Speed of convergence to stationarity for stochastically monotone Markov chains

A DOCTORAL DISSERTATION

BY

PAWEŁ LOREK

MATHEMATICAL INSTITUTE
UNIVERSITY OF WROCŁAW

SUPERVISOR: PROF. RYSZARD SZEKLI

WROCŁAW, MAY 2007

Acknowledgements

I would like to express my deep and sincere gratitude to my supervisor Professor Ryszard Szekli for his guidance and assistance through the whole process of writing this thesis.

I also want to thank Cornelia Wichelhaus née Sauer and Hans Daduna for inspiring research cooperation.

Contents

1 Introduction

The main aim of this thesis is to study convergence rate to stationarity of selected Markov chains. Some stress we put on the property of stochastic monotonicity introduced by Daley in [12]. Expressing speed of convergence depends on how we measure the distance. The measures we use are total variation distance, separation distance and $L_2(\pi)$ distance. If we write about some distance at time instant t (either real or discrete), we always mean the distance between the distribution of the chain at time t and its stationary distribution.

Many authors faced such problems. For example, Fisz [21] using matrix analysis showed geometric rate of convergence in total variation distance for finite state space assuming only that transition matrix $\mathbf{P}$ is regular. Feller [19] and Kingman [29] used another approach: they incorporated renewal equations to Markov chains and using some facts from the number theory he showed when ergodicity holds. The classical theorem of Perron and Frobenius (which can be found for example in Seneta [43]) can be used to show that for finite state spaces, geometric rate of convergence holds (see e.g. Cinlar [9]). Having second-largest eigenvalue one can have upper bound for total variation distance. However, calculating eigenvalues, even in quite simple cases, is rather difficult. For reversible Markov chains there are established ways to bound it, for example the techniques using Cheeger's or Poincare's constants. We give an example of calculating Poincare's constant for symmetric random walk on the cube and both constants for random walk on $\mathbb{Z}_n$. This technique was investigated by many authors, see for example Diaconis and Stroock [16], Chen and Wang [7], Fulman and Wilmer [22], Chung [8]. However, till 1991 most of the research on random walks has concerned only reversible Markov chains. In 1991, Fill [20] showed that one can use symmetrization: multiply matrix of original (potentially non-reversible) process $\mathbf{X}$ and matrix of time-reversed process $\widetilde{\mathbf{X}}$. That way one obtains a matrix of some new reversible process which is useful to bound total variation distance in original $\mathbf{X}$. We give details of proofs in this thesis.

The coupling method is well known and widely used method in examining the rate of convergence in Markov chains. The idea is to construct a joint process $\{(X_n, Y_n), n \geq 0\}$ such that marginally $\{X_n, n \geq 0\}$ is a Markov chain with given matrix $\mathbf{P}$ and arbitrary initial distribution μ, and $\{Y_n, n \geq 0\}$ is also a Markov chain with the same matrix $\mathbf{P}$ but with initial distribution π being its stationary distribution. Then observe the chains till the first time they meet (this is random variable T called coupling time). By estimating the tail of coupling times, one is able to bound the total variation distance for convergence. Classical references for coupling are for example Lindvall [34], Liggett [32] and Thorisson [46].

Strong Stationary Times give a probabilistic approach to bounding the rate of convergence to stationarity for Markov chains. They were introduced by Aldous and Diaconis in [2], were they gave sharp bounds in examples which were not amenable to other techniques such as spectral analysis or coupling. The same authors in [3] develop this technique, showing for example that Minimal Strong Stationary Time do always exists for ergodic Markov chains. Other references on Strong Stationary Times are: book of Diaconis [13], thesis by I. Pak [37], Saloff-Coste [40] and Diaconis and Bayer [14]. Diaconis and Fill in [15] show that there always exists a dual chain, in which first hitting time to some absorbing state is equal, in distribution, to Strong Stationary Time in the original chain. We give detailed proofs of most important theorems, some known examples to illustrate the method and use it to show that $2n \log n + cn$ steps are enough to make the total variation distance in example of "matching in graph" smaller then $1/c^2$, where $2n$ is the number of vertices and c a positive constant.

We study also factorization of first passage times distributions in presence of some monotonicity properties in Markov chains. Under some assumptions the first passage time from initial distribution μ to a selected state $\tilde{\mathbf{e}}$ has the same distribution as the sum of two independent random variables: $Y_{\tilde{\mathbf{e}}}$ (with distribution function $P(Y_{\tilde{\mathbf{e}}} \leq n) = \frac{\mu \mathbf{P}^n(\tilde{\mathbf{e}})}{\tilde{\mathbf{e}}}$) plus the first passage time from stationary distribution to $\tilde{\mathbf{e}}$. Such factorization leads to explicit formulas for Minimal Strong Stationary Time and for separation distance. We give some special cases for such factorizations. One of them is when the time-reversed process is stochastically monotone. Such monotonicity was studied for example by Keilson and Kester [28], Aldous [1] and Tweedie and Roberts [47]. We exploit this property in two examples: non-symmetric

random walk on $d-$dimensional cube and in the closed tandem with 3 servers and N customers. While studying a preprint of Brown [5] we found a false statement. The conditions on the parameters of the random walk given by Brown do not guarantee stochastic monotonicity. We elaborate this topic and give conditions for stochastic monotonicity and conditions to have factorizations. For random walk on the cube we also indicate the following phenomenon: in discrete time we need some (quite sophisticated) conditions on parameters in order to have described factorizations, whereas in continuous time we always have them (assuming just things which guarantee irreducibility and ergodicity).
It is also worth noting, that some monotonicity properties in Markov chain play important role in "coupling from the past", the algorithm which returns unbiased sample from distribution being stationary distribution of given chain, for details see Propp and Wilson [38], Häggström [23].

In more complicated systems, especially for chains with infinite state space, like queueing networks, it is difficult to estimate exact convergence rate in general. It is desirable to know if there exists geometric rate of convergence under some natural assumptions. Fayolle, Malyshev, Menshikov, Sidorenko [18] showed (using Foster-Lyapunov criteria) that there exists geometric rate of convergence in total variation distance in Jackson networks, provided the service intensities $\mu_i(\cdot) = \mu_i$ are constant, i.e. they do not depend on the number of customers waiting to be served and that mean service time is shorter then mean time of waiting for new customer. In the same year, Hordijk and Spieksma [24] showed similar result also obtaining geometric rate of convergence, but measured in a different distance.
The unreliable Jackson network is the network where some nodes can be broken or repaired. The breakdowns and repairs events are of rather general structure. In case when nodes are broken, there are several rules what can be done. Most often routing matrix is changed. For full details of such networks see Sauer and Daduna [42], and Sauer [41]. In the latter also a rich historical overview of such networks can be found. Sauer in her thesis [41] showed (using techniques of Malyshev *et al.*) that in case of constant service rates $\mu_i(\cdot) = \mu_i$ and without rerouting (i.e. customers are allowed to join the queue at broken server, where it waits till repair, the stationary distribution is not known in this case) there is geometric rate of convergence in total variation distance under some conditions.
Stationary distribution of Jackson network has one nice property: it is of product form. Sauer and Daduna [42] showed that the same property has stationary distribution of unreliable Jackson network. Product form of stationary distribution is the crucial property used in proof of our main result on the existence of spectral gap for general networks.
The only article about spectral gap for Jackson networks with non-constant service rates $\mu_i(\cdot)$, as far as we know, is McDonald and Iscoe [36]. They show, working on a different problem, that if a stationary distribution is a product of light-tailed distributions, then the spectral gap exists.
In this thesis we present the stronger, "if and only if", result which holds in both standard and unreliable Jackson networks: spectral gap exists if and only if stationary distribution is a product of light-tailed distributions, i.e. there is surely no spectral gap when at least one of these distributions is heavy-tailed and surely there is spectral gap when all of them are light-tailed. It is worth noting that in general it does not need to be the case: there are processes with positive spectral gap, although stationary distribution is heavy-tailed. Our proof differs from the one of McDonald and Iscoe, for comparison we also include their proof.

All facts taken from literature have bibliographic citations.

In section 3 we prove Theorem 3.1.4, which is a known fact from the number theory, but we did not succeed in finding a complete proof of it. Theorem 3.4.1 is a reformulated result of Fisz [21], where in our proof we stressed the role of total variation distance. We introduce Lemma 3.8.5 (for reversible chains and 3.9.3 for non-reversible), where we give necessary number of steps needed to make total variation distance arbitrary small, given we have Poincare's or Cheeger's constant calculated. In subsection 3.8.1 we calculate Poincare's bound for symmetric random walk on d-dimensional cube. In subsection 3.9.2 (clockwise random walk on $\mathbb{Z}_n$, example of Fill [20]) we give full details in calculating Poincare's and Cheeger's constant and we compare them. Subsection 3.11 is fully taken from paper of the author of this thesis *et al.* [17], where classical methods to study speed of convergence in an applied model are used.

Section 4 is a reformulation and recompilation of papers by Aldous and Diaconis [2], [3], Diaconis and Fill [15] and thesis of Pak [37]. The proof of Lemma 4.1.5 is given with full details. Besides known examples, where Strong Stationary Time was used to find the mixing time, we give new example "matching in graph" is subsection 4.6.

In section 5 we give some factorization of passage time distributions and exact formulas for separation distance. It is based on a preprint by Brown [5], but we introduce notion of ratio minimality and utilize it in statements. Proofs of Lemma 5.1.3 and 5.3.3 are completely different then the ones of Brown. Brown proved everything in continuous time. In Lemma 5.1.13 we show that positive spectrum of transition matrix implies (by Lemma 5.1.5) some factorizations, and in Lemma 5.1.14 we show that if we observe only every second step of a reversible Markov chain, then we always have some factorizations.
In subsection 5.4 we investigate example of Brown "non-symmetric random walk on d-dimensional cube" and indicate that he was wrong in his preprint [5], where his conditions did not imply stochastic monotonicity. We introduce another partial ordering and give sufficient conditions in Theorem 5.4.2. We also discover a phenomena in non-symmetric random walk on cube: the difference in behaviour between continuous and discrete time. In discrete time we have to have some conditions on parameters in order to have ratio minimality (we give an example which did not fulfill these conditions and in which there is no ratio minimality), whereas these conditions are not needed in continuous time, as stated in Theorem 5.4.5.
In subsection 5.5 we give one more example illustrating tools introduced in section 5. We investigate closed tandem of 3 servers and N customers and we introduce partial ordering of state space and show that time-reversed process is stochastically monotone what implies ratio minimality and thus exact formula for separation distance.

Further results of the thesis are in section 6. Under some natural assumptions, we show that spectral gap exists (and thus we have geometric rate of convergence) in Jackson network if and only if each of marginal distributions is light-tailed. This is done for both: standard (Theorem 6.5.2) and unreliable (Theorem 6.5.6) Jackson networks. The idea of proof is to take process which is a product of independent birth and death processes. This process has the same stationary distribution as original network process, and it is easier to show that spectral gap for this process exists if and only if stationary distributions of birth and death processes are light-tailed. The main part of proof is to conclude that if spectral gap for this process is equal to zero, then the same holds for original process, and that existence of spectral gap for this process implies existence of spectral gap for original process. This is done by comparing Cheeger's constants of both processes, and can be done due to the fact, that stationary distribution is of product form.

2 Markov chain definitions and notation, discrete time

2.1 Notation

For a Markov chain $\mathbf{X} = \{X_n, n \geq 0\}$ with a state space E, by π we denote the stationary measure, and by, μ we denote initial measure on E. We assume E is finite or enumerable and its elements will be denoted $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots\}$, and in the case when we have linear ordering: $E = \{1, 2, \ldots\}$.
We use the following notation:

$\delta_{\mathbf{e}}\mathbf{P}^n(B)$	$=$	$\mathbf{P}_{\mathbf{e}}(X_n \in B) = P(X_n \in B \mid X_0 = \mathbf{e})$	$-$	Probability of hitting B at, time n starting from $X_0 = \mathbf{e}$ ($\delta_{\mathbf{e}}$ denotes atom at $\mathbf{e}$)
$\tau_{\mathbf{e}}(B)$	$=$	$\inf\{n : X_n \in B, X_0 = \mathbf{e}\}$	$-$	First passage time to B starting from $X_0 = \mathbf{e}$
$\tau_\mu(B)$	$=$	$\int_E \tau_{\mathbf{e}}(B) d\mu(\mathbf{e})$	$-$	First passage time to B starting with $X_0 \sim \mu$
$\mu\mathbf{P}^n(B)$	$=$	$P_\mu(X_n \in B)$	$-$	Probability of hitting B at time n starting with $X_0 \sim \mu$

If set B will be one state set, i.e. $B = \{\mathbf{e}\}$, it will be shortly written $B := \mathbf{e}$.

We denote the Laplace transform of discrete, real-valued random variable Y by

$$\psi_Y(s) = \sum_{n=-\infty}^{\infty} e^{-sn} P(Y = n).$$

We assume that the state space E will be partially ordered with an ordering $\prec$. Set A is an upper (denoted by $A \uparrow$), if $x \in A$ and $x \prec y$ implies $y \in A$. Similarly set B is a lower set ($B \downarrow$), if $x \in B$ and $y \prec x$ implies $y \in B$.
Function $h \colon E \to R$ is increasing relative to $\prec$ if $x \prec y \Rightarrow h(x) \leq h(y)$.

We define stochastic ordering by:

$$\mu \prec_{st} \nu \qquad \Longleftrightarrow \qquad \forall (A \uparrow) \quad \mu(A) \leq \nu(A) \qquad \Big(\equiv \qquad \forall (f \text{ increasing} \int f d\mu \leq \int f d\nu \Big). \tag{2.1}$$

Define transition matrix $\mathbf{P}$ with elements being probabilities of getting from one state to another. We will shortly write $\mathbf{P}(i,j) := \mathbf{P}(\mathbf{e}_i, \mathbf{e}_j) = P(X_{n+1} = \mathbf{e}_j \mid X_n = \mathbf{e}_i)$. If $E = \{1, 2, \ldots, \}$ then we will use $p_{ij} := \mathbf{P}(i,j)$.
The distribution of the chain, which started with initial distribution μ, at $k-$th step will be denoted by $\mu\mathbf{P}^k(\cdot)$.

Markov chain $\{X_n, n \geq 0\}$ with transition matrix $\mathbf{P}$ is **stochastically monotone** if

$$\forall (\mu \prec_{st} \nu) \qquad \mu\mathbf{P} \prec_{st} \nu\mathbf{P} \qquad \Big(\equiv \qquad \forall (f \text{ increasing}) \quad \mathbf{P}f \text{ is increasing} \Big). \tag{2.2}$$

If $|E| = N$, and $\prec$ is the linear ordering, then stochastic monotonicity is equivalent to condition:

$$\mathbf{t}^{-1}\mathbf{P}\mathbf{t} \geq 0 \qquad \equiv \qquad (\mathbf{t}^{-1})^T \mathbf{P} \mathbf{t}^T \geq 0, \qquad \mathbf{P} = [p_{ij}], \tag{2.3}$$

where $\mathbf{t}$ is of size $N \times N$ and is defined as follows:

$$\mathbf{t} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \ldots & \ldots & \ldots & \ldots & 1 \end{pmatrix}, \quad \mathbf{t}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & -1 & 1 \end{pmatrix}$$

2.2 Distances: total variation norm, separation distance

Let μ, ν be measures on $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots\}$. One of the most popular distance between two probability measures is given by the **total variation norm**:

$$d(\mu,\nu) = \max_{B\subset E} |\mu(B) - \nu(B)|. \tag{2.4}$$

On Figure 1,

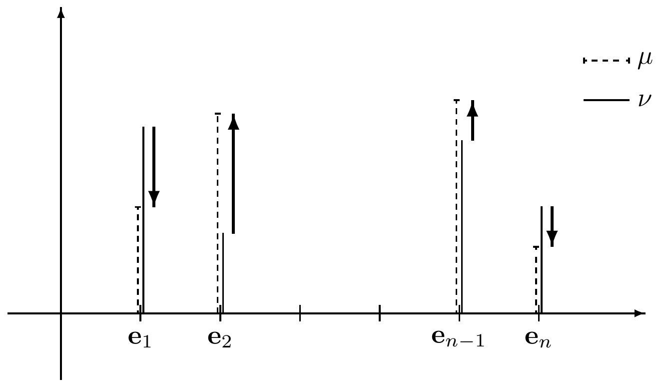


Figure 1: Example: $d(\mu,\nu)$

the length of arrows denotes $\mu(\mathbf{e}_i) - \nu(\mathbf{e}_i)$ and orientation show if it is positive or negative. Of course all the arrows sum up to 0, so looking closer at the definition of $d(\mu,\nu)$ we have the following equivalent equalities:

$$d(\mu,\nu) = \max_{B\subset E} |\mu(B) - \nu(B)| = \sum_{\mathbf{e}:\mu(\mathbf{e})\geq\nu(\mathbf{e})} (\mu(\mathbf{e}) - \nu(\mathbf{e})) = \sum_{\mathbf{e}:\nu(\mathbf{e})\geq\mu(\mathbf{e})} (\nu(\mathbf{e}) - \mu(\mathbf{e})) =$$
$$\sum_{\mathbf{e}\in E} (\mu(\mathbf{e}) - \nu(\mathbf{e}))_+ = \sum_{\mathbf{e}\in E} (\nu(\mathbf{e}) - \mu(\mathbf{e}))_+ = \frac{1}{2}\sum_{\mathbf{e}\in E} |\mu(\mathbf{e}) - \nu(\mathbf{e})|. \tag{2.5}$$

We define **separation distance** as follows:

$$s(\mu,\nu) = \max_{\mathbf{e}\in E} \left(1 - \frac{\mu(\mathbf{e})}{\nu(\mathbf{e})}\right), \tag{2.6}$$

where $\frac{1}{0} := \infty$. Or equivalently $s(\mu,\nu)$ is the smallest $s \geq 0$ such that:

$$\mu = (1-s)V + sV, \tag{2.7}$$

for some distribution V.
Note that s is not a metric, it is possible that $s(\mu,\nu) \neq s(\nu,\mu)$. If the support of μ is contained in that of ν then $s(\mu,\nu)$ is equal to 1.
A useful property of s is that it is an upper bound for d:

$$d(\mu,\nu) = \sum_{\mathbf{e}\in E} (\nu(\mathbf{e}) - \mu(\mathbf{e}))_+ = \sum_{\mathbf{e}\in E} \nu(\mathbf{e}) \left(1 - \frac{\mu(\mathbf{e})}{\nu(\mathbf{e})}\right)_+ \leq \sum_{\mathbf{e}\in E} \nu(\mathbf{e}) \max_k \left(1 - \frac{\mu(\mathbf{e}_k)}{\nu(\mathbf{e}_k)}\right)$$
$$= s(\mu,\nu) \sum_{\mathbf{e}\in E} \nu(\mathbf{e}) = s(\mu,\nu)\cdot 1.$$

So we have

$$d(\mu,\nu) \leq s(\mu,\nu). \tag{2.8}$$

There is no general inverse inequality: let for example ν be uniform distribution on finite E and μ uniform distribution on $E \setminus \{\mathbf{e}\}$. Then $d(\mu,\nu) = \frac{1}{|E|}$, while $s(\mu,\nu) = 1$.

3 Classical ergodic theorems for Markov chains

3.1 Facts from number theory

We will use 3 equalities:

$$(*) \qquad a \bmod n = a - \left\lfloor \tfrac{a}{n} \right\rfloor n,$$

$$(**) \qquad d|a \text{ i } d|b \quad \Rightarrow \quad d|(ax+by) \qquad \forall(x,y \in \mathbb{Z}),$$

$$(***) \qquad \gcd\{a,b,c\} = \gcd\{\gcd\{a,b\},c\},$$

where $\lfloor x \rfloor = \sup\{n \in \mathbb{Z} : x \leq n\}$.
We shall use the following theorem, which we give with a proof for completeness.

Theorem 3.1.1 (Cormen *et al.* [10])**.** *If a and b are any integer numbers not both equal to 0, then $\gcd\{a,b\}$ is the smallest* **positive** *element of the set $\{ax+by : x,y \in \mathbb{Z}\}$ - linear combination of a i b.*

Proof. Let s be the smallest such combination, i.e. let $s = ax+by$ for some $x,y \in \mathbb{Z}$. Let $q = \left\lfloor \frac{a}{s} \right\rfloor$. Then from equation $(*)$:

$$a \bmod \mathrm{s} = \mathrm{a} - \mathrm{qs} = \mathrm{a} - \mathrm{q(ax+by)} = \mathrm{a(1-qx)} + \mathrm{b(-qy)},$$

thus a mod s is also a linear combination of a and b. But a mod s < s therefore a mod s = 0, because s is the smallest positive linear combination. So $s|a$ and analogically (i.e. considering b mod s) $s|b$, so s is a common divisor of both numbers and therefore $s \leq \gcd\{a,b\}$, thus from $(**)$ we have that $\gcd\{a,b\}|s$ and it is enough to take $d = \gcd\{a,b\}$ - of course then $d|a$ and $d|b$, and s is a linear combination of a and b, so we have $s \geq \gcd\{a,b\}$.
If $s \geq \gcd\{a,b\}$ and $s \leq \gcd\{a,b\}$ then $s = \gcd\{a,b\}$. □

Conclusion 3.1.2. *For any integers a and b such that they both are not equal 0 there exist $x,y \in \mathbb{Z}$ such that*

$$\gcd\{a,b\} = ax+by.$$

□

Lemma 3.1.3. *For any positive integers a,b and k and for $c = k \cdot \gcd\{a,b\}$ there exist $x,y \in \mathbb{N} \setminus \{0\}$ such that*

$$ax - by = c. \tag{3.9}$$

Proof. From Conclusion 3.1.2 there exist $x', y' \in \mathbb{Z}$ such that

$$\gcd\{a,b\} = ax' + by'.$$

Multiplying both sides by k we obtain

$$c = k \cdot \gcd\{a,b\} = akx' + bky'.$$

Observe that $x = bh + kx'$, $y = ah - ky'$ for any $h \in \mathbb{N}$ are solutions to the (3.9):

$$ax - by = a(bh+kx') - b(ah-ky') = abh + akx' - abh + bky' = akx' + bky' = k \cdot \gcd\{a,b\} = c,$$

and we can take any h, so take such that both x and y are positive.

□

The next theorem is a known fact from the number theory, but the author of this thesis did not succeed in finding a complete proof of it, therefore a proof is provided here.

Theorem 3.1.4. *Let $a_1, \ldots, a_m$ be any natural numbers ≥ 2 such that $a_i \neq a_j$ for $i \neq j$. Then any number k being multiplicity of $\gcd\{a_1,\ldots,a_m\}$ and $k > a_1 \cdots a_m$ can be written in the following representation: $k = a_1x_1 + + \ldots + a_mx_m$, where $x_i, j = 1,\ldots,m$ are integer and* **positive.**

Proof. By induction:
For $\mathbf{m = 2}$
We want to write k as a linear combination: $k = r \cdot \gcd\{a_1, a_2\}$. From Conclusion 3.1.2 there exist y_1' and y_2' such that $\gcd\{a_1, a_2\} = a_1 y_1' + a_2 y_2'$. Multiplying both sides by r we have

$$k = r \cdot \gcd\{a_1, a_2\} = a_1 r y_1' + a_2 r y_2' = a_1 y_1 + a_2 y_2,$$

where $y_1 = r y_1',\ y_2 = r y_2'$.

If y_1 and y_2 are non-negative the proof is finished. Otherwise one of y_1 and y_2 is negative (both cannot). Assume $y_1 > 0$ and $y_2 < 0$. Then add $+a_1 a_2 - a_1 a_2$:

$$k = a_1 y_1 + a_2 y_2 + a_1 a_2 - a_1 a_2 = a_1(y_1 - a_2) + a_2(y_2 + a_1).$$

If $y_2 + a_1 > 0$ we are done, otherwise again we add $+a_1 a_2 - a_1 a_2$. Generally we can add $+s a_1 a_2 - s a_1 a2$:

$$k = a_1 y_1 + a_2 y_2 + s a_1 a_2 - s a_1 a_2 = a_1(y_1 - s a_2) + a_2(y_2 + s a_1)$$

as long as $y_1 - s a_2 > 0$, so $s \leq \left\lfloor \frac{y_1}{a_2} \right\rfloor$.
Consider two cases:

- $\left\lfloor \frac{y_1}{a_2} \right\rfloor < \frac{y_1}{a_2}$

  Take $s = \left\lfloor \frac{y_1}{a_2} \right\rfloor$. So we have $y_1 - s a_2 > 0$ and

  $$k = a_1 \left(y_1 - \left\lfloor \frac{y_1}{a_2} \right\rfloor a_2 \right) + a_2 \left(y_2 + \left\lfloor \frac{y_1}{a_2} \right\rfloor a_1 \right).$$

  What is left is to show :

  $$y_2 + \left\lfloor \frac{y_1}{a_2} \right\rfloor a_1 > 0, \qquad / \cdot a_2$$

  $$a_2 y_2 + \left\lfloor \frac{y_1}{a_2} \right\rfloor a_2 a_1 > 0.$$

  Using $\left\lfloor \frac{b}{a} \right\rfloor a \geq (\frac{b}{a} - 1)a = b - a$ we obtain

  $$a_2 y_2 + \left\lfloor \frac{y_1}{a_2} \right\rfloor a_2 a_1 \geq a_2 y_2 + (y_1 - a_2) a_1 = a_2 y_2 + a_1 y_1 - a_1 a_2 = k - a_1 a_2 > 0.$$

- $\left\lfloor \frac{y_1}{a_2} \right\rfloor = \frac{y_1}{a_2}$

  Then take $s = \frac{y_1}{a_2} - 1$. We have $y_1 - s a_2 > 0$ and

  $$k = a_1 \left(y_1 - (\frac{y_1}{a_2} - 1) a_2 \right) + a_2 \left(y_2 + (\frac{y_1}{a_2} - 1) a_1 \right).$$

  What is left is to show:

  $$y_2 + (\frac{y_1}{a_2} - 1) a_1 > 0, \qquad / \cdot a_2$$

  $$a_2 y_2 + (\frac{y_1}{a_2} - 1) a_2 a_1 > 0,$$

  $$a_2 y_2 + (\frac{y_1}{a_2} - 1) a_2 a_1 = a_2 y_2 + a_1 y_1 - a_1 a_2 = k - a_1 a_2 > 0.$$

The proof for $m = 2$ is finished.

For m > 2
Assumption of induction: Each natural number k' greater than $a_1 \cdot \ldots \cdot a_{m-1}$ being multiplicity of $\gcd\{a_1, \ldots, a_{m-1}\}$ (i.e. $k = r \cdot \gcd\{a_1, \ldots, a_{m-1}\}$) can be written as

$$k' = x_1 a_1 + x_2 a_2 + \ldots + x_{m-1} a_{m-1}. \tag{3.10}$$

And we want to show that any number $k > \gcd\{a_1, \ldots, a_m\}$ and $k = s \cdot \gcd\{a_1, \ldots, a_m\}$ can be written as

$$k = x_1 a_1 + x_2 a_2 + \ldots + x_m a_m.$$

Denote $d = \gcd\{a_1, \ldots, a_{m-1}\}$, so that $k = s \cdot \gcd\{d, a_m\}$. Because of Lemma 3.1.3 there exist $x'_m, y' \in \mathbb{N} \setminus \{0\}$ such that

$$a_m x'_m - dy' = k. \tag{3.11}$$

Dividing both sides by da_m we obtain

$$\frac{x'_m}{d} - \frac{y'}{a_m} = \frac{k}{a_m d} > \frac{a_1 a_2 \cdots a_{m-1}}{d} \in \mathbb{Z}.$$

Denote $h := \frac{a_1 a_2 \cdots a_{m-1}}{d}$. From above inequality there exist h consecutive integer numbers $t, t+1, \ldots, t+h-1$ such that

$$\frac{y'}{a_m} < t < t+1 < \ldots < t+h-1 < \frac{x'_m}{d}.$$

Thus we have

$$\begin{aligned} y' &< ta_m < (t+h-1)a_m, \\ (t+h-1)d &< x'_m. \end{aligned}$$

Take:

$$\begin{aligned} x_m &:= x'_m - (t+h-1)d, \\ y &:= (t+h-1)a_n - y'. \end{aligned}$$

We have $x_m > 0, y > 0$ and

$$a_m x_m + dy = a_m \left[x'_m - (t+h-1)d\right] + d\left[(t+h-1)a_m - y'\right] = a_m x'_m - dy' = k.$$

Thus it is now enough to find natural positive numbers $x_1, x_2, \ldots, x_{m-1}$ such that

$$a_1 x_1 + a_2 x_2 + \ldots + a_{m-1} x_{m-1} = dy, \tag{3.12}$$

because combining this with $a_m x_m + dy = k$ would finish the proof:

$$a_m x_m + dy = a_1 x_1 + a_2 x_2 + \ldots + a_{m-1} x_{m-1} + a_m x_m = k.$$

To prove (3.12) we will use induction assumption (3.10). We have $k' = dy = y \cdot \gcd\{a_1, a_2, \ldots, a_{m-1}\}$ so it is a multiplicity of this gcd, what is left to show is that $k' > a_1 a_2 \cdots a_{m-1}$.

$$dy = d\left[(t+h-1)a_m - y'\right] = d(ta_m + ha_m - a_m - y') > d(y' + ha_m - a_m - y') = dha_m - da_m$$

$$= a_1 a_2 \cdots a_m - da_m \geq a_1 a_2 \cdots a_{m-1} = a_m \left(a_1 a_2 \cdots a_{m-1} - d\right) > a_1 a_2 \cdots a_{m-1} - d > a_1 a_2 \cdots a_{m-1},$$

because $a_m \geq 2$.

□

Conclusion 3.1.5. *Let $a_1, \ldots, a_m$ be any natural numbers ≥ 2 such that $a_i \neq a_j$ for $i \neq j$ and $\gcd\{a_1, \ldots, a_m\} = 1$. Then any natural number $k > a_1 \cdots a_m$ can be written in the following representation:*

$$k = a_1 x_1 + + \ldots + a_m x_m,$$

where x_i are integer and **positive.**

3.2 Renewal equation

The following theorem gives the limit of so-called renewal equation.

Theorem 3.2.1 (Feller [19]). *Let a sequence of* $\{f_n\}_{n\geq 0}$ *such that* $f_0 = 0, f_n \geq 0, \sum_{n\geq 0} f_n = 1$ *and* $\gcd\{n : f_n > 0\} = 1$ *be given. Define* **renewal sequence***:*

$$u_0 = 1, \qquad u_n = \sum_{k=1}^{n} f_k u_{n-k} = f_1 u_{n-1} + f_2 u_{n-2} + \ldots + f_n u_0 \tag{3.13}$$

and set $\mu = \sum_{n\geq 1} n f_n$. *Then*

$$u_n \underset{n\to\infty}{\longrightarrow} \frac{1}{\sum_{n\geq 1} n f_n} = \frac{1}{\mu},$$

moreover, if $\sum_{n\geq 1} n f_n = \infty$ *then* $u_n \underset{n\to\infty}{\longrightarrow} 0$.

Define generating function: $F(s) = \sum_{n=0}^{\infty} f_n s^n, U(s) = \sum_{n=0}^{\infty} u_n s^n = 1 + \sum_{n=1}^{\infty} f_n s^n$. *Note that*

$$1 + F(s)U(s) = 1 + \left(\sum_{n=1}^{\infty} f_n s^n\right)\left(\sum_{k=1}^{\infty} u_k s^k\right) = 1 + \sum_{n=1}^{\infty} s^n \left(\sum_{k=1}^{n} f_k u_{n-k}\right) = 1 + \sum_{n=1}^{\infty} u_n s^n = U(s),$$

i.e. $U(s) = \frac{1}{1-F(s)}$. *Then coefficients of the generating function* $U(s)$ *at* s^n *(i.e.* u_n*) converge to* $\frac{1}{\mu}$ *with* $n \to \infty$.

Proof. Define

$$r_n = f_{n+1} + f_{n+2} + \ldots = \sum_{i=n+1}^{\infty} f_i, \tag{3.14}$$

so that:

$$\sum_{n=0}^{\infty} n f_n = f_1 + 2\cdot f_2 + 3\cdot f_3 + \ldots = (f_1 + f_2 + f_3 + \ldots) + (f_2 + f_3 + \ldots) + (f_3 + \ldots) + \ldots$$

i.e.

$$\mu = \sum_{n=0}^{\infty} r_n. \tag{3.15}$$

We have: $f_n = r_{n-1} - r_n$. Putting it to (3.13) we have ($r_0 = 1$):

$$u_n = (r_0 - r_1)u_{n-1} + (r_1 - r_2)u_{n-2} + \ldots + (r_{n-1} - r_n)u_0$$

$$= r_0 u_n + r_1 u_{n-1} + r_2 u_{n-2} + \ldots + r_n u_0 = r_0 u_{n-1} + r_1 u_{n-2} + \ldots + r_{n-1} u_0.$$

If we denote left hand side by A_n then the right hand side is A_{n-1} and we have:

$$A_n = A_{n-1} = A_{n-2} = \ldots = A_0 = r_0 u_0 = 1.$$

For each n we have:

$$r_0 u_n + r_1 u_{n-1} + \ldots + r_n u_0 = 1. \tag{3.16}$$

We also have $u_n \leq 1$ (by induction: $u_0 = 1 \leq 1$, assume that $u_{n-1} \leq 1$ then $u_n = f_1 u_{n-1} + f_2 u_{n-2} + \ldots + f_n u_0 \leq f_1 \cdot 1 + f_2 \cdot 1 + \ldots + f_n \cdot 1 \leq f_1 + \ldots + f_n \leq 1$) and u_n is bounded, so the limit

$$\lambda = \limsup_n u_n$$

exists such that for each $\varepsilon > 0$ and n large enough, all elements are less or equal to $\lambda+\varepsilon$ i.e.: $u_n < \lambda+\varepsilon$ and there exists convergent subsequence $u_{n_v} \to \lambda$, as $n \to \infty$.
Let j be such that $f_j > 0$. We will prove that then $u_{n_v-j} \to \lambda$. Assume contradiction. Then we could find n as big as we want that simultaneously the following conditions would hold:

$$u_n > \lambda - \varepsilon, \quad u_{n-j} < \lambda' < \lambda. \tag{3.17}$$

Let N be such that $r_N < \varepsilon$. $u_k \leq 1$ so for all $n > N$ we have:

$$u_n = f_0 u_n + f_1 u_{n-1} + \ldots + f_N u_{n-N} + f_{N+1} u_{n-N-1} + \ldots + f_n u_0$$

$$\leq f_0 u_n + f_1 u_{n-1} + \ldots + f_N u_{n-N} + f_{N+1} \cdot 1 + f_{N+2} \cdot 1 + \ldots + f_n \cdot 1,$$

i.e.

$$u_n \leq f_0 u_n + f_1 u_{n-1} + \ldots + f_N u_{n-N} + \varepsilon. \tag{3.18}$$

For n big enough each u_{n-k} on the right hand side is smaller than $\lambda + \varepsilon$ and from our assumption $u_{n-j} < \lambda'$, so we have:

$$\begin{gathered} u_n \leq f_0 u_n + f_1 u_{n-1} + \ldots f_{j-1} u_{n-(j-1)} + f_j u_{n-1} + f_{j+1} u_{n-(j+1)} + \\ \ldots + f_N u_{n-N} + \varepsilon < (f_0 + \ldots + f_{j-1} + f_{j+1} + \ldots + f_N)(\lambda+\varepsilon) + f_j \lambda' + \varepsilon \leq \\ (1 - f_j)(\lambda + \varepsilon) + f_j \lambda' + \varepsilon < \lambda + 2\varepsilon - f_j(\lambda - \lambda'). \end{gathered} \tag{3.19}$$

Take ε such that $3\varepsilon < f_j(\lambda - \lambda')$, then $u_n < \lambda + 2\varepsilon - 3\varepsilon = \lambda - \varepsilon$, so $u_n < \lambda - \varepsilon$, what is a contradiction with the assumption that $u_n > \lambda - \varepsilon$ ((3.17)), so the assumption that $\lambda' < \lambda$ was false. So we proved that if $u_{n_v} \to \lambda$ then also $u_{n_v-j} \to \lambda$. Iterating this reasoning: if $f_j > 0$ and $u_{n_v} \to \lambda = \limsup_n u_n$ then

$$u_{n_v-j} \to \lambda, \quad u_{n_v-2j} \to \lambda, \quad u_{n_v-3j} \to \lambda, \quad \ldots$$

Consider two cases::

Case: $f_1 > 0$.
We can take $j = 1$ and then $u_{n_v-k} \to \lambda$ for each fixed k. Put $n := n_v$ to the (3.16) :

$$r_0 u_{n_v} + r_1 u_{n_v-1} + \ldots + r_{n_v} u_0 \leq r_0 u_{n_v} + r_1 u_{n_v-1} + \ldots + r_N u_{n_v-N} \leq 1. \tag{3.20}$$

For fixed N all above $u_{n_v-k} \to \lambda$, so $\lambda(r_0+r_1+\ldots+r_N) \leq 1$. N was chosen arbitrary, so $\lambda(r_0+r_1+\ldots) \leq 1$, i.e. $\lambda \leq \frac{1}{\mu} = \frac{1}{\sum n f_n}$. If $\mu = \sum n f_n = \infty$ the proof is finished, because $u_n \to 0$.
Let $\mu = \sum n f_n < \infty$ and let $\gamma = \liminf u_n$, then for large enough n we have $u_n > \gamma - \varepsilon$. The same reasoning shows that if $u_{n_v} \to \gamma$ then also $u_{n_v-k} \to \gamma$: if opposite was true we would find so large n that simultaneously following two conditions would hold:

$$u_n < \gamma + \varepsilon, \quad u_{n-1} > \gamma' > \gamma. \tag{3.21}$$

Let again N be such that $r_N < \varepsilon$:

$$u_n = f_1 u_{n-1} + \ldots + f_N u_{n-N} + \ldots + f_n u_0 \geq f_1 u_{n-1} + \ldots + f_N u_{n-N} >$$

$$(f_1 + \ldots + f_{j-1} + f_{j+1} + \ldots + f_N)(\gamma - \varepsilon) + f_j \gamma' =$$

$$(f_1 + \ldots + f_j + \ldots + f_N)(\gamma - \varepsilon) - f_j(\gamma - \varepsilon) + f_j \gamma' =$$

$$(1 - r_N)(\gamma - \varepsilon) - f_j(\gamma - \varepsilon) + f_j \gamma' \geq (1 - \varepsilon)(\gamma - \varepsilon) - f_j(\gamma - \varepsilon) + f_j \gamma' \geq$$

$$\gamma + f_j(\gamma' - \gamma) - \varepsilon - \gamma\varepsilon \geq \gamma + f_j(\gamma' - \gamma) - 2\varepsilon.$$

If we take ε such that $f_j(\gamma' - \gamma) > 3\varepsilon$ then we obtain:

$$u_n > \gamma + 3\varepsilon - 2\varepsilon = \gamma + \varepsilon,$$

which contradicts the first inequality in (3.21). Thus $u_{n-j} \to \gamma'$.
Now further: from (3.16):

$$r_0 u_{n_v} + r_1 u_{n_v-1} + \ldots + r_N u_{n_v-N} + \varepsilon \geq 1. \tag{3.22}$$

All above $u_{n_v-k} \to \lambda$ so that $(r_0 + \ldots + r_N)\gamma + \varepsilon \geq 1$, thus $\mu\gamma \geq 1$, but we had $\mu\lambda \leq 1$ and from definition $\gamma \leq \lambda$ so $\gamma = \lambda = \frac{1}{\sum n f_n}$.

Case $f_1 = 0$.
Consider set of numbers j for which $f_j > 0$. From this set we can chose finite subset $a_1, a_2, \ldots a_m$ such that $\gcd\{a_1, \ldots, a_m\} = 1$ and $a_i \geq 2, i = 1, \ldots, m$. We know that if $u_{n_v} \to \lambda$ then also $u_{n_v - x_1 \cdot a_1} \to \lambda, u_{n_v - x_2 \cdot a_2} \to \lambda$ etc. - for each fixed $x_1 > 0, x_2 > 0, \ldots, x_m > 0$ we also have

$$u_{n_v - x_1 \cdot a_1 - x_2 \cdot a_2 - \ldots - x_m \cdot a_m} \to \lambda.$$

In another way: if k is of form: $k = x_1 a_1 + \ldots x_m a_m$ $(x_i > 0)$ then $u_{n_v-k} \to \lambda$. From Conclusion 3.1.5 any number larger than product $a_1 a_2 \ldots a_m$ can be written in this representation. It means that for $k > a_1 a_2 \ldots a_m$ we have $u_{n_v-k} \to \lambda$. Use (3.16) with $n = n_v + a_1 a_2 \ldots a_m$ to obtain (3.20).

□

3.3 Ergodic theorem for enumerable state space

We will utilize Theorem 3.2.1 for ergodicity of Markov chains. Consider a Markov chain $\mathbf{X} = \{X_n, n \geq 0\}$ with enumerable state space $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots\}$. We focus on a fixed state $\mathbf{e}_i \in E$ and will examine asymptotic behaviour of $\mathbf{P}^n(i,i)$ as $n \to \infty$. Denote $u_n^{(i)} = \mathbf{P}^n(i,i)$ for any $n \geq 0, \mathbf{e}_i \in E$ (setting $\mathbf{P}^0(i,i) \equiv 1$), i.e. the probability of being in state $\mathbf{e}_i$ after n steps if we started also in $\mathbf{e}_i$, and for any $n \geq 1$, let $f_n^{(i)}$ denote the probability that Markov chain $\mathbf{X}$ returns from state $\mathbf{e}_i$ to $\mathbf{e}_i$, for the first time exactly after n steps. Let us also denote $d^{(i)} = \gcd\{n : u_n^{(i)} > 0\}$. Of course series $u_n^{(i)}$ is a renewal sequence as defined in (3.13), i.e:

$$u_0^{(i)} = 1, \quad u_n^{(i)} = \sum_{k=1}^{n} f_k^{(i)} u_{n-k}^{(i)}.$$

Definition 3.3.1. *We call state $\mathbf{e}_i$ aperiodic if $d^{(i)} = 1$ and we call Markov chain $\mathbf{X}$* **aperiodic** *if all states are aperiodic.*

Before proceeding with the main theorem, we will need the following

Lemma 3.3.2. $d^{(i)} = \gcd\{n : u_n^{(i)} > 0\} = \gcd\{n : f_n^{(i)} > 0\}$.

Proof. Let s be the smallest integer number such that $f_s^{(i)} > 0$. Then it is also the smallest integer number such that $\mathbf{P}^s(i,i) > 0$ (because it can return for first time after s steps). Denote:
$d_N^{(i)'} = \gcd\{n : s \leq n \leq N : f_n^{(i)} > 0\}$,
$d_N^{(i)''} = \gcd\{n : s \leq n \leq N : \mathbf{P}^n(i,i) > 0\}$.

Note that $d_s^{(i)'} = d_s^{(i)''}$, because $\mathbf{P}^s(i,i) = f_s^{(i)}$. We will show by induction that: $d_N^{(i)'} = d_N^{(i)''}$. Assume that $d_N^{(i)'} = d_N^{(i)''}$. We will show that $d_{N+1}^{(i)'} = d_{N+1}^{(i)''}$:

- $f_{N+1} > 0 \ \to \ \mathbf{P}^{N+1}(i,i) > 0 \ \Rightarrow \ d_{N+1}^{(i)'} = d_{N+1}^{(i)''}$,
- $f_{N+1} = 0 = \mathbf{P}^{N+1}(i,i) \ \Rightarrow \ d_{N+1}^{(i)'} = d_{N+1}^{(i)''}$,
- $f_{N+1} = 0, \ \mathbf{P}^{N+1}(i,i) > 0$, then $\exists_{1 \leq v \leq N} : \ f_v p_{N+1-v} > 0$ (because v can be either 1 or N) $\Rightarrow d_N^{(i)'}$ divides v, $d_N^{(i)''}$ divides $N+1-v$ and both numbers divide $N+1$ (because of $d_N^{(i)'} = d_N^{(i)''}$). Thus $d_{N+1}^{(i)'} = d_N^{(i)'} = d_N^{(i)''} = d_{N+1}^{(i)''}$.

□

Definition 3.3.3. *The state* $\mathbf{e}_i \in E$ *is called* **recurrent** *if and only if*

$$\sum_{n=1}^{\infty} f_n^{(i)} = 1.$$

Markov chain is called recurrent, if all its states are recurrent.

Theorem 3.3.4 (Feller [19])**.** *Let* $\mathbf{e}_i$ *be a recurrent state.*

a) If $d^{(i)} = 1$ *then*

$$\lim_{n\to\infty} \mathbf{P}^n(i,i) = \frac{1}{\sum_{n\geq 1} n f_n^{(i)}} = \frac{1}{\mu^{(i)}}.$$

b) For general:

$$\lim_{n\to\infty} \mathbf{P}^{nd}(i,i) = \frac{d}{\sum_{n\geq 1} n f_n^{(i)}} = \frac{d}{\mu^{(i)}},$$

where $\mu^{(i)} = \sum_{n\geq 1} n f_n^{(i)}$.

Proof.
a) Taking $u_n = u_n^{(i)}$ and $u_n = u_n^{(i)}$ in the Theorem 3.2.1 we have:

$$u_n = \mathbf{P}^n(i,i) \to \frac{1}{\sum n f_n} = \frac{1}{\mu}.$$

b) From Lemma 3.3.2, $d^{(i)} = \gcd\{n : f_n^{(i)} > 0\}$, so $F^{(i)}(s) = \sum f_n^{(i)} s^n$ has non zero's coefficients only at powers $(s^d)^0, (s^d)^1, (s^d)^2, \ldots$. Thus $F^{(i)}(s) = \sum f_n^{(i)} s^n = \sum f_{nd}^{(i)} s^{nd}$ and $F^{(i)}(s^{\frac{1}{d}}) = \sum f_{nd}^{(i)} s^n$. Let $F_1^{(i)}(s) = F^{(i)}(s^{\frac{1}{d}})$. $F_1^{(i)}(s)$ is a power series with positive coefficients and $F_1^{(i)}(1) = 1$, so the assumptions of Theorem 3.2.1 are fulfilled. This implies that coefficients of $U_1^{(i)}(s) = \frac{1}{1-F_1^{(i)}(s)}$ converge to $\frac{1}{\mu_1^{(i)}}$. Keeping in mind that $\mu^{(i)} = \sum_{n\geq 1} n f_n^{(i)} = F^{(i)'}(1)$, we have:

$$\mu_1^{(i)} = F_1^{(i)'}(1) = \frac{1}{d^{(i)}} F^{(i)'}(1) = \frac{\mu^{(i)}}{d^{(i)}}.$$

We have shown that $u_{nd}^{(i)} \to \frac{d^{(i)}}{\mu^{(i)}} = \frac{d^{(i)}}{\sum_{n\geq 1} n f_n^{(i)}}$, because $U^{(i)}(s) = U_1^{(i)}(s^d)$.

□

Definition 3.3.5. *Markov chain is called* **irreducible** *if for any* $\mathbf{e}_i, \mathbf{e}_j \in E$*,* $i \neq j$ *there exists finite* $N_{i,j}$ *such that* $\mathbf{P}^{N_{i,j}}(i,j) > 0$*. In other words: for any* $\mathbf{e}_i, \mathbf{e}_j \in E$*,* $\mathbf{e}_i \neq \mathbf{e}_j$ *there is positive probability that at some time in the future the chain will be in state* j *given it is was* i*.*

Definition 3.3.6. *We call Markov chain* **ergodic** *if it is both: irreducible and aperiodic.*

Lemma 3.3.7. *If Markov chain with transition matrix* $\mathbf{P}$ *is ergodic, then*

$$\forall(\mathbf{e}_k, \mathbf{e}_j \in E) \quad \lim_{n\to\infty} \mathbf{P}^n(k,j) = \lim_{n\to\infty} \mathbf{P}^n(j,j) = \frac{1}{\mu^{(j)}}.$$

Proof. Let us denote $f_n^{(i,j)}$ - the probability of getting from i to j for the first time exactly after n steps. We have $\sum_{k=1}^{\infty} f_k^{(i,j)} = 1$ and

$$\mathbf{P}^n(i,j) = \sum_{k=1}^{n} f_k^{(i,j)} \mathbf{P}^{n-k}(j,j).$$

Taking $\lim_{n\to\infty}$ and using theorem 3.3.4 we have

$$\lim_{n\to\infty} \mathbf{P}^n(i,j) = \sum_{k=1}^{\infty} f_k^{(i,j)} \cdot \lim_{n\to\infty} \mathbf{P}^{n-k}(j,j) = 1 \cdot \lim_{n\to\infty} \mathbf{P}^{n-k}(j,j) = \frac{1}{\mu^{(j)}}.$$

□

3.4 Geometric ergodicity for finite state space

The next theorem provides us with information about speed of convergence. The presented proof follows ideas of Fisz [21], but is modified to stress a role of total variation distance in it.

Theorem 3.4.1 (Fisz [21]). *Let* $\mathbf{P} = [\mathbf{P}(i,j)]$ *be a transition matrix of a Markov chain* $\mathbf{X} = \{X_n, n \geq 0\}$ *with finite state space* $E : |E| = N$.
If there exists r_0 *such that:*

$$\#\{j : \min_{1\leq i\leq N} \mathbf{P}^{r_0}(i,j) = \delta > 0\} = s_{r_0} \geq 1 \tag{3.23}$$

then the following equalities hold:

$$\lim_{n\to\infty} \mathbf{P}^n(i,j) = \pi(j) \quad (j = 1, \ldots, N), \tag{3.24}$$

where $\pi(j) \geq \delta$ *for those* j *for which (3.23) holds. Moreover* $\sum_{j=1}^{N} \pi(j) = 1$ *and*

$$|\mathbf{P}^n(i,j) - \pi(j)| \leq (1 - s_{r_0}\delta)^{n/r_0 - 1}. \tag{3.25}$$

Proof. Denote *min* and *max* in $j-$th column after v steps:

$$b_j^v = \min_{1\leq i\leq N} \mathbf{P}^v(i,j), \quad B_j^v = \max_{1\leq i\leq N} \mathbf{P}^v(i,j).$$

We have:

$$b_j^{v+1} = \min_{1\leq i\leq N} \mathbf{P}^{v+1}(i,j) = \min_{1\leq i\leq N} \sum_{k=1}^{N} \mathbf{P}(i,k)\mathbf{P}^v(k,j) \geq \min_{1\leq i\leq N} \sum_{k=1}^{N} \mathbf{P}(i,k) b_j^v \geq b_j^v,$$

i.e.:

$$b_j^{v+1} \geq b_j^v.$$

Similarly

$$B_j^{v+1} \leq B_j^v,$$

therefore

$$b_j^1 \leq b_j^2 \leq \ldots\ldots \leq B_j^2 \leq B_j^1. \tag{3.26}$$

Let $p_{i\cdot}^{r_0}$ denote the i-th row of matrix $\mathbf{P}^{r_0}$. Note that it is a probability distribution. Denote maximal total variation distance between two rows in matrix $\mathbf{P}$ by

$$d^*(\mathbf{P}) = \max_{1\leq i,m\leq N} d(p_{i\cdot}, p_{m\cdot}).$$

Let $n > r_0$. Consider difference:

$$B_j^n - b_j^n = \max_{1\le i\le N} \mathbf{P}^n(i,j) - \min_{1\le m\le N} \mathbf{P}^n(m,j) = \max_{1\le i\le N} \sum_{k=1}^N \mathbf{P}^{r_0}(i,k)\mathbf{P}^{n-r_0}(k,j) - \min_{1\le m\le N} \sum_{k=1}^N \mathbf{P}^{r_0}(m,k)\mathbf{P}^{n-r_0}(k,j)$$

$$= \max_{1\le i,m\le N} \sum_{k=1}^N \left(\mathbf{P}^{r_0}(i,k) - \mathbf{P}^{r_0}(m,k)\right)\mathbf{P}^{n-r_0}(k,j)$$

$$= \max_{1\le i,m\le N} \Bigg(\sum_{k:\mathbf{P}^{r_0}(i,k)\ge \mathbf{P}^{r_0}(m,k)} (\mathbf{P}^{r_0}(i,k) - \mathbf{P}^{r_0}(m,k))\mathbf{P}^{n-r_0}(k,j) +$$

$$\sum_{k:\mathbf{P}^{r_0}(i,k)> \mathbf{P}^{r_0}(m,k)} (\mathbf{P}^{r_0}(i,k) - \mathbf{P}^{r_0}(m,k))\mathbf{P}^{n-r_0}(k,j)\Bigg)$$

$$\le \max_{1\le i,m\le N} \left(\sum_{k:\mathbf{P}^{r_0}(i,k)\ge \mathbf{P}^{r_0}(m,k)} (\mathbf{P}^{r_0}(i,k) - \mathbf{P}^{r_0}(m,k))B_j^{n-r_0} - \sum_{k:\mathbf{P}^{r_0}(i,k)< \mathbf{P}^{r_0}(m,k)} (\mathbf{P}^{r_0}(m,k) - \mathbf{P}^{r_0}(i,k))b_j^{n-r_0}\right).$$

From (2.5) we have

$$= \max_{1\le i,m\le N} \left(d(p_{i\cdot}^{r_0}, p_{m\cdot}^{r_0})B_j^{n-r_0} - d(p_{i\cdot}^{r_0}, p_{m\cdot}^{r_0})b_j^{n-r_0}\right).$$

$$= (B_j^{n-r_0} - b_j^{n-r_0}) \max_{1\le i,m\le N} d(p_{i\cdot}^{r_0}, p_{m\cdot}^{r_0}) = (B_j^{n-r_0} - b_j^{n-r_0})d^*(\mathbf{P}^{r_0})$$

So for $n > r_0$ we have

$$B_j^n - b_j^n \le (B_j^{n-r_0} - b_j^{n-r_0})d^*(\mathbf{P}^{r_0}). \tag{3.27}$$

We need yet to estimate $d^*(\mathbf{P}^{r_0})$:

$$d^*(\mathbf{P}^{r_0}) = \max_{1\le i,m\le N} d(p_{i\cdot}^{r_0}, p_{m\cdot}^{r_0}) = \max_{1\le i,m\le N} \sum_k (\mathbf{P}^{r_0}(i,k) - \mathbf{P}^{r_0}(m,k))_+ \le \max_{1\le i\le N} \left\{\sum_k \mathbf{P}^{r_0}(i,k)\right\} - s_{r_0}\delta = 1 - s_{r_0}\delta.$$

(estimation in $\le$: replace s_{r_0} expressions $-\mathbf{P}^{r_0}(m,k)$ with δ (minimum of these expressions, the rest set to 0).
Finally:

$$d^*(\mathbf{P}^{r_0}) \le 1 - s_{r_0}\delta. \tag{3.28}$$

From (3.27) and (3.28) we have the inequality:

$$B_j^n - b_j^n \le (1 - s_{r_0}\delta)(B_j^{n-r_0} - b_j^{n-r_0}).$$

Similarly (with $n > 2r_0$)

$$B_j^n - b_j^n \le (1 - s_{r_0}\delta)(B_j^{n-r_0} - b_j^{n-r_0}) \le (1 - s_{r_0}\delta)(1 - s_{r_0}\delta)(B_j^{n-2r_0} - b_j^{n-2r_0})$$

$$= (1 - s_{r_0}\delta)^2(B_j^{n-2r_0} - b_j^{n-2r_0}).$$

Iterating it $[\frac{n}{r_0}]$ times we get:

$$B_j^n - b_j^n \le (1 - s_{r_0}\delta)^{[n/r_0]}(B_j^{n-[\frac{n}{r_0}]r_0} - b_j^{n-[\frac{n}{r_0}]r_0}). \tag{3.29}$$

From the fact that $\mathbf{P}^{r_0}$ is stochastic it follows that $0 < s_{r_0}\delta \le 1$, i.e.

$$0 \le 1 - s_1\delta < 1.$$

From formula (3.26) we conclude existence of limits of series $\{b_j^n\}$ and $\{B_j^n\}$, and from (3.29) we conclude that they have a common limit. So:

$$\lim_{n\to\infty} \max_{1\le i\le N} \mathbf{P}^n(i,j) = \lim_{n\to\infty} \min_{1\le i\le N} \mathbf{P}^n(i,j) = \pi(j), \tag{3.30}$$

and we have proved (3.24).

