



## When Does Memory Speed-up Mixing?

Simon Apers, Alain Sarlette, Francesco Ticozzi

### ► To cite this version:

Simon Apers, Alain Sarlette, Francesco Ticozzi. When Does Memory Speed-up Mixing?. IEEE Conference on Decision and Control, Dec 2017, Melbourne, Australia. hal-01634630

**HAL Id: hal-01634630**

**<https://inria.hal.science/hal-01634630>**

Submitted on 14 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# When Does Memory Speed-up Mixing?\*

Simon Apers, Alain Sarlette, and Francesco Ticozzi

**Abstract**—We investigate under which conditions a higher-order Markov chain, or more generally a Markov chain on an extended state space, can mix faster than a standard Markov chain on a graph of interest. We find that, depending on the constraints on the dynamics, two very different scenarios can emerge: under strict invariance of the target marginal and for general initialization of the lifted chain no speedup is possible; on the other hand, if these requirements are both relaxed, the lifted dynamics can achieve mixing in a time that corresponds to the diameter of the graph, which is optimal.

## I. INTRODUCTION

More than one hundred years after Markov started studying these stochastic processes, the role of Markov chains, not just as modeling tools, but as key components of algorithms for counting, sampling and simulating is nowadays well established [1], [2]. The key feature that is at the heart of many of these algorithms is the convergence of the stochastic process towards its equilibrium distribution. Famous examples can be found in physics (the Metropolis-Hastings algorithms [3], [4] and Glauber dynamics [5]) and complexity theory (approximating the volume of a convex body [6] or the permanent of a non-negative matrix [7]). Gradually, the questions of interest for the mathematical and engineering community shifted from characterizing the rates of convergence to determining their *mixing times*, i.e. the time needed to get close to the target distribution by a prescribed distance. This type of analysis is particularly relevant when one tries to understand how the mixing behavior is affected by the size of the system, and gives rise to interesting behavior, as the cutoff phenomenon [8].

In seeking fast mixing behaviors, an interesting approach has been proposed and studied in [9], [10], [11], [12] for Markov chains on graphs: the actual Markov dynamics is designed to take place on an extended graph, which allows for faster mixing – leading to the idea of *lifted Markov chains*. This can be equivalently seen as the result of a *memory effect*.

From a system-theoretic viewpoint, it is like designing a hidden Markov model that allows for an induced (non-Markov) stabilizing dynamics on the system of interest.

Work partially by the University of Padua under the project CPDA140897/14 and QCON.

S. Apers is with the Department of Electronics and Information Systems, Ghent University, Technologiepark 914, 9052 Zwijnaard-Ghent, Belgium (simon.apers@ugent.be); A. Sarlette is with Ghent University and with the QUANTIC lab, INRIA Paris, rue Simone Iff 2, 75012 Paris, France (alain.sarlette@inria.fr); F. Ticozzi is with the Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy (ticozzi@dei.unipd.it), and the Department of Physics and Astronomy, Dartmouth College, 3127 Wilder, Hanover, NH (USA).

The key results regarding fast mixing with lifted chains include a proof of concept on some key examples [9], followed by a generic construction for lifting arbitrary graphs, which reaches the conductance bounds on the mixing time [10]. The lift construction of [10] is very extensive in resources, requiring amongst others to solve a complex optimization problem over the graph. A more recent string of papers addresses this problem, proposing more efficient approaches towards lifting Markov chains on general graphs [13], [14], [15], [16], [11], [17].

In this paper, we focus on the following question: When can we expect an advantage in the mixing speed by using lifted Markov chains instead of standard Markov dynamics? We show that the answer depends heavily on the details on the problem at hand. We focus on time-homogeneous Markov chains that are *locality constrained*, namely whose allowed transitions are determined by a reference underlying graph. For this class of dynamics, we investigate in particular the effects of the design constraints regarding: (1) the ability of initializing the lifted chain (choosing the initial state of the memory); and (2) the invariance of the target marginal distribution. In fact, we show that if the lifted Markov chain can not be initialized, and thus convergence needs to happen for all initial states on the expanded state space, while preserving the target marginal distribution at all times it is attained, then *no advantage in mixing speed can be gained by using lifted chains*. The result is derived by simple algebraic manipulations, reminiscent of those involved in the observability analysis for linear systems.

On the other hand, if we have the ability of properly initializing the lifted chain, so that convergence needs to happen only for a subset of the expanded state space, and we suitably relax the invariance constraint and only require asymptotic convergence, then *we show that it is possible to achieve mixing in finite time*, essentially corresponding to the diameter of the graph. This is achieved by adapting some design tools from the theory of stochastic bridges [18]. The resulting lift has a dimension, and hence requires resources, that are polynomial in the number of initial graph nodes. Being the diameter time the minimal possible mixing time, the lift performance is optimal and can significantly accelerate convergence when the performance of the non-lifted Markov chain is poor. While the construction of such lift relies on global information and may not be practical, it still shows where lifts can provide an advantage, and it opens the possibility of finding fast-mixing lifts that rely only on local information.

