Adv. Appl. Prob. **48,** 631–647 (2016)
*doi:*10.1017/apr.2016.20

CHEEGER INEQUALITIES FOR ABSORBING MARKOV CHAINS

GARY FROYLAND * AND
ROBYN M. STUART,** *University of New South Wales*

Abstract

We construct Cheeger-type bounds for the second eigenvalue of a substochastic transition probability matrix in terms of the Markov chain's conductance and metastability (and vice versa) with respect to its quasistationary distribution, extending classical results for stochastic transition matrices.

Keywords: Absorbing Markov chain; transient Markov chain; substochastic transition matrix; quasistationary distribution; Cheeger constant; conductance; metastability

2010 Mathematics Subject Classification: Primary 60J10

1. Introduction

Let $Z = (Z_t,\ t = 0, 1, \dots)$ be a time-homogeneous discrete-time Markov chain taking values in a finite set $\Omega = \{0, 1, 2, \dots, n\}$ with transition matrix $\bar{P} = (\bar{P}_{ij})$, where $\bar{P}_{ij} = \mathbb{P}(Z_{t+1} = j \mid Z_t = i)$ for $i, j \in \Omega$. We suppose that $X := \Omega \setminus \{0\}$ with

$$\mathbb{P}(Z_{t+1} = j \mid Z_t = 0) = 0 \quad \text{for all } j \in X,$$

and that $a_i := \mathbb{P}(Z_{t+1} = 0 \mid Z_t = i) > 0$ for at least one $i \in X$. Thus, 0 is an *absorbing state*, and we write $\bar{P}$ in the canonical form

$$\bar{P} = \begin{pmatrix} 1 & 0 \\ a & P \end{pmatrix}. \tag{1}$$

We assume that P is an irreducible matrix; that is, given $i, j \in X$, there is an $m \geq 1$ such that $(P^m)_{ij} > 0$. A *quasistationary distribution* p is a probability measure on X with the property that if the distribution at time 0 is p then the distribution conditional on nonabsorption up until time $t > 0$ is still p, i.e.

$$\mathbb{P}_p(Z_t = j \mid T > t) = p_j, \qquad j \in X, \tag{2}$$

where $\mathbb{P}_p(\cdot) = \sum_{i \in X} p_i \mathbb{P}_i(\cdot)$ and $T := \inf\{t \geq 0 \colon Z_t = 0\}$. It is well known that in the above setting there is a unique quasistationary distribution, given by the normalised leading left eigenvector of P [7], [8].

The notion of modelling the long-term behaviour of absorbing Markov processes was first explored by Bartlett [3] and Ewans [12], [13] in the context of population modelling. Subsequent work developed conditions under which the quasistationary distribution is equal to the Yaglom limit [38], where the latter is defined as the $t \to \infty$ limiting probability of the Markov process

Received 1 October 2013; revision received 19 September 2015.
* Postal address: School of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052, Australia. Email address: g.froyland@unsw.edu.au
** Current address: Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark.

generated by $\bar{P}$ being in state i at time t, conditioned on the fact that the process has remained in X until time t (see, e.g. [8]).

The equivalence of the quasistationary distribution and the Yaglom limit has been developed in the discrete-time, finite-state setting [8], [23]–[25], and for countable-state processes [32], [35]. The continuous-time setting has been analysed in [9], [16], [29], [30]; in particular, [15] states a necessary and sufficient condition for the existence of the Yaglom limit based on the distribution of the absorption times. Yaglom limits for continuous-time Markov processes have been analysed in [9], [16], [29], [30]. The Yaglom limit has also been studied via Fleming–Viot processes [18] in the continuous-time Markov process setting [2], [14]. Here, a group of N particles evolve independently according to the underlying Markov process, and when a particle enters the absorbing state it chooses uniformly one of the other particles and immediately jumps to its location. In the finite-state case, the authors in [2] show that the empirical profile of the unique invariant measure of the Fleming–Viot process with N particles converges as $N \to \infty$ to the unique quasistationary distribution of the one-particle motion. Other settings that have been considered include semi-Markov processes [1], [5], [17], continuous-state processes [34], and settings where P may be reducible [36], [37]. Applications of substochastic Markov chains include biological, environmental, and financial models, where frequently a path terminates when it enters a death or valueless state.

Our particular focus in this paper is on *conductance* and *metastability* in Markov processes. The conductance of a Markov chain (see Definition 2) is the minimum conditional probability of leaving a set B in one step of the process, minimised over all sets $B \subset X$ with probability no greater than $\frac{1}{2}$. The conductance has many other names within the literature, such as the *Cheeger constant* [4], the *isoperimetric constant* [26], and the *bottleneck ratio* [27]. The metastability of a Markov chain (see Definition 2) is the maximum conditional probability of remaining in a set B for one step of the process, maximised over all sets $B \subset X$ with probability no greater than $\frac{1}{2}$. The study of metastability has foundations in both Markov chain theory [26] and in the study of dynamical systems, where it is also called *almost-invariance* [10], [19].

If P is stochastic (i.e. $a = 0$ in (1)) and satisfies the detailed balance condition $p_i P_{ij} = p_j P_{ji}$, $i, j \in X$, the second eigenvalue of P (and, therefore, the mixing rate) can be bounded by expressions involving the conductance $\varphi_{p,P}$ [26], [33], which can be rearranged and expressed in terms of the metastability $\omega_{p,P}$, i.e.

$$1 - 2\varphi_{p,P} \le \lambda_{P,2} \le 1 - \frac{\varphi_{p,P}^2}{2}, \tag{3}$$

$$2\omega_{p,P} - 1 \le \lambda_{P,2} \le 1 - \frac{(1 - \omega_{p,P})^2}{2}. \tag{4}$$

The upper bound for $\lambda_{P,2}$ is the deeper of the two inequalities in (3). It states that if there does not exist a set $B \subset X$ from which the conditional probability of escape in one step is small, the Markov chain must have rapid convergence to equilibrium.

For stochastic matrices that are not time-reversible, the authors in [6] and [19] developed identical bounds for $\lambda_{R,2}$, the second largest eigenvalue of the time-symmetrised Markov chain with transition matrix $R = (P + \hat{P})/2$, where $\hat{P}_{ij} = p_j P_{ji}/p_i$, i.e.

$$1 - 2\varphi_{p,P} \le \lambda_{R,2} \le 1 - \frac{\varphi_{p,P}^2}{2}, \tag{5}$$

$$2\omega_{p,P} - 1 \le \lambda_{R,2} \le 1 - \frac{(1 - \omega_{p,P})^2}{2}. \tag{6}$$

Simple estimates of conductance and metastability [6], [19] immediately provide bounds for the mixing rate of P or its time-symmetrisation R. By switching the focus from bounding $\lambda_{P,2}$ or $\lambda_{R,2}$ to bounding conductance and metastability in terms of $\lambda_{P,2}$ or $\lambda_{R,2}$ (and lower eigenvalues [22]), the inequalities (3) and (4) can be exploited in Markov models of dynamical systems to estimate the sets $B \subset X$ that minimise the conductance and maximise the metastability [11], [19], [20], [22].

The literature on conductance and metastability is almost wholly concerned with Markov chains over a single communicating class. A Markov chain with killing was treated in Lawler and Sokal [26], and we will discuss their results in detail in Section 5. The metastability of an absorbing Markov process was defined and bounded by means of 'resurrecting' the chain whenever it is about to leave X [21]. Resurrection techniques (see [30] for general Markov processes) nowadays play a fundamental role in the construction of the Fleming–Viot particle systems approach to analysing long-term behaviour of Markov processes [2], [14]. In this paper our main contribution is to extend the arguments of [33] to create bounds for second eigenvalues in terms of conductance and metastability fo rthe general substochastic P.

In Section 2 we define flux, conductance, metastability, and reversibility for substochastic transition matrices. Section 3 contains our main results: extensions of the bounds (3)–(6) to the substochastic transition probability matrix case. The proofs of our main results appear in Section 4, and concluding remarks are contained in Section 5.

