Bounding spectral gaps of Markov chains: a novel exact multi-decomposition technique

2003 J. Phys. A: Math. Gen. 36 3647

(http://iopscience.iop.org/0305-4470/36/13/301)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **36** (2003) 3647–3670 PII: S0305-4470(03)56214-X

Bounding spectral gaps of Markov chains: a novel exact multi-decomposition technique

N Destainville

Laboratoire de Physique Théorique—IRSAMC, CNRS/Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex 04, France

Received 15 November 2002, in final form 6 February 2003
Published 19 March 2003
Online at stacks.iop.org/JPhysA/36/3647

Abstract

We propose an exact technique to calculate lower bounds of spectral gaps of discrete time reversible Markov chains on finite state sets. Spectral gaps are a common tool for evaluating convergence rates of Markov chains. As an illustration, we successfully use this technique to evaluate the 'absorption time' of the 'Backgammon model', a paradigmatic model for glassy dynamics. We also discuss the application of this technique to the 'contingency table problem', a notoriously difficult problem from probability theory. The interest of this technique is that it connects spectral gaps, which are quantities related to dynamics, with static quantities, calculated at equilibrium.

PACS numbers: 02.50.Ga, 05.40.−a, 05.10.Ln

1. Introduction

Markov chains [1–4] have applications in many areas, ranging from pure probability theory to theoretical or numerical statistical physics. For example, in out-of-equilibrium statistical physics, they are commonly used to write in a formalized form the time evolution of physical models [5]. We study in this paper—as a particular application of our technique—the 'Backgammon model' [6] which is a well-known paradigmatic mean-field model for glassy dynamics. Markov chains are also widespread in numerical statistical physics: Monte Carlo algorithms [5] with Metropolis or Glauber dynamics, which are examples of Markov processes, are of great importance in the computational study of complex systems. The efficiency of such algorithms relies on their rates of convergence towards their equilibrium distributions. The ergodic times must be as short as possible to save computational time. Usually, such algorithms are said to be *rapid* if their ergodic times are polynomial in the system sizes whereas (generally speaking) the numbers of configurations grow exponentially with system sizes. Monte Carlo algorithms are also more and more widespread in applied mathematics (see [4]): the 'contingency table problem' [7, 8] that we examine at the end of the present paper will provide an example.

Roughly speaking, a Markov $\mathcal{M}$ process needs a characteristic time τ to be close to equilibrium (the equilibriation time). If P denotes the transition matrix (defined below) associated with $\mathcal{M}$, if $g(P)$ is the spectral gap of P, that is to say the difference between the two largest eigenvalues of P, then $\tau \sim 1/g(P)$. We explain this point and give references below. Therefore if one calculates a lower bound on $g(P)$, one also gets an upper bound on τ. More precisely, if $1/g(P)$ is polynomial in the system size, then τ is also polynomial and the chain is rapid.

Many efficient techniques have been developed to estimate (or to bound) mixing times and spectral gaps, when elementary techniques from linear algebra do not apply. Among many others, we mention the 'coupling' technique (see [9]), the 'canonical path' and 'conductance' arguments (see [3, 4]), or the 'chain decomposition' method [10]. This last method relies on the following basic idea: (i) decompose the state space into disjoint smaller pieces; (ii) prove that the dynamics is rapid on each piece considered in isolation; (iii) re-compose the dynamics on the whole state space from that on pieces and prove it is rapid in its turn.

We insist on this method because our 'multi-decomposition' technique is an extension of the previous one. More precisely, the idea of the chain decomposition method is schematically the following: suppose that the state space Ω of the Markov chain can be 'naturally' decomposed into several disjoint subsets $\Omega_a, a = 1, \ldots, A$; then 'cut' all the transitions between different subsets Ω_a and only keep the transitions inside each subset. The so-obtained restricted Markov chain on each Ω_a is denoted by $\mathcal{M}_a$. Suppose then that one can prove that each $\mathcal{M}_a$ is rapid on Ω_a, and that in addition the original dynamics is rapid 'between' subsets. Then, under certain conditions, the original dynamics on Ω will also be rapid. Unfortunately, even if one can easily prove that the dynamics is rapid on each Ω_a, it can be extremely difficult or even impossible to handle the second step of the proof. In [11], an idea which bypasses this difficulty was presented in the case of random rhombus tilings. In the present paper, we generalize this idea in a formalized form and show that it has a much vaster domain of application than random tilings.

As compared to the previous 'simple' decomposition method, the present 'multi-decomposition' scheme uses several (m) decompositions of Ω, namely $(\Omega_{1;a_1})$, $(\Omega_{2;a_2}), \ldots, (\Omega_{m;a_m})$. For each decomposition indexed by $k = 1, \ldots, m$, the set Ω is also decomposed in disjoint subsets $\Omega_{k;a}$. As in the decomposition method, one first needs to prove that the dynamics is rapid in each subset $\Omega_{k;a}$, for each k and each a. Then if any two decompositions 'overlap sufficiently', in a sense that will be precised below, the dynamics will also be rapid on the whole set Ω. In other words, given a decomposition $(\Omega_{k;a})$, one does not have to prove any more that the original dynamics is rapid between subsets of this decomposition. This point is ensured instead by the remaining subsets $\Omega_{l;b}, l \neq k$. These ideas are schematically exemplified in figure 1.

We illustrate this delicate point with a simple example that will also be developed in greater detail below: consider a two-dimensional random walker on a square grid $p_1 \times p_2$ (see figure 2). At each step, the walker performs either a vertical or a horizontal move, with equal probability $1/4$ in each direction (we neglect the question of boundaries at this level of discussion). Then in our formalism, we define two decompositions $(\Omega_{1;a})$ and $(\Omega_{2;b})$, respectively a vertical and a horizontal one. In the vertical decomposition, each subset $\Omega_{1;a}, a = 1, \ldots, p_1$ is a vertical segment, with no possible transitions towards the neighbouring vertical segments: the horizontal edges are 'cut' in this vertical decomposition. Symmetrically, vertical edges are 'cut' in the horizontal decomposition, and the subsets $\Omega_{2;b}, b = 1, \ldots, p_2$ are horizontal segments. Now on each subset $\Omega_{1;a}$ or $\Omega_{2;b}$, the induced dynamics on vertical or horizontal segments is nothing but that of a one-dimensional random walker which stays at the same place with probability $1/2$ and moves in either direction with probability $1/4$. For such a

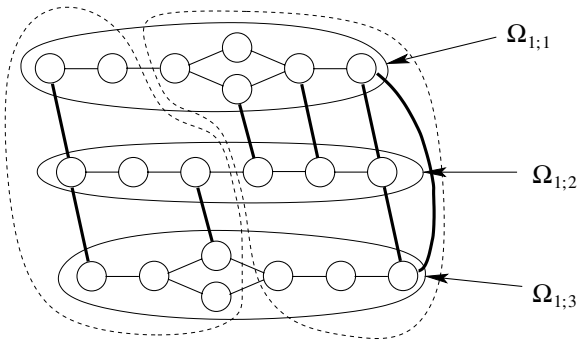

Figure 1. Schematic representation of a multi-decomposition in the case of two decompositions. There are 20 possible states in Ω represented by small circles, and the edges between these circles represent possible transitions between states. The first decomposition contains three subsets $\Omega_{1;a}$ (full lines) and the second one, two subsets $\Omega_{2;b}$ (dashed lines). In the first decomposition, we have emphasized (thick lines) the edges that are cut because the transitions between states of different subsets $\Omega_{1;a}$ are forbidden. If one can prove that the dynamics is rapid in each subset of each decomposition then, under certain conditions, one can conclude that the dynamics is rapid on the whole state set Ω.