From (3.26) and (3.29) we have:

$$|p_{ij}^n - \pi(j)| \le B_j^n - b_j^n \le (1 - s_{r_0}\delta)^{n/r_0 - 1}.$$

Using the above inequalities we can bound the total variation distance between $\delta_{\mathbf{e}_i}\mathbf{P}^n$ and π:

$$d(\delta_{\mathbf{e}_i}\mathbf{P}^n, \pi) = \frac{1}{2}\sum_{j=1}^{N} |\mathbf{P}^n(i,j) - \pi(j)| \le \frac{1}{2}\sum_{j=1}^{N}(1 - s_{r_0}\delta)^{n/r_0-1} = \frac{1}{2}N(1 - s_{r_0}\delta)^{n/r_0-1}.$$

□

3.5 Eigenvectors and eigenvalues

Another classical approach to ergodic theorems for Markov chains goes via spectral theory. We recall basic facts.
Consider $n \times n$ matrix $\mathbf{A}$. Recall that eigenvalues are the solutions to the characteristic equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. There are n solutions: $\lambda_1, \lambda_2 \ldots, \lambda_n$. and some of them can be complex not real.
Assume they are labelled in such a way that $|\lambda_1| \ge |\lambda_2| \ge \ldots \ge |\lambda_n|$.

If f_1 is an eigenvector which corresponds to eigenvalue λ_1 then f_1 is linearly independent of all eigenvectors which correspond to eigenvalues different than λ_1. So if all $\lambda_1, \ldots, \lambda_n$ are different, then corresponding eigenvectors $f_1, \ldots, f_n$ create linearly independent set. Then there exists inverse matrix to $\mathbf{N} = [f_1, \ldots, f_n]$ i.e. $\mathbf{N}^{-1}$. For one f_1 we have $\mathbf{A}f_1 = \lambda_1 f_1$ - and for all $f_1, \ldots, f_n$ we can write: $\mathbf{AN} = \mathbf{ND}$ - where $\mathbf{D}$ - diagonal matrix with values $\lambda_1, \ldots, \lambda_n$ on the diagonal. So we have:

$$\mathbf{A} = \mathbf{NDN}^{-1}. \tag{3.31}$$

In the other direction assume there exist: $\mathbf{D}$ - diagonal and invertible matrix $\mathbf{N}$ such that $\mathbf{A} = \mathbf{NDN}^{-1}$. Then all numbers on the diagonal of $\mathbf{D}$ are eigenvalues $\mathbf{A}$, j-th column of $\mathbf{N}$ is a right eigenvector corresponding to λ_j in $\mathbf{D}$, and i-th row in $\mathbf{N}^{-1}$ is the i-th left eigenvector corresponding to λ_i (because $\mathbf{AN} = \mathbf{ND}$)
If a matrix can be written in the form (3.31) it is said to be **diagonalizable**.
So, if all eigenvalues of matrix $\mathbf{A}$ are different, then the matrix is diagonalizable. But the condition that all eigenvalues are distinct is not a necessary for $\mathbf{A}$ to be diagonalizable (all is needed is that there exists m linearly independent eigenvectors for eigenvalues of multiplicity m).
Representation (3.31) is useful while calculating powers of $\mathbf{A}$,
if $\mathbf{A} = \mathbf{NDN}^{-1}$, then $\mathbf{A}^2 = \mathbf{NDN}^{-1}\mathbf{NDN}^{-1} = \mathbf{ND}^2\mathbf{N}^{-1}$ and by induction:

$$\mathbf{A}^k = \mathbf{ND}^k\mathbf{N}^{-1}, \quad k = 0, 1, \ldots \quad . \tag{3.32}$$

In order to calculate eigenvalues it is not necessary to calculate determinants (can be computationally hard). The following equality is useful:

$$tr(\mathbf{A}^k) = \sum_{i=1}^{n} \lambda_i^k, \quad k = 0, 1, \ldots \quad . \tag{3.33}$$

Note that $tr(\mathbf{AB}) = tr(\mathbf{BA})$, indeed

$$tr(\mathbf{AB}) = \sum_i (\mathbf{AB})(i,i) = \sum_i \sum_j \mathbf{A}(i,j)\mathbf{B}(j,i) =$$

$$\sum_j \sum_i \mathbf{B}(j,i)\mathbf{A}(i,j) = \sum_j (\mathbf{BA})(j,j) = tr(\mathbf{BA}).$$

Thus: $tr(\mathbf{A}^k) = tr(\mathbf{N}(\mathbf{D}^k\mathbf{N}^{-1})) = tr(\mathbf{D}^k\mathbf{N}^{-1}\mathbf{N}) = tr(\mathbf{D}^k) = \sum_{i=1}^n \lambda_i^k$.

3.6 Spectral representation

We limit ourselves to diagonalizable matrices. Assume that $\mathbf{A}$ can be written $\mathbf{A} = \mathbf{NDN}^{-1}$, where $\mathbf{D}$ is diagonal. Denote:

$$\mathbf{N} = \begin{bmatrix} f_1(1) & \dots & f_n(1) \\ \vdots & \dots & \vdots \\ f_1(n) & \dots & f_n(n) \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}, \tag{3.34}$$

$$\mathbf{N}^{-1} = \begin{bmatrix} \pi_1(1) & \dots & \pi_1(n) \\ \vdots & \dots & \vdots \\ \pi_n(1) & \dots & \pi_n(n) \end{bmatrix}.$$

$\mathbf{N}^{-1}\mathbf{N} = \mathbf{I}$ so we have

$$\pi_j f_k = \left(\sum_i \pi_j(i) f_k(i) \right) = \begin{cases} 0 & if \;\; j \neq k, \\ 1 & if \;\; j = k. \end{cases} \tag{3.35}$$

Let $\mathbf{B}_k$ - be a matrix received from multiplying vector f_k and π_k, i.e.

$$\mathbf{B}_k = \begin{bmatrix} f_k(1) \\ \vdots \\ f_k(n) \end{bmatrix} [\pi_k(1), \dots, \pi_k(n)] = \begin{bmatrix} f_k(1)\pi_k(1) & \dots & f_k(1)\pi_k(n) \\ \vdots & & \vdots \\ f_k(n)\pi_k(1) & \dots & f_k(n)\pi_k(n) \end{bmatrix}. \tag{3.36}$$

From (3.35) it follows that:

$$\mathbf{B}_j\mathbf{B}_k = f_j\pi_j f_k\pi_k = \begin{cases} 0 & if \;\; j \neq k, \\ \mathbf{B}_j & if \;\; j = k. \end{cases} \tag{3.37}$$

Finally:

$$\mathbf{A} = \mathbf{NDN}^{-1} = \dots = \begin{bmatrix} \sum_{i=1}^n \lambda_i f_i(1)\pi_i(1) & \dots & \sum_{i=1}^n \lambda_i f_i(1)\pi_i(n) \\ \vdots & & \vdots \\ \sum_{i=1}^n \lambda_i f_i(n)\pi_i(1) & \dots & \sum_{i=1}^n \lambda_i f_i(n)\pi_i(n) \end{bmatrix}.$$

In other words:

$$\mathbf{A} = \lambda_1\mathbf{B}_1 + \dots + \lambda_n\mathbf{B}_n. \tag{3.38}$$

The above representation is called **spectral representation**.

From (3.37) we have:

$$\mathbf{A}^k = (\lambda_1\mathbf{B}_1 + \dots + \lambda_n\mathbf{B}_n)^k = \lambda_1^k\mathbf{B}_1 + \dots + \lambda_n^k\mathbf{B}_n. \tag{3.39}$$

Example: Let

$$\mathbf{P} = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$$

$\lambda_1 = 1$ is an eigenvalue corresponding to eigenvector:
$f_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, because $\mathbf{P}1 = 1$.
From (3.33) $tr(\mathbf{P}) = 0.8 + 0.7 = 1.5 = \lambda_1 + \lambda_2 = 1 + \lambda_2$ - thus $\lambda_2 = 0.5$. $\pi_1 = (0.6, 0.4)$ is a left eigenvector, because $\pi_1\mathbf{P} = 1\pi_1$ corresponding to $\lambda_1 = 1$, so $\mathbf{B}_1 = f_1\pi_1$ thus:

$$\mathbf{B}_1 = \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix}.$$

$\mathbf{B}_2$ we will compute using (3.39) for $k = 0$ i.e. $\mathbf{P}^0 = \mathbf{I} = \mathbf{B}_1 + \mathbf{B}_2$, and from here::

$$\mathbf{B}_2 = \begin{bmatrix} 0.4 & -0.4 \\ -0.6 & 0.6 \end{bmatrix}.$$

So the spectral representation for $\mathbf{P}^k$ is following:

$$\mathbf{P}^k = \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix} + \left(\frac{1}{2}\right)^k \begin{bmatrix} 0.4 & -0.4 \\ -0.6 & 0.6 \end{bmatrix}.$$

$\left(\frac{1}{2}\right)^k \to 0$, when $k \to \infty$, thus we have:

$$\mathbf{P}^\infty = \lim_k \mathbf{P}^k = \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix}.$$

Look at the difference:

$$|\mathbf{P}^\infty - \mathbf{P}^k| = \left| \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix} + \left(\frac{1}{2}\right)^k \begin{bmatrix} 0.4 & -0.4 \\ -0.6 & 0.6 \end{bmatrix} - \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix} \right| =$$

$$= \left| \begin{bmatrix} 0.6(\frac{1}{2})^k & -0.4(\frac{1}{2})^k \\ -0.6(\frac{1}{2})^k & 0.4(\frac{1}{2})^k \end{bmatrix} \right| \leq \begin{bmatrix} 0.6(\frac{1}{2})^k & 0.4(\frac{1}{2})^k \\ 0.6(\frac{1}{2})^k & 0.4(\frac{1}{2})^k \end{bmatrix}.$$

Each element of last matrix can be bounded by $(0.6)(\frac{1}{2})^k$ so we have exact speed of convergence:

$$|\mathbf{P}^k(i,j) - \mathbf{P}^\infty(i,j)| \leq (0.6)\left(\frac{1}{2}\right)^k, \quad k = 0, 1, \ldots.$$

3.7 Geometric ergodicity via spectral gap

Definition 3.7.1. *A square matrix* $\mathbf{P}$ *is called* **regular** *if for some integer* k_0 *all entries of* $\mathbf{P}^{k_0}$ *are strictly positive.*

The proof of the following classical theorem can be found in [9]. Recall that $\mathbf{P}(i,j)$ stands for $\mathbf{P}(\mathbf{e}_i, \mathbf{e}_j)$.

Theorem 3.7.2 (Perron - Frobenius)**.** *If* $n \times n$ *matrix* $\mathbf{P}$ *is regular then:*

- $|\lambda_1| > |\lambda_i|, \ i = 2, \ldots, n$
- $\lambda_1 \in \mathbb{R}, \quad \lambda_1 > 0$
- *eigenvector* f_1 *corresponding to* λ_1 *has all components strictly positive (* > 0 *) and is designated uniquely up to multiplying constant*

We will limit ourselves to the case when Markov chain $\mathbf{X}$ is irreducible and aperiodic.

Theorem 3.7.3 (Cinlar [9])**.** *Let* $\mathbf{P}$ *be irreducible and aperiodic transition matrix of Markov chain* $\mathbf{X}$*. Then:*

$$\forall(\mathbf{e}_i, \mathbf{e}_j \in E) \quad \lim_{k\to\infty} \mathbf{P}^k(i,j) = \pi(j) > 0. \tag{3.40}$$

Vector π *is the only solution to the:*

$$\pi\mathbf{P} = \pi, \quad \sum_i \pi(j) = 1. \tag{3.41}$$

Moreover, speed of convergence in (3.40) is geometric.: There exists constant $\alpha > 0$ *such that:*

$$|\mathbf{P}^k(i,j) - \pi(j)| \leq \alpha|\lambda_2|^k, \quad k = 1, 2, \ldots \tag{3.42}$$

Proof. We will only proof the case when $\mathbf{P}$ is diagonalizable. The largest eigenvalue of $\mathbf{P}$ is $\lambda_1 = 1$, its right eigenvector is $f_1 = [1, \ldots, 1]^T$, and its left eigenvector is the one fulfilling (3.41). We have in spectral representation (3.38):

$$\lambda_1 = 1, \quad \mathbf{B}_1 = f_1\pi_1 = \pi_1, \text{ thus } \mathbf{B}_1(i,j) = \pi_1(j). \tag{3.43}$$

So we have:

$$\mathbf{P}^k = \mathbf{B}_1 + \lambda_2^k \mathbf{B}_2 + \ldots + \lambda_n^k \mathbf{B}_n, \quad k = 0, 1, \ldots \tag{3.44}$$

Matrix $\mathbf{P}$ is aperiodic - thus from Perron-Frobienius Theorem ((3.7.2)) eigenvalues $|\lambda_2|, \ldots, |\lambda_n|$ are all strictly less than 1, what finishes the proof of (3.40), because we have:

$$\lim_k \mathbf{P}^k = \mathbf{B}_1.$$

Look at the difference:

$$|\mathbf{P}^k(i,j) - \pi(j)| = |\lambda_2^k \mathbf{B}_2(i,j) + \ldots + \lambda_n^k \mathbf{B}_n(i,j)| \leq$$

$$|\lambda_2|^k |\mathbf{B}_2(i,j)| + \ldots + |\lambda_n|^k |\mathbf{B}_n(i,j)|$$

We had eigenvalues labelled: $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$, thus we have

$$|\mathbf{P}^k(i,j) - \pi(j)| \leq |\lambda_2|^k (|\mathbf{B}_2(i,j)| + \ldots + |\mathbf{B}_n(i,j)|)$$

Finally taking $\alpha = \sup_{i,j}\{|\mathbf{B}_2(i,j)| + \ldots + |\mathbf{B}_n(i,j)|\}$ we have

$$|\mathbf{P}^k(i,j) - \pi(j)| \leq \alpha |\lambda_2|^k.$$

□

As the conclusion from previous Theorem we have:

Lemma 3.7.4. *Let $\mathbf{P}$ be a strictly substochastic (i.e. there is at least one row which sums up to number < 1), irreducible and symmetric. Then*

$$\mathbf{P}^k(i,j) \leq \alpha |\lambda_1|^k,$$

where α is as in (3.42).

Proof. Goes exactly like proof of Theorem 3.7.3 but involves fact that all eigenvalues (including λ_1) are strictly less than 1. □

Remark: $-L = \mathbf{I} - \mathbf{P}$ is called discrete Laplacian of $\mathbf{P}$. If $\mathbf{P}$ has eigenvalues $\lambda_1, \ldots, \lambda_N$ (assume they are ordered as earlier: in absolute values), then $-L$ has eigenvalues $s_1 = (1-\lambda_1), s_2 = (1-\lambda_2), \ldots, s_N = (1-\lambda_N)$. We call s_2 the **spectral gap**. When it is positive, we say that spectral gap exists, what is equivalent to $|\lambda_2| < 1$ and implies, as seen in this section, geometric rate of convergence.

3.8 Bounding second largest eigenvalue: Cheeger's and Poincare' constants for reversible Markov chains

In this section we consider finite state space $|E| = N < \infty$. We start with

Definition 3.8.1. *We say that Markov chain with transition matrix $\mathbf{P}$ and stationary distribution π is* **reversible** *if $\forall \mathbf{e}_i, \mathbf{e}_j \in E$ we have $\pi(i)\mathbf{P}(i,j) = \pi(j)\mathbf{P}(j,i)$.*

In this section we consider only reversible Markov chains (comprehensive source on such chains is Aldous and Fill [4]).

Denote the eigenvalues of $\mathbf{P}$ by λ's ordered: $1 = |\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_N| \geq -1$.

As we have seen in Theorem 3.7.3, having λ_2 we have a bound of convergence rate in L^1 i.e. $|\mathbf{P}^n(i,j) - \pi(j)| \leq \alpha|\lambda_2|^n, n = 1, 2, \ldots$. In that Theorem α is hard to calculate. The exact bound on the total variation distance is given in the following corollary from Theorem 3.9.2 which will be given later.

Corrolary 3.8.2. *Let* $\mathbf{P}$ *be a transition matrix of a reversible Markov chain with finite state space. Then*

$$d(\delta_{\mathbf{e}}\mathbf{P}^n, \pi) \le \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e})}}|\lambda_2|^n.$$

Unfortunately, in most cases calculating λ_2 is rather difficult. But it is enough to have upper bounds on this eigenvalue.

Often, eigenvalues of Laplacian $\mathbf{I} - \mathbf{P}$ are considered. We have the following minimax characterization of the second largest eigenvalue which is defined for enumerable state space (see e.g. Horn & Johnson [25]):

$$|\beta_2| = 1 - |\lambda_2| = \inf\left\{\frac{\mathcal{E}(\phi,\phi)}{Var(\phi)} : \phi \text{ is non-constant}\right\}, \tag{3.45}$$

where for $\Lambda(i,j) = \pi(i)\mathbf{P}(i,j)$ and $\phi(\mathbf{e}_j) \equiv \phi(j)$

$$\begin{aligned}
\mathcal{E}(\phi,\phi) &= \tfrac{1}{2}\sum_{\mathbf{e}_i,\mathbf{e}_j\in E}(\phi(j)-\phi(i))^2\pi(i)\mathbf{P}(i,j) = \tfrac{1}{2}\sum_{\mathbf{e}_i,\mathbf{e}_j\in E}(\phi(j)-\phi(i))^2\Lambda(i,j),\\
Var(\phi,\phi) &= \tfrac{1}{2}\sum_{\mathbf{e}_i,\mathbf{e}_j\in E}(\phi(j)-\phi(i))^2\pi(i)\pi(j).
\end{aligned}$$

In case of finite state space λ_2 coincides with $1-\beta_2$ and we have $\beta_i := 1-\lambda_i$ for $i = 1,\ldots,N$.
Define **Cheeger's constant** by

$$k = \min_{A:\pi(A)\le 1/2}\frac{\Lambda(A\times A^C)}{\pi(A)} = \min_{S:\pi(A)\le 1/2}\frac{\sum_{\mathbf{e}_i\in A}\sum_{\mathbf{e}_j\in A^C}\pi(i)\mathbf{P}(i,j)}{\pi(A)}.$$

Often (as it will be done in section 6.2 (Markov Jump Process)) there is used another version of Cheeger's constant:

$$k' = \min_{A:\pi(A)\in(0,1)}\frac{\Lambda(A\times A^C)}{\pi(A^c)\pi(A)} = \min_{A:\pi(S)\in(0,1)}\frac{\sum_{\mathbf{e}_i\in A}\sum_{\mathbf{e}_j\in A^C}\pi(i)\mathbf{P}(i,j)}{\pi(A^c)\pi(A)}.$$

We have then:

$$\frac{k'}{2} \le k \le k'. \tag{3.46}$$

Next theorem gives bounds (lower and upper) on $|\lambda_2|$ using Cheeger's constant k. Since this result is useful to show the existence of spectral gap in a class of Markov processes we give it with a proof.

Theorem 3.8.3 (Diaconis & Stroock [16])**.** *Let* $\lambda_2 : |\lambda_2| = 1 - |\beta_2|$ *be the second largest eigenvalue of a reversible, ergodic Markov chain* $\mathbf{X}$*. Then*

$$1 - 2k \le |\lambda_2| \le 1 - \frac{1}{2}k^2.$$

The above inequality is often called **Cheeger's inequality**.

Proof.
First, we will show: $1 - 2k \le |\lambda_2|$.
To prove it take for a function ϕ in (3.45), $\phi_A(i) \equiv \phi_A(\mathbf{e}_i) := I_A(\mathbf{e}_i)$ (indicator function) with a fixed A such that $\pi(A) \le 1/2$. Then we have

$$\mathcal{E}(\phi_A,\phi_A) = \frac{1}{2}\sum_{\mathbf{e}_i\in E}\sum_{\mathbf{e}_j\in E}(\phi_A(j)-\phi_A(i))^2\Lambda(i,j)$$

$$= \frac{1}{2}\Big(\sum_{\mathbf{e}_i \in A} \sum_{\mathbf{e}_j \in E} (1-1)^2 \Lambda(i,j) + \sum_{\mathbf{e}_i \in A} \sum_{\mathbf{e}_j \in A^C} (1-0)^2 \Lambda(i,j) + \sum_{\mathbf{e}_i \in A^C} \sum_{\mathbf{e}_j \in A} (0-1)^2 \Lambda(i,j) +$$

$$\sum_{\mathbf{e}_i \in A^C} \sum_{\mathbf{e}_j \in A^C} (0-0)^2 \Lambda(i,j) \Big)$$

(and because $\Lambda(i,j) = \Lambda(j,i)$)

$$= \sum_{\mathbf{e}_i \in A} \sum_{\mathbf{e}_j \in A^C} \Lambda(i,j) = \Lambda(A \times A^C).$$

And similarly

$$Var(\phi_A) = \sum_{\mathbf{e}_i \in A} \sum_{\mathbf{e}_j \in A^C} \pi(i)\pi(j) = \pi(A)\pi(A^C),$$

and using $\pi(A)\pi(A^C) \geq \frac{1}{2}\pi(A)$ we obtain

$$1 - |\lambda_2| = |\beta_2| \leq \frac{\mathcal{E}(\phi_A, \phi_A)}{Var(\phi_A)} = \frac{\Lambda(A \times A^C)}{\pi(A)\pi(A^C)} \leq 2\frac{\Lambda(A \times A^C)}{\pi(A)}.$$

Thus

$$|\lambda_2| \geq 1 - 2\frac{\Lambda(A \times A^C)}{\pi(A)}.$$

Above is true for any A so especially:

$$|\lambda_2| \geq 1 - 2k.$$

Now we prove the second part:

$$|\lambda_2| \leq 1 - \frac{1}{2}k^2.$$

Recall we have $L\psi(i) = \sum_j L(i,j)\psi(j)$, $-L = \mathbf{I} - \mathbf{P}$.
Define inner product

$$(\phi, \psi)_\pi = \sum_{\mathbf{e} \in E} \phi(\mathbf{e})\psi(\mathbf{e})\pi(\mathbf{e}).$$

We have:

$$L\psi(i) = \sum_{\mathbf{e}_i} L(i,j)\psi(j) = \sum_{\mathbf{e}_j} (\mathbf{I} - \mathbf{P})\psi(j) = \sum_{\mathbf{e}_j} \mathbf{I}(i,j)\psi(j) - \sum_{\mathbf{e}_j} \mathbf{P}(i,j)\psi(j)$$

$$= \psi(i) - \sum_{\mathbf{e}_j} P(i,j)\psi(j)$$

and

$$
\begin{aligned}
(\phi, L\psi)_\pi &= \sum_{\mathbf{e}_i} \pi(i)\phi(i) \sum_{\mathbf{e}_j} L(i,j)\psi(j) \\
&= \sum_{\mathbf{e}_i} \pi(i)\phi(i)[\psi(i) - \sum_{\mathbf{e}_j} \mathbf{P}(i,j)\psi(j)] \\
&= \sum_{\mathbf{e}_i} \pi(i)\phi(i)\psi(i) - \sum_{\mathbf{e}_i,\mathbf{e}_j} \pi(i)\mathbf{P}(i,j)\phi(i)\psi(j) \\
&= \sum_{\mathbf{e}_i} \pi(i)\phi(i)\psi(i) - \sum_{\mathbf{e}_i,\mathbf{e}_j} \Lambda(i,j)\phi(i)\psi(j) \\
(\textstyle\sum_{\mathbf{e}_j} \mathbf{P}(i,j) = 1) \quad &= \sum_{\mathbf{e}_i} \pi(i)\phi(i)\psi(i) \sum_{\mathbf{e}_j} \mathbf{P}(i,j) - \sum_{\mathbf{e}_i,j} \Lambda(i,j)\phi(i)\psi(j) \\
&= \sum_{\mathbf{e}_i,\mathbf{e}_j} \phi(i)\psi(i)\pi(i)\mathbf{P}(i,j) - \sum_{\mathbf{e}_i,\mathbf{e}_j} \Lambda(i,j)\phi(i)\psi(j) \\
&= \sum_{\mathbf{e}_i,\mathbf{e}_j} \Big(\phi(i)\psi(i) - \phi(i)\psi(j)\Big)\Lambda(i,j) \\
(revers.) \quad &= \tfrac{1}{2} \sum_{\mathbf{e}_i,\mathbf{e}_j} \Big(\phi(i)\psi(i) - \phi(i)\psi(j) - \phi(j)\psi(i) + \phi(j)\psi(j)\Big)\Lambda(i,j) \\
&= \tfrac{1}{2} \sum_{\mathbf{e}_i,\mathbf{e}_j} (\phi(j) - \phi(i))(\psi(j) - \psi(i))\Lambda(i,j) = \mathcal{E}(\phi, \psi).
\end{aligned}
\tag{3.47}
$$

Fix $\psi \in L^2(\pi)$, denote by ψ_+ a positive part of ψ, set $S(\psi) = \{\mathbf{e}_i \in E : \psi(i) > 0\}$.
FIRST OBSERVATION: For any $\psi \in L^2(\pi)$ such that $S(\psi) \neq \emptyset$ and $\beta \in [0, \infty)$ we have

$$\beta||\psi_+||^2_\pi \geq \mathcal{E}(\psi_+, \psi_+) \quad \text{ if } L\psi \leq \beta\psi \text{ on } S(\psi). \tag{3.48}$$

For $\beta\psi(i) \geq L\psi(i)$ for any $\mathbf{e}_i \in S(\psi)$ we have

$$\beta||\psi_+||^2_\pi = \beta(\psi_+, \psi_+)_\pi = \beta \sum_{\mathbf{e}_i \in E} \psi_+(i)^2\pi(i) = \beta \sum_{\mathbf{e}_i \in S} \psi(i)^2\pi(i) + \beta \sum_{\mathbf{e}_i \in S^C} 0^2\pi(i)$$

$$= \sum_{\mathbf{e}_i \in S} \beta\pi(i)\psi(i)\psi(i) \geq \sum_{\mathbf{e}_i \in S} \pi(i)L\psi(i)\psi(i) = \sum_{\mathbf{e}_i \in S} \pi(i)L\psi(i)\psi(i) + \sum_{\mathbf{e})i \in S^C} \pi(i)L\psi(i) \cdot 0$$

$$= \sum_{\mathbf{e}_i \in E} \pi(i)L\psi(i)\psi_+(i) = \sum_{\mathbf{e}_i \in S} \pi(i)L\psi(i)\psi_+(i) = (\psi_+, L\psi)_\pi$$

and using (3.47) we get

$$\beta||\psi_+||^2\pi \geq (\psi_+, L\psi)_\pi = \mathcal{E}(\psi_+, \psi).$$

Note that

$$
\begin{aligned}
\mathcal{E}(\psi_+, \psi) &= \frac{1}{2} \sum_{\mathbf{e}_i,\mathbf{e}_j} (\psi_+(j) - \psi_+(i))(\psi(j) - \psi(i))\Lambda(i,j) \\
&\geq \frac{1}{2} \sum_{\mathbf{e}_i,\mathbf{e}_j} (\psi_+(j) - \psi_+(i))^2\Lambda(i,j) = \mathcal{E}(\psi_+, \psi_+),
\end{aligned}
$$

where we used

$$(\psi_+(j) - \psi_+(i))(\psi(j) - \psi(i)) \geq (\psi_+(j) - \psi_+(i))^2.$$

Thus we proved $\beta||\psi_+||^2_\pi \geq \mathcal{E}(\psi_+, \psi_+)$ i.e. (3.48).
SECOND OBSERVATION: For any $\psi \in L^2(\pi)$ with $S(\psi) \neq \emptyset$ we have

$$\mathcal{E}(\psi_+, \psi_+) \geq \frac{k(\psi)^2||\psi_+||^2_\pi}{2}, \quad \text{where} \quad k(\psi) \equiv \inf\left\{\frac{\Lambda(S \times S^C)}{\pi(S)} : \emptyset \neq S \subseteq S(\psi)\right\}. \tag{3.49}$$

We will use Cauchy-Schwartz inequality:

$$\left(\sum_{\mathbf{e}_1,\mathbf{e}_2} f(\mathbf{e}_1, \mathbf{e}_2)g(\mathbf{e}_1, \mathbf{e}_2)\right)^2 \leq \sum_{\mathbf{e}_1,\mathbf{e}_2} f^2(\mathbf{e}_1, \mathbf{e}_2) \cdot \sum_{\mathbf{e}_1,\mathbf{e}_2} g^2(\mathbf{e}_1, \mathbf{e}_2).$$

Taking: $f(\mathbf{e}_i,\mathbf{e}_j) = |\psi(i)-\psi(j)|\sqrt{\Lambda(i,j)}$ and $g(i,j) = |\psi(i)+\psi(j)|\sqrt{\Lambda(i,j)}$ we obtain

$$\left(\sum_{\mathbf{e}_i,\mathbf{e}_j} f(i,j)g(i,j)\right)^2 = \left(\sum_{\mathbf{e}_i,\mathbf{e}_j} |\psi^2(i)-\psi^2(j)|\Lambda(i,j)\right)^2$$

$$\leq 2\cdot\frac{1}{2}\sum_{\mathbf{e}_i,\mathbf{e}_j}(\psi(i)-\psi(j))^2\Lambda(i,j)\cdot\sum_{\mathbf{e}_i,\mathbf{e}_j}(\psi(i)+\psi(j))^2\Lambda(i,j)$$

$$= 2\mathcal{E}(\psi,\psi)\sum_{\mathbf{e}_i,\mathbf{e}_j}(\psi(i)+\psi(j))^2\Lambda(i,j).$$

Thus

$$\sum_{\mathbf{e}_i,\mathbf{e}_j}|\psi^2(i)-\psi^2(j)|\Lambda(i,j) \leq \sqrt{2}\sqrt{\mathcal{E}(\psi,\psi)}\left(\sum_{\mathbf{e}_i,\mathbf{e}_j}(\psi(i)+\psi(j))^2\Lambda(i,j)\right)^{1/2}$$

$$\leq \sqrt{2}\sqrt{\mathcal{E}(\psi,\psi)}\left(2\cdot\sum_{\mathbf{e}_i,\mathbf{e}_j}(\psi^2(i)+\psi^2(j))\Lambda(i,j)\right)^{1/2} = \sqrt{2}\sqrt{\mathcal{E}(\psi,\psi)}\left(2\cdot 2\cdot\sum_{\mathbf{e}_i,\mathbf{e}_j}\psi^2(i)\pi(i)\mathbf{P}(i,j)\right)^{1/2}$$

$$= 2^{3/2}\sqrt{\mathcal{E}(\psi,\psi)}\left(\sum_{\mathbf{e}_i}\psi^2(i)\pi(i)\underbrace{\sum_{\mathbf{e}_j}\mathbf{P}(i,j)}_{=1}\right)^{1/2} = 2^{3/2}\sqrt{\mathcal{E}(\psi,\psi)}\ ||\psi||_\pi.$$

At the same time left hand side can be rewritten as

$$\sum_{\mathbf{e}_i,\mathbf{e}_j}|\psi^2(i)-\psi^2(j)|\Lambda(i,j) = 2\cdot\sum_{\substack{\mathbf{e}_i,\mathbf{e}_j:\\ \psi(j)>\psi(i)}}(\psi^2(j)-\psi^2(i))\Lambda(i,j)$$

$$= 4\cdot\sum_{\substack{\mathbf{e}_i,\mathbf{e}_j:\\ \psi(j)>\psi(i)}}\left(\int_{\psi(i)}^{\psi(j)} tdt\right)\Lambda(i,j) = \int_0^\infty t\left(\sum_{\substack{\mathbf{e}_i,\mathbf{e}_j:\\ \psi(i)\leq t<\psi(j)}}\Lambda(i,j)\right)dt.$$

And using

$$\sum_{\substack{\mathbf{e}_i,\mathbf{e}_j:\\ \psi(i)\leq t<\psi(j)}}\Lambda(i,j) = \Lambda(S\times S^C) \quad \text{with} \quad S\equiv\{i:\psi(i)>t\}\subseteq S(\psi),$$

we have

$$\int_0^\infty t\left(\sum_{\substack{\mathbf{e}_i,\mathbf{e}_j:\\ \psi(i)\leq t<\psi(j)}}\Lambda(i,j)\right)dt = \int_0^\infty t\Lambda(S\times S^C)dt = \int_0^\infty t\pi(S)\frac{\Lambda(S\times S^C)}{\pi(S)}$$

$$\geq k(\psi)\int_0^\infty t\pi(\{i:\psi(i)>t\})dt = \frac{k(\psi)||\psi||_\pi^2}{2}.$$

Combining (3.48) and (3.49) we obtain:

$$\beta||\psi_+||_\pi^2 \geq \mathcal{E}(\psi_+,\psi_+) \geq \frac{k(\psi)^2||\psi_+||_\pi^2}{2} \quad \text{if } L\psi\leq\beta\psi \text{ on } S(\psi).$$

Thus

$$\frac{k(\psi)^2}{2}\leq\beta \tag{3.50}$$

for any $\beta\in[0,\infty)$ and any $\psi\in L^2(\pi)$ on $S(\psi)\neq\emptyset$. To get lower bound take $\beta=|\beta_2|=1-|\lambda_2|$ and ψ as normalized eigenfunction for β_2. Because ψ must have π-mean-value 0, we can always arrange that

$0 < \pi(S(\psi)) \leq \frac{1}{2}$ and therefore $k(\psi) \geq k$. Hence, the desired lower bound follows directly from (3.50) with this choice of β and ψ.

□

Define a graph $G = (V, E)$ with vertex set E and edges $V = \{\{\mathbf{e}_i, \mathbf{e}_j\} : \Lambda(i,j) > 0\}$. For given vertices $\mathbf{e}_i$ and $\mathbf{e}_j$ we deterministically choose path (without loops and repeated edges) and refer to it as the *canonical path* $\Gamma(\mathbf{e}_i, \mathbf{e}_j)$ from $\mathbf{e}_i$ to $\mathbf{e}_j$. Define *length* of the path

$$|\Gamma(\mathbf{e}_i, \mathbf{e}_j)| := \sum_{\tilde{e} \in \Gamma(\mathbf{e}_i, \mathbf{e}_j)} 1,$$

where the sum is over oriented edges $\tilde{e}$ in the path. Define $\Lambda(\tilde{e}) = \Lambda(i,j) = \pi(i)\mathbf{P}(i,j)$ if $\tilde{e} = (\mathbf{e}_i, \mathbf{e}_j)$. Define **Poincare's constant**

$$K := \max_{\tilde{e}} \left\{ \frac{1}{\Lambda(\tilde{e})} \left[\sum_{(\mathbf{e}_i, \mathbf{e}_j) : \tilde{e} \in \Gamma(\mathbf{e}_i, \mathbf{e}_j)} |\Gamma(\mathbf{e}_i, \mathbf{e}_j)| \pi(i)\pi(j) \right] \right\}$$

if $\mathbf{P}$ is irreducible and $K := \infty$ otherwise. So for fixed edge $\tilde{e}$ we have to sum up over all possible paths from $\mathbf{e}_1$ to $\mathbf{e}_2$ containing $\tilde{e}$ and calculate the above constant.

Next theorem gives an upper bound for β_1 using K.

Theorem 3.8.4 (Diaconis & Stroock [16]). *The second largest eigenvalue β_2 of a finite state, reversible Markov chain $\mathbf{P}$ with respect to everywhere positive stationary distribution π satisfies*

$$|\lambda_2| \leq 1 - \frac{1}{K}.$$

The above inequality is called **Poincare's inequality**.

Proof. For an edge $\tilde{e} = (\mathbf{e}_i, \mathbf{e}_j)$ denote $\tilde{e}^+ \equiv \mathbf{e}_i$, $\tilde{e}^- \equiv \mathbf{e}_j$ and $\Lambda(\tilde{e}) = \pi(i)\mathbf{P}(i,j)$. Define $\phi(\tilde{e}) = \phi(\tilde{e}^+) - \phi(\tilde{e}^-)$.

$$Var(\phi) = \frac{1}{2} \sum_{\mathbf{e}_i, \mathbf{e}_j \in E} (\phi(i) - \phi(j))^2 \pi(i)\pi(j) = \frac{1}{2} \sum_{\mathbf{e}_i, \mathbf{e}_j \in E} \left(\sum_{\tilde{e} \in \Gamma(\mathbf{e}_i, \mathbf{e}_j)} \phi(\tilde{e}) \right)^2 \pi(i)\pi(j) =$$

$$= \frac{1}{2} \sum_{\mathbf{e}_i, \mathbf{e}_j \in E} \left(\sum_{\tilde{e} \in \Gamma(\mathbf{e}_i, \mathbf{e}_j)} \frac{1}{\sqrt{\Lambda(\tilde{e})}} \sqrt{\Lambda(\tilde{e})} \phi(\tilde{e}) \right)^2 \pi(i)\pi(j)$$

$$\leq \frac{1}{2} \sum_{\mathbf{e}_i, \mathbf{e}_j \in E} \left(\sum_{e \in \Gamma(\mathbf{e}_i, \mathbf{e}_j)} \frac{1}{\Lambda(\tilde{e})} \right) \left(\sum_{\tilde{e} \in \Gamma(\mathbf{e}_i, \mathbf{e}_j)} \Lambda(\tilde{e}) \phi^2(\tilde{e}) \right) \pi(i)\pi(j) \leq$$

$$\max_{\tilde{e}} \left\{ \frac{1}{\Lambda(\tilde{e})} \sum_{\mathbf{e}_i, \mathbf{e}_j : \tilde{e} \in \Gamma(\mathbf{e}_i, \mathbf{e}_j)} |\Gamma(\mathbf{e}_i, \mathbf{e}_j)| \pi(i)\pi(j)| \right\} \cdot \frac{1}{2} \sum_{\tilde{e}} \Lambda(\tilde{e}) \phi^2(\tilde{e}) = K \cdot \frac{1}{2} \sum_{\tilde{e}} \Lambda(\tilde{e}) \phi^2(\tilde{e})$$

$$= K \cdot \frac{1}{2} \sum_{\mathbf{e}_i, \mathbf{e}_j \in E} (\phi(i) - \phi(j))^2 \pi(i)\mathbf{P}(i,j) = K \cdot \mathcal{E}(\phi, \phi)$$

i.e.

$$\frac{1}{K} \leq \frac{\mathcal{E}(\phi, \phi)}{Var(\phi)}$$

for any ϕ so especially for ϕ which realizes infimum in (3.45), thus

$$\frac{1}{K} \leq 1 - |\lambda_2| \quad \Rightarrow \quad |\lambda_2| \leq 1 - \frac{1}{K}.$$

□

Next lemma gives us information how many steps should be done in order to make $d(\delta_{\mathbf{e}}\mathbf{P}^k,\pi)$ less or equal to ε.

Lemma 3.8.5. *For a finite state, reversible, aperiodic and irreducible, Markov chain on finite state space E with transition matrix $\mathbf{P}$ with finite state space we have*

$$d(\delta_{\mathbf{e}_0}\mathbf{P}^{n_k(\varepsilon)},\pi) \le \varepsilon, \qquad \textit{for } n_k(\varepsilon) = \frac{2}{k^2}\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)$$

and

$$d(\delta_{\mathbf{e}_0}\mathbf{P}^{n_K(\varepsilon)},\pi) \le \varepsilon, \qquad \textit{for } n_K(\varepsilon) = K\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right),$$

where h is the Cheeger's constant and K is the Poincare's constant.

Proof. :

In $(*)$ we will use fact that $1-x<e^{-x}$ for $x\in\mathbb{R}\setminus\{0\}$.
Using Corrolary 3.8.2 and Theorem 3.8.3 we have

$$d(\delta_{\mathbf{e}_0}\mathbf{P}^{n_k(\varepsilon)},\pi) \le \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}|\lambda_2|^{n_k(\varepsilon)} \le \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}\left(1-\frac{1}{2}k^2\right)^{n_k(\varepsilon)} = \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}\left(1-\frac{1}{2}k^2\right)^{\frac{2}{k^2}\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)}$$

$$\overset{(*)}{\le} \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}\left[e^{-\frac{1}{2}k^2}\right]^{\frac{2}{k^2}\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)} = \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}e^{-\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)} = \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}2\varepsilon\sqrt{\pi(\mathbf{e}_0)} = \varepsilon.$$

Using Corrolary 3.8.2 and Theorem 3.8.4 we have

$$d(\delta_{\mathbf{e}_0}\mathbf{P}^{n_K(\varepsilon)},\pi) \le \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}|\lambda_2|^{n_K(\varepsilon)} \le \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}\left(1-\frac{1}{K}\right)^{n_K(\varepsilon)} = \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}\left(1-\frac{1}{K}\right)^{K\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)}$$

$$\overset{(*)}{\le} \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}\left[e^{-\frac{1}{K}}\right]^{K\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)} = \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}e^{-\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)} = \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}2\varepsilon\sqrt{\pi(\mathbf{e}_0)} = \varepsilon.$$

□

Remark: A number $n(\varepsilon)$ for which $d(\delta_{\mathbf{e}_0}\mathbf{P}^{n(\varepsilon)},\pi)\le\varepsilon$ is called **mixing time** and usually denoted by $\tau_{\mathbf{e}_0}(\varepsilon)$.

3.8.1 Poincare bound for symmetric random walk on d-dimensional cube

In order to illustrate Poincare's technique we present this example.
Take the state space $E=\{0,1\}^d$. Let $x\in E$. Transition probabilities of symmetric random walk on cube are defined as follows:

$$\begin{aligned} \delta_x\mathbf{P}(x+s_i) &= \tfrac{1}{d+1} \text{ for } x_i=0,\\ \delta_x\mathbf{P}(x-s_i) &= \tfrac{1}{d+1} \text{ for } x_i=1,\\ \delta_x\mathbf{P}(x) &= \tfrac{1}{d+1}, \end{aligned} \tag{3.51}$$

where $s_i=(0,\ldots,0,1,0,\ldots,0)$ with 1 at position i. It means that when random walk is in the state $x=(x_1,\ldots,x_d), x_i\in\{0,1\}$ then it changes i-th coordinate from x_i to $1-x_i$ with probability $1/(d+1)$ or with the same probability it stays at the same position. The stationary distribution is uniform : $\pi(x)=\frac{1}{2^d}$.
We will compute Poincare's constant for deterministic canonical paths.

Let $x=(x_1,\ldots,x_d), y=(y_1,\ldots,y_n)\in E$ and define canonical path $\Gamma(x,y)$ in such a way: change consecutively coordinates in x to get y from left to right. For example for $d=8$ and $x=(1,1,0,1,1,1,0,1), y=(1,0,0,0,0,0,0,0)$ the canonical path is:

$$\begin{array}{rcllll} x & = & (1,1,0,1,1,1,0,1) & \to & (1,0,0,1,1,1,0,1) & \to & (1,0,0,0,1,1,0,1) & \to & (1,0,0,0,0,1,0,1) \\ & \to & (1,0,0,0,0,0,0,1) & \to & (1,0,0,0,0,0,0,0) & = y \end{array}$$

We want to calculate

$$K:=\max_{\tilde e}\left\{\frac{1}{\Lambda(\tilde e)}\left[\sum_{(x,y):e\in\Gamma(x,y)}|\Gamma(x,y)|\pi(x)\pi(y)\right]\right\}.$$

Fix $\tilde e=(a,b)$, where $a=(a_1,\ldots,a_d)$ and $b=(u_1,\ldots,u_{i-1},1-u_i,u_{i+1},\ldots u_d)$. We have: $\Lambda(\tilde e)=\pi(a)\mathbf{P}(a,b)=\frac{1}{2^d}\cdot\frac{1}{d+1}$, $\pi(x)=\pi(y)=\frac{1}{2^d}$.
Thus

$$K:=\max_{\tilde e}\left\{\frac{1}{\Lambda(\tilde e)}E\left[\sum_{(x,y):\tilde e\in\Gamma(x,y)}|\Gamma(x,y)|\pi(x)\pi(y)\right]\right\}=2^d(d+1)\cdot\frac{1}{2^d}\cdot\frac{1}{2^d}\sum_{(x,y):\tilde e\in\Gamma(x,y)}|\Gamma(x,y)|.$$

We have to calculate how many states x and y there are such that $\tilde e\in\Gamma(x,y)$. Look at previous example once more (say $a=(1,0,0,0,1,1,0,1), b'=(1,0,0,0,0,1,0,1)$, thus $i=5$)

$$\begin{array}{rcllll} x & = & (1,1,0,1,\boxed{1},\underbrace{1,0,1}_{x_{i+1},\ldots,x_d}) & \to & (1,0,0,1,\boxed{1},1,0,1) & \to & \underbrace{(1,0,0,0,\boxed{1},1,0,1)}_{a} & \to & \underbrace{(1,0,0,0,\boxed{0},1,0,1)}_{b} \\ & \to & (1,0,0,0,\boxed{0},0,0,1) & \to & (\underbrace{1,0,0,0}_{y_1,\ldots,y_{i-1}},\boxed{0},0,0,0) & = y \end{array}$$

Thus we see that x and y must fulfill the following conditions:

$$\begin{array}{rcl} x_{i+1},\ldots,x_d & = & u_{i+1},\ldots,u_d, \\ x_i & = & u_i, \\ y_i & = & 1-u_i, \\ y_1,\ldots,y_{i-1} & = & u_1,\ldots,u_{i-1}. \end{array}$$

So only $x_1,\ldots,x_{i-1}$ and $y_{i+1},\ldots,y_d$ are not determined. Thus the number of possible x and y such that $\tilde e=(u,v)\in\Gamma(x,y)$ is $2^{i-1}\cdot 2^{d-i}=2^{d-1}$. The length $|\Gamma(x,y)|$ is at most d, so we have:

$$K:=2^d(d+1)\cdot\frac{1}{2^d}\cdot\frac{1}{2^d}\sum_{(x,y):\tilde e\in\Gamma(x,y)}|\Gamma(x,y)|\le\frac{(d+1)}{2^d}\sum_{(x,y):\tilde e\in\Gamma(x,y)}d=\frac{d(d+1)}{2^d}\cdot 2^{d-1}=\frac{d(d+1)}{2}.$$

Let us calculate $n_K(\varepsilon)$ from Lemma 3.8.5 for some fixed state $x_0\in E$:

$$n_K(\varepsilon):=K\log\left(\frac{1}{2\varepsilon\sqrt{\pi(x_0)}}\right)\le\frac{d(d+1)}{2}\log\left(\frac{1}{2\varepsilon\sqrt{\frac{1}{2^d}}}\right)=\frac{d(d+1)}{2}\log\left(2^{\frac{d}{2}}\cdot\frac{1}{2\varepsilon}\right)=$$

$$\frac{d(d+1)}{2}\left(\frac{d}{2}+\log\left(\frac{1}{2\varepsilon}\right)\right)\approx\frac{d^3}{4}+\frac{d^2}{2}\log\left(\frac{1}{2\varepsilon}\right).$$

This tells us that mixing takes about $d^3/4$ steps (i.e. after so many steps we have $d(\delta_{x_0}\mathbf{P}^{n_K},\pi)\le\varepsilon$). It is of order $\approx d^3$.

Later we will see better bounds (via coupling or strong uniform times) and correct answer: mixing takes in this example about $d\log d$ steps.

For another example where both, Poincare's and Cheeger's constants are calculated see subsection 3.9.2, where problem of non-reversible Markov chain is reduced to calculating these constants for modified reversible chain.

3.9 Bounding second largest eigenvalue: non-reversible Markov chains

3.9.1 Symmetrization

The results of subsection 3.8 had not been applicable to non-reversible Markov chains till 1991 when J. Fill developed bounds for these chains in [20]. Some of results are interesting in their own.

For any transition matrix $\mathbf{R}$ define:

$$\mathbf{R}\phi(\mathbf{e}_i) := \sum_{\mathbf{e}_j \in E} \mathbf{R}(\mathbf{e}_i, \mathbf{e}_j)\phi(\mathbf{e}_j).$$

Given functions ϕ and ψ define the Dirichlet form

$$\mathcal{E}_{\mathbf{R}}(\phi, \psi) := \frac{1}{2} \sum_{\mathbf{e}_i, \mathbf{e}_j \in E} (\phi(\mathbf{e}_i) - \phi(\mathbf{e}_j))(\psi(\mathbf{e}_i) - \psi(\mathbf{e}_j))\pi(\mathbf{e}_i)\mathbf{R}(\mathbf{e}_i, \mathbf{e}_j)$$

based on a given reversible $\mathbf{R}$ with stationary distribution π.
For $\mathbf{P}$ define $\widetilde{\mathbf{P}}$ to be time-reversal of $\mathbf{P}$, i.e.:

$$\widetilde{\mathbf{P}}(\mathbf{e}_i, \mathbf{e}_j) := \frac{\pi(\mathbf{e}_j)\mathbf{P}(\mathbf{e}_j, \mathbf{e}_i)}{\pi(\mathbf{e}_i)},$$

and define *multiplicative reversiblization* $M(\mathbf{P})$:

$$M(\mathbf{P}) := \mathbf{P}\widetilde{\mathbf{P}}.$$

It has the same stationary distribution as $\mathbf{P}$ and is time-reversible, i.e. $\pi(\mathbf{e}_i)M(\mathbf{P})(\mathbf{e}_i, \mathbf{e}_j) = \pi(\mathbf{e}_j)M(\mathbf{P})(\mathbf{e}_j, \mathbf{e}_i)$.

Theorem 3.9.1 (Mihail's identity)**.** *Let $\mathbf{P}$ be an ergodic transition matrix on a finite state space E with stationary distribution π. Then for any function ϕ we have*

$$Var(\phi) = Var(\widetilde{\mathbf{P}}\phi) + \mathcal{E}_{M(\mathbf{P})}(\phi, \phi). \tag{3.52}$$

Proof. Without loss of generality we can assume that ϕ has 0 mean under π, i.e. $(\phi, \mathbf{1})_\pi = 0$. Note that for reversible $\mathbf{R}$ we have:

$$(\phi, (\mathbf{I} - \mathbf{R})\phi)_\pi = \sum_{\mathbf{e}} \phi(\mathbf{e})(\mathbf{I} - \mathbf{R})\phi(\mathbf{e})\pi(\mathbf{e}) = \sum_{\mathbf{e}} \phi(\mathbf{e})\pi(\mathbf{e}) - \sum_{\mathbf{e}_1, \mathbf{e}_2} \pi(\mathbf{e}_1)\mathbf{P}(\mathbf{e}_i, \mathbf{e}_j)\phi(\mathbf{e}_i)\phi(\mathbf{e}_j) =$$

$$\frac{1}{2} \sum_{\mathbf{e}_i, \mathbf{e}_j} (\phi(\mathbf{e}_i) - \phi(\mathbf{e}_j))^2 \pi(\mathbf{e}_i)\mathbf{P}(\mathbf{e}_i, \mathbf{e}_j) = \mathcal{E}_{\mathbf{R}}(\phi, \phi),$$

thus

$$\mathcal{E}_{M(\mathbf{P})}(\phi, \phi) = (\phi, (\mathbf{I} - M(\mathbf{P}))\phi)_\pi = Var(\phi) - (\phi, \mathbf{P}\widetilde{\mathbf{P}}\phi) = Var(\phi) - (\widetilde{\mathbf{P}}\phi, \widetilde{\mathbf{P}}\phi)_\pi = Var(\phi) - Var(\widetilde{\mathbf{P}}\phi),$$

where we have used fact that $\widetilde{\mathbf{P}}$ is the adjoint of $\mathbf{P}$ on $L^2(\pi)$.

□

Define **chi-square** distance from stationarity for a non-reversible Markov chain:

$$\chi_n^2 := \sum_{\mathbf{e}} \frac{(\mu\mathbf{P}^n(\mathbf{e}) - \pi(\mathbf{e}))^2}{\pi(\mathbf{e})}.$$

Theorem 3.9.2 (Fill [20])**.** *Let $\mathbf{P}$ be an ergodic transition matrix on a finite state space E with stationary distribution π. Let $\lambda_2(M)$ denotes the second largest eigenvalue of $M(\mathbf{P})$. Then*

$$d(\mu\mathbf{P}^n, \pi) \le \frac{1}{2}\chi_n \le \frac{1}{2}|\lambda_2(M)|^{n/2}\chi_0.$$

In particular, for $\mu := \delta_{\mathbf{e}}$ we have

$$d(\delta_{\mathbf{e}}\mathbf{P}^n, \pi) \le \frac{1}{2}|\lambda_2(M)|^{n/2}\sqrt{\frac{1}{\pi(\mathbf{e})} - 1} \le \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e})}}|\lambda_2(M)|^{n/2}.$$

Proof. Take function $\rho_n(\mathbf{e}) = \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}$. Its mean value is $\sum_{\mathbf{e}} \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}\pi(\mathbf{e}) = 1$. Calculate its variance:

$$Var(\rho_n) = \sum_{\mathbf{e}} \left(\frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})} - 1\right)^2 \pi(\mathbf{e}) = \sum_{\mathbf{e}} \left(\frac{\mu\mathbf{P}^n(\mathbf{e}) - \pi(\mathbf{e})}{\pi(\mathbf{e})}\right)^2 \pi(\mathbf{e}) = \sum_{\mathbf{e}} \frac{(\mu\mathbf{P}^n(\mathbf{e}) - \pi(\mathbf{e}))^2}{\pi(\mathbf{e})} = \chi_n^2.$$

Calculate also

$$(\widetilde{\mathbf{P}}\rho_n)(\mathbf{e}) = \sum_{\mathbf{e}_2} \widetilde{\mathbf{P}}(\mathbf{e}, \mathbf{e}_2)\frac{\mu\mathbf{P}^n(\mathbf{e}_2)}{\pi(\mathbf{e}_2)} = \sum_{\mathbf{e}_2} \frac{\mu\mathbf{P}^n(\mathbf{e}_2)\mathbf{P}(\mathbf{e}_2, \mathbf{e})}{\pi(\mathbf{e})} = \rho_{n+1}(\mathbf{e}).$$

Thus from Mihail's identity (3.52) we have

$$\chi_n^2 = \chi_{n+1}^2 + \mathcal{E}_{M(\mathbf{P})}(\rho_n, \rho_n).$$

By the minimax characterization (3.45) for $\beta_1(M)$ we have that

$$\mathcal{E}_{M(\mathbf{P})}(\rho_n, \rho_n) \geq (1 - |\lambda_2(M)|)Var(\rho_n) = (1 - |\lambda_2(M)|)\chi_n^2.$$

Thus

$$\chi_n^2 = \chi_{n+1}^2 + \mathcal{E}_{M(\mathbf{P})}(\rho_n, \rho_n) \geq \chi_n^2 + (1 - |\lambda_2(M)|)\,\chi_n^2.$$

From above we obtain

$$\chi_{n+1}^2 \leq |\lambda_2(M)|\chi_n^2.$$

Iterating it:

$$\chi_n^2 \leq |\lambda_2(M)|^n \chi_0^2.$$

Now from Cauchy-Schwarz inequality:

$$d(\mu\mathbf{P}^n, \pi) = \frac{1}{2}\left(\sum_{\mathbf{e}} |\mu\mathbf{P}^n(\mathbf{e}) - \pi(\mathbf{e})|\right) = \frac{1}{2}\left(\left(\sum_{\mathbf{e}} \sqrt{\pi(\mathbf{e})} \cdot \frac{|\mu\mathbf{P}^n(\mathbf{e}) - \pi(\mathbf{e})|}{\sqrt{\pi(\mathbf{e})}}\right)^2\right)^{1/2} \leq$$

$$= \frac{1}{2}\left(\left(\sum_{\mathbf{e}} \pi(\mathbf{e})\right) \cdot \left(\sum_{\mathbf{e}} \frac{|\mu\mathbf{P}^n(\mathbf{e}) - \pi(\mathbf{e})|^2}{\pi(\mathbf{e})}\right)\right)^{1/2} = \frac{1}{2}\left(\chi_n^2\right)^{1/2} = \frac{1}{2}\chi_n \leq \frac{1}{2}(\beta_1(M))^{n/2}\chi_0.$$

For $\mu := \delta_{\mathbf{e}_0}$ we have

$$\chi_0^2 = \sum_{\mathbf{e}} \frac{|\delta_{\mathbf{e}_0}\mathbf{P}^0(\mathbf{e}) - \pi(\mathbf{e})|^2}{\pi(\mathbf{e})} = \frac{|1 - \pi(\mathbf{e}_0)|^2}{\pi(\mathbf{e}_0)} + \sum_{\mathbf{e} \neq \mathbf{e}_0} \frac{|0 - \pi(\mathbf{e})|^2}{\pi(\mathbf{e})} = \frac{1}{\pi(\mathbf{e}_0)} - 2 + \pi(\mathbf{e}_0) + 1 - \pi(\mathbf{e}_0) = \frac{1}{\pi(\mathbf{e}_0)} - 1$$

□

Remark: When $\mathbf{P}$ is time-reversible then $\mathbf{P} = \widetilde{\mathbf{P}}$ and $|\lambda_2(M)| = |\lambda_2|^2$ and we have $d(\mu\mathbf{P}^n, \pi) \leq \frac{1}{2}|\lambda_2|^n\chi_0$.