2. Flux, conductance, metastability, and reversibility

Let $\bar{P}$ be as in (1), and let p be the quasistationary distribution of the process conditioned to remain on X. We define a *cut* to be a division of X into complementary sets B and $B^{\mathrm{c}} := X \setminus B$.

Definition 1. (*Flux and invariance ratios.*) Given subsets $B, C \subset X$, the p-flux from B to C under P is the probability, according to p, that the process is in B at time t and in C at time $t+1$, i.e.

$$Q_{p,P}(B,C) := \sum_{i\in B,\, j\in C} p_i P_{ij}.$$

We also define the *p-flux ratio* under P, as $Q_{p,P}(B,C)/p(B)$.

Definition 2. (*Conductance and metastability.*) With the restriction $C = B^{\mathrm{c}}$, the minimal p-flux ratio under P is called the *conductance*, i.e.

$$\varphi_{p,P} := \min_{\{B\subset X:\ p(B)\le 1/2\}} \frac{Q_{p,P}(B,B^{\mathrm{c}})}{p(B)}.$$

The conductance is the smallest p-flux ratio from a set B of p-measure no greater than $\frac{1}{2}$ to B^{c}. With the restriction $B = C$, the maximal p-flux ratio under P is called the *metastability* of P, defined by

$$\omega_{p,P} := \max_{\{B\in X:\ p(B)\le 1/2\}} \frac{Q_{p,P}(B,B)}{p(B)}.$$

The metastability is the largest conditional probability of the process remaining in a subset of no greater than p-measure $\frac{1}{2}$ for one time-step.

Definition 3. We say that P is *p-reversible* if $p_i P_{ij} = p_j P_{ji}$.

If P were stochastic, Definition 3 would be the usual detailed balance condition, which implies that the underlying random walk is identical if the time-direction of the walk is reversed (see, e.g. [23, p. 5, Theorem 1.2]). This corresponds to the usual definition of reversibility for Markov chains, but we use the terminology p-reversible so that it is clear that we mean reversible with respect to the probability distribution p. As P is substochastic, satisfying the detailed balance condition does not imply that the random walk is identical when reversed. Indeed, there are several possibilities for what a 'reversed process' might mean (see, e.g. [9] and [31]). Hence, our use of the terminology 'p-reversible' only implies that $p_i P_{ij} = p_j P_{ji}$ is satisfied, and does not necessarily correspond to a statement about the distribution of the underlying random walk.

3. Main results

Define $\rho_i := \sum_{j\in X} P_{ij}$ and $\underline{\rho} := \min_i \rho_i \le \rho_i \le \max_i \rho_i =: \bar{\rho}$. We have

$$\underline{\rho} \le \lambda_{P,1} \le \bar{\rho} \tag{7a}$$

in the general case, and

$$\underline{\rho} = \lambda_{P,1} = \bar{\rho} \tag{7b}$$

if and only if $P\mathbf{1} = \lambda_{P,1}\mathbf{1}$, where $\mathbf{1}$ is the vector of all entries 1, both elementary results of linear algebra (see, e.g. [28]). Furthermore, let $p_\rho(B) := \sum_{i\in B} p_i\rho_i$ and as a direct consequence of (7a), we have, for all $B \subset X$,

$$\underline{\rho}\, p(B) \le p_\rho(B) \le \bar{\rho}\, p(B)$$

in the general case, and

$$p_\rho(B) = \lambda_{P,1} p(B) \quad \Longleftrightarrow \quad P\mathbf{1} = \lambda_{P,1}\mathbf{1}. \tag{8}$$

We next note some basic properties of the p-flux and p-invariance.

Lemma 1. *For a substochastic transition matrix P, the following facts hold.*

(i) $Q_{p,P}(X, X) = \lambda_{P,1}$.

(ii) $Q_{p,P}(B, B) = \lambda_{P,1} p(B) - Q_{p,P}(B^{\mathrm{c}}, B)$ *for all* $B \subset X$.

(iii) $Q_{p,P}(B, B^{\mathrm{c}}) = Q_{p,P}(B^{\mathrm{c}}, B) + p_\rho(B) - \lambda_{P,1} p(B)$ *for all* $B \subset X$*; if* $P\mathbf{1} = \lambda_{P,1}\mathbf{1}$, $Q_{p,P}(B, B^{\mathrm{c}}) = Q_{p,P}(B^{\mathrm{c}}, B)$ *for all* $B \subset X$.

(iv) $\underline{\rho} - \varphi_{p,P} \le \omega_{p,P} \le \bar{\rho} - \varphi_{p,P}$*; if* $P\mathbf{1} = \lambda_{P,1}\mathbf{1}$, $\omega_{p,P} = \lambda_{P,1} - \varphi_{p,P}$.

Proof. (i) The proof is obvious.

(ii) Write $Q_{p,P}(B, B) = \sum_{i\in X,\, j\in B} p_i P_{ij} - \sum_{i\notin B,\, j\in B} p_i P_{ij} = \lambda_{P,1} p(B) - Q_{p,P}(B^{\mathrm{c}}, B)$.

(iii) To show the general case, note that

$$\begin{aligned} Q_{p,P}(B, B^{\mathrm{c}}) &= \sum_{i\in B,\, j\in X} p_i P_{ij} - \sum_{i,j\in B} p_i P_{ij} \\ &= p_\rho(B) - Q_{p,P}(B, B) \\ &= p_\rho(B) - \lambda_{P,1} p(B) + Q_{p,P}(B^{\mathrm{c}}, B), \end{aligned}$$

where the last equality follows by using (ii). If $P\mathbf{1} = \lambda_{P,1}\mathbf{1}$ then $p_\rho(B) = \lambda_{P,1} p(B)$ by (8), so the general case reduces to $Q_{p,P}(B, B^{\mathrm{c}}) = Q_{p,P}(B^{\mathrm{c}}, B)$.

(iv) The proof follows by combining (ii) and (iii) to obtain

$$Q_{p,P}(B,B) = p_\rho(B) - Q_{p,P}(B,B^c), \tag{9}$$

which implies that

$$\underline{\rho}\, p(B) - Q_{p,P}(B,B^c) \le Q_{p,P}(B,B) \le \bar{\rho}\, p(B) - Q_{p,P}(B,B^c). \tag{10}$$

If $P\mathbf{1} = \lambda_{P,1}\mathbf{1}$ then by (8), we obtain

$$Q_{p,P}(B,B) = \lambda_{P,1}\, p(B) - Q_{p,P}(B,B^c). \tag{11}$$

In the general (respectively $P\mathbf{1} = \lambda_{P,1}\mathbf{1}$) case, dividing (10) (respectively (11)) by $p(B)$ and then taking the minimum of both sides over all sets $B \subset X$ such that $p(B) \le \frac{1}{2}$ yields the result. □

Following the methodology used in the stochastic case, if P is not p-reversible, we construct the adjoint of P in $(\mathbb{R}^{|X|}, \langle\cdot,\cdot\rangle_p)$, given by $\hat{P}_{ij} = p_j P_{ji}/p_i$. Then $R = (P+\hat{P})/2$ is self-adjoint on $(\mathbb{R}^{|X|}, \langle\cdot,\cdot\rangle_p)$. One can easily show that

$$Q_{p,R}(B,B^c) = \frac{1}{2(Q_{p,P}(B,B^c) + Q_{p,P}(B^c,B))} \tag{12}$$

holds for substochastic R. Substituting $Q_{p,P}(B^c,B) = Q_{p,P}(B,B^c) + \lambda_{P,1}p(B) - p_\rho(B)$ from Lemma 1(iii) into (12), one can derive the relationship

$$Q_{p,R}(B,B^c) = Q_{p,P}(B,B^c) + \tfrac{1}{2}(\lambda_{P,1}p(B) - p_\rho(B)). \tag{13}$$

For $j \in X$, one has $(pR)_j = \frac{1}{2}(\rho_j + \lambda_{P,1})p_j$ and $(R\mathbf{1})_j = \frac{1}{2}(\rho_j + \lambda_{P,1})$.