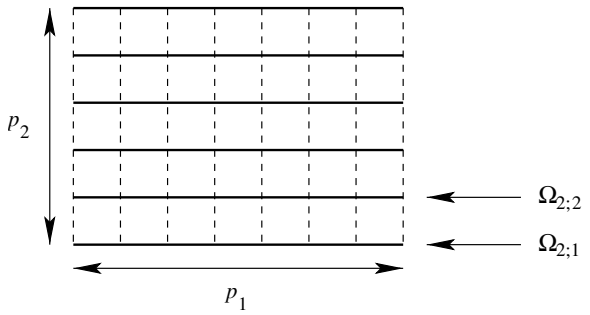

Figure 2. A 2D random walker on a square grid $p_1 \times p_2$. At time t, the walker is situated on a vertex. At time $t + 1$, he chooses with the same probability to move on a nearest neighbour, horizontally (full lines) or vertically (dashed lines), if he can. There are two natural decompositions of this Markov chain, a vertical and a horizontal one. In the horizontal decomposition, for example, the walker can only move horizontally on subsets $\Omega_{2;b}$. Vertical edges are 'cut'.

random walker on segments of lengths p_k, the spectral gaps are $g_k \simeq Cst/p_k^2$: the dynamics are rapid on each subset of each decomposition. Then the method developed in this paper can be applied to this problem. The underlying idea is that any trajectory of the walkers is a succession of vertical and horizontal moves. Therefore the dynamics on Ω is a combination of the dynamics on the different subsets. If the dynamics are rapid on each subset, it will certainly be rapid on Ω. More formally, one proves *via* the present technique that

$$g(P) \geqslant \inf(g_1, g_2) \simeq \frac{Cst}{\sup\left(p_1^2, p_2^2\right)}. \tag{1}$$

In other words, the dynamics on the whole set Ω are faster than the dynamics on the slowest subset. In this case, this inequality can be checked by independent means and appears to be

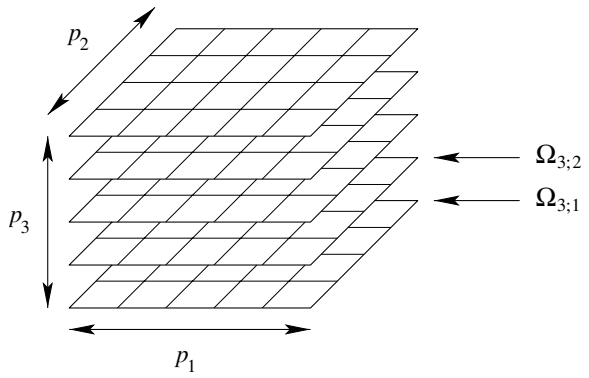

Figure 3. A decomposition (among three possible ones) in the case of a random walker on a three-dimensional cubic grid of sides $p_1 \times p_2 \times p_3$. The vertical moves, which are allowed for the three-dimensional walker, are now forbidden in this decomposition. The walker is constrained to move on one of the p_3 two-dimensional grids of sides $p_1 \times p_2$, denoted by $\Omega_{3;a}$.

an equality. In the present paper, we rigorously formalize these ideas in a more general point of view.

The interest of this method is that it can be applied inductively. This idea will be used throughout the paper. In the previous example, one can iterate the multi-decomposition scheme in the case of random walkers on larger and larger m-dimensional hyper-cubic grids of side lengths $p_1 \times \cdots \times p_m$. At each stage m, there are m decompositions, one for each dimension of space, and the subsets are $(m-1)$-dimensional hyper-cubic grids in their turn (see figure 3). Then one proves by induction on m that

$$g\big(P^{(m)}\big) \geqslant \inf_{k=1,\ldots,m} g\big(P_k^{(m-1)}\big) \simeq \frac{Cst}{\sup_k \big(p_k^2\big)} \tag{2}$$

where $P^{(q)}$ is the transition matrix of a q-dimensional random walker. Once again, this lower bound is in fact exact.

Another interest of the method is at the computational level: in the cases where analytical calculations cannot be completed to their end, the method can provide numerical estimates of the spectral gaps with much higher efficiency than direct diagonalizations of transition matrices because it requires to diagonalize much smaller matrices. This point is exemplified in section 6.

The organization of this paper is as follows: section 2 introduces the background definitions and notations used throughout the paper. In particular, the notion of discrete time reversible Markov chain on a finite state set is briefly recalled. Section 3 presents the specific definitions and the main results of the paper, which are applied to several pedagogical elementary examples in section 4, namely random walkers on hyper-cubic lattices and on the symmetric group S_n.

In section 5, we successfully apply the present technique to two 'urn models', the 'Backgammon model' and the 'Monkey urn model'. In both cases, we relate the dynamics on m urns to the dynamics on two urns. We show that the source of the slow dynamics of the Backgammon model is entirely contained in the two-urn problem. In other words, considering many urns instead of two does not bring any additional slowness [12, 13]. At last, the contingency table problem is analysed in section 6. The last section 7 is devoted to conclusions.

2. Generalities

2.1. Transition matrix P of a Markov chain $\mathcal{M}$

The present paper deals with discrete time reversible Markov chains on finite state sets. Let Ω be a finite set of cardinality N, $\mathcal{M}$ a Markov process on Ω and P its transition matrix of transition probabilities $P(x, y)$: $P(x, y) \geqslant 0$ is the conditional probability that the chain is in the state x at time $t + 1$ given that it was in the state y at time t. Note that these transition probabilities themselves do not depend on time. Probabilities are conserved so that $\sum_x P(x, y) = 1$ for all y: P is *stochastic*.

We define the state vector $\boldsymbol{e}(t)$ at time t: it has coordinates $(e(t, x))_{x\in\Omega}$ where $e(t, x)$ is the probability that the chain is in the state x at time t. Then by definition of the matrix P,

$$\boldsymbol{e}(t + 1) = P\boldsymbol{e}(t). \tag{3}$$

A state vector $\boldsymbol{\pi}$ is said to be an equilibrium (or stationary) distribution if $\boldsymbol{\pi} = P\boldsymbol{\pi}$, in other words if $\boldsymbol{\pi}$ is an eigenstate of P with eigenvalue 1.

The Markov process $\mathcal{M}$ is said to be reversible if it satisfies the detailed balance condition [1, 5]

$$\pi(x)P(y, x) = \pi(y)P(x, y) \tag{4}$$

for all states x and y. This condition ensures that $\boldsymbol{\pi}$ is effectively an equilibrium distribution for $\mathcal{M}$. It will be useful in the following. Furthermore, we suppose that this equilibrium distribution $\boldsymbol{\pi}$ exists and is unique (see for instance [1] for further detail on this point). In addition, we assume in the sequel that the 'loop' transitions $P(x, x)$ are at least $1/2$ for all states x.

2.2. Eigenvalues and spectral gap of P

We basically assume for the moment that $\pi(x) > 0$ for all states x: the chain can reach any state at equilibrium. Then we define the scalar product

$$\langle \boldsymbol{f}|\boldsymbol{g}\rangle \equiv \sum_{x\in\Omega} \frac{1}{\pi(x)} f(x)g(x). \tag{5}$$

Because of reversibility, P is self-adjoint for this scalar product:

$$\langle P\boldsymbol{f}|\boldsymbol{g}\rangle = \sum_x \frac{1}{\pi(x)} \sum_y P(x, y)f(y)g(x) \tag{6}$$

$$= \sum_y \frac{1}{\pi(y)} \sum_x P(y, x)g(x)f(y) \quad \text{(see equation (4))} \tag{7}$$

$$= \langle \boldsymbol{f}|P\boldsymbol{g}\rangle. \tag{8}$$

Thus the eigenvalues $\beta_i, i = 0, \ldots, N - 1$, of P are real. Moreover, $|\beta_i| \leqslant 1$ for all i by standard Perron–Frobenius theory [14]. Since the equilibrium distribution exists and is unique, $\beta_0 = 1$ and

$$1 > \beta_1 \geqslant \beta_2 \geqslant \cdots \geqslant \beta_{N-1} \geqslant -1. \tag{9}$$

In addition, we have supposed that $P(x, x) \geqslant 1/2$ for all x. As a consequence, $\tilde{P} = 2P - \mathbb{1}$ is also a transition matrix with non-negative transition probabilities, the eigenvalues $2\beta_i - 1$ of which are larger than -1 according to (9). Hence $\beta_i \geqslant 0$ for all i. Now we can define the *spectral gap* of P:

$$g(P) = 1 - \beta_1 > 0. \tag{10}$$

2.3. Spectral gap and equilibriation time τ

In the introduction, we claimed that the characteristic time τ to be close to equilibrium is of order $1/g(P)$, whatever the initial probability distribution $e(0)$. Let us precise this point and in particular what we mean by 'close'. If the chain was in the state $x \in \Omega$ at time 0, we denote by $e_x(t)$ the state vector at time t.

