J. Math. Anal. Appl. 481 (2020) 123488

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

On the fastest finite Markov processes [☆]

Vivek Borkar [a,*], Laurent Miclo [b,c,*]

[a] Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India
[b] Toulouse School of Economics, UMR 5314, Toulouse, France
[c] Institut de Mathematiques de Toulouse, UMR 5219, CNRS and University of Toulouse, Toulouse, France

ARTICLE INFO

Article history:
Received 9 March 2019
Available online 10 September 2019
Submitted by J.A. Filar

Keywords:
Fastest Markov chains/processes
Communication speed
Spectra of Markov operators
Hamiltonian cycles
Dynamic programming
Differentiation of Markov operators

ABSTRACT

Consider a finite state irreducible Markov process with transition graph G and invariant probability distribution π. Its inverse communication speed is defined as the expectation of the time to go from x to y when x, y are sampled independently according to π. We study this in the context of both discrete and continuous time. One of our goals is to show (under a suitable normalization condition on the transition rates for the continuous time case) that the shortest inverse communication speed among all Markov processes compatible with G and π is attained on those whose successive positions follow a Hamiltonian cycle, assuming one exists, and that π is close enough to the uniform distribution υ. This result is no longer true when π is sufficiently different from υ. Another purpose of the paper is to prove that when the invariance with respect to π is dropped and the inverse communication speed is replaced by the unweighted sum of the hitting times, then the Hamiltonian cycles, when they exist, are still the minimizers over all processes compatible with the prescribed directed graph G of permitted transitions, not only Markov processes, thus extending some previous results.

1. Introduction

Our overall goal in this paper is to investigate, among certain classes of chains and processes, those which are the fastest, in the sense that their communication times are the smallest possible. A formal statement of this claim needs some notation to be set up, which we do next.

Given a finite oriented (strongly) connected graph $G = (V, E)$, a random chain $(X_n)_{n\in\mathbb{Z}_+}$ or a random process $(X_t)_{t\in\mathbb{R}_+}$ is said to be **compatible** with G if it takes values in the vertex set V and its transitions are restricted to the edge set E (assumed to contain all the self-loops to allow for the random chain/process to stay at the same place). To indicate that the random chain/process starts from $x \in V$, we put x as

☆ Work supported by the Indo-French Centre for Applied Mathematics grant INT/FRANCE/IFCAM/2017 (for VB and LM), a J.C. Bose Fellowship (for VB), and Agence Nationale de la Recherche (France) grant ANR-12-BS01-0019 (for LM).

* Corresponding authors.
E-mail addresses: borkar@ee.iitb.ac.in (V. Borkar), miclo@cnrs.fr (L. Miclo).

https://doi.org/10.1016/j.jmaa.2019.123488
0022-247X/

a subscript of the underlying probability $\mathbb{P}_x$ and expectation $\mathbb{E}_x$. Another useful notation is τ_y, the first hitting time of a point $y \in V$ by the random chain/process. When in addition we are given a probability measure π on V, we will also be interested in Markov chains/processes compatible with G and admitting π as invariant measure or fixed marginal.

More precisely, we are interested in the following two problems:

- **(P1)**: To minimize $\mathfrak{F} := \sum_{x\in V}(\sum_{y\in V} \mathbb{E}_x[\tau_y])$ over all G-compatible Markov chains or more general random processes.
- **(P2)**: To minimize the expectation $\mathfrak{F}_\pi$ of the time needed to go from x to y when x and y are independently sampled according to a probability distribution π on V, over all Markov chains that admit π as stationary distribution. That is, we seek to minimize

$$\mathfrak{F}_\pi := \sum_{x\in V} \pi(x) \left(\sum_{y\in V} \pi(y)\mathbb{E}_x[\tau_y] \right) \tag{1}$$

where π is the prescribed stationary distribution.

Throughout this paper, υ will stand for the uniform distribution over V. Note that the first problem is *not* a special case of the second with $\pi = \upsilon$, because the minimization in the first case will be over *all* controlled Markov chains/processes, not necessarily those that have υ as the stationary distribution. In either case, this is distinct from the notion of rapid mixing [30]. It is known that the term in parentheses in (1) is independent of x and is called the Kemeny constant. The so-called *eigentime identity* relates this constant to the spectrum of the transition matrix P, in fact it turns out that it equals the trace of the fundamental matrix. That is, letting Π denote the rank one matrix whose rows are identical and equal to the stationary distribution π of a stochastic matrix P written as a row vector,

$$\sum_{x\in V} \pi(x) \left(\sum_{y\in V} \pi(y)\mathbb{E}_x\left[\tau_y\right] \right) = \mathrm{tr}\left((I - P + \Pi)^{-1}\right) - 1. \tag{2}$$

See [24] for some early contributions to this circle of ideas and a historical perspective.

The above equivalence connects this problem to another strand of research initiated in [19], [20], which in fact has motivated the present work. These and subsequent works mentioned below embed the Hamiltonian cycle problem into a Markov decision process for singularly perturbed Markov chains. The hope is to have the well developed techniques of Markov decision theory bear upon the Hamiltonian cycle problem, both in order to get additional theoretical insights and to suggest novel heuristics. We summarize below the key developments in this strand of research.

This embedding is done as follows. The state space of the Markov chain is the node set V of the directed graph and at each node i, the control set is the edge set $E_i := \{(i, j) : (i, j) \in E\}$ of outgoing edges at i. When the controlled Markov chain is in state i, a randomized control, depending on the past state-control sequence and possibly some independent randomization, chooses a probability distribution on E_i and picks an edge (i, j) according to it. The chain then moves to j. Thus the controlled transition probabilities are $P(X_{t+1} = k|X_t = i, U_t = (i, j)) = \delta_{jk}$, the Kronecker delta. This transition probability is further modified by taking a convex combination thereof with another transition probability with a weight $1 - \epsilon$ for the former, where $0 < \epsilon << 1$, making it a singularly perturbed Markov chain. The choice of the latter transition probability corresponds to a deterministic (i.e., probability 1) transition to a prescribed 'home' node. The aim of the perturbation is to render the chain uni-chain in the sense that there is a nonzero probability path from any node to the home node regardless of the control choices.

In [19], such an embedding is used in order to obtain a new linear programming relaxation of the Hamiltonian cycle problem, whereas [20] characterized Hamiltonian cycles (when they exist) as the minimizers over all deterministic stationary policies, for $\epsilon > 0$ sufficiently small, of the $(1,1)$-th element (i.e., the top diagonal term) of the fundamental matrix $(I - P_\epsilon + \Pi_\epsilon)^{-1}$, where $I, P_\epsilon, \Pi_\epsilon$ denote respectively the identity matrix, the perturbed transition matrix, and the rank one matrix with identical rows equal to the unique stationary distribution under P_ϵ, written as a row vector.

It so happens that this matrix also appears in an asymptotic expansion due to Blackwell for infinite horizon discounted cost Markov decision process. A precise bound on the perturbation parameter for which the above optimality claim holds was obtained in [12]. For the special case of uniform π, i.e., doubly stochastic P, this was refined to 'minimum over all stationary policies', deterministic or otherwise, in [5]. The minimum of the functional under consideration is strictly higher for a non-Hamiltonian graph with the same number of nodes and this gap was estimated in [6].

Another important work in this vein is [17] which related the Hamiltonian cycle problem to constrained Markov decision processes, proving as a result the NP-hardness of the latter.

The culmination of this line of research is in the work of Litvak and Ejov [28] who used the trace of the fundamental matrix rather than its $(1,1)$th element, as the 'cost' to be minimized, and showed that Hamiltonian cycles, when they exist, minimize it over *all* stochastic matrices, not just doubly stochastic, that are compatible with G, regardless of whether they are irreducible or not. In Markov decision theoretic parlance, this implies optimality over all policies for the unperturbed (i.e., $\epsilon = 0$) Markov decision process wherein the control at any time is a prescribed function of the current state, i.e., a 'stationary Markov policy'. In view of the known relation (2) of this trace with hitting times, they essentially established that if G contains a Hamiltonian cycle, then the fastest Markov chains are exactly those following deterministically the succession of states given by a Hamiltonian cycle (the corresponding value of the objective function or cost $\mathfrak{F}$ does not depend on the choice of the admissible Hamiltonian cycle). This also improves upon the results of [24] which considered only ergodic chains. The latter work also includes a historical account of the so-called 'random target lemma' or 'eigentime identity' described in the next section, going back to [27].

As already mentioned, a major motivation for this line of research has been to come up with good heuristic algorithms for the Hamiltonian cycle problem and there are a number of articles that deal with this, such as [11]. We do not dwell on this issue here. These have been extensively surveyed in [7], [18], which also surveys a related and similarly motivated optimization problem wherein one seeks to maximize the determinant of a certain rank one perturbation of $I - P_0$.

In yet another related work [8], the Markov decision theoretic formulation is leveraged to map the Hamiltonian cycle problem to that of comparing volumes of two convex sets.

There have also been other theoretical spin-offs leading to fine analysis of the structure of the problem. This includes exploration of matrices corresponding to Hamiltonian cycles [3], [10], [13] and studies into the structure of discounted occupation measures arising from these problems [15], [21], [16].

We can expect the optimality of Hamiltonian cycle to persist under small perturbations of the Markov chains considered above. For a different class of perturbations, such 'stability' results were established in Ejov et al. [14]. Specifically, they considered perturbations obtained by taking convex combination with a stochastic matrix whose all elements are identical, with a small weight for the latter. This in particular renders the chain irreducible. They calculate the exact perturbation error, proving simultaneously the optimality of the Hamiltonian cycle and its robustness to perturbation.

To summarize, the two key issues (though not necessarily the only ones) that have been central to the above line of research are: the optimality of the Hamiltonian cycle for a suitable Markov decision process, and the robustness of this claim vis-a-vis a suitable class of perturbations. In this paper we strengthen both these claims. Thus we consider $\mathfrak{F}$ above as the basic object which makes sense even for general non-Markov processes compatible with the given graph. This leads to the problem **(P1)** above, where our objectives are:

- (A) *To establish the above optimality over a larger class of processes:*
  We strengthen the above claim to show the optimality of the stationary Markov policy corresponding to a Hamiltonian cycle, when one exists, over all policies, i.e., policies that can depend on the entire past and exogenous independent randomness. Observe that this is as general as it can get, because any G-compatible process can be cast as a controlled Markov chain by considering the conditional law of the next state given the past itself as the control depending on the entire past.
  Also, the technique we use to prove this generalization is different, viz., a dynamic programming based argument that goes back to Held and Karp [23]. This requires converting this non-classical Markov decision process to a classical one by suitably re-defining the state space. This strengthens the result of Litvak and Ejov [28] which establishes optimality among stationary Markov policies. This improvement is facilitated by an intemediate technical lemma which allows us to identify the trace maximization problem with that of maximizing a sum of hitting times to all nodes starting from a fixed node.
- (B) *To extend the above result to the continuous time framework under an appropriate normalization of the jump rates:*
  The extension of the above problem to continuous time is of independent interest because it is neither automatic nor trivial. This is because there is also an exponentially distributed sojourn time at each state. Unless the rate matrix is constrained, the problem becomes ill posed because we can make the rates arbitrarily high to push mean hitting times as close to zero as desired. Another difference with the discrete time set-up is that for any stationary Markov chain, the stationary distribution can be changed arbitrarily keeping the support constant by simply modulating the sojourn times without altering the graph or the transition probabilities. Hence we need to normalize the rates in order to make the problem well-posed. We impose a condition that the sum of rates, i.e., the ℓ_1-norm $\|\Lambda\|_1$ of the rate vector Λ, is kept constant. This can be motivated as follows. In order to make the problem scale-invariant so as to avoid the well-posedness issue above, it makes sense to set a symmetric, positive, positively 1-homogeneous smooth function F (i.e., $F(a\Lambda) = aF(\Lambda)\ \forall\ a > 0$) equal to a constant. Our choice of the constraint $\|\Lambda\|_1 = 1$ is a specific instance of this, corresponding to $F(\Lambda) = \|\Lambda\|_1$.

This done, we take up next the problem of perturbations in the spirit of Ejov et al. [14], but again with a key difference. This is our problem **(P2)**. We consider a different kind of perturbation here, viz., one that keeps the stationary distribution π in a small neighborhood of the uniform distribution υ. The objective can be summarized as:

- (C) *To begin an investigation of the situation where π is not the uniform distribution:*
  We show that:
  - (C1) when G contains a Hamiltonian cycle and that π is close to the uniform distribution υ, the fastest Markov chains/processes are still those associated with Hamiltonian cycles, and,
  - (C2) that this is no longer true when π is 'far away' from υ.

  In establishing this, we first employ the continuous time framework which proves to be more convenient because of the handy tools of calculus of Markov generators one is able to draw upon. In fact, we finally treat the discrete time problem by mapping it to a continuous time one in Section 5. This highlights further differences between the two: the greater flexibility of the continuous time framework in modulating sojourn times compared to discrete time chains with self-loops leads to a potentially lower minimum.
  Since, as in [28], we show optimality of Hamiltonian cycles which corresponds in particular to uniform stationary distribution among all choices, the same is also optimal among all transition matrices corresponding to uniform stationary distribution, i.e., doubly stochastic matrices. The problem, however, becomes hard if we *fix* as above a stationary distribution π other than uniform and seek an optimum

among Markov chains with π as their stationary distribution. In continuous time, given any transition graph, the stationary distribution can be modified at will by changing the mean sojourn times, as long as the support is kept fixed. This suggests in particular the possibility of reducing the problem to one for uniform distribution. Our results above show that this is not possible.

That said, the criterion of minimizing the sum of π-weighted or unweighted mean hitting time of every other state from a base state is an optimization problem of independent interest. A potential application is to broadcasting messages over a communication network where this suggests using a ring sub-network. Another application is to a negative result in search discussed later in this article. Our results show that to search for a binary string of length l among all such in absence of prior knowledge, no search algorithm, random and history-dependent or otherwise, can do better than to list the strings and check them one by one.

The plan of the paper is as follows. The above results (A) and (B) are proved in the next section via a dynamic programming approach, which also provides as a corollary an alternative proof of the discrete time result of Litvak and Ejov [28] for the special case of $\pi = \upsilon$. In Section 3, we prove (C), see Theorem 8, by using small perturbations of the uniform probability measure. At the other extreme, the proof of Theorem 13 in Section 4 deals with large perturbations. Section 5 contains some observations about the links between continuous time and discrete time. The first part of the appendix is quite long, as it presents the technical tools needed in the proof of Theorem 8: we decompose the generators leaving π invariant into convex sums of generators associated with (not necessarily Hamiltonian) cycles and we differentiate the expectations of hitting times with respect to the generators. In the second part, we compute the fastest normalized birth and death generators leaving invariant any fixed probability measure π on $\{0, 1, 2\}$. The underlying graph is the segment graph of length 2, which is the simplest example not containing a Hamiltonian cycle.

2. The dynamic programming approach

2.1. Introduction

The aim of this section is to address the first problem proposed above, flagged as **(P1)** in the introduction, i.e., to show that the Hamiltonian cycles, when one exists, are the fastest in the sense we have defined, among *all* processes compatible with the given graph, not just Markov chains. The proof uses dynamic programming. We first recall the eigentime identity in the next subsection and then establish the desired result for resp., discrete and continuous time in the subsections that follow.

2.2. The eigentime identity

We shall use the notation $\mathcal{L}(X)$ to denote the law of a random variable X and $|A|$ for the cardinality of a finite set A. Consider a discrete time Markov chain $(X_n)_{n\in\mathbb{Z}_+}$ on a finite state space V with transition matrix $P = (p(x, y))_{x,y\in V}$. We assume it to be **irreducible**, i.e., for any $x, y \in V$, there exists a path $x_0 = x, x_1, \cdots, x_{n-1}, x_n = y$ such that $p(x_k, x_{k+1}) > 0$, for $k \in [\![0, n-1]\!] := \{0, 1, ..., n-1\}$. Let $\pi := (\pi(x))_{x\in V}$ denote its unique **stationary distribution**, which is the left eigenvector of P corresponding to the Perron-Frobenius eigenvalue $\theta_1 = 1$. In particular, if $\mathcal{L}(X_0) = \pi$, then for any $n \in \mathbb{Z}_+$, we have $\mathcal{L}(X_n) = \pi$. This justifies the term 'stationary', its uniqueness being a well-known consequence of the irreducibility hypothesis. Denote by $\theta_2, \cdots, \theta_N$ the remaining eigenvalues of P. Also define the hitting times

$$\forall\, x \in V, \qquad \tau_x := \min\{n \in \mathbb{Z}_+ \,:\, X_n = x\}. \tag{3}$$

The **eigentime identity** states that

$$\forall\, x\in V,\qquad \sum_{y}\pi(y)\mathbb{E}_x[\tau_y] \;=\; \sum_{m=2}^{N}\frac{1}{1-\theta_m}. \tag{4}$$

where $N := |V|$ and each eigenvalue is counted as many times as its (algebraic) multiplicity. For reversible chains, this is Proposition 3.13, p. 75, of Aldous and Fill [2]. It was extended to the general case in Cui and Mao [9], see also [29] for a simple proof and further extensions. The left hand side gives the mean hitting time of a target state picked randomly with distribution π. A minor modification is to consider instead the stopping times

$$\forall\, x\in V,\qquad T_x \;:=\; \min\{n\in\mathbb{N}\;:\;X_n=x\},$$

where $\mathbb{N} := \mathbb{Z}_+\setminus\{0\}$, the set of positive integers. Since $\pi(x)\mathbb{E}_x[T_x] = 1$, we can replace (4) by

$$\forall\, x\in V,\qquad \sum_{y}\pi(y)\mathbb{E}_x[T_y] \;=\; 1+\sum_{m=2}^{N}\frac{1}{1-\theta_m}. \tag{5}$$

This variant is essentially contained in Theorem 2.4 of Hunter [24]. An immediate corollary is the 'symmetrized' version

$$\sum_{x,y}\pi(x)\pi(y)\mathbb{E}_x[\tau_y] = \sum_{m=2}^{N}\frac{1}{1-\theta_m}, \tag{6}$$

$$\sum_{x,y}\pi(x)\pi(y)\mathbb{E}_x[T_y] = 1+\sum_{m=2}^{N}\frac{1}{1-\theta_m}. \tag{7}$$

Let Π be the rank one matrix whose rows are all equal to π. From Theorem 2.4 of Hunter [24], (7) equals the trace of the 'fundamental matrix' $(I-P+\Pi)^{-1}$, where I is the identity matrix.