Using these relationships, and the fact that R is self-adjoint in $(\mathbb{R}^{|X|}, \langle\cdot,\cdot\rangle_p)$, one can construct bounds for $\lambda_{R,2}$ in terms of $\varphi_{p,P}$, $\omega_{p,P}$, and the basic fundamental properties of P. Our main result is the following theorem.

Theorem 1. *Let P be an irreducible and substochastic matrix and $R = (P+\hat{P})/2$. Denote by $\lambda_{P,1}$ the leading eigenvalue of P and by $\lambda_{R,2}$ the second largest eigenvalue for R. Define $\underline{\rho} = \min_{i\in X}\sum_{j\in X} P_{ij}$ and $\bar{\rho} = \max_{i\in X}\sum_{j\in X} P_{ij}$. We have*

$$\frac{2\lambda_{P,1}k_1^2 - k_1(\lambda_{P,1}^2 + \bar{\rho}^2 - 2\lambda_{P,1}\underline{\rho})}{2(4\underline{\rho}(\bar{\rho}-\underline{\rho}) + k_1(\bar{\rho}+\lambda_{P,1} - 2\underline{\rho}))} - 2\varphi_{p,P} \le \lambda_{R,2} \le \frac{\bar{\rho}+\lambda_{P,1}}{2} - \frac{(\varphi_{p,P} - (\bar{\rho}-\lambda_{P,1})/2)^2}{2}, \tag{14}$$

where $k_1 = \sqrt{2((\lambda_{P,1}-\underline{\rho})^2 + (\bar{\rho}-\underline{\rho})^2)}$.

Also

$$\lambda_{R,2} \le \frac{\bar\rho + \lambda_{P,1}}{2} - \frac{(\underline{\rho} - \omega_{p,P})^2}{2} + \frac{(\bar\rho - \lambda_{P,1})(3\bar\rho + \lambda_{P,1} - 4\omega_{p,P})}{8}, \tag{15}$$

and

$$\lambda_{R,2} \le \sqrt{1 + \bar\rho + \lambda_{P,1} - 2\omega_{p,P}} - (1 - \omega_{p,P}). \tag{16}$$

Proof. Equations (14) and (16) are proved in Section 4. To show inequality (15), we use the right-hand inequality of (14); thus,

$$\begin{aligned}
\lambda_{R,2} &\le \frac{\bar\rho + \lambda_{P,1}}{2} - \frac{\varphi_{p,P}^2}{2} + \frac{\varphi_{p,P}(\bar\rho - \lambda_{P,1})}{2} - \frac{(\bar\rho - \lambda_{P,1})^2}{8} \\
&\le \frac{\bar\rho + \lambda_{P,1}}{2} - \frac{(\underline{\rho} - \omega_{p,P})^2}{2} + \frac{(\bar\rho - \omega_{p,P})(\bar\rho - \lambda_{P,1})}{2} \\
&\quad - \frac{(\bar\rho - \lambda_{P,1})^2}{8} \quad \text{by Lemma 1(iv)} \\
&\le \frac{\bar\rho + \lambda_{P,1}}{2} - \frac{(\underline{\rho} - \omega_{p,P})^2}{2} + \frac{(\bar\rho - \omega_{p,P})(3\bar\rho + \lambda_{P,1} - 4\omega_{p,P})}{8}.
\end{aligned}$$ □

In situations where $\underline{\rho}$ and $\bar\rho$ are substantially different, (16) may produce a tighter bound; however, when $\underline{\rho} = \bar\rho$ $(= \lambda_{P,1})$, (15) is the stronger bound. In the case where P is p-reversible, the bounds take more familiar forms.

Corollary 1. *For an irreducible, substochastic, p-reversible matrix P with leading eigenvalues $\lambda_{P,1} > \lambda_{P,2}$, we have*

$$\lambda_{P,1} - 2\varphi_{p,P} \le \lambda_{P,2} \le \lambda_{P,1} - \frac{\varphi_{p,P}^2}{2} \tag{17}$$

and

$$2\omega_{p,P} - \lambda_{P,1} \le \lambda_{P,2} \le \lambda_{P,1} - \frac{(\lambda_{P,1} - \omega_{p,P})^2}{2}. \tag{18}$$

Proof. Summing the equality $p_i P_{ij} = p_j P_{ji}$ over $i \in X$, we obtain $\lambda_{P,1} p_j = p_j \sum_{i \in X} P_{ji}$. This implies that $\sum_{i \in X} P_{ji} = \lambda_{P,1}$ for all j, or that $P\mathbf{1} = \lambda_{P,1}\mathbf{1}$, and so $\underline{\rho} = \lambda_{P,1} = \bar\rho$ by (7b). Making this substitution into the upper bound of (14), noting $P = R$ and $\lambda_{R,2} = \lambda_{P,2}$, yields the upper bound of (17). To obtain the lower bound of (17) one can check that as $\bar\rho \to \lambda_{P,1}$ and $\underline{\rho} \to \lambda_{P,1}$, the limit of the lower bound in (14) is $\lambda_{P,1} - 2\varphi_{p,P}$. One can then obtain (18) by substituting $\omega_{p,P} = \lambda_{P,1} - \varphi_{p,P}$ (as shown in Lemma 1(iv)) into (17). □

We note that if P is stochastic, (17) and (18) revert to the bounds (3) and (4), respectively. Similarly, (14) and (15) become (5) and the right-hand side of (6), respectively, as P approaches a stochastic non-p-reversible matrix.

To end this section, we remark that there is a vast literature of using Markov chain models as convenient numerical approximations of differentiable dynamical systems arising from models in fluid dynamics, molecular dynamics, astrodynamics, and physical oceanography. In these settings the computation of λ_2 is straightforward; the difficulty is in identifying a collection of states that are maximally metastable or 'almost-invariant'. The rewriting of the inequalities (3)–(6) and (14)–(18) to bound $\phi_{p,P}$ and $\omega_{p,P}$ provide rigorous ranges for these important dynamic quantities. For completeness, we state a theorem for the nonreversible case. The upper bound on $\varphi_{p,P}$ and the lower bound on $\omega_{p,P}$ are straightforward rearrangements of the corresponding upper bound in (14) and upper bound in (16). Furthermore, by introducing the additional constraint that $\lambda_{P,1} > \lambda_{R,2}$ we can derive a lower bound on $\varphi_{p,P}$ and an upper bound on $\omega_{p,P}$.

Theorem 2. *Under the conditions of Theorem 1,*

$$\varphi_{p,P} \leq \sqrt{\bar{\rho} + \lambda_{P,1} - 2\lambda_{R,2}} + \frac{\bar{\rho} - \lambda_{P,1}}{2} \tag{19}$$

and

$$\omega_{p,P} \geq \max\left(\lambda_{R,2}, \frac{2\underline{\rho} - \bar{\rho} + \lambda_{P,1}}{2}\right) - \sqrt{\bar{\rho} + \lambda_{P,1} - 2\lambda_{R,2}}. \tag{20}$$

In addition, if $\lambda_{P,1} > \lambda_{R,2}$ *then*

$$\varphi_{p,P} \geq \frac{2\underline{\rho} - \bar{\rho} - \lambda_{R,2}}{2} - \frac{(\bar{\rho} - \lambda_{P,1})^2}{8(\lambda_{P,1} - \lambda_{R,2})} \tag{21}$$

and

$$\omega_{p,P} \leq \frac{\bar{\rho} + \lambda_{R,2}}{2} + \frac{(\bar{\rho} - \lambda_{P,1})^2}{8(\lambda_{P,1} - \lambda_{R,2})}. \tag{22}$$

We do not prove (19) and (20) as the proof technique is identical to the one used to prove the corresponding bounds in (14) and (16). A proof of (22) is contained in Section 4, and (21) follows from combining the upper bound on $\omega_{p,P}$ in (22) with Lemma 1(iv).

4. Proofs of Theorems 1 and 2

Starting with the proof of Theorem 1, the proofs build on techniques from [33], which treats the stochastic p-reversible case.