Further requirements on the lifts, including matching of some target ergodic flows as in [10] and convergence to the

lift's own equilibrium distribution are not treated here, but will be the object of further study. The concluding section highlights open research directions which are impacted by the results and techniques developed in this paper.

## II. PROBLEM FORMULATION

### A. Markov chains on graphs and their lifts

Denote the graph of interest, where mixing should occur, as  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , with  $\mathcal{V}$  a set of  $N$  nodes and  $\mathcal{E}$  the set of edges i.e. node-pairs. Let  $\mathbb{P}_N$  be the set of probability vectors in  $\mathbb{R}^N$ , i.e. each  $y \in \mathbb{P}_N$  satisfies  $y_i \geq 0 \forall i = 1, 2, \dots, N$  and  $\sum_{i \in \mathcal{V}} y_i = 1$ . Each such  $y$  represents a probability distribution on  $\mathcal{V}$ . The slightly unusual notation  $y$  for probability distribution is inspired by systems theory, hinting at  $y$  being the “output of interest”, whose evolution can be induced by the evolution of some larger state  $x$ .

The targets of interest in what follows will be a distribution  $\pi$  with full support, namely such that  $\pi_i > 0$  for all  $i \in [1, \dots, N]$ . These are a dense set in  $\mathbb{P}_N$ , and the full-support hypothesis allows us to simplify the analysis.

A common approach to address this mixing problem is to approach the steady-state distribution  $\pi \in \mathbb{P}_N$  by iterating a linear, stochastic discrete-time map

$$y(t+1) = P y(t).$$

Here  $y(t)$  is the distribution over the graph nodes at time  $t$ , satisfying  $y(t) \in \mathbb{P}_N \forall t$  (and thus as well the target steady state  $\pi \in \mathbb{P}_N$ ). The transition map  $P$  contains the probabilities  $P_{i,j}$  to jump from node  $j$  to node  $i$  and must satisfy the *locality constraints* induced by  $\mathcal{G}$ : namely,  $P_{i,j} = 0$  if  $\mathcal{E}$  contains no edge from  $j$  to  $i$ . The goal is to converge towards  $\pi$  as fast as possible from any initial  $y(0)$ . Notice that, being  $\pi_i > 0$  for all  $i$ ,  $P$  must be irreducible in order to allow  $y(t)$  to converge to  $\pi$  from any initial distribution. The evolution with  $P$  defines a Markov chain on the node set  $\mathcal{V}$ .

In order to quantify the speed of convergence, we here consider mixing time. A time-homogenous Markov chain on  $\mathcal{G}$  associated to a transition matrix  $P$  is said to mix to  $\pi$  if  $P\pi = \pi$  and for all  $\epsilon > 0$  there exists  $\tau(\epsilon) > 0$  such that for all  $p \in \mathbb{P}_N$  we have:

$$\|P^t p - \pi\|_{TV} \leq \epsilon \text{ for all } t \geq \tau(\epsilon).$$

We call  $\tau(\epsilon)$  its  $\epsilon$ -mixing time<sup>1</sup>. It is typical to consider  $\tau(1/4)$  as a reference mixing time.

It has been shown that convergence can be accelerated, under the same locality constraints, by adding memory to the Markov chain, see e.g. [9], from which we borrow a particularly clear example.

**Example 1 (Diaconis lift on the cycle):** Consider a stochastic process on the finite cycle graph, i.e. the graph with nodes  $V = \{1, \dots, N\}$ , and where node  $k$  is connected by an edge to nodes  $k+1$  and  $k-1$  modulo  $N$ . The idea is to modify

<sup>1</sup>For completeness, we recall that the total variation distance between two distributions  $y(1)$  and  $y(2)$  is  $1/2$  times the 1-norm of their difference  $\|y(1) - y(2)\|_1 = \sum_{i=1}^N |y_i(1) - y_i(2)|$ .

the standard random walk on the graph, where the current state can move to either node connected to it with equal probability, in order to make the walker's next step depend on the current position *and the last move* on the cycle  $\mathcal{G}$ . Explicitly, an extended graph is constructed by associating to each node of the original graph  $V$  two nodes,  $(\pm 1, k)$ , that indicate if the current state  $v$  has been reached from  $k+1$  or  $k-1$  respectively. The full graph becomes:  $\hat{\mathcal{V}} = \{(s, k) : k = 1, 2, \dots, N \text{ and } s \in \pm 1\}$ . On this extended graph, a new dynamics is constructed, maintaining circular symmetry. In the following  $e_{\pm}$  denotes the column vectors  $(1, 0)^T$  and  $(0, 1)^T$ ,  $P^{(\pm 1)}$  denote clockwise and anti-clockwise rotation on the cycle  $\mathcal{G}$ , respectively, and  $\otimes$  is the Kronecker product. In a matrix representation, the new transition map has the form:

$$A = \sum_{i,j \in \pm 1} Q_{i,j} e_i e_j^{\dagger} \otimes P^{(j)}.$$

The crucial element is  $Q$ , a stochastic matrix which defines a Markov chain on the additional states  $\pm 1$ , with  $Q_{+1,+1} = Q_{-1,-1} = 1 - 1/N$  and  $Q_{+1,-1} = Q_{-1,+1} = 1/N$ . With this choice, keeping the same “direction” of movement is preferred, but there is a small  $1/N$  probability of switching. The allowed transitions and the relative probabilities are depicted in Figure 1.