The notion of a fundamental matrix extends to the chains that are not irreducible if we define Π by the Cesaro sum

$$\Pi := \lim_{n\uparrow\infty}\frac{1}{n+1}\sum_{m=0}^{n}P^n.$$

A **Hamiltonian cycle** A of V is an ordering $(a_0, a_1, ..., a_{N-1})$ of the elements of V and $a_N = a_0$, as the indices should be seen as elements of $\mathbb{Z}_N := \mathbb{Z}/(N\mathbb{Z})$. More precisely, the cycles $(a_0, a_1, ..., a_{N-1})$ and $(a_k, a_{k+1}, ..., a_{k+N-1})$ should be identified as the same cycle, for all $k\in\mathbb{Z}_N$. This will be implicit in the sequel, even though for notational convenience, we will represent a cycle A as $(a_0, a_1, ..., a_{N-1})$. Consider an irreducible directed graph $G = (V, E)$ where V, E denote respectively its node and edge sets. In the discrete time setting, such graphs will always be assumed to contain all the self-loops, i.e. $(i, i)\in E$ for any $i\in V$. This assumption is used mainly in Section 5. The Hamiltonian cycle $A := (a_0, a_1, ..., a_{N-1})$ is said to be **admissible** for G if $(a_k, a_{k+1})\in E$ for all $k\in\mathbb{Z}_N$. It means that G_A is a subgraph of G, where G_A is the oriented graph on V whose edges are the (a_k, a_{k+1}), for $k\in\mathbb{Z}_N$. The set of all Hamiltonian cycles (respectively, those admissible for G) is denoted $\mathcal{H}$ (resp., $\mathcal{H}(G)$) and the graph G is said to be **Hamiltonian** if $\mathcal{H}(G)\neq\emptyset$. Obviously, we have $\mathcal{H} = \mathcal{H}(K_V)$, where K_V is the complete oriented graph on V.

Consider the optimization problem of minimizing (4)/(6), or equivalently, (5)/(7), where π is equal to υ, over all irreducible P **compatible** with the given graph G (in the sense that for any $x\neq y\in V$, $P(x, y) > 0 \Rightarrow (x, y)\in E$). Since these quantities will be infinite for reducible P, we might as well consider the problem of minimizing it over all stochastic matrices P compatible with G. Say that P is Hamiltonian if there exists a Hamiltonian cycle $(a_1, \cdots, a_N)$ such that $p(a_k, a_{k+1}) = 1 = p(a_N, a_1)$ for $1 \leq k < N$. That is, the transitions deterministically trace a Hamiltonian cycle. Recall that a Hamiltonian cycle need not exist in general and the problem of determining whether one does is NP-hard (see, e.g., Garey and Johnson [22]). We then have:

Theorem 1. *The quantity* $\mathrm{tr}(I-P+\Pi)^{-1}$ *or either of the quantities (4), (5) is minimized by a Hamiltonian P, if there exists one, over all stochastic matrices compatible with the graph G.*

In case of irreducible matrices, the minimizations of $\mathrm{tr}(I - P + \Pi)^{-1}$, (4) and (5) are equivalent by the results of [24] and the fact was proved in [24]. For the general case, the fact was proved for the trace by Litvak and Ejov [28], Proposition 2.1, whereas the corresponding extension for (4), (5) is trivial because these quantities are infinite in absence of irreducibility.

In the next subsection, we give a variant of the cited result of Litvak and Ejov [28] inspired by the Held-Karp algorithm for scheduling problems [23]. Specifically, we replace $\pi(y)$ by 1 for all y in the above and show the optimality of deterministically traversing the Hamiltonian cycle over *all* G-compatible chains/processes, Markov or otherwise. This has interesting implications for random search.

2.3. A dynamic programming solution

As in Held and Karp [23], a natural state space for the dynamic program is

$$V^* := \{(x, A) : x \in V,\ A \subset V\},$$

where the first coordinate x will play the role of the current vertex and the second coordinate A the set of vertices not yet visited. With each $(x, A) \in V^*$, we associate an action space $U_x :=$ the set of probability vectors on the set

$$V_x := \{y \in V : (x, y) \in E\} \subset V$$

of successors of x in G. Note that this does not depend on A. Suppose V_x is enumerated as $(y_1, \cdots, y_{m_x})$. Given a 'control' $q = (q(y_1), \cdots, q(y_{m_x}))$, the transition probability

$$\hat{p}((y, B)|(x, A), q) \tag{8}$$

of going from $(x, A) \in V^*$ to $(y, B) \in V^*$ under control q is zero if either $y \notin V_x$ or $B \neq A\backslash\{y\}$. Otherwise it equals $q(y)$. Consider a V^*-valued controlled Markov chain $(X_n, A_n)_{n\in\mathbb{Z}_+}$ governed by a control process $(q_n)_{n\in\mathbb{Z}_+}$ with $q_n \in U_{X_n}$ for all $n \in \mathbb{Z}_+$, evolving according to the above controlled transition probability function. That is, for any $n \in \mathbb{Z}_+$,

$$\begin{aligned} &\mathbb{P}((X_{n+1}, A_{n+1}) = (y, B)|(X_m, A_m, q_m), m \leq n) \\ &= \mathbb{P}((X_{n+1}, A_{n+1}) = (y, B)|(X_n, A_n), q_n) \\ &= q_n(y)\delta_{B, A_n\backslash\{y\}}, \end{aligned}$$

where $\delta_{\cdot,\cdot}$ denotes the Kronecker delta. Since we are allowed to choose any past dependent transition probability compatible with G, this covers *all* V-valued random processes that are compatible with G, i.e., that make transitions only along the edges in E. To see this, consider *any* V-valued process $(X_n)_{n\in\mathbb{Z}_+}$ compatible with the graph, i.e., $X_{n+1} \in V_{X_n},\ \forall n \in \mathbb{Z}_+$ a.s., and define the corresponding $(A_n)_{n\in\mathbb{Z}_+}$ recursively starting from some A_0 by: $A_{n+1} = A_n\backslash\{X_n\},\ \forall n \in \mathbb{Z}_+$. Then this fits the above framework when q_n is the regular conditional law of X_{n+1} given $(X_m, A_m)_{m\leq n}$. If $(q_n)_{n\in\mathbb{Z}_+}$ is a deterministic sequence, then both $(X_n)_{n\in\mathbb{Z}_+}$ and $(X_n, A_n)_{n\in\mathbb{Z}_+}$ will be Markov.

Our objective is to minimize for a prescribed initial state x_0[1] the quantity

$$\mathbb{E}\left[\sum_y T_y \;\middle|\; X_0 = x_0, A_0 = V\right], \tag{9}$$

which is proportional to (4) when $(X_n)_{n\in\mathbb{Z}_+}$ is an irreducible Markov chain with stationary distribution $\pi = \upsilon$, the uniform distribution. Note, however, that we do not require $(X_n)_{n\in\mathbb{Z}_+}$ to be even Markov. Let

$$\zeta := \min\{n \geq 0 : A_n = \emptyset\}. \tag{10}$$

Then (9) equals

$$\begin{aligned}
\mathbb{E}\left[\sum_y T_y \;\middle|\; X_0 = x_0, A_0 = V\right] &= \mathbb{E}\left[\sum_y \sum_{m=0}^{\infty} \mathbb{1}_{\{m<T_y\}} \;\middle|\; X_0 = x_0, A_0 = V\right] \\
&= \mathbb{E}\left[\sum_{m=0}^{\infty} \sum_y \mathbb{1}_{\{m<T_y\}} \;\middle|\; X_0 = x_0, A_0 = V\right] \\
&= \mathbb{E}\left[\sum_{m=0}^{\zeta} |A_m| \;\middle|\; X_0 = x_0, A_0 = V\right].
\end{aligned} \tag{11}$$

Here A_n is nothing but 'the states not yet reached' where we do not count the initial state x_0. This allows us to apply the dynamic programming principle to the 'value function' or 'cost to go function'

$$V(x, A) := \inf \mathbb{E}\left[\sum_{m=0}^{\zeta} |A_m| \;\middle|\; X_0 = x, A_0 = A\right],$$

where the infimum is over all admissible controls. Standard arguments yield the dynamic programming equation

$$V(x, A) = \min_{q\in U_x}\left(|A| + \sum_{y\in V_x} q(y)V(y, A\backslash\{y\})\right),\ A \neq \emptyset, \tag{12}$$

$$V(\cdot, \emptyset) \equiv 0. \tag{13}$$

This is a special case of the well known 'stochastic shortest path' problem [4]. The optimal control in state (x, A) is any minimizer of the right hand side of (12). Since the expression being minimized is affine in q, this minimum will be attained at a Dirac measure, implying that the optimal choice in state (x, A) has a solution which corresponds to a deterministic move to a certain $y \in V_x$. In other words, there is an optimal trajectory which is deterministic and necessarily visits each node at least once, otherwise the cost would be infinite. Since at most one new node can be visited each time, the total cost is at least $\sum_{i=1}^{N-1} i = \frac{N(N-1)}{2}$, which equals the cost for tracing a Hamiltonian cycle if one exists. Note also that if a state is not reachable from x_0 under some policy, the corresponding cost is ∞.

[1] More generally, for a prescribed initial distribution.

We have proved:

Theorem 2. *The minimum of (9) over all V-valued random processes compatible with G is attained by tracing a Hamiltonian cycle when one exists.*

Since i_0 was arbitrary, the claim holds for all initial distributions. This has interesting implications to some random search schemes. For example, consider the problem of searching for an l bit binary password given a device or 'oracle' that can verify whether a password is correct or not. Random search schemes for this problem have been proposed, involving Markov chains on the discrete l-cube $\{0,1\}^l$, where any two strings differing in one position are deemed to be neighbors. This undirected graph can be rendered directed by replacing each undirected edge by two directed edges. A simple induction argument shows that it has a Hamiltonian cycle. Then the foregoing leads to the conclusion that no random search scheme can do strictly better on average than simply listing the l-strings and checking them one by one.

Remark 3. The Markov chains tracing a Hamiltonian cycle may not be the unique minimizer of (9), since in the proof of Theorem 2, when we minimized an affine function over a convex domain, the minimizer may also be attained at points of the boundary of the domain which are not extremal. Indeed, here is an example. Consider G the complete graph over $V = \mathbb{Z}/(3\mathbb{Z})$. The stochastic chain described below is also a minimizer of (9):

- When the initial point is $X_0 = 0$.
  - From 0 choose as next position 1 or 2 with equal probability 1/2.
  - From 1, knowing the previous position was 0, go to 2.
  - From 1, knowing the previous position was not 0, go to 0.
  - From 2, knowing the previous position was 0, go to 1.
  - From 2, knowing the previous position was not 0, go to 0.
- When the initial point is $X_0 = 1$ or $X_0 = 2$, from a current point $x \in \mathbb{Z}/(3\mathbb{Z})$, go to $x + 1$.

Of course the above process is not Markovian. Theorem 2.1 of Litvak and Ejov [28] states that the unique minimizers among Markov chains are given by Hamiltonian cycles, when they exist. Let us present another proof which will be adaptable to the continuous time setting. Consider a Markov chain X minimizing (9) and let P be its transition matrix. Make the hypothesis that G is Hamiltonian. From the above arguments, we get that starting from any $x \in V$, a.s. X must visit all the points of $X \setminus \{x\}$ exactly once before returning to x. Assume on the contrary that P does not induce a Hamiltonian cycle. Then there exist three distinct points $x, x_1, x_2 \in V$ such that $P(x, x_1) > 0$ and $P(x, x_2) > 0$. Starting with $X_0 = x$, if $X_1 = x_1$, afterwards X must visit x_2 before returning to x, namely, there is a path going from x_1 to x_2 without visiting x with transitions permitted by P. Similarly, there is such a path going from x_2 to x_1. Thus x_1 and x_2 are included in a cycle of transitions permitted by P which does not touch x. It follows that with positive probability, X will visit x_1 several times before returning to x, a contradiction.

Remark 4. The time ζ defined in (10) is the covering time, namely the first time all the vertices from V have been visited. Replace (11) by

$$\mathbb{E}\left[\zeta \;\middle|\; X_0 = x_0, A_0 = V\right] = \mathbb{E}\left[\sum_{m=0}^{\zeta} 1 \;\middle|\; X_0 = x_0, A_0 = V\right],$$

so that the above arguments show that for any fixed $x_0 \in V$, the minimum of $\mathbb{E}_{x_0}[\zeta]$ over all G-compatible random chains is attained by tracing a Hamiltonian cycle when one exists. Note that the latter Markov chain does not satisfy the assumptions of Aldous [1], especially the reversibility, that is why his bounds cannot be applied here.

2.4. Continuous time problem

We now consider the continuous time counterparts of the foregoing. Recall that a **Markov generator** on V can be represented by a matrix $L := (L(x,y))_{x,y\in V}$ whose off-diagonal entries are non-negative and whose row sums all vanish. For any $x \in V$, we will denote $L^*(x) := -L(x,x) = \sum_{y\neq x} L(x,y)$. The corresponding Markov processes, defined through the corresponding martingale problems, will be denoted $X := (X_t)_{t\geq 0}$. The probabilistic description of the evolution of X is as follows: sample X_0 according to its law $\mathcal{L}(X_0)$, wait an exponential time σ_1 with parameter $L^*(X_0)$ and choose a new position X_{σ_1} according to the probability $(L(X_0,x)/L^*(X_0))_{x\in V\setminus\{X_0\}}$. For any time $t \in [0,\sigma_1)$, we take $X_t := X_0$. Wait an independent exponential time σ_2 with parameter $L^*(X_{\sigma_1})$ and choose a new position $X_{\sigma_1+\sigma_2}$ according to the probability $(L(X_{\sigma_1},x)/L^*(X_{\sigma_1}))_{x\in V\setminus\{X_{\sigma_1}\}}$. For any time $t \in [\sigma_1, \sigma_1+\sigma_2)$, we take $X_t := X_{\sigma_1}$, and so on. Thus the law of X only depends on its **initial distribution**, namely on the law $\mathcal{L}(X_0)$. The Markov generator L is said to be **compatible** with G, if we have

$$\forall\, x \neq y \in V, \qquad L(x,y) > 0 \Rightarrow (x,y) \in E.$$

The probability measure π, viewed as a row vector, is said to be **invariant** for the generator L, if $\pi L = 0$. Its probabilistic interpretation is that if initially $\mathcal{L}(X_0) = \pi$, then for any $t \geq 0$, $\mathcal{L}(X_t) = \pi$, as in the discrete time case. The generator L is said to be **irreducible** if for any $x, y \in V$, there exists a path $x_0 = x, x_1, ..., x_l = y$, with $l \in \mathbb{Z}_+$ the length of the path, such that $L(x_k, x_{k+1}) > 0$ for all $k \in [\![0, l-1]\!]$. In our finite setting, a Markov generator L always admits an invariant probability measure and the irreducibility of L ensures that it is unique. The irreducible Markov generator L is said to be **normalized**, if

$$\sum_{x\in V} L^*(x)\pi(x) = 1, \tag{14}$$

where π is the invariant measure of L. It means that at its equilibrium π (i.e., for the stationary X starting with $\mathcal{L}(X_0) = \pi$), the *mean* jump rate of X is 1. Denote by $\mathcal{L}(G,\pi)$ the convex set of irreducible normalized Markov generators L compatible with G and admitting π for invariant probability. To simplify notation, we will also write $\mathcal{L}(\pi) := \mathcal{L}(K_V,\pi)$ when G is the complete graph K_V on V. For $y \in V$, let τ_y be the **hitting time** of y:

$$\tau_y := \inf\{t \geq 0 \,:\, X_t = y\}.$$

We are particularly interested in the functional

$$F \,:\, \mathcal{L}(G,\pi) \ni L \mapsto \sum_{x,y\in V} \pi(x)\pi(y)\mathbb{E}_x^L[\tau_y], \tag{15}$$

where we recall that the subscript x (respectively, the superscript L) in the expectation indicates that X is starting from x (resp., is generated by L). The probabilistic interpretation of $F(L)$ is the mean time to go from x to y for the Markov process generated by L, when x and y are sampled independently according to its invariant probability π.

The quantity $F(L)$ also admits a nice spectral formulation: for any $L \in \mathcal{L}(G,\pi)$, let $\Lambda(L)$ be the spectrum of $-L$, removing the eigenvalue 0. To take into account the possible multiplicities of the eigenvalues, $\Lambda(L)$ should be seen a **multiset** (i.e., an eigenvalue of $-L$ of algebraic multiplicity m appears m times in $\Lambda(L)$).

