le \infty$. X is the associated Markov chain on E. The decay parameter λ_1 is defined by (1.1) whose existence is established in [17]. It is easy to check that

$$\lambda_1 = \inf\left\{\alpha \ge 0 : \int_0^\infty P_{ii}(t)e^{\alpha t}\,dt = \infty\right\}. \tag{2.1}$$

$P(t)$ (or X) is classified as λ_1-recurrent or λ_1-transient depending on whether $\int_0^\infty P_{ii}(t)e^{\lambda_1 t}\,dt = \infty$ or $< \infty$. λ_1-recurrent chains are further divided into λ_1-positive or λ_1-null, depending on whether $\lim_{t\to\infty} e^{\lambda_1 t}P_{ij}(t) > 0$ or $= 0$, respectively (cf. [1], Section 5.2).

Remark 2.1. (1) If $P(t)$ is honest and positively recurrent, then $\lambda_1 = 0$. Thus in this case, 0-positivity is the usual positivity. On the other hand, if the chain is null-recurrent or transient, it is also possible that $\lambda_1 = 0$. Examples can be easily constructed by applying the fact that

$$\lambda_1 \le \inf_i q_i,$$

where $q_i = \sum_{j\ne i} q_{ij}$ (see Theorem 1.9 in [1], page 164).

(2) If $\lambda_1 > 0$, then the chain is transient. In this situation, the chain can be honest. The following example can be found in [1]. Let Q be defined on the state space $E = \mathbb{Z}$ by

$$q_{i,i+1} = pc, \qquad q_{ii} = -c, \qquad q_{i,i-1} = qc,$$

where $0 < p < 1, q = 1 - p$, and $c > 0$. Then the process is irreducible and the decay parameter is $\lambda_1 = (1 - 2\sqrt{pq})c$. Thus if $p \neq \frac{1}{2}$, $\lambda_1 > 0$.

(3) From the irreducibility assumption (2b) made in Remark 1.1 and (1.2), it is easy to see that any irreducible finite chain (i.e., when $N < \infty$) is λ_1-positive. This can also be checked directly.

Concerning λ_1, we summarize some relevant definitions and known results as follows, which can be regarded as extension of the Perron–Frobenius theorem in the case of countably infinite states. more details can be found in [1,24,28] and [29].

Definition 2.2. *For a given $\lambda \geq 0$, a collection of strictly positive numbers, $\alpha = \{\alpha_i, i \in E\}$, is called a λ-subinvariant measure for $P(t)$, if*

$$\sum_{j \in E} \alpha_j P_{ji}(t) \leq e^{-\lambda t} \alpha_i \tag{2.2}$$

for all $t \geq 0$ and all $i \in E$. α is called λ-invariant if equality holds in (2.2) for all $i \in E$. A collection $\beta = \{\beta_i, i \in E\}$ of strictly positive numbers is called a λ-subinvariant vector for $P(t)$, if

$$\sum_{j \in E} P_{ij}(t) \beta_j \leq e^{-\lambda t} \beta_i \tag{2.3}$$

for all $t \geq 0$ and all $i \in E$. β is called λ-invariant if equality holds for all $i \in E$.

Theorem 2.3. (1) *Nonnegative numbers $x = \{x_i, i \in E\}$ satisfying (2.2) or (2.3) are either all zero, or all strictly positive.*

(2) *For $\lambda \geq 0$, there exists a λ-subinvariant measure and a λ-subinvariant vector iff $\lambda \leq \lambda_1$.*

(3) *If $P(t)$ is λ_1-recurrent, then the λ_1-subinvariant measure and the λ_1-subinvariant vector are both unique up to constant multiples, and in fact are both λ_1-invariant. Furthermore, $P(t)$ is λ_1-positive if and only if the λ_1-invariant measure $\{\alpha_i, i \in E\}$ and λ_1-invariant vector $\{\beta_i, i \in E\}$ are strictly positive and satisfy*

$$\sum_{k \in E} \alpha_k \beta_k < +\infty. \tag{2.4}$$

Moreover, in this case,

$$\lim_{t \to \infty} e^{\lambda_1 t} P_{ij}(t) = \frac{\alpha_j \beta_i}{\sum_{k \in E} \alpha_k \beta_k}. \tag{2.5}$$

Remark 2.4. From the above theorem, we see that

$$\begin{aligned} \lambda_1 &= \sup\bigl\{r : \exists x \in \mathcal{U} \text{ as a column, s.t. } P(t)x \leq e^{-rt} x, \forall t \geq 0\bigr\} \\ &= \sup\bigl\{r : \exists y \in \mathcal{U} \text{ as a row, s.t. } yP(t) \leq e^{-rt} y, \forall t \geq 0\bigr\}. \end{aligned} \tag{2.6}$$

λ_1 can also be regarded as an eigenvalues of the Q-matrix Q. This can be seen from the following proposition (see Proposition 2.13 in [1], page 88). We recall that the minimal Q-function for a given Q-matrix Q is the minimal nonnegative solution to the corresponding backward (or forward) Kolmogorov equation, see Section 2 of [1] or Section 2.2 of [7] for details.

Proposition 2.5. *Let Q be a Q-matrix and $P_{ij}(t)$ be the minimal Q-function. $\{x_i, i \in E\}$ (resp. $\{y_i, i \in E\}$) is a non-negative column (resp. row) and c is a real number. Then the following statements are equivalent*:

(1) $P(t)x \le e^{ct}x$ (*resp.* $yP(t) \le e^{ct}y$);
(2) $Qx \le cx$ (*resp.* $yQ \le cy$).

Remark 2.6. From this proposition, it follows that if an irreducible transition function $P_{ij}(t)$ associated with the matrix Q is minimal, then the decay parameter λ_1 of $P_{ij}(t)$ satisfies

$$\begin{aligned}\lambda_1 &= \sup\bigl\{r : \exists \text{ positive column } \{x_i, i \in E\},\ \text{s.t. } Qx \le -rx\bigr\} \\ &= \sup\bigl\{r : \exists \text{ positive row } \{y_i, i \in E\},\ \text{s.t. } yQ \le -ry\bigr\}.\end{aligned} \tag{2.7}$$

Therefore, for every irreducible Q-matrix Q, we can define a parameter by (2.7). For convenience, we also call it the decay parameter of Q. In particular, if $P_{ij}(t)$ is λ_1-recurrent, then $-\lambda_1$ is an eigenvalue of Q. Furthermore, if E is finite, then Q is an irreducible ML-matrix and $-\lambda_1$ is the principal eigenvalue of Q as we have pointed out in Section 1.