In the literature, the distance between the probability distribution $e_x(t)$ and the equilibrium one π can be measured by various means. Among them, two norms are commonly used, the Euclidean norm associated with the above scalar product and the 1-norm or *variation distance* [9]:

$$\Delta_x(t) = \frac{1}{2}\sum_{y\in\Omega} |e_x(t, y) - \pi(y)|. \tag{11}$$

Note that $\Delta_x(t) = \sup_{A\in\Omega} |p(A, t) - \pi(A)|$ where $p(A, t)$ (resp. $\pi(A)$) is the probability that the system is in the subset A of Ω at time t (resp. at equilibrium). Then given $\varepsilon > 0$, we define the *mixing* time $\tau_{\rm var}(\varepsilon)$ associated with this distance:

$$\tau_{\rm var}(\varepsilon) = \max_x \min_{t_0} \{t_0/\forall t \geqslant t_0, \Delta_x(t) \leqslant \varepsilon\}. \tag{12}$$

In other words, whatever the initial state x, one is sure to stay within distance ε of equilibrium after $\tau_{\rm var}(\varepsilon)$ steps. Then elementary algebra (see e.g. [16]) shows that

$$\tau_{\rm var}(\varepsilon) \leqslant \left(\ln\left(\frac{1}{\varepsilon}\right) + \max_x \ln\left(\frac{1}{\pi(x)}\right)\right)\frac{1}{g(P)}. \tag{13}$$

This upper bound is of order $1/g(P)$ even though the second term of the prefactor of $1/g(P)$ can be substantially large. However, this second term is usually polynomial in the system size, at least at high temperature, where the distribution π is uniform: this second term is simply the entropy of the system and if it is extensive, it is polynomial in the system size. If $1/g(P)$ is also polynomial, the dynamics remain rapid. The reader can refer to Wilson [15] for an interesting discussion on this point.

In contrast, at vanishing temperature, only the low energy levels are occupied and some probabilities $\pi(x)$ tend to zero and this second term diverges, even for finite systems. This is not physically relevant, since $\tau_{\rm var}$ remains finite in general. The previous upper bound (13) is not adapted.

Should we have chosen the Euclidean norm to measure the distance under interest, we would have been led to the similar conclusion that the equilibriation time is of order $1/g(P)$ (see also [15] for further discussion). In the following sections, we shall focus on spectral gaps rather than such equilibriation times.

3. Main result

3.1. Definitions

Generally speaking, given a discrete time Markov chain $\mathcal{M}$ on a finite set Ω, a decomposition [10] of $\mathcal{M}$ is both a partition of Ω in disjoint subsets $\Omega_a, a = 1, \ldots, A$, and a new natural dynamical rule on Ω. This dynamical rule is defined by its transition matrix P' as follows: for all $x \neq y \in \Omega$, if x and y belong to the same subset Ω_a then $P'(x, y) = P(x, y)$, else $P'(x, y) = 0$. We say that we 'cut transitions between the different subsets Ω_a'. The diagonal terms $P'(x, x)$ are suitably modified to ensure that P' remains stochastic. If M_a denotes the restriction of P' to Ω_a, then P' is block-diagonal and it is the direct sum of the M_a:

$$P' = M_1 \oplus M_2 \oplus \cdots \oplus M_A. \tag{14}$$

We naturally suppose that the decomposition is designed so that each subset Ω_a is connected with respect to the new dynamical rule and that for each Ω_a seen in isolation, there also exists a unique stationary distribution. Then the equilibrium state of P' is degenerate of dimension A. We call P' the *restriction* of P to the decomposition (Ω_a).

As was announced in the introduction, we shall consider the case where we have m different decompositions $\left(\Omega_{1;a_1}\right), \left(\Omega_{2;a_2}\right), \ldots, \left(\Omega_{m;a_m}\right)$, $a_k = 1, \ldots, A_k$. *A priori*, the different A_k need not be equal. Then we define as above m different restrictions for each decomposition $(\Omega_{k;a})$, denoted by P_k. We say that we have a multi-decomposition or a m-decomposition.

Now we need to introduce a property of regularity. We say that a transition of the original chain $P(x, y) > 0$ 'belongs' to the decomposition $(\Omega_{k;a})$ if $P_k(x, y) > 0$, in other words if this transition is not cut in that decomposition.

Definition. *A m-decomposition $\left(\Omega_{1;a_1}\right), \left(\Omega_{2;a_2}\right), \ldots, \left(\Omega_{m;a_m}\right)$ of $\mathcal{M}$ is said to be regular if there exists an integer $r \leqslant m$ such that each transition $P(x, y) > 0$ belongs to exactly r decompositions. The integer r is called the redundancy of the regular multi-decomposition.*

If the m-decomposition is regular of redundancy r then we have the following simple but important relation:

$$P = \frac{P_1 + P_2 + \cdots + P_m}{r} + \frac{r - m}{r}\mathbb{1} \tag{15}$$

which interconnects the different decompositions [11]. The second term of the sum restores the diagonal coefficients $P(x, x)$.

Note that the restriction of $\mathcal{M}$ to each subset $\Omega_{k;a}$ of each decomposition remains reversible, because if $P(x, y)$ is cut, then $P(x, y)$ is also cut:

$$\pi(x)P_k(y, x) = \pi(y)P_k(x, y). \tag{16}$$

In addition, π remains a stationary distribution for each P_k, but it is not unique; we denote by E_k the eigenspace of eigenvectors of P_k corresponding to the eigenvalue 1. Its dimension is A_k because we also suppose that on each subset $\Omega_{k;a}$ there is a unique stationary distribution. Equation (16) also leads us to the conclusion that P_k remains self-adjoint for the scalar product (5). If we denote by Π_k the orthogonal projection (with respect to this scalar product) onto E_k, then any eigenvector (other than vectors of E_k) projects onto $\mathbf{0}$. We define the self-adjoint *multi-projection*:

$$\Pi = \Pi_1 + \Pi_2 + \cdots + \Pi_m. \tag{17}$$

Then $\Pi(\pi) = m\pi$. We say that the multi-decomposition is *non-degenerate* if this eigenvector is non-degenerate itself, which is satisfied in practice. We denote by ν the second largest eigenvalue of Π (in modulus) and call it the *norm of the multi-projection* Π. In the following, the difficult step will often be to calculate this norm. We shall discuss this point in subsection 3.3 and give several examples in sections 4–6.

Definition. *A non-degenerate m-decomposition, which is regular of redundancy r and which has an associated multi-projection of norm ν, is called a non-degenerate (m, r, ν)-multi-decomposition.*

3.2. Main theorem

The eigenspace E_k of P_k corresponding to the eigenvalue 1 is degenerate: we denote by $g(P_k)$ the difference between 1 and the second largest eigenvalue of P_k (strictly smaller than 1).

Theorem 1. *Let P be the transition matrix of a Markov chain $\mathcal{M}$ on Ω and $\left(\Omega_{1;a_1}\right)$, $\left(\Omega_{2;a_2}\right), \ldots, \left(\Omega_{m;a_m}\right)$ be a non-degenerate (m, r, ν)-multi-decomposition of $\mathcal{M}$. If P_k denotes the restriction of P to the decomposition $\left(\Omega_{k;a_k}\right)$, then the following gap relation holds:*

$$g(P) \geqslant \frac{m-\nu}{r} \inf_k g(P_k). \tag{18}$$

Proof. For the sake of convenience, we work in this proof with the matrix

$$\tilde{P} = \frac{P_1 + P_2 + \cdots + P_m}{m} = \frac{r}{m} P + \frac{m-r}{m} \mathbb{1}. \tag{19}$$

At the end of the calculation, the gap of P will simply be $g(P) = m/r\, g(\tilde{P})$.

Given a state vector $\boldsymbol{e}$, we note $\boldsymbol{e}_k^{\mathrm{eq}} = \Pi_k(\boldsymbol{e})$ (the superscript 'eq' stands for 'equilibrium'); $\boldsymbol{e}_k^{\mathrm{eq}}$ depends on k and $\boldsymbol{e}$, since the eigenspace of P_k corresponding to the eigenvalue 1 is degenerate. All the difficulty in the following lies in this degeneracy.

We suppose now that we have sorted altogether all the eigenvalues of the m matrices P_k: $1 > \mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_q \geqslant 0$. We denote by $\boldsymbol{f}_j$ the normalized eigenstate corresponding to the eigenvalue μ_j; $\boldsymbol{f}_j$ can *a priori* be the eigenstate of *any* matrix P_k. The vector $\boldsymbol{e} - \boldsymbol{\pi}$ is orthogonal to $\boldsymbol{\pi}$ and we write

$$\begin{aligned} \boldsymbol{e} - \boldsymbol{\pi} &= \frac{1}{m} \sum_{k=1}^{m} \left(\boldsymbol{e} - \boldsymbol{e}_k^{\mathrm{eq}}\right) + \frac{1}{m} \sum_{k=1}^{m} \left(\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi}\right) \\ &= \frac{1}{m} \sum_{j=1}^{q} \alpha_j \boldsymbol{f}_j + \frac{1}{m} \sum_{k=1}^{m} \left(\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi}\right) \end{aligned} \tag{20}$$

where each vector $\left(\boldsymbol{e} - \boldsymbol{e}_k^{\mathrm{eq}}\right)$ has been projected on the eigenbasis of P_k, and

$$\begin{aligned} \tilde{P}(\boldsymbol{e} - \boldsymbol{\pi}) &= \frac{1}{m} \sum_{k=1}^{m} P_k (\boldsymbol{e} - \boldsymbol{\pi}) \\ &= \frac{1}{m} \sum_{k=1}^{m} P_k \left(\boldsymbol{e} - \boldsymbol{e}_k^{\mathrm{eq}}\right) + \frac{1}{m} \sum_{k=1}^{m} \left(\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi}\right) \\ &= \frac{1}{m} \sum_{j=1}^{q} \mu_j \alpha_j \boldsymbol{f}_j + \frac{1}{m} \sum_{k=1}^{m} \left(\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi}\right) \end{aligned} \tag{21}$$

by definition of $\boldsymbol{f}_j$ and μ_j. Now, thanks to a suitable Abel transform,

$$\begin{aligned} \tilde{P}(\boldsymbol{e} - \boldsymbol{\pi}) = \frac{1}{m}(1-\mu_1) \sum_{k=1}^{m} \left(\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi}\right) + \frac{1}{m} \sum_{r=1}^{q-1} (\mu_r - \mu_{r+1}) \\ \times \left(\sum_{k=1}^{m} \left(\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi}\right) + \sum_{j=1}^{r} \alpha_j \boldsymbol{f}_j \right) + \frac{1}{m} \mu_q \left(\sum_{k=1}^{m} \left(\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi}\right) + \sum_{j=1}^{q} \alpha_j \boldsymbol{f}_j \right). \end{aligned} \tag{22}$$