By irreducibility of L, $\Lambda(L)$ is a priori a sub(multi)set of $\mathbb{C}_+ := \{z \in \mathbb{C} : \Re(z) > 0\}$ that is invariant under conjugation. The **eigentime relation** asserts that

$$F(L) = \sum_{\lambda\in\Lambda(L)} \frac{1}{\lambda}. \tag{16}$$

The references Cui and Mao [9] and [29] cited in the discrete time setting also deal with the continuous time case. The quantity $F(L)$ can also be written in terms of return times. Define for any $y \in V$,

$$\begin{aligned} \sigma &:= \min\{t > 0 : X_t \neq X_0\}, \\ T_y &:= \min\{t \geq \sigma : X_t = y\}. \end{aligned}$$

By irreducibility of L, we have the following eigentime identities (see Cui and Mao [9]), for any $y \in V$:

$$\sum_{y\in V} \pi(y)\mathbb{E}_x[\tau_y] \;=\; \sum_{\lambda\in\Lambda(L)} \frac{1}{\lambda} \;=\; \sum_{x,z\in V} \pi(x)\pi(z)\mathbb{E}_x[\tau_z], \tag{17}$$

$$\sum_{y} \pi(y)\mathbb{E}_x[T_y] \;=\; 1 + \sum_{\lambda\in\Lambda(L)} \frac{1}{\lambda} \;=\; \sum_{x,z\in V} \pi(x)\pi(z)\mathbb{E}_x[T_z]. \tag{18}$$

Remark 5. The smaller the $F(L)$, the faster the underlying Markov process goes between the elements of V. It does not necessarily imply that the time-marginal distributions approach equilibrium faster (especially in the discrete time analogue, where due to periodicity, this convergence may not even occur). In particular, our notion of 'fast' is distinct from the extensively studied notion of fast mixing. It is related rather to the asymptotic behavior of the variances associated with the convergence of the empirical measures. Without getting into complete details of this point of view here, we consider the special case where π is reversible for L. Choose an orthonormal basis $(\varphi_l)_{l\in[\![0,N-1]\!]}$ of L with respect to the scalar product of $\mathbb{L}^2(\pi)$, with $\varphi_0 = \mathbb{1}$, the vector of all 1's. Any function $f : V \to \mathbb{R}$ can be uniquely written under the form $f = \sum_{l\in[\![0,N-1]\!]} f_l\varphi_l$, with real coefficients $(f_l)_{l\in[\![0,N-1]\!]}$. Consider the set

$$\mathcal{B} := \{f \,:\, \forall\, l \in [\![1, N-1]\!],\ |f_l| \leq 1\}$$

For any function $f : V \to \mathbb{R}$, we have a central limit theorem for the associated empirical means, namely the convergence in law for large $t \geq 0$:

$$\frac{1}{\sqrt{t}}\left(\int_0^t (f(X_s) - \mu[f])\, ds\right) \to \mathcal{N}(0, \sigma^2(f))$$

where the r.h.s. is the centered normal distribution with asymptotic variance $\sigma^2(f)$. Classical considerations of the related Poisson equation (see also Subsection 6.1 below) show that

$$\sigma^2(f) = \sum_{\lambda\in\Lambda(L)} \frac{1}{\lambda} f_l^2$$

It follows from (16) that

$$\max_{f\in\mathcal{B}} \sigma^2(f) = F(L)$$

As in Subsection 2.3, our goal is to find the minimizers of F on $\mathcal{L}(G,\pi)$, when $\pi(x) \equiv 1\ \forall x$, over all G-compatible processes. Consider therefore a continuous time V-valued *controlled* Markov chain, denoted $(X_t, A_t)_{t\geq 0}$ again by abuse of notation, controlled by a control process $(U_t)_{t\geq 0}$. The latter takes values in $\mathcal{U}_{X_t}$, where $\mathcal{U}_x := [0,\infty)^{V_x}$, identified with the instantaneous transition rates of X_t. That is, as δ goes to 0_+,

$$\mathbb{P}(X_{t+\delta} = j | X_s, U_s, s \leq t, X_t = i) = \mathbb{P}(X_{t+\delta} = j | X_t = i, U_t)$$
$$= \begin{cases} U_t(X_t, j)\delta + o(\delta) & , \text{ if } j \in V_{X_t}, \\ -\sum_{l\in V_{X_t}} U_t(X_t, l)\delta + o(\delta) & , \text{ if } j = X_t, \\ 0 & , \text{ if } j \notin V_{X_t} \cup \{X_t\}, \end{cases}$$

where we write $U_t = (U_t(X_t, y_1), \cdots, U_t(X_t, y_{m_{X_t}}))$ for a suitable enumeration $(y_1, \cdots, y_{m_{X_t}})$ of V_{X_t}. Here it was not necessary that V_x contain x, we rather took $V_x := \{y \in V\ :\ L(x,y) > 0\}$.

For the remaining part of this subsection, we consider the case where $\pi = \upsilon$, the uniform measure on V. There is no loss of generality in requiring that L is irreducible, because the functional F is infinite for non-irreducible Markov generators admitting υ as invariant measure. The normalization condition (14) can then be written in the form

$$\sum_{x\neq y} U_t(x,y) \;=\; N. \tag{19}$$

If for any $t \geq 0$, U_t is a function of X_t alone, say $U_t = r(X_t, \cdot) \in \mathcal{U}_{X_t}$, then $(X_t)_{t\geq 0}$ is a time-homogeneous Markov process with rate matrix $R = (r(x,y))_{x,y\in V}$, where we set $r(x,y) = 0$ for $y \notin V_x$. Consider the problem of minimizing (17). As before, we augment the state process to obtain the V^*-valued process $(\hat{X}_t)_{t\geq 0} = (X_t, A_t)_{t\geq 0}$, with the understanding that A_t can change only when X_t does and a transition of X_t from x to y leads to a transition of A_t to $A_t\backslash\{y\}$. Consider the control problem of minimizing the cost

$$\mathbb{E}\left[\int_0^{\zeta} |A_t| dt \;\middle|\; X_0 = x_0, A_0 = V\right], \tag{20}$$

for

$$\zeta := \{t \geq 0 : A_t = \emptyset\},$$

which is equivalent to (17), subject to the normalization constraint (19). The constraint (19) couples decisions across different states, so dynamic programming arguments cannot be directly applied. Therefore we modify the formulation for the time being, this modification will be dropped later. The modification is as follows. Let $(a_x)_{x\in V}$ be scalars in $(0,N)$ such that $\sum_x a_x = N$. For state x, we restrict the rates to be from the set

$$\tilde{\mathcal{U}}_x := \{r(x,y) : r(x,y) = 0\ \forall\ y \notin V_x,\ \sum_{y\in V_x} r(x,y) = a_x\}.$$

Consider the value function

$$V(x,A) := \inf \mathbb{E}\left[\int_0^{\zeta} |A_t| dt \;\middle|\; X_0 = x, A_0 = A\right],$$

where the infimum is over all admissible controls. The dynamic programming equation then is

$$\min_{r(x,\cdot)\in\tilde{U}_x}\left(|A| + \sum_{y\in V_x,A} r(x,y)(V(y,A\backslash\{y\}) - V(x,A))\right) = 0,\ V(\cdot,\emptyset)\equiv 0. \tag{21}$$

Once again it is clear that the quantity being minimized is affine in the variables it is being minimized over and hence the optimum is attained for a deterministic choice of $r(x,\cdot)$ in the sense that $r(x,y)$ is non-zero for at most one $y \in V_x$. Thus there is an optimal path tracing the nodes of G in a deterministic manner, visiting each of them at least once. This is true for any choice of $(a_x)_{x\in V}$ and therefore true in general for the constraint (19). Unlike the discrete time case, this does not, however, mean that the trajectory is deterministic, because the sojourn time in each node is still random.

Since it is clear again that at most one node that was not already seen is visited at each jump time, the cost is bounded below by

$$\min_{\mathcal{A}}\sum_{n=1}^{N}\left(\sum_{m=1}^{n}\frac{1}{a_m}\right) \tag{22}$$

where

$$\mathcal{A} := \left\{[a_1,\cdots,a_N] : a_n \geq 0,\ \forall 1\leq n\leq N,\ \sum_{n=1}^{N} a_n = N\right\}.$$

The expression in the parentheses in (22) is a lower bound on time to hit the n-th node in the sequence if the nodes are arranged in the order in which they were hit. This is legitimate because we can and do restrict ourselves to deterministic transitions as argued above. But (22) is a strictly convex function being minimized on a polytope and since both the function and the polytope are symmetric in the variables, it attains its unique minimum at $[1,\cdots,1]\in\mathcal{A}$. This proves the optimality of a Hamiltonian cycle with equal mean sojourn times in each state, among all possible rate matrices subject to the rate constraint (19).

Thus we have:

Theorem 6. *The minimum of (20) over all V-valued random processes compatible with G and satisfying the normalization condition (19), is attained by tracing a Hamiltonian cycle when one exists, with equal mean sojourn times, regardless of the initial distribution.*

Here the fact that the optimality is independent of the initial condition for X follows from the arbitrary choice of the latter, as is common in application of the dynamic programming principle.

Remark 7. When G is Hamiltonian, the Markov processes tracing Hamiltonian cycles, with jump rates all equal to 1, may not be the unique minimizers of (20), an example is obtained by Poissonization of the random chain described in Remark 3. Nevertheless, they are the unique minimizers among the class of Markov processes, since the arguments at the end of Remark 3 are readily adaptable to the continuous time setting.

3. Small perturbations of the invariant measure

The rest of this article addresses the second problem stated in the introduction, flagged as **(P2)**. We can expect the optimality of Hamiltonian cycle to persist under small perturbations of the Markov chains considered above. We do this first for the continuous time framework (which turns out to be more natural for the kind of techniques we employ, viz., the calculus of Markov generators) and then for the discrete case in Section 5. Our results are distinct from previous 'persistence under perturbations' results such as those of [14] in the nature of perturbations used, as explained in the introduction.

When $A := (a_0, a_1, ..., a_{N-1}) \in \mathcal{H}$ and a probability measure π on V are fixed, the set $\mathcal{L}(G_A, \pi)$ is reduced to a singleton, its element will be denoted L_A. It is indeed given by

$$\forall\ x, y \in V, \qquad L_A(x, y) = \begin{cases} \frac{1}{N\pi(x)} & , \text{if } x = a_k \text{ and } y = a_{k+1} \text{ for some } k \in \mathbb{Z}_n \\ -\frac{1}{N\pi(x)} & , \text{if } x = y \\ 0 & , \text{otherwise} \end{cases}$$

More generally, we have the following. A cycle A in V is a finite sequence $(a_0, a_1, ..., a_{n-1})$ of distinct elements of V, with $n \in \mathbb{N} \setminus \{1\}$, up to the identification with $(a_k, a_{k+1}, ..., a_{k+n-1})$, for all $k \in \mathbb{Z}_n$. Indeed, as with Hamiltonian cycles (corresponding to $n = N$), the indices should be seen as elements of $\mathbb{Z}_n$. The set of all cycles is denoted by $\mathcal{A}$. Let $\bar{\mathcal{L}}(\pi)$ be the convex set of normalized Markov generators L admitting π for invariant probability. The difference with $\mathcal{L}(\pi)$ is that the elements of $\bar{\mathcal{L}}(\pi)$ are not required to be irreducible. For any $A \in \mathcal{A}$, there is a unique element $L \in \bar{\mathcal{L}}(\pi)$ such that

$$\forall\ x, y \in V, \qquad L(x, y) > 0 \Leftrightarrow \exists\ l \in \mathbb{Z}_n\ :\ x = a_l \text{ and } y = a_{l+1}.$$

It is indeed the generator, denoted L_A in the sequel, given by

$$\forall\ x, y \in V, \qquad L_A(x, y) = \begin{cases} \frac{1}{n\pi(x)} & , \text{if } x = a_l \text{ and } y = a_{l+1} \text{ for some } l \in \mathbb{Z}_n, \\ -\frac{1}{n\pi(x)} & , \text{if } x = y, \\ 0 & , \text{otherwise.} \end{cases} \tag{23}$$

The interest of this kind of generators will appear below in Sub-appendix 6.2, where it will be convenient to differentiate in the direction of cycles whose support is not the whole set V.

Our main result in this section is:

Theorem 8. *Assume that the graph G is Hamiltonian. Then there exists a neighborhood $\mathcal{N}$ of υ in the set $\mathcal{P}_+(V)$ of probability measures on V (endowed with the topology inherited from that of $[0, 1]^V$) such that for any $\pi \in \mathcal{N}$, the set of minimizers of F on $\mathcal{L}(G, \pi)$ is exactly $\{L_A\ :\ A \in \mathcal{H}(G)\}$.*

The complete proof of Theorem 8 is quite technical, with several intermediate lemmas. But the underlying heuristic idea is simple: when the uniform probability measure υ is slightly perturbed into an invariant probability measure π, the energy landscape of F is not significantly modified. In particular a compact set of generators not containing a global minimum of F on $\mathcal{L}(\upsilon)$ does not get continuously distorted abruptly into a set of generators containing a global minimum of F on $\mathcal{L}(\pi)$: there is no spontaneous emergence of global minimizers. This continuity property (partially included in Lemma 12) implies that it is sufficient to show that the Hamiltonian generators remain local minimizers of F on $\mathcal{L}(\pi)$ for π close enough to υ (see Lemma 11). This approach needs a convenient notion of differentiation on $\mathcal{L}(\pi)$. The corresponding calculus is introduced and developed in the Appendix, where Poisson equations play an important role. The remaining part of this section is devoted to the proof of Theorem 8 and its preliminaries.

First we check that all Hamiltonian cycles have the same speed in $\mathcal{L}(\pi)$. This was stated in the introduction in the discrete time setting and for the uniform distribution υ, but it is true more generally.

Lemma 9. *Let $A = (a_0, ..., a_{N-1}) \in \mathcal{H}$ be a Hamiltonian cycle, we have*

$$F(L_A) = \frac{N}{2} \sum_{x \neq y} \pi(x)\pi(y).$$

In particular this quantity does not depend on the choice of the Hamiltonian cycle A.

Proof. The generator L_A can be represented by the matrix

$$\begin{pmatrix} -\frac{1}{N\pi(a_0)} & \frac{1}{N\pi(a_0)} & 0 & \cdots & \cdots & 0 \\ 0 & -\frac{1}{N\pi(a_1)} & \frac{1}{N\pi(a_1)} & 0 & \cdots & 0 \\ & & \cdots & \cdots & & \\ 0 & \cdots & \cdots & 0 & -\frac{1}{N\pi(a_{N-2})} & \frac{1}{N\pi(a_{N-2})} \\ \frac{1}{N\pi(a_{N-1})} & 0 & \cdots & \cdots & 0 & -\frac{1}{N\pi(a_{N-1})} \end{pmatrix}.$$

It follows that the polynomial in X given by

$$P(X) := \det \begin{pmatrix} X-\frac{1}{N\pi(a_0)} & \frac{1}{N\pi(a_0)} & 0 & \cdots & \cdots & 0 \\ 0 & X-\frac{1}{N\pi(a_1)} & \frac{1}{N\pi(a_1)} & 0 & \cdots & 0 \\ & & \cdots & \cdots & & \\ 0 & \cdots & \cdots & 0 & X-\frac{1}{N\pi(a_{N-2})} & \frac{1}{N\pi(a_{N-2})} \\ \frac{1}{N\pi(a_{N-1})} & 0 & \cdots & \cdots & 0 & X-\frac{1}{N\pi(a_{N-1})} \end{pmatrix},$$

is equal to $X\prod_{\lambda\in\Lambda(L_A)}(X-\lambda)$. Expanding the latter expression into $X(\alpha_0+\alpha_1 X+\cdots+\alpha_{N-1}X^{N-1})$, we get that

$$\sum_{\lambda\in\Lambda(L_A)}\frac{1}{\lambda} = -\frac{\alpha_1}{\alpha_0}. \tag{24}$$

Indeed, by the usual relations between the roots and the coefficients of the polynomial $\prod_{\lambda\in\Lambda(L_A)}(X-\lambda) = \alpha_0+\alpha_1 X+\cdots+\alpha_{N-1}X^{N-1}$, we have

$$\alpha_0 = (-1)^{N-1}\prod_{m\in[\![N-1]\!]}\lambda_m, \tag{25}$$

$$\alpha_1 = (-1)^{N-2}\sum_{k\in[\![N-1]\!]}\prod_{m\in[\![N-1]\!]\setminus\{k\}}\lambda_m, \tag{26}$$

where $\Lambda(L_A)$ is parametrized as the multiset consisting of the λ_m, for $m\in[\![N-1]\!] := \{1,2,...,N-1\}$. As a consequence, (24) is deduced by taking the ratio of (26) by (25).

On the other hand, we compute directly from the definition of $P(X)$, by expanding the determinant, that

$$P(X) = \prod_{l\in\mathbb{Z}_N}\left(X-\frac{1}{N\pi(a_l)}\right) - \prod_{l\in\mathbb{Z}_N}\left(-\frac{1}{N\pi(a_l)}\right).$$

It follows that

$$\alpha_0 = \sum_{k\in\mathbb{Z}_N}\prod_{m\in\mathbb{Z}_N\setminus\{k\}}\left(-\frac{1}{N\pi(a_m)}\right),$$

$$\alpha_1 = \frac{1}{2}\sum_{k\neq l\in\mathbb{Z}_N}\prod_{m\in\mathbb{Z}_N\setminus\{k,l\}}\left(-\frac{1}{N\pi(a_m)}\right)$$

(the factor 1/2 is due to the fact that the pair (k,l) also appears as (l,k)). Multiplying the numerator and the denominator by $\prod_{m\in\mathbb{Z}_N}(-N\pi(a_m))$, we get that

$$-\frac{\alpha_1}{\alpha_0}=\frac{\frac{N^2}{2}\sum_{k\neq l\in\mathbb{Z}_N}\pi(a_k)\pi(a_l)}{N\sum_{m\in\mathbb{Z}_N}\pi(a_m)}$$

and this leads to the announced result. □

In particular, for $\pi=\upsilon$, the uniform probability measure on V, we get:

Corollary 10. *For $\pi=\upsilon$, we have for any Hamiltonian cycle A,*

$$F(L_A)=\frac{N-1}{2}.$$

The next result is the crucial step in the proof of Theorem 8. For its statement, introduce for any $A\in\mathcal{H}$ and $\epsilon\in(0,1)$,

$$\mathcal{N}_{A,\epsilon}:=\{L=(1-t)L_A+t\widetilde{L}\ :\ t\in[0,\epsilon)\text{ and }\widetilde{L}\in\bar{\mathcal{L}}(\pi)\}. \tag{27}$$

This set is a neighborhood of L_A in $\mathcal{L}(\pi)$. Any $\widetilde{L}\in\bar{\mathcal{L}}(\pi)$ can be decomposed as

$$\sum_{\widetilde{A}\in\mathcal{A}}p(\widetilde{A})L_{\widetilde{A}}=p(A)L_A+\sum_{\widetilde{A}\in\mathcal{A}\setminus\{A\}}p(\widetilde{A})L_{\widetilde{A}}$$

where p is a probability on $\mathcal{A}$ ([25], see also Lemma 25 below). Thus $\mathcal{N}_{A,\epsilon}$ remains the same set if we had required in (27) that $\widetilde{L}$ belong to the convex hull generated by the $L_{\widetilde{A}}$, $\widetilde{A}\in\mathcal{A}\setminus\{A\}$.