2.2. Donsker–Varadhan's *I*-functional

We recall the definition (1.8) of Donsker–Varadhan's I-functional for a (sub-)Markov process X with transition function $P(t)$ and generator L. It is easy to see from its definition that I is convex and lower semi-continuous, see also Section 7 of [25]. Furthermore, if $P(t)$ is conservative, then $I(\mu) = 0$ iff μ is invariant for $P(t)$ as we have pointed out in Section 1. To explore further the power of I, we first note that on $\mathcal{D}(L)^+$, $Lu = Qu$ with Qu being defined as in Section 1, that is,

$$(Qu)_i = \sum_{j \ne i} q_{ij}(u_j - u_i), \qquad i \in E.$$

Thus,

$$I(\mu) = -\inf_{u \in \mathcal{D}(L)^+} \int \frac{Qu}{u}\, d\mu.$$

Though the domain, $\mathcal{D}(L)$ is generally not known explicitly, the following lemma makes it possible to avoid the use of $\mathcal{D}(L)$.

Lemma 2.7. *Let E be a Polish space, $\{P(t), t \ge 0\}$ be a Markov semigroup (which may be non-conservative) on $C_b(E)$ with generator L. I is defined by (1.8). For each $h > 0$ and $\mu \in M_1(E)$, we define*

$$I_h(\mu) = -\inf_{u \in \mathcal{U}_b^+} \int \log \frac{P(h)u}{u}\, d\mu, \tag{2.8}$$

where, as defined in Section 1, $\mathcal{U}$ is the set of all strictly positive functions on E, and $\mathcal{U}_b^+ = \{u \in \mathcal{U} : 0 < \inf_{x\in E} u(x) \le \sup_{x\in E} u(x) < \infty\}$. Then for each $\mu \in M_1(E)$ and $h > 0$, we have

$$I_h(\mu) \le hI(\mu) \quad \textit{and} \quad I(\mu) = \lim_{h\to 0^+} \frac{1}{h} I_h(\mu). \tag{2.9}$$

Proof. Such a result is proved in [12] for every compact metric space E, see Lemma 3.1 in that paper. If we notice that in (1.8), each $u \in \mathcal{D}(L)^+$ has a strictly positive lower bound, and that L is the (strong) generator of $\{P(t), t \ge 0\}$, it can be seen that the proof of Lemma 3.1 of [12] is valid in the present situation. We provide the necessary details as follows. On the one hand, for $u \in \mathcal{D}(L)^+$ and $\mu \in M_1(E)$, define

$$\Phi(t) = \int_E \log \frac{P(t)u}{u}\, d\mu.$$

Since $dP(t)u/dt = LP(t)u = P(t)Lu$, we see that

$$\frac{d\Phi(t)}{dt} = \int_E \frac{LP(t)u}{P(t)u}\, d\mu \ge -I(\mu).$$

Now since $\Phi(0) = 0$, it follows for any $h > 0$ that

$$\Phi(h) = \int_0^h \frac{d\Phi(t)}{dt}\, dt \ge -\int_0^h I(\mu)\, dt = -hI(\mu).$$

The inequality in (2.9) then follows which further implies that

$$\limsup_{h\to 0+} \frac{1}{h} I_h(\mu) \le I(\mu).$$

On the other hand, since each $u \in \mathcal{D}(L)^+$ is bounded away from 0 and

$$\lim_{h\to 0} \frac{P(h)u - u}{h} = Lu$$

in $C_b(E)$, so $P(h)u = u + hLu + o(h)$ with $o(h)$ uniform in E. Thus

$$\log \frac{P(h)u}{u} = \log\left[1 + h\frac{Lu}{u} + o(h)\right] = h\frac{Lu}{u} + o(h),$$

which implies that

$$\frac{1}{h} I_h(\mu) \ge -\frac{1}{h}\int_E \log \frac{P(h)u}{u}\, d\mu = -\int_E \frac{Lu}{u}\, d\mu + o(1).$$

From this, we obtain

$$\liminf_{h\to 0+} \frac{1}{h} I_h(\mu) \ge I(\mu),$$

completing the proof of the equality in (2.9). □

Since we are considering Markov chains on countable state spaces, there is another way to avoid the use of $\mathcal{D}(L)$. We need to make the following modification of I:

$$J(\mu) = \begin{cases} -\inf_{u\in\mathcal{U}} \int \frac{Qu}{u}\, d\mu, & \mu \in M_1^q(E), \\ +\infty, & \text{otherwise}, \end{cases} \tag{2.10}$$

where as defined in Section 1, $M_1^q(E) = \{\mu \in M_1(E) : \mu(q) = \sum_i q_i \mu_i < \infty\}$. As we will show in Theorem 3.2 in the next section, the infimum of J over $M_1(E)$ is the same as that of I, which is exactly λ_1.

Remark 2.8. It should be noted that if the matrix Q is stable, that is, $q_i < \infty\ \forall i \in E$, then Qu/u is well defined for every $u \in \mathcal{U}$ since

$$\frac{(Qu)_i}{u_i} = \sum_{j:j\neq i} q_{ij}\frac{u_j}{u_i} - q_i > -\infty, \qquad \forall i \in E.$$

Furthermore, for $\mu \in M_1^q(E)$, the integral $\int \frac{Qu}{u}\, d\mu$ is also well defined for $u \in \mathcal{U}$ since

$$\int \frac{Qu}{u}\, d\mu = \sum_{i\neq j} \mu_i q_{ij} \frac{u_j}{u_i} - \mu(q) > -\infty.$$

Thus J is well defined, and it is easily seen that $I \le J$. Theorem 8.8 of [7] shows that for $\mu \in M_1^q(E)$, $I(\mu) = J(\mu)$.

The following proposition is an easy consequence of definition (2.10).