Thus, if $\|\cdot\|$ stands for the Euclidean norm associated with the scalar product (5),

$$\begin{aligned} \|\tilde{P}(\boldsymbol{e} - \boldsymbol{\pi})\| &\leqslant \frac{1}{m}(1-\mu_1) \left\| \sum_{k=1}^{m} \left(\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi}\right) \right\| + \sum_{r=1}^{q-1} (\mu_r - \mu_{r+1}) \|\boldsymbol{e} - \boldsymbol{\pi}\| + \mu_q \|\boldsymbol{e} - \boldsymbol{\pi}\| \\ &\leqslant \frac{1}{m}(1-\mu_1) \left\| \sum_{k=1}^{m} \left(\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi}\right) \right\| + \mu_1 \|\boldsymbol{e} - \boldsymbol{\pi}\|. \end{aligned} \tag{23}$$

Indeed, $\mu_k - \mu_{k+1} \geqslant 0$ and $\mu_q \geqslant 0$; moreover, for any $r \leqslant q$, $\sum_{k=1}^{m} \left(\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi}\right) + \sum_{j=1}^{r} \alpha_j \boldsymbol{f}_j$ is the sum of m orthogonal projections of $\boldsymbol{e} - \boldsymbol{\pi}$ on suitable spaces, the norm of each of them being therefore smaller than $\|\boldsymbol{e} - \boldsymbol{\pi}\|$. In addition, $\sum_k \left(\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi}\right) = \Pi(\boldsymbol{e} - \boldsymbol{\pi})$ and by definition of ν,

$$\left\| \sum_{k=1}^{m} \left(\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi}\right) \right\| \leqslant \nu \|\boldsymbol{e} - \boldsymbol{\pi}\| \tag{24}$$

because Π is self-adjoint. As a consequence,

$$\|\tilde{P}(\boldsymbol{e} - \boldsymbol{\pi})\| \leqslant \left[\frac{\nu}{m}(1 - \mu_1) + \mu_1\right] \|\boldsymbol{e} - \boldsymbol{\pi}\| \tag{25}$$

from which it follows that

$$g(\tilde{P}) \geqslant \frac{m - \nu}{m}(1 - \mu_1) \geqslant \frac{m - \nu}{m} \inf_k g(P_k) \tag{26}$$

because $\tilde{P}$ is self-adjoint. Thus

$$g(P) \geqslant \frac{m - \nu}{r} \inf_k g(P_k). \tag{27}$$

□

3.3. How to calculate the norm ν of the multi-projection

As mentioned above, the difficult step in the present multi-decomposition technique is generally to calculate the norm ν. In sections 4 and 5, we list several examples where this calculation is feasible exactly. The idea is to construct a matrix which has the same eigenvalues as Π, which is much smaller than Π and the elements of which depend only on the equilibrium distribution $\boldsymbol{\pi}$ and the subsets $\Omega_{k;a}$. This is an important point: ν is calculated *at equilibrium* and the spectral gap of P, that is its *dynamical* rate of convergence, is eventually monitored by this *equilibrium* quantity.

However, even in the cases where this calculation is out of reach, the present technique is also a great advance as compared to direct diagonalization of the matrix P to obtain $g(P)$, because in order to calculate ν numerically, one has to diagonalize a much smaller matrix than P. This point of view is used at the end of section 6 to support a conjecture about the 'contingency table problem'.

As was already remarked in section 3.1, given a decomposition, $(\Omega_{k;a})_{a=1,\ldots,A_k}$, since the different subsets have been 'disconnected', the vector space E_k of eigenstates corresponding to the eigenvalue 1 is degenerate of dimension A_k. In order to write the projections $\boldsymbol{e}_k^{\mathrm{eq}} - \boldsymbol{\pi} = \Pi_k(\boldsymbol{e} - \boldsymbol{\pi})$, which belong to E_k, we need a suitable basis of E_k. For each k and each a, we define the vector $\boldsymbol{g}_{k;a}$, the coordinates of which are equal to $g_{k;a}(x) = \pi(x)$ for any state $x \in \Omega_{k;a}$ and to 0 anywhere else. Note that

$$\|\boldsymbol{g}_{k;a}\| = \left(\sum_{x \in \Omega_{k;a}} \pi(x) \right)^{1/2}. \tag{28}$$

We set

$$\boldsymbol{u}_{k;a} = \frac{\boldsymbol{g}_{k;a}}{\|\boldsymbol{g}_{k;a}\|}. \tag{29}$$

Then $(\boldsymbol{u}_{k;a})_{a=1,\ldots,A_k}$ is an orthonormal basis of E_k and by definition of Π_k (section 3.1),

$$\Pi_k(\boldsymbol{f}) = \sum_{a=1}^{A_k} \langle \boldsymbol{f} | \boldsymbol{u}_{k;a} \rangle \boldsymbol{u}_{k;a} \tag{30}$$

for any vector $\boldsymbol{f}$. Hence

$$\Pi(\boldsymbol{f}) = \sum_k \Pi_k(\boldsymbol{f}) = \sum_{k=1}^{m} \sum_{a=1}^{A_k} \langle \boldsymbol{f} | \boldsymbol{u}_{k;a} \rangle \boldsymbol{u}_{k;a}. \tag{31}$$

Let us denote by U the matrix which has the different vectors $\boldsymbol{u}_{k;a}$ for all k and all a as column vectors; this matrix has N lines and $\sum_k A_k$ columns (the total number of subsets $\Omega_{k;a}$). We also denote by $\mathcal{P}$ the diagonal matrix of diagonal elements the $\pi(x)$, $x \in \Omega$. Then

$$\Pi(\boldsymbol{f}) = U U^t \mathcal{P}^{-1} \boldsymbol{f} \tag{32}$$

where U^t is the transpose of U with the vectors $\boldsymbol{u}_{k;a}$ as line vectors. We are interested in the eigenvalues of Π and therefore of $U U^t \mathcal{P}^{-1}$. Now given any two matrices A and B, a common theorem in basic linear algebra [14] says that AB and BA have the same non-zero eigenvalues, even if A and B are not square. As a consequence, Π has the same non-zero eigenvalues as $\bar{\Pi} = U^t \mathcal{P}^{-1} U$, the coefficients of which are

$$\bar{\Pi}_{(k;a),(l;b)} = \langle \boldsymbol{u}_{k,a} | \boldsymbol{u}_{l,b} \rangle = \frac{\sum_{x \in \Omega_{k;a} \cap \Omega_{l;b}} \pi(x)}{\left(\sum_{x \in \Omega_{k;a}} \pi(x) \right)^{1/2} \left(\sum_{x \in \Omega_{l;b}} \pi(x) \right)^{1/2}}. \tag{33}$$

This matrix is positive semi-definite; its eigenvalue is non-negative.