The mixing time  $\tau(1/4)$  of the whole distribution on  $\hat{\mathcal{V}}$  with this lift is of order  $N$ , while the mixing time  $\tau(1/4)$  of any non-lifted walk on the cycle (see e.g. [19], showing the best is the simple walk with probability  $1/2$  to take each edge) would be of order  $N^2$ .

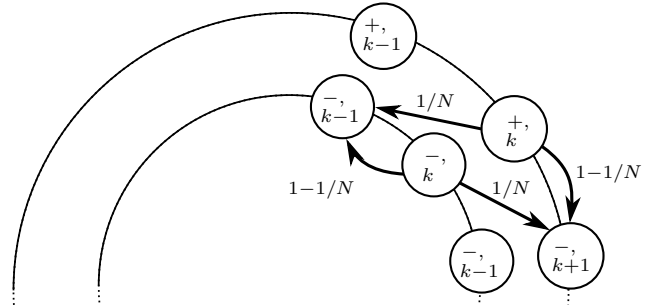


Fig. 1. The Diaconis lift on a cycle of  $N$  nodes.

Extending this construction to general graphs leads to Markov chains on lifted graphs or, for short, *lifted Markov chains*.

**Definition 1:** A graph  $\hat{\mathcal{G}} = (\hat{\mathcal{V}}, \hat{\mathcal{E}})$  on  $\hat{N}$  nodes is said to be a *lift* of  $\mathcal{G}$  if there exists a surjective map  $\mathfrak{c} : \hat{\mathcal{V}} \mapsto \mathcal{V}$ , such that:

$$(i, j) \in \hat{\mathcal{E}} \implies (\mathfrak{c}(i), \mathfrak{c}(j)) \in \mathcal{E}.$$

We denote by  $\mathfrak{c}^{-1}$  the map that takes as input a single node  $k \in \mathcal{V}$  and outputs all the nodes  $j \in \hat{\mathcal{V}}$  for which  $\mathfrak{c}(j) = k$ . We will denote by  $x \in \mathbb{P}_{\hat{N}}$  a distribution over the lifted graph nodes  $\hat{\mathcal{V}}$ . The associated marginal distribution over  $\mathcal{V}$  is given by  $y_k = \sum_{j \in \mathfrak{c}^{-1}(k)} x_j$ . Notice that it is a linear map, and can be represented as

$$y = Cx, \quad (1)$$

with  $C$  a matrix of zeroes and ones. In a *lifted Markov chain* for  $\mathcal{G}$ , the distribution  $y(t)$  on  $\mathcal{V}$  at time  $t$  is obtained as the marginal of  $x(t)$ , whose evolution is generated by a linear, stochastic discrete-time map

$$x(t+1) = Ax(t) \quad (2)$$

where  $A$  satisfies the locality constraints of  $\hat{\mathcal{G}}$ , and thereby also of  $\mathcal{G}$  in the following sense: for each  $x$ , there exists a stochastic matrix  $P^{(x)}$  satisfying the locality constraints of  $\mathcal{G}$  and such that (1),(2) corresponds to  $y(t+1) = P^{(x)}y(t)$ . Explicitly, we have  $P^{(x)} = C A B^{(x)}$  where  $B^{(x)}$  is a linear stochastic map from  $\mathbb{R}^N$  to  $\mathbb{R}^{\hat{N}}$ , with

$$B_{i,j}^{(x)} = \frac{x_i(t)}{\sum_{k \in c^{-1}(j)} x_k(t)} \text{ if } c(i) = j,$$

and  $B_{i,j}^{(x)} = 0$  otherwise.

In Markov modeling,  $A$  would be called a hidden Markov chain [20]. In algorithmic applications, which are our main motivation, the pair  $(\hat{\mathcal{G}}, A)$  is to be *designed* towards accelerating the convergence towards  $\pi$  with respect to the (best) Markov chain  $P$  on the original graph  $\mathcal{G}$ .