Proposition 2.9. *If $E_1 \subset E_2$ are two countable sets, $Q^{(1)}$ and $Q^{(2)}$ are two Q-matrices defined on E_1 and E_2, respectively, such that, $q_{ij}^{(1)} = q_{ij}^{(2)}$ for all $i, j \in E_1$. Let $J^{(1)}$ and $J^{(2)}$ be defined as* (2.10), *corresponding to $Q^{(1)}$ and $Q^{(2)}$, respectively. Then if $\mu \in M_1(E_1) \cap M_1(E_2)$, that is, if $\mu(E_2 \setminus E_1) = 0$, we have*

$$J^{(2)}(\mu) \le J^{(1)}(\mu). \tag{2.11}$$

Proof. Given $\mu \in M_1(E_1) \cap M_1(E_2)$ with $\mu(q^{(2)}) < \infty$, since $\mu(q^{(2)}) = \mu(q^{(1)})$ by the assumption, we have that

$$\begin{aligned} J^{(2)}(\mu) &= -\inf_{u\in\mathcal{U}(E_2)} \left[\sum_{i\in E_1} \sum_{j\in E_2:j\neq i} \mu_i q_{ij}^{(2)} \frac{u_j}{u_i} - \mu\big(q^{(2)}\big)\right] \\ &\le -\inf_{u\in\mathcal{U}(E_2)} \left[\sum_{i\in E_1} \sum_{j\in E_1:j\neq i} \mu_i q_{ij}^{(2)} \frac{u_j}{u_i} - \mu\big(q^{(2)}\big)\right] \\ &= -\inf_{u\in\mathcal{U}(E_1)} \left[\sum_{i\in E_1} \sum_{j\in E_1:j\neq i} \mu_i q_{ij}^{(1)} \frac{u_j}{u_i} - \mu\big(q^{(1)}\big)\right] = J^{(1)}(\mu). \end{aligned}$$

□

Remark 2.10. Summarizing the above discussion, we assume in this paper that the Q-matrix under consideration is stable, and the corresponding transition function is the minimal one.

3. The main results

Now we start presenting the main results of this article. Our first new observation is that conservativeness is necessary for the existence of a probability measure μ satisfying $I(\mu) = 0$.

Proposition 3.1. *Let $P(t)$ be an irreducible (sub-)Markov transition function. If there exists some $\mu \in M_1(E)$ such that $I(\mu) = 0$, then $P(t)$ is conservative and μ is its stationary distribution.*

Proof. Fix an $h > 0$. By adding an additional state 0, we define $\tilde{E} = E \cup \{0\}$ and let

$$\tilde{P}(h) = \begin{pmatrix} 1 & 0 \\ \rho(h) & P(h) \end{pmatrix},$$

where $\rho(h)$ is chosen so that $\tilde{P}(h)$ is conservative. Let $\tilde{I}_h$ and I_h be defined as in (2.8) for $\tilde{P}(h)$ and $P(h)$ respectively. Then, viewing $\mu \in M_1(E)$ as a probability measure on $\tilde{E}$ supported on E, it is easy to check that $\tilde{I}_h(\mu) \le I_h(\mu)$. Thus, if $I(\mu) = 0$ for some $\mu \in M_1(E)$, then by Lemma 2.7 we see that $\tilde{I}_h(\mu) = 0$. Therefore, as we have pointed out in Section 1, it follows from Corollary 7.26 in [25], page 142, that

$$\mu P_h = \mu \tilde{P}_h = \mu. \tag{3.1}$$

A standard argument then shows that $P(h)$ is conservative: Write (3.1) in its components as follows:

$$\sum_i \mu_i P_{ij}(h) = \mu_j. \tag{3.2}$$

The irreducibility of $P(t)$ implies that $\mu_j > 0$ for every j. Then summing over j in (3.2) we get the desired conclusion. □

The next theorem gives the fundamental variational representation for λ_1 in terms of I and J.

Theorem 3.2. *For any irreducible transition function on E,*

$$\inf_{\mu \in M_1(E)} I(\mu) = \lambda_1 = \inf_{\mu \in M_1(E)} J(\mu). \tag{3.3}$$

Recall the definition of J, we have the following corollary to the above theorem, which is more applicable in estimating λ_1.

Corollary 3.3. *For any irreducible transition function on E,*

$$-\lambda_1 = \sup_{\mu \in M_1^q(E)} \inf_{u \in \mathcal{U}} \int \frac{Qu}{u}\, d\mu \tag{3.4}$$

$$= \inf_{u \in \mathcal{U}} \sup_{\mu \in M_1^q(E)} \int \frac{Qu}{u}\, d\mu = \inf_{u \in \mathcal{U}} \sup_{j \in E} \frac{(Qu)_j}{u_j}. \tag{3.5}$$

Proof. The first equality follows from Theorem 3.2 and the definition of J. The third one follows easily from the fact that

$$\sup_{\mu \in M_1^q(E)} \int \frac{Qu}{u}\, d\mu = \sup_{j \in E} \frac{(Qu)_j}{u_j}.$$

Thus we only need to prove the second equality. To this end, we note that on the one hand, it is trivial that the *lhs* $\leq$ *rhs*. On the other hand, by Theorem 2.3 and Remark 2.4, there exists a $\tilde{u} \in \mathcal{U}$ such that

$$\frac{(Q\tilde{u})_i}{\tilde{u}_i} \leq -\lambda_1.$$

Thus

$$-\lambda_1 \geq \sup_{\mu \in M_1^q(E)} \int \frac{Q\tilde{u}}{\tilde{u}}\, d\mu \geq \inf_{u \in \mathcal{U}} \sup_{\mu \in M_1^q(E)} \int \frac{Qu}{u}\, d\mu,$$

which is precisely what we need. □

We now start preparing to prove Theorem 3.2. Since

$$\inf_{\mu \in M_1(E)} I(\mu) \leq \inf_{\mu \in M_1(E)} J(\mu) = \inf_{\mu \in M_1^q(E)} J(\mu),$$

we only need to prove the following two inequalities:

$$\inf_{\mu \in M_1(E)} I(\mu) \geq \lambda_1 \quad \text{and} \quad \inf_{\mu \in M_1(E)} J(\mu) \leq \lambda_1. \tag{3.6}$$

The proof of the first inequality involves studying a series of transition functions tilted with respect to the original one. For the proof of the second inequality, a truncation and approximation procedure is used.

The approach by properly tilting transition functions is a typical technique in the study of large deviations. It is also a standard approach in the study of recurrence of Markov and sub-Markov processes, and is often known as h-transformation. For example, it was used in [17] to study decay rates of Markov chains (see also [1]). By using such an approach here, we obtain a linear relationship among the corresponding I functionals, which may be of independent interests. To define the tilted transition functions, let $\beta = \{\beta_i, i \in E\}$ be a (positive) λ_1-subinvariant vector.