Theorem 2. *Let* $\left(\Omega_{1;a_1}\right), \left(\Omega_{2;a_2}\right), \ldots, \left(\Omega_{m;a_m}\right)$ *be a non-degenerate* (m, r, ν)*-multi-decomposition of a Markov chain* $\mathcal{M}$*, of stationary distribution* $\pi(x)$*, the norm of the associated multi-projection is the second largest eigenvalue of the matrix* $\bar{\Pi}$ *defined by equation (33).*

The dimension of the square matrix $\bar{\Pi}$ is $\sum_k A_k$. This quantity is in general much smaller than the dimension N of P. If the distribution $\boldsymbol{\pi}$ is uniform (for example at infinite temperature), then

$$\bar{\Pi}_{(k;a),(l;b)} = \frac{|\Omega_{k;a} \cap \Omega_{l;b}|}{|\Omega_{k;a}|^{1/2} |\Omega_{l;b}|^{1/2}} \tag{34}$$

where $|S|$ stands for the cardinality of any set S. One may also choose to work in a different basis where the above matrix may have a more practical expression: if $\mathcal{Q}$ is the diagonal matrix with diagonal coefficients $|\Omega_{k;a}|^{1/2}$, then $\hat{\Pi} = \mathcal{Q} \bar{\Pi} \mathcal{Q}^{-1}$ has the same eigenvalues as $\bar{\Pi}$ and Π. Its coefficients are

$$\hat{\Pi}_{(k;a),(l;b)} = \frac{|\Omega_{k;a} \cap \Omega_{l;b}|}{|\Omega_{l;b}|}. \tag{35}$$

The same kind of transformation can also be used in the case where $\boldsymbol{\pi}$ is not uniform:

$$\hat{\Pi}_{(k;a),(l;b)} = \frac{\sum_{x \in \Omega_{k;a} \cap \Omega_{l;b}} \pi(x)}{\sum_{x \in \Omega_{l;b}} \pi(x)}. \tag{36}$$

This is the conditional probability, at equilibrium, that the system is in the subset $\Omega_{k;a}$ given that it is in the subset $\Omega_{l;b}$.

4. Elementary examples

In this section, we display several examples where the multi-decomposition technique is particularly efficient. In these three examples, the lower bound provided by the method can be compared to exact values of the gap calculated by independent means. In these three cases, the lower bound of theorem 1 turns out to be an equality.

4.1. Random walk on a m-dimensional cube

We consider a random walker on a m-dimensional hyper-cube. We first derive exactly the spectral gap of this Markov chain. Then we prove that the lower bound provided by the multi-decomposition method is in fact an equality.

Our random walker moves on the vertices of a cube, $\Omega_m = \{0, 1\}^m$. He moves through the edges of the cube. His position is denoted by $(x_1, x_2, \ldots, x_m)$. At each step, he chooses randomly an index i among m as well as a direction $b = \pm 1$. If the move $x_i \to x_i + b$ is possible then he performs this move, else he stays at the same place. The allowed transition rates are all equal to $1/2m$. Then the transition matrix $P^{(m)}$ of this chain can be written as a tensor product of m one-dimensional walkers on $\{0, 1\}$:

$$P^{(m)} = \frac{1}{m} \sum_{k=1}^{m} (\mathbb{1} \otimes \cdots \otimes \mathbb{1} \otimes t_k \otimes \mathbb{1} \otimes \cdots \otimes \mathbb{1}) \tag{37}$$

because the walker can choose one of the m directions with equal probability $1/m$ and then he performs a one-dimensional walk in that direction. The transition matrix of each one-dimensional walker in the direction k is

$$t_k = t = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{38}$$

The eigenvalues of t are 1 and 0, and $g(P^{(m)}) = 1/m$.

We now use the multi-decomposition technique in order to compare the obtained lower bound with this exact value: we relate the spectral gap of $P^{(m)}$ to that of a random walker on a $(m-1)$-dimensional hyper-cube Ω_{m-1}. There are m natural ways to build the decompositions, by preventing moves in one (and only one) of the m directions of space. Then there are two subsets in each decomposition, and they are isomorphic to a hyper-cube Ω_{m-1}.

Furthermore, each transition belongs to exactly $m-1$ decompositions among m, more precisely to every decomposition save the one where this transition is forbidden. Therefore this multi-decomposition is regular of redundancy $m-1$.

Now we consider the multi-projection associated with this multi-decomposition. Let us prove with the help of relation (34) (and theorem 2) that its norm $\nu = 1$. In this case, $|\Omega_{k;a}| = 2^{m-1}$, the cardinality of a $(m-1)$-dimensional hyper-cube, for all $k = 1, \ldots, m$ and all $a = 1, 2$; and $|\Omega_{k;a} \cap \Omega_{l;b}| = 2^{m-2}$ whenever $k \neq l$. Hence $\bar{\Pi}_{(k;a),(l;b)} = 1/2$ if $k \neq l$ and $\bar{\Pi}$ is a block-matrix of 2×2 blocks. Diagonal blocks are equal to the identity and non-diagonal ones to t. It is an elementary exercise to compute the eigenvalues of such a matrix: m is the largest eigenvalue, 1 is m-fold degenerate, and the $m-1$ remaining eigenvalues are equal to 0. We conclude that $\nu = 1$. According to theorem 1, $g(P^{(m)}) \geqslant \inf_{k=1,\ldots,m} g(P_k)$. All matrices P_k have the same gap.

Now it is time to draw the attention of the reader to the following important point: the (allowed) non-diagonal transition rates in matrices $P^{(m)}$ and P_k are equal. Therefore P_k is the direct sum (see equation (14)) of two transition matrices of a $(m-1)$-dimensional random walker, but with non-diagonal transition rates $1/2m$ instead of $1/2(m-1)$ in $P^{(m-1)}$. As a consequence,

$$g(P_k) = \frac{m-1}{m} g(P^{(m-1)}) \tag{39}$$

and

$$g(P^{(m)}) \geqslant \frac{m-1}{m} g(P^{(m-1)}) \geqslant \cdots \geqslant \frac{1}{m} g(P^{(1)}) = \frac{1}{m}. \tag{40}$$

Here, the lower bound calculated *via* the multi-decomposition technique is the exact spectral gap.

4.2. Random walk on a m-dimensional box

We extend this calculation to the case considered in the introduction, where the m-dimensional box Ω_m is not necessarily a cube any longer. Its side lengths are denoted by $p_1, p_2, \ldots, p_m$. We follow the same line as in the previous example except that now $p_k \geqslant 2$. However, before going on, we need to clarify the question of boundary conditions, which was temporarily postponed in the introduction. In fact, we shall demonstrate the following striking property: the details of the proof do not depend precisely on boundary conditions provided the subsets $\Omega_{k;a}$ are the same, as well as the stationary distribution π. This is the first illustration of a general feature of the multi-decomposition technique that was already mentioned in section 3.3, and that will be largely discussed in the following sections: ν is an equilibrium quantity.

We consider both 'von Neumann' and periodic boundary conditions. In the first case, and as in our previous example, if the walker tries to move outside the grid, he stays at the same place. In the second case, the grid lies on a torus and the walker has always $2m$ possible moves; all the vertices play the same role. In both cases, the allowed transition rates are equal to $1/4m$. As compared to the previous examples, where they were set to $1/2m$, these transition rates ensure that $P(x, x) \geqslant 1/2$. (In the previous example, this property was due to the fact that all vertices lay on the boundary.)

The exact calculation of the gap of the transition matrix $P^{(m)}$ is similar to the previous one: $P^{(m)}$ can also be written as a tensor product of m one-dimensional walkers (equation (37)), since the walker chooses one of the m directions with equal probability and performs a one-dimensional walk in that direction. However, the transition matrix t_k of the one-dimensional walker now depends on both the side length p_k and the boundary conditions. For von Neumann ones,

$$t_k = \frac{1}{4}\begin{pmatrix} 3 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 1 & 2 & 1 \\ 0 & \cdots & \cdots & \cdots & 1 & 3 \end{pmatrix} \tag{41}$$

is a $p_k \times p_k$ matrix. Its eigenvalues are $\lambda_j = \cos(j\pi/p_k)/2 + 1/2$, where $j = 0, 1, \ldots, p_k - 1$. The eigenstates of $P^{(m)}$ are the tensor products of the one-dimensional ones and its eigenvalues are all the combinations

$$\lambda_{j_1, j_2, \ldots, j_m} = \frac{\lambda_{j_1} + \lambda_{j_2} + \cdots + \lambda_{j_m}}{m}. \tag{42}$$

If $\sup_k p_k = p_{k_0}$, the second largest eigenvalue is obtained when all the j_k are set to 0 except for the k_0th one which is set to 1. Thus $g(P^{(m)}) = (1 - \cos(\pi/p_{k_0}))/2m$. In the case of periodic boundary conditions, a similar calculation leads to $g\big(P^{(m)}_{\text{periodic}}\big) = (1 - \cos(2\pi/p_{k_0}))/2m$, with

$$t_{k,\text{periodic}} = \frac{1}{4}\begin{pmatrix} 2 & 1 & 0 & \cdots & \cdots & 1 \\ 1 & 2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 1 & 2 & 1 \\ 1 & \cdots & \cdots & \cdots & 1 & 2 \end{pmatrix} \tag{43}$$

instead of t_k.