4.1. Spectral properties of P, $\hat{P}$, and R

We present a brief collection of some important results concerning the spectra of P, $\hat{P}$, and R. Denote the leading left and right eigenvectors of R by u_1 and r, respectively, and scale r so that $\langle r, 1\rangle_p = 1$. Denote the leading right eigenvector of P by v_1, and form the vector pv_1 by elementwise multiplication, i.e. $(pv_1)_i = p_i v_{1,i}$. The leading left and right eigenvector properties are presented in Table 1.

TABLE 1: Spectra of P, $\hat{P}$, and R.

Matrix	Left eigenvector property	Right eigenvector property	Other properties
P	$pP = \lambda_{P,1}p$	$Pv_1 = \lambda_{P,1}v_1$	$(P\mathbf{1})_i = \rho_i$ and $\underline{\rho}\mathbf{1} \leq P\mathbf{1} \leq \bar{\rho}\mathbf{1}$
$\hat{P}$	$(pv_1)\hat{P} = \lambda_{P,1}(pv_1)$	$\hat{P}\mathbf{1} = \lambda_{P,1}\mathbf{1}$	$(p\hat{P})_j = p_j\rho_j$ and $\underline{\rho}p \leq p\hat{P} \leq \bar{\rho}p$
R	$u_1R = \lambda_{R,1}u_1$	$Rr = \lambda_{R,1}r$	$(pR)_j = \dfrac{\lambda_{P,1} + p_j\rho_j}{2}$ $\dfrac{\underline{\rho} + \lambda_{P,1}}{2}p \leq pR \leq \dfrac{\bar{\rho} + \lambda_{P,1}}{2}p$ $(R\mathbf{1})_i = \dfrac{\rho_i + \lambda_{P,1}}{2}$ $\dfrac{\underline{\rho} + \lambda_{P,1}}{2}\mathbf{1} \leq R\mathbf{1} \leq \dfrac{\bar{\rho} + \lambda_{P,1}}{2}\mathbf{1}$

4.2. Upper bounds on $\lambda_{R,2}$

Since $\langle Rx, y\rangle_p = \langle x, Ry\rangle_p$ (i.e. R is self-adjoint in $(\mathbb{R}^{|X|}, \langle\cdot,\cdot\rangle_p)$), its top two left eigenvectors u_1, u_2 are orthogonal with respect to $\langle\cdot,\cdot\rangle_{1/p}$. Since $\langle u_1, u_2\rangle_{1/p} = 0$ and $u_1 > 0$, u_2 must have both positive and negative values. Define $x_i = u_{2,i}/p_i$ and order x as $x_1 \geq x_2 \geq \cdots \geq x_k > 0 \geq x_{k+1} \geq \cdots \geq x_n$. Let $B = \{1,\dots,k\}$, and assume that $\sum_{i\in B} p_i \leq \frac{1}{2}$ (if not, switch the parity of u_2). Define y as

$$y_i = \begin{cases} x_i & \text{if } u_{2,i} \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

We now prove two preliminary auxiliary lemmas.

Lemma 2. *For y defined by (23), we have*

$$\lambda_{R,2} \leq \frac{\langle Ry, y\rangle_p}{\langle y, y\rangle_p}.$$

Proof. Since u_2 is a left eigenvector of R, we have

$$u_2\lambda_{R,2}y = u_2Ry. \tag{24}$$

The left-hand side of (24) is

$$u_2\lambda_{R,2}y = \lambda_{R,2}\sum_{i\in B} p_i y_i^2 = \lambda_{R,2}\langle y, y\rangle_p, \tag{25}$$

while the right-hand side is

$$\begin{aligned} u_2Ry &= \sum_{i\in X, j\in B} u_{2,i}R_{ij}y_j \\ &= \sum_{i\in X, j\in B} u_{2,i}R_{ij}y_j \\ &\geq \sum_{i,j\in B} u_{2,i}R_{ij}y_j \quad \text{since } u_{2,i} \leq 0 \text{ for } i \notin B \\ &= \langle Ry, y\rangle_p. \end{aligned} \tag{26}$$

Combining (24)–(26) yields the desired result. □

Lemma 3. *For arbitrary $z \in \mathbb{R}^n$, we have*

$$\left\langle z, \left(\frac{\lambda_{P,1}+\bar\rho}{2} - R\right)z\right\rangle_p \geq \frac{1}{2}\sum_{i,j\in X} p_i R_{ij}(z_i - z_j)^2.$$

Proof. We have

$$\begin{aligned} \left\langle z, \left(\frac{\lambda_{P,1}+\bar\rho}{2} - R\right)z\right\rangle_p &= \sum_{i\in X} z_i^2\left(\frac{\lambda_{P,1}+\bar\rho}{2}\right)p_i - \sum_{i,j\in X} z_iR_{ij}z_jp_i \\ &\geq \sum_{i,j\in X} z_i^2R_{ji}p_j - \sum_{i,j\in X} z_iR_{ij}z_jp_i \quad \text{by Table 1} \\ &= \frac{1}{2}\sum_{i,j\in X} p_iR_{ij}(z_i - z_j)^2. \end{aligned}$$

□

Since Lemma 3 holds for all $z \in \mathbb{R}^n$, in particular it holds for y defined by (23). Thus,

$$\left\langle y, \left(\frac{\lambda_{P,1} + \bar{\rho}}{2} - R\right) y \right\rangle_p \geq \frac{1}{2} \sum_{i,j \in X} p_i R_{ij} (y_i - y_j)^2. \tag{27}$$

Apply Lemma 2 to the left-hand side of (27) to obtain

$$\frac{\lambda_{P,1} + \bar{\rho}}{2} \langle y, y \rangle_p - \langle y, Ry \rangle \leq \left(\frac{\lambda_{P,1} + \bar{\rho}}{2} - \lambda_{R,2}\right) \langle y, y \rangle,$$

and rearrange to obtain

$$\lambda_{R,2} \leq \frac{\bar{\rho} + \lambda_{P,1}}{2} - \frac{\sum_{i,j \in X} p_i R_{ij} (y_i - y_j)^2}{\sum_{i \in B} p_i y_i^2}. \tag{28}$$

Using the inequality $(a+b)^2 \leq 2(a^2 + b^2)$, we have

$$\begin{aligned}
\sum_{i<j} p_i R_{ij} (y_i + y_j)^2 &\leq 2 \sum_{i<j} p_i R_{ij} (y_i^2 + y_j^2). \\
&= 2 \sum_{i<j} p_i R_{ij} y_i^2 + 2 \sum_{j<i} p_i R_{ij} y_i^2 \quad \text{(by } p\text{-reversibility)} \\
&= 2 \sum_{i \neq j} p_i R_{ij} y_i^2 \\
&\leq 2 \sum_{i \in B} p_i y_i^2.
\end{aligned}$$

Hence,

$$\frac{\sum_{i<j} p_i R_{ij} (y_i + y_j)^2}{2 \sum_{i \in B} p_i y_i^2} \leq 1. \tag{29}$$

We multiply the second term on the right-hand side of (28) by the left-hand side of (29) to obtain

$$\lambda_{R,2} \leq \frac{\bar{\rho} + \lambda_{P,1}}{2} - \frac{\sum_{i<j} p_i R_{ij} (y_i - y_j)^2}{\sum_{i \in B} p_i y_i^2} \frac{\sum_{i<j} p_i R_{ij} (y_i + y_j)^2}{2 \sum_{i \in B} p_i y_i^2}.$$

By the Cauchy–Schwarz inequality,

$$\left(\sum_{i<j} p_i R_{ij} (y_i^2 - y_j^2)\right)^2 \leq \left(\sum_{i<j} p_i R_{ij} (y_i - y_j)^2\right) \left(\sum_{i<j} p_i R_{ij} (y_i + y_j)^2\right),$$

and so

$$\lambda_{R,2} \leq \frac{\bar{\rho} + \lambda_{P,1}}{2} - \frac{1}{2} \left(\frac{\sum_{i<j} p_i R_{ij} (y_i^2 - y_j^2)}{\sum_{i \in B} p_i y_i^2}\right)^2. \tag{30}$$

We now examine the numerator of the squared term on the right-hand side of (30). Recall that $B := \{1, \dots, k\}$. For $l \leq k$, we define a new set $B_l = \{1, \dots, l\}$. Then

$$\sum_{i<j} p_i R_{ij} (y_i^2 - y_j^2) = \sum_{i<j} p_i R_{ij} \left(\sum_{i \leq l < j} (y_l^2 - y_{l+1}^2)\right) = \sum_{l=1}^{k} (y_l^2 - y_{l+1}^2) \sum_{i \in B_l,\, j \in B_l^{\mathrm{c}}} p_i R_{ij}. \tag{31}$$

We now split the proof into two subsections, one which establishes the upper bound on $\lambda_{R,2}$ in terms of $\varphi_{p,P}$ and one which establishes the upper bound on $\lambda_{R,2}$ in terms of $\omega_{p,P}$.