Then for each $0 \le \theta \le \lambda_1$, β is also a θ-subinvariant vector, thus we can define a tilted (sub-Markov) transition function $\{P^\theta(t), t \ge 0\}$ as follows:

$$P^\theta_{ij}(t) = \frac{e^{\theta t} P_{ij}(t)\beta_j}{\beta_i}, \qquad i, j \in E, t \ge 0. \tag{3.7}$$

The Q-matrix Q^θ of this new transition function is given by

$$q^\theta_{ij} = \frac{\beta_j}{\beta_i}(q_{ij} + \theta\delta_{ij}), \qquad i, j \in E. \tag{3.8}$$

Let λ^θ_1 and I^θ be the corresponding decay parameter and I functional, respectively. Then we have:

Theorem 3.4. *With the above notations*,

$$I^\theta = I - \theta. \tag{3.9}$$

Proof. Let I^θ_h be defined as in (2.8) for $P^\theta(h)$. Then by Lemma 2.7, it suffices to prove that

$$I^\theta_h = I_h - \theta h. \tag{3.10}$$

To this end, we note that it is easy to check that

$$\begin{aligned} I^\theta_h(\mu) &= -\inf_{u \in \mathcal{U}^+_b} \int \log \frac{P^\theta(h)u}{u}\, d\mu \\ &= -\inf_{u \in \mathcal{U}^+_b} \int \log \frac{P(h)(u\beta)}{u\beta}\, d\mu - \theta h. \end{aligned}$$

For every $u \in \mathcal{U}^+_b$, there exists constants $0 < c_1 < c_2$, such that $c_1 < u_i < c_2$, for all $i \in E$. Define $u^{(n)} = (\frac{1}{n} \vee (u\beta)) \wedge n$ for $n \ge 1$, then $u^{(n)} \in \mathcal{U}^+_b$ and $u^{(n)} \to u\beta$ point-wise. Thus from the fact that

$$P_h\beta \le e^{-\theta h}\beta$$

coordinate-wise, we obtain

$$\begin{aligned} \log \frac{P(h)(u^{(n)})}{u^{(n)}} &= \log \frac{P(h)((1/n \vee (u\beta)) \wedge n)}{(1/n \vee (u\beta)) \wedge n} \\ &\le \log \frac{P(h)(1/n + (u\beta)) \wedge n}{(1/n \vee (u\beta)) \wedge n} \\ &\le \log \frac{(1/n + c_2 e^{-\theta h}\beta) \wedge n}{(1/n \vee (c_1\beta)) \wedge n} \\ &\le \log\left(1 + \frac{c_2}{c_1} e^{-\theta h}\right), \end{aligned}$$

where the last inequality can be directly verified by considering the following cases

$$(1)\quad c_1\beta_i \le \frac{1}{n};\qquad (2)\quad \frac{1}{n} < c_1\beta_i < n\quad \text{and}\quad (3)\quad c_1\beta_i \ge n$$

respectively. It then follows from Fatou's lemma that

$$-\int \log \frac{P(h)(u\beta)}{u\beta}\,d\mu \le -\limsup_{t\to\infty}\int \log \frac{P(h)u^{(n)}}{u^{(n)}}\,d\mu \le -\inf_{u\in\mathcal{U}_b^+}\int \log\frac{P(h)u}{u}\,d\mu.$$

Thus $I_h^\theta(\mu) \le I_h(\mu) - \theta h$.

To prove the reversed inequality, note that

$$\begin{aligned} I_h(\mu) &= -\inf_{u\in\mathcal{U}_b^+}\int \log\frac{P(h)u}{u}\,d\mu \\ &= -\inf_{u\in\mathcal{U}_b^+}\int \log\frac{P^\theta(h)(u/\beta)}{u/\beta}\,d\mu + \theta h \end{aligned}$$

and that

$$\left(P_h^\theta \beta^{-1}\right)_i = \sum_{j\in E} P_{ij}^\theta(h)\beta_j^{-1} = \beta_i^{-1}e^{\theta h}\sum_{j\in E} P_{ij}(h) \le e^{\theta h}\beta_i^{-1}.$$

An argument similar to the previous one gives that $I_h(\mu) \le I_h^\theta(\mu) + \theta h$, completing the proof. □

As will be seen soon, the first inequality in (3.6) follows from (3.9). To prove the second inequality in (3.6), we need a truncation and approximation procedure. Let (E_n) be an increasing sequence of finite subsets of E, such that

$$\varnothing \subset E_1 \subset E_2 \subset \cdots \subseteq E = \bigcup_n E_n.$$

The E_n-truncated Q-matrix $Q_n = (q_{ij}^{(n)})_{i,j\in E_n}$, is defined by

$$q_{ij}^{(n)} = q_{ij}, \qquad i, j \in E_n.$$

Associated with the matrix $(q_{ij}^{(n)})_{i,j\in E_n}$, there is a unique (and hence minimal) transition function $P_{ij}^{(n)}(t)$. If it is irreducible, we denote its decay parameter denoted by $\lambda_1^{(n)}$. The following lemma will be used in the proof of Theorem 3.2.

Lemma 3.5 ([3], Section 3, Lemma 1 and Section 4, Lemma 2). *There exists an increasing sequence (E_n) of finite subsets of E, such that the transition function $P_{ij}^{(n)}(t)$ associated with the matrix $(q_{ij}^{(n)})_{i,j\in E_n}$ is irreducible, and $\lambda_1 = \lim_{n\to\infty}\downarrow \lambda_1^{(n)}$.*

Now we can prove Theorem 3.2.

Proof of Theorem 3.2. First, applying Theorem 3.4 we see that for any $\mu \in M_1(E)$

$$I(\mu) - \lambda_1 = I_1^{\lambda}(\mu) \geq 0,$$

this implies that

$$\inf_{\mu \in M_1(E)} I(\mu) \geq \lambda_1.$$

To prove the second inequality in (3.6), we note that the truncated chain $P^{(n)}(t)$ is $\lambda_1^{(n)}$-positive. This implies that the $\lambda_1^{(n)}$-tilted chain $P^{n,\lambda_1^{(n)}}(t)$ is conservative and ergodic. Thus, still denote by I the I-functional for $P^{(n)}$ without confusion, there is a unique $\mu^{(n)} \in M_1(E_n)$ such that

$$\inf_{\mu \in M_1(E_n)} I(\mu) = I\big(\mu^{(n)}\big) = I^{\lambda_1^{(n)}}\big(\mu^{(n)}\big) + \lambda_1^{(n)} = \lambda_1^{(n)}.$$

Now from Proposition 2.9, it follows that

$$\inf_{\mu \in M_1(E)} J(\mu) \leq \inf_{\mu \in M_1(E_n)} J(\mu) \leq \inf_{\mu \in M_1(E_n)} J^{(n)}(\mu) = \inf_{\mu \in M_1(E_n)} I^{(n)}(\mu) = \lambda_1^{(n)}.$$

Letting $n \to \infty$ and applying Lemma 3.5 we conclude that

$$\inf_{\mu \in M_1(E)} J(\mu) \leq \lambda_1,$$

which is the desired assertion. □

Theorem 3.4 can be further applied to study when and where the infimum in (3.3) is attained. To see this, let $\alpha = (\alpha_1, \alpha_2, \dots)$ and $\beta = (\beta_1, \beta_2, \dots)$ be the λ_1-subinvariant measure and vector, respectively. Denote $\alpha\beta = (\alpha_1\beta_1, \alpha_2\beta_2, \dots)$.