Define

$$\begin{aligned}\pi_\wedge&:=\min_{x\in V}\pi(x),\\ \epsilon_1(N,\pi_\wedge)&:=\pi_\wedge^4\ln\left(1+\frac{1}{N\pi_\wedge^2}\right),\\ \epsilon_2(\pi_\wedge)&:=\frac{1}{56}\pi_\wedge^{12},\\ \epsilon(N,\pi_\wedge)&:=\epsilon_1(N,\pi_\wedge)\wedge\epsilon_2(\pi_\wedge).\end{aligned}$$

Lemma 11. *For $N\geq 2$ and any $A\in\mathcal{H}$, L_A is the unique minimizer of F over $\mathcal{N}_{A,\epsilon(N,\pi_\wedge)}$.*

Proof. Assume that for some given $A\in\mathcal{H}$, L_A is not the unique minimizer of F over $\mathcal{N}_{A,\epsilon(N,\pi_\wedge)}$. Then we can find $t\in(0,\epsilon(N,\pi_\wedge))$ and a probability p on $\mathcal{A}\setminus\{A\}$, such that $F(L_t)\leq F(L_A)$, with

$$\begin{aligned}L_t&:=(1-t)L_A+t\widetilde{L},\\ \widetilde{L}&:=\sum_{\widetilde{A}\in\mathcal{A}\setminus\{A\}}p(\widetilde{A})L_{\widetilde{A}}.\end{aligned}$$

Applying Taylor-Lagrange formula to the function $[0,t]\ni s\mapsto F(L_s)$, we get: there exists $s\in[0,t]$ such that

$$F(L_t)=F(L_A)+tD_{\widetilde{L}}F(L_A)+\frac{t^2}{2}D^2_{\widetilde{L}}F(L_s),$$

where the derivatives $D_{\widetilde{L}}F(L_A)$ and $D^2_{\widetilde{L}}F(L_s)$ are defined in the Appendix, respectively in (39) and just before Lemma 19. Taking into account Propositions 26 and 27 and (56), we obtain

$$F(L_t) \geq F(L_A) + t\frac{N-1}{2N} - t^2\left(\frac{F(L_s)}{\pi_\wedge^2} + \frac{F(L_s)^2}{\pi_\wedge^4} + \frac{F(L_s)^3}{\pi_\wedge^6}\right). \tag{28}$$

To evaluate $F(L_s)$, note that for $s \in (0,t)$, by Proposition 26 and (56) again,

$$\begin{aligned}\partial_s F(L_s) &= D_{\widetilde{L}}F(L_s)\\ &\leq M(L_s) + M(L_s)^2\\ &\leq \frac{F(L_s)}{\pi_\wedge^2} + \frac{F(L_s)^2}{\pi_\wedge^4},\end{aligned}$$

where $M(L_s)$ is defined in (55). Classical computations show that if a $\mathcal{C}^1$ function $f : [0,t] \to (0,+\infty)$ satisfies $\partial_s f(s) \leq af(s) + bf^2(s)$ for all $s \in [0,t]$, where $a, b > 0$, then assuming $f(0)\exp(bt) < f(0) + b/a$, we get

$$\forall\, s \in [0,t], \qquad f(s) \leq \frac{bf(0)\exp(bt)}{b + af(0)(1-\exp(bt))}.$$

In particular, if

$$\exp(bt) < 1 + b/(2af(0)), \tag{29}$$

then

$$\forall\, s \in [0,t], \qquad f(s) \leq 2\left(f(0) + \frac{b}{2a}\right).$$

Let us apply this observation to the mapping $[0,t] \ni s \mapsto F(L_s)$ and $a := 1/\pi_\wedge^2$, $b = 1/\pi_\wedge^4$. Since

$$\begin{aligned}F(L_0) &= F(L_A)\\ &= \frac{N}{2}\sum_{x\neq y\in V}\pi(x)\pi(y)\\ &\leq \frac{N}{2},\end{aligned}$$

we get that condition (29) is satisfied, in view of the definition of $\epsilon_1(N,\pi_\wedge)$ and due to the fact that $t \in (0, \epsilon_1(N,\pi_\wedge))$. It follows that,

$$\begin{aligned}\forall\, s \in [0,t], \qquad F(L_s) &\leq N + \frac{1}{\pi_\wedge^2}\\ &\leq \frac{2}{\pi_\wedge^2},\end{aligned}$$

since $\pi_\wedge \leq 1/N$. Substituting this bound in (28), we deduce that

$$\begin{aligned}F(L_t) &\geq F(L_A) + t\frac{N-1}{2N} - t^2\left(\frac{2}{\pi_\wedge^4} + \frac{4}{\pi_\wedge^8} + \frac{8}{\pi_\wedge^{12}}\right)\\ &\geq F(L_A) + t\frac{1}{4} - \frac{14}{\pi_\wedge^{12}}t^2.\end{aligned}$$

The r.h.s. is strictly larger than $F(L_A)$ if $t < \epsilon_2(\pi_\wedge)$ and this is in contradiction with our initial assumption. □

Denote for any $\pi \in \mathcal{P}_+(V)$,

$$F_\wedge(\pi) := \inf\{F(L) : L \in \mathcal{L}(\pi)\}. \tag{30}$$

Another ingredient in the proof of Theorem 8 is:

Lemma 12. *The mapping $\mathcal{P}_+(V) \ni \pi \mapsto F_\wedge(\pi)$ is continuous.*

Proof. Let $\mathcal{L}$ be the set of irreducible and normalized Markov generators (so that $\mathcal{L} = \sqcup_{\pi\in\mathcal{P}_+(V)}\mathcal{L}(\pi)$), endowed with the topology inherited from $\mathbb{R}^{V^2}$. The functional F is defined on $\mathcal{L}$ and (16) is valid on $\mathcal{L}$. As a consequence, F is continuous on $\mathcal{L}$. Indeed, if $(L_n)_{n\in\mathbb{N}}$ is a sequence of elements of $\mathcal{L}$ converging to $L \in \mathcal{L}$, then according to Paragraph 5 of Chapter 2 of Kato [26], we have $\lim_{n\to\infty}\Lambda(L_n) = \Lambda(L)$ and so $\lim_{n\to\infty} F(L_n) = F(L)$. Next consider a sequence $(\pi_n)_{n\in\mathbb{N}}$ of elements from $\mathcal{P}_+(V)$ converging to $\pi \in \mathcal{P}_+(V)$ and such that the sequence $(F_\wedge(\pi_n))_{n\in\mathbb{N}}$ admits a limit. For all $n \in \mathbb{N}$, let L_n be an element from $\mathcal{L}(\pi_n)$ such that

$$F_\wedge(\pi_n) \ \le\ F(L_n) \ \le\ F_\wedge(\pi_n) + \frac{1}{n}.$$

Due to the normalization condition and the fact that $\pi \in \mathcal{P}_+(V)$, we can extract a subsequence from $(L_n)_{n\in\mathbb{N}}$ (still denoted $(L_n)_{n\in\mathbb{N}}$ below) converging to some generator L. It is clear that L is normalized and that π is invariant for L. Let us check that L is irreducible. Fix $x \in V$. For any $n \in \mathbb{N}$, let $X^{(n)} := (X_t^{(n)})_{t\ge0}$ be a Markov process starting from x and whose generator is L_n. It is not difficult to deduce from the corresponding martingale problems, that $X^{(n)}$ converges in law (with respect to the Skorokhod topology) to a Markov process starting from x and whose generator is L. Thus for any $y \in V$ and $T \ge 0$,

$$\lim_{n\to\infty} \mathbb{E}_x[T \wedge \tau_y^{(n)}] = \mathbb{E}_x[T \wedge \tau_y]$$

(with obvious notation). It follows that

$$\begin{aligned}\mathbb{E}_x[T \wedge \tau_y] &\le \liminf_{n\to\infty} \mathbb{E}_x[\tau_y^{(n)}]\\ &\le \liminf_{n\to\infty} \frac{1}{\pi_n(x)\pi_n(y)} F(L_n)\\ &= \frac{1}{\pi(x)\pi(y)} \liminf_{n\to\infty} F_\wedge(\pi_n)\\ &\le \frac{N}{2\pi(x)\pi(y)}\end{aligned}$$

by Lemma 9. Letting T go to infinity, we get $\mathbb{E}_x[\tau_y] \le N/(2\pi_\wedge^2)$. This bound, valid for all $x, y \in V$, implies that L is irreducible and thus $L \in \mathcal{L}(\pi)$. Furthermore, the above arguments show that

$$\begin{aligned}\lim_{n\to\infty} F_\wedge(\pi_n) &= \lim_{n\to\infty} F(L_n)\\ &= F(L)\\ &\ge F_\wedge(\pi).\end{aligned}$$

So $F_\wedge$ is lower continuous on $\mathcal{P}_+(V)$. By considering the sequence $(\pi_n)_{n\in\mathbb{N}}$ identically equal to π, we also get that the infimum defining $F_\wedge(\pi)$ is attained.

To show that $F_\wedge$ is upper continuous on $\mathcal{P}_+(V)$, let again $(\pi_n)_{n\in\mathbb{N}}$ be a sequence of elements from $\mathcal{P}_+(V)$ converging to some $\pi \in \mathcal{P}_+(V)$ and such that the sequence $(F_\wedge(\pi_n))_{n\in\mathbb{N}}$ admits a limit. According to the previous remark, there exists $L \in \mathcal{L}(\pi)$ such that $F(L) = F_\wedge(\pi)$. For any $n \in \mathbb{N}$, consider the matrix $\widetilde{L}_n$ given by

$$\forall\ x, y \in V, \qquad \widetilde{L}_n(x, y) := \frac{\pi(x)}{\pi_n(x)} L(x, y).$$

It is easy to prove that $\widetilde{L}_n$ is an irreducible Markov generator leaving π_n invariant. But it may not be normalized, so let $\kappa_n > 0$ be such that $L_n := \kappa_n \widetilde{L}_n$ belongs to $\mathcal{L}(\pi_n)$. There is no difficulty in checking that L_n converges to L and thus that $\lim_{n\to\infty} F(L_n) = F(L) = F_\wedge(\pi)$. Thus passing to the limit in $F(L_n) \geq F_\wedge(\pi_n)$, we deduce that

$$F_\wedge(\pi) \geq \lim_{n\to\infty} F_\wedge(\pi_n)$$

as desired. □

With all these ingredients, we can now come to the

Proof of Theorem 8. Note that it is sufficient to consider the case where G is the complete graph over V, because then $F_\wedge(\pi) \leq \min\{F(L) : L \in \mathcal{L}(G, \pi)\}$ for any graph G and probability measure π on V.

The main argument is by contradiction. Assuming that the statement of Theorem 8 is not true, we can find a sequence $(\pi_n)_{n\in\mathbb{N}}$ from $\mathcal{P}_+(V)$ converging to υ, such that for all $n \in \mathbb{N}$, there exists $L_n \in \mathcal{L}(\pi_n) \setminus \{L_{\pi_n,A} : A \in \mathcal{H}\}$ with $F(L_n) = F_\wedge(\pi_n)$. (Here we have included π_n in the index of $L_{\pi_n,A}$ to underscore the fact that this generator, associated to a Hamiltonian cycle A, also depends on the underlying invariant probability π_n.) As seen in the proof of Lemma 12, a subsequence (still denoted $(L_n)_{n\in\mathbb{N}}$) converging toward some $L \in \mathcal{L}(\upsilon)$ can be extracted from $(L_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} F(L_n) = F(L)$$

and by Lemma 12,

$$\lim_{n\to\infty} F_\wedge(\pi_n) = F_\wedge(\upsilon).$$

It follows that $F(L) = F_\wedge(\upsilon)$. From Remark 7, we deduce that there exists $A \in \mathcal{H}$ such that $L = L_{\upsilon,A}$. Using again the fact that

$$\lim_{n\to\infty} \pi_n = \upsilon, \tag{31}$$

we get that $\lim_{n\to\infty} L_{\pi_n,A} = L_{\upsilon,A}$ and thus

$$\lim_{n\to\infty} (L_n - L_{\pi_n,A}) = 0. \tag{32}$$

Consider $r := \min_{n\in\mathbb{N}} \pi_{n,\wedge}$, which is positive due to (31), and let $\epsilon := \epsilon(N, r)$, with the notation introduced before Lemma 11. From (32), we deduce that for $n \in \mathbb{N}$ large enough, L_n belongs to $\mathcal{N}(\pi_n, A, \epsilon)$, defined as in (27), with π replaced by π_n. Then Lemma 11 asserts that $L_n = L_{\pi_n,A}$, because L_n is a minimizer of L over $\mathcal{L}(\pi_n)$. This is in contradiction with our initial assumption. □

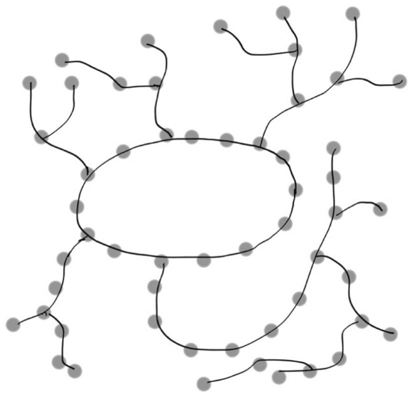

Fig. 1. The graph $\breve{G}$.

4. Large perturbations of the invariant measure

Here we consider large perturbations of the uniform probability measure υ. We show that Theorem 8 cannot be extended to all probability measures π, at least for the graphs which are not a Hamiltonian cycle, a situation where $\mathcal{L}(G,\pi)$ is not reduced to a singleton. (In particular, this requires $N \geq 3$.) Our main result here is:

Theorem 13. *Assume that G is not a Hamiltonian cycle. Then there exist probability measures π on V such that none of the elements of $\{L_A : A \in \mathcal{H}(G)\}$ is a minimizer of F on $\mathcal{L}(G,\pi)$.*

Thus for some (G,π), the minimizers of F on $\mathcal{L}(G,\pi)$ are (spatially) **hesitating** Markov processes: at some vertex, the next visited point is not chosen deterministically. For a given Hamiltonian graph G which is not a Hamiltonian cycle, it would be interesting to describe the probability measures π leading to a transition between non-hesitating and hesitating minimizers. This issue remains open at present.

Proof of Theorem 13. The heuristic idea behind the following arguments is very simple: when G is not a Hamiltonian cycle, it is possible to find an oriented sub-graph $\breve{G}$ looking as in Fig. 1. By choosing probability measures π close to the uniform measure on the sub-cycle of $\breve{G}$ and the associated fastest Markov processes, we get values of F which are strictly smaller than those obtained on the Hamiltonian generators on the whole graph G, because they almost correspond to the minimization problem over the sub-cycle endowed with the uniform measure.

More precisely, let $G = (V,E)$ be a finite oriented connected graph which is not a Hamiltonian cycle. Then we can find a cycle $A := (a_0, a_1, ..., a_{n-1}) \in \mathcal{A}(G)$ with $n < \mathrm{card}(V)$. Denote $\widetilde{V} := \{a_0, a_1, ..., a_{n-1}\}$ and $\widehat{V} := V \setminus \widetilde{V}$. By the strong connectivity of G, we can find a subset $\widehat{E}$ of oriented edges from E, such that $\mathrm{card}(\widehat{E}) = \mathrm{card}(\widehat{V})$ and for any $x \in \widehat{V}$ we can find exactly one $y \in V$ with $(x,y) \in \widehat{E}$. Putting together the edges from A and those from $\widehat{E}$, we get a graph $\breve{G}$ on V looking like the following picture, where the cycle is oriented clockwise and the trees are oriented toward the cycle. For $r > 0$, consider the Markov generator L_r defined by

$$\forall\, x \neq y \in V, \qquad L_r(x,y) := \begin{cases} 1 & \text{, if there exists } l \in [\![0, n-1]\!] \text{ such that } x = a_l \text{ and } y = a_{l+1} \\ r & \text{, if } (x,y) \in \widehat{E} \\ 0 & \text{, otherwise.} \end{cases}$$

This generator is not irreducible, since it does not allow the chain to go from the cycle A to $\widehat{V}$. Nevertheless, its unique invariant probability measure π is the uniform probability measure on $\widetilde{V}$. The generator L_r then satisfies an **extended normalization condition**, in the sense that

$$\sum_{x\neq y\in V}\pi(x)L_r(x,y)=1.$$

The eigenvalues of L_r are easy to find:

$$\Lambda(L_r)=\Lambda(\widetilde{L}_A)\sqcup\{r[|\widehat{V}|]\},$$

where $\widetilde{L}_A$ is the generator corresponding to the Hamiltonian cycle given by A on $\widetilde{V}$ and $\{r[|\widehat{V}|]\}$ is the multiset consisting of the value r with the multiplicity $|\widehat{V}|$. This identity is an immediate consequence of following decomposition of L_r, where all the elements of $\widetilde{V}$ have been put before those of $\widehat{V}$ and where the elements of $\widehat{V}$ have been ordered so that the (oriented) distance to $\widetilde{V}$ is non-decreasing (in particular the last element corresponds to a leaf of $\breve{G}$):

$$L_r=\begin{pmatrix}\widetilde{L}_A & 0\\ C & D\end{pmatrix}.$$

In the r.h.s., the $\widehat{V}\times\widehat{V}$ matrix D is sub-diagonal and its diagonal consists only of $-r$. Formula (16) enables us to extend the functional F to L_r and we get

$$F(L_r)=F(\widetilde{L}_A)+\frac{N}{r}.$$

In particular, it follows that

$$\lim_{r\to+\infty}F(L_r)\ =\ F(\widetilde{L}_A)\ =\ \frac{n-1}{2}\ <\ F(L_H)$$

for any Hamiltonian cycle $H\in\mathcal{H}(G)$, where we used twice Corollary 10. From now on, we fix $r>0$ large enough, so that

$$F(L_r)<F(L_H) \tag{33}$$

for any Hamiltonian cycle $H\in\mathcal{H}(G)$.

For any $\epsilon>0$, consider the Markov generator

$$L_{r,\epsilon}:=Z_{r,\epsilon}^{-1}(L_r+\epsilon L_G)$$

where

- the Markov generator L_G is defined by

$$\forall\ x\neq y\in V,\qquad L_G(x,y):=\begin{cases}1 & \text{, if } (x,y)\in E\\ 0 & \text{, otherwise}\end{cases}$$

- the constant $Z_{r,\epsilon}>0$ is such that $L_{r,\epsilon}$ is normalized (observe that $L_r+\epsilon L_G$ is irreducible on V).