We now decompose the m-dimensional walker in $(m-1)$-dimensional ones (figure 3). The argument is exactly the same as in the previous section 4.1, except that $g(P_k)$ now depends on k since p_k does. For each k, the decomposition $(\Omega_{k,a})$ is obtained by cutting the edges in the direction k of the grid. For both boundary conditions, this results in constraining the walker to move on a $(m-1)$-dimensional grid, with the same boundary conditions as in the original grid Ω_m. In addition, the matrix $\bar{\Pi}$ (or $\hat{\Pi}$) is slightly more complex: we calculate

$$|\Omega_{k;a}| = \prod_{j\neq k} p_j \qquad |\Omega_{k;a} \cap \Omega_{l;b}| = \prod_{\substack{j\neq k \\ j\neq l}} p_j \qquad \text{and} \qquad \hat{\Pi}_{(k;a),(l;b)} = \frac{1}{p_k} \tag{44}$$

whenever $k \neq l$, whatever a and b. The eigenvalues of $\hat{\Pi}$ can also be exactly calculated and one also gets $\nu = 1$. The redundancy is $r = m-1$ as it is in the previous example. Theorem 1 can be applied and one gets the gap relation $g(P^{(m)}) \geqslant \inf_k g(P_k)$. We recall that P_k is the direct sum (see equation (14)) of p_k transition matrices of a $(m-1)$-dimensional random walker, with transition rates $1/4m$ instead of $1/4(m-1)$. Taking this fact into account, one gets by induction

$$g(P^{(m)}) \geqslant \frac{1}{2m} \inf_{k=1,\ldots,m} (1 - \cos(\pi/p_k)) \tag{45}$$

for von Neumann boundary conditions. The exact gap was calculated above and this lower bound is in fact exact.

The conclusion is the same for periodic boundary conditions: since the subsets $\Omega_{k;a}$ are the same as in the von Neumann case, the matrix $\hat{\Pi}$ is identical for both boundary conditions. Therefore ν and r are also equal in both cases, and the gap relations are identical. Finally, the two proofs are exactly similar except that the first steps of the proof by induction, that is to say the spectral gaps of the *one*-dimensional random walkers, differ. One also finally gets the exact gap calculated above for periodic boundary conditions.

4.3. Random walk on the symmetric group with random transpositions

This Markov chain belongs to the 'Card shuffling models' [9]. In this example, we re-derive the spectral gap which is usually calculated by group representation theory arguments [17]. In this case, the multi-decomposition technique is remarkably simple and efficient and provides again an exact lower bound.

Given an integer $m \geqslant 2$ and the symmetric group S_m, the 'random transposition Markov chain' [9, 17] is defined on $\Omega = S_m$ as follows: given a state (a permutation) $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_m)$, pick up uniformly two different positions k and l at random, then transpose σ_k and σ_l with probability $1/2$ (so that $P(x,x) \geqslant 1/2$). This Markov chain converges towards the uniform distribution on S_m (see [9, 17]). The transition matrix $P^{(m)}$ is defined as follows: if two different permutations x and y differ by a single transposition then $P(x,y) = 1/[m(m-1)]$. In [17], it is proven that

$$g(P^{(m)}) = \frac{1}{m-1} \tag{46}$$

with arguments based on the theory of representations of S_m (the result is slightly different in [17] because transition rates are slightly different).

Now we decompose the random walk on S_m in random walks on S_{m-1} in order to compute inductively the lower bound on the gap. The decomposition is natural: for a given $k = 1, \ldots, m$, we fix the position of σ_k and prevent transpositions involving this index k. The subsets $\Omega_{k;a}$ are defined by the fixed value $\sigma_k = a$. The resulting chain on each $\Omega_{k;a}$ is the

same as the original one on a set of cardinality $m-1$, with transition rates $1/[m(m-1)]$ instead of $1/[(m-1)(m-2)]$. Now we calculate $\bar{\Pi}$:

$$|\Omega_{k;a}| = (m-1)! \qquad |\Omega_{k;a} \cap \Omega_{l;b}| = (m-2)! \qquad \bar{\Pi}_{(k;a),(l;b)} = \frac{1}{m-1} \tag{47}$$

when $k \neq l$ and $a \neq b$. If $k \neq l$ but $a = b$, $|\Omega_{k;a} \cap \Omega_{l;b}| = 0$ since σ_k and σ_l cannot be equal, and $\bar{\Pi}_{(k;a),(l;b)} = 0$.

Hence $\bar{\Pi}$ is a block-matrix with m lines and m columns of blocks:

$$\bar{\Pi} = \left(\begin{array}{c|c|c|c|c} \mathbb{1} & B & \cdots & \cdots & B \\ \hline B & \mathbb{1} & B & \cdots & B \\ \hline \vdots & \ddots & \ddots & \ddots & \vdots \\ \hline B & \cdots & B & \mathbb{1} & B \\ \hline B & \cdots & \cdots & B & \mathbb{1} \end{array}\right) \tag{48}$$

where each block is itself a $m \times m$ matrix and

$$B = \begin{pmatrix} 0 & \frac{1}{m-1} & \cdots & \cdots & \frac{1}{m-1} \\ \frac{1}{m-1} & 0 & \frac{1}{m-1} & \cdots & \frac{1}{m-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \frac{1}{m-1} & \cdots & \frac{1}{m-1} & 0 & \frac{1}{m-1} \\ \frac{1}{m-1} & \cdots & \cdots & \frac{1}{m-1} & 0 \end{pmatrix}. \tag{49}$$

The eigenvalues of B are 1 (non-degenerate) and $-1/(m-1)$, $(m-1)$-fold degenerate, with respective eigenvectors $X_0 = (1, \ldots, 1)$ and $X_i = (0, \ldots, 0, 1, -1, 0, \ldots, 0)$, $i = 1, \ldots, m-1$. Then one computes the m^2 eigenvectors of $\bar{\Pi}$: the $(m-1)^2$ vectors $(0, \ldots, 0, X_i, -X_i, 0, \ldots, 0)$, where $i > 0$, with the same eigenvalue $m/(m-1)$; the $(m-1)$ vectors $(0, \ldots, 0, X_0, -X_0, 0, \ldots, 0)$, with eigenvalue 0; the $(m-1)$ vectors $(X_i, X_i, \ldots, X_i)$ with eigenvalue 0; and the vector $(X_0, X_0, \ldots, X_0)$ with eigenvalue m. Finally, the multi-decomposition is non-degenerate and

$$\nu = \frac{m}{m-1}. \tag{50}$$

This value tends to 1 when m is large but is not equal to 1. This nuance is crucial in order to get the good behaviour of the gap with m.

On the other hand, this multi-decomposition is regular with redundancy $r = m-2$. Indeed, each transposition of indices k and l is forbidden in exactly two decompositions, namely $(\Omega_{k;a})$ and $(\Omega_{l;b})$. Then theorem 1 provides the gap relation

$$g(P^{(m)}) \geqslant \frac{m}{m-1} \inf_k g(P_k). \tag{51}$$

Taking into account the difference of transition rates in P_k and $P^{(m-1)}$, we have $g(P_k) = (m-2)/m\, g(P^{(m-1)})$ and

$$g(P^{(m)}) \geqslant \frac{m-2}{m-1} g(P^{(m-1)}) \geqslant \cdots \geqslant \frac{1}{m-1} g(P^{(2)}) = \frac{1}{m-1} \tag{52}$$

because $g(P^{(2)}) = 1$. Once again, this bound is the exact gap.

5. Urn problems

In this section, we show that the multi-decomposition technique can be applied to a sub-class of the larger class of 'urn models' [18, 19]. The latter have been designed as toy models for

glassy dynamics, to which belongs the maybe more notorious 'Backgammon model' [6]. For the first time, we explicitly introduce temperature in our method. However, for the sake of simplicity, we shall essentially focus on dynamics at vanishing temperature $T \to 0$.

In urn models, m identical urns contain n identical (distinguishable or not) particles, or balls. In the historical Ehrenfest urn model, $m = 2$ and n is large. Then balls can be exchanged from one urn to another, which will define the elementary moves. The different rules of exchange characterize the different urn models. In general, they are defined so that the system prefers to have the maximum of empty urns, and eventually only an urn containing all the balls. Usually, these urn models are endowed with a Hamiltonian: the energy E of a configuration is the number of non-empty boxes.