4.2.1. *Upper bound on* $\lambda_{R,2}$ *in terms of* $\varphi_{p,P}$. Combining the expression for $Q_{p,R}(B, B^{\mathrm{c}})$ given in (13) with (31), we obtain

$$\sum_{i<j} p_i R_{ij}(y_i^2 - y_j^2) \geq \sum_{l=1}^{k}(y_l^2 - y_{l+1}^2)\biggl(\sum_{i\in B_l,\, j\in B_l^{\mathrm{c}}} p_i P_{ij} - \frac{p(B_l)(\bar\rho - \lambda_{P,1})}{2}\biggr).$$

Since $l \leq k$, $p(B_l) \leq p(B) \leq \frac{1}{2}$, so $Q_{p,P}(B_l, B_l^{\mathrm{c}}) \geq \varphi_{p,P}\, p(B_l)$. Therefore,

$$\begin{aligned}
\sum_{i<j} p_i R_{ij}(y_i^2 - y_j^2) &\geq \sum_{l=1}^{k}(y_l^2 - y_{l+1}^2)p(B_l)\biggl(\varphi_{p,P} - \frac{\bar\rho - \lambda_{P,1}}{2}\biggr)\\
&= \biggl(\varphi_{p,P} - \frac{\bar\rho - \lambda_{P,1}}{2}\biggr)\sum_{l=1}^{k}(y_l^2 - y_{l+1}^2)\sum_{i=1}^{l} p_i\\
&= \biggl(\varphi_{p,P} - \frac{\bar\rho - \lambda_{P,1}}{2}\biggr)\sum_{i=1}^{k} p_i \sum_{l=1}^{k}(y_l^2 - y_{l+1}^2)\\
&= \biggl(\varphi_{p,P} - \frac{\bar\rho - \lambda_{P,1}}{2}\biggr)\sum_{i\in B} p_i y_i^2.
\end{aligned}$$

Therefore, combining (30) with (27),

$$\begin{aligned}
\lambda_{R,2} &\leq \frac{\bar\rho + \lambda_{P,1}}{2} - \biggl(\frac{(\varphi_{p,P} - (\bar\rho - \lambda_{P,1})/2)\sum_{i\in B} p_i y_i^2}{2\sum_{i\in B} p_i y_i^2}\biggr)^2\\
&= \frac{\bar\rho + \lambda_{P,1}}{2} - \frac{(\varphi_{p,P} - (\bar\rho - \lambda_{P,1})/2)^2}{2}.
\end{aligned}$$

4.2.2. *Upper bound on* $\lambda_{R,2}$ *in terms of* $\omega_{p,P}$. Returning to (31), we have

$$\begin{aligned}
\sum_{i<j} p_i R_{ij}(y_i^2 - y_j^2) &= \sum_{l=1}^{k}(y_l^2 - y_{l+1}^2)\sum_{i\in B_l, j\in B_l^{\mathrm{c}}} p_i R_{ij}\\
&= \sum_{l=1}^{k}(y_l^2 - y_{l+1}^2)\sum_{i\in B_l}\biggl(\sum_{j\in X} p_i R_{ij} - \sum_{j\in B_l} p_i R_{ij}\biggr)\\
&= \sum_{l=1}^{k}(y_l^2 - y_{l+1}^2)\biggl(\sum_{i\in B_l} p_i \sum_{j\in X} R_{ij} - Q_{p,P}(B, B)\biggr).
\end{aligned}$$

Since $l \leq k$, $p(B_l) \leq p(B) \leq \frac{1}{2}$, so $Q_{p,P}(B_l, B_l) \leq \omega_{p,P}\, p(B_l)$. Therefore,

$$\begin{aligned}
\sum_{i<j} p_i R_{ij}(y_i^2 - y_j^2) &\geq \sum_{l=1}^{k}(y_l^2 - y_{l+1}^2)\sum_{i\in B_l} p_i\biggl(\sum_{j\in X} R_{ij} - \omega_{p,P}\biggr)\\
&= \sum_{i=1}^{k} p_i \sum_{l=1}^{k}(y_l^2 - y_{l+1}^2)\biggl(\sum_{j\in X} R_{ij} - \omega_{p,P}\biggr)
\end{aligned}$$

$$= \sum_{i\in B}\sum_{j\in X} p_i y_i^2 R_{ij} - \omega_{p,P}\sum_{i\in B} p_i y_i^2$$
$$\geq \sum_{i,j\in B} p_i y_i^2 R_{ij} - \omega_{p,P}\sum_{i\in B} p_i y_i^2.$$

Since y is sorted in descending order and the sum is over $i < j$, we have $y_i^2 \leq y_i y_j$ and so

$$\begin{aligned}\sum_{i<j} p_i R_{ij}(y_i^2 - y_j^2) &\geq \sum_{i,j\in B} p_i y_i y_j R_{ij} - \omega_{p,P}\sum_{i\in B} p_i y_i^2 \\ &= \langle y, Ry\rangle_p - \omega_{p,P}\sum_{i\in B} p_i y_i^2 \\ &\geq \lambda_{R,2}\langle y, y\rangle_p - \omega_{p,P}\sum_{i\in B} p_i y_i^2 \quad \text{(by Lemma 2)} \\ &= \lambda_{R,2}\sum_{i\in B} p_i y_i^2 - \omega_{p,P}\sum_{i\in B} p_i y_i^2. \end{aligned} \tag{32}$$

Returning to (30) and substituting (32), we have

$$\begin{aligned}\lambda_{R,2} &\leq \frac{\bar{\rho} + \lambda_{P,1}}{2} - \frac{1}{2}\bigg(\frac{\lambda_{R,2}\sum_{i\in B} p_i y_i^2 - \omega_{p,P}\sum_{i\in B} p_i y_i^2}{\sum_{i\in B} p_i y_i^2}\bigg)^2 \\ &= \frac{\bar{\rho} + \lambda_{P,1}}{2} - \frac{(\lambda_{R,2} - \omega_{p,P})^2}{2}.\end{aligned}$$

Subtracting $\omega_{p,P}$ from both sides and completing the square for $(\lambda_{R,2} - \omega_{p,P})$, we obtain

$$\lambda_{R,2} \leq \sqrt{1 + \bar{\rho} + \lambda_{P,1} - 2\omega_{p,P}} - (1 - \omega_{p,P}).$$

4.3. Lower bounds on $\lambda_{R,2}$

For arbitrary $x \in \mathbb{R}^n$, we have

$$\langle Rx, x\rangle_p = \sum_{i,j\in B} p_i R_{ij} x_i x_j + 2\sum_{i\in B,\, j\in B^c} p_i R_{ij} x_i x_j + \sum_{i,j\in B^c} p_i R_{ij} x_i x_j. \tag{33}$$

Define $p_r(B) := \sum_{i\in B} p_i r_i$, and let

$$x_i = \begin{cases} p_r(B^c) & \text{if } i \in B, \\ -p_r(B) & \text{otherwise.}\end{cases}$$

We have $\langle x, r\rangle_p = 0$ and $\langle x, x\rangle_p = p(B)p_r(B^c)^2 + p(B^c)p_r(B)^2$. Substituting this specific x into (33), and noting that $Q_{p,R}(B, B^c) = Q_{p,R}(B^c, B)$, we obtain

$$\langle Rx, x\rangle_p = p_r(B^c)^2 Q_{p,R}(B, B) - 2p_r(B^c)p_r(B)Q_{p,R}(B, B^c) + p_r(B)^2 Q_{p,R}(B^c, B^c). \tag{34}$$