For $r,\epsilon>0$, denote $\pi_{r,\epsilon}$ the invariant probability measure of $L_{r,\epsilon}$. It is clear that as ϵ goes to 0_+, $\pi_{r,\epsilon}$ converges toward π. It follows that

$$\lim_{\epsilon \to 0_+} Z_{r,\epsilon} = 1,$$
$$\lim_{\epsilon \to 0_+} L_{r,\epsilon} = L_r.$$

From the general theory of perturbation of spectra of finite operators (see, e.g., the beginning of the second chapter of the book of Kato [26]), we have

$$\lim_{\epsilon \to 0_+} F(L_{r,\epsilon}) = F(L_r).$$

Taking into account (33), we can thus find $\epsilon > 0$ small enough so that

$$F(L_{r,\epsilon}) < F(L_H)$$

for any Hamiltonian cycle $H \in \mathcal{H}(G)$. That is, the probability measure $\pi_{r,\epsilon}$ satisfies the statement of Theorem 13. Observe that this probability measure $\pi_{r,\epsilon}$ is quite far away from υ, because it gives very small weight to the elements of $\widehat{V}$. □

Note that this result does not contradict Theorem 6, which considered the minimization of $\sum_y E_x[\tau_y]$ and not $\sum_y \pi(y) E_x[\tau_y]$.

5. The discrete time framework

Here we discuss the links between the search of the fastest continuous-time Markov processes with the analogous problem in discrete time.

Let a graph $G = (V, E)$ and a probability measure π on V be fixed and denote by $\mathcal{K}(G, \pi)$ the set of irreducible Markov kernels K on V whose permitted transitions are edges from E plus self-loops (i.e., the possibility to stay at the same node which allows for a nontrivial Bernoulli distributed random sojourn time) and having π as the unique invariant distribution (i.e., satisfying $\pi K = \pi$). For any $K \in \mathcal{K}(G, \pi)$, let $\mathfrak{X} := (\mathfrak{X}_n)_{n \in \mathbb{Z}_+}$ be a Markov chain whose transitions are dictated by K. For any $y \in V$, recall (see (3)) the first hitting time of y defined as

$$\mathfrak{r}_y := \inf\{n \in \mathbb{Z}_+ \, : \, \mathfrak{X}_n = y\}.$$

On $\mathcal{K}(G, \pi)$, we consider the functional $\mathfrak{F}_\pi$ defined by

$$\forall \; K \in \mathcal{K}(G, \pi), \qquad \mathfrak{F}_\pi(K) := \sum_{x,y \in V} \pi(x)\pi(y)\mathbb{E}_x[\mathfrak{r}_y],$$

where subscript x in the expectation indicates that $\mathfrak{X}$ is starting from $x \in V$. As before, this serves as the objective function we seek to minimize.

To any $K \in \mathcal{K}(G, \pi)$, we associate $\Theta(K)$ the multi-set consisting of the spectrum of K, removing the eigenvalue 1 which has multiplicity 1. In other words, this is the set of all eigenvalues except 1, repeated as many times as their respective multiplicities. It is a priori a sub(multi-)set of the closed unit disk centered at 0 of $\mathbb{C}$ left invariant by conjugation, since complex eigenvalues will occur in complex conjugate pairs. In analogy with the continuous-time situation, we have the eigentime relation already introduced in subsection 2.2 (see (4)):

$$\forall \; K \in \mathcal{K}(G, \pi), \qquad \mathfrak{F}_\pi(K) = \sum_{\theta \in \Theta(K)} \frac{1}{1 - \theta}. \tag{34}$$

Recall the notation $L^*(x) := -L(x,x)$ for an infinitesimal generator $L(\cdot,\cdot)$ of a continuous time Markov chain. To any such $L \in \mathcal{L}(G,\pi)$, associate

$$\begin{aligned} l &:= \max\{L^*(x) \,:\, x \in V\}, \\ K &:= I + \frac{L}{l} \text{ i.e., } K(x,y) := \delta_{xy} + \frac{L(x,y)}{l}. \end{aligned} \tag{35}$$

The number l is chosen so that all the diagonal entries of K are non-negative, since the only nonpositive elements of L are the diagonal elements $-L^*(x)$. Since the row sums of L are zero, the row sums of K are equal to those of I, so are all equal to 1. Furthermore, since $\pi L =$ the zero vector, we have $\pi K = \pi I = \pi$, thus $K \in \mathcal{K}(G,\pi)$. Furthermore, from (35), we have $\Theta(K) = 1 - \Lambda(L)/l$, so that by (34),

$$\mathfrak{F}_\pi(K) = lF(L). \tag{36}$$

Taking into account that

$$l \;\geq\; \sum_{x\in V} \pi(x)L^*(x) \;=\; 1,$$

it follows that $\mathfrak{F}_\pi(K) \geq F(L)$. We will denote $\Phi \,:\, \mathcal{L}(G,\pi) \to \mathcal{K}(G,\pi)$ the mapping $L \mapsto K$ defined above.

Conversely, for any $K \in \mathcal{K}(G,\pi)$,

$$\begin{aligned} \sum_{x\in V} \pi(x)(1 - K(x,x)) &= \sum_{x\in V} \pi(x)\left(1 - \left(1 + \frac{L(x,x)}{l}\right)\right) \\ &= \sum_{x\in V} \pi(x)\left(\frac{-L(x,x)}{l}\right) \in (0,1]. \end{aligned}$$

Hence to any $K \in \mathcal{K}(G,\pi)$, associate

$$\begin{aligned} k &:= \frac{1}{\sum_{x\in V} \pi(x)(1 - K(x,x))}, \\ L &= k(K - I). \end{aligned}$$

It is immediate to check that $L \in \mathcal{L}(G,\pi)$. Furthermore, we get $\Lambda(L) = k(1 - \Theta(K))$ and it follows that

$$F(L) = \mathfrak{F}_\pi(K)/k.$$

Taking into account that

$$k \;\geq\; \frac{1}{\sum_{x\in V} \pi(x)} \;= 1,$$

we get that $F(L) \leq \mathfrak{F}_\pi(K)$. Denote $\Psi \,:\, \mathcal{K}(G,\pi) \to \mathcal{L}(G,\pi)$ the mapping $K \mapsto L$ as above.

Remark 14. The mappings Φ and Ψ are not inverse of each other, because the image of $\mathcal{L}(G,\pi)$ by Φ is included in $\mathcal{K}_0(G,\pi) := \{K \in \mathcal{K}(G,\pi) \,:\, \exists\, x \in V \text{ with } K(x,x) = 0\}$. Nevertheless, we have that Φ and Ψ_0 are inverse of each other, where Ψ_0 is the restriction of Ψ to $\mathcal{K}_0(G,\pi)$.

When one is looking for the minimal value of $\mathfrak{F}_\pi$ on $\mathcal{K}(G,\pi)$, one can restrict attention to $\mathcal{K}_0(G,\pi)$, because

$$\min\{\mathfrak{F}_\pi(K) : K \in \mathcal{K}(G,\pi)\} = \min\{\mathfrak{F}_\pi(K) : K \in \mathcal{K}_0(G,\pi)\}.$$

Indeed, for any $K \in \mathcal{K}(G,\pi)$, there exist a unique $\widetilde{K} \in \mathcal{K}_0(G,\pi)$ and $\alpha \in [0,1)$ such that $K = (1-\alpha)\widetilde{K} + \alpha I$. Then we get $\Theta(K) = (1-\alpha)\Theta(\widetilde{K}) + \alpha$, i.e. $\Theta(K) - 1 = (1-\alpha)(\Theta(\widetilde{K}) - 1)$. This implies that

$$\begin{aligned}\mathfrak{F}_\pi(\widetilde{K}) &= (1-\alpha)\mathfrak{F}_\pi(K)\\ &\leq \mathfrak{F}_\pi(K).\end{aligned}$$

As in (30), denote

$$\begin{aligned}F_\wedge(G,\pi) &:= \inf\{F(L) : L \in \mathcal{L}(G,\pi)\},\\ \mathfrak{F}_\wedge(G,\pi) &:= \inf\{\mathfrak{F}_\pi(K) : K \in \mathcal{K}(G,\pi)\}.\end{aligned}$$

From the above considerations, we deduce:

Proposition 15. *We always have*

$$F_\wedge(G,\pi) \leq \mathfrak{F}_\wedge(G,\pi).$$

If in addition we assume that there is a minimizer $L \in \mathcal{L}(G,\pi)$ of F such that $L^*(x)$ does not depend on $x \in V$ (it is then equal to 1), then we get $F_\wedge(G,\pi) = \mathfrak{F}_\wedge(G,\pi)$.

Proof. Consider $K \in \mathcal{K}(G,\pi)$. We have seen that

$$\begin{aligned}\mathfrak{F}_\pi(K) &\geq F(\Psi(K))\\ &\geq F_\wedge(G,\pi),\end{aligned}$$

so taking the infimum over $K \in \mathcal{K}(G,\pi)$, we get the first bound.

Conversely, if $L \in \mathcal{L}(G,\pi)$ is a minimizer of F whose diagonal is constant, then $l = 1$ in (36), namely $\mathfrak{F}_\pi(\Phi(L)) = F(L) = F_\wedge(G,\pi)$. From the previous inequality, it follows that $\Phi(L)$ is indeed a minimizer of $\mathfrak{F}_\pi$ on $\mathcal{K}(G,\pi)$ and we conclude that $F_\wedge(G,\pi) = \mathfrak{F}_\wedge(G,\pi)$. □

Remark 16. This result implies in particular that the (normalized) fastest continuous-time Markov process is as least as fast as the fastest discrete-time Markov chain. The superior performance (vis-a-vis our objective function) for the continuous time Markov chains over discrete time Markov chain is due to greater flexibility of modulating the continuous-time exit rate from each vertex and therefore the sojourn times, as opposed to what is possible by using self-loops in the discrete-time.

In association with Theorem 6, the above proposition also enables us to recover the result of Litvak and Ejov [28] stating that for any Hamiltonian graph G, the permutation matrices associated to the Hamiltonian cycles of G are the unique minimizers of $\mathfrak{F}_\upsilon$ on $\mathcal{K}(G,\upsilon)$. Proposition 15 does not enable us to extend directly Theorem 13 to the discrete time setting, because the diagonal of the generator associated to a Hamiltonian cycle is constant if and only if the underlying invariant probability measure is uniform. This extension is nevertheless true. To show it, note that the differentiation technique of Section 3 can be adapted to $\mathcal{K}(G,\pi)$.

Appendices

6. Differentiation on $\mathcal{L}(\pi)$

This part of the appendix introduces some elements of differential calculus on $\mathcal{L}(\pi)$, which are helpful in the proof of Theorem 8. Here we will be working mainly with the complete graph K_V. We will first differentiate with respect to general generators. Next we provide more precise formulas for the first and second derivatives with respect to cycles. Finally we will bound these derivatives.

6.1. Differentiation with respect to general generators

We begin by presenting a more analytical expression for the functional F. For $y \in V$, consider the function

$$f_y \,:\, V \ni x \mapsto \frac{\mathbb{1}_{\{y\}}(x)}{\pi(y)} - 1.$$

Note that $\pi[f_y] = 0$, so for any $L \in \mathcal{L}(\pi)$, by irreducibility, there exists a unique function φ_y^L on V satisfying the Poisson equation

$$\begin{cases} L[\varphi_y^L] = f_y, \\ \varphi_y^L(y) = 0. \end{cases} \tag{37}$$

In the first equality, $L[\varphi_y^L]$ is a matrix-vector product, where functions are seen as column vectors (measures will be interpreted as row vectors), i.e.,

$$\forall\, x \in V, \qquad L[\varphi_y^L](x) := \sum_{z \in V} L(x,z)\varphi_y^L(z)$$

The uniqueness of the solution to (37) comes from the fact that the difference χ between two solutions satisfies $L[\chi] = 0$ and $\chi(y) = 0$. By irreducibility of L, $L[\chi] = 0$ implies that χ is constant and $\chi(y) = 0$ enables us to conclude that $\chi = 0$. The following relation with the functional F is equally well-known:

Lemma 17. *For any $L \in \mathcal{L}(\pi)$ and any $x, y \in V$, we have*

$$\varphi_y^L(x) = \mathbb{E}_x^L[\tau_y],$$

so that

$$F(L) = \sum_{y \in V} \pi(y)\pi[\varphi_y^L]. \tag{38}$$

To simplify notation, from now on, we will remove the superscript L of $\mathbb{E}_x^L$ and φ_y^L when the underlying generator L is clear from the context.

Proof. Let us recall a simple argument, which will be used again in the sequel. Through the martingale problem characterization of X, we have that for any given function φ on V, the process $(M_t)_{t\geq 0}$ defined by

$$\forall\, t \geq 0, \qquad M_t := \varphi(X_t) - \varphi(X_0) - \int_0^t L[\varphi](X_s)\, ds$$

is a martingale. In particular, for any stopping time τ, the process $(M_{\tau\wedge t})_{t\geq 0}$ is also a martingale. Thus, starting from $x \in V$, we get,

$$\begin{aligned}
\mathbb{E}_x[M_{\tau_y\wedge t}] &= 0 \\
&= \mathbb{E}_x\left[\varphi(X_{\tau\wedge t}) - \varphi(X_0) - \int_0^{\tau\wedge t} L[\varphi](X_s)\,ds\right] \\
&= \mathbb{E}_x\left[\varphi(X_{\tau\wedge t})\right] - \varphi(x) - \mathbb{E}_x\left[\int_0^{\tau\wedge t} L[\varphi](X_s)\,ds\right].
\end{aligned}$$

Since τ is a.s. finite, we obtain by dominated convergence that as t goes to infinity,

$$\mathbb{E}_x\left[\varphi(X_\tau)\right] - \varphi(x) - \mathbb{E}_x\left[\int_0^{\tau} L[\varphi](X_s)\,ds\right] = 0.$$

For any $y \in V$, consider $\varphi := \varphi_y$ and $\tau := \tau_y$. From (37) and from the fact that $f_y(z) = -1$ for any $z \in V \setminus \{y\}$, we deduce

$$\varphi_y(x) = \mathbb{E}_x[\tau_y].$$

The last identity of the lemma comes from

$$\begin{aligned}
\sum_{x\in V} \pi(x)\mathbb{E}_x[\tau_y] &= \sum_{x\in V} \pi(x)\varphi_y(x) \\
&= \pi[\varphi_y]. \quad \square
\end{aligned}$$

Since we are looking for minimizers of F on $\mathcal{L}(\pi)$, it is natural to differentiate this functional. Recall the definition of $\bar{\mathcal{L}}(\pi)$ from Section 3. For $L \in \mathcal{L}(\pi)$, $\widetilde{L} \in \bar{\mathcal{L}}(\pi)$ and $\epsilon \in [0, 1)$, let $L_\epsilon := (1-\epsilon)L + \epsilon\widetilde{L} \in \bar{\mathcal{L}}(\pi)$. Define

$$D_{\widetilde{L}}F(L) := \lim_{\epsilon\to 0_+} \frac{F(L_\epsilon) - F(L)}{\epsilon}, \tag{39}$$

where $F(L)$ is the functional considered in (38). In the proof of the following result, it will be shown that the above limit exists.

Lemma 18. *With the above notation, we have*

$$\begin{aligned}
D_{\widetilde{L}}F(L) &= \sum_{y\in V} \pi(y)(\pi[\varphi_y] - \pi[\psi_y]) \\
&= F(L) - \sum_{y\in V} \pi(y)\pi[\psi_y].
\end{aligned}$$

where ψ_y is the unique solution of another Poisson equation

$$\begin{cases} L[\psi_y] = \widetilde{L}[\varphi_y], \\ \psi_y(y) = 0. \end{cases} \tag{40}$$

Proof. Let $\mathcal{F}_\pi$ stand for the space of functions f on V whose mean with respect to π vanishes. By restriction to $\mathcal{F}_\pi$, $L \in \mathcal{L}(\pi)$ can be seen as an invertible endomorphism of $\mathcal{F}_\pi$. Denote by $L^{-1}_{|\mathcal{F}_\pi}$ its inverse. Similarly, for $\epsilon \in [0,1)$, let $L^{-1}_{\epsilon,\mathcal{F}_\pi}$ be the inverse of L_ϵ on $\mathcal{F}_\pi$. The mapping $[0,1) \ni \epsilon \mapsto L_\epsilon$ being analytic, the same is true for $[0,1) \ni \epsilon \mapsto L^{-1}_{\epsilon,\mathcal{F}_\pi}$. Since we have

$$\forall\ \epsilon \in [0,1),\ \forall\ y \in V, \qquad \varphi_y^{L_\epsilon} = L^{-1}_{\epsilon,\mathcal{F}_\pi}[f_y] - L^{-1}_{\epsilon,\mathcal{F}_\pi}[f_y](y)$$

(where the number $L^{-1}_{\epsilon,\mathcal{F}_\pi}[f_y](y)$ stands for the function $L^{-1}_{\epsilon,\mathcal{F}_\pi}[f_y](\cdot)$ on V evaluated at value y), we deduce that the mapping

$$[0,1) \ni \epsilon \mapsto \varphi_y^{L_\epsilon}$$

is analytic. The same is true for $[0,1) \ni \epsilon \mapsto F(L_\epsilon)$, due to the equality

$$\forall\ \epsilon \in [0,1), \qquad F(L_\epsilon) = \sum_{y\in V} \pi(y)\pi[\varphi_y^{L_\epsilon}].$$

In particular its derivative $D_{\widetilde{L}}F(L)$ exists and is equal to $\sum_{y\in V}\pi(y)\pi[\varphi'_y]$, where φ'_y is the derivative of $\varphi_y^{L_\epsilon}$ at $\epsilon = 0$. Differentiating the relation $L_\epsilon[\varphi_y^{L_\epsilon}] = f_y$, we get

$$(\widetilde{L} - L)[\varphi_y] + L[\varphi'_y] = 0.$$

Furthermore, since $\varphi_y^{L_\epsilon}(y) = 0$ for all $\epsilon \in [0,1)$ by definition, we have that $\varphi'_y(y) = \partial_\epsilon \varphi_y^{L_\epsilon}(y)|_{\epsilon=0} = 0$, so that $\varphi_y - \varphi'_y$ satisfies the equation (40) and must be equal to ψ_y. The claim then follows from the equality $\varphi'_y = \varphi_y - \psi_y$ for all $y \in V$. □

In the above proof, we have seen that $[0,1) \ni \epsilon \mapsto F(L_\epsilon)$ is analytic, so we can differentiate it a second time at $\epsilon = 0$. Denote $D^2_{\widetilde{L}}F(L) = \partial^2_\epsilon F(L_\epsilon)|_{\epsilon=0}$.