### B. Initialization and Invariance

When a stochastic dynamics is seen as an algorithm, it is crucial to specify whether, and how, we are allowed to initialize it, or if we have no control over its initial state. We can reasonably assume that the input of such algorithm is a node of  $\mathcal{V}$ , chosen accordingly to an initial distribution  $y(0) = y_0$ , on which we have no control. We consider two possible scenarios and the associated set of initial distributions  $\mathcal{S}$ :

- (S) In a first scenario, it is possible to initialize the lifted evolution depending on the initial input. The algorithm design can then, in addition to  $\hat{\mathcal{G}}$  and  $A$ , choose how to lift the weight  $y_k(0)$  attributed to each node  $k \in \mathcal{V}$  of the original graph  $\mathcal{G}$ , onto its associated lifted nodes  $x_{c^{-1}(k)}(0)$  in agreement with the locality constraints. We further require, consistently with the dynamics, the designed initialization to be a *linear* map:  $F : y(0) \mapsto x(0)$ , satisfying

$$CFy(0) = y(0) \quad \forall y(0).$$

In particular, we have  $F_{k,j} = 0$  whenever  $c(k) \neq j$ .

Hence, the set of relevant initial conditions  $\mathcal{S}$  for the lifted Markov chain does not comprise all possible distributions  $x(0)$  on  $\hat{\mathcal{V}}$ , but only those of the form  $x(0) = Fy(0)$ , for all initial distributions  $y(0)$  on  $\mathcal{V}$ .

- (s) In some other cases, there might be no control over the initialization of the lifted dynamics. The set of relevant initial conditions  $\mathcal{S}$  are then all  $x \in \mathbb{P}_{\hat{N}}$ .

### C. Invariance of the target marginal

For a Markov chain, mixing is necessarily towards its unique *invariant* distribution, i.e.  $\pi = P\pi$ . For a lifted Markov chain, however,  $Cx(t) = \pi$  does not necessarily imply  $Cx(t+1) = \pi$ . However, it may seem natural to

request that we do not leave the target  $\pi$  if we start there at  $t = 0$ . We thus identify two possible scenarios:

- (i) We impose  $Cx(t) = \pi$  for all  $t > 0$  whenever  $Cx(0) = \pi$ , for all allowed  $x(0) \in \mathcal{S}$ .
- (l) We allow  $y(t) \neq \pi$  for some  $t \geq 0$  even when  $y(0) = \pi$ .

Notice that, for standard time-homogeneous Markov chains, if a transition matrix mixes that must have the target distribution as its unique stationary one, so the scenarios discussed here are relevant only for lifted chains.

### D. Marginal vs lift mixing time

Most papers that bound the mixing time of lifted Markov chains analyze in fact how fast the state *on the lifted space*  $x$  converges to *its own* stationary value  $\bar{x}$ , i.e. they consider  $\|A^t x - \bar{x}\|_{TV}$ . Our original algorithmic task however is to accelerate convergence of the *marginal*  $y(t) = Cx(t)$ , compared to the performance of the original chain  $P$ . To this aim, we define the marginal mixing time.

*Definition 2 (Marginal mixing time):* A lifted chain on  $\hat{\mathcal{G}}$  associated to a transition matrix  $A$  is said to mix to the marginal  $\pi$  on  $\mathcal{G}$  from initial conditions  $\mathcal{S}$ , if for all  $\epsilon > 0$  there exists  $\tau_M(\epsilon) > 0$  such that for all  $x \in \mathcal{S}$  we have:

$$\|CA^t x - \pi\|_{TV} \leq \epsilon \text{ for all } t \geq \tau(\epsilon).$$

We call  $\tau_M(\epsilon)$  its  $\epsilon$ -*marginal mixing time*.

Of course  $\tau_M(\epsilon) \leq \tau(\epsilon)$  for all  $\epsilon$ . While the convergence of  $x$  is indeed a sufficient proxy for the convergence of  $y = Cx$ , it is not truly necessary. One might argue that for generic  $A$  we expect  $x$  and  $y = Cx$  to have similar convergence speeds. However in our application, the specific lifts designed to speed up convergence are all but generic, and this distinction could become relevant. For instance, in our designs based on stochastic bridges the lifted Markov chain  $x$  in fact does not converge to a stationary value, but the projected state  $y$  does. Furthermore, it is easy to construct lifted walks where  $y$  converges much faster than  $x$ .

## III. SCENARIOS WHERE LIFTED CHAINS CAN NOT SPEED UP MIXING

We find that, under the constraints (si), the lifted Markov chain cannot go faster than the best non-lifted chain  $P$  compatible with the constraints, *even if we only look at the marginal mixing time*.

*Theorem 1:* Under constraints (si), for any lifted Markov chain  $(\hat{\mathcal{G}}, A)$  whose marginal  $y_t = CAx_t$  mixes to  $\pi$ , there exists a stochastic  $P^q$  such that  $y_{t+1} = P^q y_t$  for all  $t$ .

*Proof:* The essential idea of the proof is that to satisfy invariance of  $\pi$  for all initializations on  $\hat{\mathcal{V}}$ , it is necessary that any two  $x^{(1)}, x^{(2)}$  for which  $Cx^{(1)} = Cx^{(2)}$ , induce the same flow on  $\mathcal{G}$ .

Since we have (s), the lift can start from any distribution  $x$  over  $\hat{\mathcal{V}}$ . Invariance (i) then requires that any  $x$  for which  $Cx = \pi$ , satisfies  $CAx = \pi$ .