First note that

$$Q_{p,R}(B^c, B^c) = \lambda_{P,1} - 2Q_{p,R}(B, B^c) - Q_{p,R}(B, B),$$

and that by (9) and (13), we have

$$Q_{p,R}(B^c, B^c) = \lambda_{P,1} - 2(Q_{p,P}(B, B^c) + \tfrac{1}{2}(\lambda_{P,1}p(B) - p_\rho(B))) - p_\rho(B) + Q_{p,P}(B, B^c)$$
$$= \lambda_{P,1} - \lambda_{P,1}p(B) - Q_{p,P}(B, B^c). \tag{35}$$

Substituting (9), (13), and (35) into (34), we obtain

$$\begin{aligned}\langle Rx, x\rangle_p &= p_r(B^c)^2(p_\rho(B) - Q_{p,P}(B, B^c)) \\ &\quad - p_r(B)p_r(B^c)(2Q_{p,P}(B, B^c) + \lambda_{P,1}p(B) - p_\rho(B)) \\ &\quad + p_r(B)^2(\lambda_{P,1} - \lambda_{P,1}p(B) - Q_{p,P}(B, B^c)) \\ &= -Q_{p,P}(B, B^c) - \lambda_{P,1}p(B)p_r(B) - p_\rho(B)p_r(B) + p_\rho(B) + p_r(B)^2\lambda_{P,1} \\ &\geq -Q_{p,P}(B, B^c) - p(B)p_r(B)(\lambda_{P,1} + \bar{\rho}) + \underline{\rho}\, p(B) + p_r(B)^2\lambda_{P,1}.\end{aligned}$$

We divide both sides of this inequality by $\langle x, x\rangle_p$. Maximising over x, we obtain

$$\begin{aligned}\lambda_{R,2} &\geq \frac{-Q_{p,P}(B, B^c) - p(B)p_r(B)(\lambda_{P,1} + \bar{\rho}) + \underline{\rho}\, p(B) + p_r(B)^2\lambda_{P,1}}{p(B)p_r(B^c)^2 + p(B^c)p_r(B)^2} \\ &= \frac{-Q_{p,P}(B, B^c)/p(B)}{1 - 2p_r(B) + p_r(B)^2/p(B)} + \frac{\lambda_{P,1}p_r(B)^2/p(B) - p_r(B)(\lambda_{P,1} + \bar{\rho}) + \underline{\rho}}{1 - 2p_r(B) + p_r(B)^2/p(B)}.\end{aligned} \tag{36}$$

We have

$$\frac{-1}{1 - 2p_r(B) + p_r(B)^2/p(B)} \geq \frac{-1}{\min_{p_r(B)\in[0,1]}\{1 - 2p_r(B) + p_r(B)^2/p(B)\}}. \tag{37}$$

The denominator of the right-hand side of (37) is equal to 1 at endpoint $p_r(B) = 0$, and $p(B^c)/p(B) \geq 1$ at endpoint $p_r(B) = 1$. By differentiation, there is a critical point at $p_r(B)^* = p(B)$ which is a minimum; it is straightforward to verify that this attains the global minimum of $p(B^c)$. Furthermore, since $p(B^c) \geq \frac{1}{2}$, we have

$$\lambda_{R,2} \geq -2\frac{Q_{p,P}(B, B^c)}{p(B)} + \frac{\lambda_{P,1}p_r(B)^2/p(B) - p_r(B)(\lambda_{P,1} + \bar{\rho}) + \underline{\rho}}{1 - 2p_r(B) + p_r(B)^2/p(B)}.$$

Since this holds for all B with $p(B) \leq \frac{1}{2}$, in particular it holds for the set B^* that attains $\varphi_{p,P}$. Hence,

$$\begin{aligned}\lambda_{R,2} &\geq -2\varphi_{p,P} + \frac{\lambda_{P,1}p_r(B^*)^2 p(B^*) - p_r(B^*)(\lambda_{P,1} + \bar{\rho}) + \underline{\rho}}{1 - 2p_r(B^*) + p_r(B^*)^2/p(B^*)} \\ &\geq -2\varphi_{p,P} + \min_{\{B:\ p(B)\in[0,1/2],\ p_r(B)\in[0,1]\}} \frac{\lambda_{P,1}p_r(B)^2/p(B) - p_r(B)(\lambda_{P,1} + \bar{\rho}) + \underline{\rho}}{1 - 2p_r(B) + p_r(B)^2/p(B)}.\end{aligned} \tag{38}$$

We now evaluate the second term in (38).

Endpoint of (38) *with* $p(B) = 0$. The limit of the second term on the right-hand side of (38) as $p(B) \to 0$ is $\lambda_{P,1} + \underline{\rho}/p_r(B)^2$, which tends to a minimal value of $\lambda_{P,1}$ as $p_r(B) \to 0$.

Endpoint of (38) *with* $p(B) = \frac{1}{2}$. The second term on the right-hand side of (38) evaluates to

$$\frac{2\lambda_{P,1} p_r(B)^2 - p_r(B)(\lambda_{P,1} + \bar{\rho}) + \underline{\rho}}{1 + 2(p_r(B)^2 - p_r(B))}. \tag{39}$$

Endpoint of (39) *with* $p_r(B) = 0$. This evaluates to $\underline{\rho}$.

Endpoint of (39) *with* $p_r(B) = 1$. This evaluates to $\lambda_{P,1} - \bar{\rho} + \underline{\rho} < \underline{\rho}$.

Critical point of (39). We obtain the following equation for the value of $p_r(B)$ at the critical points of (39), i.e.

$$(\bar{\rho} - \lambda_{P,1}) p_r(B)^2 + 2(\lambda_{P,1} - \underline{\rho}) p_r(B) - \frac{\lambda_{P,1} + \bar{\rho} - 2\underline{\rho}}{2}, \tag{40}$$

which gives two critical points

$$p_r(B)^* = \frac{\pm k - 2(\lambda_{P,1} - \underline{\rho})}{2(\bar{\rho} - \lambda_{P,1})},$$

where $k = \sqrt{2((\lambda_{P,1} - \underline{\rho})^2 + (\bar{\rho} - \underline{\rho})^2)}$.

Noting that

$$k = 2\sqrt{\frac{1}{2((\lambda_{P,1} - \underline{\rho})^2 + (\bar{\rho} - \underline{\rho})^2)}} \geq 2\sqrt{(\lambda_{P,1} - \underline{\rho})^2} = 2(\lambda_{P,1} - \underline{\rho}),$$

we have

$$\frac{-k - 2(\lambda_{P,1} - \underline{\rho})}{2(\bar{\rho} - \lambda_{P,1})} < 0 \quad \text{and} \quad 0 \leq \frac{k - 2(\lambda_{P,1} - \underline{\rho})}{2(\bar{\rho} - \lambda_{P,1})} \leq 1.$$

Noting that the derivative (40) evaluates to $-(\lambda_{P,1} + \bar{\rho} - \underline{\rho})/2 < 0$ at $p_r(B) = 0$ and to $(\lambda_{P,1} + \bar{\rho} - \underline{\rho})/2 > 0$ at $p_r(B) = 1$, we conclude that

$$p_r(B)^* = \frac{k - 2(\lambda_{P,1} - \underline{\rho})}{2(\bar{\rho} - \lambda_{P,1})}$$

attains the global minimum on [0, 1]. Substituting this into (38), we obtain

$$\lambda_{R,2} \geq -2\varphi_{p,P} + \frac{2\lambda_{P,1}k^2 - k(\lambda_{P,1}^2 + \bar{\rho}^2 - 2\lambda_{P,1}\underline{\rho})}{2(4\underline{\rho}(\bar{\rho} - \underline{\rho}) + k(\bar{\rho} + \lambda_{P,1} - 2\underline{\rho}))}.$$

Critical point of (38) *with respect to* $p(B)$. Differentiating the second term in (38) with respect to $p(B)$, we obtain the following equation for the value of $p(B)$ at the critical points of the quotient, i.e.

$$p_r(B)^2(\lambda_1 - \underline{\rho} - \lambda_1 p_r(B) + \bar{\rho} p_r(B)) = 0.$$

The only possible critical point is at $p_r(B) = 0$ (the other possibility has $p_r(B) < 0$, which cannot happen). We have already examined this possibility and have shown that this is not the global minimum.