Theorem 3.6. *With the above notation, the following assertions are equivalent:*

(a) *$P(t)$ is λ_1-positive.*
(b) *α, β are unique and λ_1-invariant, and $\alpha\beta$ is summable.*
(c) *The infimum $\inf_{\mu \in M_1(E)} I(\mu) = \lambda_1$ is attained at some $\mu \in M_1(E)$, which is unique and given by the normalized $\alpha\beta$.*

Proof. The equivalence between assertions (a) and (b) is given by Theorem 2.3. To prove that they are equivalent to assertion (c), we note that if $P(t)$ is λ_1-positive, then $P^{\lambda_1}(t)$ is conservative and positively recurrent, with the normalized $\alpha\beta$, denoted by μ, as its unique stationary distribution. Thus $I^{\lambda_1}(\mu) = 0$. From this and (3.9), assertion (c) follows. On the other hand, if assertion (c) holds, then (3.9) implies that $I^{\lambda_1}(\mu) = 0$ with μ being the normalized $\alpha\beta$. From Proposition 3.1, we see that P^{λ_1} is conservative with μ as its stationary distribution. This implies that $P^{\lambda_1}(t)$ is positively recurrent, or equivalently, $P(t)$ is λ_1-positive. □

According to this theorem, the infimum of $I(\mu)$ over $M_1(E)$ is attained at some $m \in M_1(E)$ iff $P(t)$ is λ_1-positive. The latter is equivalent to the ergodicity of the tilted chain $P^{\lambda_1}(t)$ and m

is its unique stationary distribution. In the following, we give an interpretation of m from the perspective of quasi-stationarity of the chain $P(t)$ itself. The motivation comes from the following consideration: If we define $I_\tau = I - \lambda_1$, then it is nonnegative and lower semi-continuous. The fact that $I(\mu) = 0$ iff μ is stationary for $P(t)$ leads to our study of the zeros of I_τ. Our investigation shows that a zero m of I_τ is different from the classical and well studied quasi-stationary distribution. We may call such a zero m a "fractional quasi-ergodic" or simply "quasi-ergodic distribution" of $P(t)$. To explain this more explicitly, let $X = \{X_t, t \ge 0\}$ be the Markov chain on $\tilde{E} = \{0\} \cup E$ with $E = \mathbb{N}$ as described in the beginning of Section 1. Let $\{P_i, i \in \tilde{E}\}$ be the corresponding Markov family. We assume that X is irreducible on E and that 0 is an absorbing state of it. Denote by

$$\tau = \inf\{t \ge 0 : X_t = 0\}$$

the absorption time. Suppose that $P_i(\tau > t) > 0, \forall i \in E, t \ge 0$, and absorbing is certain, meaning that for all $i \in E$,

$$P_i(\tau < \infty) = 1.$$

Let λ_1 be the decay parameter of $P(t) = \{P_{ij}(t), t \ge 0\}$ on E. To see the connection between the decay rate and the quasi-stationarity of X, we impose the following two hypotheses:

(H1) X is λ_1-positive.

(H2) The λ_1-*invariant measure* $\alpha = \{\alpha_i, i \in E\}$ given in Theorem 2.3 is finite, that is, $\sum_{j \in E} \alpha_i < \infty$.

According to Theorem 2.3, under these two hypotheses, we can normalize α and the λ_1-invariant vector β so that $\sum_{j \in E} \alpha_i = \sum_{j \in E} \alpha_i \beta_i = 1$. Denote $m_i = \alpha_i \beta_i$.

The relationship among λ_1, α, β and the quasi-stationary distribution are summarized in the following, see [1,15] and [2] for details.

Proposition 3.7. *Under assumptions* (H1) *and* (H2), *we have*:

(a) $\forall i \in E$,

$$\lim_{t \to \infty} e^{\lambda_1 t} P_i(\tau > t) = \beta_i, \tag{3.11}$$

and so $\lim_{t \to \infty} \frac{1}{t} \log P_i(\tau > t) = -\lambda_1$.

(b) α *is the unique quasi-stationary distribution* (QSD) *of* X *characterized by*

$$P_\alpha(X_t = j | \tau > t) = \alpha_j, \qquad \text{for all } j \in E \text{ and } t \ge 0.$$

Furthermore, α *is the unique Yaglom limit defined by*

$$\lim_{t \to \infty} P_i(X_t = j | \tau > t) = \alpha_j, \qquad \text{for all } i \in E. \tag{3.12}$$

According to Theorems 2.3 and 3.6, under (H1) and (H2),

$$\lambda_1 = \inf_{\mu \in M_1(E)} I(\mu) = I(m).$$

By definition, m is clearly different from α, it is known to be the unique doubly limiting quasi-stationary distribution of X, that is, $\forall i, j \in E$,

$$\lim_{s\to\infty}\lim_{t\to\infty} P_i(X_s = j|\tau > t+s) = m_j. \tag{3.13}$$

A substantially new difference between the interpretations of α and m is that m is the unique "fractional" Yaglom limit of X, that is, $\forall q \in (0, 1)$,

$$\lim_{t\to\infty} P_i(X_{qt} = j|\tau > t) = m_j, \qquad \forall i \in E. \tag{3.14}$$

Proof of this assertion for more general processes can be found in [5]. But since in the present case, a direct proof is simpler, we outline it as follows: from the Markov property and (3.11),

$$\begin{aligned}
\lim_{t\to\infty} P_i(X_{qt} = j|\tau > t) &= \lim_{t\to\infty} \frac{E_i[P_{X_{qt}}(\tau > (1-q)t); X_{qt} = j, \tau > qt]}{P_i(\tau > t)} \\
&= \lim_{t\to\infty} \frac{E_i[e^{\lambda_1 t} P_j(\tau > (1-q)t); X_{qt} = j, \tau > qt]}{e^{\lambda_1 t} P_i(\tau > t)} \\
&= \lim_{t\to\infty} \frac{E_i[e^{\lambda_1 qt}\beta_j; X_{qt} = j, \tau > qt]}{\beta_i} \\
&= \lim_{t\to\infty} P_i^{\lambda_1}(X_{qt} = j) = m_j,
\end{aligned}$$

where $P_i^{\lambda_1}$ is used for the tilted process $P^{\lambda_1}(t)$, and we have used in the last equality the fact that under (H1), $P^{\lambda_1}(t)$ is ergodic with the unique stationary distribution m.