We have the extension of Lemma 3.8.5 for non-reversible Markov chains:

Lemma 3.9.3. *For a non-reversible, aperiodic and irreducible, Markov chain on finite state space E with transition matrix $\mathbf{P}$ of size $n \times n$ we have*

$$d(\delta_{\mathbf{e}_0}\mathbf{P}^{n'_k(\varepsilon)}, \pi) \leq \varepsilon, \qquad \text{for } n'_k(\varepsilon) = \frac{4}{k^2}\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)$$

and

$$d(\delta_{\mathbf{e}_0}\mathbf{P}^{n'_K(\varepsilon)}, \pi) \leq \varepsilon, \qquad \text{for } n'_K(\varepsilon) = 2K\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right),$$

where k is a Cheeger's constant and K is Poincare's constant of $M(\mathbf{P})$.

Proof. In $(*)$ we will use fact that $1 - x < e^{-x}$ for $x \in \mathbb{R} \setminus \{0\}$.
Using Theorems 3.8.3 and 3.9.2 we have

$$d(\delta_{\mathbf{e}_0}\mathbf{P}^{n'_k(\varepsilon)}, \pi) \leq \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}|\lambda_2(M)|^{n'_k(\varepsilon)/2} \leq \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}\left(1 - \frac{1}{2}k^2\right)^{n'_k(\varepsilon)/2} = \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}\left(1 - \frac{1}{2}k^2\right)^{\frac{2}{k^2}\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)}$$

$$\stackrel{(*)}{\leq} \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}\left[e^{-\frac{1}{2}k^2}\right]^{\frac{2}{k^2}\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)} = \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}e^{-\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)} = \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}2\varepsilon\sqrt{\pi(\mathbf{e}_0)} = \varepsilon$$

Using Theorems 3.8.4 and 3.9.2 we have

$$d(\delta_{\mathbf{e}_0}\mathbf{P}^{n'_K(\varepsilon)}, \pi) \leq \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}|\lambda_2(M)|^{n'_K(\varepsilon)/2} \leq \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}\left(1 - \frac{1}{K}\right)^{n'_K(\varepsilon)/2} = \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}\left(1 - \frac{1}{K}\right)^{K\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)}$$

$$\stackrel{(*)}{\leq} \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}\left[e^{-\frac{1}{K}}\right]^{K\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)} = \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}e^{-\log\left(\frac{1}{2\varepsilon\sqrt{\pi(\mathbf{e}_0)}}\right)} = \frac{1}{2}\frac{1}{\sqrt{\pi(\mathbf{e}_0)}}2\varepsilon\sqrt{\pi(\mathbf{e}_0)} = \varepsilon.$$

□

3.9.2 Example: Clockwise random walk on $\mathbb{Z}_n$

To illustrate above techniques for non-reversible Markov chain $\mathbf{X} = \{X_n, n \geq 0\}$ we give an example, originally described by Fill [20]. We will give full details here, calculating both, Poincare's and Cheeger's constants.
For $p \in (0,1)$ we have matrix $p_{j,j} = 1-p$ and $p_{j,j+1} = p$, where $j+1$ is modulo n. Thus we have random walk on circle $\{0, 1, \ldots, n-1\}$ with uniform stationary distribution, i.e. $\pi(j) = \frac{1}{n}, j = 1, \ldots, n$, which is non-reversible.

$$\mathbf{P} = \begin{pmatrix} 1-p & p & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1-p & p & 0 & 0 & \ldots & 0 & 0 \\ 0 & & 1-p & p & 0 & \ldots & 0 & 0 \\ \vdots & & & & \vdots & & & \\ p & & 0 & 0 & 0 & \ldots & 0 & 1-p \end{pmatrix}, \ \widetilde{\mathbf{P}} = \begin{pmatrix} 1-p & 0 & 0 & 0 & 0 & \ldots & 0 & p \\ p & 1-p & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & p & 1-p & 0 & 0 & \ldots & 0 & 0 \\ \vdots & & & & \vdots & & & \\ 0 & & 0 & 0 & 0 & \ldots & p & 1-p \end{pmatrix},$$

$$M(\mathbf{P}) = \begin{pmatrix} p^2+q^2 & pq & 0 & 0 & 0 & \ldots & 0 & pq \\ pq & p^2+q^2 & pq & 0 & 0 & \ldots & 0 & 0 \\ 0 & pq & p^2+q^2 & pq & 0 & \ldots & 0 & 0 \\ \vdots & & & & \vdots & & & \\ pq & & 0 & 0 & 0 & \ldots & pq & p^2+q^2 \end{pmatrix}.$$

CHEEGER
Calculating Cheeger's constant for $M(\mathbf{P})$ we obtain:

$$k = \min_{S:\pi(S)\leq 1/2} \frac{\sum_{x\in S}\sum_{y\in S^C}\pi(x)M(\mathbf{P})(x,y)}{\pi(S)} = \min_{S:\pi(S)\leq 1/2} \frac{\frac{1}{n}\sum_{x\in S}\sum_{y\in S^C}M(\mathbf{P})(x,y)}{\frac{|S|}{n}}$$

$$= \min_{S:\pi(S)\leq 1/2} \frac{\sum_{x\in S}\sum_{y\in S^C}pq}{|S|} = \frac{2pq}{\frac{n}{2}} = \frac{4pq}{n},$$

because min is achieved for for $S = \{a, a+1, \ldots, a+\frac{n}{2}\}$.

POINCARE

Let us calculate Poincare constant for our $M(\mathbf{P})$. Because of symmetry, for each $\tilde{e}$ value of $\Lambda(\tilde{e})$ is the same. We can choose $\tilde{e} = (0,1)$ and $\Lambda(\tilde{e}) = \pi(\tilde{e})M(\mathbf{P})(0,1) = \frac{1}{n}pq$, and of course $\pi(x) = \pi(y) = \frac{1}{n}$ for any x, y. Thus

$$K := \max_e \left\{ \frac{1}{\Lambda(e)} \left[\sum_{(x,y):e\in\Gamma(x,y)} |\Gamma(x,y)|\pi(x)\pi(y) \right] \right\} = \frac{1}{\frac{1}{n}pq}\frac{1}{n}\frac{1}{n} \sum_{(x,y):e\in\Gamma(x,y)} |\Gamma(x,y)|$$

We will choose paths deterministically, taking the shortest one from x to y. We have to calculate how many possible path with e on it there is, and calculate their lengths. Let $[a,b]$ denote path $\{a, a+1, a+2, \ldots, b\}$. For definiteness let n be **odd** (otherwise for $x = 0$ and $y = n/2$ we would not have the shortest path).

Of course there is one path of length 1 containing $\tilde{e}$, simply $[0,1]$. There are two paths of length 2 containing $\tilde{e}$: $[0,2]$ and $[n-1,1]$. Generally: there are i paths of length i, namely: $[n-i+1,1], [n-i+2,2], \ldots, [0,i]$ and the longest is of length $\frac{n-1}{2}$. Thus, using $\sum_{i=1}^m = \frac{m(m+1)(2m+1)}{6}$ we have

$$= \frac{1}{pq}\frac{1}{n} \sum_{(x,y):\tilde{e}\in\Gamma(x,y)} |\Gamma(x,y)| = \frac{1}{pq}\frac{1}{n}\sum_{i=1}^{\frac{n-1}{2}} i^2 = \frac{1}{pq}\frac{1}{n}\frac{\left(\frac{n-1}{2}\right)\left(\frac{n-1}{2}+1\right)\left(2\cdot\frac{n-1}{2}+1\right)}{6} = \frac{1}{pq}\frac{1}{n}\frac{\left(\frac{n-1}{2}\right)\left(\frac{n+1}{2}\right)n}{6}$$

$$= \frac{1}{24pq}(n^2-1).$$

Comparing cheeger and poincare

To compare it take n to be odd. Then Cheeger's constant is slightly different then calculated earlier, namely $k = \frac{4pq}{n-1}$ (because now $|S| = \frac{n-1}{2}$). And we can compare (using Theorems 3.8.3 and 3.8.4))

$$\text{Poincare} \quad |\lambda_2(M(\mathbf{P}))| \leq 1 - \frac{1}{K} = 1 - \frac{24pq}{n^2-1},$$

$$\text{Cheeger} \quad |\lambda_2(M(\mathbf{P}))| \leq 1 - \frac{1}{2}k^2 = 1 - \frac{8(pq)^2}{(n-1)^2}.$$

Which is better? It depends on p and n. Cheeger's is better when

$$1 - \frac{8(pq)^2}{(n-1)^2} \leq 1 - \frac{24pq}{n^2-1} \iff pq \leq \frac{1}{3(n+1)}$$

To see if there is significant difference in order of steps calculate $n'_k(\varepsilon)$ and $n'_K(\varepsilon)$:

$$n'_k(\varepsilon) = \frac{4}{h^2}\log\left(\frac{1}{2\varepsilon\sqrt{\pi(x)}}\right) = \frac{n^2}{(2pq)^2}\log\left(\frac{\sqrt{n}}{2\varepsilon}\right) = \frac{1}{(2pq)^2}n^2\left(\log\left(\sqrt{n}\right) + \log\left(\frac{1}{2\varepsilon}\right)\right),$$

$$n'_K(\varepsilon) = 2K\log\left(\frac{1}{2\varepsilon\sqrt{\pi(x)}}\right) = \frac{n^2-1}{12pq}\log\left(\frac{\sqrt{n}}{2\varepsilon}\right) = \frac{n^2-1}{12pq}\left(\log\left(\sqrt{n}\right) + \log\left(\frac{1}{2\varepsilon}\right)\right).$$

We see that in both cases number of steps of order $\approx n^2\log(\sqrt{n})$ is enough to make total variation distance small. But in this case we have:

$$n'_K \leq n'_k \iff \frac{3}{pq} \geq 1 - \frac{1}{n^2},$$

and $\frac{3}{pq} \geq 12$ for $p \in [0,1]$,thus estimation by Poincare's constant is better, regardless of value of n and p.

Remark: Note that in both cases minimum (as one could expected) is achieved for $p = \frac{1}{2}$, then $n'_k(\varepsilon) = 4n^2\left(\log\left(\sqrt{n}\right) + \log\left(\frac{1}{2\varepsilon}\right)\right)$ and (much better) $n'_K(\varepsilon) = \frac{n^2-1}{3}\left(\log\left(\sqrt{n}\right) + \log\left(\frac{1}{2\varepsilon}\right)\right)$.

Some examples where exact rates of convergence for non-reversible Markov chains are calculted can be found in Wilmer [48].

3.10 Geometric ergodicity via coupling for enumerable state space

Theorem 3.10.1 (Liggett [32])**.** *Let $\mathbf{P} = [\mathbf{P}(i,j)]$ be a transition matrix of Markov chain $\mathbf{X} = \{X_n, n \geq 0\}$ with enumerable state space E, which initial distribution is μ. If matrix $\mathbf{P}$ is regular (see Def. 3.7.1) then limit*

$$P_\mu(X_k = \mathbf{e}_j) = \sum_{\mathbf{e}_i \in E} \mu(\mathbf{e}_i)\mathbf{P}^k(i,j) \to \pi(j) \tag{3.53}$$

exists for any $\mathbf{e}_j \in E$. The limit is independent from the initial distribution μ and the following equality is fulfilled:

$$\sum_{\mathbf{e}_i \in E} \pi(i)\mathbf{P}(i,j) = \pi(j). \tag{3.54}$$

Moreover

$$d(\mu\mathbf{P}^n, \pi) \leq (1-\delta)^{n/r_0 - 1}, \tag{3.55}$$

where k_0 is from Def. 3.7.1 and $\delta = \min_{\mathbf{e}_i, \mathbf{e}_j} \mathbf{P}^{k_0}(i,j) > 0$.

Proof. The idea of coupling is to take independent chain $\mathbf{Y} = \{Y_n, n \geq 0\}$ with the same transition matrix as $\mathbf{X}$ (i.e. $\mathbf{P}$) but with the initial distribution π (stationary for $\mathbf{P}$) and observe process $\mathbf{X}$ till first time they meet, later on taking $\mathbf{Y}$ process.
Let $\mathbf{Y} = \{Y_n, n \geq 0\}$ be a Markov chain independent from $\mathbf{X}$ with transition matrix $\mathbf{P}$ and initial distribution π (this is so called stationary version of $\{X_n, n \geq 0\}$ process).
Define:

$$Z_n = \begin{cases} X_n & \text{if} \quad n < T, \\ Y_n & \text{if} \quad n \geq T, \end{cases} \tag{3.56}$$

where

$$T = \min_k \{k : X_k = Y_k\}$$

is so called **coupling time**. We state that:

i) $\{X_n, n \geq 0\}$ and $\{Z_n, n \geq 0\}$ have the same distribution,

ii) The following inequality holds:

$$\forall (n > 0, B \subset E) \qquad |P_\mu(Z_n \in B) - P_\pi(Y_n \in B)| \leq P_\mu(T > n),$$

iii) $P_\mu(T < \infty) = 1$.

Proving i), ii), iii) finishes the proof of (3.53) because

$$|P_\mu(X_n = \mathbf{e}_j) - \pi(\mathbf{e}_j)| \overset{\text{i)}}{=} |P_\mu(Z_n = \mathbf{e}_j) - P_\pi(Y_n = \mathbf{e}_j)| \overset{\text{ii)}}{\leq} P_\mu(T > n) \xrightarrow{\text{iii)}} 0.$$

Proof: i)
For $k > n$ we have

$$P(Z_0 = x_0, \ldots, Z_n = x_n, T = k) = P(X_0 = x_0, \ldots, X_n = x_n, T = k),$$

and for $k \leq n$ (using the independence of X_n and Y_n and Markov's property)

$$\begin{aligned} P(Z_0 = x_0, \ldots, Z_n = x_n, T = k) &= P(X_0 = x_0, \ldots, X_k = x_k, Y_k = x_k, \ldots, Y_n = x_n, T = k) \\ &= P(X_0 = x_0, \ldots, X_k = x_k, T = k) p_{x_k, x_{k+1}} \cdots p_{x_{n-1}, x_n} \\ &= P(X_0 = x_0, \ldots, X_n = x_n, T = k). \end{aligned}$$

Proof: ii)

$$|P_\mu(Z_n \in B) - P_\pi(Y_n \in B)| =$$

$$= |P_\mu(Z_n \in B, T \leq n) + P_\mu(Z_n \in B, T > n) - P_\pi(Y_n \in B, T \leq n) - P_\pi(Y_n \in B, T > n)| =$$

$$\overset{(*)}{=} |P_\mu(Z_n \in B, T > n) - P_\pi(Y_n \in B, T > n)| \overset{(**)}{\leq} P_\mu(T > n).$$

In $(*)$ first and third expressions where equal in absolute values, whereas in $(**)$ we were subtracting two non-negative numbers, both less than $P_\mu(T > n)$.

Proof: iii)

$$P_\mu(T \leq k_0) \geq \sum_{k \in E} \mu \mathbf{P}^{k_0}(k) \pi \mathbf{P}^{k_0}(k) = \sum_{\mathbf{e}_k \in E} \left(\sum_{\mathbf{e}_i \in E} \mu(\mathbf{e}_i) \mathbf{P}^{k_0}(i,k) \right) \pi(\mathbf{e}_k) \geq \sum_{\mathbf{e}_k \in E} \left(\sum_{\mathbf{e}_i \in E} \mu(\mathbf{e}_i) \delta \right) \pi(\mathbf{e}_k) = \delta \cdot 1 \cdot 1 = \delta.$$

So we have

$$P_\mu(T > k_0) \leq (1 - \delta) \quad \text{and} \quad P_\mu(T > m \cdot k_0) \leq (1 - \delta)^m.$$

For any n we have

$$P_\mu(T > n) = P_\mu(T > \frac{n}{k_0} \cdot k_0) \leq P_\mu(T > \left[\frac{n}{k_0}\right] \cdot k_0) \leq (1 - \delta)^{[\frac{n}{k_0}]}.$$

And because our assumption was $\delta > 0$ so we have

$$\lim_{n \to \infty} P_\mu(T > n) = 0, \quad \text{i.e.} \quad P_\mu(T < \infty) = 1.$$

Assertion (3.55) we obtain from i), ii) and iii)

$$d(\mu \mathbf{P}^n, \pi) = \sup_{B \subset E} |P_\mu(X_n \in B) - \pi(B)| \leq P_\mu(T > n) \leq (1 - \delta)^{[\frac{n}{k_0}]}.$$

□

Remark: For coupling defined in (3.56) statements i) and ii) do always hold. The problem is thus to determine whether $P_\mu(T < \infty) = 1$ or not. In the above theorem assumption $\min_{\mathbf{e}_i, \mathbf{e}_j \in E} P^{k_0}(i,j) = \delta > 0$ assured it.

One of the inequalities obtained while proving last theorem is quite important, so we state it separately.

Lemma 3.10.2. *For coupling defined in (3.56) so called* **coupling inequality** *holds:*

$$d(\mu \mathbf{P}^n, \pi) \leq P_\mu(T > n).$$

Proof. See proof of Theorem 3.10.1. □

3.10.1 Symmetric random walk on d-dimensional cube

We revisit the cube example in order to compare Poincare and coupling technique.
Consider again random walk with state space $E = \{0,1\}^d$ and transition probabilities given in (3.51). The random walk is in state $x = (x_1, \ldots, x_d), x_i \in \{0,1\}$ then it changes i-th coordinate from x_i to $1 - x_i$ with probability $1/(d+1)$ or with the same probability it stays at the same position. The stationary distribution is uniform : $\pi(x) = \frac{1}{2^d}$.
Define coupling like in (3.56) taking independent processes: $\{X_n, n \geq 0\}$, with initial distribution μ, and $\{Y_n, n \geq 0\}$ with initial distribution π being stationary for this chain. Now let these chains evolve in the following way: we choose random variables $\{U_n\}$ which are independent and uniformly distributed on $\{1, 2, \ldots, d, d+1, \ldots 2d\}$ and set:

$$X_n(k) = Y_n(k) = \begin{cases} 1 & \text{if} \quad U_n = 2k - 1, \\ 0 & \text{if} \quad U_n = 2k. \end{cases}$$

It means, that if $U_n = 2k-1$ then we set kth coordinate of both X_n and Y_n to 1; if $U_n = 2k$ then we set kth coordinate of both processes to 0 leaving other coordinates unchanged. Of course these processes are independent and both governed by the (3.51). Now let $T_\mu = \min\{n : X_n = Y_n\}$. An upper bound for this coupling time is the time until all coordinates will be chosen at least once. It is equivalent to the following scheme. We have $2d$ urns; what is the probability that after n steps there exists at least one urn not chosen so far? It is exactly well known Coupons Collector Problem. Define V to be the number of drawings till the time when each ball is drawn at least once. For each ball i define

$$A_i = \begin{cases} 1 & \text{ball} \quad i \quad \text{was not drawn in the first} \quad m \quad \text{experiments,} \\ & \\ 0 & \text{otherwise.} \end{cases} \tag{3.57}$$

Set $n = d \log d + cd$.

$$P(V > n) = P(\bigcup_{i=1}^{d} A_i) \le \sum_{i=1}^{d} P(A_i) = d\left(1 - \frac{1}{d}\right)^m \le de^{-\frac{n}{d}} = e^{-c}.$$

Thus, using Lemma 3.10.2 (coupling inequality) we have

$$d(\mu \mathbf{P}^n, \pi) \le P(T_\mu > n) \le P(V > n) \le e^{-c}, \quad \text{for } n = d \log d + cd.$$

Remark: In subsection 3.8.1 we have shown (calculating Poincare's constant). that after d^3 steps process is well mixed, whereas via coupling we obtained much better result, that it takes about $d \log d$ steps. However, in practice, finding good coupling is harder then calculating Poincare's (or Cheeger's) constant.

3.11 Maintaining Communication Between an Explorer and a Base Station

Eigenvalue aproach presented in subsecion 3.5 was used in [17] to show the speed of convergence of some Markov chain in the field of robotic intelligence. We present here the main part of the paper.

We consider a (robotic) explorer starting an exploration of an unknown terrain from its base station. As the explorer has only limited communication radius, it is necessary to maintain a line of robotic relay stations following the explorer, so that consecutive stations are within the communication radius of each other.

We construct a graph modeling the base station, the explorer and the relay stations with vertices. The vertices are always logically organized in a path $(v_1, v_2, \ldots, v_{n-1}, v_n)$, where v_1 corresponds to the base station, v_n to the explorer and $v_2, \ldots, v_{n-1}$ to the relay stations. To represent the path we introduce undirected edges (v_i, v_{i+1}) for every $i \in \{1, \ldots, n-1\}$. The communication is routed along this path from v_1 to v_n or in the other direction. The graph is embedded on a plane, thus we will use the notion of a position $p(v)$ of a vertex v. Distances between vertices are given by the L_2 norm and described by $|(v_i, v_{i+1})|$.

The goal of a strategy minimizing the distance between the relay stations is to arrange the relay stations on the line between v_1 and v_n in equal distances from each other, or, in other words, to bring the relay stations as near to this optimal positions as possible.

We require every edge on the path $v_1, \ldots, v_n$ to have at most length d, so that the maximum transmission distance of d is not exceeded and communication links between partners on the communication path can be hold up. We assume that the terrain is without obstacles.

The following GO-TO-THE-MIDDLE strategy is executed repeatedly by every relay station. Relay station i observes the positions $p(v_{i-1})$ and $p(v_{i+1})$ of its communication partners and moves itself into the middle of the interval from v_{i-1} to v_{i+1}.

For simplification of the analysis we will assume that the strategy is invoked in discrete time steps. Each time step is subdivided into two shorter substeps. In the first one, all relay stations check the positions of their neighbors. In the second substep all relay stations move to the middle of the observed positions of its neighbors as described above.

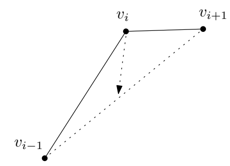


Figure 2: Node v_i executes Go-To-The-Middle strategy by moving into the middle of the interval between v_{i-1} and v_{i+1}.

Since the explorer moves, it may be necessary to extend the path of relay stations. We perform this at the end of the path, between the last relay station and the explorer. This happens every time the distance between v_{n-1} and v_n increases to more than d. We rename the vector v appropriately, so that v_{n+1} describes the explorer and v_n the new relay station. The new relay station is inserted in the middle of the interval connecting the last relay station and the explorer.
Assume that the explorer can carry a sufficiently large pool of relay stations. Then this strategy is easily executed, since new relay stations are available at the explorer's position.
It is easy to proove that with this strategy the path in each step is valid, i.e. the maximum transmission distance d is not exceeded.

We analyze the convergence rate of the Go-To-The-Middle with additional assumption: the explorer does not move. There are n stations, the positions of v_1 and v_n are fixed. For a node v_i we define $d^k(v_i)$ to be the distance of node v_i to the straight line crossing nodes v_1 and v_n before $k-$th step of the execution of Go-To-The-Middle
Distance of a point to a line is defined in the usual geometrical way, as depicted in Fig. 3.

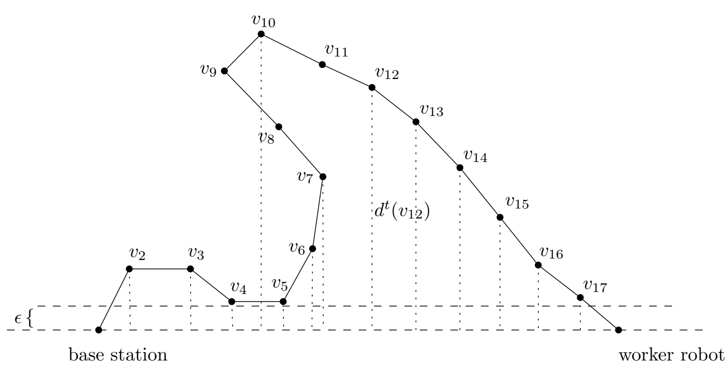


Figure 3: Relay stations and the area of diameter ε around the straight line

We assume that at the beginning all nodes (relay stations) are on one side of the line connecting the explorer and the base station. If not, the nodes can be divided into distinct segments, and the analysis can be applied in each segment separately. The case, when all nodes are on one side yields the worst case.
We will need one more

Lemma 3.11.1. *Let L be $n \times n$ matrix defined as follows: $L(i,j) = \frac{1}{2}$ for all i, j such that $|i-j| = 1$. For all other i, j we have $L(i,j) = 0$ (it is shown later in (3.62))*
The eigenvalues of the matrix L are

$$\lambda_j = \cos\left(\frac{j\pi}{n+1}\right), \quad j = 1, \ldots n \ .$$

The corresponding eigenvectors are

$$x_j(i) = \sin\left(\frac{\pi j i}{n+1}\right), \quad i = 1, \ldots, n, \quad j = 1, \ldots, n \ .$$

Proof:

We will show that the specified eigenvalues and eigenvectors actually correspond to the matrix L. Thus for each pair of eigenvalue λ_j and eigenvector x_j it must hold

$$Lx_j^T = \lambda_j x_j^T \ . \tag{3.58}$$

Let us fix some j and prove Eq. (3.58) for this pair λ_j, x_j. Recall that we claim that

$$\begin{aligned} \lambda_j &= \cos\left(\frac{j\pi}{n+1}\right), \\ x_j &= \left[\sin\left(\frac{j\pi}{n+1}\right), \sin\left(\frac{2j\pi}{n+1}\right), \sin\left(\frac{3\pi}{n+1}\right), \ldots, \sin\left(\frac{nj\pi}{n+1}\right)\right] \ . \end{aligned}$$

From Eq. (3.58) the following system of equalities must hold

$$\frac{1}{2}\sin\left(\frac{2j\pi}{n+1}\right) = \cos\left(\frac{j\pi}{n+1}\right)\sin\left(\frac{j\pi}{n+1}\right), \tag{3.59}$$

$$\frac{1}{2}\sin\left(\frac{(i-1)j\pi}{n+1}\right) + \frac{1}{2}\sin\left(\frac{(i+1)j\pi}{n+1}\right) = \cos\left(\frac{j\pi}{n+1}\right)\sin\left(\frac{ij\pi}{n+1}\right), \tag{3.60}$$

$$\frac{1}{2}\sin\left(\frac{(n-1)j\pi}{n+1}\right) = \cos\left(\frac{j\pi}{n+1}\right)\sin\left(\frac{nj\pi}{n+1}\right) \ , \tag{3.61}$$

where Eq. (3.60) must hold for all $i = 2, \ldots, n-1$.
Observe that one of the basic trigonometric identities is $\sin(2\alpha) = 2\sin\alpha\cos\alpha$. With its help Eq. (3.59) follows easily. Similarly we have

$$\sin\alpha + \sin\beta = 2\sin\left(\frac{\alpha+\beta}{2}\right)\cos\left(\frac{\alpha-\beta}{2}\right) \ ,$$

which can be used to prove Eq. (3.60) for each $i = 2, \ldots, n-1$.
For Eq. (3.61) we first note that $\sin(aj\pi) = \sin((1-a)j\pi)$ for any $j \in \mathbb{N}$ and $0 \le a < 1$. From this we have

$$\sin\left(\frac{(n-1)j\pi}{n+1}\right) = \sin\left(\frac{2j\pi}{n+1}\right) \ ,$$

and

$$\sin\left(\frac{nj\pi}{n+1}\right) = \sin\left(\frac{j\pi}{n+1}\right) \ .$$

This reduces Eq. (3.61) to Eq. (3.60) which has already been shown. □

Now we may proceed with the main theorem of the example.

Theorem 3.11.2. *Consider a valid communication path with $n-2$ relay stations. Then after at most $9n^2 \log \frac{1}{\varepsilon} n$ steps for every i it holds $d(v_i) \le \varepsilon$ for any $\varepsilon > 0$.*

Proof:

Obviously it holds $d^k(v_1) = d^k(v_n) = 0$ for all $k \ge 1$. We define $A^k := [d^k(v_2) \ldots, d^k(v_{n-1})]$ to be the vector of distances of relay stations to the straight line. A^0 describes the start configuration.
Then after one step of Go-To-The-Middle the distance

$$d^k(v_i) = \frac{d^{k-1}(v_{i-1}) + d^{k-1}(v_{i+1})}{2},$$

for all $1 < i < n$, which effectively means that $d^k(v_2) = d^{k-1}(v_3)/2$ and $d^k(v_{n-1}) = d^k(v_{n-2})/2$ since $d^k(v_1) = d^k(v_n) = 0$.
We can describe the changes of the vector A^k by multiplying it with an appropriate transition matrix L so that $A^k = A^{k-1}L = A^0 L^k$. This $n \times n$ matrix is defined as follows: $L(i,j) = \frac{1}{2}$ for all i, j such that $|i-j| = 1$. For all other i, j we have $L(i,j) = 0$.

$$
L = \begin{pmatrix}
 & 1/2 & & & & & \\
1/2 & & 1/2 & & & & \\
 & 1/2 & & \ddots & & & \\
 & & 1/2 & & 1/2 & & \\
 & & & \ddots & & 1/2 & \\
 & & & & 1/2 & & 1/2 \\
 & & & & & 1/2 &
\end{pmatrix} \tag{3.62}
$$

Matrix L is symmetric, substochastic and irreducible. By Lemma 3.11.1 all eigenvalues of L are different and thus L is diagonalizable. From Lemma 3.7.4 we have that after k steps we have $L^k(i,j) \leq \alpha\beta^k$ for any i, j, sticking to the definitions of α and β from that lemma. As all entries of L are not larger than 1 we have $\alpha \leq n \cdot 1$. The value of $|\cos \frac{j\pi}{n+1}|$ is the largest for $j/(n+1)$ approaching 0 or 1. Without loss of generality we set $j = 1$. Then we have $\beta = \cos \frac{\pi}{n+1}$.

Now assuming that the communication distance between nodes is d, we know that A^0 can contain an entry as big as dn. On the other hand we know that entries of L^k are always non-negative. Recall that $A^k = A^0 L^k$. Then to have $d^k(v_i) \leq \varepsilon$ we must have each element of L^k smaller than $\frac{\varepsilon}{dn^2}$ since $d^0(v_i) \leq dn$.

We thus have to find a k such that $L^k(i,j) \leq \frac{\varepsilon}{dn^2}$ for all i, j. Using Lemma 3.7.4 we should then have $n \cdot 1 \cdot \beta^k \leq \frac{\varepsilon}{dn^2}$ and accordingly $\beta^k \leq \frac{\varepsilon}{dn^3}$. We still have to find an upper bound on β. As argued before, β is largest, when $\pi/(n+1)$ approaches 0. Thus let us expand $\cos x$ around $x = 0$ from the Taylor series. We obtain $\cos x \leq 1 - \frac{x^2}{2} + \frac{x^4}{24}$, and set $x = \pi/(n+1)$. Since $\pi^2/2 \geq 1$ and $\pi^4/24 \leq 5$ we obtain

$$
\begin{aligned}
\cos \frac{\pi}{n+1} &\leq 1 - \frac{\pi^2}{2(n+1)^2} + \frac{\pi^4}{24(n+1)^4} \\
&\leq 1 - \frac{1}{(n+1)^2} + 5\frac{1}{(n+1)^4} \ .
\end{aligned}
$$

Since $\frac{5}{(n+1)^4} \leq \frac{1}{2(n+1)^2}$ for a sufficiently large n we have $\cos \frac{\pi}{n+1} \leq 1 - \frac{1}{2(n+1)^2}$. This lets us conclude that for $k = 2(n+1)^2$ we obtain $\beta^k \leq 1/e$ and for $k = 2(n+1)^2 \cdot \ln \frac{1}{\varepsilon} dn^3$ we get $\beta^k \leq \frac{\varepsilon}{dn^3}$. Assuming that d is constant and upper bounding $2(n+1)$ with $3n$ we have that $\beta^k \leq \frac{\varepsilon}{dn^3}$ for $k = 9n^2 \ln \frac{1}{\varepsilon} n$. This proves that after $k = 9n^2 \ln \frac{1}{\varepsilon} n$ steps we have $d^k(v_i) \leq \varepsilon$.

4 Strong Stationary Time

4.1 Introduction

Definition 4.1.1. *Random variable T is a* **stopping time** *for chain $\mathbf{X} = \{X_n, n \geq 0\}$ if it is a function:*

$$T : \mathbf{X} \to \mathbb{N} \cup \{\infty\},$$

such that $T(\mathbf{X})$ is independent of $X_{T(X)+1}, X_{T(X)+2}, \ldots$

We can think of stopping time as of an algorithm which observes Markov chain $\mathbf{X}$ and according to some stopping rule it stops it at some moment. This rule can be a very complicated function of the past, but it cannot depend on the future. The above definition of T we can rewrite as a running time of the following algorithm:

Algorithm 4.1.2 (stopping time)**.**

- At time k, $(X_0, X_1, \ldots, X_k)$ was observed
- Calculate $f_k(X_0, X_1, \ldots, X_k)$, where $f_k : (X_0, X_1, \ldots, X_k) \to \{0, 1\}$
- **If** $f_k(X_0, X_1, \ldots, X_k) = 1$ **then Stop**
- Observe next step and return to the beginning

We call **randomized stopping time** the running time of the following:

Algorithm 4.1.3 (randomized stopping time)**.**

- At time k, $(X_0, X_1, \ldots, X_k)$ was observed
- Calculate $f_k(X)$, where $f_k : (X_0, X_1, \ldots, X_k) \to [0, 1]$
- Flip a coin with probability of getting Head equal to $p = f_k(X_0, X_1, \ldots, X_k)$
- **If** we got Head **then Stop**
- Observe next step and return to the beginning

The difference is that randomized algorithm can at each step decide whether to stop or not using independent randomization, for example flipping a coin.
Note that when in the Algorithm 4.1.3 the functions f_k return only 0 or 1 then it is equivalent to the Algorithm 4.1.2. From now on we will identify stopping time with randomized stopping time.

Introduce the following

Definition 4.1.4. *Random variable T_μ is* **Strong Stationary Time** *for a Markov chain $\mathbf{X} = \{X_n, n \geq 0\}$ with transition matrix $\mathbf{P}$ and initial distribution μ, when it is a stopping time and the following equality holds:*

$$\forall(\mathbf{e} \in E) \quad P_\mu(X_k = \mathbf{e} | T_\mu = k) = \pi(\mathbf{e}).$$

When the stationary distribution is uniform, this random variable is called **Strong Uniform Time**.

We have the following very useful

Lemma 4.1.5 (Aldous, Diaconis [3])**.** *If T_μ is a Strong Stationary Time for $\mathbf{X} = \{X_n, n \geq 0\}$ then*

$$s(\mu \mathbf{P}^n, \pi) \leq P(T_\mu > n). \tag{4.63}$$

Proof.

$$\begin{aligned}
P(X_k = \mathbf{e}_i) &\geq \\
P(X_k = \mathbf{e}_i, T_\mu \leq k) &= \sum_{j=0}^{k} P(X_k = \mathbf{e}_i, T_\mu = j) \\
&= \sum_{j=0}^{k} \sum_{\mathbf{e}_s \in E} P(X_k = \mathbf{e}_i, T_\mu = j, X_j = \mathbf{e}_s) \\
&= \sum_{j=0}^{k} \sum_{\mathbf{e}_s} P(X_k = \mathbf{e}_i | T_\mu = j, X_j = \mathbf{e}_s) \cdot P(T_\mu = j, X_j = \mathbf{e}_s) \\
&= \sum_{j=0}^{k} \sum_{\mathbf{e}_s} P(X_k = \mathbf{e}_i | T_\mu = j, X_j = \mathbf{e}_s) \cdot P(X_j = \mathbf{e}_s | T_\mu = j) \cdot P(T_\mu = j) \\
&= \sum_{j=0}^{k} \sum_{\mathbf{e}_s} P(X_k = \mathbf{e}_i | T_\mu = j, X_j = \mathbf{e}_s) \cdot \pi(\mathbf{e}_s) \cdot P(T_\mu = j) \\
&\overset{(4.64)}{=} \sum_{j=0}^{k} \sum_{\mathbf{e}_s} P(X_k = \mathbf{e}_i | X_j = \mathbf{e}_s) \cdot \pi(\mathbf{e}_s) \cdot P(T_\mu = j) \\
&= \sum_{j=0}^{k} \underbrace{\sum_{\mathbf{e}_s} \pi(\mathbf{e}_s) \delta_{\mathbf{e}_s} \mathbf{P}^{k-j}(\mathbf{e}_i)}_{=\pi(\mathbf{e}_i) \quad (\pi \mathbf{P} = \pi)} \cdot P(T_\mu = j) \\
&= \pi(\mathbf{e}_i) \sum_{j=0}^{k} \cdot P(T_\mu = j) = \pi(\mathbf{e}_i) P(T_\mu \leq k) = \pi(\mathbf{e}_i)[1 - P(T_\mu > k)].
\end{aligned}$$

Thus:

$$1 - \frac{P(X_k = \mathbf{e}_i)}{\pi(\mathbf{e}_i)} \leq P(T_\mu > k).$$

The above is true for any $\mathbf{e}_i$, so it is for maximum:

$$\max_{\mathbf{e}_i} \left(1 - \frac{P(X_k = \mathbf{e}_i)}{\pi(\mathbf{e}_i)} \right) = s(\mu \mathbf{P}^k, \pi) \leq P(T_\mu > k).$$

What is left is to prove:

$$P(X_k = \mathbf{e}_i | T_\mu = j, X_j = \mathbf{e}_s) = P(X_k = \mathbf{e}_i | X_j = \mathbf{e}_s). \tag{4.64}$$

In order to do this we will use following two equalities: $(k \geq j)$

$$P(X_j = \mathbf{e}_s, T = j) = \sum_r P(X_j = \mathbf{e}_s, X_{j-1} = x_{j-1}^r, \ldots X_0 = x_0^r), \tag{4.65}$$

$$P(X_k = \mathbf{e}_i | X_j = \mathbf{e}_s, X_{j-1} = x_{j-1}, \ldots, X_0 = x_0) = P(X_k = \mathbf{e}_i | X_j = \mathbf{e}_s). \tag{4.66}$$

Equality (4.65) follows from fact, that T_μ is a stopping time, thus the event $\{T_\mu = j\}$ depends only on $X_0, X_1, \ldots, X_j$, and (4.66) is the Markov property.

Thus:

$$
\begin{aligned}
& P(X_k = \mathbf{e}_i | X_j = \mathbf{e}_s, T_\mu = j) \\
= \quad & \frac{P(X_k = \mathbf{e}_i, X_j = \mathbf{e}_s, T_\mu = j)}{P(X_j = \mathbf{e}_s, T_\mu = j)} \\
\overset{(4.65)}{=} \quad & \frac{\sum_r P(X_k = \mathbf{e}_i, X_j = \mathbf{e}_s, X_{j-1} = x^r_{j-1}, \ldots, X_0 = x^r_0)}{\sum_r P(X_j = \mathbf{e}_s, X_{j-1} = x^r_{j-1}, \ldots X_0 = x^r_0)} \\
= \quad & \frac{\sum_r P(X_k = \mathbf{e}_i | X_j = \mathbf{e}_s, X_{j-1} = x^r_{j-1}, \ldots, X_0 = x^r_0) \cdot P(X_j = \mathbf{e}_s, X_{j-1} = x^r_{j-1}, \ldots, X_0 = x^r_0)}{\sum_r P(X_j = \mathbf{e}_s, X_{j-1} = x^r_{j-1}, \ldots, X_0 = x^r_0)} \\
\overset{(4.66)}{=} \quad & \frac{P(X_k = \mathbf{e}_i | X_j = \mathbf{e}_s) \sum_r P(X_j = \mathbf{e}_s, X_{j-1} = x^r_{j-1}, \ldots, X_0 = x^r_0)}{\sum_r P(X_j = \mathbf{e}_s, X_{j-1} = x^r_{j-1}, \ldots, X_0 = x^r_0)} \\
= \quad & P(X_k = \mathbf{e}_i | X_j = \mathbf{e}_s).
\end{aligned}
$$

□

Lemma 4.1.5 together with equality (2.8) gives useful bound on total variation distance:

$$d(\mu \mathbf{P}^n, \pi) \le s(\mu \mathbf{P}^n, \pi) \le P(T_\mu > n).$$

We will need the following

Lemma 4.1.6. *For an ergodic Markov chain $\mathbf{X} = \{X_n, n \ge 0\}$ and any initial distribution μ, separation distance $s(\mu \mathbf{P}^n, \pi)$ is a decreasing function of n.*

Proof. We will use equivalent definition of separation distance (equality (2.7), i.e. $s(\mu \mathbf{P}^n, \pi)$ is the smallest number s_n such $\mu \mathbf{P}^n = (1 - s_n)\pi + s_n V_n$, where V_n is some distribution). So we have:

$$
\begin{aligned}
\mu \mathbf{P}^{n+1}(\mathbf{e}) &= (1 - s_{n+1})\pi(\mathbf{e}) + s_{n+1} V_{n+1}(\mathbf{e}), \\
\mu \mathbf{P}^n(\mathbf{e}_s) &= (1 - s_n)\pi(\mathbf{e}_s) + s_n V_n(\mathbf{e}_s).
\end{aligned}
$$

We can calculate:

$$\mu \mathbf{P}^{n+1}(\mathbf{e}) = \sum_{\mathbf{e}_s \in E} \mu \mathbf{P}^n(\mathbf{e}_s) \delta_{\mathbf{e}_s} \mathbf{P}(\mathbf{e}) = \sum_{\mathbf{e}_s \in E} \left[(1 - s_n)\pi(\mathbf{e}_s) + s_n V_n(\mathbf{e}_s)\right] \delta_{\mathbf{e}_s} \mathbf{P}(\mathbf{e})$$

$$= (1 - s_n) \sum_{\mathbf{e}_s \in E} \pi(\mathbf{e}_s) \delta_{\mathbf{e}_s} \mathbf{P}(\mathbf{e}) + s_n \sum_{\mathbf{e}_s \in E} V_n(\mathbf{e}_s) \delta_{\mathbf{e}_s} \mathbf{P}^n(\mathbf{e}) = (1 - s_n)\pi(\mathbf{e}) + s_n \sum_{\mathbf{e}_s \in E} V_n(\mathbf{e}_s) \delta_{\mathbf{e}_s} \mathbf{P}^n(\mathbf{e}),$$

thus

$$s_{n+1} = s(\mu \mathbf{P}^{n+1}(\mathbf{e}), \pi) \le s_n.$$

□

Definition 4.1.7. *Random variable T_μ is called* **Minimal Strong Stationary Time** *(shortly MSST) for Markov chain $\mathbf{X} = \{X_n, n \ge 0\}$ if it is Strong Stationary Time for which equality in (4.63) holds, i.e. for which:*

$$s(\mu \mathbf{P}^n, \pi) = P(T_\mu > n).$$

Again, when stationary distribution is uniform, T_μ is called **Minimal Strong Uniform Time***.*

Does there always exist such a Strong Stationary Time? Does there exist a Minimal Strong Uniform Time for given chain? The answer is given by D. Aldous and P. Diaconis in [3] in form of the following

Lemma 4.1.8 (Aldous, Diaconis [3]). *Let $\mathbf{X} = \{X_n, n \geq 0\}$ be ergodic Markov chain with state space E. Then, for any initial distribution μ, there exists Minimal Strong Stationary Time T_μ for this chain.*

Proof. We will construct the desired random variable T_μ. Define $\alpha_n = \min_{\mathbf{e}} \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}$ (set $\alpha_0 = \min_{\mathbf{e}} \frac{1}{\pi(\mathbf{e})}$). T_μ is defined as follows: at time n, given that random walk $\mathbb{X}$ is in state $\mathbf{e}$ flip a coin with probability of heads

$$p_n(\mathbf{e}) = \frac{\alpha_n - \alpha_{n-1}}{\frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})} - \alpha_{n-1}}.$$

If heads comes up, stop, otherwise flip again with probability $p_{k+1}(\mathbf{e})$ and so on. Denote the first time when head occurs by T_μ, i.e. $P(T_\mu = n | X_n = \mathbf{e}, T > n-1) = p_n(\mathbf{e})$. Notice that $p_n(\mathbf{e}) \geq 0$ for all $\mathbf{e} \in E, n \geq 0$. Let k be the smallest integer such that $\alpha_k > 0$, of course then we have $p_s(\mathbf{e}) = 0$ for $s < k$ and

$$p_k(\mathbf{e}) = \frac{\alpha_k}{\mu\mathbf{P}^k(\mathbf{e})} = \frac{P(T_\mu = k, X_k = \mathbf{e})}{P(X_k = \mathbf{e})} = P(T_\mu = k | X_k = \mathbf{e}).$$

We also have

$$P(T_\mu = k) = \sum_{\mathbf{e}} P(T_\mu = k, X_k = \mathbf{e}) = \sum_{\mathbf{e}} P(T_\mu = k | X_k = \mathbf{e}) P(X_k = \mathbf{e})$$

$$= \sum_{\mathbf{e}} \frac{\alpha_k \pi(\mathbf{e})}{P(X_k = \mathbf{e})} P(X_k = \mathbf{e}) = \alpha_k,$$

thus

$$P(X_k = \mathbf{e} | T_\mu = k) = \frac{P(X_k = \mathbf{e}, T_\mu = k)}{P(T_\mu = \mathbf{e})} = \frac{P(T_\mu = n | X_k = \mathbf{e}) P(X_k = \mathbf{e})}{P(T_\mu = k)} = \frac{\frac{\alpha_k \pi(\mathbf{e})}{P(X_k = \mathbf{e})} P(X_k = \mathbf{e})}{\alpha_k} = \pi(\mathbf{e}).$$

By induction we will show that

$$P(X_n = \mathbf{e}, T_\mu = n) = \pi(\mathbf{e})(\alpha_n - \alpha_{n-1}), \quad n \geq k, \quad \mathbf{e} \in E. \tag{4.67}$$

It is equivalent to

$$P(X_n = \mathbf{e} | T_\mu = n) P(T_\mu = n) = \pi(\mathbf{e})(\alpha_n - \alpha_{n-1}), \quad n \geq k, \quad \mathbf{e} \in E.$$

For $n > k$

$$P(X_n = \mathbf{e}, T_\mu = n) = P(T_\mu = n | X_n = \mathbf{e}, T_\mu > n-1) P(X_n = \mathbf{e}, T_\mu > n-1) =$$

$$= \frac{\alpha_n - \alpha_{n-1}}{\left(\frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})} - \alpha_{n-1}\right)} \cdot \left[P(X_n = \mathbf{e}) - P(X_n = \mathbf{e}, T_\mu \leq n-1)\right]. \tag{4.68}$$

We yet need to calculate

$$\begin{aligned}
P(X_n = \mathbf{e}, T_\mu \leq n-1) &= \sum_{j=1}^{n-1} P(X_n = \mathbf{e}, T_\mu = j) = \sum_{j=1}^{n-1} \sum_{\mathbf{e}_s \in E} P(X_n = \mathbf{e}, T_\mu = j, X_j = \mathbf{e}_s) \\
&= \sum_{j=1}^{n-1} \sum_{\mathbf{e}_s \in E} P(X_n = \mathbf{e} | T_\mu = j, X_j = \mathbf{e}_s) P(T_\mu = j, X_j = \mathbf{e}_s) \\
&\overset{(4.67)}{=} \sum_{j=1}^{n-1} \sum_{\mathbf{e}_s \in E} P(X_n = \mathbf{e} | T_\mu = j, X_j = \mathbf{e}_s) \pi(\mathbf{e}_s)(\alpha_j - \alpha_{j-1}) \\
&\overset{(4.64)}{=} \sum_{j=1}^{n-1} \sum_{\mathbf{e}_s \in E} P(X_n = \mathbf{e} | X_j = \mathbf{e}_s) \pi(\mathbf{e}_s)(\alpha_j - \alpha_{j-1}) \\
&= \sum_{j=1}^{n-1} \underbrace{\sum_{\mathbf{e}_s} \pi(\mathbf{e}_s) \delta_{\mathbf{e}_s} \mathbf{P}^{k-j}(\mathbf{e})}_{=\pi(\mathbf{e}) \quad (\pi \mathbf{P} = \pi)} \cdot (\alpha_j - \alpha_{j-1}) \\
&= \sum_{j=1}^{n-1} \pi(\mathbf{e})(\alpha_j - \alpha_{j-1}) = \pi(\mathbf{e}) \alpha_{n-1}.
\end{aligned}$$

Using this in (4.68) we obtain $(P(X_n = \mathbf{e}) \equiv \mu \mathbf{P}^n(\mathbf{e}))$

$$P(X_n = \mathbf{e}, T_\mu = \mathbf{e}) = \frac{\alpha_n - \alpha_{n-1}}{\left(\frac{\mu \mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})} - \alpha_{n-1}\right)} \cdot [\mu \mathbf{P}^n(\mathbf{e}) - \pi(\mathbf{e}) \alpha_{n-1}]$$

$$= \frac{\alpha_n - \alpha_{n-1}}{\left(\frac{\mu \mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})} - \alpha_{n-1}\right)} \cdot \left[\pi(\mathbf{e}) \left(\frac{\mu \mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})} - \alpha_{n-1}\right)\right] = \pi(\mathbf{e})(\alpha_{n-1} - \alpha_n). \tag{4.69}$$

So we have

$$P(T_\mu = n) = \sum_{\mathbf{e}} P(T = n, X_n = \mathbf{e}) = \sum_{\mathbf{e}} \pi(\mathbf{e})(\alpha_n - \alpha_{n-1}) = \alpha_n - \alpha_{n-1}, \tag{4.70}$$

thus

$$P(X_n = \mathbf{e} | T_\mu = n) = \frac{P(X_n = \mathbf{e}, T_\mu = n)}{P(T_\mu = n)} \overset{(4.69)}{=} \frac{\pi(\mathbf{e})(\alpha_n - \alpha_{n-1})}{P(T_\mu = n)} \overset{(4.70)}{=} \frac{\pi(\mathbf{e})(\alpha_n - \alpha_{n-1})}{\alpha_n - \alpha_{n-1}} = \pi(\mathbf{e}),$$

what finishes the proof.

□

4.2 Example: Random walk on $\mathbb{Z}_4$

0

3 1

2

Figure 4: $\mathbb{Z}_4$

Random walk starts at 0 ($X_0 = 0$), $E = \{1, 2, 3, 4\}$. Later the chain stays at the same state with probability $\frac{1}{2}$ or goes left or right with probability $\frac{1}{4}$. This can be written as:

$$X_k = X_{k-1} + \theta_k \bmod 4,$$

where $\theta_1, \theta_2, \ldots$ are iid random variables with distribution: $P(\theta_i = 1) = P(\theta_i = -1) = \frac{1}{4}, P(\theta_i = 0) = \frac{1}{2}$.
Here is an algorithm which running time is Strong Uniform Time for above random walk:

Algorithm 4.2.1 (for Z_4)**.**

- Observe $X = (X_0, X_1, \ldots, X_m)$ till first time when $X_m \in \{1, 3\}$
- Do one step more
- **Stop**

Of course the running time of above algorithm is a stopping time. To prove that it is Strong Uniform Time, let us assume that $T = k$. Because of symmetry in step $k-1$ random walk wast either in 4 or in 1 with the same probability, i.e.:

$$P(X_{k-1} = 1 | T = k) = P(X_{k-1} = 3 | T = k) = \tfrac{1}{2},$$

Thus for any $i = 0, 1, 2, 3$ we have

$$P(X_k = i | T = k) = \frac{1}{2} P(X_{k-1} = i | T = k)$$

$$+\frac{1}{4}[P(X_{k-1} = i+1 | T = k) + P(X_{k-1} = i-1 | T = k)] = \tfrac{1}{4} = \tfrac{1}{|E|}.$$

In order to have a bound on $d(\mu\mathbf{P}^n, \pi)$ and to use lemma 4.1.5 we have to calculate:

$$P(T > k) = \sum_{i=k+1}^{\infty} \left(\frac{1}{2}\right)^{i-2} \cdot \frac{1}{2} = 2 \sum_{k+1}^{\infty} \left(\frac{1}{2}\right)^{i} = 2 \cdot \frac{(\frac{1}{2})^{k+1}}{1 - \frac{1}{2}} = 2 \cdot \left(\frac{1}{2}\right)^{k}.$$

Thus:

$$d(\mu\mathbf{P}^k, \pi) \leq s(\mu\mathbf{P}^k, \pi) \leq 2 \cdot \left(\frac{1}{2}\right)^{k}.$$

4.3 Example: Random walk on $\mathbb{Z}_n$

The example is similar to the example of Diaconis and Fill [15].
Assume $n = 2^m$. For $n = 16$ it is a random walk on n-points circle:

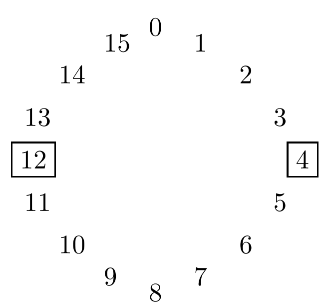


Figure 5: $\mathbb{Z}_{16}$

Similarly as in previous example we can write:

$$X_k = X_{k-1} + \theta_k \bmod n,$$

where $\theta_1, \theta_2, \ldots$ are iid with distribution: $P(\theta_i = 1) = P(\theta_i = -1) = \frac{1}{4}, P(\theta_i = 0) = \frac{1}{2}$.
Here is an algorithm which running time is Strong Uniform Time for above walk:

Algorithm 4.3.1 (for $\mathbb{Z}_n$).