This concludes the proof of Theorem 1. □

4.4. Proof of Theorem 2

We only prove (22). Return to (36) and use the lower bound from (10) to obtain

$$\lambda_{R,2} \geq \frac{Q_{p,P}(B,B) - (\bar{\rho} + \lambda_{P,1}) p_r(B) p(B) + \lambda_{P,1} p_r(B)^2}{p(B) p_r(B^c)^2 + p(B^c) p_r(B)^2}.$$

Rearranging, we obtain

$$\begin{aligned} &Q_{p,P}(B,B) \\ &\quad \leq \lambda_{R,2}(p(B)p_r(B^c)^2 + p(B^c)p_r(B)^2) + (\bar{\rho} + \lambda_{P,1}) p_r(B) p(B) - \lambda_{P,1} p_r(B)^2 \\ &\quad = \lambda_{R,2}(p(B) - 2p(B)p_r(B) + p_r(B)^2) + (\bar{\rho} + \lambda_{P,1}) p_r(B) p(B) - \lambda_{P,1} p_r(B)^2. \end{aligned}$$

Dividing through by $p(B)$,

$$\begin{aligned} \frac{Q_{p,P}(B,B)}{p(B)} &\leq \lambda_{R,2}\bigg(1 - 2p_r(B) + \frac{p_r(B)^2}{p(B)}\bigg) + (\bar{\rho} + \lambda_{P,1}) p_r(B) - \frac{\lambda_{P,1} p_r(B)^2}{p(B)} \\ &= \lambda_{R,2}(1 - 2p_r(B)) + \frac{(\lambda_{R,2} - \lambda_{P,1}) p_r(B)^2}{p(B)} + (\bar{\rho} + \lambda_{P,1}) p_r(B). \end{aligned}$$

The right-hand side is a monotonic function of $p(B)$, which lies in the interval $(0, \frac{1}{2}]$. If $\lambda_{R,2} - \lambda_{P,1} > 0$, the right-hand side is a decreasing function of $p(B)$, which approaches $+\infty$ as $p(B) \to 0$. On the other hand, if $\lambda_{R,2} - \lambda_{P,1} < 0$ as per the conditions of Theorem 2 then the right-hand side is an increasing function of $p(B)$. The global maximum on the interval $(0, \frac{1}{2}]$ occurs when $p(B) = \frac{1}{2}$, at which value the second term equals $2(\lambda_{R,2} - \lambda_{P,1}) p_r(B)^2$. Hence,

$$\frac{Q_{p,P}(B,B)}{p(B)} \leq \lambda_{R,2}(1 - 2p_r(B)) + 2(\lambda_{R,2} - \lambda_{P,1}) p_r(B)^2 + (\bar{\rho} + \lambda_{P,1}) p_r(B).$$

We differentiate the right-hand side with respect to $p_r(B)$ to find a maximum at $p_r(B)^* = (\lambda_{P,1} + \bar{\rho} - 2\lambda_{R,2})/4(\lambda_{P,1} - \lambda_{R,2}) > 0$. Thus,

$$\begin{aligned} &\frac{Q_{p,P}(B,B)}{p(B)} \\ &\quad \leq -\frac{(\lambda_{P,1} + \bar{\rho} - 2\lambda_{R,2})^2}{8(\lambda_{P,1} - \lambda_{R,2})} + \frac{(\lambda_{P,1} + \bar{\rho} - 2\lambda_{R,2})^2}{4(\lambda_{P,1} - \lambda_{R,2})} + \lambda_{R,2} \\ &\quad = \frac{(\lambda_{P,1} + \bar{\rho} - 2\lambda_{R,2})^2}{8(\lambda_{P,1} - \lambda_{R,2})} + \lambda_{R,2} \\ &\quad = \frac{(\lambda_{P,1}^2 + 2\bar{\rho}\lambda_{P,1} - 4\lambda_{P,1}\lambda_{R,2} - 4\bar{\rho}\lambda_{R,2} + \bar{\rho}^2 + 4\lambda_{R,2}^2) + 8\lambda_{R,2}(\lambda_{P,1} - \lambda_{R,2})}{8(\lambda_{P,1} - \lambda_{R,2})} \\ &\quad = \frac{\lambda_{P,1}^2 - 2\bar{\rho}\lambda_{P,1} + \bar{\rho}^2 + 4\bar{\rho}(\lambda_{P,1} - \lambda_{R,2}) + 4\lambda_{R,2}(\lambda_{P,1} - \lambda_{R,2})}{8(\lambda_{P,1} - \lambda_{R,2})} \\ &\quad = \frac{(\bar{\rho} - \lambda_{P,1})^2 + 4(\bar{\rho} + \lambda_{R,2})(\lambda_{P,1} - \lambda_{R,2})}{8(\lambda_{P,1} - \lambda_{R,2})} \\ &\quad = \frac{\bar{\rho} + \lambda_{R,2}}{2} + \frac{(\bar{\rho} - \lambda_{P,1})^2}{8(\lambda_{P,1} - \lambda_{R,2})}. \end{aligned}$$

This concludes the proof of Theorem 2 □

5. Discussion

We contrast our results with the conductance-based bounds for Markov processes with killing investigated by Lawler and Sokal [26] for continuous-time Markov processes defined on a measurable set of states. We briefly recap their results and translate them to our finite-state setting. Consider a continuous-time Markov process with transition rate matrix $\bar{A}_{ij} = \eta_{ij}$, where $\eta_{ii} := -\sum_{j\neq i} \eta_{ij}$. Lawler and Sokal considered processes that are positive-recurrent and irreducible; these two conditions ensure there is a unique finite invariant probability measure for the process [23], denoted here by $\bar{p}$. Next suppose that the process at state i is killed (i.e. exits the set of states permanently) at rate $k_i \geq 0$, and define the killing matrix K with entries k_i on the diagonal and 0 elsewhere. The rate matrix $A = \bar{A} - K$ for the process with killing is given by

$$A_{ij} = \begin{cases} \eta_{ii} - k_i & \text{if } i = j, \\ \eta_{ij} & \text{otherwise,} \end{cases}$$

for $1 \leq i, j \leq n$. The definition of the minimal $\bar{p}$-flux rate under A for the process with killing is given in [26, Equations (3.3) and (3.4)], as $\phi_{\bar{p},A} := \min_{B\in X,\ \bar{p}(B)>0} \phi_{\bar{p},A}(B)$, where

$$\begin{aligned} \phi_{\bar{p},A}(B) :&= \frac{\sum_{i\in B,\ j\notin B} \bar{p}_i \eta_{ij} + \sum_{i\in B} \bar{p}_i k_i}{\bar{p}(B)} \\ &= \frac{-\sum_{i,j\in B} \bar{p}_i \eta_{ij} + \sum_{i\in B} \bar{p}_i k_i}{\bar{p}(B)} \\ &= -\frac{\sum_{i,j\in B} \bar{p}_i A_{ij}}{\bar{p}(B)}. \end{aligned} \tag{41}$$

The numerator in (41) measures both the rate of flow to $B^{\mathrm{c}} := X \setminus B$ *and* the rate of killing. Bounds on the largest eigenvalue of A in terms of $\phi_{\bar{p},A}$ are given in [26, Theorem 3.1].