Comparing this assertion with the those in Proposition 3.7, we see a phase transition in the conditional limit

$$\lim_{t\to\infty} P_i(X_{qt} = j|\tau > t)$$

when q varies from $0 < q < 1$ to $q = 1$.

A simple but non-trivial consequence of (3.12) is that $\forall i \in E$, $f \in \mathcal{B}_b(E)$,

$$\lim_{t\to\infty} E_i\left[\frac{1}{t}\int_0^t f(X_s)\,ds \middle| \tau > t\right] = m(f). \tag{3.15}$$

This is the reason that we call m a "quasi-ergodic measure" of X. The above limit can be extended to the following higher moment cases:

$$\lim_{t\to\infty} E_i\left[\left(\frac{1}{t}\int_0^t f(X_s)\,ds\right)^n \middle| \tau > t\right] = \big[m(f)\big]^n,$$

for any $n \geq 1$.

Finite Markov chains are trivial examples for which both (H1) and (H2) are satisfied. To give a simple but non-trivial example, we consider birth–death processes. This is a typical family of Markov chains which has been extensively studied from various aspects, including the quasi-stationary behavior. Here we only consider the linear birth–death process with birth and death rates given by

$$b_n = nb, \qquad n \geq 0, \qquad d_n = nd, \qquad n \geq 1,$$

where $b, d > 0$. 0 is the only absorbing state, and $E = \mathbb{N}$ is a transient irreducible class with decay parameter $\lambda_1 = |b - d|$. The chain is always λ_1-positive for $b \neq d$. However, the λ_1-invariant measure π is summable, and hence (H1) and (H2) are fulfilled, if and only if $b < d$.

The above approach can be generalized to give variational representation for the decay parameter of the Q-matrix with a potential. To describe this explicitly, let $Q = (q_{ij})$ be a standard Q-matrix on E, and $V \in C_b(E)$. Let $Q + V = (q^V_{ij})_{i,j \in E}$, with $q^V_{ij} = q_{ij} + \delta_{ij} V(i)$. $Q + V$ is a "quasi" Q-matrix in the sense that for some constant C,

$$\sum_j q^V_{ij} \leq C \qquad \forall i$$

which determine a "minimal" quasi-transition function $P^V(t) = (P^V_{ij}(t))$ with

$$\sum_j P^V_{ij}(t) \leq e^{Ct} \qquad \forall t, i.$$

If Q is conservative, regular and $X = \{X_t, t \geq 0\}$ is the Markov chain constructed from Q, then $P^V(t)$ is the Feynman–Kac semigroup given by

$$P^V(t) g(i) = E_i\bigl[g(X_t) e^{\int_0^t V(X_s)\,ds}\bigr], \qquad g \in C_b(E)$$

and $P^V_{ij}(t) = P^V_t \delta_j(i)$. Let $\lambda_1(V)$ be defined as in (2.7) for $Q + V$, and I the I-functional defined for Q.

Theorem 3.8. *With the above notations*,

$$-\lambda_1(V) = \sup_{\mu \in M_1(E)} \left\{ \int V\,d\mu - I(\mu) \right\}. \tag{3.16}$$

Remark. When E is finite, the above assertion follows from the Perron–Frobenius theorem.

Proof of Theorem 3.8. Let $V \leq C$. Define $Q' = Q + V - C$, then Q' is a standard Q-matrix and $Q' + C = Q + V$. Let λ'_1 be the decay parameter of Q' and I' be the I-functional for Q'. Then by Theorem 3.2,

$$\begin{aligned} -\lambda'_1 &= -\inf_{\mu \in M_1(E)} I'(\mu) \\ &= \sup_{\mu \in M_1(E)} \left\{ \inf_{u \in \mathcal{D}(L)^+} \int \frac{Q'u}{u}(x)\,d\mu(x) \right\} \\ &= \sup_{\mu \in M_1(E)} \left\{ \int V\,d\mu - I(\mu) \right\} - C. \end{aligned}$$

On the other hand, according to Remark 2.6,

$$\begin{aligned}\lambda_1' &= \sup\{r : \exists \text{ positive } \{x_i, i \in E\},\ \text{s.t. } Q'x \le -rx\}\\ &= \sup\{r : \exists \text{ positive } \{x_i, i \in E\},\ \text{s.t. } (Q+V)x \le -(r-C)x\}\\ &= \lambda_1(V) + C.\end{aligned}$$

Combining the above equalities, we get the desired assertion. □

4. Connections with discrete-time chains

In this section, we make some remarks on applying the results obtained in the last section to discrete-time Markov chains, and the other way round. Let $P = (P_{ij})_{i,j\in E}$ be an irreducible, nonnegative matrix satisfying

$$\sum_{j\in E} P_{ij} \le c \qquad \forall i \in E$$

for some constant c. $P^n = (P^n_{ij})$ denote the nth power of P. The convergence parameter of P is defined by

$$R = \inf\left\{ s > 0 : \sum_{n=0}^{\infty} P^n_{ij} s^n = \infty \right\}.$$

It is known (Corollary 4 of [28]) that

$$\begin{aligned}R &= \sup\{r : \exists \text{ positive column } \{u_i, i \in E\},\ \text{s.t. } rPu \le u\}\\ &= \sup\{r : \exists \text{ positive row } \{y_i, i \in E\},\ \text{s.t. } ryP \le y\}.\end{aligned} \tag{4.1}$$

Thus $\rho = 1/R$ is the Perron–Frobenius eigenvalue of P when E is finite, and (1.5) gives a dual form of variational representations of ρ. If E is infinite, from (4.1) it is easy to check that

$$\frac{1}{R} = \inf_{u\in\mathcal{U}} \sup_{i\in E} \frac{(Pu)_i}{u_i}. \tag{4.2}$$

Furthermore, the proof of Lemma 2, the argument in page 376 and Theorem 5.2 of [28] also show that

$$\frac{1}{R} = \sup_{u\in\mathcal{U}_\alpha} \inf_{i\in E} \frac{(Pu)_i}{u_i}, \tag{4.3}$$

where α is any fixed R-subinvariant measure, and

$$\mathcal{U}_\alpha \triangleq \left\{ u \in \mathcal{U} : \sum_j u_j \alpha_j < \infty \right\}.$$

Combining it with (4.2) we get the following "dual" formula for $1/R$:

$$\frac{1}{R} = \inf_{u \in \mathcal{U}} \sup_{i \in E} \frac{(Pu)_i}{u_i} = \sup_{u \in \mathcal{U}_\alpha} \inf_{i \in E} \frac{(Pu)_i}{u_i}, \tag{4.4}$$

which can be regarded as a generalization of (1.5) to the case where E is infinite. We will discuss the connection between the variational formulas for $1/R$ and those for λ_1 in two directions.