Lemma 19. *For $L \in \mathcal{L}(\pi)$, $\widetilde{L} \in \bar{\mathcal{L}}(\pi)$, we have*

$$\begin{aligned} D^2_{\widetilde{L}}F(L) &= \sum_{y\in V} \pi(y)(2\pi[\varphi_y] - 4\pi[\psi_y] + 2\pi[\breve{\psi}_y]) \\ &= 4D_{\widetilde{L}}F(L) - 2F(L) + 2\sum_{y\in V}\pi(y)\pi[\breve{\psi}_y], \end{aligned}$$

where $\breve{\psi}_y$ is the unique solution of

$$\begin{cases} L[\breve{\psi}_y] \ = \ \widetilde{L}[\psi_y], \\ \breve{\psi}_y(y) \ = \ 0. \end{cases} \tag{41}$$

Proof. For any $y \in V$, denote φ''_y the second derivative of $\varphi_y^{L_\epsilon}$ at $\epsilon = 0$. By differentiating twice the relation $L_\epsilon[\varphi_y^{L_\epsilon}] = f_y$ at $\epsilon = 0$, we get

$$(\partial^2_\epsilon L_\epsilon)[\varphi_y] + 2(\partial_\epsilon L_\epsilon)[\varphi'_y] + L[\varphi''_y] = 0.$$

Since $\partial^2_\epsilon L_\epsilon = 0$,

$$
\begin{aligned}
L[\varphi''_y] &= 2(L-\widetilde{L})[\varphi'_y]\\
&= 2(L-\widetilde{L})[\varphi_y-\psi_y]\\
&= 2L[\varphi_y-\psi_y]-2\widetilde{L}[\varphi_y-\psi_y]\\
&= 2L[\varphi_y-\psi_y]-2\widetilde{L}[\varphi_y]+2\widetilde{L}[\psi_y]\\
&= 2L[\varphi_y-\psi_y]-2L[\psi_y]+2\widetilde{L}[\psi_y]\\
&= L[2\varphi_y-4\psi_y]+2\widetilde{L}[\psi_y].
\end{aligned}
$$

It follows that $\varphi''_y/2-\varphi_y+2\psi_y$ satisfies the first condition of equation (41). It also vanishes at y, since $\varphi''_y(y)=0=\varphi_y(y)=\psi_y(y)$. Thus we get that $\varphi''_y=2\varphi_y-4\psi_y+2\breve{\psi}_y$. The announced result is now a consequence of the equality

$$
D^2_{\widetilde{L}}F(L)=\sum_{y\in V}\pi(y)\pi[\varphi''_y],
$$

which comes from (38). □

6.2. First derivatives with respect to cycles

It will be convenient to use these differentiations with respect to particular generators $\widetilde{L}\in\bar{\mathcal{L}}(\pi)$.

Lemma 20. *Let $A=(a_l)_{l\in\mathbb{Z}_n}\in\mathcal{A}$ be given and for $y\in V$, consider the function ψ_y defined by (40) with $\widetilde{L}=L_A$. Then we have*

$$
\forall\ x\in V,\qquad \psi_y(x)=\frac{1}{n}\sum_{l\in\mathbb{Z}_n}(\varphi_y(a_{l+1})-\varphi_y(a_l))(\varphi_{a_l}(x)-\varphi_{a_l}(y)). \tag{42}
$$

Furthermore, we get that

$$
\begin{aligned}
\sum_{y\in V}\pi(y)\pi[\psi_y] &= -\frac{1}{n}\sum_{l\in\mathbb{Z}_n}\sum_{y\in V}\pi(y)(\varphi_y(a_{l+1})-\varphi_y(a_l))\varphi_{a_l}(y)\\
&= \frac{1}{n}\left(\sum_{l\in\mathbb{Z}_n}\frac{1}{2}\mathbb{E}_{a_{l+1}}[\tau^2_{a_l}]-\mathbb{E}_\pi[\tau_{a_l}]\mathbb{E}_{a_{l+1}}[\tau_{a_l}]\right),
\end{aligned}
$$

where $\mathbb{E}_\pi$ stands for the expectation relative to the initial distribution π for X.

Proof. For any function φ on V, we have

$$
L_A[\varphi]=\sum_{l\in\mathbb{Z}_n}\frac{\varphi(a_{l+1})-\varphi(a_l)}{n\pi(a_l)}\mathbb{1}_{\{a_l\}},
$$

where $\mathbb{1}_{\{a_l\}}$ is the indicator function of the point a_l. Let ψ be a function such that $L[\psi]=L_A[\varphi]$. Using the martingale problem as in the proof of Lemma 17, we get for any $x,y\in V$,

$$
\psi(y)-\psi(x)=\mathbb{E}_x\left[\int_0^{\tau_y}L[\psi](X_s)\,ds\right]
$$

$$
\begin{aligned}
&= \mathbb{E}_x\left[\int_0^{\tau_y} L_A[\varphi](X_s)\,ds\right]\\
&= \sum_{l\in\mathbb{Z}_n} \frac{\varphi(a_{l+1})-\varphi(a_l)}{n}\mathbb{E}_x\left[\int_0^{\tau_y} \frac{\mathbb{1}_{\{a_l\}}}{\pi(a_l)}(X_s)\,ds\right]\\
&= \sum_{l\in\mathbb{Z}_n} \frac{\varphi(a_{l+1})-\varphi(a_l)}{n}\mathbb{E}_x\left[\int_0^{\tau_y} 1+f_{a_l}(X_s)\,ds\right]\\
&= \sum_{l\in\mathbb{Z}_n} \frac{\varphi(a_{l+1})-\varphi(a_l)}{n}\left(\varphi_y(x)+\mathbb{E}_x\left[\int_0^{\tau_y} f_{a_l}(X_s)\,ds\right]\right).
\end{aligned}
$$

Taking into account that $L[\varphi_{a_l}] = f_{a_l}$, we deduce that

$$
\mathbb{E}_x\left[\int_0^{\tau_y} f_{a_l}(X_s)\,ds\right] = \varphi_{a_l}(y)-\varphi_{a_l}(x),
$$

so that

$$
\psi(y)-\psi(x) = \sum_{l\in\mathbb{Z}_n} \frac{\varphi(a_{l+1})-\varphi(a_l)}{n}\left(\varphi_y(x)+\varphi_{a_l}(y)-\varphi_{a_l}(x)\right).
$$

Note that

$$
\begin{aligned}
\sum_{l\in\mathbb{Z}_n} \frac{\varphi(a_{l+1})-\varphi(a_l)}{n}\varphi_y(x) &= \frac{\varphi_y(x)}{n}\sum_{l\in\mathbb{Z}_n} \varphi(a_{l+1})-\varphi(a_l)\\
&= 0.
\end{aligned}
$$

Thus

$$
\psi(y)-\psi(x) = \frac{1}{n}\sum_{l\in\mathbb{Z}_n}\left(\varphi(a_{l+1})-\varphi(a_l)\right)\left(\varphi_{a_l}(y)-\varphi_{a_l}(x)\right).
$$

Considering for $y\in V$ the functions $\varphi=\varphi_y$ and $\psi=\psi_y$ and recalling that $\psi_y(y)=0$, gives the first relation of the lemma. Integrating this relation with respect to π in x, we get

$$
\pi[\psi_y] = \frac{1}{n}\sum_{l\in\mathbb{Z}_n}(\varphi_y(a_{l+1})-\varphi_y(a_l))(\pi[\varphi_{a_l}]-\varphi_{a_l}(y)).
$$

A well-known result (recall (4) or see, e.g., the book of Aldous and Fill [2]) asserts that the quantity $\sum_{y\in V}\pi(y)\varphi_y(x)$ does not depend on $x\in V$. It follows that

$$
\sum_{y\in V}\pi(y)(\varphi_y(a_{l+1})-\varphi_y(a_l))\pi[\varphi_{a_l}] = 0 \tag{43}
$$

and hence

$$
\sum_{y\in V}\pi(y)\pi[\psi_y] = -\frac{1}{n}\sum_{y\in V}\pi(y)\sum_{l\in\mathbb{Z}_n}(\varphi_y(a_{l+1})-\varphi_y(a_l))\varphi_{a_l}(y),
$$

which is the second equality of the lemma. For any $l \in \mathbb{Z}_n$, let ϕ_{a_l} be the function defined by:

$$\forall\ x \in V, \qquad \phi_{a_l}(x) = \sum_{y\in V} \pi(y)\varphi_{a_l}(y)(\varphi_y(x) - \varphi_y(a_l)). \tag{44}$$

We have $\sum_{y\in V} \pi(y)\pi[\psi_y] = -\frac{1}{n}\sum_{l\in\mathbb{Z}_n} \phi_{a_l}(a_{l+1})$. To compute ϕ_{a_l}, note that $\phi_{a_l}(a_l) = 0$ and that

$$\begin{aligned} L[\phi_{a_l}] &= \sum_{y\in V} \pi(y)\varphi_{a_l}(y)L[\varphi_y] \\ &= \sum_{y\in V} \pi(y)\varphi_{a_l}(y)f_y \\ &= \sum_{y\in V} \pi(y)\varphi_{a_l}(y)\left(\frac{\mathbb{1}_{\{y\}}}{\pi(y)} - 1\right) \\ &= \varphi_{a_l} - \pi[\varphi_{a_l}]. \end{aligned}$$

This observation leads us to resort once again to the martingale problem, to get for any $x \in V$,

$$\begin{aligned} \phi_{a_l}(a_l) &= \phi_{a_l}(x) + \mathbb{E}_x\left[\int_0^{\tau_{a_l}} (\varphi_{a_l}(X_s) - \pi[\varphi_{a_l}])\, ds\right] \\ &= \phi_{a_l}(x) + \mathbb{E}_x\left[\int_0^{\tau_{a_l}} \varphi_{a_l}(X_s)\, ds\right] - \pi[\varphi_{a_l}]\mathbb{E}_x[\tau_{a_l}] \\ &= \phi_{a_l}(x) - \pi[\varphi_{a_l}]\varphi_{a_l}(x) + \mathbb{E}_x\left[\int_0^{\tau_{a_l}} \varphi_{a_l}(X_s)\, ds\right] \\ &= \phi_{a_l}(x) - \pi[\varphi_{a_l}]\varphi_{a_l}(x) + \frac{1}{2}\mathbb{E}_x[\tau_{a_l}^2] \end{aligned}$$

according to Lemma 21 below. Recalling that $\phi_{a_l}(a_l) = 0$, we get

$$\phi_{a_l}(x) = \pi[\varphi_{a_l}]\varphi_{a_l}(x) - \frac{1}{2}\mathbb{E}_x[\tau_{a_l}^2] \tag{45}$$

and this leads immediately to the last equality of the lemma. □

In the previous proof, we needed the following result.

Lemma 21. *For any $x, y \in V$, we have*

$$\mathbb{E}_x\left[\int_0^{\tau_y} \varphi_y(X_s)\, ds\right] = \frac{1}{2}\mathbb{E}_x[\tau_y^2].$$

Proof. Coming back to the probabilistic interpretation of φ_y and taking into account the Markov property and Fubini theorem, we get,

$$\mathbb{E}_x\left[\int_0^{\tau_y} \varphi_y(X_s)\, ds\right] = \int_0^{+\infty} \mathbb{E}_x\left[\mathbb{1}_{\{s\le\tau_y\}}\mathbb{E}_{X_s}[\tau_y]\right]\, ds$$

$$= \int_0^{+\infty}\int_0^{+\infty} \mathbb{E}_x\left[\mathbb{1}_{\{s\leq\tau_y\}}\mathbb{E}_{X_s}[\mathbb{1}_{\{t\leq\tau_y\}}]\right]\,ds\,dt$$

$$= \int_0^{+\infty}\int_0^{+\infty} \mathbb{E}_x\left[\mathbb{1}_{\{s\leq\tau_y\}}\mathbb{E}[\mathbb{1}_{\{t\leq\tau_y\circ\theta_s\}}|\sigma(X_u\,:\,u\in[0,s])]\right]\,ds\,dt$$

$$= \int_0^{+\infty}\int_0^{+\infty} \mathbb{E}_x\left[\mathbb{1}_{\{s\leq\tau_y\}}\mathbb{1}_{\{t\leq\tau_y\circ\theta_s\}}\right]\,ds\,dt$$

$$= \int_0^{+\infty}\int_0^{+\infty} \mathbb{E}_x\left[\mathbb{1}_{\{s+t\leq\tau_y\}}\right]\,ds\,dt,$$

where θ_s is the usual shift by time s of the trajectories of X. Using the Fubini theorem, we get

$$\begin{aligned}\int_0^{+\infty}\int_0^{+\infty} \mathbb{E}_x\left[\mathbb{1}_{\{s+t\leq\tau_y\}}\right]\,ds\,dt &= \mathbb{E}_x\left[\int_0^{+\infty}\int_0^{+\infty}\mathbb{1}_{\{s+t\leq\tau_y\}}\,ds\,dt\right]\\ &= \frac{1}{2}\mathbb{E}_x\left[\int_0^{+\infty}\int_0^{+\infty}\mathbb{1}_{\{s\leq\tau_y\}}\mathbb{1}_{\{t\leq\tau_y\}}\,ds\,dt\right]\\ &= \frac{1}{2}\mathbb{E}_x\left[\left(\int_0^{\tau_y}ds\right)^2\right]\\ &= \frac{1}{2}\mathbb{E}_x[\tau_y^2]. \quad\square\end{aligned}$$

6.3. *Second derivatives: preliminaries*

Before treating the second derivative in a similar way, let us present two remarks about the quantities that appear in Lemma 20. Except for (46) and (47) below, this sub-appendix should be skipped on a first reading, as is not necessary for Theorem 8, but we believe it will be relevant for a further study of the minimizers of the mapping F on $\mathcal{L}(\pi)$.

Define the following quantities, associated with a given $L\in\mathcal{L}(\pi)$:

$$\begin{aligned}\forall\, x,y\in V, &\qquad h_L(x,y) := \frac{1}{2}\mathbb{E}_y[\tau_x^2]-\mathbb{E}_\pi[\tau_x]\mathbb{E}_y[\tau_x]\\ \forall\, A=(a_1,...,a_n)\in\mathcal{A}, &\qquad H_A(L) \;=\; \frac{1}{n}\sum_{l\in\mathbb{Z}_n} h_L(a_l,a_{l+1})\end{aligned} \tag{46}$$

Lemmas 18 and 20 allow us to write:

$$\forall\, A=(a_1,...,a_n)\in\mathcal{A}, \qquad D_AF(L)=F(L)-H_A(L) \tag{47}$$

where $D_AF(L)$ is a short hand for $D_{L_A}F(L)$.

Let us say that a cycle $A=(a_1,...,a_n)\in\mathcal{A}$ is below the generator L, if

$$\forall\, l\in\mathbb{Z}_n, \qquad L(a_l,a_{l+1})>0$$

and denote by $\mathcal{A}(L)$ the set of cycles below L. Then we have:

Lemma 22. *Assume that $L \in \mathcal{L}(\pi)$ is a minimizer of F on $\mathcal{L}(\pi)$. Then,*

$$\begin{aligned} \forall\ A \in \mathcal{A}(L), &\qquad H_A(L) = F(L), \\ \forall\ A \in \mathcal{A} \setminus \mathcal{A}(L), &\qquad H_A(L) \leq F(L). \end{aligned}$$

In particular, we get

$$F(L) = \max_{A\in\mathcal{A}} H_A(L).$$

Proof. Consider a minimizer $L \in \mathcal{L}(\pi)$ of F on $\mathcal{L}(\pi)$ and $A \in \mathcal{A}(L)$. By definition of $\mathcal{A}(L)$, for $\epsilon \in \mathbb{R}$ small enough, $L_\epsilon := (1-\epsilon)L + \epsilon L_A$ remains a Markov generator, even if ϵ is negative. It belongs to $\mathcal{L}(\pi)$ due to the fact that π is invariant for both L and L_A. Thus the mapping $\epsilon \mapsto F(L_\epsilon)$ attains a local minimum at the interior point $\epsilon = 0$, so by differentiating we get that $D_A F(L) = 0$, which implies $H_A(L) = F(L)$. For $A \in \mathcal{A} \setminus \mathcal{A}(L)$, the operator $(1-\epsilon)L + \epsilon L_A$ is not Markovian for $\epsilon < 0$. So $D_A F(L)$ only corresponds to the right derivative of $F(L_\epsilon)$ at $\epsilon = 0_+$. The minimizing assumption on L implies that $D_A F(L) \geq 0$, namely $H_A(L) \leq F(L)$. The last identity of the lemma is an immediate consequence of the previous observations and of the fact that there exists at least one cycle below L, by irreducibility. □

Next we mention a spectral relation satisfied by the quantities $(h_L(x, y))_{x,y\in V}$, reminiscent of (16). Indeed, it is proved in a similar way, as will become clear from the following proof where the arguments for (16) will be recalled.

Lemma 23. *For any $L \in \mathcal{L}(\pi)$, we have*

$$\sum_{x,y\in V} \pi(x)\pi(y)h_L(x, y) = \sum_{\lambda\in\Lambda(L)} \frac{1}{\lambda^2}. \tag{48}$$

Proof. As in the proof of Lemma 17, let $\mathcal{F}_\pi$ stand for the space of functions f on V whose mean with respect to π vanishes and denote by Π the orthogonal projection from $\mathbb{L}^2(\pi)$ to $\mathcal{F}_\pi$:

$$\forall\ f \in \mathbb{L}^2(\pi), \qquad \Pi[f] = f - \pi[f].$$

Let $(g_y)_{y\in V}$ be any orthonormal basis of $\mathbb{L}^2(\pi)$ and R be any endomorphism of $\mathcal{F}_\pi$. We have seen in Lemma 6 of [29] that

$$\mathrm{tr}(R) = \sum_{y\in V} \pi[\Pi[g_y]R[\Pi[g_y]]].$$

In [29], we considered the orthonormal basis given by

$$\forall\ y \in V, \qquad g_y := \frac{\mathbb{1}_{\{y\}}}{\sqrt{\pi(y)}}$$

and the operator $L^{-1}_{|\mathcal{F}_\pi}$ defined in the proof of Lemma 18, in order to conclude to (16), taking into account the fact that $\Pi[g_y] = \sqrt{\pi(y)} f_y$, for all $y \in V$, and that $\mathrm{tr}(L^{-1}_{|\mathcal{F}_\pi}) = -\sum_{\lambda\in\Lambda(L)} \frac{1}{\lambda}$.