- Observe $X = (X_0, X_1, \ldots, X_m)$ till first time (T_1) when $X_m \in \{2^{m-2}, 3 \cdot 2^{m-2}\}$
- Observe random walk till first time after T_1 when it gets to state which is in distance of 2^{m-3} from X_{T_1} (call this moment T_2)
- Observe random walk till first time after T_2 when it gets to state which is in distance of 2^{m-4} from X_{T_2} (call this moment T_3)

$\vdots$

- Observe random walk till first time after T_{m-2} when it gets to state which is in distance of 1 from $X_{T_{m-2}}$ (call this moment T_{m-1} - at this time random walk is in state which is an odd number)
- Observe one step more ($T = T_m$)
- **Stop**

Of course running time of this algorithm is a stopping time. Assume that $T_1 = k_1$, then, because of symmetry, X_{k_1} is in one of the states from $\{2^{m-2}, 3 \cdot 2^{m-2}\}$ with equal probability (all paths $i_1, i_2, \ldots, i_{k_1}$ such that at time $k-1$ are first time in state 2^{m-2} or $3 \cdot 2^{m-2}$ are equally likely). The same is at moment T_2, and because of symmetry X_{T_2} has uniform distribution on $\{2^{m-2} + 2^{m-3}, 2^{m-2} - 2^{m-3}, 3 \cdot 2^{m-2} + 2^{m-3}, \cdot 2^{m-2} - 2^{m-3}\}$.
So, because of symmetry, at time T_{m-1} the distribution of $X_{T_{m-1}}$ is uniform on all possible odd numbers, i.e.

$$P(X_k = 2i + 1 | T_{m-1} = k) = \frac{1}{2^{m-1}}.$$

To calculate $P(X_k = i | T = k)$ consider separately cases when i is odd and even. If i is odd, then the step earlier it had to be in the same state and had to stay in it. If i is even, then a step earlier it had to be in some neighbouring odd states and in one step it had to get to i. So we can calculate:

$$P(X_k = 2i + 1 | T = k) = \frac{1}{2} P(X_{k-1} = 2i + 1 | T_{m-1} = k) = \frac{1}{2} \cdot \frac{1}{2^{m-1}} = \frac{1}{2^m}.$$

$$\begin{aligned} P(X_k = 2i | T = k) &= \frac{1}{4} P(X_{k-1} = 2i - 1 | T_{m-1} = k) + \frac{1}{4} P(X_{k-1} = 2i + 1 | T_{m-1} = k) \\ &= \frac{1}{4} \cdot \frac{1}{2^{m-1}} + \frac{1}{4} \cdot \frac{1}{2^{m-1}} = \frac{1}{2^m}. \end{aligned}$$

Thus T is a Strong Uniform Time. To estimate $P(T > k)$ we will use Markov's inequality (for $Y \geq 0, \quad P(Y > k) \leq \frac{EY}{k}$) and the following lemma:

Lemma 4.3.2 (Feller [19]). *For random walk on integers starting from 0 and moving $+1$ or -1 with probability $\frac{\theta}{2}$, and staying in the same state with probability $1 - \theta$, expected time of getting to $\pm b$ is $\frac{b^2}{\theta}$.*

□

In random walk on $\mathbb{Z}_n$ we first have to get to state in distance $\pm 2^{m-2}$, later to state in distance $\pm 2^{m-3}$ etc., last time to state in distance 1 and to make one step more. Here $\theta := \frac{1}{2}$ so:

$$\begin{aligned} ET &= 2\left(2^{m-2}\right)^2 + 2\left(2^{m-3}\right)^2 + \ldots + 2\left(2^{m-(m-1)}\right)^2 + 2\left(2^{m-m}\right)^2 + 1 \\ &= 2(4^{m-3} + 4^{m-3} + \ldots + 1) + 1 = 2 \cdot \frac{4^{m-1} - 1}{4 - 1} = \frac{2}{3}\left(\frac{1}{4}n^2 - 1\right) + 1 = \frac{1}{6}n^2 + \frac{1}{3}. \end{aligned}$$

Let $k = cn^2$, $c \in \{1, 2, ...\}$. Then

$$d(\mu \mathbf{P}^k, \pi) \leq s(\mu \mathbf{P}^k, \pi) \leq P(T > k) \leq \frac{ET}{k} = \frac{\frac{1}{6}n^2 + \frac{1}{3}}{cn^2} \leq \frac{1}{5c}.$$

So the speed of convergence is of order n^2, we have to make $c \cdot n^2$ steps to be near to the uniform distribution.

4.4 Example: Shuffling cards : top-to-random

This example is from Aldous and Diaconis [2], calculations can be found in Diaconis [13].

We have n cards (underlines denote n gaps between cards)

$$k_1, \_, k_2, \_, k_3, \_, \ldots, \_, k_{n-1}, \_, k_n, \_$$

Shuffling cards top-to-random is a procedure which takes first card from the top (at the beginning: k_1) and puts it randomly (i.e. with uniform distribution) in gaps. So the probability it gets into fixed position is $\frac{1}{n}$, in particular this is the probability of card getting under k_n or staying in the same (first) position.
The state space here is a set of all permutations of these cards: $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N\}$, where $N = n!$. The stationary distribution is a uniform one. How many steps are required to be "near" stationary (uniform) distribution (to mix cards well)?

Algorithm 4.4.1 (for shuffling cards top-to-random)**.**

- Observe $X = (X_0, X_1, \ldots, X_m)$ till first card get under k_n (denote this moment by T_1)
- Observe random walk till second card get under k_n (denote this moment by T_2)

  ⋮

- Observe random walk till there will be $n-1$ cards under k_n (i.e. card k_n is on the top, denote this moment by T_{n-1})
- Do one step more
- **Stop**

Of course T is a stopping time. The proof that it is a Strong Uniform Time is by induction. At each moment, conditioned on cards which are under k_n, every permutation of them is equally likely.
Not formally: Let us look at the moment when second card (k_j) goes under k_n (at moment T_2). Assume, that now there are two cards k_i and k_j. Two arrangements are possible: (k_i, k_j) or (k_j, k_i). Card k_j could get under or over k_i with the same probability ($\frac{1}{n}$), so both arrangements are equally likely.
Formally, the position can be described in the following way:

$$\{M, \{k_{i_1}, \ldots, k_{i_M}\}, \Pi_M\},$$

where:

$$\begin{aligned} M &\quad \text{- number of cards under } k_n, \\ k_{i_1}, \ldots k_{i_M} &\quad \text{- cards under } k_n, \\ \Pi_M &\quad : \{1, \ldots, M\} \to \{k_{i_1}, \ldots, k_{i_M}\} \text{ - permutation.} \end{aligned}$$

Lemma 4.4.2. *At each moment, conditioned on* $M, \{k_{i_1}, \ldots, k_{i_M}\}$ *the permutation* Π_M *is uniform.*

Proof. By induction:
Of course it is true when $M = 0$ or $M = 1$.
Assume it is true for $M = r$, it remains true till next card $k_{i_{r+1}}$ gets under k_n:

$$k_{i_1}, \_, k_{i_2}, \_, k_{i_3}, \_, \ldots, \_, k_{i_{r-1}}, \_, k_{i_r}, \_$$

Because of assumption before this new card gets under k_n the distribution over r cards is uniform. New card has the same chance to get into every gap between cards $k_{i_1}, \ldots, k_{i_r}$, so each of new arrangements are equally likely, i.e. the probability of each permutation Π_{M+1} has the same probability.

□

Lemma 4.4.3. *In the above model let* $k = n \log n + cn$. *Then*

$$d(\mu \mathbf{P}^k, U) \le e^{-c} \qquad \text{for} \quad c \ge 0, n \ge 2.$$

Proof.
Notice that T can be rewritten as (defining T_0=0)

$$T = \underbrace{T_n - T_{n-1}}_{X_n} + \underbrace{T_{n-1} - T_{n-2}}_{X_{n-1}} + \ldots + \underbrace{T_1 - T_0}_{X_1}. \tag{4.71}$$

Random variable X_1 has geometric distribution with parameter $\frac{1}{n}$ (what we will denote: $X_1 \sim Geo(\frac{1}{n})$), it is waiting time for first success which is when first card gets under k_n. In general: $X_i \sim Geo(\frac{i}{n})$.

At first we proof weaker result using Chebyshev's inequality ($P(|T - ET| \geq c\sigma) \leq \frac{1}{c^2}$).
To do that calculate $\mathbb{E}T$ and $VarT$.

$$\mathbb{E}T = n + \frac{n}{2} + \frac{n}{3} + \ldots + \frac{n}{n} = n \sum_{i=1}^{n} \frac{1}{i} \approx n \log n$$

$$VarT = \sum_{i=1}^{n} VarX_i = \sum_{i=1}^{n} \frac{1 - \frac{1}{n}}{(\frac{i}{n})^2} = \sum_{i=1}^{n} \left(1 - \frac{i}{n}\right) \left(\frac{n}{i}\right)^2 =$$

$$= n^2 \sum_{i=1}^{n} \frac{1}{i^2} - n \sum_{i=1}^{n} \frac{1}{i} \approx \frac{\pi^2}{6} n^2 - n \log n \approx n^2$$

For $k = n \log n + cn$ from Chebyshev's inequality we obtain:

$$d(\mu \mathbf{P}^k, \pi) \leq s(\mu \mathbf{P}^k, \pi) \leq P(T > k) =$$

$$= P(T > n \log n + cn) \approx P(T > \mathbb{E}T + c\sqrt{VarT}) \leq \tfrac{1}{c^2}.$$

To finish the proof consider drawing balls from urn with replacement. There are n different balls. Let V denotes number of drawings till time each ball is drawn at least once. For each ball i we introduce

$$A_i = \begin{cases} 1 & \text{ball } i \text{ was not drawn in first } m \text{ experiments,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.72}$$

Set $m = n \log n + cn$.

$$P(V > m) = P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i) = n \left(1 - \frac{1}{n}\right)^m \leq n e^{-\frac{m}{n}} = e^{-c}$$

V can be rewritten as

$$V = (V - V_{n-1}) + (V_{n-1} - V_{n-2}) + \ldots + (V_2 - V_1) + V_1,$$

where V_j is the number of experiments till j different balls are chosen at least once. After j different balls are drawn the chance that new drawing will give new ball is $\frac{n-i}{n}$ so $V_{i+1} - V_i$ has a geometric distribution with parameter $1 - \frac{i}{n}$. It follows that T and V have the same distributions, thus

$$P(T > n \log n + cn) \leq e^{-c}.$$

□

4.5 Example: Symmetric walk on $d-$demensional cube

This example can be found in Aldous and Diaconis [3]. We give no prove here.

State space: $E = \{0,1\}^d$. Denote $s = (s_1, \ldots, s_d), s' = (s'_1, \ldots, s'_d) \in E$, where $s_i, s'_i \in \{0,1\}$. Probability transition matrix $\mathbf{P}$ in this case is of form:

$$\mathbf{P}(s,s') = \begin{cases} \frac{1}{d+1} & \text{if} \qquad \sum_{i=1}^d |s_i - s'_i| = 0 \quad \text{or} \quad 1, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{4.73}$$

I.e. in one step from state s only states which differs on one coordinate can be reached, each with probability $\frac{1}{d+1}$ or with the same probability chain can stay in the same state.
The state space for case $d = 3$ and possible movements (arrows) from state $(0,0,0)$ are shown in Figure 6.

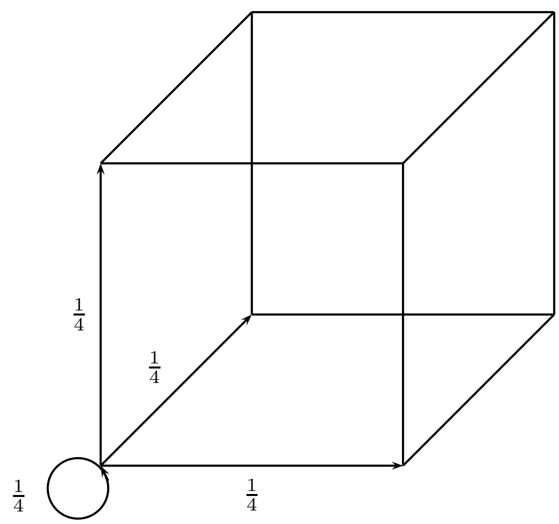


Figure 6: Example: Random Walk on 3D cube

It turns out that in this case the speed of convergence is exactly the same as in shuffling cards top-in-random. Here the Strong Uniform Time T as in (4.71) can be written in the following way:

$$T = \underbrace{T_d - T_{d-1}}_{X_d} + \underbrace{T_{d-1} - T_{d-2}}_{X_{d-1}} + \ldots + \underbrace{T_1 - T_0}_{X_1},$$

where X_i are iid with distribution $Geo(\frac{i}{d})$. So after $k = d \log d + cd$ steps we have:

$$d(\mu\mathbf{P}^k, \pi) \le s(\mu\mathbf{P}^k, \pi) \le e^{-c}.$$

□

4.6 Example: matching in graph

We start with the following

Definition 4.6.1 (Matching in graph)**.** *The set of edges having no common vertex, such that each vertex belongs to some edge from this set (graph with odd number of vertices can have one vertex not belonging to matching).*

Here are all 3 possible matchings for 4 vertices:

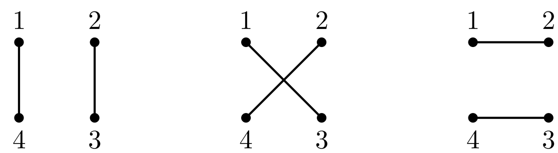


Figure 7: Possible matchings for 4 vertices

We will be looking for random matching in given complete graph with even number of vertices. In such a graph if we label vertices from 1 to $2n$ matching can be defined as division of set $\{1, 2, \ldots, 2n\}$ into n 2-elements subsets. So matchings from Figure 7 can be represented as: $\{1,2\}\{3,4\}; \{1,3\}\{2,4\}; \{1,4\}\{2,3\}$ Number of matchings can be counted in following way: we put all $2n$ numbers one next to another and take consecutive pairs as matching's elements; we can arrange $2n$ numbers in $(2n)!$ ways, this we have to devide by 2^n (the order in pairs does not count) and devide it by $n!$ (order of pairs does not count either). So we have $N = \frac{(2n)!}{2^n n!}$ matchings in graph with $2n$ vertices.

We consider Markov chain with state space $E = \{\text{set of matchings}\}$ which evolves in the following way: from state $\mathbf{s}_1 \in E$ in one step it can go to state $\mathbf{s}_2$ which differs only this, that in only two pairs vertices were exchanged, or it stays in the same state.

More precise: Being in state $\mathbf{s}_1 = \{i_1, i_2\}, \ldots, \{i_{2n-1}, i_{2n}\}$ we chose 2 pairs uniformly at random. If the same pairs are chosen we do nothing, otherwise we flip a coin: assume that pairs $\{i_1, i_2\}$ and $\{j_1, j_2\}$ were chosen: if Head then we exchange vertices in these pairs to get $\{i_1, j_1\}$ and $\{j_1, i_2\}$; if Tail then we exchange vertices to get $\{i_1, j_2\}$ and $\{j_1, i_2\}$.

So the probability transition matrix $\mathbf{P} = [\mathbf{P}(\mathbf{s}_i, \mathbf{s}_j)]_{(i,j=1\ldots N)}$ is following:

$$\mathbf{P} = \quad \mathbf{P}(\mathbf{s}_i, \mathbf{s}_j) = \begin{cases} \frac{1}{n} & \text{if } \ \mathbf{s}_i = \mathbf{s}_j, \\ \frac{1}{n^2} & \text{if } \mathbf{s}_i \ i \ \mathbf{s}_j \text{ differs by exchange,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.74}$$

First notice that indeed it is a probability matrix.

From any state we can get to one of $\binom{n}{2} \cdot 2$ other states (there are $\binom{n}{2}$ ways of choosing a pair and in each pair we can exchange vertices in two ways), so

$$\forall(\mathbf{s} \in E) \qquad \sum_i \mathbf{P}(\mathbf{s}, \mathbf{s}_i) = 1 \cdot \frac{1}{n} + \binom{n}{2} \cdot 2 \cdot \frac{1}{n^2} = \frac{1}{n} + \frac{1}{n^2} \cdot \frac{n(n-1)}{2} \cdot 2 = \frac{n^2}{n^2} = 1.$$

Matrix $\mathbf{P}$ is symmetric, so the stationary distribution is uniform.

The movement in $k-$th step can be described as $\{L_k, R_k, \theta_k\}$ where L_k is a pair chosen with left hand, R_k is a pair chosen with right one, $\theta_k \in \{Tail, Head\}$ is the result of independently flipped coin. The idea in finding a Strong Uniform Time is to observe random walk and marking some vertices in such way, that all the time distribution of matchings on already marked vertices is uniform. Here we have the algorithm which running time is a desired Strong Uniform Time:

Algorithm 4.6.2 (for matching)**.**

- **If** $L_k = R_k$ and chooses pair is unmarked - mark it
- **If** $L_k \neq R_k$ **and** one of pairs L_k, R_k is marked - marked the other
- **If** all pairs are marked then **Stop**
- Observe next step and return to the beginning

Let $T_1, T_2, \ldots, T := T_n$ denote succeeding moments when first, second, $\ldots$, $n-$th pair is marked.

Lemma 4.6.3. *At each step, conditioned on marked pairs and vertices in these pairs, distribution on them is uniform. So T is the Strong Uniform Time.*

Proof.
Induction by number of marked pairs will be used. When there is 0 or 1 pair marked of course the lemma is true.

Assume that at time T_1 a pair $\{i_1, i_2\}$ was marked - notice that before moment T_2 this pair surely was not changed - random walk could only chose another pairs and make exchanges there (otherwise another pair would be marked, but it is impossible before T_2) or it could also happen that $L_k = R_k = \underline{\{i_1, i_2\}}$, but then nothing was exchanged. Assume that in moment T_2 a pair $\underline{\{j_1, j_2\}}$ is marked. There are 3 possible matchings on marked elements:

$$\underline{\{i_1, i_2\}}\ \underline{\{j_1, j_2\}}\ ;\ \underline{\{i_1, j_2\}}\ \underline{\{j_1, i_2\}}\ ;\ \underline{\{i_1, j_1\}}\ \underline{\{i_2, j_2\}}$$

Calculate probability each of them:

$$\begin{aligned}
\Pr(\underline{\{i_1, i_2\}}\ \underline{\{j_1, j_2\}}) &= \tfrac{1}{n^2} \\
&\quad \text{it had to be } L_k = R_k = \{j_1, j_2\} \\
\Pr(\underline{\{i_1, j_2\}}\ \underline{\{j_1, i_2\}}) &= \tfrac{1}{n} \cdot \tfrac{1}{n} \cdot \tfrac{1}{2} + \tfrac{1}{n} \cdot \tfrac{1}{n} \cdot \tfrac{1}{2} = \tfrac{1}{n^2} \\
&\quad \text{or} : L_k = \{i_1, i_2\}, R_k = \{j_1, j_2\},\ \theta_k = Tail \\
&\quad \text{or} : L_k = \{j_1, j_2\}, R_k = \{i_1, i_2\},\ \theta_k = Tail \\
\Pr(\underline{\{i_1, j_1\}}\ \underline{\{i_2, j_2\}}) &= \tfrac{1}{n} \cdot \tfrac{1}{n} \cdot \tfrac{1}{2} + \tfrac{1}{n} \cdot \tfrac{1}{n} \cdot \tfrac{1}{2} = \tfrac{1}{n^2} \\
&\quad \text{or} : L_k = \{j_1, j_2\}, R_k = \{i_1, i_2\},\ \theta_k = Head \\
&\quad \text{or} : L_k = \{i_1, i_2\}, R_k = \{j_1, j_2\},\ \theta_k = Head
\end{aligned}$$

So conditioning on event that in time T_1 a pair $\underline{\{i_1, i_2\}}$ was chosen and in time T_2 a pair $\underline{\{j_1, j_2\}}$ was chosen, all three mathcings are equally likely, so the distribution on these pairs is uniform.

The reasoning with M pairs marked is similar. Assume that at time k there are M pairs marked: $\underline{\{i_1^1, i_2^1\}}, \underline{\{i_1^2, i_2^2\}}, \ldots, \underline{\{i_1^M, i_2^M\}}$ In next steps till new card is chosen the distribution on them is uniform (from induction assumption) Assume that later a new pair $\underline{\{j_1, j_2\}}$ is marked. Such situations are possible: either these two vertices will be together in pair: $\underline{\{j_1, j_2\}}$ either vertices in newly chosen pair and in some other already marked will be exchanged: new pairs are then $\underline{\{i_1^s, j_1\}}, \underline{\{i_2^s, j_2\}}$. Calculate probabilities of these all situations:

$$\begin{aligned}
\Pr(\underline{\{i_1^1, i_2^1\}}, \ldots, \qquad\qquad \underline{\{i_1^M, i_2^M\}}, \underline{\{j_1, j_2\}}) &= \tfrac{1}{n^2} \\
&\quad \text{it had to be } L_k = R_k = \{j_1, j_2\} \\
\Pr(\underline{\{i_1^1, i_2^1\}}, \ldots, \underline{\{i_1^s, j_1\}}, \ldots, \underline{\{i_1^M, i_2^M\}}, \underline{\{i_2^s, j_2\}}) &= \tfrac{1}{n} \cdot \tfrac{1}{n} \cdot \tfrac{1}{2} + \tfrac{1}{n} \cdot \tfrac{1}{n} \cdot \tfrac{1}{2} = \tfrac{1}{n^2} \\
&\quad \text{or} : L_k = \{i_1, i_2\}, R_k = \{j_1, j_2\},\ \theta_k = Tail \\
&\quad \text{or} : L_k = \{j_1, j_2\}, R_k = \{i_1, i_2\},\ \theta_k = Tail
\end{aligned}$$

So all of these $2M+1$ situations are equally likely, thus in every moment the distribution on marked vertices is uniform. It follows that random variable T (the running time of algorithm 4.6.2) is a Strong Uniform Time.

$\square$

The speed of convergence is given in the following lemma:

Lemma 4.6.4. *For $c > 0$ after $k = 2n \log n + cn$ steps total variation distance between $\mathbf{P}^k$ (the distribution of matchings in $k-th$ step) and uniform distribution is less or equal to $\frac{1}{c^2}$.*

Proof.
To estimate $P(T > k)$ notice that intervals between moment when some pair is chose are iid random variables with geometric distribution. We can write (defining $T_0 = 0$):

$$T = \underbrace{T_n - T_{n-1}}_{X_n} + \underbrace{T_{n-1} - T_{n-2}}_{X_{n-1}} + \ldots + \underbrace{T_1 - T_0}_{X_1},$$

where X_1 is waiting time for first moment in which $L_k = R_k$, so it has distribution $Geo(\frac{1}{n})$.
X_2 is a first moment after X_1 in which new pair is marked, what can happen in two ways: whether $L_k = R_k$ and this pair was not marked earlier or exactly one of pairs L_k and R_k is unmarked. Calculate:

$$P(X_2 = 1) = \frac{1}{n^2} \cdot (n-1) + \frac{1}{n^2} \cdot (n-1) = \frac{2(n-1)}{n^2}.$$

So X_2 has distribution $Geo(\frac{2}{n})$.

In general: $X_i \sim Geo\Big(\frac{(n-i)(i+1)}{n^2}\Big)$. So we can calculate $\mathbb{E}T$ and $VarT$:

$$\mathbb{E}T = \sum_{i=0}^{n-1} \frac{n^2}{(i+1)(n-i)} = n^2 \sum_{i=1}^{n} \frac{1}{i(n+1-i)} = n^2 \sum_{i=1}^{n} \frac{1}{(n+1)} \left(\frac{1}{i} + \frac{1}{n+1-i} \right)$$

$$= \frac{n^2}{n+1} \cdot 2 \sum_{i=1}^{n} \frac{1}{k} \approx 2n \log n,$$

$$VarT = \sum_{i=0}^{n-1} \frac{1 - \frac{(i+1)(n-i)}{n^2}}{\left(\frac{(i+1)(n-i)}{n^2}\right)^2} = \sum_{i=1}^{n} \frac{n^2 - i(n+1-i)}{n^2} \cdot \frac{(n^2)^2}{\big(i(n+1-i)\big)^2}$$

$$= n^2 \sum_{i=1}^{n} \frac{n^2 - i(n+1-i)}{i^2(n+1-i)^2} = \left[n^4 \sum_{i=1}^{n} \frac{1}{i^2(n+1-i)^2} - n^2 \sum_{i=1}^{n} \frac{1}{i(n+1-i)} \right].$$

Notice that second term is already calculated before, it is $\mathbb{E}T$, so $n^2 \sum_{i=1}^{n} \frac{1}{i(n+1-i)} \approx 2n \log n$.
First term we can estimate in the following way:

$$\frac{1}{i^2(n+1-i)^2} = \frac{1}{(n+1-i)^2 + i^2} \left(\frac{1}{i^2} + \frac{1}{(n+1-i)^2} \right) \leq \frac{1}{n^2} \left(\frac{1}{i^2} + \frac{1}{(n+1-i)^2} \right).$$

Thus:

$$n^4 \sum_{i=1}^{n} \frac{1}{i^2(n+1-i)^2} \leq n^4 \sum_{i=1}^{n} \frac{1}{n^2} \left(\frac{1}{i^2} + \frac{1}{(n+1-i)^2} \right) = n^2 \left(\sum_{i=1}^{n} \frac{1}{i^2} + \sum_{i=1}^{n} \frac{1}{i^2} \right)$$

$$\leq 2n^2 \sum_{i=1}^{\infty} \frac{1}{i^2} = 2n^2 \frac{\pi^2}{6} = \frac{\pi^2}{3} n^2.$$

So we have:

$$VarT \approx n^2.$$

From Chebyshev's inequality we obtain

$$P(T - \mathbb{E}T > c \cdot \sqrt{VarT}) = P(T > 2n \log n + cn) \leq \tfrac{1}{c^2}.$$

Using Lemma 4.1.5 with $k = 2n \log n + cn$ finishes the proof:

$$d(\mu \mathbf{P}^k, \pi) \leq (\mu \mathbf{P}^k, \pi) \leq P(T > k) \leq \tfrac{1}{c^2}.$$

$\square$

5 Factorization of passage time distributions

5.1 Factorization in discrete time

Let $\tau_{\mathbf{e}}(B) = \inf\{n : X_n \in B, X_0 = \mathbf{e}\}$ and $\tau_\mu(B) = \sum_{\mathbf{e}\in E} \tau_{\mathbf{e}}(B)\mu(\mathbf{e})$.

We are interested in the total variation distance between distributions: $\mu\mathbf{P}^n$ and π.
As shown in section 3 we have a bound

$$d(\mu\mathbf{P}^n, \pi) \le s(\mu\mathbf{P}^n, \pi) = \max_k \left(1 - \frac{\mu\mathbf{P}^n(\mathbf{e}_k)}{\pi(\mathbf{e}_k)}\right).$$

So if only $\frac{\mu\mathbf{P}^n(\mathbf{e}_i)}{\pi(\mathbf{e}_i)}$ achieves minimum at $\tilde{\mathbf{e}}$ we have a bound on the total variation distance. Another issue is whether we are able to compute $\frac{\mu\mathbf{P}^n(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}$ or not. Some examples will be given.

Definition 5.1.1. *For a chain $\mathbf{X} = \{X_n, n \ge 0\}$ with initial distribution μ a state $\tilde{\mathbf{e}}$ is* **ratio minimal** *for $\mu, \mathbf{X}$*

$$\textit{if} \qquad \forall(\mathbf{e} \in E, n \ge 0) \qquad \frac{\mu\mathbf{P}^n(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})} \le \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}.$$

For a chain $\mathbf{X} = \{X_n, n \ge 0\}$ a state $\tilde{\mathbf{e}}$ is **s-uniform** for $\mu, \mathbf{X}$ if

$$\forall\ (n \ge 1) \quad s(\mu\mathbf{P}^n, \pi) = 1 - \frac{\mu\mathbf{P}^n(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}.$$

Lemma 5.1.2. $\tilde{\mathbf{e}}$ *is ratio minimal for $\mu, \mathbf{X}$ $\iff$ $\tilde{\mathbf{e}}$ is s-uniform for $\mu, \mathbf{X}$.*

Proof.

$\Rightarrow$ Assume $\tilde{\mathbf{e}}$ is ratio minimal for $\mu, \mathbf{X}$. Then

$$s(\mu\mathbf{P}^n, \pi) = \max_{\mathbf{e}} \left(1 - \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}\right) = 1 - \min_{\mathbf{e}} \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})} = 1 - \frac{\mu\mathbf{P}^n(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}.$$

$\Leftarrow$ Assume $\tilde{\mathbf{e}}$ is s-uniform for $\mu, \mathbf{X}$. Then $\forall(n \ge 1, \mathbf{e} \in E)$

$$s(\mu\mathbf{P}^n, \pi) = 1 - \frac{\mu\mathbf{P}^n(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})} = \max_{\mathbf{e}} \left(1 - \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}\right) \ge \left(1 - \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}\right)$$

i.e.

$$\frac{\mu\mathbf{P}^n(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})} \le \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}$$

□

Lemma 5.1.3. *Let $\mathbf{X} = \{X_n, n \ge 0\}$ be a Markov chain with initial distribution μ. If state $\tilde{\mathbf{e}}$ is ratio minimal for $\mathbf{X}, \mu$ then*

$$\mu\mathbf{P}^{n_1}(\tilde{\mathbf{e}}) \le \mu\mathbf{P}^{n_2}(\tilde{\mathbf{e}}) \text{ for any } n_1 < n_2$$

i.e. $\mu\mathbf{P}^n(\tilde{\mathbf{e}})$ is non-decreasing in n.

Proof.

$$\mu\mathbf{P}^{n_1}(\tilde{\mathbf{e}}) = \frac{\mu\mathbf{P}^{n_1}(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}\pi(\tilde{\mathbf{e}}) \stackrel{*}{=} \frac{\mu\mathbf{P}^{n_1}(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}\sum_i \pi(\mathbf{e}_i)\delta_{\mathbf{e}_i}\mathbf{P}^{n_2-n_1}(\tilde{\mathbf{e}}) = \sum_i \frac{\mu\mathbf{P}^{n_1}(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}\pi(\mathbf{e}_i)\delta_{\mathbf{e}_i}\mathbf{P}^{n_2-n_1}(\tilde{\mathbf{e}})$$

$$\stackrel{**}{\leq} \sum_i \frac{\mu\mathbf{P}^{n_1}(\mathbf{e}_i)}{\pi(\mathbf{e}_i)}\pi(\mathbf{e}_i)\delta_{\mathbf{e}_i}\mathbf{P}^{n_2-n_1}(\tilde{\mathbf{e}}) = \sum_i \mu\mathbf{P}^{n_1}(\mathbf{e}_i)\delta_{\mathbf{e}_i}\mathbf{P}^{n_2-n_1}(\tilde{\mathbf{e}}) = \mu\mathbf{P}^{n_2}(\tilde{\mathbf{e}}),$$

where

$*$ $\sum_i \pi(\mathbf{e}_i)\delta_{\mathbf{e}_i}\mathbf{P}^{n_2-n_1}(\tilde{\mathbf{e}})$: start with stationary distribution π, then probability of getting to $\tilde{\mathbf{e}}$ after time $n_2 - n_1$ is (further stationary) $= \pi(\tilde{\mathbf{e}})$,

$**$ - from assumption.

Thus

$$\mu\mathbf{P}^{n_1}(\tilde{\mathbf{e}}) \leq \mu\mathbf{P}^{n_2}(\tilde{\mathbf{e}}) \quad \text{for} \quad n_1 < n_2.$$

□

We will need the following

Lemma 5.1.4. *Let Y be non-negative random variable. Then:*
$\phi_Y(s) = (1-e^{-s})\sum_{n=0}^{\infty} e^{-sn}P(Y \leq n)$

Proof.

$$(1-e^{-s})\sum_{n=0}^{\infty} e^{-sn}P(Y \leq n) = (1-e^{-s})\sum_{n=0}^{\infty} e^{-sn}\left(P(Y=0)+P(Y=1)+\ldots+P(Y=n)\right) =$$

$$(1-e^{-s})\big[e^{-s0}P(Y=0) + e^{-s\cdot 1}(P(Y=0)+P(Y=1))e^{-s\cdot 2}(P(Y=0)+P(Y=1)+P(Y=2))$$
$$+\ldots + e^{-s\cdot k}(P(Y=0)+\ldots+P(Y=k)) + \ldots\big] =$$

$$= (1-e^{-s})\big[P(Y=0)(e^{-s\cdot 0}+e^{-s\cdot 1}+\ldots)+P(Y=1)(e^{-s\cdot 1}+e^{-s\cdot 2}+\ldots)+\ldots+P(Y=k)(e^{-s\cdot k}+e^{-s\cdot(k+1)}+\ldots)\big]$$

$$= (1-e^{-s})\big[P(Y=0)\left(\frac{e^{-s\cdot 0}}{1-e^{-s}}\right) + P(Y=1)\left(\frac{e^{-s\cdot 1}}{1-e^{-s}}\right) + \ldots + P(Y=k)\left(\frac{e^{-s\cdot k}}{1-e^{-s}}\right) + \ldots =$$

$$= e^{-s\cdot 0}P(Y=0) + e^{-s\cdot 1}P(Y=1) + \ldots + e^{-s\cdot k}P(Y=k) + \ldots = \phi_Y(s)$$

□

The next lemma states, that in the case of $\mu\mathbf{P}^n(\mathbf{e})$ being increasing in n, the passage time to $\mathbf{e} \in E$ starting with a distribution μ can be decomposed into first passage time to stationarity, i.e. to distribution π starting from μ plus passage time to $\mathbf{e}$ starting from distribution π. We can interpret it as if $\tau_\mu(\mathbf{e})$ had the same distribution as the sum of $Y_\mathbf{e}$ and $\tau_\pi(\mathbf{e})$, i.e.: $\tau_\mu(\mathbf{e}) \stackrel{d}{=} Y_\mathbf{e} + \tau_\pi(\mathbf{e})$, where $Y_\mathbf{e}$ is a random variable with distribution function $F_{Y_\mathbf{e}}(n) = P(Y_\mathbf{e} \leq n) = \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}$ (or equivalently, density of $Y_\mathbf{e}$ is given by $f_{Y_\mathbf{e}}(n) = \frac{\mu\mathbf{P}^n(\mathbf{e})-\mu\mathbf{P}^{n-1}(\mathbf{e})}{\pi(\mathbf{e})}$).

Lemma 5.1.5 (Brown [5])**.** *Let $\mathbf{X} = \{X_n, n \geq 0\}$ be ergodic Markov chain with enumerable state space $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots\}$ and invariant measure π. If for initial distribution μ and $\mathbf{e} \in E$ we have that $\mu\mathbf{P}^n(\mathbf{e})$ is non-decreasing in n then*

$$\psi_{\tau_\mu(\mathbf{e})} = \psi_{Y_\mathbf{e}} \cdot \psi_{\tau_\pi(\mathbf{e})},$$

where $Y_\mathbf{e}$ is independent of $\mathbf{X}$ with distribution function $F_{Y_\mathbf{e}}(n) = P(Y_\mathbf{e} \leq n) = \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}$ or, equivalently, with density $f_{Y_\mathbf{e}}(n) = \frac{\mu\mathbf{P}^n(\mathbf{e})-\mu\mathbf{P}^{n-1}(\mathbf{e})}{\pi(\mathbf{e})}$.

Proof. We have:

$$\pi(\mathbf{e}) = P_\pi(X_n = \mathbf{e}) = \sum_{x=0}^{n} \delta_\mathbf{e}\mathbf{P}^{n-x}(\mathbf{e})P(\tau_\pi(\mathbf{e}) = x). \tag{5.75}$$

Multiplying both sides by e^{-sn} and summing up over all n we have (left side is independent of n):

$$\frac{\pi(\mathbf{e})}{1-e^{-s}} = \pi(\mathbf{e})\sum_{n=0}^{\infty} e^{-sn} = \sum_{n=0}^{\infty} e^{-sn}\sum_{x=0}^{n}\delta_\mathbf{e}\mathbf{P}^{n-x}(\mathbf{e})P(\tau_\pi(\mathbf{e}) = x)$$

$$\sum_{x=0}^{\infty}\sum_{n=x}^{\infty} e^{-sn}\delta_\mathbf{e}\mathbf{P}^{n-x}(\mathbf{e})P(\tau_\pi(\mathbf{e}) = x) = \sum_{x=0}^{\infty} e^{-sx}P(\tau_\pi(\mathbf{e}) = x)\sum_{n=0}^{\infty} e^{-sn}\delta_\mathbf{e}\mathbf{P}^{n}(\mathbf{e}) = \psi_{\delta_\mathbf{e}\mathbf{P}^n(\mathbf{e})}(s)\psi_{\tau_\pi(\mathbf{e})}(s).$$

Thus:

$$\psi_{\delta_\mathbf{e}\mathbf{P}^n(\mathbf{e})}(s) = \frac{\pi(\mathbf{e})}{(1-e^{-s})\psi_{\tau_\pi(\mathbf{e})}(s)}. \tag{5.76}$$

The assumption is that $\mu\mathbf{P}^n(\mathbf{e})$ is non-decreasing in n and of course $\mu\mathbf{P}^n(\mathbf{e}) \to \pi(\mathbf{e})$ as $n \to \infty$, so $\frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}$ is a distribution function of some random variable $Y_\mathbf{e}$. We can calculate its Laplace transform using Lemma 5.1.4:

$$\psi_{Y_\mathbf{e}}(s) = (1-e^{-s})\sum_{n=0}^{\infty} e^{-sn}P(Y_\mathbf{e} \le n) = (1-e^{-s})\sum_{n=0}^{\infty} e^{-sn}\frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})} = \frac{(1-e^{-s})}{\pi(\mathbf{e})}\psi_{\mu\mathbf{P}^n(\mathbf{e})}(s).$$

Thus:

$$\psi_{\mu\mathbf{P}^n(\mathbf{e})}(s) = \frac{\pi(\mathbf{e})}{(1-e^{-s})}\psi_{Y_\mathbf{e}}(s). \tag{5.77}$$

Similarly to (5.75) expanding $\mu\mathbf{P}^n(\mathbf{e})$ we obtain:

$$\mu\mathbf{P}^n(\mathbf{e}) = P_\mu(X_n = \mathbf{e}) = \sum_{x=0}^{n}\delta_\mathbf{e}\mathbf{P}^{n-x}(\mathbf{e})P(\tau_\mu(\mathbf{e}) = x)$$

and thus:

$$\psi_{\mu\mathbf{P}^n(\mathbf{e})}(s) = \psi_{\delta_\mathbf{e}P^n(\mathbf{e})}(s)\psi_{\tau_\mu(\mathbf{e})}(s). \tag{5.78}$$

Finally from (5.76), (5.77) and (5.78) we have:

$$\psi_{\tau_\mu(\mathbf{e})}(s) = \frac{\psi_{\mu\mathbf{P}^n(\mathbf{e})}(s)}{\psi_{\delta_\mathbf{e}\mathbf{P}^n(\mathbf{e})}(s)} = \frac{\pi(\mathbf{e})}{(1-e^{-s})}\psi_{Y_\mathbf{e}}(s)\cdot\frac{(1-e^{-s})}{\pi(\mathbf{e})}\psi_{\tau_\pi(\mathbf{e})}(s) = \psi_{Y_\mathbf{e}}(s)\cdot\psi_{\tau_\pi(\mathbf{e})}(s).$$

□

Next theorem is the extension of Theorem 5.1.5 in case when stationary distribution π is uniform. Then we can have factorization not only for a single state, but for set of states.

Let U denotes uniform distribution and $U_{|B}$ uniform distribution narrowed to $B \subset E$.

Lemma 5.1.6. *Let $\mathbf{X} = \{X_n, n \ge 0\}$ be ergodic Markov chain with finite state space $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots\}$ and uniform stationary distribution U. If for initial distribution μ and $B \subset E$ we have that $\mu\mathbf{P}^n(B)$ is non-decreasing in n then*

$$\psi_{\tau_\mu(B)} = \psi_{Y_B}\cdot\psi_{\tau_U(B)},$$

where Y_B is independent of $\mathbf{X}$ with distribution function $F_{Y_B}(n) = P(Y_B \le n) = \frac{\mu\mathbf{P}^n(B)}{U(B)}$ or, equivalently, with density $f_{Y_B}(n) = \frac{\mu\mathbf{P}^n(B) - \mu\mathbf{P}^{n-1}(B)}{U(B)}$.

Proof. We have:

$$U(B) = P_U(X_n \in B) = \sum_{x=0}^{n} U_{|_B}\mathbf{P}^{n-x}(B)P(\tau_U(B) = x) \tag{5.79}$$

Multiplying both sides by e^{-sn} and summing up over all n we have (left side is independent of n):

$$\frac{U(B)}{1-e^{-s}} = U(B)\sum_{n=0}^{\infty} e^{-sn} = \sum_{n=0}^{\infty} e^{-sn}\sum_{x=0}^{n} U_{|_B}\mathbf{P}^{n-x}(B)P(\tau_U(B) = x)$$

$$\sum_{x=0}^{\infty}\sum_{n=x}^{\infty} e^{-sn}U_{|_B}\mathbf{P}^{n-x}(\mathbf{e})P(\tau_U(\mathbf{e}) = x) = \sum_{x=0}^{\infty} e^{-sx}P(\tau_U(B) = x)\sum_{n=0}^{\infty} e^{-sn}U_{|_B}\mathbf{P}^{n}(B) = \psi_{U_{|_B}\mathbf{P}^n(B)}(s)\psi_{\tau_U(B)}(s).$$

Thus:

$$\psi_{U_{|_B}\mathbf{P}^n(B)}(s) = \frac{\pi(B)}{(1-e^{-s})\psi_{\tau_U(B)}(s)}. \tag{5.80}$$

Calculate Laplace transform of Y_B using Lemma 5.1.4:

$$\psi_{Y_B}(s) = (1-e^{-s})\sum_{n=0}^{\infty} e^{-sn}P(Y_B \le n) = (1-e^{-s})\sum_{n=0}^{\infty} e^{-sn}\frac{\mu\mathbf{P}^n(B)}{U(\mathbf{e})} = \frac{(1-e^{-s})}{U(B)}\psi_{\mu\mathbf{P}^n(B)}(s).$$

Thus:

$$\psi_{\mu\mathbf{P}^n(B)}(s) = \frac{U(B)}{(1-e^{-s})}\psi_{Y_B}(s). \tag{5.81}$$

Similarly to (5.79) expanding $\mu\mathbf{P}^n B)$ we obtain:

$$\mu\mathbf{P}^n(B) = P_\mu(X_n \in B) = \sum_{x=0}^{n} U_{|_B}\mathbf{P}^{n-x}(B)P(\tau_\mu(B) = x)$$

and thus:

$$\psi_{\mu\mathbf{P}^n(B)}(s) = \psi_{U_{|_B}P^n(B)}(s)\psi_{\tau_\mu(B)}(s). \tag{5.82}$$

Finally from (5.80), (5.81) and (5.82) we have:

$$\psi_{\tau_\mu(B)}(s) = \frac{\psi_{\mu\mathbf{P}^n(B)}(s)}{\psi_{U_{|_B}\mathbf{P}^n(B)}(s)} = \frac{U(B)}{(1-e^{-s})}\psi_{Y_B}(s)\cdot\frac{(1-e^{-s})}{U(B)}\psi_{\tau_U(B)}(s) = \psi_{Y_B}(s)\cdot\psi_{\tau_U(B)}(s).$$

□

Theorem 5.1.7. *Let $\mathbf{X} = \{X_n, n \ge 0\}$ be a Markov chain with stationary distribution π. If $\tilde{\mathbf{e}}$ is ratio minimal for $\mu, \mathbf{X}$ then*

(i) $\psi_{\tau_\mu(\tilde{\mathbf{e}})} = \psi_{Y_{\tilde{\mathbf{e}}}}\cdot\psi_{\tau_\pi(\tilde{\mathbf{e}})}$ $\quad(\tau_\mu(\tilde{\mathbf{e}}) \overset{d}{=} Y_{\tilde{\mathbf{e}}} + \tau_\pi(\tilde{\mathbf{e}}))$,

(ii) there exists Minimal Strong Stationary Time T_μ such that

$$s(\mu\mathbf{P}^n, \pi) = 1 - \frac{\mu\mathbf{P}^n(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})} = P(T_\mu > n) \quad \forall(n \ge 0).$$

where $Y_{\tilde{\mathbf{e}}}$ is as in Lemma 5.1.5.

Proof.

(i) From Lemma 5.1.3 we have that $\mu\mathbf{P}^n(\tilde{\mathbf{e}})$ is non-increasing in n and this is the assumption of Lemma 5.1.5 which implies $\psi_{\tau_\mu(\tilde{\mathbf{e}})} = \psi_{Y_{\tilde{\mathbf{e}}}}\cdot\psi_{\tau_\pi(\tilde{\mathbf{e}})}$.

(ii) In Lemma 4.1.8 we had that there always existed $\mu-$Minimal Strong Stationary Time. In the proof we had: $\alpha_n = \min_{\mathbf{e}} \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}$. Because $\tilde{\mathbf{e}}$ is ratio minimal, thus we have $\alpha_n = \frac{\mu\mathbf{P}^n(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}$. It was constructed in such way that

$$P(T_\mu = n) = \alpha_n - \alpha_{n-1}.$$

Since it is $\mu-$Minimal Strong Stationary Time, we have $P(T_\mu > n) = s(\mu\mathbf{P}^n, \pi) = 1 - \frac{\mu\mathbf{P}^n(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}$.

□

Next lemma says us something about the structure of a chain in case of $\delta_{\mathbf{e}}\mathbf{P}^n(\mathbf{e})$ being non-increasing in n. Then the first passage time to $\mathbf{e}$ starting with stationary distribution π is a geometric compound: $\tau_\pi(\mathbf{e}) \stackrel{d}{=} \sum_{i=1}^{N} W_{\mathbf{e},i}$, where N has geometric distribution: $P(N = k) = (1-\pi(\mathbf{e}))^k\pi(\mathbf{e}), k = 0, 1, \ldots$ and $W_{\mathbf{e},i}, i = 1, 2, \ldots$ are i.i.d. random variables with the same distribution as $W_{\mathbf{e}}$: $P(W_{\mathbf{e}} \le n) = \frac{1-\delta_{\mathbf{e}}\mathbf{P}^n(\mathbf{e})}{1-\pi(\mathbf{e})}$ or equivalently with density function $P(W_{\mathbf{e}} = n) = \frac{\delta_{\mathbf{e}}\mathbf{P}^{n-1}(\mathbf{e}) - \delta_{\mathbf{e}}\mathbf{P}^n(\mathbf{e})}{1-\pi(\mathbf{e})}$.

Lemma 5.1.8 (Brown [5])**.** *Let $\mathbf{X} = \{X_n, n \ge 0\}$ be ergodic Markov chain with enumerable state space $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots\}$ and invariant measure π. If for a state $\mathbf{e}$ we have that $\delta_{\mathbf{e}}\mathbf{P}^n(\mathbf{e})$ is non-increasing in n then*

$$\psi_{\tau_\pi(\mathbf{e})} = \frac{\pi(\mathbf{e})}{1-(1-\pi(\mathbf{e}))\psi_{W_{\mathbf{e}}}},$$

where $W_{\mathbf{e}}$ is independent of $\mathbf{X}$ and $P(W_{\mathbf{e}} > n) = \frac{\delta_{\mathbf{e}}\mathbf{P}^n(\mathbf{e}) - \pi(\mathbf{e})}{1-\pi(\mathbf{e})}$.

Proof. Using Lemma 5.1.4 ($W_{\mathbf{e}}$ is a non-negative random variable):

$$\psi_{W_{\mathbf{e}}}(s) = (1-e^{-s})\sum_{n=0}^{\infty} e^{-sn} P(W_{\mathbf{e}} \le n) = (1-e^{-s})\sum_{n=0}^{\infty} e^{-sn}(1 - P(W_{\mathbf{e}} > n)) =$$

$$(1-e^{-s})\frac{1}{1-e^{-s}} - (1-e^{-s})\sum_{n=0}^{\infty} e^{-sn} P(W_{\mathbf{e}} > n) = 1 - (1-e^{-s})\sum_{n=0}^{\infty} e^{-sn} P(W_{\mathbf{e}} > n). \tag{5.83}$$

We have

$$\sum_{n=0}^{\infty} e^{-sn} P(W_B > n) = \sum_{n=0}^{\infty} e^{-sn} \frac{\delta_B\mathbf{P}^n(\mathbf{e}) - \pi(\mathbf{e})}{1-\pi(\mathbf{e})} = \frac{1}{1-\pi(\mathbf{e})}\sum_{n=0}^{\infty} e^{-sn}\delta_B\mathbf{P}^n(\mathbf{e}) - \frac{\pi(\mathbf{e})}{1-\pi(\mathbf{e})}\sum_{n=0}^{\infty} e^{-sn} =$$

$$\frac{1}{1-\pi(B)} \cdot \psi_{\delta_B\mathbf{P}^n(\mathbf{e})}(s) - \frac{\pi(\mathbf{e})}{1-\pi(\mathbf{e})} \cdot \frac{1}{1-e^{-s}}.$$

Putting it again to (5.83) we obtain

$$\psi_{W_{\mathbf{e}}}(s) = 1 - (1-e^{-s})\left(\frac{1}{1-\pi(B)} \cdot \psi_{\delta_B\mathbf{P}^n(\mathbf{e})}(s) - \frac{\pi(\mathbf{e})}{1-\pi(\mathbf{e})} \cdot \frac{1}{1-e^{-s}}\right),$$

and from here:

$$\psi_{\delta_{\mathbf{e}}\mathbf{P}^n(\mathbf{e})}(s) = \frac{1-\psi_{W_{\mathbf{e}}}(1-\pi(\mathbf{e}))}{1-e^{-s}}.$$

Putting it to (5.76) $\left(\psi_{\tau_\pi(B)}(s) = \frac{\pi(\mathbf{e})}{(1-e^{-s})\psi_{\delta_B\mathbf{P}^n(\mathbf{e})}(s)}\right)$ we have

$$\psi_{\tau_\pi(\mathbf{e})}(s) = \frac{\pi(\mathbf{e})}{1-(1-\pi(\mathbf{e}))\psi_{W_{\mathbf{e}}}(s)}.$$

□

5.1.1 Special cases

The clue is to find out when $\tilde{\mathbf{e}}$ ratio minimality holds.
Notice that it has nothing to do with the ordering of the space. Nevertheless, in some cases when we define a partial ordering on E it is possible to find μ and $\tilde{\mathbf{e}}$ such that we have $\tilde{\mathbf{e}}$ ratio minimality for $\mu, \mathbf{X}$. Introduce the following assumptions :

A0: Markov chain $\mathbf{X} = \{X_n, n \geq 0\}$ is ergodic, $E = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ is finite and partially ordered with ordering $\prec$

A1: there exists maximal state $\mathbf{e}_M$ such that $\forall(\mathbf{e}_i \in E) \quad \mathbf{e}_i \prec \mathbf{e}_M$

A2: $\frac{\mu(\mathbf{e}_j)}{\pi(\mathbf{e}_j)} \searrow_j$ (relative to $\prec$) $\quad \left(\text{i.e. } \mathbf{e}_i \prec \mathbf{e}_j \quad \Longrightarrow \quad \frac{\mu(\mathbf{e}_i)}{\pi(\mathbf{e}_i)} \geq \frac{\mu(\mathbf{e}_j)}{\pi(\mathbf{e}_j)}\right)$

$A1$ says that there always exists maximal state comparable to all the others, and $A2$ that μ is smaller then π in some stochastic sense (likelihood ratio ordering).

• **Stochastic monotonicity of time-reversed process $\tilde{\mathbf{X}}$**

Recall the definition of stochastically monotone Markov chain.

Set A is an upper set (denoted by $A\uparrow$), if ($x \in A$ and $x \prec y$) $\Rightarrow y \in A$. Markov chain with transition matrix $\mathbf{P}$ is **stochastically monotone** if

$$\forall(A\uparrow)\forall(\mathbf{e}_1 \prec \mathbf{e}_2) \qquad \delta_{\mathbf{e}_1}\mathbf{P}(A) = \sum_{\mathbf{e}\in A} \delta_{\mathbf{e}_1} P(\mathbf{e}) \leq \delta_{\mathbf{e}_2}\mathbf{P}(A). \tag{5.84}$$

Time-reversed process $\{\tilde{\mathbf{X}}\}$ is the one with transition probabilities $\delta_{\mathbf{e}_2}\tilde{\mathbf{P}}(\mathbf{e}_1) = \frac{\pi(\mathbf{e}_1)}{\pi(\mathbf{e}_2)}\delta_{\mathbf{e}_1}\mathbf{P}(\mathbf{e}_2)$ (and process is **time-reversible** if $\tilde{\mathbf{P}} = \mathbf{P}$).

Fact 5.1.9 (Stoyan & Mueller [44]). *Let h be decreasing relative to $\prec$ and $\mathbf{P}$ stochastically monotone. Then*

$$\forall(n \geq 0)\ \forall(\mathbf{e}_i \prec \mathbf{e}_j) \qquad E_{\mathbf{e}_i} h(X_n) \geq E_{\mathbf{e}_j} h(X_n).$$

□

Theorem 5.1.10 (Brown [5]). *Assume that $A0 - A2$ are fulfilled, and in addition that*

- *Time-reversed process $\tilde{\mathbf{X}} = \{\tilde{X}_n, n \geq 0\}$ is stochastically monotone.*

Then

(i) $\quad e_M$ is ratio minimal for $\mu, \mathbf{X}$,

(ii) $\quad \psi_{\tau_\mu(\mathbf{e}_M)} = \psi_{Y_{\mathbf{e}_M}} \cdot \psi_{\tau_\pi(\mathbf{e}_M)} \qquad (\tau_\mu(\mathbf{e}_M) \stackrel{d}{=} Y_{\mathbf{e}_M} + \tau_\pi(\mathbf{e}_M))$,

(iii) $\quad$ there exists Minimal Strong Stationary Time T_μ such that

$$s(\mu\mathbf{P}^n, \pi) = 1 - \frac{\mu\mathbf{P}^n(\mathbf{e}_M)}{\pi(\mathbf{e}_M)} = P(T_\mu > n) \quad \forall(n \geq 0).$$

where $Y_{\mathbf{e}_M}$ is random variable independent of chain $\mathbf{X}$ with distribution $P(Y_{\mathbf{e}_M} \leq n) = \frac{\mu\mathbf{P}^n(\mathbf{e}_M)}{\pi(\mathbf{e}_M)}$.

Proof. Define $h(\mathbf{e}) = \frac{\mu(\mathbf{e})}{\pi(\mathbf{e})}$. For all $\mathbf{e}_i \prec \mathbf{e}_j$ using fact 5.1.9 we have

$$\frac{\mu\mathbf{P}^n(\mathbf{e}_j)}{\pi(\mathbf{e}_j)} = \sum_{\mathbf{e}_i\in E} \frac{\mu(\mathbf{e}_i)}{\pi(\mathbf{e}_j)}\delta_{\mathbf{e}_i}\mathbf{P}^n(\mathbf{e}_j) = \sum_{\mathbf{e}_i\in E} \frac{\mu(\mathbf{e}_i)}{\pi(\mathbf{e}_j)} \cdot \frac{\pi(\mathbf{e}_j)}{\pi(\mathbf{e}_i)}\delta_{\mathbf{e}_j}\tilde{\mathbf{P}}^n(\mathbf{e}_i)$$

$$= \sum_{\mathbf{e}_i\in E} \frac{\mu(\mathbf{e}_i)}{\pi(\mathbf{e}_i)}\delta_{\mathbf{e}_j}\tilde{\mathbf{P}}^n(\mathbf{e}_i) = E_{\mathbf{e}_j} h(\tilde{X}) \leq E_{\mathbf{e}_i} h(\tilde{X}) = \frac{\mu\mathbf{P}^n(\mathbf{e}_i)}{\pi(\mathbf{e}_i)}$$

and especially for all $\mathbf{e}_i$ we have $\mathbf{e}_i \prec \mathbf{e}_M$, thus

$$\frac{\mu\mathbf{P}^n(\mathbf{e}_M)}{\pi(\mathbf{e}_M)} \leq \frac{\mu\mathbf{P}^n(\mathbf{e}_i)}{\pi(\mathbf{e}_i)},$$

i.e. $\mathbf{e}_M$ is ratio minimal for $\mu, \mathbf{X}$. This (by Theorem 5.1.7) implies (ii) and (iii). □

Remark 5.1.11.

Define distance $dist(\mathbf{e}_x, \mathbf{e}_y) = \min_n \delta_{\mathbf{e}_x}\mathbf{P}^n(\mathbf{e}_y) > 0$. If $\mu = \delta_{\mathbf{e}_x}$ (we start from $\{\mathbf{e}_x\}$) than necessary condition on $\mathbf{e}_M$ to be ratio minimal for $\mu, \mathbf{X}$ is: $dist(\mathbf{e}_x, \mathbf{e}_y) \leq dist(\mathbf{e}_x, \mathbf{e}_M)$ for each $\mathbf{e}_y \in E$. It is this way, because if $\mathbf{e}_M$ is ratio minimal and $\delta_{\mathbf{e}_x}\mathbf{P}^n(\mathbf{e}_M) > 0$ then:

$$1 > s(\delta_{\mathbf{e}_x}\mathbf{P}^n, \pi) = 1 - \frac{\delta_{\mathbf{e}_x}\mathbf{P}^k(\mathbf{e}_M)}{\pi(\mathbf{e}_M)} \geq 1 - \frac{\delta_{\mathbf{e}_x}\mathbf{P}^k(\mathbf{e}_y)}{\pi(\mathbf{e}_y)},$$

thus $\delta_{\mathbf{e}_x}\mathbf{P}^k(\mathbf{e}_y) > 0$.
In applications there usually exists one potential state, which can be ratio minimal. The problem then reduces to producing a partial ordering $\prec$ under which $\mathbf{e}_M$ is the unique maximal state, $\mathbf{e}_x$ is a minimal state, and $\tilde{\mathbf{X}}$ is stochastically monotone.