First, we can apply the results obtained in Section 3 to give a second form of "dual" representations of $1/R$. In doing this, we will give a connection between R and an associate continuous-time version of R. To be precise, define a standard Q-matrix $Q = (q_{ij})$ by

$$q_{ij} = \frac{1}{c} P_{ij} - \delta_{ij}, \qquad i, j \in E.$$

Its decay parameter λ_1 is defined by (2.7). Notice that

$$P = cQ + c\mathbf{1},$$

where $\mathbf{1}$ is the identically 1 function on E. Then by Theorem 3.2 we get that

$$\begin{aligned} c(1 - \lambda_1) &= c \sup_{\mu \in M_1(E)} \left[1 + \inf_{u \in \mathcal{U}} \int \frac{Qu}{u} \, d\mu \right] \\ &= c \sup_{\mu \in M_1(E)} \left[\inf_{u \in \mathcal{U}} \int \frac{(Q + \mathbf{1})u}{u} \, d\mu \right] \\ &= \sup_{\mu \in M_1(E)} \inf_{u \in \mathcal{U}} \int \frac{Pu}{u} \, d\mu. \end{aligned} \tag{4.5}$$

From (2.7) and (4.1), we see that

$$c(1 - \lambda_1) = \frac{1}{R} \quad \text{or} \quad \lambda_1 = 1 - \frac{1}{cR}. \tag{4.6}$$

Combining this with (4.2) and (4.5), we have proved the following theorem.

Theorem 4.1. *Let P be a nonnegative and irreducible matrix satisfying*

$$\sum_{j \in E} P_{ij} \le c \qquad \forall i \in E$$

for some c, and R be the corresponding decay parameter. Then

$$\frac{1}{R} = \sup_{\mu \in M_1(E)} \inf_{u \in \mathcal{U}} \int \frac{Pu}{u} \, d\mu = \inf_{u \in \mathcal{U}} \sup_{i \in E} \frac{(Pu)_i}{u_i} = \inf_{u \in \mathcal{U}} \sup_{\mu \in M_1(E)} \int \frac{(Pu)}{u} \, d\mu. \tag{4.7}$$

As another consequence of (4.6), consider a continuous-time Markov chain $P(t)$ with the Q-matrix $Q=(q_{ij})$ and its embedded discrete time chain $P=(P_{ij})$, that is,

$$P_{ij}=\frac{q_{ij}}{q_i}+\delta_{ij}.$$

If we define a Q-matrix $\overline{Q}=(\bar{q}_{ij})$ by

$$\bar{q}_{ij}=\frac{q_{ij}}{q_i},$$

then

$$P=\overline{Q}+\mathbf{1}.$$

Thus if we denote by $\bar{\lambda}_1$ and R the decay parameters of $\overline{Q}$ and P, respectively, then from (4.6) we see that

$$\bar{\lambda}_1=1-\frac{1}{R}. \tag{4.8}$$

On the other hand, if we denote by $\bar{J}$ the modified I-functional corresponding to $\overline{Q}$ (see definition (2.10)), and notice that $M_1^{\bar{q}}(E)=M_1(E)$, then it follows from Theorem 3.2 that

$$\bar{\lambda}_1=-\sup_{\mu\in M_1(E)}\inf_{u\in\mathcal{U}}\int\frac{\overline{Q}u}{u}\,d\mu=-\sup_{\mu\in M_1(E)}\inf_{u\in\mathcal{U}}\int\frac{Qu}{qu}\,d\mu. \tag{4.9}$$

Thus, if we denote by λ_1 the decay parameter of Q, then since $Q\tilde{u}\le-\lambda_1\tilde{u}$ for some $\tilde{u}\in\mathcal{U}$, we immediately get that

$$\bar{\lambda}_1\ge\Big(\sup_i q_i\Big)^{-1}\lambda_1.$$

Furthermore, for $\mu\in M_1^q(E)$, we can define $\mu_q\in M_1(E)$ by

$$d\mu_q=\frac{q\,d\mu}{\mu(q)}.$$

Then it follows from (4.9) that

$$\begin{aligned}\bar{\lambda}_1&\le-\sup_{\mu\in M_1^q(E)}\inf_{u\in\mathcal{U}}\int\frac{Qu}{qu}\,d\mu_q\\&=-\sup_{\mu\in M_1^q(E)}\inf_{u\in\mathcal{U}}\int\frac{Qu}{u}\frac{d\mu}{\mu(q)}\le\Big(\inf_i q_i\Big)^{-1}\lambda_1.\end{aligned}$$

Combining these with (4.8) we obtain that

$$\inf_i q_i\Big(1-\frac{1}{R}\Big)\le\lambda_1\le\sup_i q_i\Big(1-\frac{1}{R}\Big),$$

recovering a known result (Proposition 3.2 in [1], page 186).

Going the other way round, we next apply (4.4) to derive its continuous-time version for λ_1 when E is finite.

Proposition 4.2. *Let E be finite, $P(t)$ be an irreducible Markov transition function on E with Q-matrix $Q = (q_{ij})$. Then*

$$-\sup_{u\in\mathcal{U}}\inf_i \frac{(Qu)_i}{u_i} = \lambda_1 = -\inf_{u\in\mathcal{U}}\sup_i \frac{(Qu)_i}{u_i}. \tag{4.10}$$

Proof. To prove the first equality, for any $h > 0$, denote by $R(h)$ the convergence parameter of the discrete time chain with transition matrix $P(h)$, then $R(h) = e^{\lambda_1 h}$. If α is the (unique) λ_1-invariant measure, then

$$\mathcal{U}_\alpha = \mathcal{U}$$

since E is finite. Thus, we see from (4.4) (which is just (1.5) in the present case) that

$$e^{-\lambda_1 h} = \sup_{u\in\mathcal{U}}\inf_i \frac{(P(h)u)_i}{u_i} \geq \inf_i \frac{(P(h)u)_i}{u_i} \tag{4.11}$$

for each $u \in \mathcal{U}$. Since

$$\log\frac{P(h)u}{u} = \log\left[1 + h\frac{Qu}{u} + o(h)\right] = h\frac{Qu}{u} + o(h)$$

with $o(h)$ uniform in E, it follows from (4.11) that for each $u \in \mathcal{U}$,

$$-\lambda_1 = \sup_{u\in\mathcal{U}}\inf_i\left[\frac{(Qu)_i}{u_i} + o(1)\right] \geq \inf_i\left[\frac{(Qu)_i}{u_i} + o(1)\right]$$

with $o(1)$ uniform in E. Thus by letting $h \to 0+$, we obtain that

$$-\lambda_1 \geq \inf_i \frac{(Qu)_i}{u_i},$$

and hence

$$-\lambda_1 \geq \sup_{u\in\mathcal{U}}\inf_i \frac{(Qu)_i}{u_i}.$$

On the other hand, if β is the λ_1-invariant vector, then

$$-\lambda_1 = \frac{Q\beta}{\beta} \leq \sup_{u\in\mathcal{U}}\inf_i \frac{(Qu)_i}{u_i}.$$

The first equality in (4.10) follows from the above two inequalities. The second equality is given in Corollary 3.3. It can also be proved in the similar way as above when E is finite. □

The above arguments can be generalized to some specific cases when E is infinite.