Given a lifted Markov chain satisfying (si), consider a map  $q : \mathcal{V} \mapsto \hat{\mathcal{V}}$  which to every  $j \in \mathcal{V}$  assigns a single node  $k_j \in \hat{\mathcal{V}}$  for which  $c(k_j) = j$ , and let  $x = q(y)$  denote the distribution with  $x_{q(j)} = y_j$  for all  $j \in \mathcal{V}$ , and  $x_i = 0$  for

all remaining  $i \in \hat{\mathcal{V}}$ . We will show that for any  $x(t)$  with  $y(t) = Cx(t)$ , the lifted Markov chain satisfies

$$y(t+1) = P^q y(t), \quad (3)$$

i.e. it behaves like the non-lifted Markov chain  $P^q$  defined by:

$$P_{i,j}^q = \sum_{\ell \in \mathfrak{c}^{-1}(i)} A_{\ell, q(j)}.$$

Proving (3) amounts to proving that

$$CAx = P^q Cx \quad (4)$$

for all  $x \in \mathbb{P}_{\hat{\mathcal{N}}}$ . For any  $x$  of the form  $x = q(y)$ , with  $y \in \mathbb{P}_N$ , we indeed have (4) by construction. For any other  $x$ , defining  $x^{(q)} = q(Cx)$ , there remains to show that  $CAx = CAx^{(q)}$ . To do so, select some  $a > 0$  such that  $a\pi_j > y_j$  for all  $j \in \mathcal{V}$  and define  $\pi' = \eta(a\pi - y)$ , with  $1/\eta = \sum_{j \in \mathcal{V}} (a\pi_j - y_j) = a - 1$  i.e.  $a = 1 + 1/\eta$ . One can easily check that  $\pi' \in \mathbb{P}_N$ . Now select any distribution  $x'$  over  $\hat{\mathcal{V}}$  such that  $Cx' = \pi'$  and let  $x^{(1)} = (x + x'/\eta)/a$ ,  $x^{(2)} = (x^{(q)} + x'/\eta)/a$  which are properly normalized distributions. We then have by construction  $a(x^{(1)} - x^{(2)}) = x - x^{(q)}$ , and with  $Cx^{(1)} = Cx^{(2)} = \pi$ . Invariance (i) requires that  $CAx^{(1)} = CAx^{(2)} = \pi$ , which readily implies  $CA(x - x^{(q)}) = 0$ . ■

In some sense, we are proving that under the constraints (si) the dynamics on the lifted graph is unobservable from the graph of interest.

#### IV. SCENARIOS WHERE LIFTED CHAINS CAN ACHIEVE OPTIMAL MIXING TIME

##### A. Stochastic bridges

Consider two distributions  $y$  and  $y'$  over the nodes  $\mathcal{V}$  of a graph  $\mathcal{G}$  with diameter  $D$ . Then, under mild assumptions, there exists a time-varying Markov chain  $\{P(t)\}_{t=1}^D$  such that  $y' = P(D)P(D-1)\dots P(1)y$ , where all the  $P(t)$  satisfy the locality constraints imposed by  $\mathcal{G}$ , see e.g. [18], [21]. We call this  $\{P(t)\}_{t=1}^D$  a *stochastic bridge* from  $y$  to  $y'$ . While in general the solution is not unique and one is typically looking for the closest solution to a reference evolution in the pseudo-distance induced by the relative entropy, we are here only concerned with the existence of such a bridge. This is guaranteed if for example  $y$  is a delta distribution, namely  $y = e_i$  for some  $i$ , and  $y'$  has full support [21]. In the following, we shall join stochastic bridges from every singular initial distribution to the target  $\pi$  into a unique lift via the node-clock lift.

##### B. Lifts from stochastic bridges

The following basic lift constructions will be used in the proof of the main results.  $v^\dagger$  denotes the adjoint (line vector) of  $v$  (column vector). We will use the notation  $e_i$  to denote a vector with all elements zero except its  $i$ 'th element equal to one, whose dimension is determined by the context.

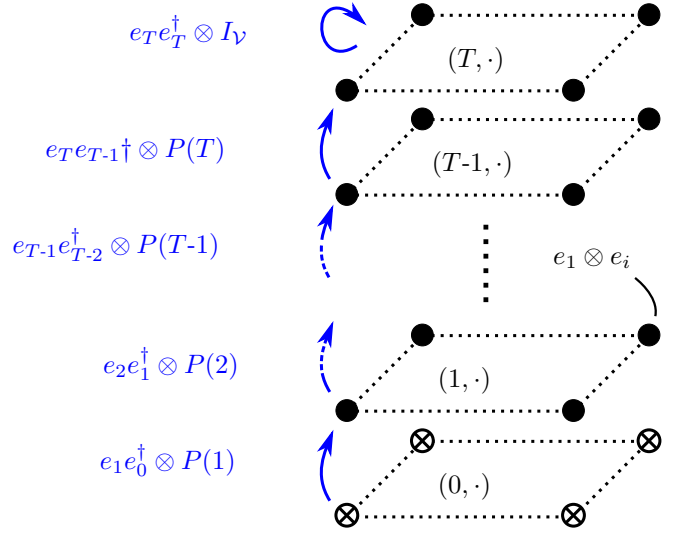


Fig. 2. A depiction of the “clock-lift” and its dynamics. The layers of nodes corresponds to copies of the reference graph, indexed by the time index. If  $x(0)$  is initialized over the nodes marked with a cross, with  $x \mapsto (0, x)$ , then  $y(t) = P(t)P(t-1)\dots P(1)y(0)$  for all  $t \leq T$ , as the effective transition matrices are those reported on the left.