• Failure Rate Monotonicity
Notice that under assumptions A0-A2 we have that $\frac{\mu\mathbf{P}^n(\mathbf{e}_i)}{\pi(\mathbf{e}_i)}$ is non-increasing in $\mathbf{e}_i$ what is stronger than our needs - in Theorem 5.1.7 only fact that $\frac{\mu\mathbf{P}^n(\mathbf{e}_i)}{\pi(\mathbf{e}_i)}$ has minimum in $\mathbf{e}_M$ is needed.
In this subsection we shall find weaker assumptions on transition probability matrix under which there exists some ratio minimal state.

Let assumptions $A0 - A2$ will still be fulfilled, but in this subsection we additionally assume that state space is totally ordered, i.e $E = \{1, 2, \ldots, M\}$. Define A_j to be an upper set of form $A_j = \{j, j+1, \ldots, M\}$. Of course $A_j \subset A_i$ if $i < j$.

Define $\mathbf{P}$ to be **Failure Rate Monotone** (denoted by $\mathbf{P} \in FRM$) for the transition probability matrix of Markov chain with linear ordering if

$$\text{For} \quad \begin{array}{c} i_1 < i_2 \\ j_1 < j_2 \end{array} \quad : \quad \delta_{i_1}\mathbf{P}(A_{j_1})\delta_{i_2}\mathbf{P}(A_{j_2}) \geq \delta_{i_1}\mathbf{P}(A_{j_2})\delta_{i_2}\mathbf{P}(A_{j_1}) \quad \equiv \quad \frac{\delta_{i_1}\mathbf{P}(A_{j_2})}{\delta_{i_1}\mathbf{P}(A_{j_1})} \leq \frac{\delta_{i_2}\mathbf{P}(A_{j_2})}{\delta_{i_2}\mathbf{P}(A_{j_1})}.$$

Theorem 5.1.12 (Brown [5]). *Let assumptions $A0 - A2$ will be fulfilled, and in addition*

- $\mathbf{P} \in FRM$

Then

(i) e_M is ratio minimal for $\mu, \mathbf{X}$,

(ii) $\psi_{\tau_\mu(\mathbf{e}_M)} = \psi_{Y_{\mathbf{e}_M}} \cdot \psi_{\tau_\pi(\mathbf{e}_M)} \qquad (\tau_\mu(\mathbf{e}_M) \stackrel{d}{=} Y_{\mathbf{e}_M} + \tau_\pi(\mathbf{e}_M))$,

(iii) there exists Minimal Strong Stationary Time T_μ such that

$$s(\mu\mathbf{P}^n, \pi) = 1 - \frac{\mu\mathbf{P}^n(\mathbf{e}_M)}{\pi(\mathbf{e}_M)} = P(T_\mu > n), \quad \forall(n \geq 0).$$

where $Y_{\mathbf{e}_M}$ is random variable independent of chain $\mathbf{X}$ with distribution $P(Y_{\mathbf{e}_M} \leq n) = \frac{\mu\mathbf{P}^n(\mathbf{e}_M)}{\pi(\mathbf{e}_M)}$.

It is enough to prove (i) which implies (by Theorem 5.1.7) (ii) and (iii). The proof can be found in Brown [5].

□

• Non-negative spectrum of P
Lemma 5.1.8 gives us some nice factorization in case when $\delta_{\mathbf{e}}\mathbf{P}^n(\mathbf{e})$ is non-increasing in n. We find some conditions which guarantee it.

Lemma 5.1.13. *Let $\{X_n, n \geq 0\}$ be a Markov chain on a finite state space $|E| = N$ with all eigenvalues real and non-negative. Then*

$$\forall(\mathbf{e} \in E), \forall(n_1 < n_2) \qquad \delta_{\mathbf{e}}\mathbf{P}^{n_1}(\mathbf{e}) \geq \delta_{\mathbf{e}}\mathbf{P}^{n_2}(\mathbf{e})$$

i.e. $\delta_{\mathbf{e}}\mathbf{P}^n(\mathbf{e})$ is non-increasing in n for every $\mathbf{e} \in E$.

Proof. From spectral representation (3.39) we have

$$\delta_{\mathbf{e}}\mathbf{P}^n(\mathbf{e}) = \sum_{i=1}^{N} \lambda_i^n \mathbf{B}_i(\mathbf{e},\mathbf{e}) = \sum_{i=1}^{N} \lambda_i^n \left(\pi(\mathbf{e})\right)^2.$$

From our assumptions for all $i = 1, \ldots, N$, we have $\lambda_i \in [0,1]$, thus $\delta_{\mathbf{e}}\mathbf{P}^n(\mathbf{e})$ in non-increasing in n. □

- **Markov chain $\{X_{2n}, n \geq 0\}$ governed by $\mathbf{P}^2$, for reversible $\mathbf{P}$.**

Lemma 5.1.14. *Let $\{X_n\}$ be a Markov chain on finite state space $|E| = N$ with transition matrix $\mathbf{P}$. Assume it is time-reversible. Let $\{Y_n\}$ be a Markov chain with transition matrix $\mathbf{R} = \mathbf{P}^2$ (i.e. we observe every second step of $\{X_n\}$). Then*

$$\forall(\mathbf{e} \in E), \forall(n_1 < n_2) \qquad \delta_{\mathbf{e}}\mathbf{R}^{n_1}(\mathbf{e}) \leq \delta_{\mathbf{e}}\mathbf{R}^{n_2}(\mathbf{e})$$

i.e. $\delta_{\mathbf{e}}\mathbf{R}^n(\mathbf{e})$ is non-increasing in n for every $\mathbf{e} \in E$.

Proof. It is known that for time-reversible Markov chain all eigenvalues are real and $\lambda_i \in [-1,1]$ for $i = 1, \ldots, N$. We have

$$\delta_{\mathbf{e}}\mathbf{R}^n(\mathbf{e}) = P(Y_n = \mathbf{e}|Y_0 = \mathbf{e}) = P(X_{2n} = \mathbf{e}|Y_0 = \mathbf{e}) \stackrel{*}{=} \sum_{i=1}^{N} \lambda_i^{2n} \left(\pi(\mathbf{e})\right)^2,$$

where $*$ is from representation (3.39). We see that all eigenvalues of matrix $\mathbf{R}$ are non-negative, thus assumption of Lemma 5.1.13 is fulfilled.

□

Remark In continuous time we always have that $\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})$ is non-decreasing in t, see for example Keilson [27].

5.2 Continuous time Markov chains

A stochastic process $\mathbf{X} = (X_t, t \geq 0)$ is called **Markov Process** if for any $t, s \geq 0$ and $\mathbf{e} \in E$ we have $P(X_{t+s} = \mathbf{e} \mid X_u; u \leq t) = P(X_{t+s} = \mathbf{e} \mid X_t)$. Denote its state space by E and its stationary distribution by π. We assume E is finite or enumerable and its elements will be denoted $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots\}$, and in the case when we have linear ordering: $E = \{1, 2, \ldots\}$.
For fixed $\mathbf{e}, \mathbf{e}'$ function $t \to \delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e}')$ is called a **transition function**. Also transition function is called the whole family of matrices $\mathbf{P}_t, t \geq 0$.
We use the following notation similarly to discrete case

$\delta_{\mathbf{e}}\mathbf{P}_t(B)$	$=$	$P_{\mathbf{e}}(X_t \in B) = P(X_t \in B \| X_0 = \mathbf{e})$	–	Probability of hitting B at time t starting from $X_0 = \mathbf{e}$
$\tau_{\mathbf{e}}(B)$	$=$	$\inf\{t : X_t \in B, X_0 = \mathbf{e}\}$	–	First passage time to B starting from $X_0 = \mathbf{e}$
$\tau_\mu(B)$	$=$	$\int_E \tau_{\mathbf{e}}(B) d\mu(\mathbf{e})$	–	First passage time to B starting with $X_0 \sim \mu$
$\mu\mathbf{P}_t(B)$	$=$	$P_\mu(X_t \in B)$	–	First passage time to B starting with $X_0 \sim \mu$

If set B will be one state set, i.e. $B = \{\mathbf{e}\}$, it will be shortly written $B := \mathbf{e}$.

We we denote the Laplace transform of continuous, real-valued random variable Y by

$$\psi_Y(s) = \int_E e^{-st} dF_Y(t).$$

We assume that for every $\mathbf{e} \neq \mathbf{e}'$ there exist limits

$$q(\mathbf{e}, \mathbf{e}') = \lim_{t \to 0} \frac{\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e}')}{t}.$$

We define $q(\mathbf{e}, \mathbf{e}) = -\sum_{\mathbf{e}' \neq \mathbf{e}} q(\mathbf{e}, \mathbf{e}')$. Matrix $\mathbf{Q} = [q(\mathbf{e}_1, \mathbf{e}_2)]_{\mathbf{e}_1, \mathbf{e}_2 \in E}$ is called **intensity matrix** (it can be written $\mathbf{Q} = \lim_{t \to 0} \frac{\mathbf{P}_t - \mathbf{I}}{t}$) We also assume, that it is **conservative**, i.e. $-q(\mathbf{e}, \mathbf{e}) < \infty$ $\forall(\mathbf{e} \in E)$. From Chapman-Kolmogorov equations we have:

$$\frac{d}{dt}\mathbf{P}_t = \mathbf{P}_t\mathbf{Q} = \mathbf{Q}\mathbf{P}_t.$$

Assuming that $\mathbf{P}_0 = \mathbf{I}$ (i.e. $\lim_{t\to 0} \delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e}') = 1$ if $\mathbf{e} = \mathbf{e}'2$ and 0 otherwise) the above equation has solution:

$$\mathbf{P}_t = e^{\mathbf{Q}t},$$

where $e^{\mathbf{A}}$ means matrix exponential: $e^{\mathbf{A}} = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!}$.

We define **separation distance** as follows:

$$s(\mu,\nu) = \max_{\mathbf{e}\in E}\left(1 - \frac{\mu(\mathbf{e})}{\nu(\mathbf{e})}\right).$$

We will be interested in $s(\mu\mathbf{P}_t, \pi)$. Similarly as for discrete time, s is an upper bound for total variation distance:

$$d(\mu\mathbf{P}_t,\pi) \le s(\mu\mathbf{P}_t,\pi).$$

Markov process $\mathbf{X} = \{X_t, t\ge 0\}_{t\ge 0}$ with finite or enumerable state space E is **stochastically monotone** if

$$\forall(t) \quad \forall(\mu \prec_{st} \nu) \qquad \mu\mathbf{P}_t \prec_{st} \nu\mathbf{P}_t \qquad \Big(\equiv \qquad \forall(f \text{ increasing}) \quad \mathbf{P}_t f \text{ jest increasing}\Big). \tag{5.85}$$

5.2.1 Uniformization

Let $\mathbf{X} = (X_t, t \ge 0)$ be a Markov chain with enumerable state space E with intensity matrix $\mathbf{Q} = [q(\mathbf{e},\mathbf{e}')]$ ($q(\mathbf{e},\mathbf{e}) = -\sum_{\mathbf{e}'\neq\mathbf{e}} q(\mathbf{e},\mathbf{e}')$). Denote $q_{\mathbf{e}} = -q(\mathbf{e},\mathbf{e})$.
The process evolves in the following way: with some initial distribution μ starting state $\mathbf{e}$ is chosen, later the process stays in this state for random time T_1 - exponentially distributed $Exp(q(\mathbf{e}))$, after that it jumps to state $\mathbf{e}'$ with probability $\frac{q(\mathbf{e},\mathbf{e}')}{q(\mathbf{e})}$, where it stays exponential time $Exp(q(\mathbf{e}'))$, and so on.
If all $q(\mathbf{e})$ are bounded, i.e. $q(\mathbf{e}) \le c < \infty$ $\forall\mathbf{e}\in E$, then the process is called **uniformizable**. (Process is always uniformizable in the case of finite state space ($|E| < \infty$)).

We assume as former that there exist $\frac{d}{dt}\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e}')\Big|_{t=0+} = q(\mathbf{e},\mathbf{e}')$ and recall that for matrix $\mathbf{A}$ we defined: $e^{\mathbf{A}} = \sum_{k=0}^{\infty}\frac{\mathbf{A}^k}{k!}$.

From Chapman-Kolmogorov we have:

$$\mu\mathbf{P}_{t+h}(\mathbf{e}) = \sum_{\mathbf{e}'\in E}\mu\mathbf{P}_t(\mathbf{e}')\delta_{\mathbf{e}'}\mathbf{P}_h(\mathbf{e}),$$

$$\mu\mathbf{P}_{t+h}(\mathbf{e}) - \mu\mathbf{P}_t(\mathbf{e}) = \sum_{\mathbf{e}'\in E}\mu\mathbf{P}_t(\mathbf{e}')\delta_{\mathbf{e}'}\mathbf{P}_h(\mathbf{e}) - \mu\mathbf{P}_t(\mathbf{e}).$$

Dividing both sides by h and making $h \to 0$

$$\frac{d}{dt}\mu\mathbf{P}_t(\mathbf{e}) = \sum_{\mathbf{e}'\in E}\mu\mathbf{P}_t(\mathbf{e}')q(\mathbf{e}',\mathbf{e}) = -q(\mathbf{e})\mu\mathbf{P}_t(\mathbf{e}) + \sum_{\mathbf{e}'\neq\mathbf{e}}\mu\mathbf{P}_t(\mathbf{e}')q(\mathbf{e}',\mathbf{e}). \tag{5.86}$$

Let us define matrix $\mathbf{A}^c$:

$$\mathbf{A}^c = \begin{cases} a^c(\mathbf{e},\mathbf{e}') &= \frac{q_{mn}}{c} \qquad \text{for } \mathbf{e}\neq\mathbf{e}', \\ a^c(\mathbf{e},\mathbf{e}) &= 1 - \frac{q(\mathbf{e})}{c}. \end{cases} \tag{5.87}$$

Then equality (5.86) can be rewritten as:

$$\frac{d}{dt}\mu\mathbf{P}_t(\mathbf{e}) = -c\mu\mathbf{P}_t(\mathbf{e}) + c\sum_{\mathbf{e}'\in E}\mu\mathbf{P}_t(\mathbf{e}')a^c(\mathbf{e}',\mathbf{e}). \tag{5.88}$$

For vector ($\mathbf{P}_t = (\mathbf{P}_t(0), \mathbf{P}_t(1), \ldots)$) we can write:

$$\frac{d}{dt}\mu\mathbf{P}_t = -c\ \mu\mathbf{P}_t\ [\mathbf{I} - \mathbf{A}^c], \tag{5.89}$$

what has the only solution: $\mathbf{P}_t = \mathbf{P}_0 e^{-ct[\mathbf{I}-\mathbf{A}^c]}$. Having $\mathbf{P}_t$ we construct matrix $\mathbf{Q}$ and later $\mathbf{A}^{\mathbf{c}}$ and we can again recover $\mathbf{P}_t$:

$$\mathbf{P}_t = e^{-ct[\mathbf{I}-\mathbf{A}^{\mathbf{c}}]} = e^{t\mathbf{Q}} = \sum_{k=0}^{\infty} e^{-ct}\frac{(ct)^k}{k!}\left(\mathbf{A}^c\right)^k, \tag{5.90}$$

i.e.

$$\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e}') = \sum_{k=0}^{\infty} e^{-ct}\frac{(ct)^k}{k!}\left(a^c(\mathbf{e},\mathbf{e}')\right)^k. \tag{5.91}$$

Let us assume we have continuous time process $\mathbf{X} = (X_t, t \geq 0)$ with intensity matrix $\mathbf{Q}^X$ and initial distribution μ (denote it by $X_t \sim [\mu, \mathbf{Q}^X]$). Define discrete time process $X_n^c \sim [\mu, \mathbf{A}^c]$, where $\mathbf{A}^c$ is defined in (5.87), $c \geq \sup_{\mathbf{e}\in E} q^X(\mathbf{e})$ (where $q^X(\mathbf{e}) = -q^X(\mathbf{e},\mathbf{e})$) and let $N_c(t)$ be a Poisson process with intensity c. Let $\mathbf{Y} = (Y_t, t \geq 0)$, where we define

$$Y_t = X^c_{N_c(t)}. \tag{5.92}$$

Then processes $(Y_t, t \geq 0)$ and $(X_t, t \geq 0)$ have the same finite dimensional distributions (these are the same processes). Let us calculate intensity matrix $\mathbf{Q}^Y$ of process $\mathbf{Y}$. We have to calculate differential at zero of $\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e}')$. From equality (5.91) for $\mathbf{e} \neq \mathbf{e}'$ we have $((a(\mathbf{e},\mathbf{e}')^0 = \delta_{\mathbf{e}}(\mathbf{e}'))$:

$$\frac{d}{dt}\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e}') = \frac{d}{dt}\left(\sum_{k=0}^{\infty} e^{-ct}\frac{(ct)^k}{k!}\left(a^c(\mathbf{e},\mathbf{e}')\right)^k\right)$$

$$= \frac{d}{dt}\left(e^{-ct}\left(a^c(\mathbf{e},\mathbf{e}')\right)^0 + e^{-ct}ct\left(a^c(\mathbf{e},\mathbf{e}')\right) + e^{-ct}\frac{(ct)^2}{2}\left(a^c(\mathbf{e},\mathbf{e}')^c\right)^2 + \ldots\right)$$

$$= \left(-ce^{-ct}ct\frac{q(\mathbf{e},\mathbf{e}')}{c} + e^{-ct}c\frac{q^X(\mathbf{e},\mathbf{e}')}{c} + t(\ldots+\ldots)\right)$$

and for $t = 0$:

$$q^Y(\mathbf{e},\mathbf{e}') = q^X(\mathbf{e},\mathbf{e}').$$

The same in the case $\mathbf{e} = \mathbf{e}'$, we have $q^Y(\mathbf{e}) = q^X(\mathbf{e})$.

We can have influence on process $N_c(t)$ setting any $c \geq \sup_{\mathbf{e}\in E} q^X(\mathbf{e})$ what means we can set arbitrary big intensity of jumps, but than the terms on diagonal of matrix $\mathbf{A^c}$ increase making probability of staying in the same state higher. As a result in given state the process stays "the same amount of time" as process X_t what is shown on picture 8.

In process $\mathbf{X}$ time of staying in $\mathbf{e}$ is a random variable with distribution $Exp(q(\mathbf{e}))$. The same is in $\mathbf{Y}$, the intervals between jumps (ones which really change the state) are mixture of: staying in state $\mathbf{e}$ (Poisson process with parameter c, so intervals - let us call their distribution by M - are exponentially distributed $Exp(c)$) and geometrical distribution (with probability $a^c(\mathbf{e},\mathbf{e})$ we stay in state $\mathbf{e}$, and with probability $1 - a^c(\mathbf{e},\mathbf{e})$ we leave it) - denote this random variable by N (geometric distribution with success parameter $p = 1 - a^c(\mathbf{e},\mathbf{e}) = 1 - (1 - \frac{q(\mathbf{e})}{c}) = \frac{q(\mathbf{e})}{c}$). Then we have moment generating function of N:

$$\psi_N(t) = \frac{pe^t}{1-(1-p)e^t}$$

and moment generating function of random variable M, which has distribution $Exp(c)$:

$$\psi_M(t) = \frac{c}{c-t}.$$

We are interested in distribution of $S = \sum_{k=1}^{N} M_i$ (M_i s i.i.d. $\sim M$):

$$\psi_S(t) = \psi_N(\log(\psi_M(t))) = \frac{p\frac{c}{c-t}}{1-(1-p)\frac{c}{c-t}} = \frac{p\frac{c}{c-t}}{\frac{pc-t}{c-t}} = \frac{pc}{c-t}\cdot\frac{c-t}{pc-t} = \frac{pc}{pc-t},$$

thus S has exponential distribution with parameter $pc = c\frac{q(\mathbf{e})}{c} = q(\mathbf{e})$, the same as process $\mathbf{X}$.

For finite state space $E : |E| = N$ we can always make continuous time process out of discrete by making transition matrix $\mathbf{P} = \mathbf{A}^c = \mathbf{I} - \frac{1}{c}\mathbf{Q}$ like as in (5.87). Analogically, one can make $\mathbf{Q}$ out of $\mathbf{P}$ taking $\mathbf{Q} = c(\mathbf{I} - \mathbf{P})$. Matrices $\mathbf{Q}$ and $\mathbf{P}$ have the same eigenvectors and eigenvalues of $\mathbf{Q}$ equal to $c(1-\lambda_i)$, where λ_i is eigenvalue of $\mathbf{P}$, because if $\mathbf{P}f = \lambda f$ then $\mathbf{Q}f = c(\mathbf{I} - \mathbf{P})f = cf - c\lambda f = c(1-\lambda)f$.
Again, denote right eigenvectors of $\mathbf{P}$ by $f_1, \ldots, f_N$ and left ones by $\pi_1, \ldots, \pi_N$ and eigenvalues $\lambda_1 = 1, \lambda_2, \ldots, \lambda_N$. Matrix $\mathbf{Q}$ has the same eigenvectors and the eigenvalues are: $s_1 = c(1-\lambda_1) = 0, s_2 = c(1-\lambda_2), \ldots, s_N = c(1-\lambda_N)$.

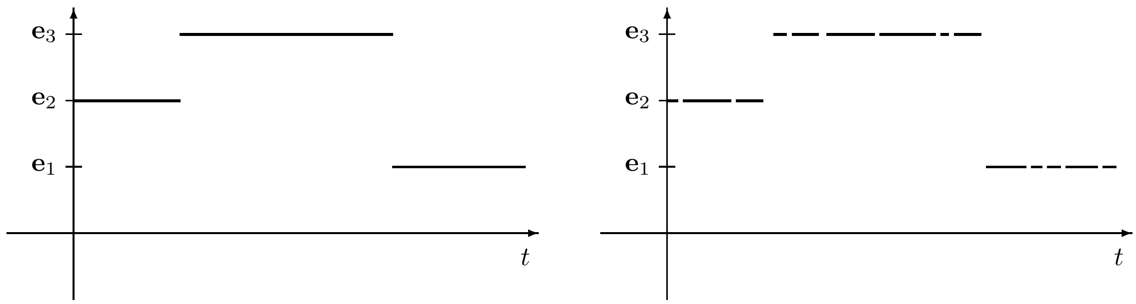


Figure 8: Exemplary "the same" realizations of processes $\mathbf{X}$ and $\mathbf{Y}$

Spectral representation (3.39) gives us $\mathbf{P}^k = \sum_{j=1}^N \lambda_j^k \mathbf{B}_j$ where $\mathbf{B}_j = f_j^T \pi_j$ as defined in (3.36). Because $\lambda_1 = 1$ and eigenvector corresponding to this eigenvalue is $f_1 = (1, 1, \dots, 1)$ thus we have

$$\mathbf{P}^k = \pi + \sum_{j=2}^N \lambda_j^k f_j \pi_j.$$

If we look at probability of getting from state $\mathbf{e}$ to $\mathbf{e}'$ after k steps we have

$$\delta_{\mathbf{e}} \mathbf{P}^k(\mathbf{e}') = \pi(\mathbf{e}') + \sum_{j=2}^N \lambda_j^k f_j(\mathbf{e}) \pi_j(\mathbf{e}'). \tag{5.93}$$

Putting $\mathbf{P}^k (= \mathbf{A}^c) = \sum_{j=1}^N \lambda_j^k \mathbf{B}_j$ into (5.90) we obtain:

$$\mathbf{P}_t = \sum_{k=0}^\infty e^{-ct} \frac{(ct)^k}{k!} (\mathbf{A}^c)^k = \sum_{k=0}^\infty e^{-ct} \frac{(ct)^k}{k!} \sum_{j=1}^N \lambda_j^k \mathbf{B}_j = \sum_{j=1}^N e^{-ct} \mathbf{B}_j \sum_{k=0}^\infty \frac{(ct\lambda_j)^k}{k!} =$$

$$\sum_{j=1}^N e^{-ct} \mathbf{B}_j e^{ct\lambda_j} = \sum_{j=1}^N e^{-ct(1-\lambda_j)} \mathbf{B}_j = \sum_{j=1}^N e^{-s_j t} \mathbf{B}_j.$$

Thus again looking at probability of getting from $\mathbf{e}$ to $\mathbf{e}'$ after time t we have

$$\delta_{\mathbf{e}} \mathbf{P}_t(\mathbf{e}') = \sum_{j=1}^N e^{-s_j t} \mathbf{B}_j(\mathbf{e}, \mathbf{e}'). \tag{5.94}$$

In particular $\Big(\mathbf{B}_j(\mathbf{e}, \mathbf{e}) = \pi_j(\mathbf{e}) \cdot \pi_j(\mathbf{e})\Big)$ we have

$$\delta_{\mathbf{e}} \mathbf{P}_t(\mathbf{e}) = \sum_{j=1}^N e^{-s_j t} (\pi_j(\mathbf{e}))^2, \tag{5.95}$$

what is a mixture of exponentials (because $s_i \geq 0$) and thus is completely monotone.
We can rewrite (5.94) putting $s_1 = 0$:

$$\delta_{\mathbf{e}} \mathbf{P}_t(\mathbf{e}') = \pi(\mathbf{e}') + \sum_{j=2}^N e^{-s_j t} f_j(\mathbf{e}) \pi_j(\mathbf{e}'). \tag{5.96}$$

5.3 Factorization in continuous time

Similarly as in section 5.1 we are interested in total variation distance between $\mu \mathbf{P}_t$ and π.

Definition 5.3.1. *For a chain $\mathbf{X} = (X_t, t \geq 0)$ with initial distribution μ a state $\tilde{\mathbf{e}}$ is* **ratio minimal** *for $\mu, \mathbf{X}$*

$$\text{if} \qquad \forall (\mathbf{e} \in E, t \geq 0) \qquad \frac{\mu \mathbf{P}_t(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})} \leq \frac{\mu \mathbf{P}_t(\mathbf{e})}{\pi(\mathbf{e})}.$$

For a chain $\mathbf{X} = (X_t, t \geq 0)$ a state $\tilde{\mathbf{e}}$ is **s-uniform** for $\mu, \mathbf{X}$ if

$$\forall\ (t \geq 0) \quad s(\mu\mathbf{P}_t, \pi) = 1 - \frac{\mu\mathbf{P}_t(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}.$$

Lemma 5.3.2. *$\tilde{\mathbf{e}}$ is ratio minimal for $\mu, \mathbf{X} \iff \tilde{\mathbf{e}}$ is s-uniform $\mu, \mathbf{X}$.*

Proof.

$\Rightarrow$ Assume $\tilde{\mathbf{e}}$ is ratio minimal for $\mu, \mathbf{X}$. Then

$$s(\mu\mathbf{P}_t, \pi) = \max_{\mathbf{e}} \left(1 - \frac{\mu\mathbf{P}_t(\mathbf{e})}{\pi(\mathbf{e})}\right) = 1 - \min_{\mathbf{e}} \frac{\mu\mathbf{P}_t(\mathbf{e})}{\pi(\mathbf{e})} = 1 - \frac{\mu\mathbf{P}_t(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}.$$

$\Leftarrow$ Assume $\tilde{\mathbf{e}}$ is s-uniform for $\mu, \mathbf{X}$. Then $\forall (n \geq 1, \mathbf{e} \in E)$

$$s(\mu\mathbf{P}_t, \pi) = 1 - \frac{\mu\mathbf{P}_t(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})} = \max_{\mathbf{e}} \left(1 - \frac{\mu\mathbf{P}_t(\mathbf{e})}{\pi(\mathbf{e})}\right) \geq \left(1 - \frac{\mu\mathbf{P}_t(\mathbf{e})}{\pi(\mathbf{e})}\right)$$

i.e.

$$\frac{\mu\mathbf{P}_t(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})} \leq \frac{\mu\mathbf{P}_t(\mathbf{e})}{\pi(\mathbf{e})}.$$

□

Lemma 5.3.3. *Let $\mathbf{X} = (X_t, t \geq 0)$ be a Markov chain with initial distribution μ. If state $\tilde{\mathbf{e}}$ is ratio minimal for $\mathbf{X}, \mu$ then*

$$\mu\mathbf{P}_{t_1}(\tilde{\mathbf{e}}) \leq \mu\mathbf{P}_{t_2}(\tilde{\mathbf{e}}) \text{ for any } t_1 < t_2$$

i.e. $\mu\mathbf{P}_t(\tilde{\mathbf{e}})$ is non-decreasing in t.

Proof. Similar to discrete-time case

$$\begin{aligned}\mu\mathbf{P}_{t_1}(\tilde{\mathbf{e}}) = \frac{\mu\mathbf{P}_{t_1}(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}\pi(\tilde{\mathbf{e}}) &\overset{*}{=} \frac{\mu\mathbf{P}_{t_1}(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})} \sum_i \pi(\mathbf{e}_i)\delta_{\mathbf{e}_i}\mathbf{P}_{t_2-t_1}(\tilde{\mathbf{e}}) = \sum_i \frac{\mu\mathbf{P}_{t_1}(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}\pi(\mathbf{e}_i)\delta_{\mathbf{e}_i}\mathbf{P}_{t_2-t_1}(\tilde{\mathbf{e}}) \\ &\overset{**}{\leq} \sum_i \frac{\mu\mathbf{P}_{t_1}(\mathbf{e}_i)}{\pi(\mathbf{e}_i)}\pi(\mathbf{e}_i)\delta_{\mathbf{e}_i}\mathbf{P}_{t_2-t_1}(\tilde{\mathbf{e}}) = \sum_i \mu\mathbf{P}_{t_1}(\mathbf{e}_i)\delta_{\mathbf{e}_i}\mathbf{P}_{t_2-t_1}(\tilde{\mathbf{e}}) = \mu\mathbf{P}_{t_2}(\tilde{\mathbf{e}}),\end{aligned}$$

where

$*$ $\sum_i \pi(\mathbf{e}_i)\delta_{\mathbf{e}_i}\mathbf{P}_{t_2-t_1}(\tilde{\mathbf{e}})$: start with stationary distribution π, then probability of getting to $\tilde{\mathbf{e}}$ after time $t_2 - t_1$ is (further stationary) $= \pi(\tilde{\mathbf{e}})$,

$**$ - from assumption.

Thus

$$\mu\mathbf{P}_{t_1}(\tilde{\mathbf{e}}) \leq \mu\mathbf{P}_{t_2}(\tilde{\mathbf{e}}) \quad \text{for} \quad t_1 < t_2.$$

□

Lemma 5.3.4. *Let X be non-negative random variable. Then:*

$\psi_X(s) = s\int_0^\infty e^{-st}P(X \leq t)dt$

Proof:

$$s\int_0^\infty e^{-st}P(X \leq t)dt = s\int_0^\infty \left(\frac{e^{-st}}{-s}\right)' P(X \leq t)dt = s\left(\frac{e^{-st}}{-s}\right)\Big|_0^\infty + \int_0^\infty e^{-st}dF(t) = \psi_X(s)$$

□

Next lemma states that in case of $\mu\mathbf{P}_t(\mathbf{e})$ being non-decreasing in t for fixed $\mathbf{e}$ we have the following factorization: $\tau_\mu(\mathbf{e}) \overset{d}{=} Y_{\mathbf{e}} + \tau_\pi(\mathbf{e})$, where $Y_{\mathbf{e}}$ has distribution funcion $F_{Y_{\mathbf{e}}}(t) = P(Y_{\mathbf{e}} \leq t) = \frac{(\mu\mathbf{P}_t)(\mathbf{e})}{\pi(\mathbf{e})}$. It is as if getting to state $\mathbf{e}$ from initial distribution μ would be done through getting to stationarity first. Then $Y_{\mathbf{e}}$ is responsible for speed of convergence to mentioned stationarity.

Lemma 5.3.5 (Brown [5]). *Let* $\mathbf{X} = (X_t, t \geq 0)$ *be ergodic Markov chain with enumerable state space* $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots\}$ *and invariant measure* π. *If for initial distribution* μ *and* $\mathbf{e} \in E$ *we have that* $\mu\mathbf{P}_t(\mathbf{e})$ *is non-decreasing in* t *then*

$$\psi_{\tau_\mu(\mathbf{e})} = \psi_{Y_\mathbf{e}} \cdot \psi_{\tau_\pi(\mathbf{e})},$$

where $Y_\mathbf{e}$ *is independent of* $\mathbf{X}$ *with distribution function* $F_{Y_\mathbf{e}}(t) = P(Y_\mathbf{e} \leq t) = \frac{(\mu\mathbf{P}_t)(\mathbf{e})}{\pi(\mathbf{e})}$.

Proof.

$$\pi(\mathbf{e}) = \mathbf{P}_\pi(X_t = \mathbf{e}) = \int_0^t \delta_\mathbf{e}\mathbf{P}_{t-x}(\mathbf{e})dF_{\tau_\pi(\mathbf{e})}(x) \tag{5.97}$$

(If we start from stationary distribution π then the distribution in any next moment does not depend on present - that is why the right hand side of the expression is independent of t)
Multiply both sides by e^{-st} and integrate it on t

$$\frac{\pi(\mathbf{e})}{s} = \int_0^\infty e^{-st}\pi(\mathbf{e})dt = \int_0^\infty e^{-st}\int_0^t \delta_\mathbf{e}\mathbf{P}_{t-x}(\mathbf{e})dF_{\tau_\pi(\mathbf{e})}(x)dt$$

$$= \int_0^\infty \int_0^t e^{-st}\delta_\mathbf{e}\mathbf{P}_{t-x}(\mathbf{e})dF_{\tau_\pi(\mathbf{e})}(x)dt = \int_0^\infty \int_x^\infty e^{-st}\delta_\mathbf{e}\mathbf{P}_{t-x}(\mathbf{e})dtdF_{\tau_\pi(\mathbf{e})}(x) =$$

$$\int_0^\infty \int_0^\infty e^{-s(t+x)}\delta_\mathbf{e}\mathbf{P}_t(\mathbf{e})dtdF_{\tau_\pi(\mathbf{e})}(x) = \int_0^\infty e^{-sx}dF_{\tau_\pi(\mathbf{e})}(x) \cdot \int_0^\infty e^{-st}\delta_\mathbf{e}\mathbf{P}_t(\mathbf{e})dt = \psi_{\delta_\mathbf{e}\mathbf{P}_t(\mathbf{e})}(s) \cdot \psi_{\tau_\pi(\mathbf{e})}(s).$$

Thus

$$\psi_{\tau_\pi(\mathbf{e})}(s) = \frac{\pi(\mathbf{e})}{s\psi_{\delta_\mathbf{e}\mathbf{P}_t(\mathbf{e})}(s)}. \tag{5.98}$$

The assumption was $\mu\mathbf{P}_t(\mathbf{e})$ non-decreasing in t and of course $\lim_{t\to\infty}\mu\mathbf{P}_t(\mathbf{e}) = \pi(\mathbf{e})$ and $\mu\mathbf{P}_t(\mathbf{e}) \leq \pi(\mathbf{e})$ for all t, thus $\frac{(\mu\mathbf{P}_t)(\mathbf{e})}{\pi(\mathbf{e})}$ is a distribution function of some random variable $Y_\mathbf{e}$. From Lemma 5.3.4 :

$$\psi_{Y_\mathbf{e}}(s) = s\int_0^\infty e^{-st}P(Y_\mathbf{e} \leq t)dt = s\int_0^\infty e^{-st}\frac{\mu\mathbf{P}_t(\mathbf{e})}{\pi(\mathbf{e})}dt = \frac{s}{\pi(\mathbf{e})}\psi_{\mu\mathbf{P}_t(\mathbf{e})}(s) \tag{5.99}$$

Similarly as in (5.97) we can write for $\mu\mathbf{P}_t(\mathbf{e})$

$$\mu\mathbf{P}_t(\mathbf{e}) = \mathbf{P}_\mu(X_t = \mathbf{e}) = \int_0^t \delta_\mathbf{e}\mathbf{P}_{t-x}(\mathbf{e})dF_{\tau_\mu(\mathbf{e})}(x) \tag{5.100}$$

Again: multiplying both sides by e^{-st} and integrating it we have

$$\psi_{\mu\mathbf{P}_t(\mathbf{e})}(s) = \psi_{\delta_\mathbf{e}\mathbf{P}_t(\mathbf{e})}(s)\psi_{\tau_\mu(\mathbf{e})}(s) \tag{5.101}$$

From (5.98) and (5.99) we obtain

$$\psi_{\delta_\mathbf{e}\mathbf{P}_t(\mathbf{e})}(s) = \frac{\pi(\mathbf{e})}{s\psi_{\tau_\pi(\mathbf{e})}} \quad \psi_{\mu\mathbf{P}_t(\mathbf{e})}(s) = \frac{\pi(\mathbf{e})}{s}\psi_{Y_\mathbf{e}}(s)$$

what together with 5.101 gives

$$\psi_{\tau_\mu(\mathbf{e})}(s) = \frac{\psi_{\mu\mathbf{P}_t(\mathbf{e})}(s)}{\psi_{\delta_\mathbf{e}\mathbf{P}_t(\mathbf{e})}(s)} = \frac{\frac{\pi(\mathbf{e})}{s}\psi_{Y_\mathbf{e}}(s)}{\frac{\pi}{s\psi_{\tau_\pi(\mathbf{e})}}} = \psi_{Y_\mathbf{e}}(s) \cdot \psi_{\tau_\mu(\mathbf{e})}.$$

□

Similar as in discrete case, we have the extension of Lemma 5.3.6 if stationary distribution is uniform.

Lemma 5.3.6. *Let* $\mathbf{X} = (X_t, t \geq 0)$ *be ergodic Markov chain with finite state space* E *and uniform stationary distribution* U. *If for initial distribution* μ *and* $B \subset E$ *we have that* $\mu\mathbf{P}_t(B)$ *is non-decreasing in* t *then*

$$\psi_{\tau_\mu(B)} = \psi_{Y_B} \cdot \psi_{\tau_U(B)},$$

where Y_B *is independent of* $\mathbf{X}$ *with distribution function* $F_{Y_B}(t) = P(Y_B \leq t) = \frac{(\mu\mathbf{P}_t)(B)}{U(\mathbf{e})}$.

Proof.

$$U(B) = \mathbf{P}_U(X_t \in B) = \int_0^t U_{|B}\mathbf{P}_{t-x}(B) dF_{\tau_U(B)}(x). \tag{5.102}$$

Multiply both sides by e^{-st} and integrate it on t

$$\frac{U(B)}{s} = \int_0^\infty e^{-st}\pi(B)dt = \int_0^\infty e^{-st}\int_0^t U_{|B}\mathbf{P}_{t-x}(B)dF_{\tau_U(B)}(x)dt$$

$$= \int_0^\infty \int_0^t e^{-st} U_{|B}\mathbf{P}_{t-x}(B)dF_{\tau_U(B)}(x)dt = \int_0^\infty \int_x^\infty e^{-st}U_{|B}\mathbf{P}_{t-x}(B)dtdF_{\tau_U(B)}(x) =$$

$$\int_0^\infty \int_0^\infty e^{-s(t+x)}U_{|B}\mathbf{P}_t(B)dtdF_{\tau_U(B)}(x) = \int_0^\infty e^{-sx}dF_{\tau_U(B)}(x) \cdot \int_0^\infty e^{-st}U_{|B}\mathbf{P}_t(B)dt = \psi_{U_{|B}\mathbf{P}_t(B)}(s) \cdot \psi_{\tau_U(B)}(s).$$

Thus:

$$\psi_{\tau_U(B)}(s) = \frac{U(B)}{s\psi_{U_{|B}\mathbf{P}_t(B)}(s)}. \tag{5.103}$$

From Lemma 5.3.4:

$$\psi_{Y_B}(s) = s\int_0^\infty e^{-st}P(Y_B \le t)dt = s\int_0^\infty e^{-st}\frac{\mu\mathbf{P}_t(B)}{U(B)}dt = \frac{s}{U(B)}\psi_{\mu\mathbf{P}_t(B)}(s). \tag{5.104}$$

Similarly as in (5.102) we can write for $\mu\mathbf{P}_t(B)$

$$\mu\mathbf{P}_t(B) = \mathbf{P}_\mu(X_t \in B) = \int_0^t U_{|B}\mathbf{P}_{t-x}(B)dF_{\tau_\mu(B)}(x). \tag{5.105}$$

Again: multiplying both sides by e^{-st} and integrating it we have

$$\psi_{\mu\mathbf{P}_t(B)}(s) = \psi_{U_{|B}\mathbf{P}_t(B)}(s)\psi_{\tau_\mu(B)}(s). \tag{5.106}$$

From (5.103) and (5.104) we obtain

$$\psi_{U_{|B}\mathbf{P}_t(B)}(s) = \frac{U(B)}{s\psi_{\tau_U(B)}} \quad \psi_{\mu\mathbf{P}_t(B)}(s) = \frac{U(B)}{s}\psi_{Y_B}(s),$$

what together with 5.106 gives

$$\psi_{\tau_\mu(B)}(s) = \frac{\psi_{\mu\mathbf{P}_t(B)}(s)}{\psi_{U_{|B}\mathbf{P}_t(B)}(s)} = \frac{\frac{U(B)}{s}\psi_{Y_\mathbf{e}}(s)}{\frac{U(B)}{s\psi_{\tau_U(B)}}} = \psi_{Y_B}(s) \cdot \psi_{\tau_\mu(B)}.$$

□

Theorem 5.3.7. *Let $\mathbf{X} = (X_t : t \ge 0)$ be a Markov chain with stationary distribution π. If $\tilde{\mathbf{e}}$ is ratio minimal for $\mu, \mathbf{X}$ then*

(*i*) $\psi_{\tau_\mu(\tilde{\mathbf{e}})} = \psi_{Y_{\tilde{\mathbf{e}}}} \cdot \psi_{\tau_\pi(\tilde{\mathbf{e}})} \qquad (\tau_\mu(\tilde{\mathbf{e}}) \overset{d}{=} Y_{\tilde{\mathbf{e}}} + \tau_\pi(\tilde{\mathbf{e}})),$

(*ii*) there exists Minimal Strong Stationary Time T_μ such that

$$s(\mu\mathbf{P}_t, \pi) = 1 - \frac{\mu\mathbf{P}_t(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})} = P(T_\mu > t) \quad \forall (t > 0).$$

Proof.

(i) From lemma 5.3.3 we have that $\mu\mathbf{P}^n(\tilde{\mathbf{e}})$ and this is the assumption of Lemma 5.1.5 which implies $\psi_{\tau_\mu(\tilde{\mathbf{e}})} = \psi_{Y_{\tilde{\mathbf{e}}}} \cdot \psi_{\tau_\pi(\tilde{\mathbf{e}})}$.

(ii) In Lemma 4.1.8 it was proven (by Aldous & Diaconis) that there always existed $\mu-$Minimal Strong Stationary Time for discrete time. In proof we had: $\alpha_n = \min_\mathbf{e} \frac{\mu\mathbf{P}^n(\mathbf{e})}{\pi(\mathbf{e})}$. Because $\tilde{\mathbf{e}}$ is ratio minimal, thus we have $\alpha_n = \frac{\mu\mathbf{P}^n(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}$. It was constructed such way that

$$P(T_\mu = n) = \alpha_n - \alpha_{n-1}.$$

Using uniformization (see section 5.2.1), by choosing $c = 2\max_i \sum_{k\neq i} q_{ik}$ we obtain discrete time skeleton $\mathbf{P} = \mathbf{I} + \frac{1}{c}\mathbf{Q}$ which satisfies the above conditions. Denote the embedded Markov chain by $\mathbf{Y} = \{Y_n, n \geq 0\}$. We can represent $\mathbf{X} = (X_t, t \geq 0)$ by $\{X_t = Y_{N_t}\}$, where $\mathbf{N} = (N_t, t \geq 0)$ is a Poisson process with rate c (independent of $\mathbf{Y}$). Denote epochs from the Poisson process as $\{S_n\}$ and define $Z = S_{Z'}$, where Z' is a μ-Minimal Strong Stationary Time for $\mathbf{Y}$. Of course Z is a Strong Stationary Time for continuous time and:

$$P(Z > t) = \sum_{n=0}^{\infty} \frac{(ct)^n e^{-ct}}{n!} P(Z' > n) = \sum_{n=0}^{\infty} \frac{(ct)^n e^{-ct}}{n!} \left(1 - \frac{\mu\mathbf{P}^n(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})}\right) = 1 - \frac{\mu\mathbf{P}_t(\tilde{\mathbf{e}})}{\pi(\tilde{\mathbf{e}})} = s(\mu\mathbf{P}_t, \pi).$$

□

Next lemma says us about the structure of a chain in the case of $\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})$ being non-increasing in t. Then first passage time to $\mathbf{e}$ starting with π is a geometric compound: $\tau_\pi(\mathbf{e}) \stackrel{d}{=} \sum_{i=1}^{N} W_{\mathbf{e},i}$, where N has geometric distribution: $P(N = k) = (1 - \pi(\mathbf{e}))^k \pi(\mathbf{e}), k = 0, 1, \ldots$ and $W_{\mathbf{e},i}, i = 1, 2, \ldots$ are i.i.d. random variables with the same distribution as $W_{\mathbf{e}}$: $P(W_{\mathbf{e}} \leq t) = \frac{1-\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})}{1-\pi(\mathbf{e})}$.

Theorem 5.3.8 (Brown [5])**.** *Let $\mathbf{X} = (X_t, t \geq 0)$ be ergodic Markov chain with enumerable state space $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots\}$ and invariant measure π. If for a state $\mathbf{e}$ we have that $\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})$ is non-increasing in t then*

$$\psi_{\tau_\pi(\mathbf{e})} = \frac{\pi(\mathbf{e})}{1 - (1 - \pi(\mathbf{e}))\psi_{W_{\mathbf{e}}}},$$

where $W_{\mathbf{e}}$ is independent of $\mathbf{X}$ and $P(W_{\mathbf{e}} > t) = \frac{\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e}) - \pi(\mathbf{e})}{1-\pi(\mathbf{e})}$.

Proof: Using lemma 5.3.4 :

$$\psi_{W_{\mathbf{e}}} = s\int_0^\infty e^{-st} P(W_{\mathbf{e}} \leq t)dt = 1 - s\int_0^\infty e^{-st} P(W_{\mathbf{e}} > t)dt = 1 - s\int_0^\infty e^{-st} \frac{\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e}) - \pi(\mathbf{e})}{1 - \pi(\mathbf{e})} dt$$

$$= 1 - \frac{s}{1-\pi(\mathbf{e})}\int_0^\infty e^{-st}\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})dt + \frac{s\pi(\mathbf{e})}{1-\pi(\mathbf{e})}\int_0^\infty e^{-st} = 1 - \frac{s}{1-\pi(\mathbf{e})}\psi_{\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})}(s) + \frac{s\pi(\mathbf{e})}{1-\pi(\mathbf{e})}\frac{1}{s}$$

$$= 1 - \frac{s}{1-\pi(\mathbf{e})}\psi_{\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})}(s) + \frac{\pi(\mathbf{e})}{1-\pi(\mathbf{e})}$$

Thus:

$$s\psi_{\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})}(s) = 1 + (1 - \pi(\mathbf{e}))\psi_{W_{\mathbf{e}}}(s)$$

Putting it to (5.98) $\left(\psi_{\tau_\pi(\mathbf{e})}(s) = \frac{\pi(\mathbf{e})}{s\psi_{\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})}(s)}\right)$ one gets:

$$\psi_{\tau_\pi(\mathbf{e})}(s) = \frac{\pi(\mathbf{e})}{1 + (1 - \pi(\mathbf{e}))\psi_{W_{\mathbf{e}}}(s)}.$$

□

5.3.1 Special cases

The clue is to find out when $\tilde{\mathbf{e}}$ ratio minimality holds. Let us adapt assumptions A0 - A2 from section 5.1.1 (Factorization for discrete time).

• Stochastic monotonicity of time-reversed process $\tilde{\mathbf{X}}$

We need a definition of stochastic monotonicity in continuous time.

Set A is an upper set (denoted by $A\uparrow$), if ($x \in A$ and $x \prec y$) $\Rightarrow y \in A$. Set B is a lower set (denoted by $B\downarrow$), if ($x \in B$ and $y \prec x$) $\Rightarrow y \in B$.

Markov chain with intensity matrix $\mathbf{Q}$ is **stochastically monotone** if both beneath conditions hold:

$$\begin{aligned} &\forall(A\uparrow: \mathbf{e}_2 \notin A)\forall(\mathbf{e}_1 \prec \mathbf{e}_2) \qquad \mathbf{Q}(\mathbf{e}_1, A) = \sum_{\mathbf{e}\in A}\mathbf{Q}(\mathbf{e}_1, \mathbf{e}) \leq \mathbf{Q}(\mathbf{e}_2, A) \\ &\text{and} \\ &\forall(A\uparrow: \mathbf{e}_2 \notin A)\forall(\mathbf{e}_1 \prec \mathbf{e}_2) \qquad \mathbf{Q}(\mathbf{e}_1, A) = \sum_{\mathbf{e}\in A}\mathbf{Q}(\mathbf{e}_1, \mathbf{e}) \leq \mathbf{Q}(\mathbf{e}_2, A). \end{aligned} \tag{5.107}$$

Lemma 5.3.9 (Brown [5]). *Let* $\mathbf{X} = (X_t, t \geq 0)$ *be a continuous time Markov chain with intensity matrix* $\mathbf{Q}$ *which is stochastically monotone. Let* $c \geq 2 \max_{\mathbf{e}} \sum_{\mathbf{e}' \neq \mathbf{e}} \mathbf{Q}(\mathbf{e}, \mathbf{e}')$ *and define* $\mathbf{P} = \mathbf{I} + \frac{1}{c}\mathbf{Q}$. *Then* $\mathbf{P}$ *is stochastically monotone in sense of definition 5.84.*

Proof. Consider three cases:

1) If $\mathbf{e}_1 \prec \mathbf{e}_2$ and A is an upper set such that $\mathbf{e}_2 \notin A$ then

$$\delta_{\mathbf{e}_1}\mathbf{P}(A) = \frac{1}{c}\mathbf{Q}(\mathbf{e}_1, A) \leq \frac{1}{c}\mathbf{Q}(\mathbf{e}_2, A) = \delta_{\mathbf{e}_2}\mathbf{P}(A).$$

2) If $\mathbf{e}_1 \prec \mathbf{e}_2$ and A is an upper set with $\mathbf{e}_1 \in A$ then

$$\delta_{\mathbf{e}_1}\mathbf{P}(A) = 1 - \frac{1}{c}\mathbf{Q}(\mathbf{e}_1, A^c) \leq 1 - \frac{1}{c}\mathbf{Q}(\mathbf{e}_2, A^c) = \delta_{\mathbf{e}_2}\mathbf{P}(A)$$

(A^c is a compliment of A and a lower set).

3) If $\mathbf{e}_1 \prec \mathbf{e}_2$ and A is an upper set such that $\mathbf{e}_1 \notin A$ and $\mathbf{e}_2 \in A$ then

$$\delta_{\mathbf{e}_1}\mathbf{P}(A) = \frac{1}{c}\mathbf{Q}(\mathbf{e}_1, A)$$

$$\delta_{\mathbf{e}_2}\mathbf{P}(A) = 1 - \frac{1}{c}\mathbf{Q}(\mathbf{e}_2, A^c),$$

thus

$$\delta_{\mathbf{e}_1}\mathbf{P}(A) - \delta_{\mathbf{e}_2}\mathbf{P}(A) = 1 - \frac{1}{c}[\mathbf{Q}(\mathbf{e}_1, A) + \mathbf{Q}(\mathbf{e}_2, A^c)] \geq 1 - \frac{\mathbf{Q}(\mathbf{e}_1, A) + \mathbf{Q}(\mathbf{e}_2, A^c)}{2 \max_{\mathbf{e}} \sum_{\mathbf{e}' \neq \mathbf{e}} \mathbf{Q}(\mathbf{e}, \mathbf{e}')} \geq 0.$$

Using stochastic monotonicity of $\mathbf{P}$ we have:

$$\forall (A \uparrow) \forall (\mathbf{e}_1 \prec \mathbf{e}_2) \qquad \delta_{\mathbf{e}_1}\mathbf{P}_t(A) = \sum_{n=0}^{\infty} \frac{(ct)^n e^{-ct}}{n!} \delta_{\mathbf{e}_1}\mathbf{P}^n(A) \leq \sum_{n=0}^{\infty} \frac{(ct)^n e^{-ct}}{n!} \delta_{\mathbf{e}_2}\mathbf{P}^n(A) = \delta_{\mathbf{e}_2}\mathbf{P}_t(A),$$

i.e.

$$\forall (A \uparrow) \forall (\mathbf{e}_1 \prec \mathbf{e}_2) \qquad \delta_{\mathbf{e}_1}\mathbf{P}_t(A) \leq \delta_{\mathbf{e}_2}\mathbf{P}_t(A). \tag{5.108}$$

□

Time-reversed process $\{\tilde{\mathbf{X}}\}$ is the one with intensity matrix $\tilde{\mathbf{Q}}(\mathbf{e}_2, \mathbf{e}_1) = \frac{\pi(\mathbf{e}_1)}{\pi(\mathbf{e}_2)}\mathbf{Q}(\mathbf{e}_1, \mathbf{e}_2)$ (and process is **time-reversible** if $\tilde{\mathbf{Q}} = \mathbf{Q}$).

Theorem 5.3.10 (Brown [5]). *Assume that* $A0 - A2$ *are fulfilled, and in addition that*

- *Time-reversed process* $\widetilde{\mathbf{X}} = (\tilde{X}_t, t \geq 0)$ *is stochastically monotone.*

Then

(i) e_M is ratio minimal for $\mu, \mathbf{X}$,

(ii) $\psi_{\tau_\mu(\mathbf{e}_M)} = \psi_{Y_{\mathbf{e}_M}} \cdot \psi_{\tau_\pi(\mathbf{e}_M)} \qquad (\tau_\mu(\mathbf{e}_M) \stackrel{d}{=} Y_{\mathbf{e}_M} + \tau_\pi(\mathbf{e}_M))$,

(iii) there exists Minimal Strong Stationary Time T_μ such that

$$s(\mu\mathbf{P}_t, \pi) = 1 - \frac{\mu\mathbf{P}_t(\mathbf{e}_M)}{\pi(\mathbf{e}_M)} \quad \forall (t \geq 0).$$

where $Y_{\mathbf{e}_M}$ *is random variable independent of chain* $\mathbf{X}$ *with distribution* $P(Y_{\mathbf{e}_M} \leq t) = \frac{\mu\mathbf{P}_t(\mathbf{e}_M)}{\pi(\mathbf{e}_M)}$.

Proof. Similar to discrete-time case.
Define $h(\mathbf{e}) = \frac{\mu(\mathbf{e})}{\pi(\mathbf{e})}$. Using Fact 5.1.9 we have

$$\frac{\mu\mathbf{P}_t(\mathbf{e}_M)}{\pi(\mathbf{e}_M)} = \sum_{\mathbf{e}_i \in E} \frac{\mu(\mathbf{e}_i)}{\pi(\mathbf{e}_M)} \delta_{\mathbf{e}_i}\mathbf{P}_t(\mathbf{e}_M) = \sum_{\mathbf{e}_i \in E} \frac{\mu(\mathbf{e}_i)}{\pi(\mathbf{e}_M)} \cdot \frac{\pi(\mathbf{e}_M)}{\pi(\mathbf{e}_i)} \delta_{\mathbf{e}_M}\tilde{\mathbf{P}}_t(\mathbf{e}_i)$$

$$= \sum_{\mathbf{e}_i \in E} \frac{\mu(\mathbf{e}_i)}{\pi(\mathbf{e}_i)} \delta_{\mathbf{e}_M}\tilde{\mathbf{P}}_t(\mathbf{e}_i) = E_{\mathbf{e}_M} h(\tilde{X}) \leq E_{\mathbf{e}_i} h(\tilde{X}) = \frac{\mu\mathbf{P}_t(\mathbf{e}_i)}{\pi(\mathbf{e}_i)},$$

i.e. $\mathbf{e}_M$ is ratio minimal for μ, X. This (by Theorem 5.3.7) implies (ii) and (iii).