Theorem 3. (Lawler and Sokal [26].) *Let $A = \bar{A} - K$ and suppose that the invariant probability measure $\bar{p}$ for the process without killing satisfies $\bar{p}_i \eta_{ij} = \bar{p}_j \eta_{ji}$ for all $i, j \in X$. Denote the largest eigenvalue of A by $\lambda_{A,1}$. Choose M so that*

$$M \geq \frac{1}{\bar{p}_j}\bigg(\sum_{i\in X} \bar{p}_i \eta_{ij} + \frac{1}{2} k_j\bigg) \quad \textit{for all } j \in X.$$

Then

$$-\frac{\phi_{\bar{p},A}^2}{2M} \leq \lambda_{A,1} \leq -\phi_{\bar{p},A}.$$

Theorem 3 bounds the leading eigenvalue of the Markov process in the reversible setting, while our Corollary 1 bounds the second eigenvalue, which is concerned with the mixing rate. Furthermore, Theorem 3 considers the conditional probability of points leaving $B \subset X$ according to $\bar{p}$ (the stationary distribution for the process *without* killing), whereas Corollary 1 considers the conditional probabilities according to p (the quasistationary distribution for the process *with* killing).

To conclude, we derived new bounds on the second eigenvalue of a substochastic transition probability matrix in terms of the conductance and metastability. Our results extend existing analogous results for stochastic transition probability matrices, and are consistent with these well-known results when the transition matrix is stochastic. If P is p-reversible, the bounds on the second eigenvalue (Corollary 1) rely only on the largest eigenvalue and either the conductance or the metastability. If P is not p-reversible, the second largest eigenvalue of a p-symmetrised chain is bounded (Theorem 1) in terms of the maximal and minimal row sums of P, the leading eigenvalue of P, and either the conductance or metastability. The bounds for the non-p-reversible case reduce to the bounds for the p-reversible case if P is p-reversible. Finally, we note that the 'inverse problem' of determining metastable subsets of states from dominant eigenvectors of P is also of interest because it reveals important information on the mixing properties of the system. The connection between the spectra of Markov chain transition matrices and the existence of metastable sets is well established for stochastic Markov chains; in Theorem 2 we extended this relationship to the substochastic setting, which is relevant for open dynamical systems.

Acknowledgements

GF was supported by an Australian Research Council Future Fellowship. RMS was supported by an Australian Postgraduate Award, the ARC Centre of Excellence for the Mathematics and Statistics of Complex Systems (MASCOS), and the University of Copenhagen (UCPH) Excellence Programme for Interdisciplinary Research.

References

[1] Arjas, E., Nummelin, E. and Tweedie, R. L. (1980). Semi-Markov processes on a general state space: α-theory and quasistationarity. *J. Austral. Math. Soc. A* **30,** 187–200.

[2] Asselah, A., Ferrari, P. A. and Groisman, P. (2011). Quasistationary distributions and Fleming–Viot processes in finite spaces. *J. Appl. Prob.* **48,** 322–332.

[3] Bartlett, M. S. (1960). *Stochastic Population Models in Ecology and Epidemiology*. Methuen, London.

[4] Cheeger, J. (1970). A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in Analysis (Papers dedicated to Salomon Bochner)*, Princeton University Press, pp. 195–199.

[5] Cheong, C. K. (1970). Quasi-stationary distributions in semi-Markov processes. *J. Appl. Prob.* **7,** 388–399, 788.

[6] Chung, F. (2005). Laplacians and the Cheeger inequality for directed graphs. *Ann. Combinatorics* **9,** 1–19.

[7] Collet, P., Martínez, S. and San Martín, J. (2013). *Quasi-Stationary Distributions: Markov Chains, Diffusions and Dynamical Systems*. Springer, Heidelberg.

[8] Darroch, J. N. and Seneta, E. (1965). On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *J. Appl. Prob.* **2,** 88–100.

[9] Darroch, J. N. and Seneta, E. (1967). On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Prob.* **4,** 192–196.

[10] Dellnitz, M. and Junge, O. (1999). On the approximation of complicated dynamical behavior. *SIAM J. Numerical Anal.* **36,** 491–515.

[11] Deuflhard, P., Huisinga, W., Fischer, A. and Schütte, C. (2000). Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Linear Algebra Appl.* **315,** 39–59.

[12] Ewens, W. J. (1963). The diffusion equation and a pseudo-distribution in genetics. *J. R. Statist. Soc. B* **25,** 405–412.

[13] Ewens, W. J. (1964). The pseudo-transient distribution and its uses in genetics. *J. Appl. Prob.* **1,** 141–156.

[14] Ferrari, P. A. and Marić, N. (2007). Quasi-stationary distributions and Fleming–Viot processes in countable spaces. *Electron. J. Prob.* **12,** 684–702.

[15] Ferrari, P. A., Kesten, H., Martinez, S. and Picco, P. (1995). Existence of quasi-stationary distributions. A renewal dynamical approach. *Ann. Prob.* **23,** 501–521.

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

[17] Flaspohler, D. C. and Holmes, P. T. (1972). Additional quasi-stationary distributions for semi-Markov processes. *J. Appl. Prob.* **9,** 671–676.

[18] Fleming, W. H. and Viot, M. (1979). Some measure-valued Markov processes in population genetics theory. *Indiana Univ. Math. J.* **28,** 817–843.

[19] Froyland, G. (2005). Statistically optimal almost-invariant sets. *Phys. D* **200,** 205–219.

[20] Froyland, G. and Padberg, K. (2009). Almost-invariant sets and invariant manifolds—connecting probabilistic and geometric descriptions of coherent structures in flows. *Phys. D* **238,** 1507–1523.
[21] Froyland, G., Pollett, P. K. and Stuart, R. M. (2014). A closing scheme for finding almost-invariant sets in open dynamical systems. *J. Comput. Dynamics* **1,** 135–162.
[22] Huisinga, W. and Schmidt, B. (2006). Metastability and dominant eigenvalues of transfer operators. In *New Algorithms for Macromolecular Simulation* (Lecture Notes Comput. Sci. Eng. **49**), Springer, Berlin, pp. 167–182.
[23] Kelly, F. P. (1979). *Reversibility and Stochastic Networks*. John Wiley, Chichester.
[24] Kelly, F. P. (1983). Invariant measures and the q-matrix. In *Probability, Statistics and Analysis* (London Math. Soc. Lecture Notes Ser. **79**), Cambridge University Press, pp. 143–160.
[25] Kingman, J. F. C. (1969). Markov population processes. *J. Appl. Prob.* **6,** 1–18.
[26] Lawler, G. F. and Sokal, A. D. (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.* **309,** 557–580.
[27] Levin, D. A., Peres, Y. and Wilmer, E. L. (2009). *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI.
[28] Meyer, C. (2000). *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, PA.
[29] Nair, M. G. and Pollett, P. K. (1993). On the relationship between μ-invariant measures and quasi-stationary distributions for continuous-time Markov chains. *Adv. Appl. Prob.* **25,** 82–102, 717–719.
[30] Pakes, A. G. (1993). Absorbing Markov and branching processes with instantaneous resurrection. *Stoch. Process. Appl.* **48,** 85–106.
[31] Pollett, P. K. (1988). Reversibility, invariance and μ-invariance. *Adv. Appl. Prob.* **20,** 600–621.
[32] Seneta, E. and Vere-Jones, D. (1966). On quasi-stationary distributions in discrete-time Markov chains with a denumerable infinity of states. *J. Appl. Prob.* **3,** 403–434.
[33] Sinclair, A. and Jerrum, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Inf. Comput.* **82,** 93–133.
[34] Steinsaltz, D. and Evans, S. N. (2004). Markov mortality models: implications of quasistationarity and varying initial distributions. *Theoret. Pop. Biol.* **65,** 319–337.
[35] Tweedie, R. L. (1974). Some ergodic properties of the Feller minimal process. *Quart. J. Math. Oxford Ser. (2)* **25,** 485–495.
[36] Van Doorn, E. A. and Pollett, P. K. (2008). Survival in a quasi-death process. *Linear Algebra Appl.* **429,** 776–791.
[37] Van Doorn, E. A. and Pollett, P. K. (2009). Quasi-stationary distributions for reducible absorbing Markov chains in discrete time. *Markov Process. Relat. Fields* **15,** 191–204.
[38] Yaglom, A. M. (1947). Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.)* **56,** 795–798.