1) *Clock lift*: This type of lift is used to convert a time-inhomogeneous chain into a time-homogeneous one. It is the discrete equivalent of adding time as a state variable in dynamical systems to make them formally time-invariant.

The following construction is depicted in Figure 2, and converts a Markov chain  $\{P(t)\}_{t \in \mathbb{N}}$  on a node space  $\mathcal{V}$ , i.e. where the state undergoes  $y(t) = P(t)y(t-1)$ , into a time-homogeneous lifted chain whose marginal evolution matches exactly the evolution of  $y(t)$  over a time interval  $[0, T]$ , by essentially adding a time-index to the original nodes. The lifted node space becomes  $\hat{\mathcal{V}} = \{(t, v) : t \in \{0, 1, \dots, T\} \text{ and } v \in \mathcal{V}\}$ , and the surjective map  $\mathfrak{c} : \hat{\mathcal{V}} \rightarrow \mathcal{V}$  is defined by  $\mathfrak{c}((t, v)) = v$ . We define

$$A = \sum_{t=1}^T e_t e_{t-1}^\dagger \otimes P(t) + e_T e_T^\dagger \otimes I_V,$$

where  $e_t$  is the canonical unit vector whose component  $t$  equals 1 and all other components are 0;  $\otimes$  is the Kronecker product, such that  $A \otimes B$  acts on vectors of  $\mathbb{R}^{(T+1)|\mathcal{V}|} = \mathbb{R}^{\hat{\mathcal{V}}}$  when  $A$  acts on  $\mathbb{R}^{(T+1)}$  and  $B$  on  $\mathbb{R}^{|\mathcal{V}|}$ ; and  $I_V$  is the identity on  $\mathbb{R}^{|\mathcal{V}|}$ . The additional term with the identity ensures that once we reach time  $T$  the evolution becomes trivial, and in particular if we reach the desired  $\pi$  we remain there. This  $A$  should be associated to the specific initialization (see aspect (S))

$$F : y(0) \mapsto x(0) = e_0 \otimes y(0).$$

The full state evolution then obeys

$$x(t) = A^t x(0) = e_t \otimes P(t)P(t-1)\dots P(1)y(0),$$

so that  $Cx(t) = P(t)P(t-1)\dots P(1)y(0) = y(t)$  for  $0 \leq t \leq T$  and  $Cx(t) = y(T)$  for  $t \geq T$ .



2) *Node-clock lift*: To any given lifted Markov chain  $x(t+1) = Ax(t)$ , we can associate a non-Markov process  $y(t+1) = P^{(x)}y(t)$ , where  $P^{(x)}$  depends on some hidden variable  $x$ . Conversely, consider that separately for each initial state  $y(0) = e_i$ , we have built stochastic evolutions  $y^{(i)}(t) = P^{(i)}(t)P^{(i)}(t-1)\dots P^{(i)}(1)e_i$ , with all  $P^{(i)}(k)$  satisfying the locality constraints of  $\mathcal{G}$ . We would like essentially that an initial distribution  $y(0)$  over the nodes implies that the system follows *the whole trajectory of the particular chain*  $y^{(i)}(t)$  with a probability  $y_i(0)$ .

We can capture this process over any bounded time-frame  $[0, T]$  using a lifted Markov chain which we call a *node-clock-lift*, corresponding to a clock-lift that is further augmented with an additional node-index — see figure 3. The lifted node space becomes  $\hat{\mathcal{V}} = \{(t, v_0, v) : t \in \{0, 1, \dots, T\} \text{ and both } v_0, v \in \mathcal{V}\} \cup \{(t+1, v, v) : v \in \mathcal{V}\}$ , the surjective map  $c : \hat{\mathcal{V}} \rightarrow \mathcal{V}$  is defined by  $c((t, j, i)) = i$ . We define

$$A = \sum_{t=1}^T \sum_{i \in \mathcal{V}} e_t e_{t-1}^\dagger \otimes e_i e_i^\dagger \otimes P^{(i)}(t) + e_{T+1} e_T^\dagger \otimes \Pi^{(\mathcal{V} \leftarrow \mathcal{V}^2)} + e_{T+1} e_{T+1}^\dagger \otimes I_{\mathcal{V}} \quad (5)$$

with the same notation as in the previous paragraph and an initial state  $x(0) = e_0 \otimes \sum_{i \in \mathcal{V}} y_i(0) e_i \otimes e_i$ . In the last term,  $\Pi^{(\mathcal{V} \leftarrow \mathcal{V}^2)}$  denotes the “projection” such that  $\Pi_{(v), (v_0, v)}^{(\mathcal{V} \leftarrow \mathcal{V}^2)} = 1$  for all  $v, v_0 \in \mathcal{V}$ . This means, after all the different initial states have followed the  $P^{(i)}(k)$  for  $T$  steps, the resulting contributions are summed up in a single copy of  $\mathcal{V}$ .