□

- **Time-reversible processes**

Lemma 5.3.8 gives us factorization in case when $\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})$ is non-increasing in t. We find some conditions which guarantee it.

Lemma 5.3.11. *Let $\mathbf{X} = (X_t, t \geq 0)$ be a Markov chain with finite state space $E = |N|$ which is time-reversible, i.e. for all $\mathbf{e}_1, \mathbf{e}_2 \in E$ we have $\pi(\mathbf{e}_1)\mathbf{Q}(\mathbf{e}_1, \mathbf{e}_2) = \pi(\mathbf{e}_2)\mathbf{Q}(\mathbf{e}_2, \mathbf{e}_1)$. Then*

$$\forall(\mathbf{e} \in E), \forall(t_1 < t_2) \qquad \delta_{\mathbf{e}}\mathbf{P}_{t_1}(\mathbf{e}) \geq \delta_{\mathbf{e}}\mathbf{P}_{t_2}(\mathbf{e})$$

i.e. $\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})$ is non-increasing in t for every $\mathbf{e} \in E$.

Proof. As seen in subsection 5.2.1(Uniformization) any transition matrix of discrete time Markov chain $\mathbf{P}$ can be constructed out of intensity matrix $\mathbf{Q}$ and vice versa. We have the following dependencies:

$$\mathbf{P} \text{ has eigenvalues } \lambda_1, \ldots, \lambda_N \qquad \Longleftrightarrow \qquad \mathbf{Q} \text{ has eigenvalues } s_1 = c(1-\lambda_1), \ldots, s_N = c(1-\lambda_N)$$

for some constant $c > 0$.
Recall equality (5.95)

$$\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e}) = \sum_{j=1}^{N} e^{-s_j t}(\pi_j(\mathbf{e}))^2,$$

where all s_i are non-negative what implies required monotonicity of $\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})$ in t. □

Remark: It is worth noting the difference between discrete and continuous time. In discrete time the assumption of $\mathbf{P}$ being time-reversible was not enough to guarantee monotonicity of $\delta_{\mathbf{e}}\mathbf{P}_t(\mathbf{e})$ (this is the case when we observe only every second step of such chain), whereas in continuous time this condition is sufficient.

5.4 Non-symmetric random walk on d-dimensional cube

Consider discrete time Markov chain $\mathbf{X} = \{X_n, n \geq 0\}$, $E = \{0,1\}^d$ with transition probabilities:

$$\begin{aligned} \delta_x \mathbf{P}(x + s_i) &= \alpha_i \text{ for } x_i = 0, \\ \delta_x \mathbf{P}(x - s_i) &= \beta_i \text{ for } x_i = 1, \\ \delta_x \mathbf{P}(x) &= 1 - \sum_{i:x_i=0} \alpha_i - \sum_{i:x_i=1} \beta_i, \end{aligned} \tag{5.109}$$

where $x = (x_1, \ldots, x_d) \in E, x_i \in \{0,1\}$ and $s_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 at position i.
Assume that α_i and β_i are such that the chain is ergodic. Note that for $\alpha_i = \beta_i = \frac{1}{d+1}$ for $i = 1, \ldots, d$ we have a symmetric random walk on cube as defined in (3.51).
From remark 5.1.11 the only potential state which can be ratio minimal is state $x+1$, where $(x+1)_i = x_i + 1 \mod 2$. This suggests ordering through distance. Recall $dist(x, y) = \min_n \delta_x \mathbf{P}^n(y)$,

$$y \prec z \iff dist(x, y) < dist(x, z). \tag{5.110}$$

Under this ordering x is a unique minimal state, and $x + 1$ unique maximal. As minimal state we can fix $x_{min} = (0, \ldots, 0)$, then the maximal state is $x_{max} = (1, \ldots, 1)$. The problem reduces to finding conditions for α_i i β_i under which $\tilde{\mathbf{X}}$ is stochastically monotone, for then we could use Theorem 5.3.10.

Such chain is time-reversible with stationary distribution:

$$\pi(x) = \prod_{i:x_i=1} \frac{\alpha_i}{\alpha_i + \beta_i} \prod_{i:x_i=0} \frac{\beta_i}{\alpha_i + \beta_i} \tag{5.111}$$

Let us check time-reversibility: let x be such that $x_k = 0$. Denote $G = \prod_{i=1}^d (\alpha_i + \beta_i)$.

$$\pi(x)\delta_x \mathbf{P}(x + \delta_k) = \frac{1}{G} \left(\prod_{\substack{i:x_i=1 \\ i \neq k}} \alpha_i \right) \left(\prod_{\substack{i:x_i=0 \\ i \neq k}} \beta_i \cdot \beta_k \right) \cdot \alpha_k,$$

$$\pi(x + \delta_k)\delta_{(x+\delta_k)} \mathbf{P}(x) = \frac{1}{G} \left(\prod_{\substack{i:x_i=1 \\ i \neq k}} \alpha_i \cdot \alpha_k \right) \left(\prod_{\substack{i:x_i=0 \\ i \neq k}} \beta_i \right) \cdot \beta_k,$$

so we have equality $\pi(x)\delta_x \mathbf{P}(x + \delta_k) = \pi(x + \delta_k)\delta_{(x+\delta_k)} \mathbf{P}(x)$.
Because of time-reversibility it is enough to check whether X_n is stochastically monotone.

Brown in [5] stated that the following conditions assured ergodicity and stochastic monotonicity of the chain:

$$\alpha_i > 0, \beta_i > 0, \; i = 1, \ldots, d, \tag{5.112}$$

$$\exists (A \subset \{1, 2, \ldots, d\}) \quad \sum_{i \in A} \alpha_i + \sum_{i \in A^c} \beta_i < 1, \tag{5.113}$$

$$\max(\sum_{i=1}^d \alpha_i + \max_i \beta_i, \quad \sum_{i=1}^d \beta_i + \max_i \alpha_i) \leq 1, \tag{5.114}$$

however we have

Lemma 5.4.1. *The chain given in (5.109) is not always stochastically monotone under ordering (5.110).*

Proof. It is enough to consider the example: Take $d = 3$ and

$$\alpha_1 = \frac{1}{1000}, \alpha_2 = \frac{5}{100}, \alpha_3 = \frac{799}{1000}, \quad \beta_1 = \beta_2 = \frac{1}{10}, \beta_3 = \frac{1}{1000}$$

and take

$$x = (1, 0, 0) \preceq (1, 1, 0) = y, \quad A = \{(1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 1, 1)\}.$$

Then

$$\delta_x \mathbf{P}(A) = \alpha_2 + \alpha_3, \quad \delta_y \mathbf{P}(A) = 1 - \beta_1 - \beta_2$$

and stochastic monotonicity would mean that $\delta_x \mathbf{P}(A) \le \delta_y \mathbf{P}(A)$ i.e.

$$\alpha_2 + \alpha_3 + \beta_1 + \beta_2 \le 1,$$

but we have

$$\alpha_2 + \alpha_3 + \beta_1 + \beta_2 = \frac{50}{1000} + \frac{799}{1000} + \frac{100}{1000} + \frac{100}{1000} = \frac{1049}{1000} > 1.$$

□

Let us introduce the following **coordinate-wise** ordering:

$$x \preceq y \iff x_i \le y_i \quad \text{for} \quad i = 1, 2, \ldots d. \tag{5.115}$$

Theorem 5.4.2. *Suppose* $\mathbf{X} = \{X_n, n \ge 0\}$ *is a random walk as defined in (5.109). Assume that:*

$$(a) \qquad \alpha_i > 0, \beta_i > 0, \; i = 1, \ldots, d, \tag{5.116}$$

$$(b) \qquad \exists (A \subset \{1, 2, \ldots, d\}) \quad \sum_{i \in A} \alpha_i + \sum_{i \in A^c} \beta_i < 1, \tag{5.117}$$

$$(c) \qquad \forall (B \subseteq \{1, 2, \ldots, d\}, r \in B) \quad \sum_{k \in B} \alpha_k + \sum_{k \in B^c} \beta_k + \beta_r \le 1. \tag{5.118}$$

Then we have

(*i*) x_{max} is ratio minimal for $\delta_{x_{min}}, \mathbf{X}$,

(*ii*) $\psi_{\tau_{\delta_{x_{min}}}(x_{max})} = \psi_{Y_{x_{max}}} \cdot \psi_{\tau_\pi(x_{max})} \qquad (\tau_{\delta_{x_{min}}}(x_{max}) \stackrel{d}{=} Y_{x_{max}} + \tau_\pi(x_{max}))$,

(*iii*) there exists Minimal Strong Stationary Time T_μ such that

$$s(\delta_{x_{min}} \mathbf{P}^n, \pi) = 1 - \frac{\delta_{x_{min}} \mathbf{P}^n(x_{max})}{\pi(x_{max})} = P(T_\mu > n) \qquad \forall (n \ge 0).$$

where $Y_{x_{max}}$ *is a random variable independent of* $\mathbf{X}$ *with distribution* $P(Y_{x_{max}} \le n) = \frac{\delta_{x_{min}} \mathbf{P}^n(x_{max})}{\pi(x_{max})}$.

Proof. Condition (5.116) is sufficient and necessary for irreducibility and (5.117) for aperiodicity. It is enough to prove that (5.118) implies stochastic monotonicity under coordinate-wise ordering, because: under this ordering x_{min} is unique minimal state and x_{max} is unique maximal state, thus condition $A1$ (see section 5.1.1) holds, $A2$ also holds because for initial distribution being an atom we always have $\frac{\delta_{x_{min}}(\mathbf{e})}{\pi(\mathbf{e})}$ non-increasing in $\mathbf{e}$. It is enough to have stochastic monotonicity of $\mathbf{X}$ because it is time-reversible. Having this stochastic monotonicity we can conclude all three assertions from Theorem 5.3.10.

In order to prove that (5.118) implies stochastic monotonicity assume that

$$x = \begin{cases} 0 & \text{at positions} \quad \{j_1, \ldots, j_m\} \\ 1 & \text{at positions} \quad \{i_1, \ldots, i_{d-m}\} \end{cases} \preceq \quad y = \begin{cases} 0 & \text{at positions} \quad \{j_1, \ldots, j_m\} \setminus \{t_{d-m+1}, \ldots, t_{d-n}\} \\ 1 & \text{at positions} \quad \{i_1, \ldots, i_{d-m}\} \cup \{t_{d-m+1}, \ldots, t_{d-n}\} \end{cases}$$

and let A be an upper set.

Denote level of x by $|x| = \sum_{i=1}^d x_i$ and let $e_k = (0, 0, \ldots, 0, k, 0, \ldots, 0)$, where 1 is on $k-$th position. Consider three cases:

1° $x \notin A$, $y \notin A$.
Then

$$\delta_x \mathbf{P}(A) = \sum_{k \in B} \alpha_k, \quad B \subseteq \{j_1, \ldots, j_m\}.$$

Fix $k \in B$. We have $x \notin A$ and $x + e_k \in A$. Note that y cannot have 1 on $k-$th position, because then we would have $x + e_k \preceq y$ and $y \in A$ (because A is an upper set), but we assumed $y \notin A$. But it means there is a state $y + e_k$ and of course it is comparable to $x + e_k$ which is in A, thus $y + e_k$ must also be in A. Other words: $x + e_k \in A$ implies $x + e_k \in A$ and we have

$$\delta_x \mathbf{P}(A) \le \delta_y \mathbf{P}(A).$$

2° $x \in A$, $y \in A$.
Then
$$\delta_y \mathbf{P}(A) = 1 - \sum_{k \in C} \beta_k, \quad C \subseteq \{i_1, \ldots, i_{d-m}\} \cup \{t_{d-m+1}, \ldots, t_{d-n}\}.$$
Fix $k \in C$. Then $y - e_k \notin A$. Note that 0 cannot have 1 on positione k, because then we would have $x \preceq y - e_k$ and $x \in A$ would imply $y - e_k$ (because A is an upper set). Thus x has 1 at position k. It must be that alsa $x - e_k \notin A$ because if we assumed contradiction, i.e. that $x - e_k \in A$ it would imply that $y - e_k \in A$ (because $x - e_k \preceq y - e_k$ and A is an upper set). Thus
$$\delta_x \mathbf{P}(A) \leq \delta_y \mathbf{P}(A).$$

3° $x \notin A, y \in A$. Consider two cases:

a) $|x| + 1 = |y|$, i.e. y is just one level higher (equivalently: $n = m - 1$).
Then
$$\delta_x \mathbf{P}(A) = \sum_{k \in B} \alpha_k, \quad B \subseteq \{j_1, \ldots, j_m\}$$
and
$$\delta_y \mathbf{P}(A) = 1 - \sum_{k \in C} \beta_k, \quad C \subseteq \{i_1, \ldots, i_{d-m}\} \cup \{t\}.$$
The worst case is when $\delta_x \mathbf{P}(A)$ is maximized and $1 - \delta_y \mathbf{P}(A)$ minimized - let us check this case by taking $B = \{j_1, \ldots, j_m\}$ and $C = \{i_1, \ldots, i_{d-m}\} \cup \{t\} = B^c \cup \{t\}$. Stochastic monotonicity means then for some $r \in B$ corresponding to $\{t\}$) we have
$$\sum_{k \in B} \alpha_k \leq 1 - \sum_{k \in B^c} \beta_k - \beta_r,$$
what was assumed.

b) $|x| + 2 < |y|$, i.e. y is at least two levels higher.
Then similarly
$$\delta_x \mathbf{P}(A) = \sum_{k \in B} \alpha_k, \quad B \subseteq \{j_1, \ldots, j_m\}$$
and
$$\delta_y \mathbf{P}(A) = 1 - \sum_{k \in C} \beta_k, \quad C \subseteq \{i_1, \ldots, i_{d-m}\} \cup \{t_{d-m+1}, \ldots, t_{d-n}\}.$$
Moreover, fix $k \in B$. Then $x + e_k \in A$. If y has 0 on $k-$th position then surely $k \notin C$. Otherwise note that $x + e_k \in A$ implies (because A is an upper set) that $y - e_s \in A$ for $s \neq k$ and $s \in B \cap C$. Thus we conclude that $B \cap C = \emptyset$. So the worst case is $C = B^c$ and then stochastic monotonicity means
$$\delta_x \mathbf{P}(A) = \sum_{k \in B} \alpha_k \leq 1 - \sum_{k \in B^c} \beta_k = \delta_y \mathbf{P}(A),$$
i.e.
$$\sum_{k \in B} \alpha_k + \sum_{k \in B^c} \beta_k \leq 1$$
what is implied by our assumption (5.118).

□

REMARKS

1. Brown's condition (5.114), even for coordinate-wise partial ordering is not sufficient. Again, take the example given in Lemma 5.4.1
   To see this fix $d = 3$ and set
$$\alpha_1 = \frac{1}{1000}, \alpha_2 = \frac{5}{100}, \alpha_3 = \frac{799}{1000}, \quad \beta_1 = \beta_2 = \frac{1}{10}, \beta_3 = \frac{1}{1000}$$
   and take
$$x = (1, 0, 0) \preceq (1, 1, 0) = y, \quad A = \{(1, 1, 0), (1, 0, 1), (1, 1, 1)\}.$$
   Then
$$\delta_x \mathbf{P}(A) = \alpha_2 + \alpha_3, \quad \delta_y \mathbf{P}(A) = 1 - \beta_1 - \beta_2$$

and stochastic monotonicity would mean that $\delta_x\mathbf{P}(A) \le \delta_y\mathbf{P}(A)$ i.e.

$$\alpha_2+\alpha_3+\beta_1+\beta_2 \le 1,$$

but we have

$$\alpha_2+\alpha_3+\beta_1+\beta_2 = \frac{50}{1000}+\frac{799}{1000}+\frac{100}{1000}+\frac{100}{1000} = \frac{1049}{1000} > 1.$$

2. Brown's condition (5.114) is just a special case of (5.118).
   If we take $B = \{1,\dots,d\}$ and any $r \in B$, then (5.118) means: $\sum_{i=1}^d \alpha_i + \beta_r \le 1$
   If we take $B = \{r\}$ then (5.118) means: $\alpha_r + \sum_{i=1}^d \beta_i \le 1$. Putting above couple together we have Brown's condition, i.e.

$$\max(\sum_{i=1}^d \alpha_i + \max_i \beta_i, \quad \sum_{i=1}^d \beta_i + \max_i \alpha_i) \le 1.$$

Theorem 5.4.3. *Suppose $\mathbf{X} = \{X_n, n \ge 0\}$ is a random walk as defined in (5.109) and let assumptions of Theorem 5.4.2 hold, i.e. (5.116), (5.117) and (5.118). Define A_k to be a set of $\binom{d}{k}$ subsets of size k from $\{1,\dots,d\}$ and $s_\gamma = \sum_{i\in\gamma}(\alpha_i+\beta_i)$ for γ a subset of $1,\dots,d$. Then*

$$(i) \qquad s(\delta_{x_{min}}\mathbf{P}^n, \pi) = \sum_{k=1}^d (-1)^{k-1} \sum_{\gamma\in A_k} (1-s_\gamma)^n \qquad \forall(n \ge 0), \tag{5.119}$$

$$(ii) \qquad \text{all } 2^d \text{ eigenvalues of } \mathbf{P} \text{ are } \{1-s_\gamma, \gamma \subset \{1,\dots,d\}\}. \tag{5.120}$$

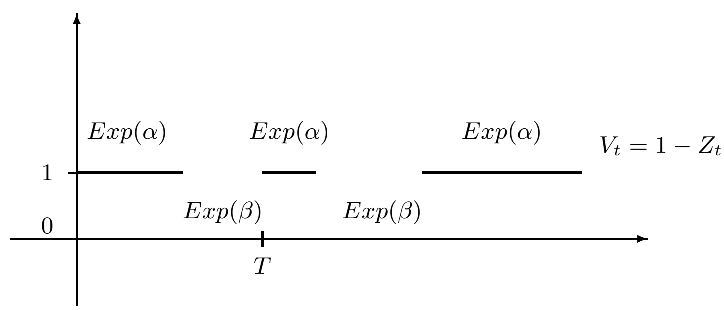
If instead of (5.118) we assume that $\sum_{i=1}^d(\alpha_i+\beta_i) \le 1$ then

$$(iii) \qquad \forall(x\in E) \qquad \delta_x\mathbf{P}^n(x) \text{ is non-increasing in } n, \tag{5.121}$$

$$(iv) \qquad \forall(x\in E) \qquad \psi_{\tau_\pi(x)} = \frac{\pi(x)}{1-(1-\pi(x))\psi_{W_x}}, \tag{5.122}$$

where W_x is independent of $\mathbf{X}$ and $P(W_x > n) = \frac{\delta_x\mathbf{P}^n(x)-\pi(x)}{1-\pi(x)}$.

Proof. To calculate $s(\delta_{(x_{min})}\mathbf{P}^n, \pi)$ we have to calculate $\delta_{(x_{min})}\mathbf{P}^n(x_{max})$. Let us switch to continuous time using uniformization (with $c = 1$) $\mathbf{Q} = \mathbf{P} - \mathbf{I}$. Then on the diagonal of the intensity matrix we have $\mathbf{Q}(x,x) = -(\sum_{i:x_i=0}\alpha_i + \sum_{i:x_i=1}\beta_i)$, the process stays in state x for time which is exponentially distributed with parameter $((\sum_{i:x_i=0}\alpha_i + \sum_{i:x_i=1}\beta_i)$. Notice that it is a distribution of minimum of random variables with parameters α_i and β_i. Exponential distribution has memoryless property - that is why this process is a sum of d independent zero-one processes.
Look closer at one such zero-one process, which stays in state 0 time $Exp(\alpha)$, and in 1 $Exp(\beta)$. Denote this process by $\{Z_t, t \ge 0\}$. To calculate $\delta_{(x_{min})}\mathbf{P}^n(x_{max})$ we have to calculate probability $P(Z_t = 1)$. I will be easier to consider process $V_t = 1 - Z_t$.



We will write renewal equation for $q(t) = P(V_t = 1)$. Let F be the distribution of sum of two random variables: $Exp(\alpha) + Exp(\beta)$, let $X_1 \sim Exp(\alpha), X_2 \sim Exp(\beta)$ and $T = X_1 + X_2 \ \sim F$:

$$q(t) = P(V_t = 1, t < T) + P(V_t = 1, t \ge T) = z(t) + \int_0^t q(t-u)F(du)$$

Where $z(t) = P(V_t = 1, t < X_1+X_2) = P(t \le X_1, t < X_1+X_2) = P(t \le X_1) = 1-P(X_1 \le t) = 1-(1-e^{-\alpha t}) = e^{-\alpha t}$

The solution to this equation is convolution $U * z$, where $U = \sum_{n=0}^{\infty} F^{n*}$.
Laplace transform of $z(t)$: $\psi_{z(t)}(\theta) = \int_0^\infty z(t)e^{-\theta t}dt = \frac{1}{\alpha}\frac{\alpha}{\alpha+\theta} = \frac{1}{\alpha+\theta}$. F is the distribution of two independent random variables $Exp(\alpha)+Exp(\beta)$, so the Laplace transform of F is: $\psi_F(\theta) = \frac{\alpha\beta}{(\alpha+\theta)(\beta+\theta)}$. The Laplace transform of renewal function U is given by:

$$\psi_U(\theta) = \frac{1}{1 - \frac{\alpha\beta}{(\alpha+\theta)(\beta+\theta)}} = \frac{(\alpha+\theta)(\beta+\theta)}{\theta(\alpha+\beta+\theta)}.$$

Therefore here we have transform of $q(t)$:

$$\psi_{q(t)}(\theta) = \psi_U(\theta) \cdot \psi_{z(t)}(\theta) = \frac{(\alpha+\theta)(\beta+\theta)}{\theta(\alpha+\beta+\theta)} \cdot \frac{1}{\alpha+\theta} = \frac{\beta+\theta}{\theta(\alpha+\beta+\theta)}.$$

But we were interested in process $Z_t = 1 - V_t$, so $p(t) = 1 - q(t)$ has Laplace transform:

$$\psi_{p(t)}(\theta) = \int_0^\infty p(t)e^{-\theta t}dt = \int_0^\infty e^{-\theta t}dt - \int_0^\infty q(t)e^{-\theta t}dt = \frac{1}{\theta} - \frac{\beta+\theta}{\theta(\alpha+\beta+\theta)} = \frac{\alpha}{\theta(\alpha+\beta+\theta)}.$$

It is easy to check that this is the Laplace transform of function $\frac{\alpha}{\alpha+\beta}(1 - e^{-(\alpha+\beta)t})$, hence

$$p(t) = P(Z_t = 1) = \frac{\alpha}{\alpha+\beta}(1 - e^{-(\alpha+\beta)t}).$$

Thus we have:

$$\delta_{(x_{min})}\mathbf{P}_t(x_{max}) = \prod_{i=1}^{d} \frac{\alpha_i}{\alpha_i+\beta_i}\left(1 - e^{-(\alpha_i+\beta_i)t}\right) = \pi(x_{max})\prod_{i=1}^{d}\left(1 - e^{-(\alpha_i+\beta_i)t}\right)$$

and a formula for separation distance:

$$s(\delta_{(x_{min})}\mathbf{P}_t, \pi) = 1 - \frac{\delta_{(x_{min})}\mathbf{P}_t(x_{max})}{\pi(x_{max})} = 1 - \prod_{i=1}^{d}\left(1 - e^{-(\alpha_i+\beta_i)t}\right).$$

But we also have

$$\delta_{(x_{min})}\mathbf{P}_t(x_{max}) = \prod_{i=1}^{d}\left(1 - e^{-(\alpha_i+\beta_i)t}\right) = \pi(x_{max})\prod_{i=1}^{d}\left(1 - e^{-(\alpha_i+\beta_i)t}\right) = \pi(x_{max})\sum_{k=0}^{d}(-1)^k\sum_{\gamma\in A_k} e^{-s_\gamma t}.$$

From the above expression for $\delta_{(x_{min})}\mathbf{P}_t(x_{max})$ and spectral representation (5.96):

$$\delta_{(x_{min})}\mathbf{P}_t(x_{max}) = \pi(x_{max}) + \sum_{j=2}^{d} f_j(x_{min})\pi_j(x_{max})e^{-s_j t}.$$

We see that the eigenvalues of $\mathbf{Q}$ are $\{-s_\gamma, \gamma \subset \{1, \ldots, d\}\}$ and thus the eigenvalues of $\mathbf{P} = \mathbf{I} - \mathbf{Q}$ are $\{1 - s_\gamma, \gamma \subset \{1, \ldots, d\}\}$ what proves (ii).
The spectral representation for discrete time (5.93):

$$\delta_{(x_{min})}\mathbf{P}^n(x_{max}) = \pi(x_{max}) + \sum_{j=2}^{d} f_j(x_{min})\pi_j(x_{max})(1-\lambda_j)^n.$$

We conclude that

$$s(\delta_{(x_{min})}\mathbf{P}^n, \pi) = 1 - \frac{\delta_{(x_{min})}\mathbf{P}^n(x_{max})}{\pi(x_{max})} = \sum_{k=1}^{d}(-1)^{k-1}\sum_{\gamma\in A_k}(1 - s_\gamma)^n,$$

i.e. (i).
From (ii) we have that the eigenvalues of $\mathbf{P}$ are $\{1 - s_\gamma, \gamma \subset \{1, \ldots, d\}\}$. From our additional assumption $\sum_{i=1}^{d}(\alpha_i + \beta_i) \le 1$ we have that all of them are non-negative, thus assertion (iii) follows from Lemma 5.1.13 and (iii) implies (iv) by Lemma 5.1.8. □

Remark: In the case $\sum_{i=1}^d (\alpha_i + \beta_i) \le 1$ we recognize (i) as an inclusion-exclusion formula. Consider n multinomial trials with cell probabilities $p_i = \alpha_i + \beta_i$, $i = 1, \ldots, d$ and $p_{d+1} = 1 - \sum_{i=1}^d (\alpha_i + \beta_i)$. Let A_n be the event that at least one of the cells from $1, \ldots, d$ is empty. Then $A_n = \cup_{i=1}^d C_i$, where $C_i = 1$ if cell i is empty, 0 otherwise. Then above expression represents inclusion-exclusion formula for $Pr(\cup_{i=1}^d C_i)$. Thus if we denote T to be the waiting time for all of cells $1, \ldots, d$ to be occupied, then

$$s(\delta_{x_{min}} \mathbf{P}^n, \pi) = P(T > n) = P(A_n).$$

Next lemma shows, that if we only observe every second step of random walk on cube, we always have $\delta_x \mathbf{P}^n(x)$ non-increasing in n without the assumption that $\sum_{i=1}^d (\alpha_i + \beta_i) \le 1$.

Lemma 5.4.4. *Let $\mathbf{Y} = \{Y_n, n \ge 0\}$ be random walk with transition matrix $\mathbf{R} = \mathbf{P}^2$, where $\mathbf{P}$ is defined as in (5.109). Assume (5.116) and (5.117). Then*

$$\begin{aligned} &(i) \quad \forall (x \in E) \qquad \delta_x \mathbf{P}^n(x) \text{ is non}-\text{increasing in } n \\ &(ii) \qquad \forall (x \in E) \qquad \psi_{\tau_\pi(x)} = \frac{\pi(x)}{1 - (1 - \pi(x))\psi_{W_x}}, \end{aligned}$$

Proof. (i) follows from Lemma 5.1.14, and this implies (ii) by Lemma 5.1.8. □

Continuous time

Theorem 5.4.5. *Let $\mathbf{X} = (X_t, t \ge 0)$ be a non-symmetric continuous time random walk on cube, i.e. its intensity matrix is given by $\mathbf{Q} := \mathbf{P} - \mathbf{I}$, where $\mathbf{P}$ is given in (5.109). Let initial distribution be $\mu := \delta_{x_{min}}$. Assume that*

$$(a) \qquad \alpha_i > 0, \beta_i > 0, \ i = 1, \ldots, d, \tag{5.123}$$

$$(b) \qquad \exists (A \subset \{1, 2, \ldots, d\}) \quad \sum_{i \in A} \alpha_i + \sum_{i \in A^c} \beta_i < 1. \tag{5.124}$$

Then

$$\begin{aligned} &(i) && x_{max} \text{ is ratio minimal for } \mu, \mathbf{X}, \\ &(ii) && \forall (t \ge 0) \quad s(\delta_{x_{min}} \mathbf{P}_t, \pi) = 1 - \frac{\delta_{x_{min}} \mathbf{P}_t(x_{max})}{\pi(x_{max})} = 1 - \prod_{i=1}^d \left(1 - e^{-(\alpha_i + \beta_i)t}\right), \\ &(iii) && \psi_{\tau_{\delta_{x_{min}}}(x_{max})} = \psi_{Y_{x_{max}}} \cdot \psi_{\tau_\pi(x_{max})}, \end{aligned}$$

where $Y_{x_{max}}$ is a random variable independent of $\{X_n\}$ with distribution $P(Y_{x_{max}} \le n) = \frac{\delta_{x_{min}} \mathbf{P}^n(x_{max})}{\pi(x_{max})}$.

Proof. In the proof of the Theorem 5.4.3 we had: for one 1-0 process (start at 0, in 0 $Exp(\alpha)$ and in 1 $Exp(\beta)$)

$$\delta_0 \mathbf{P}_t(1) = P(Z_t = 1) = \frac{\alpha}{\alpha + \beta}(1 - e^{-(\alpha+\beta)t}).$$

Thus for d such independent processes we had

$$\delta_{x_{min}} \mathbf{P}_t(x_{max}) = \prod_{i=1}^d \frac{\alpha_i}{\alpha_i + \beta_i}(1 - e^{-(\alpha_i+\beta_i)t}) = \pi(x_{max}) \prod_{i=1}^d (1 - e^{-(\alpha_i+\beta_i)t}).$$

Now if we look at process starting at 0, the probability that it is at time t at position 0 is:

$$\delta_0 \mathbf{P}_t(0) = 1 - P(Z_t = 1) = 1 - \frac{\alpha}{\alpha + \beta}(1 - e^{-(\alpha+\beta)t}) = \frac{\beta}{\alpha + \beta}\left(1 + \frac{\alpha}{\beta} e^{-(\alpha+\beta)t}\right).$$

Thus we can calculate for any x

$$\begin{aligned} \delta_{x_{min}} \mathbf{P}_t(x) &= \prod_{i:x_i=1} \frac{\alpha_k}{\alpha_k + \beta_k}\left(1 - e^{-(\alpha_k+\beta_k)t}\right) \prod_{i:x_i=0} \frac{\beta_k}{\alpha_k + \beta_k}\left(1 + \frac{\alpha_k}{\beta_k} e^{-(\alpha_k+\beta_k)t}\right) \\ &= \pi(x) \prod_{i:x_i=1}\left(1 - e^{-(\alpha_k+\beta_k)t}\right) \prod_{i:x_i=0}\left(1 + \frac{\alpha_k}{\beta_k} e^{-(\alpha_k+\beta_k)t}\right). \end{aligned} \tag{5.125}$$

Independently of the choice of $\alpha_i, \beta_i, i = 1, \ldots, d$ i.e. they just must fulfill (a) and (b) we have that x_{max} is ratio minimal for $\mu = \delta_{x_{min}}$, X, i.e.

$$\frac{\delta x_{min}\mathbf{P}^t(x_{max})}{\pi(x_{max})} \leq \frac{\delta x_{min}\mathbf{P}^t(x)}{\pi(x)}.$$

To check it just use (5.125)

$$\frac{\pi(x_{max})\prod_{i=1}^d(1-e^{-(\alpha_i+\beta_i)t})}{\pi(x_{max})} \leq \frac{\pi(x)\prod_{i:x_i=1}\left(1-e^{-(\alpha_k+\beta_k)t}\right)\prod_{i:x_i=0}\left(1+\frac{\alpha_k}{\beta_k}e^{-(\alpha_k+\beta_k)t}\right)}{\pi(x)},$$

$$\prod_{i=1}^d(1-e^{-(\alpha_i+\beta_i)t}) \leq \prod_{i:x_i=1}\left(1-e^{-(\alpha_k+\beta_k)t}\right)\prod_{i:x_i=0}\left(1+\frac{\alpha_k}{\beta_k}e^{-(\alpha_k+\beta_k)t}\right),$$

$$\prod_{i:x_i=0}(1-e^{-(\alpha_i+\beta_i)t}) \leq \prod_{i:x_i=0}\left(1+\frac{\alpha_k}{\beta_k}e^{-(\alpha_k+\beta_k)t}\right),$$

what is obviously always true, since all the coefficients on the left hand side are always ≤ 1 and all on the right hand side are always ≥ 1. □

Remark:
In Theorem 5.4.5 we proved that in continuous time we always have that x_{max} is ratio minimal for $\mu := \delta_{x_{min}}$, $\mathbf{X}$, i.e.

$$\frac{\delta x_{min}\mathbf{P}_t(x_{max})}{\pi(x_{max})} \leq \frac{\delta x_{min}\mathbf{P}_t(x)}{\pi(x)} \quad \text{for any } x \in \{0,1\}^d,$$

for any choice of $\alpha_i, \beta_i, i = 1, \ldots, d$ (they only must fulfill some natural conditions which guarantee irreducibility and aperiodicity).

In discrete time, we had a condition $\forall(B \subseteq \{1,2,\ldots,d\}, r \in B) \quad \sum_{k\in B}\alpha_k + \sum_{k\in B^c}\beta_k + \beta_r \leq 1$, which guaranteed it. It was only sufficient. One could think that this ratio minimality in discrete time also always hold, but this is not a case as shown in next

Example: Let $\mathbf{X}$ be a random walk on 3-dimensional cube with parameters.

$$\alpha_1 = \frac{3}{12}, \alpha_2 = \frac{4}{12}, \alpha_3 = \frac{4}{12}, \beta_1 = \frac{3}{12}, \beta_2 = \frac{4}{12}, \beta_3 = \frac{4}{12}$$

with initial distribution $\mu := \delta_{x_{min}}$.
Then the conditions (5.116) and (5.117) hold, the chain is ergodic. But note that the condition 5.118 does not hold, i.e. we do not have

$$\forall(B \subseteq \{1,2,\ldots,d\}, r \in B) \quad \sum_{k\in B}\alpha_k + \sum_{k\in B^c}\beta_k + \beta_r \leq 1,$$

because for $B = \{1,\ldots,d\}$ and $r = 3$ above equals to $\frac{15}{12} \geq 1$. In this case we have (calculated in MAPLE):

$$\frac{\delta_{x_{min}}\mathbf{P}^3((111))}{\pi(111)} = \frac{4}{3} > \frac{\delta_{x_{min}}\mathbf{P}^3((101))}{\pi(101)} = \frac{1}{3}$$

$$\frac{\delta_{x_{min}}\mathbf{P}^4((111))}{\pi(111)} = \frac{4}{9} < \frac{\delta_{x_{min}}\mathbf{P}^4((101))}{\pi(101)} = \frac{38}{27}$$

i.e. $x_{max} = (1,1,1)$ is not ratio minimal for $\mu, \mathbf{X}$.

5.5 Tandem

We consider continuous time Markov chain $\mathbf{X} = (X_t, t \geq 0)$ with state space
$E = \{(n_1, n_2, n_3) : n_1 + n_2 + n_3 = N, n_i \in \mathbb{N}, i = 1,2,3\}$.
The intensity matrix is following:

$$\mathbf{Q}((n_1,n_2,n_3),(m_1,m_2,m_3)) = \begin{cases} \mu_1 & \text{if} \quad n_1>0, m_1=n_1-1, m_2=n_2+1, m_3=n_3, \\ \mu_2 & \text{if} \quad n_2>0, m_1=n_1, m_2=n_2-1, m_3=n_3+1, \\ \mu_3 & \text{if} \quad n_3>0, m_1=n_1-1, m_2=n_2, m_3=n_3-1, \\ -\sum_{(r_1,r_2,r_3)\neq(n_1,n_2,n_3)} \mathbf{Q}((n_1,n_2,n_3),(r_1,r_2,r_3)) \text{ if } n_1=m_1, n_2=m_2, n_3=m_3 \\ 0 & \text{otherwise} \end{cases}.$$

This is a tandem process: there are 3 servers and N customers. The intensities of service at node i are μ_i and customer serviced at server i goes directly to the queue at server $(i \mod 3)+1$. We assume that waiting room at each server is of infinite capacity..
For $N=3$ the state space and possible transitions are depicted in Fig. 9.

We will consider embedded chain, i.e. observe at at moments when some change occurs. Then the probability matrix of this discrete time process is given by:

$$\mathbf{P} = \mathbf{I} + \frac{1}{c}\mathbf{Q}, \tag{5.126}$$

where $\mathbf{I}$ is identity matrix and $c \geq \sup_{n_1,n_2,n_3}(-\mathbf{Q}(n_1,n_2,n_3),(n_1,n_2,n_3))$, in our case $c \geq \mu_1+\mu_2+\mu_3$.

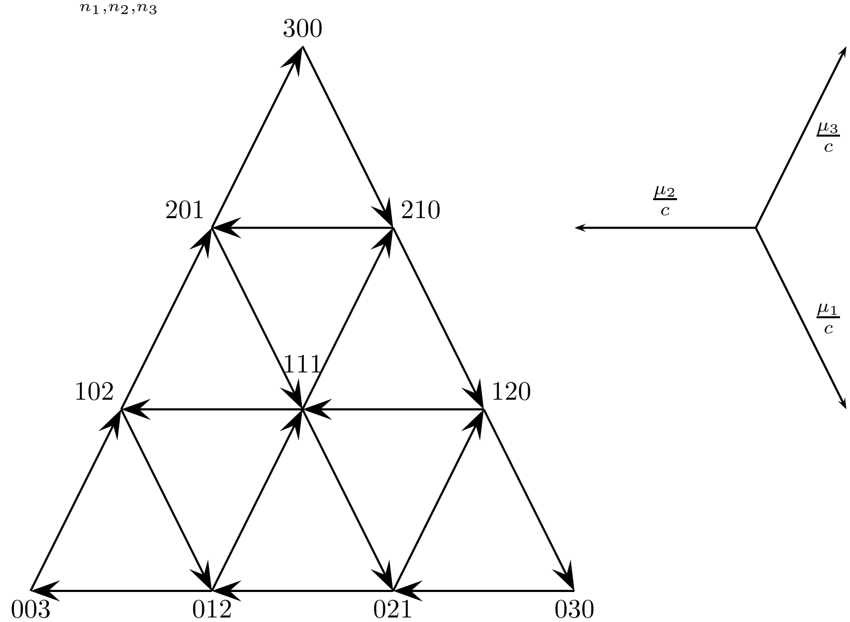


Figure 9: Closed Tandem, 3 servers, 3 customers

Denote process governed by $\mathbf{P}$ by $\mathbf{Z} = \{Z_n, n \geq 0\}$. The stationary distribution is

$$\pi(n_1,n_2,n_3) = K\rho_1^{n_1}\rho_2^{n_2}\rho_3^{n_3}, \ \rho_i = \frac{c}{\mu_i}, \ n_1+n_2+n_3 = N, \tag{5.127}$$

where K is normalisation constant.

We will need to have probability transition matrix of time-reversed process, i.e.:

$$\widetilde{\mathbf{P}}(x,y) = \frac{\pi(y)}{\pi(x)}\mathbf{P}(y,x).$$

$(x=(x_1,x_2,x_3), y=(y_1,y_2,y_3))$.
Of course the diagonal does not change: $\widetilde{\mathbf{P}}(x,x) = \frac{\pi(x)}{\pi(x)}\mathbf{P}(x,x) = \mathbf{P}(x,x)$.

$$\widetilde{\mathbf{P}}((n_1,n_2,n_3),(m_1,m_2,m_3)) = \frac{\pi((m_1,m_2,m_3))}{\pi((n_1,n_2,n_3))}\mathbf{P}((m_1,m_2,m_3),(n_1,n_2,n_3))$$

Consider three cases:

- $m_1 > 0, n_1 = m_1 - 1, n_2 = m_2 + 1, n_3 = m_3$:

$$\widetilde{\mathbf{P}}((m_1-1, m_2+1, m_3), (m_1, m_2, m_3)) = \frac{\pi((m_1, m_2, m_3))}{\pi((m_1-1, m_2+1, m_3))}\mathbf{P}((m_1, m_2, m_3), (m_1-1, m_2+1, n_3)) =$$

$$\frac{K\rho_1^{m_1}\rho_2^{m_2}\rho_3^{m_3}}{K\rho_1^{m_1-1}\rho_2^{m_2+1}\rho_3^{m_3}}\frac{\mu_1}{c} = \rho_1\frac{1}{\rho_2}\frac{1}{\rho_1} = \frac{1}{\rho_2} = \frac{\mu_2}{c}.$$

- $m_2 > 0, n_1 = m_1, n_2 = m_2 - 1, n_3 = m_3 + 1$:

$$\widetilde{\mathbf{P}}((m_1-1, m_2+1, m_3), (m_1, m_2, m_3)) = \frac{\pi((m_1, m_2, m_3))}{\pi((m_1, m_2-1, m_3+1))}\mathbf{P}((m_1, m_2, m_3), (m_1, m_2-1, n_3+1)) =$$

$$\frac{K\rho_1^{m_1}\rho_2^{m_2}\rho_3^{m_3}}{K\rho_1^{m_1}\rho_2^{m_2-1}\rho_3^{m_3+1}}\frac{\mu_2}{c} = \rho_2\frac{1}{\rho_3}\frac{1}{\rho_2} = \frac{1}{\rho_3} = \frac{\mu_3}{c}.$$

- $m_3 > 0, n_1 = m_1 + 1, n_2 = m_2, n_3 = m_3 - 1$:

$$\widetilde{\mathbf{P}}((m_1+1, m_2, m_3-1), (m_1, m_2, m_3)) = \frac{\pi((m_1, m_2, m_3))}{\pi((m_1+1, m_2, m_3-1))}\mathbf{P}((m_1, m_2, m_3), (m_1+1, m_2, n_3-1)) =$$

$$\frac{K\rho_1^{m_1}\rho_2^{m_2}\rho_3^{m_3}}{K\rho_1^{m_1+1}\rho_2^{m_2}\rho_3^{m_3-1}}\frac{\mu_3}{c} = \frac{1}{\rho_1}\rho_3\frac{1}{\rho_3} = \frac{1}{\rho_1} = \frac{\mu_1}{c}.$$

In other words, the time-reversed process dynamics correspond to the picture when we change the orientations of the arrows and change the μs: $\widetilde{\mu_1} = \mu_2, \widetilde{\mu_2} = \mu_3, \widetilde{\mu_3} = \mu_3$. The example for $N = 3$ is depicted on Fig. 10.

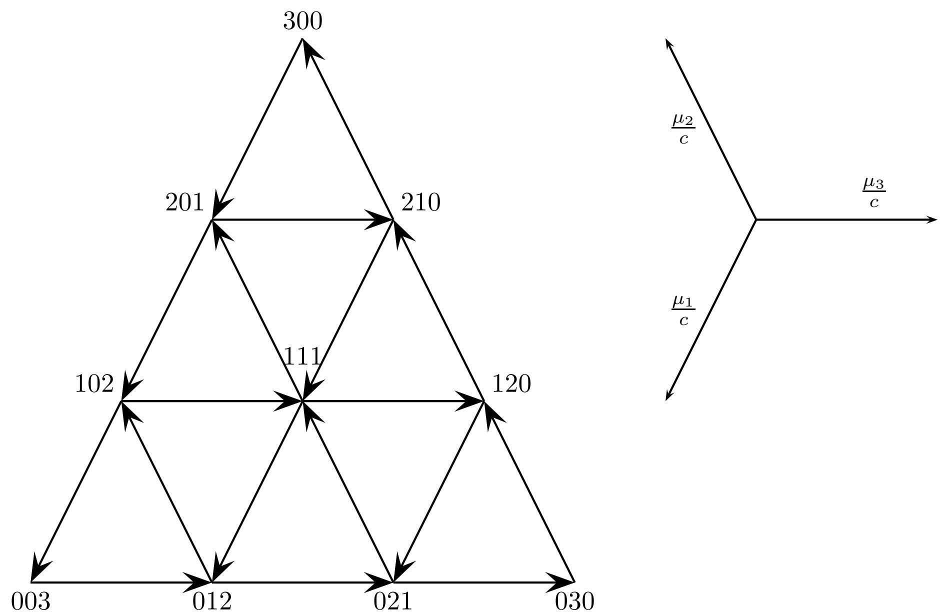


Figure 10: Time-reversed chain for tandem, 3 servers, 3 customers

Theorem 5.5.1. *Let $\mathbf{Z} = \{Z_n, n \geq 0\}$ be a embedded Markov chain for tandem queue process with transition matrix $\mathbf{P}$ given in (5.126). Take initial distribution $\mu := \delta_{x_{min}}$, where $x_{min} := (0, 0, N)$. Assume that $c \geq (\mu_1 + \mu_2 + \mu_3)$. Then*

$$(i) \qquad x_{max} := (N, 0, 0) \text{ is ratio minimal for } \mu, X, \tag{5.128}$$

$$(ii) \qquad \forall (n \geq 0) \qquad s(\delta_{(0,0,N)}\mathbf{P}^n, \pi) = 1 - \frac{\delta_{(0,0,N)}\mathbf{P}^n(N, 0, 0)}{\pi(N, 0, 0)} \tag{5.129}$$

$$= 1 - \tfrac{1}{K}\left(\tfrac{c}{\mu_1}\right)^N \delta_{(0,0,N)}\mathbf{P}^n(N, 0, 0),$$

$$(iii) \qquad \psi_{\tau_\mu(x_{max})} = \psi_{Y_{x_{max}}} \cdot \psi_{\tau_\pi(x_{max})} \quad (\tau_{\delta_{x_{min}}}(x_{max}) \stackrel{d}{=} Y_{x_{max}} + \tau_\pi(x_{max})), \tag{5.130}$$

where $Y_{x_{max}}$ is a random variable independent of $\mathbf{Z}$ with distribution $P(Y_{x_{max}} \le n) = \frac{\delta_{x_{min}}\mathbf{P}^n(x_{max})}{\pi(x_{max})}$ and K is a normalising constant from (5.127).

Proof. Consider partial sum ordering for this chain $\preceq$, i.e:

$$(n_1, n_2, n_3) \preceq (m_1, m_2, m_3) \quad \iff n_1 \le m_1, n_1 + n_2 \le m_1 + m_2, n_1 + n_2 + n_3 \le m_1 + m_2 + m_3.$$

Under this ordering there exist: unique minimal state $x_{min} = (0, 0, N)$ and unique maximal state $x_{max} = (N, 0, 0)$. With initial distribution $\mu := \delta_{x_{min}}$ the assumptions $A0 - A2$ of Theorem 5.3.10 are fulfilled. It is enough to have stochastic monotonicity of time-reversed process $\{\widetilde{X}_n, n \ge 0\}$ under this ordering. Having this we can conclude all three assertions from Theorem 5.3.10. Second equality in (ii) is obtained using $\pi(N, 0, 0) = K\left(\frac{c}{\mu_1}\right)^N$.

Recall that A is an upper set if: $x \preceq y, x \in A \Rightarrow y \in A$

Stochastic monotonicity of $\widetilde{\mathbf{P}}$ means:

$$\forall (x \preceq y) \quad \forall A - \text{upper set} \qquad \delta_x\widetilde{\mathbf{P}}(A) \le \delta_y\widetilde{\mathbf{P}}(A).$$

Consider 3 cases:

- $x \prec y$, $x \notin A, y \notin A$ and $\delta_x\widetilde{\mathbf{P}}(A) > 0$.
  The only possibility of getting from $x \notin A$ to A in one step is to go from $x = (x_1, x_3, x_3)$ to $x' = (x_1 + 1, x_2 - 1, x_3) \in A$ or to $x'' = (x_1, x_2 + 1, x_3 - 1) \in A$. Thus $\delta_x\widetilde{\mathbf{P}}(A)$ is at most $\frac{\mu_2+\mu_3}{c}$.
  But if $\delta_x\widetilde{\mathbf{P}}(A) = \frac{\mu_2+\mu_3}{c}$ then we cannot have y which is $y \succ x$ (i.e. and $y \succeq x$ and $y \neq x$) such that $y \notin A$.
  Thus we have $\delta_x\widetilde{\mathbf{P}}(A) = \frac{\mu_2}{c}$ or $\delta_x\widetilde{\mathbf{P}}(A) = \frac{\mu_3}{c}$.

  - If $\delta_x\widetilde{\mathbf{P}}(A) = \frac{\mu_2}{c}$ then possible ys : $y \succ x$ are of form $y = (x_1, x_2 + k, x_3 - k)$ for some $k \ge 1$. But then surely $y' = (y_1 + 1, y_2 - 1, y_3) \in A$ what means that $\delta_y\widetilde{\mathbf{P}}(A)$ is at least $\frac{\mu_2}{c}$ (it is possible that $\delta_y\widetilde{\mathbf{P}}(A) = \frac{\mu_2+\mu_3}{c}$).
  - If $\delta_x\widetilde{\mathbf{P}}(A) = \frac{\mu_3}{c}$ then possible ys: $y \succ x$ are of form $y = (x_1 + k, x_2 - k, x_3)$ for some $k \ge 1$. But then surely $y' = (y_1, y_2 + 1, y_3 - 1) \in A$ what means that $\delta_y\widetilde{\mathbf{P}}(A)$ is at least $\frac{\mu_3}{c}$ (again, it is possible to have $\delta_y\widetilde{\mathbf{P}}(A) = \frac{\mu_2+\mu_3}{c}$).

- $x \prec y$, $x \in A$ and $y \in A$
  The only possibility of getting out from A is to change from $x = (x_1, x_2, x_3)$ to $(x_1 - 1, x_2, x_3 + 1)$. Thus $\delta_x\widetilde{\mathbf{P}}(A) = 1 - \frac{\mu_1}{c}$. But then there are only to possibilities for $y \succ x$ such that $\delta_y\widetilde{\mathbf{P}}(A) > 0$: either $\delta_y\widetilde{\mathbf{P}}(A) = 1 - \frac{\mu_1}{c}$ or $\delta_y\widetilde{\mathbf{P}}(A) = 1$. Of course in both cases we have $\delta_x\widetilde{\mathbf{P}}(A) \le \delta_y\widetilde{\mathbf{P}}(A)$.
- $x \prec y$, $x \notin A$ and $y \in A$
  $\delta_x\widetilde{\mathbf{P}}(A)$ for $x \notin A$ can be at most $\frac{\mu_2+\mu_3}{c}$ (in case when $x' = (x_1 + 1, x_2 - 1, x_3) \in A$ and $x'' = (x_1, x_2 + 1, x_3 - 1) \in A$. The smallest value of $\delta_y\widetilde{\mathbf{P}}(A)$ for $y \in A$ is for y being on border of A, then $\delta_y\widetilde{\mathbf{P}}(A) = 1 - \frac{\mu_1}{c}$. Thus we always have $\delta_x\widetilde{\mathbf{P}}(A) \le \delta_y\widetilde{\mathbf{P}}(A)$, because $\frac{\mu_2+\mu_3}{c} \le 1 - \frac{\mu_1}{c}$, which follows from assumption that $c \ge (\mu_1 + \mu_2 + \mu_3)$.

□

Remark: In order to have an upper bound on mixing time for this process it is needed to calculate or estimate $\delta_{(0,0,N)}\mathbf{P}^n((N, 0, 0))$ for all $n \ge 1$.

Remark: The similar result can easily be formulated for continuous time process $\mathbf{X}$.

6 Spectral gap in queueing networks

6.1 Hazard rate function & Heavy Tail

Let us start with general definition of **heavy tail**

Definition 6.1.1. *Random variable X is* **heavy-tailed** *if*

$$\forall s > 0 \quad E[e^{sX}] = \infty$$

and **light-tailed** *if*

$$\exists s_0 > 0 \quad E[e^{s_0 X}] < \infty.$$

We make the following assumption: $F(0) = 0$.

Continuous case

Define hazard rate function for $t : \bar{F}(t) > 0$

$$h(t) = \frac{f(t)}{\bar{F}(t)}, \text{ and } h(t) = +\infty \text{ otherwise.}$$

Equivalently,

$$h(t) = (-\log(\bar{F}(t))'.$$

Define cumulative hazard rate function:

$$H(t) = \int_0^t h(s)ds.$$

We have

$$H(t) = \int_0^t h(s)ds = \int_0^t (-\log(\bar{F}(t))' = -\log(\bar{F}(t)) + \log \underbrace{\bar{F}(0)}_{=1} = -\log(\bar{F}(t)),$$

thus

$$\bar{F}(t) = e^{-H(t)} = e^{-\int_0^t h(s)ds}.$$

Discrete case

For random variable X distributed on $0, 1, \ldots$, define hazard rate function:

$$h(k) = \frac{P(X = k)}{\sum_{j=k}^{\infty} P(X = j)} = \frac{P(X = k)}{P(X \geq k)} = \frac{P(X > k-1) - P(X > k)}{P(X > k-1)} =$$

$$= \frac{\bar{F}(k-1) - \bar{F}(k)}{\bar{F}(k-1)} = 1 - \frac{\bar{F}(k)}{\bar{F}(k-1)}.$$

Define cumulative hazard rate function:

$$H(m) = \sum_{k=0}^{m} h(k) = \sum_{k=0}^{m} \left(1 - \frac{\bar{F}(k)}{\bar{F}(k-1)}\right).$$

We have

$$\bar{F}(m) = \prod_{j=1}^{m} \frac{\bar{F}(j)}{\bar{F}(j-1)} = \prod_{j=1}^{m} (1 - h(j))\,. \tag{6.131}$$

Next theorem characterizes heavy-tail distribution by hazard rates. We assume $\liminf_{k\to\infty} h(k) = \limsup_{k\to\infty} h(k)$ to make it simpler, this assumption can be relaxed allowing $\liminf_{k\to\infty} h(k) = \limsup_{k\to\infty} h(k)$, for such characterization in case $\liminf_{k\to\infty} h(k) \neq \limsup_{k\to\infty} h(k)$ see for example T. Rolski *et al.* [39].

Moreover, throughout the paper we assume that for any fixed $k > 0$ we have $h(k) > 0$ (what is not a restriction in most distributions we are interested in).

Lemma 6.1.2. *If the hazard rate h fulfills $\liminf\limits_{k\to\infty} h(k) = \limsup\limits_{k\to\infty} h(k)$, then*

$$\lim_{k\to\infty} h(k) = 0 \qquad \Longleftrightarrow \text{ X is heavy-tailed.}$$

Proof.