To prove (48), we use $R = (L^{-1}_{|\mathcal{F}_\pi})^2$. Observe that for any $y \in V$,

$$\begin{aligned}(L^{-1}_{|\mathcal{F}_\pi})^2[f_y] &= L^{-1}_{|\mathcal{F}_\pi}[\varphi_y - \pi[\varphi_y]]\\ &= \phi_y - \pi[\phi_y]\end{aligned}$$

where ϕ_y is the unique solution of

$$\begin{cases} L[\phi_y] \,=\, \varphi_y - \pi[\varphi_y] \\ \phi_y(y) \,=\, 0. \end{cases} \tag{49}$$

(This notation agrees with that introduced in (44)). Thus we get

$$\begin{aligned}\sum_{\lambda\in\Lambda(L)} \frac{1}{\lambda^2} &= \mathrm{tr}((L^{-1}_{|\mathcal{F}_\pi})^2)\\ &= \sum_{y\in V} \pi[\Pi[g_y](L^{-1}_{|\mathcal{F}_\pi})^2[\Pi[g_y]]]\\ &= \sum_{y\in V} \pi(y)\pi[f_y(L^{-1}_{|\mathcal{F}_\pi})^2[f_y]]\\ &= \sum_{y\in V} \pi(y)\pi[f_y(\phi_y - \pi[\phi_y])]\\ &= \sum_{y\in V} \pi(y)\pi[f_y\phi_y]\\ &= \sum_{y\in V} \pi(y)\phi_y(y) - \pi(y)\pi[\phi_y]\\ &= -\sum_{y\in V} \pi(y)\pi[\phi_y]\\ &= -\sum_{x,y\in V} \pi(x)\pi(y)\phi_y(x).\end{aligned}$$

In the proof of Lemma 20 (see (45)), it was shown that

$$\forall\, x, y \in V, \qquad \phi_x(y) = -h_L(x, y),$$

which leads immediately to (48). □

6.4. Second derivatives with respect to cycles

Lemma 20 can be extended to the second derivative presented in Lemma 19, by computing similarly the function $\breve{\psi}_y$ defined by (41) with $\widetilde{L} = L_A$, for fixed $A = (a_l)_{l\in\mathbb{Z}_n} \in \mathcal{A}$ and $y \in V$. For our purposes, it is convenient to consider a generalization of this situation. Given another cycle $A' = (a'_l)_{l\in\mathbb{Z}_{n'}} \in \mathcal{A}$, consider the equation in the function Ψ_y:

$$\begin{cases} L[\Psi_y] \,=\, L_{A'}[\psi_y], \\ \Psi_y(y) \,=\, 0, \end{cases} \tag{50}$$

where ψ_y is still associated with L, A and y as in Lemma 20. Of course, when $A' = A$, we recover $\Psi_y = \breve{\psi}_y$.

Lemma 24. *For $A=(a_l)_{l\in\mathbb{Z}_n}\in\mathcal{A}$, $A'=(a'_l)_{l\in\mathbb{Z}_{n'}}\in\mathcal{A}$ and $y\in V$ given as above, consider the function Ψ_y defined by (50). Then we have, for any $x\in V$,*

$$\Psi_y(x)=\frac{1}{nn'}\sum_{l\in\mathbb{Z}_n,k\in\mathbb{Z}_{n'}}(\varphi_y(a_{l+1})-\varphi_y(a_l))(\varphi_{a_l}(a'_{k+1})-\varphi_{a_l}(a'_k))(\varphi_{a'_k}(x)-\varphi_{a'_k}(y)). \tag{51}$$

Furthermore, we get that

$$\begin{aligned}&\sum_{y\in V}\pi(y)\pi[\Psi_y]\\&=\frac{1}{nn'}\sum_{l\in\mathbb{Z}_n,k\in\mathbb{Z}_{n'}}(h_L(a'_k,a_{l+1})-h_L(a'_k,a_l))(\varphi_{a_l}(a'_{k+1})-\varphi_{a_l}(a'_k))\\&=\frac{1}{nn'}\sum_{l\in\mathbb{Z}_n,k\in\mathbb{Z}_{n'}}\left(\frac{1}{2}(\mathbb{E}_{a_{l+1}}[\tau^2_{a'_k}]-\mathbb{E}_{a_l}[\tau^2_{a'_k}])-\mathbb{E}_\pi[\tau_{a'_k}](\mathbb{E}_{a_{l+1}}[\tau_{a'_k}]-\mathbb{E}_{a_l}[\tau_{a'_k}])\right)\\&\left(\mathbb{E}_{a'_{k+1}}[\tau_{a_l}]-\mathbb{E}_{a'_k}[\tau_{a_l}]\right).\end{aligned} \tag{52}$$

Proof. From Lemma 20, we have

$$\begin{aligned}L_{A'}[\psi_y]&=\frac{1}{n}\sum_{l\in\mathbb{Z}_n}(\varphi_y(a_{l+1})-\varphi_y(a_l))L_{A'}[\varphi_{a_l}]\\&=\frac{1}{n}\sum_{l\in\mathbb{Z}_n}(\varphi_y(a_{l+1})-\varphi_y(a_l))\sum_{k\in\mathbb{Z}_{n'}}\frac{\varphi_{a_l}(a'_{k+1})-\varphi_{a_l}(a'_k)}{n'\pi(a'_k)}\mathbb{1}_{\{a'_k\}}\\&=\frac{1}{nn'}\sum_{l\in\mathbb{Z}_n}(\varphi_y(a_{l+1})-\varphi_y(a_l))\sum_{k\in\mathbb{Z}_{n'}}(\varphi_{a_l}(a'_{k+1})-\varphi_{a_l}(a'_k))\frac{\mathbb{1}_{\{a'_k\}}}{\pi(a'_k)}\\&=\frac{1}{nn'}\sum_{l\in\mathbb{Z}_n}(\varphi_y(a_{l+1})-\varphi_y(a_l))\sum_{k\in\mathbb{Z}_{n'}}(\varphi_{a_l}(a'_{k+1})-\varphi_{a_l}(a'_k))f_{a'_k},\end{aligned}$$

where we used that for any $l\in\mathbb{Z}_n$,

$$\sum_{k\in\mathbb{Z}_{n'}}(\varphi_{a_l}(a'_{k+1})-\varphi_{a_l}(a'_k))=0.$$

Thus, denoting

$$\xi_y(\cdot):=\frac{1}{nn'}\sum_{l\in\mathbb{Z}_n}(\varphi_y(a_{l+1})-\varphi_y(a_l))\sum_{k\in\mathbb{Z}_{n'}}(\varphi_{a_l}(a'_{k+1})-\varphi_{a_l}(a'_k))(\varphi_{a'_k}(\cdot)-\varphi_{a'_k}(y)),$$

we get that $L[\xi_y]=L_{A'}[\psi_y]$ and $\xi_y(y)=0$. It follows that $\Psi_y=\xi_y$, as announced.

We deduce that

$$\pi[\breve{\psi}_y]=\frac{1}{nn'}\sum_{l\in\mathbb{Z}_n}(\varphi_y(a_{l+1})-\varphi_y(a_l))\sum_{k\in\mathbb{Z}_{n'}}(\varphi_{a_l}(a'_{k+1})-\varphi_{a_l}(a'_k))(\pi[\varphi_{a'_k}]-\varphi_{a'_k}(y))$$

and

$$\sum_{y\in V}\pi(y)\pi[\breve{\psi}_y] = \frac{1}{nn'}\sum_{y\in V}\pi(y)\sum_{l\in\mathbb{Z}_n,k\in\mathbb{Z}_{n'}}(\varphi_y(a_{l+1})-\varphi_y(a_l))(\varphi_{a_l}(a'_{k+1})-\varphi_{a_l}(a'_k))(\pi[\varphi_{a'_k}]-\varphi_{a'_k}(y))$$
$$= -\frac{1}{nn'}\sum_{y\in V}\sum_{l\in\mathbb{Z}_n,k\in\mathbb{Z}_{n'}}\pi(y)(\varphi_y(a_{l+1})-\varphi_y(a_l))(\varphi_{a_l}(a'_{k+1})-\varphi_{a_l}(a'_k))\varphi_{a'_k}(y), \tag{53}$$

where we used again (recall (43)) that

$$\sum_{y\in V}\pi(y)(\varphi_y(a_{l+1})-\varphi_y(a_l)) = 0.$$

Remember also (cf. (44)) that

$$\forall\ x\in V,\qquad \sum_{y\in V}\pi(y)(\varphi_y(x)-\varphi_y(a'_k))\varphi_{a'_k}(y) = \phi_{a'_k}(x)$$
$$= -h_L(a'_k,x).$$

Thus substituting in (53)

$$\varphi_y(a_{l+1})-\varphi_y(a_l) = \varphi_y(a_{l+1})-\varphi_y(a'_k)-(\varphi_y(a_l)-\varphi_y(a'_k)),$$

we deduce (52). The last equality of the lemma is obtained by expressing h_L and φ_x, for $x\in V$, in terms of expectations of hitting times. □

Denote by $H_{A',A}(L)$ the expression given by (52). Considering the case $A'=A$, Lemma 19 leads to

$$D_{A,A}F(L) = 2F(L)-4H_A(L)+2H_{A,A}(L) \tag{54}$$

where $D_{A,A}F(L)$ is a shorthand for $D^2_{L_A}F(L)$. Lemma 24 enables us to extend (54) to the case where $A\neq A'$: if we define for any $A,A'\in\mathcal{A}$, $D_{A',A}F(L) := D_{A'}(D_AF(L))$, then we get, via computations similar to those of the proof of Lemma 19,

$$D_{A',A}F(L) = 2(F(L)-H_A(L)-H_{A'}(L)+H_{A',A}(L)).$$

The previous expressions for differentiation up to order 2 with respect to Markov generators associated to cycles can be extended to general Markov generators from $\bar{\mathcal{L}}(\pi)$. To go in this direction, we need to recall a simple result:

Lemma 25. *The extreme points of the convex set $\bar{\mathcal{L}}(\pi)$ are exactly the generators L_A for $A\in\mathcal{A}$.*

As a consequence, any $L\in\bar{\mathcal{L}}(\pi)$ can be decomposed into a barycentric sum

$$L = \sum_{A\in\mathcal{A}}p(A)L_A,$$

where p is a probability measure on $\mathcal{A}$. For an extensive discussion of such decompositions, see the book of Kalpazidou [25]. Note that the above decomposition is not unique in general, because $\bar{\mathcal{L}}(\pi)$ is not a simplex for $N\geq 3$. For instance, the generator

$$L := \frac{1}{2}\begin{pmatrix}-2 & 1 & 1\\ 1 & -2 & 1\\ 1 & 1 & -2\end{pmatrix}$$

of the simple random walk on $\mathbb{Z}_3$ can be written in the form $L = \frac{1}{2}L_{(0,1,2)} + \frac{1}{2}L_{(0,2,1)}$ and $L = \frac{1}{3}L_{(0,1)} + \frac{1}{3}L_{(1,2)} + \frac{1}{3}L_{(2,0)}$.

Nevertheless, given $\widetilde{L}, \widehat{L} \in \bar{\mathcal{L}}(\pi)$, decompose them into

$$\widetilde{L} = \sum_{A\in\mathcal{A}} \widetilde{p}(A)L_A,$$
$$\widehat{L} = \sum_{A\in\mathcal{A}} \widehat{p}(A)L_A,$$

where $\widetilde{p}, \widehat{p}$ are probability measures on $\mathcal{A}$. Then we get for any $L \in \mathcal{L}(\pi)$.

$$D_{\widetilde{L}}F(L) = \sum_{A\in\mathcal{A}} \widetilde{p}(A)D_A F(L),$$
$$D_{\widehat{L}}D_{\widetilde{L}}F(L) = \sum_{A,A'\in\mathcal{A}} \widetilde{p}(A)\widehat{p}(A')D_{A,A'}F(L).$$

It follows that we can write

$$D_{\widetilde{L}}F(L) = F(L) - H_{\widetilde{L}}(L),$$
$$D_{\widehat{L}}D_{\widetilde{L}}F(L) = 2(F(L) - H_{\widetilde{L}}(L) - H_{\widehat{L}}(L) + H_{\widetilde{L},\widehat{L}}(L)),$$

where

$$H_{\widetilde{L}}(L) = \sum_{x\neq y} \pi(x)\widetilde{L}(x,y)h_L(x,y),$$
$$H_{\widehat{L},\widetilde{L}}(L) = \sum_{x\neq y,\, x'\neq y'} \pi(x')\widehat{L}(x',y')\pi(x)\widetilde{L}(x,y)(h_L(x',y) - h_L(x',x))(\varphi_x(y') - \varphi_x(x'))$$

(definitions which conform to (46) and (52) when $\widetilde{L} = L_A$ and $\widehat{L} = L_{A'}$).

6.5. Bounds on derivatives

In view of (42) and (51), the following quantity seems to play an important role in bounding the derivatives:

$$M(L) := \max_{y,x,x'\in V} |\varphi_y(x) - \varphi_y(x')| \tag{55}$$
$$= \max_{y,x\in V} \varphi_y(x),$$

where the last equality is due to the fact that φ_y, for $y \in V$, is non-negative and vanishes at y.

Proposition 26. *We have for any $L \in \mathcal{L}(\pi)$ and $\widetilde{L}$, $\widehat{L} \in \bar{\mathcal{L}}(\pi)$,*

$$F(L) \le M(L),$$
$$\left|D_{\widetilde{L}}F(L)\right| \le M(L) + M(L)^2,$$
$$\left|D_{\widehat{L}}D_{\widetilde{L}}F(L)\right| \le 2(M(L) + M(L)^2 + M(L)^3).$$

Proof. The first bound follows directly from the definitions of $F(L)$ and φ_y, see Lemma 17. For the second, note that (42) can be extended to the solution of (40) for general $\widetilde{L} \in \bar{\mathcal{L}}(\pi)$: we get

$$\forall\ y, x \in V, \qquad \psi_y(x) = \sum_{z \neq z' \in V} \pi(z)\widetilde{L}(z,z')(\varphi_y(z') - \varphi_y(z))(\varphi_z(x) - \varphi_z(y)).$$

Taking into account the renormalization of $\widetilde{L}$, it follows that for any $y \in V$, we have for the supremum norm:

$$\|\psi_y\|_\infty \leq M(L)^2.$$

For the third bound of the lemma, note that (51) can also be extended to the more general function Ψ_y, which is the solution of

$$\begin{cases} L[\Psi_y] = \widehat{L}[\psi_y] \\ \Psi_y(y) = 0, \end{cases}$$

for any given $y \in V$ and where ψ_y is the solution of (40). It follows that for any $x, y \in V$, this generalized function Ψ_y is given by

$$\Psi_y(x) = \sum_{u \neq v,\, u' \neq v'} \pi(u')\widehat{L}(u',v')\pi(u)\widetilde{L}(u,v)(\varphi_y(v) - \varphi_y(u))(\varphi_u(v') - \varphi_u(u'))(\varphi_{u'}(x) - \varphi_{u'}(y)).$$

Thus

$$\|\Psi_y\|_\infty \leq M(L)^3. \quad \square$$

A natural question is how to upper bound $M(L)$. A first answer is to use the operator norm $|||\cdot|||_{\infty\to\infty}$ from $\mathbb{L}^\infty(\pi)$ to $\mathbb{L}^\infty(\pi)$ with the operator $L^{-1}_{|\mathcal{F}_\pi}$ introduced in the proof of Lemma 18:

$$\begin{aligned} M(L) &= \max_{y\in V} \|\varphi_y\|_\infty \\ &\leq \max_{y\in V} \|\varphi_y - \pi[\varphi_y]\|_\infty + \max_{y\in V} \pi[\varphi_y] \\ &\leq |||L^{-1}_{|\mathcal{F}_\pi}|||_{\infty\to\infty} \max_{y\in V} \|f_y\|_\infty + \max_{y\in V}(1-\pi(y))M(L) \\ &\leq |||L^{-1}_{|\mathcal{F}_\pi}|||_{\infty\to\infty} \frac{1}{\pi_\wedge} + (1-\pi_\wedge)M(L) \end{aligned}$$

where $\pi_\wedge := \min_{x\in V} \pi(x)$. It follows that

$$M(L) \leq \frac{|||L^{-1}_{|\mathcal{F}_\pi}|||_{\infty\to\infty}}{\pi_\wedge^2}.$$

But the norm $|||L^{-1}_{|\mathcal{F}_\pi}|||_{\infty\to\infty}$ does not seem so easy to evaluate. One can instead resort to the operator norm from $\mathbb{L}^2(\pi)$ to $\mathbb{L}^2(\pi)$ as follows. Denoting I the identity operator on $\mathcal{F}_\pi$, we have as above

$$\begin{aligned} M(L) &\leq |||I|||_{2\to\infty}|||L^{-1}_{|\mathcal{F}_\pi}|||_{2\to 2} \max_{y\in V} \|f_y\|_2 + (1-\pi_\wedge)M(L) \\ &\leq \frac{1}{\sqrt{\pi_\wedge}}|||L^{-1}_{|\mathcal{F}_\pi}|||_{2\to 2} \max_{y\in V} \sqrt{\frac{1}{\pi(y)} - 1} + (1-\pi_\wedge)M(L) \end{aligned}$$

$$\leq \frac{1}{\sqrt{\pi_\wedge}} |||L^{-1}_{|\mathcal{F}_\pi}|||_{2\to 2}\sqrt{\frac{1}{\pi_\wedge}} + (1-\pi_\wedge)M(L)$$
$$\leq |||L^{-1}_{|\mathcal{F}_\pi}|||_{2\to 2}\frac{1}{\pi_\wedge} + (1-\pi_\wedge)M(L).$$

As a consequence, we get

$$M(L) \leq \frac{|||L^{-1}_{|\mathcal{F}_\pi}|||_{2\to 2}}{\pi_\wedge^2}.$$

This expression is advantageous when L is reversible with respect to π, since in this situation, $|||L^{-1}_{|\mathcal{F}_\pi}|||_{2\to 2} = 1/\lambda$, where λ is the spectral gap of L, namely the smallest element of $\Lambda(L)$ (which is then in $(0,+\infty)$). Nevertheless, since we are interested in $F(L)$, there is a simple comparison:

$$M(L) \leq \frac{F(L)}{\pi_\wedge^2}. \tag{56}$$

Indeed, let $y, z \in V$ be such that $M(L) = \varphi_y(z)$. We have

$$\begin{aligned} F(L) &= \sum_{y'\in V} \pi(y')\pi[\varphi_{y'}] \\ &= \sum_{y',z'\in V} \pi(y')\pi(z')\varphi_{y'}(z') \\ &\geq \pi(y)\pi(z)\varphi_y(z) \\ &\geq \pi_\wedge^2\varphi_y(z) \\ &= \pi_\wedge^2 M(L). \end{aligned}$$

We now concentrate on the case $\pi = \upsilon$, the uniform measure and $L = L_A$, with A a Hamiltonian cycle. The following differentiation of a cyclic generator L_A in the direction of another cycle $\widetilde{A}$ has been crucial in the proof of Theorem 8.