Note that in fact some of the lifted nodes will never be populated (e.g.  $(0, i, j) \in \hat{\mathcal{V}}$  with  $i \neq j$ ), so in fact  $\hat{\mathcal{V}}$  can be slightly reduced modulo a somewhat less compact description.

After this preparatory work, we are now ready to readily prove the main result of the section.

**Theorem 2:** Under constraints (SI) a lifted Markov chain can be constructed, such that  $\tau_M(1/4) < D_G + 1$ , with  $D_G$  the graph diameter; the associated lifted graph has of order  $D_G N^2$  nodes.

*Proof:* Given any  $\mathcal{G}$  and  $\pi$ , for each node  $i \in \mathcal{V}$  we can build the stochastic bridge from a  $y(0)$  concentrated on  $i$ , towards the target  $y(D_G) = \pi$ . We can then combine these bridges via a node-clock-lift (see Section IV-B) into a single lifted Markov chain which, when initialized (S) with  $F_{(s=0, v_0=i, v=i), i} = 1 \forall i \in \mathcal{V}$  and all other  $F_{i,j} = 0$ , converges exactly ( $\epsilon = 0$ ) to  $\pi$  in  $D_G$  time steps (and stays there in the additional  $D_G + 1$  step). ■

*Remark:* Note that also the lifted dynamics, for  $x(t)$ , converges in  $D_G$  steps from any initialized  $x(0)$  towards exactly  $x(D_G + 1) = e_{D_G+1} \otimes \pi$ .

## V. CONCLUSIONS AND OUTLOOK

In this paper, we argue that the role of lifted Markov chains, in regard to obtaining fast mixing dynamics on graphs, highly depends on the capability of properly initializing the lifted chain depending on the random initial state of

the reference graph, as well as the invariance requirements on the target marginal distribution. With respect to these two properties, we highlight two “extreme” cases, the (si) scenario, where extending the state space does not yield any advantage over a time-homogeneous Markov transition on the reference graph, and the (SI) scenario, where diameter-time mixing is possible. While the first result may seem in apparent contradiction with those of [10], their results *do not require invariance* of the marginal.

Of course, a number of interesting situations remain to be explored in between these two extremes. By modifying the lift structure, techniques based on stochastic bridges can still be used to address the situations where the invariance requirement is tightened, namely Si. The performance of the complementary case sl, as well as some refinements that take into account irreducibility of the lifted dynamics and constraints on the ergodic flows could be studied using *conductance-type bounds* [2], and it will be the object of future work.

Distinguishing these various scenarios becomes important when comparing the power of lifted Markov chains to other algorithmic acceleration techniques. We are particularly motivated by comparing the ultimate mixing performance of classical versus quantum Markov chains [22]: while in the relevant literature the latter are typically compared to Markov chains on the reference graph, it appears that a fairer comparison should consider lifted Markov chains. Towards this, we need to characterize lifted chains that correspond to the typical quantum chain through which acceleration is obtained. In fact, the quantum evolution of interest typically include an auxiliary (coin) system, that effectively acts as an extension of the original state space and introduces memory effects. The present work represents the basis to establish a framework towards such comparisons.

Lastly, stochastic bridges can also be used to construct lifted Markov chains which (at least on some marginal) match the behavior of other algorithms or stochastic dynamics. Thus, the mixing properties of lifted chain detailed here, and adaptations thereof, would enable indirect proofs for the achievable mixing performance of the algorithm they simulate.

## REFERENCES

- [1] D. Aldous and J. Fill, “Reversible markov chains and random walks on graphs,” 2002.
- [2] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*. American Mathematical Soc., 2009.
- [3] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *The journal of chemical physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [4] W. K. Hastings, “Monte carlo sampling methods using markov chains and their applications,” *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
- [5] F. Martinelli, “Lectures on glauher dynamics for discrete spin models,” in *Lectures on probability theory and statistics*. Springer, 1999, pp. 93–191.
- [6] M. Dyer, A. Frieze, and R. Kannan, “A random polynomial-time algorithm for approximating the volume of convex bodies,” *Journal of the ACM (JACM)*, vol. 38, no. 1, pp. 1–17, 1991.
- [7] M. Jerrum, A. Sinclair, and E. Vigoda, “A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries,” *Journal of the ACM (JACM)*, vol. 51, no. 4, pp. 671–697, 2004.