$(\Longrightarrow)$ Assume $\lim\limits_{k\to\infty} h(k) = 0$, what means

$$\forall(\varepsilon \in (0,1]) \quad \exists(k') \quad \forall(k \geq k') \quad h(k) < \varepsilon$$

and thus

$$\forall(k \geq k') \quad 1 - h(k) > 1 - \varepsilon.$$

And on the contrary assume X is light-tailed, this means $\exists s_0 > 0 : Ee^{s_0 X} < \infty$. Then:
Using (6.131) we obtain

$$\bar{F}(k) = \prod_{j=1}^{k}(1 - h(j)) = \prod_{j=1}^{k'-1}(1 - h(j)) \cdot \prod_{j=k'}^{k}(1 - h(j)) \geq c_2^{k'-1}(1-\varepsilon)^{k-k'},$$

where $c_2 = \min\limits_{1\leq j\leq k'-1}(1 - h(j))$. Set $c = c_2^{k'-1}(1-\varepsilon)^{-k'}$, then

$$\bar{F}(k) \geq c(1-\varepsilon)^k$$

Setting any $\varepsilon_0 = 1 - e^{-s_0+2} < 1 - e^{-s_0+1}$ we have (denoting finite sum $d = \sum\limits_{k=0}^{k'-1} e^{s_0 k}\bar{F}(k)$)

$$\sum_{k=0}^{\infty} e^{s_0 k}\bar{F}(k) = \sum_{k=0}^{k'-1} e^{s_0 k}\bar{F}(k) + \sum_{k=k'}^{\infty} e^{s_0 k}\bar{F}(k) \geq d + c\sum_{k=k'}^{\infty} e^{s_0 k}(1-\varepsilon_0)^k = d + c\sum_{k=k'}^{\infty} e^{s_0 k} e^{k\cdot\log(1-\varepsilon_0)}$$

$$= d + c\sum_{k=k'}^{\infty} e^{(s_0+\log(1-\varepsilon_0))k} \geq d + c\sum_{k=k'}^{\infty} e^{(s_0+\log(e^{-s_0+2}))k} = d + c\sum_{k=k'}^{\infty} e^{2k} = \infty,$$

but

$$\sum_{k=0}^{\infty} e^{s_0 k}\bar{F}(k) = \sum_{k=0}^{\infty} e^{s_0 k} P(X > k) = \sum_{k=0}^{\infty} e^{s_0 k} \sum_{m=k+1}^{\infty} P(X = m) = \sum_{m=1}^{\infty}\sum_{k=0}^{m-1} e^{sk} P(X = m) =$$

$$\sum_{m=1}^{\infty} P(X = m)\sum_{k=0}^{m-1} e^{s_0 k} = \sum_{m=1}^{\infty} P(X = m)\frac{e^{s_0 m} - 1}{e_0^s - 1} =$$

$$\frac{1}{e^{s_0} - 1}\left[\sum_{m=0}^{\infty} e^{s_0 m} P(X = m) - e^0 P(X = 0) - (1 - P(X = 0))\right] = \frac{Ee^{s_0 X} - 1}{e^{s_0} - 1},$$

i.e. we have

$$Ee^{s_0 X} = 1 + (e^{s_0} - 1)\sum_{k=0}^{\infty} e^{s_0 k}\bar{F}(k) \tag{6.132}$$

and thus $Ee^{s_0 X} = \infty$ what contradicts the finiteness of $Ee^{s_0 X}$ or equivalently, contradicts that X is heavy-tailed.

($\Longleftarrow$) Assume X is heavy-tailed, i.e. $\forall(s>0): Ee^{sX}=\infty$ and in contradiction assume $\lim_{k\to\infty} h(k)>0$, this means

$$\forall(\varepsilon\in(0,1)) \quad \exists(k') \quad \forall(k\geq k') \quad h(k)>\varepsilon$$

i.e.

$$\forall(\varepsilon\in(0,1)) \quad \exists(k') \quad \forall(k\geq k') \quad 1-h(k)\leq 1-\varepsilon$$

Using (6.131) we obtain

$$\bar{F}(k)=\prod_{j=1}^{k}(1-h(j))=\prod_{j=1}^{k'-1}(1-h(j))\cdot\prod_{j=k'}^{k}(1-h(j))\leq c_2^{k'-1}(1-\varepsilon)^{k-k'},$$

where $c_2=\max_{1\leq j\leq k'-1}(1-h(j))$. Set $c=c_2^{k'-1}(1-\varepsilon)^{-k'}$, then

$$\bar{F}(k)\leq c(1-\varepsilon)^k$$

Setting $s_0=\log\left(\frac{1}{\sqrt{1-\varepsilon}}\right)>0$ we have

$$\sum_{k=0}^{\infty}e^{s_0k}\bar{F}(k)=\sum_{k=0}^{k'-1}e^{s_0k}\bar{F}(k)+\sum_{k=k'}^{\infty}e^{s_0k}\bar{F}(k)\leq d+c\sum_{k=k'}^{\infty}e^{s_0k}(1-\varepsilon)^k=d+c\sum_{k=k'}^{\infty}e^{s_0k}e^{k\cdot\log(1-\varepsilon)}$$

$$=d+c\sum_{k=k'}^{\infty}e^{(s_0+\log(1-\varepsilon))k}=d+c\sum_{k=k'}^{\infty}e^{\left(\log\left(\frac{1}{\sqrt{1-\varepsilon}}\right)+\log(1-\varepsilon)\right)k}=d+c\sum_{k=k'}^{\infty}e^{\left(\log\sqrt{1-\varepsilon}\right)k}=M<\infty,$$

because $\log\sqrt{1-\varepsilon}<0$ but from (6.132) and assumption that $Ee^{s_0X}=\infty$ for every $s>0$ we have

$$\infty=Ee^{s_0X}=1+(e^{s_0}-1)\sum_{k=0}^{\infty}e^{s_0k}\bar{F}(k)\leq 1+(e^{s_0}-1)M<\infty$$

which is in contradiction to the fact that X is heavy-tailed.

□

6.2 Markov Jump Process

Consider a Markov Jump process $\mathbf{X}=(X_t,t\geq 0)$ with enumerable state space E and intensity matrix J. Corresponding infinitesimal generator $-L$ can be defined by:

$$Lf(\mathbf{e})=\sum_{\mathbf{e}'\in E}[f(\mathbf{e})-f(\mathbf{e}')]J(\mathbf{e},\mathbf{e}'), \quad \mathbf{e}\in E.$$

Definition 6.2.1. *We call Markov jump process reversible if*

$$\forall(\mathbf{e},\mathbf{e}'\in E) \qquad \pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')=\pi(\mathbf{e}')J(\mathbf{e}',\mathbf{e}).$$

The scalar product on $L_2(\pi)$ is given by

$$(f,g)_\pi=\sum_{\mathbf{e}}f(\mathbf{e})g(\mathbf{e})\pi(\mathbf{e}), \qquad ||f||_\pi^2=(f,f)_\pi.$$

Lemma 6.2.2. *For a reversible Markov jump process and f satisfying $(f,\mathbf{1})_\pi=0$ ($\mathbf{1}$ denotes the constant function with value 1) we have*

$$(f,Lf)_\pi=\frac{1}{2}\sum_{\mathbf{e}}\sum_{\mathbf{e}'}(f(\mathbf{e})-f(\mathbf{e}'))^2\pi(\mathbf{e})J(\mathbf{e},\mathbf{e}').$$

Proof. Use $\forall(\mathbf{e})\int_{\mathbf{e}'}J(\mathbf{e},\mathbf{e}')=0$ and $\forall(\mathbf{e}')\int_{\mathbf{e}}\pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')=0.$. □

Definition 6.2.3 (Spectral Gap)**.** *Suppose $-L$ is the infinitesimal generator of* $\mathbf{X}$.
As **spectral gap** *we define*

$$Gap(L):=\inf\left\{-(f,Lf)_\pi:||f||_\pi=1,(f,\mathbf{1})_\pi=0\right\}. \tag{6.133}$$

Define

$$M = \sup_{\mathbf{e}\in E} J(\mathbf{e}, \{\mathbf{e}\}^c).$$

Definition 6.2.4. Cheeger's constant *is defined as follows:*

$$k := \inf_{A\subset E, \pi(A)\in(0,1)} k(A), \qquad k(A) := \frac{\int_{x\in A}\pi(dx)J(x,A^c)}{\pi(A)\pi(A^c)}. \tag{6.134}$$

Lemma 6.2.5 (Lawler & Sokal [30])**.** *For a reversible Markov process with infinitesimal generator $-L$ we have*

$$Gap(L) \le k.$$

Proof. Consider

$$f(\mathbf{e}) = \begin{cases} \pi(A^c), & \mathbf{e}\in A, \\ -\pi(A), & \mathbf{e}\in A^c. \end{cases}$$

We have

$$(f,\mathbf{1})_\pi = \sum_{\mathbf{e}} f(\mathbf{e})\pi(\mathbf{e}) = \sum_{\mathbf{e}\in A} f(\mathbf{e})\pi(\mathbf{e}) + \sum_{\mathbf{e}\in A^c} f(\mathbf{e})\pi(\mathbf{e}) = \pi(A)\pi(A^c) - \pi(A^c)\pi(A) = 0.$$

$$(f,f)_\pi = \sum_{\mathbf{e}} f^2(\mathbf{e})\pi(\mathbf{e}) = \sum_{\mathbf{e}\in A} f^2(\mathbf{e})\pi(\mathbf{e}) + \sum_{\mathbf{e}\in A} f^2(\mathbf{e})\pi(\mathbf{e}) = \pi(A)\pi^2(A^c) + \pi(A^c)\pi^2(A) = \pi(A)\pi(A^c).$$

Using Lemma 6.2.2:

$$(f,Lf)_\pi = \frac{1}{2}\sum_{\mathbf{e}}\sum_{\mathbf{e}'}(f(\mathbf{e})-f(\mathbf{e}'))^2\pi(\mathbf{e})J(\mathbf{e},\mathbf{e}') =$$

$$\frac{1}{2}\left(0+0+\sum_{\mathbf{e}\in A}\sum_{\mathbf{e}'\in A^c} 1\cdot\pi(\mathbf{e})J(\mathbf{e},\mathbf{e}') + \sum_{\mathbf{e}\in A^c}\sum_{\mathbf{e}'\in A} 1\cdot\pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')\right) \stackrel{revers.}{=} \sum_{\mathbf{e}\in A}\pi(\mathbf{e})J(\mathbf{e},A^c).$$

Thus

$$k(A) = \frac{(f,Lf)_\pi}{(f,f)_\pi}$$

and because it was just a special function, thus

$$k = \inf_{A\subset E,\pi(A)\in(0,1)} k(A) = \inf_{A\subset E,\pi(A)\in(0,1)}\left\{\frac{(f,Lf)_\pi}{(f,f)_\pi}\right\} \ge Gap(L).$$

□

Before stating next theorem we define

$$\eta = \inf_{\mathfrak{L}}\sup_{c}\frac{(E|(X+c)^2-(Y+c)^2|)^2}{E(X+c)^2},$$

where the infimum is taken over all distributions of i.i.d r.v. (X,Y) with variance 1. It is not obvious that $\eta \ne 0$, but indeed

$$\eta \ge 1, \tag{6.135}$$

as proved in Lawler & Sokal [30].

Theorem 6.2.6 (Lawler & Sokal [30])**.** *For any reversible, stationary Markov jump process with bounded infinitesimal generator $-L$ and stationary measure π we have:*

$$Gap(L) \ge \frac{\eta k^2}{8M} \ge \frac{k^2}{8M}.$$

Proof. Let $f \in L^2(\pi)$, and $g = f + c$ (constant c will be determined later). Using Lemma 6.2.2 and the Schwarz inequality we have:

$$(f,Lf)_\pi = \frac{1}{2}\sum_{\mathbf{e},\mathbf{e}'}\pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')[g(\mathbf{e})-g(\mathbf{e}')]^2 \ge \frac{1}{2}\frac{\left(\sum_{\mathbf{e},\mathbf{e}'}\pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')[g^2(\mathbf{e})-g^2(\mathbf{e}')]^2\right)^2}{\sum_{\mathbf{e},\mathbf{e}'}\pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')[g(\mathbf{e})+g(\mathbf{e}')]^2}$$

$$\ge \frac{1}{2}\frac{\left(\sum_{\mathbf{e},\mathbf{e}'}\pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')[g^2(\mathbf{e})-g^2(\mathbf{e}')]^2\right)^2}{\sum_{\mathbf{e},\mathbf{e}'}\pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')[2g^2(\mathbf{e})+2g^2(\mathbf{e}')]} \ge \frac{1}{2}\frac{\left(\sum_{\mathbf{e},\mathbf{e}'}\pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')[g^2(\mathbf{e})-g^2(\mathbf{e}')]^2\right)^2}{4M\sum_{\mathbf{e}}\pi(\mathbf{e})g^2(\mathbf{e})}. \tag{6.136}$$

By reversibility

$$\sum_{\mathbf{e},\mathbf{e}'} \pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')[g^2(\mathbf{e}) - g^2(\mathbf{e}')] = 2 \sum_{\substack{\mathbf{e},\mathbf{e}':\\ g^2(\mathbf{e})>g^2(\mathbf{e}')}} \pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')[g^2(\mathbf{e}) - g^2(\mathbf{e}')].$$

$$= 2\int_0^\infty d\alpha \sum_{\substack{\mathbf{e},\mathbf{e}':\\ g^2(\mathbf{e})>\alpha\geq g^2(\mathbf{e}')}} \pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')[g^2(\mathbf{e}) - g^2(\mathbf{e}')] = 2\int_0^\infty d\alpha - (\mathbf{I}_{(A^c)}, L\mathbf{I}_{(A^c)})_\pi$$

(where $\mathbf{I}$ is the indicator function, and $A_\alpha := \{\mathbf{e} : g^2(\mathbf{e}) > \alpha\}$)

$$\geq 2\int_0^\infty d\alpha k\pi(A_\alpha)\pi(A_\alpha^c) = 2k\int_0^\infty d\alpha \sum_{\substack{\mathbf{e},\mathbf{e}':\\ g^2(\mathbf{e})>\alpha\geq g^2(\mathbf{e}')}} \pi(\mathbf{e})\pi(\mathbf{e}')$$

$$= 2k\int_0^\infty \sum_{\substack{\mathbf{e},\mathbf{e}':\\ g^2(\mathbf{e})>g^2(\mathbf{e}')}} \pi(\mathbf{e})\pi(\mathbf{e}')(g^2(\mathbf{e}) - g^2(\mathbf{e}')) = k\sum_{\mathbf{e},\mathbf{e}'} \pi(\mathbf{e})\pi(\mathbf{e}')|g^2(\mathbf{e}) - g^2(\mathbf{e}')|$$

i.e.

$$\sum_{\mathbf{e},\mathbf{e}'} \pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')|g^2(\mathbf{e}) - g^2(\mathbf{e}')| \geq k\sum_{\mathbf{e},\mathbf{e}'} \pi(\mathbf{e})\pi(\mathbf{e}')|g^2(\mathbf{e}) - g^2(\mathbf{e}')| \tag{6.137}$$

Combining (6.136) and (6.137) we have

$$(f, Lf)_\pi \geq \frac{k^2}{8M} \frac{\left(\sum_{\mathbf{e},\mathbf{e}'} \pi(\mathbf{e})J(\mathbf{e},\mathbf{e}')[g^2(\mathbf{e}) - g^2(\mathbf{e}')]^2\right)^2}{\sum_{\mathbf{e}\in E} \pi(\mathbf{e})g^2(\mathbf{e})}$$

We now optimize the choice of c. By definition of η we have

$$(f, Lf)_\pi \geq \frac{\eta k^2}{8M}\left[\sum_{\mathbf{e}\in E} f^2(\mathbf{e})\pi(\mathbf{e}) - \left(\sum_{\mathbf{e}\in E} f(\mathbf{e})\pi(\mathbf{e})\right)\right]$$

In particular, if $(f, \mathbf{1})_\pi = 0$:

$$(f, Lf)_\pi \geq \frac{\eta k^2}{8M}||f||_\pi^2,$$

which (using (6.135)) implies

$$Gap(L) \geq \frac{\eta k^2}{8M} \geq \frac{k^2}{8M}.$$

□

Lemma 6.2.7 (Liggett [31]). *Let $-\widetilde{L}$ be the infinitesimal generator of a vector Markov process with stationary distribution π whose components are independent Markov processes on the state space $E = E_1 \times E_2 \times E_3 \times \ldots$, whose components are independent Markov processes: on $i-th$ component there is Markov process with generator $-L_i$, state space E_1 and and invariant probability measure. Let π be the product of π_i's. Then*

$$Gap(\widetilde{L}) = \inf_i Gap(L_i).$$

Proof. To show that $Gap(\widetilde{L}) \leq Gap(L_i)$ for $i = 1, \ldots, m$, simply take in definition of spectral gap (6.133) as functions f, functions which depend only on the i-th coordinate.

For $Gap(\widetilde{L}) \geq Gap(L_i)$, it is enough to consider the case with two components only, for then iterating this proof gives it for infinitely many components, since functions which depend on finitely many coordinates are dense in $L_2(\pi)$.
Thus we consider L_1, L_2 and coordinates x and y, number them 1 and 2. Set $\varepsilon = \min(Gap(L_1), Gap(L_2))$. Let f be the function on the product space which satisfies $\int f d\pi = 0$ (i.e. $(f, \mathbf{1})_\pi = 0$) and $||f||_\pi = 1$. Write

$$f(x, y) = h(x, y) + h_1(x) + h_2(y),$$

where $\int h(x, y)d\pi_1$ for a.e. y, $\int h(x, y)d\pi_2 = 0$ for a.e. x, $\int h_1(x)d\pi_1 = 0, \int h_2(x)d\pi_2 = 0$. Then h, h_1 and h_2 are orthogonal in $L_2(\pi)$, so that

$$||h||_\pi^2 + ||h_1||_\pi^2 + ||h_2||_\pi^2 = ||f||_\pi^2 = 1. \tag{6.138}$$

So are $P_t h, P_t h_1$ and $P_t h_2$:

$$||P_t h||_\pi^2 + ||P_t h_1||_\pi^2 + ||P_t h_2||_\pi^2 = ||P_t f||_\pi^2. \tag{6.139}$$

Since $P_t h_1 = P_t^{(1)} h_1$, $P_t h_2 = P_t^{(2)} h_2$ and fact that $Gap(L)$ is the largest ε' for which $||P_t g||_\pi \le e^{-\varepsilon' Gap(L)} ||g||_\pi$ (see subsection 6.5.3), we have

$$||P_t h_1||_\pi \le e^{-\varepsilon t} ||h_1||_\pi, \quad ||P_t h_2||_\pi \le e^{-\varepsilon t} ||h_2||_\pi. \tag{6.140}$$

On the other hand:

$$||P_t h||_\pi = ||P_t^{(1)} P_t^{(2)} h||_\pi \le e^{-\varepsilon t} ||P_t^{(2)} h||_\pi \le e^{-2\varepsilon t} ||h||_\pi \tag{6.141}$$

Combining (6.138)-(6.141) we get

$$||P_t f||_\pi^2 \le e^{-4\varepsilon t} ||h||_\pi^2 + e^{-2\varepsilon t} ||h_1||_\pi^2 + e^{-2\varepsilon t} ||h_2||_\pi^2 \le e^{-2\varepsilon t}$$

i.e.

$$||P_t f||_\pi \le e^{-\varepsilon t},$$

which means that $Gap(\widetilde{L}) \ge \varepsilon$. □

6.3 Markov chains on $\mathbb{Z}_+$

We consider a continuous time Markov chain $\mathbf{X} = (X_t, t \ge 0)$ on the state space $E = \mathbb{Z}_+$, with transition rates $q_{x,y}$. Infinitesimal generator $-\Omega$ of process on $L^2(\mathbb{N}, \pi)$ is given by

$$\Omega f(x) = \sum_y [f(x) - f(y)] q_{x,y}$$

for $f \in G$ - the set of functions on $\mathbb{Z}_+$ with finite support. Let π be the stationary distribution of the process. Assume that

$$\sum_{x \in E} |q_{x,x}| \pi(x) < \infty.$$

Then $\pi(x) q_{x,y} = 0$ for all $y \in E$.

Lemma 6.3.1. *If $f \in G$ then*

$$\sum_x f(x) \Omega f(x) \pi(x) = -\frac{1}{2} \sum_{x,y} q_{x,y} [f(y) - f(x)]^2 \pi(x),$$

Thus

$$Gap(\Omega) = \frac{1}{2} \inf \left\{ \sum_{x,y} [f(y) - f(x)]^2 q_{x,y} \pi(x) : f \in G, \sum_x f(x) \pi(x) = 0, \sum_x f^2(x) \pi(x) = 1 \right\}$$

Proof. Use $\forall (x) \sum_y q_{x,y} = 0$ and $\forall (y) \sum_x \pi(x) q_{x,y} = 0$ □

Theorem 6.3.2 (Liggett [31])**.**

$$Gap(\Omega) \le \frac{1}{2} \inf_{n \ge 0} \frac{\sum_{x \le n \le y} [\pi(x) q_{x,y} + \pi(y) q_{y,x}]}{(\sum_{x \le n} \pi(x))(\sum_{x > n} \pi(x))},$$

and if process is reversible, then $\frac{1}{2} \inf_{n \ge 0} \frac{\sum_{x \le n \le y} \left[\pi(x) q_{x,y} + \pi(y) q_{y,x}\right]}{(\sum_{x \le n} \pi(x))(\sum_{x > n} \pi(x))} = \inf_{n \ge 0} \frac{\sum_{x \le n \le y} \left[\pi(x) q_{x,y}\right]}{(\sum_{x \le n} \pi(x))(\sum_{x > n} \pi(x))}$

Proof. In Lemma 6.3.1 use $f(x) = c \cdot K(x) - d$, where $K(x) := 1_{\{x \in A_n\}}$, $A_n := \{0, 1, \ldots, n\}$. Constants c and d are chosen so that $\sum f(x) \pi(x) = 0, \sum f^2(x) \pi(x) = 1$.

$$\begin{aligned} \sum_{x,y} q_{x,y} \left[f(y) - f(x)\right]^2 \pi(x) &= \sum_{x \in A_n, y \in A_n} 0 + \sum_{x \in A_n, y \in A_n^c} c^2 q_{x,y} \pi(x) + \sum_{x \in A_n^c, y \in A_n} c^2 q_{y,x} \pi(y) \\ &= c^2 \sum_{x \le n \le y} \left[\pi(x) q_{x,y} + \pi(y) q_{y,x}\right]. \end{aligned}$$

What is left is to calculate c^2. It is done from:

$$\sum f(x) \pi(x) = \sum_{x \le n} (c - d) \pi(x) + \sum_{x > n} -d \pi(x) = 0$$

and

$$\sum f^2(x) \pi(x) = \sum_{x \le n} (c - d)^2 \pi(x) + \sum_{x > n} d^2 \pi(x) = 1$$

obtaining:

$$c^2 = \left(\sum_{x \le n} \pi(x) \right) \left(\sum_{x > n} \pi(x) \right).$$

□

6.4 Birth & Death process

Again we take $E = \mathbb{Z}_+$.
Birth & Death (B & D) process $\mathbf{X} = (X_t, t \geq 0)$ is a Markov chain on $\mathbb{Z}_+$ such that $q_{x,y} = 0$ if $|x - y| > 1$. Thus we only have birth rates $q_{j,j+1} = \lambda(j)$ and death rates $q_{j,j-1} = \mu(j)$.
Infinitesimal generator $-\Omega$ of such process on $L^2(\mathbb{N}, \pi)$ is given by

$$\Omega f(j) := [f(j) - f(j+1)]\lambda(j) + [f(j) - f(j-1)]\mu(j). \tag{6.142}$$

Similarly we consider $\mathbf{Q} = [q_{x,y}]_{x,y\in\mathbb{N}}$, the intensity matrix:

$$q_{x,y} = \begin{cases} \lambda(x) & \text{if } y = x+1 \\ \mu(y) & \text{if } y = x-1 \\ -\mu(y) - \lambda(y) & \text{if } y = x \\ 0 & \text{otherwise} \end{cases}$$

Define

$$m(0) := 1, m(n) := \frac{\lambda(0)\dots\lambda(n-1)}{\mu(1)\dots\mu(n)}.$$

Stationary distribution for a B & D process $\mathbf{X}$ is then given by

$$\pi(i) = \frac{m(i)}{m}, \text{ where } m := \sum_{i=0}^{\infty} m(i).$$

The proof of next theorem uses the Schwarz inequality and can be found in Liggett [31].

Theorem 6.4.1 (Liggett [31]). *Assume that $\mathbf{X}$ is a Birth & Death process with $q_{i,i+1} > 0$, for all $i \geq 0$ and for some $b, c > 0$ we have*

$$\sum_{j>i} \pi(j) \leq c\pi(i)q_{i,i+1} \quad \text{and} \quad \sum_{j>i} \pi(j)q_{j,j+1} \leq b\pi(i)q_{i,i+1}.$$

Then

$$Gap(\Omega) \geq \frac{(\sqrt{b+1} - \sqrt{b})^2}{c} \geq \frac{1}{2c(1+2b)}.$$

□

Let us calculate hazard rate of stationary distribution of birth and death process:

$$h(i) = \frac{\pi(i)}{\sum_{j\geq i}\pi(j)} = \frac{\frac{m(i)}{m}}{\sum_{j\geq i}\frac{m(j)}{m}} = \frac{m(i)}{\sum_{j\geq i} m(j)} = \frac{\lambda(0)\cdots\lambda(i-1)}{\mu(1)\cdots\mu(i)} \frac{1}{\sum_{j\geq i}\frac{\lambda(0)\cdots\lambda(j-1)}{\mu(1)\cdots\mu(j)}} = \frac{1}{1 + \sum_{j>i}\frac{\lambda(i)\cdots\lambda(j-1)}{\mu(i+1)\cdots\mu(j)}}$$

Next theorem is reformulation of theorem of Chen, [6] into hazard rates.

Theorem 6.4.2 (Chen [6]). *If $0 < \inf_i \lambda(i) \leq \sup_i \lambda(i) =: C < \infty$ then*

$$Gap(\Omega) > 0 \qquad \Longleftrightarrow \qquad \inf_{i\geq 0} h(i) > 0$$

Proof.
($\Longleftarrow$) Assume $\inf_{i\geq 0} h(i) > 0$.
It means $\sup_{i\geq 0} \frac{1}{h(i)} > 0$.
We have

$$\sup_{i\geq 0}\frac{1}{h(i)} = \sup_{i\geq 0}\left(1 + \sum_{j>i}\frac{\lambda(i)\cdots\lambda(j-1)}{\mu(i+1)\cdots\mu(j)}\right) = \sup_{i\geq 0}\left(1 + \frac{\lambda(i)}{\mu(i+1)} + \sum_{j>i+1}\frac{\lambda(i)\cdots\lambda(j-1)}{\mu(i+1)\cdots\mu(j)}\right) < \infty.$$

Because of assumption it means that

$$\sup_{i\geq 0}\left(\frac{1}{\mu(i+1)} + \sum_{j>i+1}\frac{\lambda(i)\cdots\lambda(j-1)}{\mu(i+1)\cdots\mu(j)}\right) < \infty$$

what is equivalent to

$$\sup_{i\geq 0}\left(\sum_{j>i+1}\frac{\lambda(i+1)\dots\lambda(j-1)}{\mu(i+1)\dots\mu(j)}\right)=\sup_{i\geq 0}\left(\sum_{j>i}\frac{m(j)}{m(i)\lambda(i)}\right)<\infty.$$

Therefore for every i we have

$$\sum_{j>i}\frac{\pi(j)}{\lambda(i)\pi(i)}\leq c<\infty,\quad \text{i.e.}\quad \sum_{j>i}\pi(j)\leq c\pi(i)\lambda(i),$$

and this is the assumption of Theorem 6.4.1, thus $Gap(\Omega)>0$.

($\Longrightarrow$) Assume $Gap(\Omega)>0$
From Theorem 6.3.2 we have

$$0<Gap(\Omega)\leq\inf_{n\geq 0}\frac{\sum_{i\leq n\leq j}\pi(i)q(i,j)}{(\sum_{i\leq n}\pi(i))(\sum_{i>n}\pi(i))}=\inf_{n\geq 0}\frac{\pi(n)q(n,n+1)}{(\sum_{i\leq n}\pi(i))(\sum_{i>n}\pi(i))}$$

$$\leq C\inf_{n\geq 0}\frac{\pi(n)}{(\sum_{i\leq n}\pi(i))(\sum_{i>n}\pi(i))}\leq C\inf_{n\geq 0}\frac{\pi(n)}{(\pi(0))(\sum_{i>n}\pi(i))}=\frac{C}{\pi(0)}\inf_{n\geq 0}\frac{\pi(n)}{(\sum_{i>n}\pi(i))}$$

i.e.

$$\inf_{i\geq 0}\frac{\pi(i)}{(\sum_{j>i}\pi(j))}>0,$$

which is of course equivalent to

$$\inf_{i\geq 0}\frac{\pi(i)}{(\sum_{j\geq i}\pi(j))}=\inf_{i\geq 0}h(i)>0.$$

□

Remark 6.4.3.

For constant birth intensities, i.e. $\lambda(i)\equiv\lambda$ we can easily have any distribution $\{p_i\}$ as a stationary distribution, as can be seen in next Theorem.

Theorem 6.4.4. *Consider a Birth & Death process* $\mathbf{X}$ *with* $\lambda(k)\equiv\lambda,\mu(k)=\lambda\cdot\frac{p_{k-1}}{p_k}(p_0\equiv 1)$. *Let* h *be a hazard rate of its statinary distribution and* $-\Omega$ *its infinitesimal generator. Assume that* $\liminf_{k\to\infty}h(k)=\limsup_{k\to\infty}h(k)$ *Then*

$$\begin{array}{lc} (i) & \pi(i)=p_i,\\ (ii) & \{p_i\}\ \textit{is heavy-tailed}\iff Gap(\Omega)=0,\\ (iii) & \{p_i\}\ \textit{is light-tailed}\iff Gap(\Omega)>0.\end{array}$$

Proof.
(i)

$$\pi(i)=\frac{\lambda\cdots\lambda}{\lambda\cdots\lambda\frac{p_0}{p_1}\frac{p_1}{p_2}\cdots\frac{p_{i-1}}{p_i}}=\frac{\lambda^i}{\lambda^i\frac{1}{p_i}}=p_i.$$

(ii) If π is heavy-tailed, then by Lemma 6.1.2 we have that $\lim_{i\to\infty}h(i)=0$ what implies $\inf_{i\geq 0}h(i)=0$ (because we assumed also throughout the paper that $h(i)>0$), what holds if and only if $Gap(\Omega)=0$ (Theorem 6.4.2).
(ii) If π is light-tailed, then by Lemma 6.1.2 we have that $\lim_{i\to\infty}h(i)>0$ what implies $\inf_{i\geq 0}h(i)>0$, what holds if and only if $Gap(\Omega)>0$ (again, Theorem 6.4.2).

□

It is worth noting that for general $\lambda(i)$ and $\mu(i)$ the fact that stationary distribution is heavy-tailed does not imply that spectral gap is equal to zero.

Example (Chen & Wang [7]). Let $E=\mathbb{Z}_+$ and $\lambda(i)=\mu(i)=i^\gamma(i\geq 1)$ for some $\gamma>0$ and $\mu(0)=0$ (then we have equilibrium rate $r(i)=1$). Denote its infinitesimal generator by Ω. Stationary distribution is $\pi(n)=C\cdot\frac{1}{n^\gamma}$, where c is a normalisation constant. This distribution is heavy-tailed (for any $\gamma>0$), but $Gap(\Omega)>0$ if and only if $\gamma\geq 2$.

6.5 Queueing networks

6.5.1 Spectral gap for Jackson network

Consider a **Jackson network** which consists of m numbered servers, denoted by $\bar{M} := \{1, \dots, m\}$. Station $j \in \bar{M}$, is a single server queue with infinite waiting room under FCFS (First Come First Served) regime. All the customers in the network are indistinguishable. There is an external Poisson arrival stream with intensity $\bar{\lambda}$ and arriving customers are sent to node j with probability r_{0j}, $\sum_{j=1}^{m} r_{0j} = r \le 1$. The quantity $r_{00} := 1 - r$ is then the rejection probability with that customers immediately leave the network. Customers arriving at node j from the outside or from other nodes request a service which is exponentially distributed with mean 1. Service at node j is provided with intensity $\mu_j(x_j)$ ($\mu_j(0) := 0$), where x_j is the number of customers at node j including the one being served. All the service times and arrival processes are assumed to be independent.
A customer departing from node i immediately proceeds to node j with probability $r_{ij} \ge 0$ or departs from the network with probability r_{i0}. The routing is independent of the past system given the momentary node where the customer is. Let $\bar{M}_0 := \{0, 1, \dots, m\}$. We assume that the matrix $R := (r_{ij},\ i, j \in \bar{M}_0)$ is irreducible.
Let $X'_j(t)$ be the number of customers present at node j at time $t \ge 0$. Then $X'(t) = (X'_1(t), \dots, X'_m(t))$ is the joint queue length vector at time instant $t \ge 0$ and $\mathbf{X} := (X'(t), t \ge 0)$ is the joint queue length process with state space $E = \mathbb{Z}_+^m$.
Denote possible transformations from one state to another:

$$\begin{aligned} T_{ij}x &:= (x_1, \dots, x_i - 1, \dots, x_j + 1, \dots, x_m), \\ T_{\cdot j}x &:= (x_1, \dots, x_j + 1, \dots, x_m), \\ T_{i\cdot}x &:= (x_1, \dots, x_i - 1, \dots, x_m). \end{aligned} \tag{6.143}$$

The following theorem is classical.

Theorem 6.5.1 (Jackson [26]). *Under the above assumptions the queueing process* $\mathbf{X}$ *is a continuous time Markov process with transition matrix* $\mathbf{Q} = (q(x,y),\ x, y \in E)$ *given by*

$$q(x,y) = \begin{cases} \mu_i(x_i) r_{ij} & \text{if } y = T_{ij}x, \\ \bar{\lambda} r_{0j} & \text{if } y = T_{\cdot j}x, \\ \mu_i(x_i) r_{i0} & \text{if } y = T_{i\cdot}x, \\ -\sum_{x \ne y} q(x,y) & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases} \tag{6.144}$$

The corresponding infinitesimal generator is given by

$$Lf(x) = \sum_{j=1}^{m} [f(x) - f(T_{\cdot j}x)] \bar{\lambda} r_{0j} + \sum_{i=1}^{m} \sum_{j=1}^{m} [f(x) - f(T_{ij}x)] \mu_i(x_i) r_{ij} + \sum_{j=1}^{m} [f(x) - f(T_{j\cdot}x)] \mu_j(x_j) r_j. \tag{6.145}$$

The unique stationary distribution exists if and only if the unique solution of the **traffic equation**

$$\lambda_i = \bar{\lambda} r_{0j} + \sum_{j=1}^{m} \lambda_j r_{ji}, \quad i = 1, \dots, m \tag{6.146}$$

satisfies

$$b_i := \sum_{n=0}^{\infty} \frac{\lambda_i^n}{\prod_{y=1}^{n} \mu_i(y)} < \infty, \quad 1 \le i \le m.$$

what we henceforth assume. Then, the stationary distribution $\pi(x), x = (x_1, \dots, x_m)$ *is given by product:*

$$\pi(x) = \prod_{i=1}^{m} \pi_i(x_i), \text{ where } \pi_i(x_i) := \frac{1}{b_i} \frac{\lambda_i^{x_i}}{\prod_{y=1}^{x_i} \mu_i(y)}. \tag{6.147}$$

□

The parameters of a Jackson network are: the arrival intensity $\bar{\lambda}$, the routing matrix R (with its traffic vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$), the vector of service rates $\boldsymbol{\mu} = (\mu_1(\cdot), \ldots, \mu_m(\cdot))$ and the numbers of servers m.

Remark: We will shortly write $\bar{\lambda}_j$ for $\bar{\lambda} r_{0j}$.

For some monotonicities in Jackson networks see Szekli [45], Daduna and Szekli [11] or Lindvall [35].

Let

$$h_i(k) = \frac{\pi_i(k)}{\sum_{j=k}^{\infty} \pi_i(j)}$$

be the hazard rate of marginal distribution π_i.
We make the following assumptions:

$$A1 \qquad \text{There exist unique (up to a multiplying constant) solution to the traffic equation (6.146) i.e. } \lambda_1, \ldots, \lambda_m, \tag{6.148}$$

$$A2 \qquad \forall i: \quad \mu^i = \sup_y \mu_i(y) < \infty, \quad \mu_i = \inf_y \mu_i(y) > 0, \tag{6.149}$$

$$A3 \qquad \forall i: \quad \liminf_{k\to\infty} h_i(k) = \limsup_{k\to\infty} h_i(k). \tag{6.150}$$

Let $R > 0$ denotes the fact that all elements of R are positive numbers .
Recall that the spectral gap for this process (with generator given in 6.145) is given by

$$Gap(L) := \inf\{-(f, Lf)_\pi : ||f||_\pi = 1, (f, \mathbf{1})_\pi = 0\}.$$

We say that spectral gap exists if $Gap(L) > 0$.

Next theorem states that spectral gap exists for Jackson network if and only if each of marginal distributions π_i, $i = 1, \ldots, m$ is light-tailed.
Remark Assumption $A3$ in next theorem can be relaxed allowing $\liminf_{k\to\infty} h_i(k) \neq \limsup_{k\to\infty} h_i(k)$, the proof would be then similar, just some more cases to consider would be needed.

Theorem 6.5.2. *Consider Jackson network with routing such that $R^k > 0$ for some $k > 0$ and let $-L$ be the infinitesimal generator of this network. Assume in addition $A1$, $A2$ and $A3$. Then*

$$Gap(L) > 0 \textit{ if and only if each of marginal distributions } \pi_i, i = 1, \ldots, m \textit{ is light-tailed.}$$

Proof. Let J be the kernel corresponding to $-L$ i.e. such that $Lf(\mathbf{e}) = \sum_{\mathbf{e}' \in E} (f(\mathbf{e}) - f(\mathbf{e}'))J(\mathbf{e}, \mathbf{e}')$.
We start with defining m birth and death processes. Define infinitesimal generator $-L_i$ on $L^2(\mathbb{N}, \pi)$ by

$$L_i f(y) := [f(y) - f(y-1)]\lambda_i + [f(y) - f(y-1)]\mu_i(y) \tag{6.151}$$

Let $J^{(i)} = [J^{(i)}_{x,y}]_{x,y\in\mathbb{N}}$ be:

$$J^{(i)}_{x,y} = \begin{cases} \lambda_i & \text{if } y = x+1, \\ \mu_i(y) & \text{if } y = x-1, \\ -\mu_i(y) - \lambda_i & \text{if } y = x, \\ 0 & \text{otherwise.} \end{cases} \tag{6.152}$$

On $E = Z^m$ consider the Markov chain $\widetilde{\mathbf{X}} = (\widetilde{X}_t, t \geq 0)$, whose components are m independent, stationary birth and death processes, each having infinitesimal generator $-L_i$. Denote the infinitesimal generator of $\widetilde{\mathbf{X}}$ by $-\widetilde{L}$. Let $\widetilde{J}$ be the corresponding kernel such that $\widetilde{L}f(\mathbf{e}) = \sum_{\mathbf{e}' \in E} (f(\mathbf{e}) - f(\mathbf{e}'))\widetilde{J}(\mathbf{e}, \mathbf{e}')$.

The stationary distribution of process with generator $-\widetilde{L}$ is the product of stationary distributions of processes with generators $-L_i$, $i = 1, 2, \ldots, m$. Thus, by Theorem 6.2.7, we have $Gap(\widetilde{L}) = \inf_i Gap(L_i)$.
Note that both, J and $\widetilde{J}$, have the same stationary distribution being a product of stationary distributions of $J^{(i)}$.

Define $\widetilde{k}$ to be the Cheeger's constant for $\widetilde{\mathbf{X}}$ process and k to be Cheeger's constant for $\mathbf{X}$, i.e.

$$\widetilde{k} := \inf_{A\subset E, \pi(A)\in(0,1)} \widetilde{k}(A), \qquad \widetilde{k}(A) := \frac{\int_{x\in A}\pi(dx)\widetilde{J}(x,A^c)}{\pi(A)\pi(A^c)},$$

$$k := \inf_{A\subset E, \pi(A)\in(0,1)} k(A), \qquad k(A) := \frac{\int_{x\in A}\pi(dx)J(x,A^c)}{\pi(A)\pi(A^c)}.$$

To establish the theorem it is enough to have the following

$$\exists(w_1>0, w_2>0) \quad \forall(A\subset E) \qquad w_2\sum_{x\in A}\pi(x)\widetilde{J}(x,A^c) \geq \sum_{x\in A}\pi(x)J(x,A^c) \geq w_1\sum_{x\in A}\pi(x)\widetilde{J}(x,A^c). \tag{6.153}$$

Having above is enough because:

1. If $\exists i : \pi_i$ is heavy-tailed.
   Then according to Theorem 6.4.4 we have that $Gap(L_i) = 0$ and thus (Lemma 6.2.7) $Gap(\widetilde{L}) = 0$. Next, from Theorem 6.2.6 for $\widetilde{L}$ we conclude that $\widetilde{k} = 0$ ($\widetilde{k}^2 \leq 8M \cdot Gap(\widetilde{L}) = 0$). Now using (6.153) we have that
   $$k(A) := \frac{\int_{x\in A}\pi(dx)J(x,A^c)}{\pi(A)\pi(A^c)} \leq w_2\frac{\int_{x\in A}\pi(dx)\widetilde{J}(x,A^c)}{\pi(A)\pi(A^c)} = w_2\widetilde{k}(A)$$
   i.e.
   $$k \leq w_2\widetilde{k} = 0$$
   thus $k = 0$ and from Lemma 6.2.5 we have that $Gap(L) = 0$.
2. If $\forall i : \pi_i$ is light-tailed.

   Then according to Theorem 6.4.4 we have that $Gap(L_i) > 0$ for all i and thus (Lemma 6.2.7) $Gap(\widetilde{L}) > 0$. Applying Lemma 6.2.5 to $\widetilde{L}$ we conclude that $\widetilde{k} > 0$. Now using the existence of w_1 in (6.153) we have
   $$k(A) = \frac{\int_{x\in A}\pi(dx)J(x,A^c)}{\pi(A)\pi(A^c)} \geq w_1\frac{\int_{x\in A}\pi(dx)\widetilde{J}(x,A^c)}{\pi(A)\pi(A^c)} = w_1\widetilde{k}(A)$$
   i.e.
   $$k \geq w_1\widetilde{k} > 0$$
   thus (using Lemma 6.2.6) we get
   $$Gap(L) > 0.$$

Note, that (6.153) is equivalent to two conditions:

$$\inf_{\substack{A\subset E\\ \pi(A)\in(0,1)}}\left\{\frac{\sum_{x\in A}\pi(x)J(x,A^c)}{\sum_{x\in A}\pi(x)\widetilde{J}(x,A^c)}\right\} > 0 \quad \text{and} \quad \sup_{\substack{A\subset E\\ \pi(A)\in(0,1)}}\left\{\frac{\sum_{x\in A}\pi(x)J(x,A^c)}{\sum_{x\in A}\pi(x)\widetilde{J}(x,A^c)}\right\} < \infty. \tag{6.154}$$

Denote

$$W(A) = \frac{\sum_{x\in A}\pi(x)J(x,A^c)}{\sum_{x\in A}\pi(x)\widetilde{J}(x,A^c)} = \frac{\sum_{x\in A}\pi(x)\left(\sum_{T_{ij}x\in A^c}\mu_i(x_i)r_{ij} + \sum_{T_{\cdot j}x\in A^c}\bar{\lambda}_j + \sum_{T_{i\cdot}x\in A^c}\mu_i(x_i)r_{i0}\right)}{\sum_{x\in A}\pi(x)\left(\sum_{T_{\cdot j}x\in A^c}\lambda_j + \sum_{T_{i\cdot}x\in A^c}\mu_i(x_i)\right)}.$$

Recall that $\mu^i = \sup_i \mu_i(x_i) < \infty$ and $\mu_i = \inf_y \mu_i(y) > 0$.
Let $\mu^{max} = \max_{0\leq i\leq m}\mu^i < \infty$ and $\mu_{min} = \min_{0\leq i\leq m}\mu_i > 0$.
Denote

$$M_J(A,x) = \sum_{T_{ij}x\in A^c}\mu_i(x_i)r_{ij} + \sum_{T_{\cdot j}x\in A^c}\bar{\lambda}_j + \sum_{T_{i\cdot}x\in A^c}\mu_i(x_i)r_{i0}$$

and

$$M_J^{min} = \inf_{A,x:M_J(A,x)>0}\{M_J(A,x)\}, \quad M_J^{max} = \sup_{A,x:M_J(A,x)>0}\{M_J(A,x)\}.$$

We have

$$\begin{aligned} M_J^{min} &\geq \min\{\mu_{min} \cdot \min_{i,j: r_{ij}>0} r_{ij}, \min_{j:\bar{\lambda}_j>0} \bar{\lambda}_j, \mu_{min} \cdot \min_{i:r_{i0}>0} r_{i0}\} > 0, \\ M_J^{max} &\leq m(m-1)\mu^{max} \cdot 1 + \sum_{j=1}^m \bar{\lambda}_j + m\mu^{max} \cdot 1 = m^2 \mu^{max} + \sum_{j=1}^m \bar{\lambda}_j < \infty. \end{aligned}$$

Similarly

$$M_{\widetilde{J}}(A,x) = \sum_{T_{\cdot j} x \in A^c} \lambda_j + \sum_{T_{i\cdot} x \in A^c} \mu_i(x_i)$$

and

$$M_{\widetilde{J}}^{min} = \inf_{A,x: M_{\widetilde{J}}(A,x)>0} \left\{ M_{\widetilde{J}}(A,x) \right\}, \quad M_{\widetilde{J}}^{max} = \sup_{A,x: M_{\widetilde{J}}(A,x)>0} \left\{ M_{\widetilde{J}}(A,x) \right\}.$$

We have

$$\begin{aligned} M_{\widetilde{J}}^{min} &\geq \min\{\min_j \lambda_j, \mu_{min}\} > 0, \\ M_{\widetilde{J}}^{max} &\leq \sum_{j=1}^m \lambda_j + m \cdot \mu^{max} < \infty. \end{aligned}$$

Note that both M_J and $M_{\widetilde{J}}$ do not depend on A. Define:

$$\partial \widetilde{A} = \{x : M_{\widetilde{J}}(A,x) > 0\}, \qquad \partial A = \{x : M_J(A,x) > 0\}.$$

$\partial \widetilde{A}$ is a set of all states from which process with kernel $\widetilde{J}$ can get in one step to A^c, whereas ∂A is a set of states from which original process with kernel J can get in one step to A^c. We have

$$W(A) = \frac{\sum_{x \in \partial A} \pi(x) M_J(A,x)}{\sum_{x \in \partial \widetilde{A}} \pi(x) M_{\widetilde{J}}(A,x)}$$

and

$$\frac{M_J^{max}}{M_{\widetilde{J}}^{min}} \cdot \frac{\sum_{x \in \partial A} \pi(x)}{\sum_{x \in \partial \widetilde{A}} \pi(x)} \geq \frac{\sum_{x \in \partial A} \pi(x) M_J(A,x)}{\sum_{x \in \partial \widetilde{A}} \pi(x) M_{\widetilde{J}}(A,x)} \geq \frac{M_J^{min}}{M_{\widetilde{J}}^{max}} \cdot \frac{\sum_{x \in \partial A} \pi(x)}{\sum_{x \in \partial \widetilde{A}} \pi(x)}$$

To show (6.154) we have to show that

$$\infty > \frac{\sum_{x \in \partial A} \pi(x)}{\sum_{x \in \partial \widetilde{A}} \pi(x)} > 0$$

or equivalently

$$\infty > \frac{\sum_{x \in \partial \widetilde{A}} \pi(x)}{\sum_{x \in \partial A} \pi(x)} > 0. \tag{6.155}$$

Before proceeding let us examine how much can differ $\pi(x)$ and $\pi(x')$, where x' is a state to which one can get from x in at most $k \leq m$ steps.
Recall (6.147)

$$\pi(x) = \prod_{i=1}^m \pi_i(x_i), \text{ where } \pi_i(x_i) := \frac{1}{b_i} \frac{\lambda_i^{x_i}}{\prod_{y=1}^{x_i} \mu_i(y)}.$$

Note that x and x' on position i can only differ by at most k. We have (assume $x_i \geq k$)

$$\pi_i(x_i + k) = \frac{1}{b_i} \frac{\lambda_i^{x_i+k}}{\prod_{y=1}^{x_i+k} \mu_i(y)} = \pi_i(x_i) \frac{\lambda_i^k}{\mu_i(x_i+1) \cdots \mu_i(x_i+k)}$$

and

$$\pi_i(x_i - k) = \frac{1}{b_i} \frac{\lambda_i^{x_i-k}}{\prod_{y=1}^{x_i-k} \mu_i(y)} = \pi_i(x_i) \frac{\lambda_i^{-k}}{\frac{1}{\mu_i(x_i-m+1)\cdots\mu_i(x_i)}}$$

and thus we have bounds:

$$\left(\frac{\lambda_i}{\mu^i}\right)^k \pi_i(x_i) \quad \le \quad \pi_i(x_i+m) \quad \le \pi_i(x_i)\left(\frac{\lambda_i}{\mu_i}\right)^k,$$

$$\left(\frac{\lambda_i}{\mu_i}\right)^{-k} \pi_i(x_i) \quad \le \quad \pi_i(x_i-k) \quad \le \pi_i(x_i)\left(\frac{\lambda_i}{\mu^i}\right)^{-k}.$$

Let $c_i = \max\left(\left(\frac{\lambda_i}{\mu^i}\right), \left(\frac{\lambda_i}{\mu^i}\right)^{-1}\left(\frac{\lambda_i}{\mu_i}\right), \left(\frac{\lambda_i}{\mu_i}\right)\right)$. We have

$$\left(\frac{1}{c_i}\right)^k \pi_i(x_i) \le \pi_i(x_i \pm k) \le c_i^k \pi_i(x_i)$$

Denote yet $c = \prod_{i=1}^k c_i$. So if we take x and x' which can differ from x' by $\pm k$ at each position, we have

$$\frac{1}{c}\pi(x) \le \pi(x') \le c\pi(x),$$

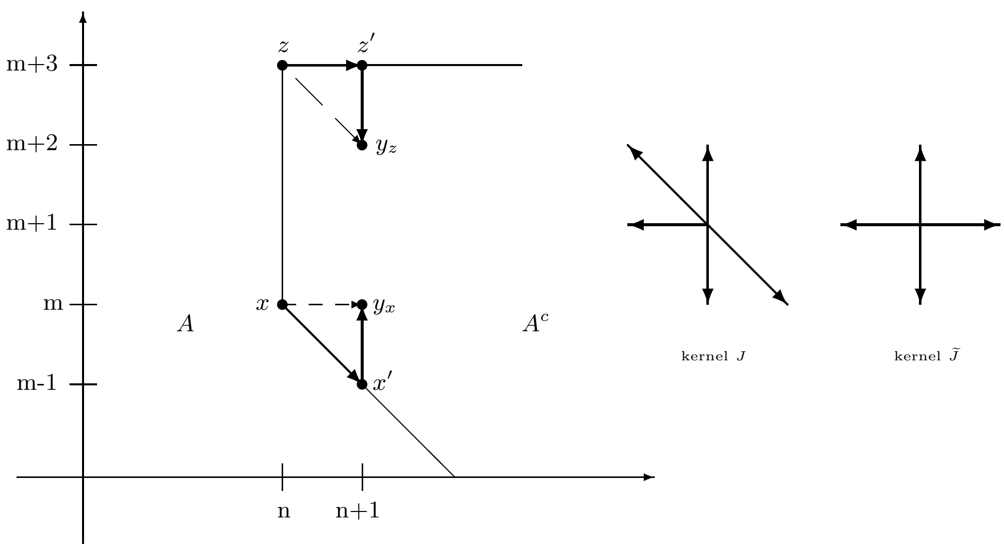
i.e.

$$\frac{1}{c}\pi(x') \le \pi(x) \le c\pi(x') \tag{6.156}$$

Of course we can have states which are in ∂A and not in $\partial\widetilde{A}$ and also such that are in $\partial\widetilde{A}$ but not in ∂A.
If we take $z \in \partial A \setminus \partial\widetilde{A}$, i.e. there exists some $y_z \in A^c$ such that process with kernel $\widetilde{J}$ cannot move from z to y_z in one step, but process with kernel J can. State y_z must be of form $y_z = T_{ij}z$. But note that there always exists path $T_{i\cdot}, T_{\cdot j}$ such that $z' = T_{i\cdot}z \in A$ and $y_z = T_{\cdot j}z' \in A^c$. See the following

Example. Consider Jackson network with two stations and no transition $T_{\cdot 1}$, i.e. with $\bar{\lambda}_1 = 0$, assume all other transitions are possible.
Take A as on the picture (poly-line going through x', x, z, z' belongs to A) and $x = (n, m), y_x = (n+1, m), x' = (n+1, m-1), z = (n, m+3), y_z = (n+1, m+2), z' = (n+1, m+3)$.

6.5.1



In the above example we have $z \in \partial A \setminus \partial\widetilde{A}$ and $x \in \partial\widetilde{A} \setminus \partial A$. In the example we have that $z' = T_{\cdot 1}z \in A$ and $y_z = T_{2\cdot}z' \in A^c$ (there is also, not drawn on the picture, another such path: $z'' = T_{2\cdot} \in A$ and then $y_z = T_{\cdot 1}z''$). Of course either $z' \in \partial\widetilde{A} \setminus \partial A$ or $z' \in \partial\widetilde{A} \cap \partial A$. For any such z' there are no more than $2m$ corresponding z, thus using (6.156) (where $c \ge 1$) we have

$$\frac{\sum_{x\in\partial\widetilde{A}} \pi(x)}{\sum_{x\in\partial A} \pi(x)} = \frac{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial\widetilde{A}\setminus\partial A} \pi(x)}{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial A\setminus\partial\widetilde{A}} \pi(x)} \ge \frac{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial\widetilde{A}\setminus\partial A} \pi(x)}{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + 2mc\left(\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial\widetilde{A}\setminus\partial A} \pi(x)\right)}$$

$$\geq \frac{\sum_{x\in\partial\widetilde{A}\cap\partial A}\pi(x)+\sum_{x\in\partial\widetilde{A}\setminus\partial A}\pi(x)}{(2mc+1)\left(\sum_{x\in\partial\widetilde{A}\cap\partial A}\pi(x)+\sum_{x\in\partial\widetilde{A}\setminus\partial A}\pi(x)\right)}=\frac{1}{2mc+1}$$

Similarly fix $x\in\partial\widetilde{A}\setminus\partial A$, i.e. there exists some $y_x\in A^c$ such that original process with kernel J cannot move there in one step, but process with kernel $\widetilde{J}$ can. State y_x must be of form $y_x=T_{\cdot i}x$ or $y_x=T_{j\cdot}x$. Then there exists path of length at most k such that process with kernel J can, moving along the path, get to y_x. There must exists state x' on this path such that $x'\in A$ and next state on the path is in A^c. Of course either $x'\in\partial\widetilde{A}\cap\partial A$ or $x'\in\partial A\setminus\partial\widetilde{A}$. For any such x' there can be surely no more than $(2k)^m$ states corresponding to it (because there is at most so many any other states: they can differ by $\pm k$ at each of m positions). Thus, using (6.156) (where $c\geq 1$) we have

$$\frac{\sum_{x\in\partial\widetilde{A}}\pi(x)}{\sum_{x\in\partial A}\pi(x)}=\frac{\sum_{x\in\partial\widetilde{A}\cap\partial A}\pi(x)+\sum_{x\in\partial\widetilde{A}\setminus\partial A}\pi(x)}{\sum_{x\in\partial\widetilde{A}\cap\partial A}\pi(x)+\sum_{x\in\partial A\setminus\partial\widetilde{A}}\pi(x)}\leq\frac{\sum_{x\in\partial\widetilde{A}\cap\partial A}\pi(x)+c(2k)^m\left(\sum_{x\in\partial\widetilde{A}\cap\partial A}\pi(x)+\sum_{x\in\partial A\setminus\partial\widetilde{A}}\pi(x)\right)}{\sum_{x\in\partial\widetilde{A}\cap\partial A}\pi(x)+\sum_{x\in\partial A\setminus\partial\widetilde{A}}\pi(x)}$$

$$\leq\frac{(1+c(2k)^m)\left(\sum_{x\in\partial\widetilde{A}\cap\partial A}\pi(x)+\sum_{x\in\partial A\setminus\partial\widetilde{A}}\pi(x)\right)}{\sum_{x\in\partial\widetilde{A}\cap\partial A}\pi(x)+\sum_{x\in\partial A\setminus\partial\widetilde{A}}\pi(x)}=(1+c(2k)^m)<\infty.$$

This way we have shown that $W(A)$ is bounded uniformly (for any $A:\pi(A)\in(0,1)$) from below and above:

$$\infty>(1+c(2k)^m)\frac{M_J^{max}}{M_{\widetilde{J}}^{min}}\geq\frac{\sum_{x\in\partial\widetilde{A}}\pi(x)M_J(A,x)}{\sum_{x\in\partial A}\pi(x)M_{\widetilde{J}}(A,x)}\geq\frac{M_J^{min}}{M_{\widetilde{J}}^{max}}\cdot\frac{1}{2mc+1}>0$$

i.e. (6.154).