We shall focus on two different models governed by a Metropolis algorithm at $T \to 0$ based upon the previous Hamiltonian. These two models are identical, except that balls are indistinguishable in the first model and distinguishable in the second one. We shall see that this difference has dramatic consequences on their dynamics. The first model is usually known as the 'Monkey model' and the second one as the 'Backgammon model'. Their elementary moves are defined as follows:

(i) 'Monkey model' [19]: balls are *indistinguishable;* two distinct urns are chosen uniformly at random among the m possible ones, a 'departure' urn D, and an 'arrival' urn A. If a ball can be removed from D to A without increasing the energy E, that is to say if D *and* A are not empty, this move is performed[1];

(ii) 'Backgammon model' [6]: balls are *distinguishable;* among all the balls contained in the m urns, one ball b is chosen at random and its urn is denoted by D. A second urn is chosen uniformly at random *among the* $m - 1$ *remaining ones*. denoted by A. The ball b is taken from D to A if this move does not increase the energy (that is if A is not empty).

By construction, in both models, the number of empty urns can only increase. The dynamic is frozen when all the urns are empty save one. We can already foresee that model (ii) will be slower to reach this frozen state than model (i) because it is more difficult to empty an urn entirely: when one urn is nearly empty, the ball b belongs to the other urns with much higher probability.

Note that at $T > 0$, this Metropolis algorithm must be modified as follows [5]: an elementary move is accepted with probability $\min(1, \exp(-\Delta E)/T)$, where ΔE is the energy variation if the move *was* accepted. The canonical distributions corresponding to the above Hamiltonian are

$$\pi_{(\mathrm{i})}(n_1, \ldots, n_m) = \frac{1}{Z_{(\mathrm{i})}} \exp(-E/T) \tag{53}$$

for model (i), and

$$\pi_{(\mathrm{ii})}(n_1, \ldots, n_m) = \frac{1}{Z_{(\mathrm{ii})}} \frac{n!}{n_1! \ldots n_m!} \exp(-E/T) \tag{54}$$

for model (ii), where n_i represents the number of balls in the urn number i. These distributions satisfy the detailed balance (4) for the Metropolis Markov chain.

First we treat the backgammon model which displays slow dynamics. The monkey model will be studied at the end of the section. We will prove below that the spectral gap of the backgammon model at $T \to 0$ decays exponentially with n. By contrast, the monkey model is rapid: $1/g(P)$ is polynomial in m and n. As a consequence, all the glassy character of the backgammon model comes from the distinguishable character of its particles.

[1] In the monkey model of [19], the arrival box is non-empty; we get rid of this constraint for the sake of convenience with respect to our method; this modification does not alter the overall conclusions.

5.1. Backgammon model: case $m = 2$

To begin with, we study the case of $m = 2$ urns, denoted by U_1 and U_2. At this stage, let us discuss the important following point: at zero temperature, the fundamental state is 2-fold degenerate: all the balls can be in U_1 (state x_1) or in U_2 (state x_2). In other words, the eigenvalue 1 of P is degenerate and at vanishingly small temperature, the spectral gap tends to 0 and the equilibriation time diverges.

But what does this equilibriation time measure? At very low temperature, the energy landscape is essentially a bistable potential with an energy barrier of height 1. Thus $1/g(P)$ measures the time that the system needs to equilibriate between the two fundamental states x_1 and x_2, in other words to explore the two potential minima with almost equal probabilities. Naturally, this time diverges like $\exp(1/T)$. But we are not interested in such a time but rather in *the typical time to reach one of the two minima*, the 'absorption time' in [12]. In other words, at zero temperature, we consider that the system has reached equilibrium when one urn is empty (and when all urns are empty save one if there are several urns). For a finite system, this time remains finite, even at $T = 0$ [12, 13].

Therefore we would like to design a new transition matrix P', the spectral gap of which measures efficiently the previous time at $T = 0$. A solution consists in mixing up both stationary states as follows: when the system has reached one of the states x_1 or x_2 for the first time at t_0, at each time $t > t_0$, he can be flipped to state x_1 or x_2 with probability $1/2$. In other words, we set $P'(x_1, x_2) = P'(x_2, x_1) = P'(x_1, x_1) = P'(x_2, x_2) = 1/2$. As a consequence, at time $t_0 + 1$, the system can have explored the two potential minima with equal probabilities. The unique equilibrium distribution is now $\pi = (x_1 + x_2)/2$. And the spectral gap $g(P')$ now measures the equilibriation time we are interested in. We denote by $P^{(2)}$ the previous matrix P'. It is a $(n + 1) \times (n + 1)$ matrix, with elements proportional to the number of balls in the urn D:

$$P^{(2)} = \begin{pmatrix} \frac{1}{2} & \frac{1}{n} & 0 & \cdots & \cdots & 0 & \frac{1}{2} \\ 0 & 0 & \frac{2}{n} & 0 & \cdots & \cdots & 0 \\ 0 & \frac{n-1}{n} & \ddots & \frac{3}{n} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{3}{n} & 0 & \frac{n-1}{n} & 0 \\ 0 & \cdots & \cdots & 0 & \frac{2}{n} & 0 & 0 \\ \frac{1}{2} & 0 & \cdots & \cdots & 0 & \frac{1}{n} & \frac{1}{2} \end{pmatrix}. \tag{55}$$

We have not succeeded in diagonalizing this matrix. It is not specifically our purpose since we principally desire to exemplify how the multi-decomposition method relates m-urn problems to two-urn ones. However, high precision numerical diagonalization up to $n = 200$ shows without ambiguity that $g(P^{(2)}) \simeq 1/2^{n-1}$ when n is large. This value seems to be asymptotically exact. We shall see in the following subsection that all the slow character of the backgammon model lies in this 2-urn exponentially small gap: adding more urns will only modify it by a polynomial factor.

5.2. Backgammon model: case $m > 2$

Now we consider the case $m > 2$ and relate its dynamics to the previous one $m = 2$. We denote the urns by $U_1, \ldots, U_m$. We apply the same kind of trick as in the two-urn case to get rid of the degeneracy of the fundamental state: the m lower energy states, where all urns are empty save one, are mixed up and form a single state. Once all urns are empty save one, all

the balls can be transferred to any urn U_k with equal probability $1/m$. This trick ensures that at $T = 0$ the equilibrium state is non-degenerate and that the spectral gap of the new transition matrix measures the good equilibriation time, that is to say the typical time needed to empty all urns save one.

The multi-decomposition is defined as follows: $\Omega_{k,a}, k = 1, \ldots, m$, is the set of all configurations where the urn U_k contains exactly $n_k = a$ balls. In this subset, the number of balls of U_k cannot vary. The integer a can take all the values ranging from 0 to n. Now we prove that the multi-decomposition is not degenerate and that $\nu = 2$, whatever $m > 2$ and $n \geqslant 2$. We work with the matrix $\hat{\Pi}$ of equation (36).

However, one must keep in mind that all the method relies on the scalar product (5) where $\pi(x)$ cannot vanish for any x. Therefore we must work at $T > 0$ where $\pi(x) > 0$ and then take the limit $T \to 0$ in the gap relation (18). We recall that the coefficients of $\hat{\Pi}$ are the conditional probabilities at equilibrium that U_k contains a balls given that U_l contains b balls. When $T \to 0$, in regard to equation (54), the system condenses on its energy minimum *inside* $\Omega_{l;b}$. The energy is minimized when all the $n - b$ remaining balls are in the same urn. This minimum is $(m - 1)$-fold degenerate. Therefore $a = n - b$ with probability $1/(m - 1)$ and $a = 0$ with probability $1 - 1/(m - 1)$.

Therefore the limiting matrix $\hat{\Pi}$ is a $m \times m$ block-matrix like (48) where each block is now a $(n + 1) \times (n + 1)$ matrix and

$$B = \begin{pmatrix} \frac{m-2}{m-1} & \cdots & \cdots & \frac{m-2}{m-1} & 1 \\ 0 & \cdots & 0 & \frac{1}{m-1} & 0 \\ \vdots & & \iddots & & \vdots \\ 0 & \frac{1}{m-1} & 0 & \cdots & 0 \\ \frac{1}{m-1} & 0 & \cdots & \cdots & 0 \end{pmatrix}. \tag{56}$$

The matrix B can be diagonalized and its eigenvalues are 1 (non-degenerate), $1/(m - 1)$ and $-1/(m - 1)$. The corresponding eigenvectors are $(m - 1, 0, \ldots, 0, 1)$; the vectors $(1, 0, \ldots, 0, 1)$, $(0, 1, 0, \ldots, 0, 1, 0)$, and so forth; the vectors $(1, 0, \ldots, 0, -1)$, $(0, 1, 0, \ldots, 0, -1, 0)$, and so forth. The exact number of vectors of the two last species depends on the parity of n: if n is even, one should add the vector $(0, \ldots, 0, 1, 0, \ldots, 0)$, where the one is at the middle of the vector. The corresponding eigenvalue is $1/(m - 1)$. Then the same kind of calculation as in the previous example 4.3 leads to the conclusion that the multi-decomposition is non-degenerate and that $\nu = 2$.