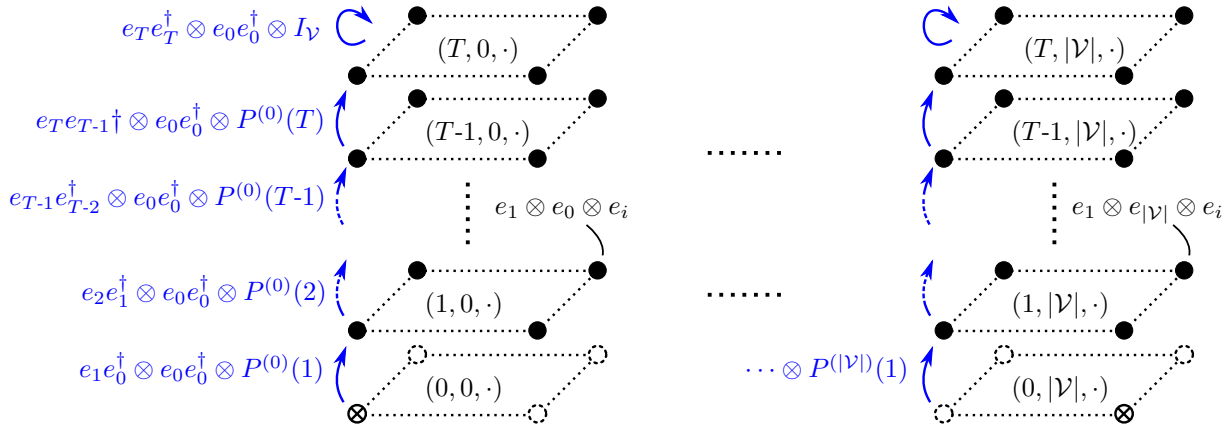


Fig. 3. A depiction of the idea underlying the “node-clock-lift”. A different time-dependent stochastic bridge is enacted depending on the initial state on the reference graph  $y(0) = e_i$ , and each is made time-homogeneous by a clock lift.

- [8] P. Diaconis, “The cutoff phenomenon in finite markov chains,” *Proceedings of the National Academy of Sciences*, vol. 93, no. 4, pp. 1659–1664, 1996.
- [9] P. Diaconis, S. Holmes, and R. M. Neal, “Analysis of a nonreversible markov chain sampler,” *Annals of Applied Probability*, pp. 726–752, 2000.
- [10] F. Chen, L. Lovász, and I. Pak, “Lifting markov chains to speed up mixing,” in *Proceedings of the thirty-first annual ACM symposium on Theory of computing*. ACM, 1999, pp. 275–281.
- [11] P. Diaconis and L. Miclo, “On the spectral analysis of second-order markov chains,” *Ann. Fac. Sci. Toulouse Math.*, vol. 22, no. 3, pp. 573–621, 2013.
- [12] M. Vucelja, “Lifting a nonreversible markov chain monte carlo algorithm,” *American Journal of Physics*, vol. 84, no. 12, pp. 958–968, 2016.
- [13] N. Alon, I. Benjamini, E. Lubetzky, and S. Sodin, “Non-backtracking random walks mix faster,” *Communications in Contemporary Mathematics*, vol. 9, no. 04, pp. 585–603, 2007.
- [14] K. Jung, D. Shah, and J. Shin, “Distributed averaging via lifted markov chains,” *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 634–647, 2010.
- [15] K. S. Turitsyn, M. Chertkov, and M. Vucelja, “Irreversible monte carlo algorithms for efficient sampling,” *Physica D: Nonlinear Phenomena*, vol. 240, no. 4, pp. 410–414, 2011.
- [16] H. C. Fernandes and M. Weigel, “Non-reversible monte carlo simulations of spin models,” *Computer Physics Communications*, vol. 182, no. 9, pp. 1856–1859, 2011.
- [17] Y. Sakai and K. Hukushima, “Eigenvalue analysis of an irreversible random walk with skew detailed balance conditions,” *Physical Review E*, vol. 93, no. 4, p. 043318, 2016.
- [18] M. Pavon and F. Ticozzi, “Discrete-time classical and quantum markovian evolutions: Maximum entropy problems on path space,” *Journal of Mathematical Physics*, vol. 51, no. 4, p. 042104, 2010.
- [19] B. Gerencsér, “Markov chain mixing time on cycles,” *Stochastic Processes and their Applications*, vol. 121, no. 11, pp. 2553–2570, 2011.
- [20] L. Rabiner and B. Juang, “An introduction to hidden markov models,” *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4–16, 1986.
- [21] T. T. Georgiou and M. Pavon, “Positive contraction mappings for classical and quantum schrödinger systems,” *Journal of Mathematical Physics*, vol. 56, no. 3, p. 033301, 2015.
- [22] J. Kempe, “Quantum random walks: an introductory overview,” *Contemporary Physics*, vol. 44, no. 4, pp. 307–327, 2003.