Proposition 27. *For any $A \in \mathcal{H}$ and $\widetilde{A} \in \mathcal{A} \setminus \{A\}$, we have on $\mathcal{L}(\upsilon)$,*

$$D_{\widetilde{A}}F(L_A) \geq \frac{N-1}{2N}.$$

Proof. There is no loss of generality in assuming that $V = \mathbb{Z}_N$ and that $A = (0, 1, 2, ..., N-1)$. To simplify the notation, let us write $L = L_A$. By invariance of L and υ through the rotations $\mathbb{Z}_N \ni x \mapsto x + y \in \mathbb{Z}_N$ for any fixed $y \in \mathbb{Z}_N$, we recover that the quantity $\mathbb{E}_\upsilon[\tau_x]$ does not depend on the choice of $x \in \mathbb{Z}_N$ and that it is equal to $F(L)$. Furthermore, since under L, the Markov process waits an exponential time with parameter 1 before adding 1 to the current state, we get that for any $x, y \in \mathbb{Z}_N$, $\mathbb{E}_x[\tau_y] = \rho(x, y)$, where

$$\forall\, x, y \in \mathbb{Z}_N, \qquad \rho(x, y) := \min\{n \in \mathbb{Z}_+ \,:\, y = x + n\}.$$

It follows that $F(L) = (N-1)/2$ (for an alternative proof, see Corollary 10). Thus we get that

$$\begin{aligned} \forall\, x, y \in \mathbb{Z}_N, \qquad h_L(x, y) &= \frac{1}{2}\left(\mathbb{E}_y[\tau_x^2] - (N-1)\mathbb{E}_y[\tau_x]\right) \\ &= \frac{1}{2}\left(\mathbb{E}_y[\tau_x^2] - (N-1)\rho(y, x)\right). \end{aligned}$$

Since under $\mathbb{P}_y$, τ_x is a sum of $\rho(y,x)$ independent exponential random variables of parameter 1, we compute that

$$\mathbb{E}_y[\tau_x^2] = \rho(y,x)^2 + \rho(y,x).$$

It follows that for any $x, y \in \mathbb{Z}_N$, $h_L(x,y) = h_N(\rho(y,x))$, where

$$h_N \,:\, [0, N-1] \ni m \mapsto \frac{1}{2}\left(m^2 - (N-2)m\right).$$

This function h_N is decreasing on $[0,(N-2)/2]$, increasing on $[(N-2)/2, N-1]$ and we have $h_N(0) = 0 < (N-1)/2 = h_N(N-1)$.

Thus from the definition (46), we get that

$$\begin{aligned} \forall\, \widetilde{A} \in \mathcal{A}, \qquad H_{\widetilde{A}}(L) &\leq h_N(N-1) \\ &= H_A(L). \end{aligned}$$

Let us improve this bound when $A \neq \widetilde{A}$ (up to the cyclic identification). Consider $\widetilde{A} := (\widetilde{a}_0, \widetilde{a}_1, ..., \widetilde{a}_n) \in \mathcal{A}$, and assume first there exists at least one $l \in \mathbb{Z}_n$, such that $\rho(\widetilde{a}_{l+1}, \widetilde{a}_l) \neq N-1$. We get

$$\begin{aligned} H_{\widetilde{A}}(L) &\leq \frac{n-1}{n} h_N(N-1) + \frac{1}{n}\max\{h_N(0), h_N(N-2)\} \\ &= \frac{n-1}{n} H_A(L) \\ &\leq \frac{N-1}{N} H_A(L), \end{aligned}$$

where in the equality, we used that $h_N(0) = h_N(N-2) = 0$ and that $h_N(N-1) = H_A(L)$, according to (46). Next assume on the contrary that for all $l \in \mathbb{Z}_n$, we have $\rho(\widetilde{a}_{l+1}, \widetilde{a}_l) = N-1$, which means that $\widetilde{a}_l = \widetilde{a}_{l+1} + N - 1$, i.e. $\widetilde{a}_{l+1} = \widetilde{a}_l + 1$ in $\mathbb{Z}_N$. Since this must be true for all $l \in \mathbb{Z}_n$, it follows that $n = N$ and that $\widetilde{A}$ must be of the form $(k, k+1, ..., k+N-1)$, for some $k \in \mathbb{Z}_N$, namely, it is identified with A.

As a consequence, taking into account (47), we obtain,

$$\begin{aligned} \forall\, \widetilde{A} \in \mathcal{A} \setminus \{A\}, \qquad D_{\widetilde{A}}F(L) &= F(L) - H_{\widetilde{A}}(L) \\ &\geq F(L) - \frac{N-1}{N} H_A(L) \\ &= D_A F(L) + \frac{1}{N} H_A(L) \\ &= \frac{N-1}{2N} \end{aligned}$$

where we took into account that $D_A F(L) = 0$, because $L = L_A$ is not modified in the direction of the cycle A. □

Above we worked with the complete graph K_V and the associated set of Markov generators $\mathcal{L}(\pi)$. All the previous considerations can be extended to the case of $\mathcal{L}(G,\pi)$, where the graph G is as in the introduction. The only difference is that $\mathcal{A}$ has to be replaced by $\mathcal{A}(G)$, the set of cycles using only edges from E. For instance, Lemma 25 has to be replaced by

Lemma 28. *The extreme points of the convex set $\bar{\mathcal{L}}(G,\pi)$ (the set of normalized Markov generators L, compatible with G and admitting π for invariant probability) are exactly the generators L_A for $A \in \mathcal{A}(G)$.*

7. Computations on the simplest example of non-Hamiltonian connected graph

The length 2 segment $S_2 := (\{0,1,2\}, \{(0,1),(1,0),(1,2),(2,1)\})$ is the simplest non-Hamiltonian (strongly) connected graph. We compute here the minimizer of F on $\mathcal{L}(S_2, \pi)$, for any probability measure π on $\{0,1,2\}$. We hope this example will motivate further investigation of the minimizers of F in the challenging non-Hamiltonian framework.

To simplify the notation, write $x = \pi(0), y = \pi(1)$ and $z = \pi(2)$, by assumption we have that $x, y, z > 0$ and $x + y + z = 1$. Modulo exchanging the vertices 0 and 2, we assume that $|x - 1/2| \geq |z - 1/2|$.

Any Markov generator L from $\mathcal{L}(S_2, \pi)$ has the form

$$L := \begin{pmatrix} -a & a & 0 \\ \alpha & -\alpha - \beta & \beta \\ 0 & b & -b \end{pmatrix}$$

where the coefficients $a, \alpha, \beta, b > 0$ satisfy,

$$\begin{aligned} xa &= y\alpha, \\ y\beta &= zb, \\ xa + y(\alpha + \beta) + zb &= 1. \end{aligned}$$

The first two equalities correspond to the invariance of π for L (here π is even reversible for the birth and death generator L) and the third one is the normalization condition, which can be rewritten as

$$2xa + 2zb = 1 \tag{57}$$

Denote $\Lambda(L) = \{\lambda_1, \lambda_2\}$, its elements are the non-zero roots in X of the polynomial $\det(X + L)$. We compute that

$$\det(X + L) = X(X^2 - (a + \alpha + \beta + b)X + ab + a\beta + \alpha b),$$

so that

$$\begin{aligned} \lambda_1 + \lambda_2 &= a + \alpha + \beta + b, \\ \lambda_1 \lambda_2 &= ab + a\beta + \alpha b. \end{aligned}$$

From (16), we have

$$\begin{aligned} F(L) &= \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \\ &= \frac{\lambda_1 + \lambda_2}{\lambda_1 \lambda_2} \\ &= \frac{a + \alpha + \beta + b}{ab + a\beta + \alpha b} \\ &= \frac{a\left(1 + \frac{x}{y}\right) + b\left(1 + \frac{z}{y}\right)}{ab\left(1 + \frac{x}{y} + \frac{z}{y}\right)} \\ &= \frac{a(x + y) + b(y + z)}{ab(y + x + z)} \end{aligned}$$

$$
\begin{aligned}
&= \frac{a(x+y)+b(y+z)}{ab} \\
&= \frac{x+y}{b} + \frac{y+z}{a} \\
&= \frac{1-z}{b} + \frac{1-x}{a}.
\end{aligned}
$$

Taking into account (57), the minimizer of F on $\mathcal{L}(S_2, \pi)$ corresponds to the minimizer of

$$
(0, 1/(2x)) \ni a \mapsto 2z\frac{1-z}{1-2xa} + \frac{1-x}{a}. \tag{58}
$$

We are thus led to the second order equation in a:

$$
4x(z(1-z) - x(1-x))a^2 + 4x(1-x)a - (1-x) = 0. \tag{59}
$$

Due to the assumption $|x - 1/2| \geq |z - 1/2|$, the first coefficient is non-negative. We consider two cases.

• If $|x-1/2| = |z-1/2|$, then (59) degenerates into a first order equation and $a := 1/(4x)$ is the minimizer of the mapping (58). It follows that the minimizer of F on $\mathcal{L}(S_2, \pi)$ is

$$
L_\wedge := \begin{pmatrix} -\frac{1}{4x} & \frac{1}{4x} & 0 \\ \frac{1}{4y} & -\frac{1}{2y} & \frac{1}{4y} \\ 0 & \frac{1}{4z} & -\frac{1}{4z} \end{pmatrix}
$$

and the minimal value $F_\wedge(S_2, \pi)$ of F on $\mathcal{L}(S_2, \pi)$ is

$$
F(L_\wedge) \;=\; 4(1-z)z + 4(1-x)x \;=\; 8x(1-x).
$$

In particular, for $\pi = \upsilon$, the uniform distribution on $\{0, 1, 2\}$, we get

$$
L_\wedge := \frac{1}{4}\begin{pmatrix} -3 & 3 & 0 \\ 3 & -6 & 3 \\ 0 & 3 & -3 \end{pmatrix}
$$

and $F_\wedge(S_2, \upsilon) = 16/9$.

• If $|x - 1/2| > |z - 1/2|$, then (59) admits two solutions

$$
a_\pm := \frac{-x(1-x) \pm \sqrt{x(1-x)z(1-z)}}{2x(z(1-z) - x(1-x))},
$$

but only a_+ belongs to $(0, 1/(2x))$ and is in fact the minimizer of the mapping (58). This value can be simplified into

$$
a_+ = \frac{1}{2x}\frac{\sqrt{x(1-x)}}{\sqrt{x(1-x)} + \sqrt{z(1-z)}}.
$$

It follows that the minimizer of F on $\mathcal{L}(S_2, \pi)$ is

$$
\begin{aligned}
L_\wedge &:= \frac{1}{\sqrt{x(1-x)}+\sqrt{z(1-z)}}\begin{pmatrix} -\frac{\sqrt{x(1-x)}}{2x} & \frac{\sqrt{x(1-x)}}{2x} & 0 \\ \frac{\sqrt{x(1-x)}}{2y} & -\frac{\sqrt{x(1-x)}+\sqrt{z(1-z)}}{2y} & \frac{\sqrt{z(1-z)}}{2y} \\ 0 & \frac{\sqrt{z(1-z)}}{2z} & -\frac{\sqrt{z(1-z)}}{2z} \end{pmatrix} \\
&= pL_{(0,1)} + (1-p)L_{(1,2)}
\end{aligned}
$$

with the notation introduced in (23) and

$$
p := \frac{\sqrt{x(1-x)}}{\sqrt{x(1-x)}+\sqrt{z(1-z)}}.
$$

The minimal value $F_\wedge(S_2,\pi)$ of F on $\mathcal{L}(S_2,\pi)$ is

$$
F(L_\wedge) = 2\left(\sqrt{x(1-x)}+\sqrt{z(1-z)}\right)^2.
$$

Letting $|x-1/2|$ converge to $|z-1/2|$, we recover the values of $L_\wedge$ and $F(L_\wedge)$ obtained in the previous case.

8. Conclusions and future directions

Our main contributions in this article have been the following. First, we revisit the Markov decision theoretical framework for the Hamiltonian cycle problem initiated by Filar and collaborators, and establish the optimality of a Hamiltonian cycle for the problem of minimizing the mean hitting time to a state picked according to the stationary distribution for the chain. The latter quantity also equals the trace of the fundamental matrix associated with the chain. Our result is the most general of this kind in so far as the optimality is over all random processes compatible with the graph. This is done for both discrete and continuous time chains. We then took up a detailed study of the continuous state scenario, with suitable normalized exit rates. Because of the additional degrees of freedom in terms of sojourn times, this theory is much richer. We establish the optimality of Hamiltonian cycle as before when the stationary distribution is uniform, and establish its robustness to small perturbation of the uniform stationary distribution. We give a counterexample to show that if the perturbation is large, the claim fails.

For future work, an important direction is: to initiate research on continuous time Markov chains in this context, as the field is wide open, and furthermore, the problem is non-trivially distinct from the well studied discrete case, calling upon novel techniques for proofs.

References

[1] David J. Aldous, Lower bounds for covering times for reversible Markov chains and random walks on graphs, J. Theoret. Probab. 2 (1) (1989) 91–100.

[2] D. Aldous, J.A. Fill, Reversible Markov Chains and Random Walks on Graphs, Unfinished monograph, available at http://www.stat.berkeley.edu/~aldous/RWG/book.html, 2002, recompiled 2014.

[3] K. Avrachenkov, A. Eshragh, J. Filar, On transition matrices of Markov chains corresponding to Hamiltonian cycles, Ann. Oper. Res. 243 (1) (2016) 19–35.

[4] D.P. Bertsekas, J.N. Tsitsiklis, An analysis of stochastic shortest path problems, Math. Oper. Res. 16 (3) (1991) 580–595.

[5] V.S. Borkar, V. Ejov, J.A. Filar, Directed graphs, Hamiltonicity and doubly stochastic matrices, Random Structures Algorithms 25 (2004) 376–395.

[6] V.S. Borkar, V. Ejov, J.A. Filar, On the Hamiltonicity gap and doubly stochastic matrices, Random Structures Algorithms 34 (4) (2009) 502–519.

[7] V.S. Borkar, V. Ejov, J.A. Filar, G.T. Nguyen, Hamiltonian Cycle Problem and Markov Chains, Springer Science & Business Media, New York, 2012.

[8] V.S. Borkar, J.A. Filar, Markov chains, Hamiltonian cycles, and volumes of convex bodies, J. Global Optim. 55 (3) (2013) 633–639.

[9] H. Cui, Y.-H. Mao, Eigentime identity for asymmetric finite Markov chains, Front. Math. China 5 (4) (2010) 623–634.
[10] V. Ejov, J.A. Filar, W. Murray, G.T. Nguyen, Determinants and longest cycles of graphs, SIAM J. Discrete Math. 22 (3) (2008) 1215–1225.
[11] V. Ejov, J.A. Filar, M. Haythoroe, G.T. Nguyen, Refined MDP-based branch-and-fix algorithm for the hamiltonian cycle problem, Math. Oper. Res. 34 (3) (2009) 758–768.
[12] V. Ejov, J.A. Filar, M. Nguyen, Hamiltonian cycles and singularly perturbed Markov chains, Math. Oper. Res. 19 (1) (2004) 223–237.
[13] V. Ejov, G.T. Nguyen, Consistent behavior of certain determinants induced by graphs, Linear Algebra Appl. 431 (5–7) (2009) 543–552.
[14] V. Ejov, N. Litvak, G.T. Nguyen, P.G. Taylor, Proof of the Hamiltonicity-trace conjecture for singularly perturbed Markov chains, J. Appl. Probab. 48 (4) (2011) 901–910.
[15] A. Eshragh, J.A. Filar, Hamiltonian cycles, random walks, and the geometry of the space of discounted occupation measures, Math. Oper. Res. 36 (2) (2011) 258–278.
[16] A. Eshragh, J.A. Filar, T. Kalinowski, S. Mohammadian, Hamiltonian cycles and subsets of discounted occupational measures, Math. Oper. Res. (2019), in press, https://arxiv.org/abs/1805.04725.
[17] E.A. Feinberg, Constrained discounted Markov decision processes and Hamiltonian cycles, Math. Oper. Res. 25 (1) (2000) 130–140.
[18] J.A. Filar, Controlled Markov chains, graphs, and Hamiltonicity, Found. Trends Stoch. Syst. 1 (2) (2007) 77–162.
[19] J.A. Filar, D. Krass, Hamiltonian cycles and Markov chains, Math. Oper. Res. 19 (1994) 223–237.
[20] J.A. Filar, K. Liu, Hamiltonian cycle problem and singularly perturbed decision processes, in: Statistics, Probability and Game Theory: Papers in Honor of David Blackwell, in: IMS Lecture Notes – Monograph Series, vol. 30, Institute of Mathematical Statistics, Hayward, CA, 1996, pp. 44–63.
[21] J.A. Filar, A. Moeini, Hamiltonian cycle curves in a space of discounted occupational measures, Ann. Oper. Res. (2015), https://doi.org/10.1007/s10479-015-2030-2.
[22] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman, San Francisco, 1979.
[23] M. Held, R.M. Karp, A dynamic programming approach to sequencing problems, J. Soc. Indust. Appl. Math. 10 (1) (1962) 196–210.
[24] J.J. Hunter, Mixing times with applications to perturbed Markov chains, Linear Algebra Appl. 417 (1) (2006) 108–123.
[25] S.L. Kalpazidou, Cycle Representations of Markov Processes, second edition, Springer Verlag, New York, 2006.
[26] T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, Berlin, 1995, reprint of the 1980 edition.
[27] J.G. Kemeny, J.L. Snell, Finite Markov Chains, van Nostrand, Princeton, NJ, 1960.
[28] N. Litvak, V. Ejov, Markov chains and optimality of the Hamiltonian cycle, Math. Oper. Res. 34 (1) (2009) 71–82.
[29] L. Miclo, An absorbing eigentime identity, Markov Process. Related Fields 21 (2) (2015) 249–262.
[30] R. Montenegro, P. Tetali, Mathematical aspects of mixing times in Markov chains, Found. Trends Theor. Comput. Sci. 1 (3) (2006) 237–354.