For comparison we will recall theorem with proof by McDonald, Iscoe [36] in our settings. The theorem states that if each π_i, $i=1,\ldots,m$ is light-tailed, then there exists spectral gap for Jackson network. The idea of proof is different from the one of Theorem 6.5.2, authors construct some birth and death process and then compare it to the original Jackson network.
Remark As in case of standard Jackson network, assumption $A3$ in next theorem can be relaxed allowing $\liminf_{k\to\infty}h_i(k)\neq\limsup_{k\to\infty}h_i(k)$, the proof would be then similar, just some more cases to consider would be needed.

Theorem 6.5.3 (McDonald & Iscoe [36])**.** *Consider Jackson network with routing such that $R^k>0$ for some $k>0$ and let $-L$ be the infinitesimal generator of this network. Assume in addition A1, A2 and A3. Then*

if each of marginal distributions $\pi_i, i=1,\ldots,m$ is light-tailed, then $Gap(L)>0$.

Proof. Let $-L, J, -\widetilde{L}, -\widetilde{J}$ be defined as in Theorem 6.5.2.
It is enough to find constants $\nu>0$ such that

$$\forall(A\subset E,\pi(A)>0)\qquad\sum_{x\in A}\pi(x)J(x,A^c)\geq\nu\sum_{x\in A}\pi(x)\widetilde{J}(x,A^c),\tag{6.157}$$

because then according to Theorem 6.4.4 we have that $Gap(L_i)>0$ for all i and thus (Lemma 6.2.7) $Gap(\widetilde{L})>0$. Now applying Lemma 6.2.5 to $\widetilde{L}$ we conclude that $\widetilde{k}>0$. Now using the existence of $\nu>0$ in (6.157) we have

$$k(A):=\frac{\int_{x\in A}\pi(dx)J(x,A^c)}{\pi(A)\pi(A^c)}\geq\nu_2\frac{\int_{x\in A}\pi(dx)\widetilde{J}(x,A^c)}{\pi(A)\pi(A^c)}=\nu\widetilde{k}(A)$$

i.e.

$$k\geq\nu\widetilde{k}>0$$

thus (using Lemma 6.2.6) we get

$$Gap(L) > 0.$$

To prove existence of μ consider two cases:

a) $\lambda := \min(\frac{\bar{\lambda}_i}{\lambda_i} : 1 \le i \le m) > 0$ and $\mu := \min(r_{i0} : 1 \le i \le m) > 0$.

Take $\nu = \min(\lambda, \mu)$, for then we have $\bar{\lambda}_j \ge \nu \lambda_j$ and $r_{i0} \ge \nu$ and we have for $x \in A$

$$J(x, A^c) = \sum_{y \in A^c} J(x, y) = \sum_{y = T_{ij} x \in A^c} J(x, y) + \sum_{y = T_{\cdot j} x \in A^c} J(x, y) + \sum_{y = T_{i \cdot} x \in A^c} J(x, y)$$

$$= \sum_{T_{ij} x \in A^c} \mu_i(x_i) r_{ij} + \sum_{T_{\cdot j} x \in A^c} \bar{\lambda}_j + \sum_{T_{i \cdot} x \in A^c} \mu_i(x_i) r_{i0} \ge$$

$$\nu \sum_{T_{\cdot j} x \in A^c} \lambda_j + \nu \sum_{T_{i \cdot} x \in A^c} \mu_i(x_i) = \nu \widetilde{J}(x, A^c).$$

(Terms at T_{ij} were simply dropped).

b) $\min(\frac{\bar{\lambda}_i}{\lambda_i} : 1 \le i \le m) = 0$ or $\min(r_{i0} : 1 \le i \le m) = 0$.

This time it is possible that $\bar{\lambda}_i = 0$ or some $r_{i0} = 0$. We will describe ν by constructed birth and death process with matrix $Q = Q^+ + Q^-$ and such that

$$\sum_{x \in A} \pi(x) J(x, A^c) \ge \sum_{x \in A} \pi(x) Q(x, A^c) \ge \nu \sum_{x \in A} \pi(x) \widetilde{J}(x, A^c).$$

We start with Q^+.

Define polytope $P_+(x) = \{x\} \cup \{T_{\cdot k} x : 1 \le k \le m\}$ and define $\widetilde{C}_{\cdot i}(x)$ to be the set of all nonself-intersecting *probable* paths on $P_+(x)$ ("probable" means that transition from any point on the path to next one can occur). A typical path consists of x followed by an exogenous arrival at some node, say a_1, followed by departure from a_1 into some a_2 etc and finally a transition from let say a_s to i. Because the Jackson network is exogenously supplied, there must exist such path. (of course $t = (x, T_{\cdot i} x) \in \widetilde{C}_{\cdot i}(x)$ if $\bar{\lambda}_i > 0$). The transitions corresponding to the path described above are

$$T_{\cdot a_1}, T_{a_1 a_2}, \ldots, T_{a_s i}$$

For such a path t, define

$$\lambda(t) := \min(\bar{\lambda}_{a_1}, \lambda_{a_1} r_{a_1 a_2}, \ldots, \lambda_{a_s} r_{a_s i}).$$

Of course $\lambda(t)$ does not depend on x. We reduce $\widetilde{C}_{\cdot i}(x)$ to

$$C_{\cdot i}(x) = \{t \in \widetilde{C}_{\cdot i}(x) : \text{no two paths have any transition in common}\} = \bigcup_{k=1}^{n} t_k^{(i)}.$$

Note that $|C_{\cdot i}(x)| = n \le m$ (because if $t_k^{(i)}$ starts with $T_{\cdot a}$, then any other cannot start with the same a).

Set the transition (birth) rate of process Q^+ at i:

$$Q^+(i) = Q^+(x, T_{\cdot i} x)) := \frac{1}{2m} \sum_{t \in C_{\cdot i}(x)} \lambda(t) = \frac{1}{2m} \sum_{k=1}^{n} \lambda(t_k^{(i)}).$$

(Factor $\frac{1}{2}$ will become clear at the and of proof).

Fix $x \in A$ such that $J(x, A^c) > 0$, then we have $Q^+(x, T_{\cdot i} x) > 0$ for all i such that $T_{\cdot i} x \in A^c$. If we take any path $t \in C_{\cdot i}(x)$ we have $x \in A$ and $T_{\cdot i} x \in A^c$, so at least one transition in t, say T_{jk} crosses from A to A^c (i.e. $T_{\cdot j} x \in A, T_{\cdot k} x \in A^c$) and is in the sum $\sum_{y \in A} \pi(y) J(y, A^c)$, namely in term $\pi(y) \mu_j(y_j) r_{jk}$, where $y = T_{\cdot j} x$.

Reversibility implies:

$$\bar{\lambda}_j = \lambda_j r_{j0}, \text{ and } \lambda_i r_{ij} = \lambda_j r_{ji}, \tag{6.158}$$

Using this we have:

$$\pi(x) \bar{\lambda}_j = \pi_j(x_j) \bar{\lambda}_j = \pi_j(x_j) \lambda_j r_{j0},$$

i.e.

$$\pi_j(x_j) \lambda_j r_{j0} = \pi_j(x_j + 1) \mu_j(x_j + 1) r_{j0}$$

and

$$\pi_j(x_j)\lambda_j = \pi_j(x_j+1)\mu_j(x_j+1).$$

Thus

$$\pi(y)\mu_j(y_j)r_{jk} = \pi(T_{\cdot j}x)\mu_j(x_j+1)r_{jk} = \pi_j(x_j+1)\mu_j(x_j+1)r_{jk} = \pi(x)\lambda_j r_{jk} \geq \pi(x)\lambda(t).$$

But $y = T_{\cdot j}x$ can only be used (at most) once for each i (because no two paths in $C_{\cdot i}(x)$ have any transitions in common), i.e. at most m times in total. This is also true with initial transition $x \to T_{\cdot a}x$ for which contribution to previous sum is

$$\pi(x)\bar{\lambda}_a \geq \pi(x)\lambda(t).$$

Thus

$$\sum_{x\in A}\pi(x)Q^+(x,A^c) = \sum_{x\in A}\pi(x)\sum_{T_{\cdot i}x\in A^c}Q^+(x,T_{\cdot i}x) = \frac{1}{2m}\sum_{x\in A}\sum_{T_{\cdot i}x\in A^c}\sum_{t\in C_{\cdot i}(x)}\lambda^{(2)}(t)\pi(x) \leq \frac{1}{2m}\sum_{y\in A}m\pi(y)J(y,A^c)$$

and thus

$$\frac{1}{2}\sum_{x\in A}\pi(x)J(x,A^c) \geq \sum_{x\in A}\pi(x)Q^+(x,A^c). \tag{6.159}$$

In the similar way for x different from $\mathbf{0}$ vector we define $P_-^{(i)}(x) := \{x\} \cup \{T_{i\cdot}x\} \cup \{T_{ij}x : j \neq i\}$. Define $\widetilde{C}_{i\cdot}(x)$ to be the set of all probable non-self-intersecting paths on $P_-^{(i)}(x)$ having common initial point x and terminal point $T_{i\cdot}x$. A typical path t consists of transition from i to a_1, then from a_1 to a_2 and so on, till leaving the network from a_s. Clearly path $(x, T_{i\cdot}x) \in \widetilde{C}_{i\cdot}$ if $r_{i0} > 0$. The transitions corresponding to above path are:

$$T_{ia_1}, T_{a_1a_2}, \ldots, T_{a_s\cdot}$$

For such path t define

$$\mu(t) = \min\left(\mu_i r_{ia_1}\ ,\frac{\mu_i}{\lambda_i}\lambda_{a_1}r_{a_1a_2}\ ,\ldots,\ \frac{\mu_i}{\lambda_i}\lambda_{a_s}r_{a_s}\right) \quad \left(\text{recall} : \mu_i := \inf_y \mu_i(y)\ \right).$$

If $r_{i0} > 0$, then $\mu(t) = \mu_i r_{i0}$.
Of course $\mu(t)$ does not depend on x. Denote

$$C_{i\cdot}(x) = \{t \in \widetilde{C}_{i\cdot}(x) : \text{no two paths have any transition in common}\} = \bigcup_{k=1}^{n} t_k^{(i\cdot)}.$$

Of course $n \leq m$. Define

$$Q^-(i) = Q^-(x, T_{i\cdot}x) = \frac{1}{2m}\sum_{t\in C_{i\cdot}x}\mu(t) = \frac{1}{2m}\sum_{k=1}^{n}\mu(t_k^{(i\cdot)}).$$

Above is then independent of x and $Q^-(x,y) > 0 \iff y = T_{i\cdot}x$ for some i.
Fix $x \in A$ such that $Q^+(x,A^c) > 0$, then for each $i \in \{1,\ldots,m\}$ such that $T_{i\cdot}x \in A^c$ we have $Q^-(x, T_{i\cdot}x)$. For every $t \in C_{i\cdot}(x)$ we have $x \in A$ and $T_{i\cdot}x \in A^c$.
It follows that there is at least one transition in t, say T_{jk} crosses from A to A^c (i.e. $T_{ij}x \in A$ and $T_{ik}x \in A^c$) and contributes the term $\pi(y)\mu_j(y_j)r_{jk}$ with $y = T_{ij}x$ to the sum $\sum_{y\in A}\pi(y)J(y,A^c)$.
By reversibility (equation: (6.158)):

$$\pi(x)Q(x,T_{ij}x) = \pi(x)\mu_i(x_i)r_{ij} = \pi(T_{ij}x)Q(T_{ij}x,x) = \pi(T_{ij}x)\mu_j(x_j+1)r_{ji}$$

and using $\lambda_i r_{ij} = \lambda_j r_{ji}$ we have

$$\pi(y)\mu_j(y_j)r_{jk} = \pi(T_{ij}x)\mu_j(x_j+1)r_{ji}r_{jk} = \pi(x)\frac{\mu_i(x_i)}{\lambda_i}\lambda_j r_{jk} \geq \pi(x)\mu(t).$$

Further:

If we consider $T_{a_s\cdot}$ then it contributes the term $\pi(y)\mu_{a_s}(y_{a_s})r_{a_s0}$, where $y = T_{ia_s}x$ to the sum $\sum_{y\in A}\pi(y)J(y,A^c)$, then we also have

$$\pi(y)\mu_{a_s}(y_{a_s})r_{a_s} = \pi(T_{ia_s}x)\mu_{a_s}(x_{a_s}+1)r_{a_s} = \pi(x)\frac{\mu_i(x_i)}{\lambda_i}\lambda_{a_s}r_{a_s} \geq \pi(x)\mu(t).$$

If $r_{i0} > 0$, then $T_{i\cdot}$ is possible, then in the sum $\sum_{y\in A}\pi(y)J(y,A^c)$ there is a term $\pi(x)\mu_i(x_i)r_i \geq \pi(x)\mu(t)$. In $C_{i\cdot}(x)$ there are no two transitions in common, so above y is only being used at most once, and at most m times in total, thus

$$\sum_{x\in A}\pi(x)Q^-(x,A^c) = \frac{1}{2m}\sum_{x\in A}\sum_{i:T_{i\cdot}x\in A^c}\sum_{t\in C_{i\cdot}(x)}\mu(t)\pi(x) \leq \frac{1}{2m}\sum_{y\in A}m\pi(y)J(y,A^c),$$

which gives

$$\frac{1}{2}\sum_{x\in A}\pi(x)J(x,A^c) \geq \sum_{x\in A}\pi(x)Q^-(x,A^c). \tag{6.160}$$

Adding (6.159) and (6.160) gives

$$\sum_{x\in A}\pi(x)J(x,A^c) \geq \sum_{x\in A}\pi(x)Q(x,A^c), \quad Q := Q^+ + Q^- \tag{6.161}$$

Now Q is a transition kernel of some multidimensional birth and death process. It's stationary distribution can differ from π, but simply redefine $\lambda := \min(\frac{Q^+(i)}{\lambda_i}, 1 \leq i \leq m)$, $\mu := \min(\frac{Q^-(i)}{\mu^i}; 1 \leq i \leq m)$ and take $\nu := \min(\lambda,\mu) > 0$. Then we always have $Q^+(x,T_{\cdot j}x) = Q^+(j) \geq \nu\lambda_j$ and $Q^-(x,T_{i\cdot}x)Q^-(i) \geq \nu\mu_i(x_i)$, using these we have

$$\begin{aligned}\sum_{x\in A}\pi(x)Q(x,A^c) &= \sum_{\substack{x\in A\\ y=T_{\cdot j}x}}\pi(x)Q^+(x,T_{\cdot j}x) + \sum_{\substack{x\in A\\ y=T_{i\cdot}x}}\pi(x)Q^-(x,T_{i\cdot}x)\\ &\geq \nu\sum_{\substack{x\in A\\ y=T_{\cdot j}x}}\pi(x)\lambda_j + \nu\sum_{\substack{x\in A\\ y=T_{i\cdot}x}}\pi(x)\mu_i(x_i) = \nu\sum_{x\in A}\pi(x)\widetilde{J}(x,A^c).\end{aligned}$$

Using this and (6.161) we have

$$\sum_{x\in A}\pi(x)J(x,A^c) \geq \sum_{x\in A}\pi(x)Q(x,A^c) \geq \nu\sum_{x\in A}\pi(x)\widetilde{J}(x,A^c).$$

□

6.5.2 Unreliable Jackson network & spectral gap

In this subsection we investigate the existence of spectral gap in unreliable Jackson networks, i.e. a Jackson network in which servers may break down.

The breakdowns events are of rather general structure and may occur in different ways: they can break down as an isolated event or in groups, the same with repairs. It is not required that those servers which stopped together return to service at the same time.

Denote $\bar{M} := \{1,2,\dots,m\}$ and $\bar{M}_0 := \{0,1,2,\dots,m\}$.

Behaviour of breakdowns and repair:

- Let $\bar{D}\subset\bar{M}$ be the set of servers in down status and $\bar{I}\subset\bar{M}\setminus\bar{D}, \bar{I}\neq\emptyset$ be the subset of nodes in up status. Then the servers in $\bar{I}$ break down with intensity $\alpha^{\bar{D}}_{\bar{D}\cup\bar{I}}(x_i : i\in\bar{M})$, if there are x_i customers at server $i, i\in\bar{D}\cup\bar{I}$. Thus the breakdown of servers $\bar{I}$ depends on local loads at servers in $\bar{I}$ and of those which are already under repair $\bar{D}$.
- Let $\bar{D}\subset\bar{M}$ be the set of servers in down status and $\bar{H}\subset\bar{D}, \bar{H}\neq\emptyset$. The broken servers from $\bar{H}$ return from repair with intensity $\beta^{\bar{D}}_{\bar{D}\setminus\bar{H}}(x_i : i\in\bar{M})$.
- The routing is changed according to so-called Repetitive Service - Random Destination Blocking (RS-RD BLOCKING) rule: For $\bar{D}$ - set of servers under repair routing probabilities are restricted to nodes from $\bar{M}_0\setminus\bar{D}$ as follows:

$$r^{\bar{D}}_{ij} = \begin{cases} r_{ij}, & i,j\in\bar{M}_0\setminus\bar{D}, i\neq j\\ r_{ii}+\sum_{k\in\bar{D}}r_{ik}, & i\in\bar{M}_0\setminus\bar{D}, i=j\end{cases}$$

The external arrival rates are:

$$\begin{aligned}\bar{\lambda}^{\bar{D}}_j &= \bar{\lambda}r^{\bar{D}}_{0j} = \bar{\lambda}r_{0j} = \bar{\lambda}_j \text{ for nodes } j\in\bar{M}\setminus\bar{D}\\ \bar{\lambda}^{\bar{D}}_j &= \bar{\lambda}r^{\bar{D}}_{0j} = 0 \text{ for nodes } j\in\bar{D}, (j\neq 0).\end{aligned} \tag{6.162}$$

Denote $R^{\bar{D}} = (r^{\bar{D}}_{ij})_{i,j\in\bar{M}_0}$. Note that $R^{\emptyset} = R$.

Intensities $\alpha^{\bar{D}}_{\bar{D}\cup\bar{I}}(x_i : i \in \bar{M})$ and $\beta^{\bar{D}}_{\bar{D}\setminus\bar{H}}(x_i : i \in \bar{M})$ cannot be general.
Indices at αs and βs are to mean that subset of broken nodes changes from $\bar{D}$ to $\bar{D} \cup \bar{I}$ or $\bar{D} \setminus \bar{H}$ respectively.
Rules for general classes of suitable intensities were found by Sauer ([41], [42]), such that breakdown and repair events are controlled by a multidimensional birth & death process: if we represent for each $j \in \bar{M}$ its normal status as 0 and its repair status as 1, then the intensities of births and deaths are:

$$q(\vec{s}, \vec{s}') := \begin{cases} \alpha(\vec{s}, \vec{a}), & \text{if} \quad \vec{s}' = \vec{s} + \vec{a} \\ \beta(\vec{s}, \vec{a}), & \text{if} \quad \vec{s}' = \vec{s} - \vec{a} \end{cases} \qquad \text{for} \quad \vec{s}, \vec{s}', \vec{a} \in \{0,1\}^M.$$

Definition 6.5.4. *The intensities of breakdowns and repairs for $\emptyset \neq \bar{I} \subset \bar{D}$ and $\emptyset \neq \bar{H} \subset \bar{M} \setminus \bar{D}$ are*

$$\alpha^{\bar{D}}_{\bar{D}\cup\bar{I}}(x_i : i \in \bar{M}) := \frac{A(x_i : i \in \bar{D}\cup\bar{I})}{A(x_i : i \in \bar{D})}$$
and
$$\beta^{\bar{D}}_{\bar{D}\setminus\bar{H}}(x_i : i \in \bar{M}) := \frac{B(x_i : i \in \bar{D})}{B(x_i : i \in \bar{D}\setminus H)}$$

where A and B are any non-negative functions :

$$A, B : \bigcup_{\bar{I}\subset\bar{M}} \left(\{\bar{I}\} \times \mathbb{N}^{|\bar{I}|} \right) \to (0, \infty)$$

with $A(x_i : i \in \emptyset) = B(x_i : i \in \emptyset) = 1$ such that all feasible intensities $\alpha^{\bar{D}}_{\bar{D}\cup\bar{I}}$ and $\beta^{\bar{D}}_{\bar{D}\setminus\bar{H}}$ are finite and assuming $\frac{0}{0} := 0$.

In this subsection we will consider breakdowns and repairs which are state independent.

State independent breakdowns and repairs: Breakdown and repair intensities depend on the servers but are independent of the number of customers. Then the function A and B are of form:

$$A(x_i : i \in \bar{I}) = A(\bar{I}), \quad B(x_i, i :\in \bar{H}) = B(\bar{H}), \text{for all } \bar{I}, \bar{H} \subset \bar{M}$$

Thus, if $\bar{D} \subset \bar{M}$ is in down status then the intensities of breakdown of set $\bar{I} \neq \emptyset$ and repair of set $\bar{H} \neq \emptyset$ are:

$$\alpha^{\bar{D}}_{\bar{D}\cup\bar{I}}(x_i : i \in \bar{M}) = \frac{A(\bar{D}\cup\bar{I})}{A(\bar{D})}$$

$$\beta^{\bar{D}}_{\bar{D}\setminus\bar{H}}(x_i : i \in \bar{M}) = \frac{B(\bar{D})}{B(\bar{D}\setminus\bar{H})}$$

where A and B are any non-negative functions,

$$A, B : \mathcal{P}(\bar{M}) \to (0, \infty)$$

with $A(\emptyset) = B(\emptyset) = 1$ such that all feasible intensities $\alpha^{\bar{D}}_{\bar{D}\cup\bar{I}}$ and $\beta^{\bar{D}}_{\bar{D}\setminus\bar{H}}$ are finite.
The important characteristic is that here breakdown/repair process is a Markov process on its own state space $\mathcal{P}(\bar{M})$ of all subsets of $\bar{M}$.

Product form

In order to describe unreliable Jackson network we need to attach to the state space $\mathbb{Z}^m_+$ of the corresponding standard network process $\mathbf{X}$ an additional component which includes information of availability behaviour of the system described by a process $\mathbf{Y}$. We introduce states of the form:

$$(\bar{I}, x_1, x_2, \ldots, x_m) \in \mathcal{P}(\bar{M}) \times \mathbb{Z}^m_+ =: \widetilde{E}$$

The set $\bar{I}$ is the set of servers in down status. At node $i \in \bar{I}$ there are x_i customers waiting for server being repaired.

Denote intensity matrix of this process by $\mathcal{J}$. We can write down these intensities.
First extend the notion of possible transformation from one state to another. Let $x = (\bar{D}, x_1, \ldots, x_m) \in \mathcal{P}(\bar{M}) \times \mathbb{Z}^m_+$

$$\begin{aligned} T_{ij}x &:= (\bar{D}, x_1, \ldots, x_i - 1, \ldots, x_j + 1, \ldots, x_m), \\ T_{\cdot j}x &:= (\bar{D}, x_1, \ldots, x_j + 1, \ldots, x_m), \\ T_{i\cdot}x &:= (\bar{D}, x_1, \ldots, x_i - 1, \ldots, x_m), \\ T_{\cdot\bar{H}}x &:= (\bar{D} \setminus \bar{H}, x_1, \ldots, x_m), \\ T_{\bar{I}\cdot}x &:= (\bar{D} \cup \bar{I}, x_1, \ldots, x_m). \end{aligned} \tag{6.163}$$

For $x = (\emptyset, x_1 \ldots, x_m)$:

$$\mathcal{J}(x,y) = \begin{cases} \mu_i(x_i) r_{ij} & \text{if } y = T_{ij}x \\ \bar{\lambda}_j & \text{if } y = T_{\cdot j}x \\ \mu_i(x_i) r_{i0} & \text{if } y = T_{i\cdot}x \\ A(\bar{I}) & \text{if } y = T_{\bar{I}\cdot}x \\ -\sum_{x \neq y} \mathcal{J}(x,y) & \text{if } y = x \\ 0 & \text{otherwise} \end{cases} \tag{6.164}$$

And for general $x = (\bar{D}, x_1 \ldots, x_m)$:

$$\mathcal{J}(x,y) = \begin{cases} \mu_i(x_i) r_{ij}^{\bar{D}} & \text{if } y = T_{ij}x \\ \bar{\lambda}_j^{\bar{D}} & \text{if } y = T_{\cdot j}x \\ \mu_i(x_i) r_{i0}^{\bar{D}} & \text{if } y = T_{i\cdot}x \\ \frac{A(\bar{D}\cup\bar{I})}{A(\bar{D})} & \text{if } y = T_{\bar{I}\cdot}x \\ \frac{B(\bar{D})}{B(\bar{D}\setminus\bar{H})} & \text{if } y = T_{\cdot\bar{H}}x \\ -\sum_{x \neq y} \mathcal{J}(x,y) & \text{if } y = x \\ 0 & \text{otherwise} \end{cases} \tag{6.165}$$

Theorem 6.5.5 (Sauer & Daduna [42]). *Let $\widetilde{\mathbf{X}} = (\mathbf{Y}, \mathbf{X})$ be the process described above operating on the state space $\widetilde{E}$ with breakdown/repair intensities from Definition 6.5.4.*
Assume rule RS-RD-BLOCKING is used for changing routing matrix when there are some broken servers. In addition let us assume that original routing matrix R is reversible, i.e.:

$$\lambda_j r_{ji} = \lambda_i r_{ij}, \qquad i,j \in \bar{M}_0$$

(λ_j - solution to the traffic equation).
Then the stationary distribution of process $\widetilde{\mathbf{X}}$ is of product form:

For $x = (\bar{I}, x_1, \ldots, x_m) \in \mathcal{P}(\bar{M}) \times \mathbb{Z}_+^m$ we have:

$$\pi(x) = \pi(\bar{I}, x_1, \ldots, x_m) = \frac{1}{C} \frac{A(\bar{I})}{B(\bar{I})} \prod_{i=1}^{m} \pi_i(x_i) \tag{6.166}$$

(note that for $\bar{I} = \emptyset$ we have $\frac{A(\emptyset)}{B(\emptyset)} = \frac{1}{1} = 1$), where

$$\pi_i(x_i) = \frac{1}{b_i} \frac{\lambda^{x_i}}{\prod_{y=1}^{x_i} \mu_i(y)}, \qquad b_i = \sum_{n=0}^{\infty} \frac{\lambda_i^n}{\prod_{y=1}^{n} \mu_i(y)}$$

and C is a normalisation constant.
Constants $b_i, i = 1, \ldots, m$ are finite if and only if network is ergodic.

Next theorem is the extension of Theorem 6.5.2 for unreliable Jackson network. Sauer in [41] showed geometric rate of convergence in total variation distance for unreliable Jackson networks without rerouting (i.e. customers do traverse according to the routing matrix and are allowed to join the queue at broken server waiting till it is repaired, the stationary distribution in this case is not known) with constant service rates, i.e. $\mu_i(y) = \mu_i$ under some assumptions.

Theorem 6.5.6 (Spectral Gap for unreliable Jackson network)**.** *Let $\mathcal{L}$ be the infinitesimal generator of unreliable Jackson network process with routing matrix satisfying $R^k > 0$ for some $k \geq 1$. Denote its kernel by $\mathcal{J}$. Let intensities of breakdown and repairs be of form given in Definition 6.5.4 and use RS-RD BLOCKING as a rerouting rule. Assume in addition A1 and A2 (i.e. (6.149) and (6.149)). Then*

$$Gap(\mathcal{L}) > 0 \text{ if and only if each of marginal distributions } \pi_i, i = 1, \ldots, m \text{ is light-tailed.}$$

Proof. Define $\widetilde{\mathcal{J}}$ to be a Q matrix associated with $(m+1)$-dimensional vector $\widetilde{\mathbf{Z}}_t = (\widetilde{\mathbf{Y}}_t, \widetilde{\mathbf{X}}_t)$, where $\widetilde{\mathbf{X}}_t$ is a vector of m independent birth and death processes with Q matrices given in (6.152), i.e. ones with stationary distribution π_i and generators $-L_i$, and let $\widetilde{\mathbf{Y}}_t$ be the process on state space $\mathcal{P}(\bar{M})$ with infinitesimal generator denoted by $-L_{m+1}$ and stationary distribution:

$$\pi_{m+1}(\bar{I}) = \frac{1}{C'} \frac{A(\bar{I})}{B(\bar{I})}, \qquad C' := \left(\sum_{\bar{I} \subset \bar{M}} \frac{A(\bar{I})}{B(\bar{I})} \right)$$

Denote the infinitesimal generator of process $\widetilde{\mathcal{J}}$ by $-\widetilde{\mathcal{L}}$.
From Lemma 6.2.7 we have that $Gap(\widetilde{\mathcal{L}}) = \min_{1 \leq i \leq m+1} Gap(L_i)$. The state space $\mathcal{P}(\bar{M})$ is finite, thus $Gap(L_{m+1}) > 0$.

Define Cheeger's constants:

$$\kappa := \inf_{A, \pi(A) \in (0,1)} \kappa(A), \qquad \kappa(A) := \frac{\int_{x \in A} \pi(dx) \mathcal{J}(x, A^c)}{\pi(A)\pi(A^c)}.$$

$$\widetilde{\kappa} := \inf_{A, \pi(A) \in (0,1)} \widetilde{\kappa}(A), \qquad \widetilde{\kappa}(A) := \frac{\int_{x \in A} \pi(dx) \widetilde{\mathcal{J}}(x, A^c)}{\pi(A)\pi(A^c)}.$$

The proof is similar to case of standard Jackson network.
To establish the theorem it is enough that the following two conditions hold:

$$\exists (v_1 > 0, v_2 > 0) \quad \forall (A \subset E) \qquad v_2 \sum_{x \in A} \pi(x) \widetilde{\mathcal{J}}(x, A^c) \geq \sum_{x \in A} \pi(x) \mathcal{J}(x, A^c) \geq v_1 \sum_{x \in A} \pi(x) \widetilde{\mathcal{J}}(x, A^c). \tag{6.167}$$

Having this is enough, because:

1. If $\exists i : \pi_i$ is heavy-tailed.
   Then according to Theorem 6.4.4 we have that $Gap(L_i) = 0$ and thus (Lemma 6.2.7) $Gap(\widetilde{\mathcal{L}}) = 0$. Next, from Theorem 6.2.6 for $\widetilde{\mathcal{L}}$ we conclude that $\widetilde{\kappa} = 0$ ($\widetilde{\kappa}^2 \leq 8M \cdot Gap(\widetilde{\mathcal{L}}) = 0$). Now using (6.167) we have that

   $$\kappa(A) := \frac{\int_{x \in A} \pi(dx) \mathcal{J}(x, A^c)}{\pi(A)\pi(A^c)} \leq v_2 \frac{\int_{x \in A} \pi(dx) \widetilde{\mathcal{J}}(x, A^c)}{\pi(A)\pi(A^c)} = v_2 \widetilde{\kappa}(A)$$

   i.e.

   $$\kappa \leq v_2 \widetilde{\kappa} = 0$$

   thus $\kappa = 0$ and from Lemma 6.2.5 we have that $Gap(\mathcal{L}) = 0$.
2. If $\forall i : \pi_i$ is light-tailed.
   Then according to Theorem 6.4.4 we have that $Gap(L_i) > 0$ for all $i = 1, \ldots, m$ and and $Gap(L_{m+1}) > 0$ (because state space of this process $\mathcal{P}(\bar{M})$ is finite) thus (Lemma 6.2.7) $Gap(\widetilde{\mathcal{L}}) > 0$. Now applying Lemma 6.2.5 to $\widetilde{\mathcal{L}}$ we conclude that $\widetilde{\kappa} > 0$. Using the existence of $v_1 > 0$ in (6.167) we have

   $$\kappa := \frac{\int_{x \in A} \pi(dx) \mathcal{J}(x, A^c)}{\pi(A)\pi(A^c)} \geq v_1 \frac{\int_{x \in A} \pi(dx) \widetilde{\mathcal{J}}(x, A^c)}{\pi(A)\pi(A^c)} = v_1 \widetilde{\kappa} > 0$$

   i.e.

   $$\kappa > 0$$

   what implies (using Lemma 6.2.6) $Gap(\mathcal{J}) > 0$.

Note that (6.167) is equivalent to two conditions:

$$\inf_{\substack{A\subset E\\ \pi(A)\in(0,1)}}\left\{\frac{\sum_{x\in A}\pi(x)\mathcal{J}(x,A^c)}{\sum_{x\in A}\pi(x)\widetilde{\mathcal{J}}(x,A^c)}\right\}>0 \quad\text{and}\quad \sup_{\substack{A\subset E\\ \pi(A)\in(0,1)}}\left\{\frac{\sum_{x\in A}\pi(x)\mathcal{J}(x,A^c)}{\sum_{x\in A}\pi(x)\widetilde{\mathcal{J}}(x,A^c)}\right\}<\infty. \tag{6.168}$$

Denote

$$W(A)=\frac{\sum_{x\in A}\pi(x)\mathcal{J}(x,A^c)}{\sum_{x\in A}\pi(x)\widetilde{\mathcal{J}}(x,A^c)}$$

$$=\frac{\sum_{x\in A}\pi(x)\left(\sum_{T_{ij}x\in A^c}\mu_i(x_i)r_{ij}^{\bar D}+\sum_{T_{\cdot j}x\in A^c}\bar\lambda_j^{\bar D}+\sum_{T_{i\cdot}x\in A^c}\mu_i(x_i)r_{i0}^{\bar D}+\sum_{T_{\bar I\cdot}x\in A^c}\frac{A(\bar D\cup\bar I)}{A(\bar D)}+\sum_{T_{\cdot\bar H}x\in A^c}\frac{B(\bar D)}{B(\bar D\setminus\bar H)}\right)}{\sum_{T_{i\cdot}x\in A^c}\mu_i(x_i)+\sum_{T_{\cdot j}x\in A^c}\lambda_j+\sum_{T_{\bar I\cdot}x\in A^c}\frac{A(\bar D\cup\bar I)}{A(\bar D)}+\sum_{T_{\cdot\bar H}x\in A^c}\frac{B(\bar D)}{B(\bar D\setminus\bar H)}}.$$

Recall that $\mu^i=\sup_i\mu_i(x_i)<\infty$ and $\mu_i=\inf_y\mu_i(y)>0$.
Let $\mu^{max}=\max_{0\le i\le m}\mu^i<\infty$ and $\mu_{min}=\min_{0\le i\le m}\mu_i>0$.
Denote

$$M_{\mathcal{J}}(A,x)=\sum_{T_{ij}x\in A^c}\mu_i(x_i)r_{ij}^{\bar D}+\sum_{T_{\cdot j}x\in A^c}\bar\lambda_j^{\bar D}+\sum_{T_{i\cdot}x\in A^c}\mu_i(x_i)r_{i0}^{\bar D}+\sum_{T_{\bar I\cdot}x\in A^c}\frac{A(\bar D\cup\bar I)}{A(\bar D)}+\sum_{T_{\cdot\bar H}x\in A^c}\frac{B(\bar D)}{B(\bar D\setminus\bar H)}$$

and

$$M_{\mathcal{J}}^{min}=\inf_{A,x:M_{\mathcal{J}}(A,x)>0}\{M_{\mathcal{J}}(A,x)\},\quad M_{\mathcal{J}}^{max}=\sup_{A,x:M_{\mathcal{J}}(A,x)>0}\{M_{\mathcal{J}}(A,x)\}.$$

We have

$$M_{\mathcal{J}}^{min}\ \ge\ \min\left\{\mu_{min}\cdot\min_{i,j:r_{ij}>0}r_{ij},\ \min_{j:\bar\lambda_j>0}\bar\lambda_j,\ \mu_{min}\cdot\min_{i:r_{i0}>0}r_{i0},\ \min_{\substack{\bar D\subset\bar M,\bar I\subset\bar M\setminus\bar D\\ \frac{A(\bar D\cup\bar I)}{A(\bar D)}>0}}\left(\frac{A(\bar D\cup\bar I)}{A(\bar D)}\right),\ \min_{\substack{\bar D\subset\bar M,\bar H\subset\bar D\\ \frac{B(\bar D)}{B(\bar D\setminus\bar H)}>0}}\left(\frac{B(\bar D)}{B(\bar D\setminus\bar H)}\right)\right\}>0,$$

$$M_{\mathcal{J}}^{max}\ \le\ m(m-1)\mu^{max}\cdot1+\sum_{j=1}^m\bar\lambda_j+m\mu^{max}\cdot1+\sum_{\bar D\subset\bar M,\bar I\subset\bar M\setminus\bar D}\left(\frac{A(\bar D\cup\bar I)}{A(\bar D)}\right)+\sum_{\bar D\subset\bar M,\bar H\subset\bar D}\left(\frac{B(\bar D)}{B(\bar D\setminus\bar H)}\right)<\infty.$$

Similarly

$$M_{\widetilde{\mathcal{J}}}(A,x)=\sum_{T_{\cdot j}x\in A^c}\lambda_j+\sum_{T_{i\cdot}x\in A^c}\mu_i(x_i)++\sum_{T_{\bar I\cdot}x\in A^c}\frac{A(\bar D\cup\bar I)}{A(\bar D)}+\sum_{T_{\cdot\bar H}x\in A^c}\frac{B(\bar D)}{B(\bar D\setminus\bar H)}$$

and

$$M_{\widetilde{\mathcal{J}}}^{min}=\inf_{A,x:M_{\widetilde{\mathcal{J}}}(A,x)>0}\{M_{\widetilde{\mathcal{J}}}(A,x)\},\quad M_{\widetilde{\mathcal{J}}}^{max}=\sup_{A,x:M_{\widetilde{\mathcal{J}}}(A,x)>0}\{M_{\widetilde{\mathcal{J}}}(A,x)\}.$$

We have

$$M_{\widetilde{\mathcal{J}}}^{min}\ \ge\ \min\left\{\min_j\lambda_j,\mu_{min},\ \min_{\substack{\bar D\subset\bar M,\bar I\subset\bar M\setminus\bar D\\ \frac{A(\bar D\cup\bar I)}{A(\bar D)}>0}}\left(\frac{A(\bar D\cup\bar I)}{A(\bar D)}\right),\ \min_{\substack{\bar D\subset\bar M,\bar H\subset\bar D\\ \frac{B(\bar D)}{B(\bar D\setminus\bar H)}>0}}\left(\frac{B(\bar D)}{B(\bar D\setminus\bar H)}\right)\right\}>0,$$

$$M_{\widetilde{\mathcal{J}}}^{max}\ \le\ \sum_{j=1}^m\lambda_j+m\cdot\mu^{max}+\sum_{\bar D\subset\bar M,\bar I\subset\bar M\setminus\bar D}\left(\frac{A(\bar D\cup\bar I)}{A(\bar D)}\right)+\sum_{\bar D\subset\bar M,\bar H\subset\bar D}\left(\frac{B(\bar D)}{B(\bar D\setminus\bar H)}\right)<\infty.$$

Note that both $M_{\mathcal{J}}$ and $M_{\widetilde{\mathcal{J}}}$ do not depend on A. Define:

$$\partial\widetilde{A}=\{x:M_{\widetilde{\mathcal{J}}}(A,x)>0\},\qquad \partial A=\{x:M_{\mathcal{J}}(A,x)>0\}.$$

$\partial\widetilde{A}$ is a set of all states from which process with kernel $\widetilde{\mathcal{J}}$ can get in one step to A^c, whereas ∂A is a set of states from which original process with kernel $\mathcal{J}$ can get in one step to A^c. We have

$$W(A) = \frac{\sum_{x\in\partial A} \pi(x) M_{\mathcal{J}}(A,x)}{\sum_{x\in\partial \widetilde{A}} \pi(x) M_{\widetilde{\mathcal{J}}}(A,x)}$$

and

$$\frac{M_{\mathcal{J}}^{max}}{M_{\widetilde{\mathcal{J}}}^{min}} \cdot \frac{\sum_{x\in\partial A} \pi(x)}{\sum_{x\in\partial \widetilde{A}} \pi(x)} \geq \frac{\sum_{x\in\partial A} \pi(x) M_{\mathcal{J}}(A,x)}{\sum_{x\in\partial \widetilde{A}} \pi(x) M_{\widetilde{\mathcal{J}}}(A,x)} \geq \frac{M_{\mathcal{J}}^{min}}{M_{\widetilde{\mathcal{J}}}^{max}} \cdot \frac{\sum_{x\in\partial A} \pi(x)}{\sum_{x\in\partial \widetilde{A}} \pi(x)}$$

To show (6.154) we have to show that

$$\infty > \frac{\sum_{x\in\partial A} \pi(x)}{\sum_{x\in\partial \widetilde{A}} \pi(x)} > 0$$

or equivalently

$$\infty > \frac{\sum_{x\in\partial \widetilde{A}} \pi(x)}{\sum_{x\in\partial A} \pi(x)} > 0. \tag{6.169}$$

Before proceeding let us examine how much can differ $\pi(x)$ and $\pi(x')$, where x' is a state to which one can get from x in at most $k \leq m$ steps movements within the and both x and x' have the same set of broken nodes $\bar{I}$. Recall (6.166) that for $x = (\bar{I}, x_1, \ldots, x_m) \in \mathcal{P}(\bar{M}) \times \mathbb{Z}_+^m$ we have:

$$\pi(x) = \pi(\bar{I}, x_1, \ldots, x_m) = \frac{1}{C}\frac{A(\bar{I})}{B(\bar{I})}\prod_{i=1}^{m} \pi_i(x_i), \qquad \text{where } \pi_i(x_i) := \frac{1}{b_i}\frac{\lambda_i^{x_i}}{\prod_{y=1}^{x_i} \mu_i(y)}$$

and for $x' = (\bar{I}, x'_1, \ldots, x'_m) \in \mathcal{P}(\bar{M}) \times \mathbb{Z}_+^m$ we have

$$\pi(x') = \pi(\bar{I}, x'_1, \ldots, x'_m) = \frac{1}{C}\frac{A(\bar{I})}{B(\bar{I})}\prod_{i=1}^{m} \pi_i(x'_i), \qquad \text{where } \pi_i(x'_i) := \frac{1}{b_i}\frac{\lambda_i^{x'_i}}{\prod_{y=1}^{x'_i} \mu_i(y)}.$$

Note that x_i and x'_i can only differ by at most k for $i, i = 1, \ldots, m$. Exactly the same calculations as for standard Jackson networks show that (6.156) holds, i.e.

$$\frac{1}{c}\pi(x') \leq \pi(x) \leq c\pi(x'), \tag{6.170}$$

where $c = \prod_{i=1}^{k} c_i, \quad c_i = \max\left(\left(\frac{\lambda_i}{\mu^i}\right), \left(\frac{\lambda_i}{\mu^i}\right)^{-1}\left(\frac{\lambda_i}{\mu_i}\right), \left(\frac{\lambda_i}{\mu_i}\right)\right)$.

Of course we can have states which are in ∂A and not in $\partial\widetilde{A}$ and also such that are in $\partial\widetilde{A}$ but not in ∂A.
If we take $z \in \partial A \setminus \partial\widetilde{A}$, i.e. there exists some $y_z \in A^c$ such that process with kernel $\widetilde{\mathcal{J}}$ cannot move from z to y_z in one step, but process with kernel $\mathcal{J}$ can. State y_z must be of form $y_z = T_{ij}z$, it cannot be of form $y_z = T_{\bar{I}\cdot}z$ or $y_z = T_{\cdot\bar{H}}z$ because changing only set of broken nodes in z is always possible in both processes, i.e. one driven by $\mathcal{J}$ and $\widetilde{\mathcal{J}}$. Of course neither it can be of form $y_z = T_{i\cdot}z$ nor $y_z = T_{\cdot j}z$, because then process driven by $\widetilde{\mathcal{J}}$ could also move in this direction in one step.
But note that there always exists path $T_{i\cdot}, T_{\cdot j}$ such that $z' = T_{i\cdot z} \in A$ and $y_z = T_{\cdot j}z' \in A^c$.
Of course either $z' \in \partial\widetilde{A} \setminus \partial A$ or $z' \in \partial\widetilde{A} \cap \partial A$.
For any such z' there are surely no more than $2m$ corresponding z, thus using (6.170) (where $c \geq 1$) we have

$$\frac{\sum_{x\in\partial\widetilde{A}} \pi(x)}{\sum_{x\in\partial A} \pi(x)} = \frac{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial\widetilde{A}\setminus\partial A} \pi(x)}{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial A\setminus\partial\widetilde{A}} \pi(x)} \geq \frac{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial\widetilde{A}\setminus\partial A} \pi(x)}{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + 2mc\left(\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial\widetilde{A}\setminus\partial A} \pi(x)\right)}$$

$$\geq \frac{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial\widetilde{A}\setminus\partial A} \pi(x)}{(2mc+1)\left(\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial\widetilde{A}\setminus\partial A} \pi(x)\right)} = \frac{1}{2mc+1}$$

Similarly fix $x \in \partial\widetilde{A} \setminus \partial A$, i.e. there exists some $y_x \in A^c$ such that original process with kernel $\mathcal{J}$ cannot move there in one step, but process with kernel $\widetilde{\mathcal{J}}$ can. State y_x must be of form $y_x = T_{\cdot i}x$ or $y_x = T_{j\cdot}x$, it cannot be of form $y_x = T_{\bar{I}\cdot}y$ or $y_x = T_{\cdot\bar{H}}$, because changing only set of broken nodes in z is always possible in both processes, i.e. driven by $\mathcal{J}$ and $\widetilde{\mathcal{J}}$. Then there exists path of length at most k such that process with kernel $\mathcal{J}$ can, moving along the path, get to y_x. The path consists only of $T_{\cdot i}, T_{j\cdot}, T_{ij}$ (it does not contain any $T_{\bar{I}\cdot}, T_{\cdot\bar{H}}$, because such transitions are always possible in both processes, with kernel $\widetilde{\mathcal{J}}$ and $\mathcal{J}$). There must exists state x' on this path such that $x' \in A$ and next state on the path is in A^c. Of course either $x' \in \partial\widetilde{A} \cap \partial A$ or $x' \in \partial A \setminus \partial\widetilde{A}$. For any such x' there can be surely no more than $(2k)^m$ states corresponding to it. Thus, using (6.156) (where $c \geq 1$) we have

$$\frac{\sum_{x\in\partial\widetilde{A}} \pi(x)}{\sum_{x\in\partial A} \pi(x)} = \frac{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial\widetilde{A}\setminus\partial A} \pi(x)}{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial A\setminus\partial\widetilde{A}} \pi(x)} \leq \frac{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + c(2k)^m \left(\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial A\setminus\partial\widetilde{A}} \pi(x)\right)}{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial A\setminus\partial\widetilde{A}} \pi(x)}$$

$$\leq \frac{(1 + c(2k)^m)\left(\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial A\setminus\partial\widetilde{A}} \pi(x)\right)}{\sum_{x\in\partial\widetilde{A}\cap\partial A} \pi(x) + \sum_{x\in\partial A\setminus\partial\widetilde{A}} \pi(x)} = (1 + c(2k)^m) < \infty.$$

This way we have shown that $W(A)$ is bounded uniformly (for any $A : \pi(A) \in (0,1)$) from below and above:

$$\infty > (1 + c(2k)^m) \cdot \frac{M_{\mathcal{J}}^{max}}{M_{\widetilde{\mathcal{J}}}^{min}} \geq \frac{\sum_{x\in\partial\widetilde{A}} \pi(x) M_{\mathcal{J}}(A, x)}{\sum_{x\in\partial A} \pi(x) M_{\widetilde{\mathcal{J}}}(A, x)} \geq \frac{M_{\mathcal{J}}^{min}}{M_{\widetilde{\mathcal{J}}}^{max}} \cdot \frac{1}{2mc + 1} > 0$$

i.e. (6.154).

□

6.5.3 Consequences of existence of spectral gap for rate of convergence

Spectral gap controls the exponential rate of convergence in $L_2(\pi)$ in the sense that it is the largest ε for which

$$||P_t f - \int f d\pi||_\pi \leq e^{-\varepsilon t}||f - \int f d\pi||_\pi \tag{6.171}$$

for all $f \in L_2(\pi)$ and $t \geq 0$. To see this suppose $||f||_\pi = 1$ and $(f, \mathbf{1})_\pi = 0$, then we have

$$\frac{d}{dt}||P_t f||_\pi^2 = 2(P_t f, LP_t f)_\pi \leq -2Gap(L)||P_t f||^2,$$

i.e.

$$\frac{\frac{d}{dt}||P_t f||_\pi^2}{||P_t f||_\pi^2} \leq -2Gap(L),$$

which means $||P_t f||_\pi^2 \leq e^{-2Gap(L)}$. Equivalently (because $||f||_\pi = 1$)

$$||P_t f||_\pi \leq e^{-Gap(L)}||f||_\pi.$$

We say that exponential L_2 convergence occurs if there exists spectral gap, i.e. $Gap(L) > 0$.
Thus in previous sections we showed that for Jackson network, both standard and unreliable, we have exponential convergence rate if and only if each of marginal stationary distribution is light-tailed (plus some other natural assumptions on the parameters of the network).

Bibliography

[1] Aldous, D.J. Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes. *Probability Theory and Related Fields*, **77**, 137–145, (1988).

[2] Aldous, D.J., Diaconis, P. Shuffling Cards and Stopping Times. *American Mathematical Monthly*, **93**, 333–348, (1986).

[3] Aldous, D.J., Diaconis, P. Strong uniform times and finite random walks. *Advances in Applied Mathematics*, **8**, 69–97, (1987).

[4] Aldous, D.J., Fill, J.A. Reversible Markov Chains and Random Walks on Graphs. *Monograph in preperation:* `http://www.stat.berkeley.edu/users/aldous/RWG/book.html`

[5] Brown, M. Consequences of Monotonicity for Markov Transition Functions. *Technical report, City college, CUNY*, (1990).

[6] Chen, Mu-Fa *From Markov Chains to Non-Equilibrium Particle Systems.* World Scientific, Singapore, (1992).

[7] Chen, M., Wang, F. Cheeger's inequality for general symmetric forms and existence criteria for spectral gap. *The Annals of Applied Probability*, **28**, 235–257, (2000).

[8] Chung, Fan R.K. Laplacians and Cheeger's inequality for directed graphs. *Preprint.*

[9] Cinlar, E. *Introduction to Stochastic Processes.* Springer-Varlag, New York, (1975).

[10] Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C. *Introduction to Algorithms.*

[11] Daduna, H., Szekli, R. Dependencies in Markovian Networks. *The Annals of Applied Probability*, **25**, 226–254, (1995).

[12] Daley, D.J. Stochastically Monotone Markov Chains. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, **10**, 305–317, (1968).

[13] Diaconis, P. *Group Representations in Probability and Statistics.* Harvard University, (1988).

[14] Diaconis, P., Bayer D. Trailing the Dovetail Shuffle to Its Lair. *The Annals of Probability*, **2**, 294–313, (1992).

[15] Diaconis, P., Fill, J.A. Strong stationary times via a new form of duality. *The Annals of Probability*, **18**, 1483-1522, (1990).

[16] Diaconis, P., Stroock, D. Geometric bounds for eigenvalues of Markov Chains. *The Annals of Probability*, **1**, 36–61, (1991).

[17] Dynia, M., Kutyłowski, J., Lorek, P., Meyer auf der Heide, F. Maintaining Communication Between an Explorer and a Base Station. *In Proceedings of the 1st IFIP International Conference on Biologically Inspired Cooperative Computing (BICC 2006), volume **216** of IFIP International Federation for Information Processing*, Boston, MA, USA (2006).

[18] Fayolle, G., Malyshev, V.A., Menshikov, M.V., Sidorenko, A.F. Lyaponov functions for Jackson networks. *Rapports de Rechere*, (1991).

[19] Feller, W. *Wstęp do Rachunku Prawdopodobieństwa.* Państwowe Wydawnictwo Naukowe, Warszawa, (1977).

[20] Fill, J.A. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *The Annals of Applied Probability*, **1**, 92–87, (1991).

[21] Fisz, M. *Rachunek Prawdopodobieństwa i Statystyka Matematyczna.* Państwowe Wydawnictwo Naukowe, Warszawa (1969).

[22] Fulman, J., Wilmer, E.L Comparing eigenvalue bounds for Markov chains: When does Poincare beat Cheeger? *Preprint.*

[23] Häggström, O. *Finite Markov Chains and Alogrithmic Applications.* Cambridge University Press, (2002).

[24] Hordijk, A., Spieksma, F. d On ergodicity and recurrence properties of Markov Chain with an application to an open Jackson Network. *Adv. Appl. Prob.*, **24**, 343–376, (1992).

[25] Horn, R., Johnson, C. *Matrix Analysis.* Cambridge Univ. Press., (1985).

[26] Jackson, J.R. Networks of waiting lines. *Operations Research*, **4**, 518–521, (1957).

[27] Keilson, J. *Markov Chain Models - Rarity and Exponentiality.* Springer-Verlag, New York Heidelberg Berlin, (1979).

[28] Keilson, J., Kester, A. Monotone Matrices and Monotone Markov Processes. *Stochastic Processes and their Applications*, **5**, 231–241, (1977).

[29] Kingman, J.F.C. *Regenerative Phenomena.* John Wiley & Sons, INC, New York, (1972).

[30] Lawler, G.F., Sokal, A.D. Rapid convergence to equilibrium in one dimensional stochastic Ising models. *Transaction of the Amercian Society*, **309** , 557–580, (1988).

[31] Liggett, T.H. Exponential L_2 convergence of attractive reversible nearest particle systems. *The Annals of Probability*, **17**, 403–432, (1989).

[32] Liggett, T.H. *Interacting Particle Systems.* Springer-Varlag , New York, (1985).

[33] Liggett, T.H. L_2 rates of convergence of attractive reversible nearest particle systems: critical case. *he Annals of Probability*, **19**, 935–959, (1991).

[34] Lindvall, T. *Lectures on the Coupling Method.* John Wiley & Sons, INC, New York (1992).

[35] Lindvall, T. Stochastic Monotonicities in Jackson Queueing Networks. *Probability in the Engineering and Informational Sciences*, **11**, 1–9, (1997).

[36] McDonald, D., Iscoe, I. Asymptotics of exit times for markov jump processes II: applications to jackson networks. *The Annals of Probability*, **22**, 2168–2182, (1994).

[37] Pak, I. Random Walk on Groups : Strong Uniform Time Approach. *Ph.D. Thesis, Harvard University* `http://www-math.mit.edu/~pak/time57.ps`, (1997).

[38] Propp, J., Wilson, D. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Alogrithms*, **9**, 232–252, (1996).

[39] Rolski, T., Schmidli, H., Schmidt, V., Teugels, J. *Stochastic Processes for Insurance and Finance.* John Wiley & Sons, (1999).

[40] Saloff-Coste, L. Random Walks on Finite Groups. *Preprint* `http://www-stat.stanford.edu/~cgates/PERSI/papers/rwfg.pdf`, (1997).

[41] Sauer, C. Stochastic Product From Networks with Unreliable Nodes: Analysis of Performance and Avaibility *Ph.D. Thesis, Hamburg*, (2006).

[42] Sauer, C., Daduna, H. Availability formulas and performance measures for seperable degredable networks. *Economic Quality Control*, **18(2)**, 165–194, (2003).

[43] Seneta, E. *Non-negative matrices and Markov Chains.* Springer-Verlag, (1973).

[44] Stoyan, D., Mueller, A. *Comparison Methods for Stochastic Models and Risks.* J. Wiley & Sons, Chichester, (2002).

[45] Szekli, R. *Stochastic Ordering and Dependence in Applied Probability.* Springer-Verlag, New York, (1995).

[46] Thorisson, H. *Coupling, Stationarity, and Regeneration.* Springer-Verlag, New York, (2000).

[47] Tweedie, R.L., Roberts, G.O. Rates of convergence of stochastically monotone and continuous time Markov models. *Journal of Applied Probability*, **37**, 359–373, (2000).

[48] Wilmer, E.L. Exact Rates of Convergence for Some Simple Non-Reversible Markov Chains. *Ph.D. Thesis*