Acknowledgements

The authors are grateful to the referees for the vary careful reading of the first version of the paper, for pointing out the recent book [22] and some other relevant references, especially for the many helpful comments and suggestions. The work is supported by the NSFC 11271022 and SRFDP 20120002110045.

References

[1] Anderson, W.J. (1991). *Continuous-Time Markov Chains*: *An Applications-Oriented Approach*. *Springer Series in Statistics*: *Probability and Its Applications*. New York: Springer. MR1118840

[2] Breyer, L.A. (1997). Quasistationarity and conditioned Markov processes. Ph.D. thesis, Department of Mathematics, The University of Queensland.

[3] Breyer, L.A. and Hart, A.G. (2000). Approximations of quasi-stationary distributions for Markov chains. *Math*. *Comput*. *Modelling* **31** 69–79. MR1768783

[4] Chen, J. and Deng, X. (2013). Large deviations and related problems for absorbing Markov chains. *Stochastic Process*. *Appl*. **123** 2398–2418. MR3038511

[5] Chen, J.W., Li, H.T. and Jian, S.Q. (2012). Some limit theorems for absorbing Markov processes. *J*. *Phys*. *A*: *Math*. *Theor.* **45** 3747–3762.

[6] Chen, M. (1996). Estimation of spectral gap for Markov chains. *Acta Math*. *Sinica* (*N*.*S*.) **12** 337–360. MR1457859

[7] Chen, M. (2004). *From Markov Chains to Non-equilibrium Particle Systems*, 2nd ed. River Edge, NJ: World Scientific Co. MR2091955

[8] Chen, M. (2005). *Eigenvalues*, *Inequalities*, *and Ergodic Theory*. *Probability and Its Applications* (*New York*). London: Springer. MR2105651

[9] Chen, M. (2010). Speed of stability for birth–death processes. *Front*. *Math*. *China* **5** 379–515. MR2660525

[10] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. New York: Springer. MR1619036

[11] Deuschel, J. and Stroock, D.W. (1989). *Large Deviations*. *Pure and Applied Mathematics* **137**. Boston, MA: Academic Press. MR0997938

[12] Donsker, M.D. and Varadhan, S.R.S. (1975). Asymptotic evaluation of certain Markov process expectations for large time. I. II. *Comm*. *Pure Appl*. *Math*. **28** 1–47; ibid. **28** (1975), 279–301. MR0386024

[13] Donsker, M.D. and Varadhan, S.R.S. (1975). On a variational formula for the principal eigenvalue for operators with maximum principle. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **72** 780–783. MR0361998

[14] Donsker, M.D. and Varadhan, S.R.S. (1976). On the principal eigenvalue of second-order elliptic differential operators. *Comm*. *Pure Appl*. *Math*. **29** 595–621. MR0425380

[15] Flaspohler, D.C. (1974). Quasi-stationary distributions for absorbing continuous-time denumerable Markov chains. *Ann*. *Inst*. *Statist*. *Math*. **26** 351–356. MR0368155

[16] Georgiou, N., Rassoul-Agha, F. and Seppäläinen, (2015). Variational formulas and cocycle solutions for directed polymer and percolation models. Preprint. Available at arXiv:1311.3016.

[17] Kingman, J.F.C. (1963). Ergodic properties of continuous-time Markov processes and their discrete skeletons. *Proc*. *Lond*. *Math*. *Soc*. (3) **13** 593–604. MR0154334

[18] Kosygina, E., Rezakhanlou, F. and Varadhan, S.R.S. (2006). Stochastic homogenization of Hamilton–Jacobi–Bellman equations. *Comm*. *Pure Appl*. *Math*. **59** 1489–1521. MR2248897

[19] Kosygina, E. and Varadhan, S.R.S. (2008). Homogenization of Hamilton–Jacobi–Bellman equations with respect to time-space shifts in a stationary ergodic medium. *Comm. Pure Appl. Math.* **61** 816–847. MR2400607

[20] Pollett, P.K. Quasi-stationary distributions: A bibliography. Available at http://www.maths.uq.edu.au/~pkp/papers/qsds.html.

[21] Rassoul-Agha, F. and Seppäläinen, T. (2014). Quenched point-to-point free energy for random walks in random potentials. *Probab. Theory Related Fields* **158** 711–750. MR3176363

[22] Rassoul-Agha, F. and Seppäläinen, T. (2015). *A Course on Large Deviations with an Introduction to Gibbs Measures*. *Graduate Studies in Mathematics* **162**. Providence, RI: Amer. Math. Soc. MR3309619

[23] Rassoul-Agha, F., Seppäläinen, T. and Yilmaz, A. (2013). Quenched free energy and large deviations for random walks in random potentials. *Comm. Pure Appl. Math.* **66** 202–244. MR2999296

[24] Seneta, E. (2006). *Non-negative Matrices and Markov Chains*. *Springer Series in Statistics*. New York: Springer. MR2209438

[25] Stroock, D.W. (1984). *An Introduction to the Theory of Large Deviations*. *Universitext*. New York: Springer. MR0755154

[26] van Doorn, E.A. (2002). Representations for the rate of convergence of birth–death processes. *Theory Probab. Math. Statist.* **65** 37–43. MR1936126

[27] van Doorn, E.A. and Pollett, P.K. (2011). Quasi-stationary distributions. Available at http://eprints.eemcs.utwente.nl/20245/.

[28] Vere-Jones, D. (1967). Ergodic properties of nonnegative matrices. I. *Pacific J. Math.* **22** 361–386. MR0214145

[29] Vere-Jones, D. (1968). Ergodic properties of nonnegative matrices. II. *Pacific J. Math.* **26** 601–620. MR0236745

Received March 2015 and revised December 2015