As in the previous example, this multi-decomposition is regular with redundancy $r = m-2$ since each transposition of indices k and l is forbidden in exactly two decompositions. Since $\nu = 2$, theorem 1 reads

$$g(P^{(m)}) \geqslant \inf_{k=1,\ldots,m} g(P_k). \tag{57}$$

As compared to the previous examples, the situation is more complex here, because the matrix P_k is *not* the direct sum of identical matrices. According to the number $n_k = a$ of balls that are stuck in the urn U_k, the number of particles that are effectively involved in the dynamics on the subset $\Omega_{k;a}$ varies. It is equal to $n - a$. Now we prove by induction on m that

$$g(P^{(m)}) \geqslant \frac{1}{m-1} g(P^{(2)}) \tag{58}$$

where $g(P^{(2)})$ is the gap of the backgammon model on two urns with the same n balls.

This relation is clear for $m = 2$. Suppose it is true for $m - 1$. The matrix P_k in (57) is the direct sum of $n + 1$ matrices of backgammon models on $m - 1$ urns with $n - a$ balls,

$a = 0, \ldots, n$. However, their non-diagonal transition rates must be re-scaled by a factor $(n-a)/n$ to get P_k, so as to take into account that the *number of balls* differs in both matrices; and by a factor $(m-2)/(m-1)$ so as to take into account the *number of arrival urns*. Moreover, up to urn permutations, all matrices P_k are identical and they have the same gaps. Hence

$$g(P^{(m)}) \geqslant \inf_{a=0,\ldots,n-1} \frac{n-a}{n}\frac{m-2}{m-1} g\big(P^{(m-1)}_{n-a}\big) = \frac{1}{m-1} \inf_{a<n} \frac{n-a}{n} g\big(P^{(2)}_{n-a}\big) \tag{59}$$

where the subscript $n-a$ means that the system contains $n-a$ balls[2]. Since $g\big(P^{(2)}_{n-a}\big) \simeq 1/2^{n-a-1}$, the previous infimum is reached when $a = 0$. Hence (58) is verified.

This relation is of crucial importance in the understanding of the origin of glassy dynamics: all the slow character (the exponential decay of the gap with n) originates from the $m = 2$ case; adding additional boxes only divides the gap by a polynomial factor (see the discussion at the end of this subsection, and [12]).

We have tested numerically this gap relation on small systems where the transition matrix can easily be fully diagonalized numerically, namely for all m and n such that $m + n \leqslant 14$ as well as all $n \leqslant 20$ for $m \leqslant 4$. In all cases, the previous inequality turns out to be an equality. We are led to the following:

Conjecture 1. *Inequality (58) is in fact an equality:*

$$g(P^{(m)}) = \frac{1}{m-1} g(P^{(2)}). \tag{60}$$

We shall return to this conjecture below.

Usually, the time unit in discrete time Markov processes is chosen to depend on the system size in order to be more physical. In the present case, we define the time unit so that the time increment is $\delta t = 1/(m-1)$ at each step of the Markov chain, as in [6, 12]. Then on average, one tries one move per urn and per time unit and one expects that typical times will not depend on the system size at the large size limit; we will see below that it is precisely what happens. Spectral gaps must also be re-scaled to take this point into effect: $g'(P) = (m-1)g(P)$. Relation (58) becomes

$$g'(P^{(m)}) \geqslant g'(P^{(2)}). \tag{61}$$

In this section, we have focused on spectral gaps, whereas other references [6, 12, 13, 18, 19] consider absorption or ergodic times $\tau'(m, n)$ instead. Let us make the assumption that $\tau'(m, n) \simeq 1/g'(P^{(m)})$. In the case $m = 2$, we have seen that $g(P^{(2)}) \simeq 1/2^{n-1}$, and it has been calculated that $\tau'(2, n) \simeq 2^{n-1}$ [6, 12, 13]. Therefore the previous assumption is exact in this case. Under this assumption, relation (61) becomes $\tau'(2, n) \geqslant \tau'(m, n)$.

Since the two-urn dynamics is naturally faster than the m-urn one (see [12]), we also naturally assume $\tau'(2, n) \leqslant \tau'(m, n)$. As a conclusion,

$$\tau'(2, n) \simeq \tau'(m, n). \tag{62}$$

We recover the conclusion of Lipowski [12], which was based on numerical simulations. Under the previous assumption, this relation has the same meaning as conjecture 1.

[2] We have excluded the value $a = n$ in the relation. Indeed if U_k contains n balls, the subset $\Omega_{k;n}$ has only one trivial state and the Markov chain on $\Omega_{k;0}$ is degenerate. In equation (20), this subset does not provide any eigenvector $\boldsymbol{f}_j$ and cannot contribute to μ_1. Therefore it does not appear in the infimum of (18).

5.3. Monkey model [19]

Now we sketch the proof, that by contrast, the monkey model is rapid. This property is already present when $m = 2$. Indeed, in that case, the number of balls in each urn increases or decreases by one unit with equal probability at each time step. Therefore this model is equivalent to a random walker with $n + 1$ states with two absorbing states at its extremities. Its spectral gap is

$$g(P^{(2)}) = 1 - \cos(\pi/n) \simeq \frac{\pi^2}{2}\frac{1}{n^2}. \tag{63}$$

Taking care of transition rates, which are now equal to $1/[m(m-1)]$, all the previous proof can be transposed in the present case. In particular, in regard to equation (53), $\hat{\Pi}$ and ν are unchanged. One finally gets

$$g(P^{(m)}) \geqslant \frac{2}{m(m-1)} g(P^{(2)}) \simeq \frac{\pi^2}{m(m-1)}\frac{1}{n^2}. \tag{64}$$

The Markov chain is rapid.

6. Contingency table problem

Finally, we present partial results on a difficult problem from probability theory: the 'contingency table problem'. In this case, we will not be able to calculate exactly the norm ν because the matrices $\bar{\Pi}$ or $\hat{\Pi}$ are not as regular as in the previous examples. However ν can be computed numerically for thousands of different examples because these matrices are sufficiently small. This leads us to the conjecture that $\nu \leqslant 2$. This upper bound is sufficient to prove that the Markov chains under consideration are rapid (see equations (73)–(76)). Furthermore, as in section 4.2, the same argument can be applied to two different Markov chains commonly associated with this problem, because ν is an equilibrium quantity.

Contingency tables are $m \times n$ matrices of non-negative integers with given positive row and column sums, which arise in statistics as two-way tables to store data from sample collection (see [7, 8] for instance). To test correlations between row and column entries, χ^2-tests are applied to such tables, but they require to sample (almost) uniformly from the set of contingency tables with given row and column sums. Since no systematic way is known to perform this sampling, Monte Carlo Markov chains have been designed to explore randomly these configuration sets. We study two such chains.

Consider a positive integer Σ and two sequences of positive integers: the row sums $r_k, k = 1, \ldots, m$ and the column sums c_p, $p = 1, \ldots, n$ such that

$$\sum_{k=1}^{m} r_k = \sum_{p=1}^{n} c_p = \Sigma. \tag{65}$$

Denote by $\Omega_{m,n}$ the set of all the $m \times n$ contingency tables (z_{kp}) such that for all k and all p

$$\sum_{p=1}^{n} z_{kp} = r_k \qquad \text{and} \qquad \sum_{k=1}^{m} z_{kp} = c_p. \tag{66}$$

An example of contingency table is displayed in figure 4.

6.1. Diaconis and Saloff-Coste's chain [20] $\mathcal{M}_{DS}$

These authors define a discrete time reversible chain $\mathcal{M}_{DS}$ on $\Omega_{m,n}$ as follows: at time t the chain is in the state $x = (z_{ij})$. With probability $1/2$, do nothing (so that $P(x, x) \geqslant 1/2$); with

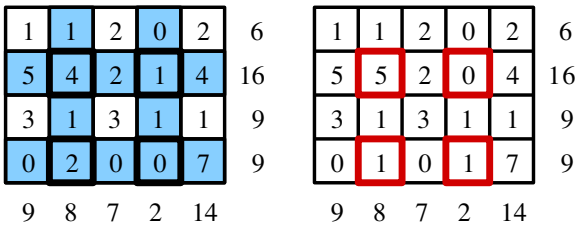

Figure 4. An example of contingency table (left) and of elementary move as defined in subsection 6.1. The row and column sums are listed on the right and below the table. The two lines and two columns chosen at random are emphasized, as well as the 'boxes' z_{ij} that decrease or increase by one unit. This elementary move does not affect the row and column sums.

probability $1/2$, pick up uniformly two different rows $k < l$ as well as two different columns $p < q$ at random. Choose a number $b \in \{-1, 1\}$ uniformly at random. Define $x' = (z'_{ij})$ by $z'_{ij} = z_{ij}$ for all i and j except that

$$z'_{kp} = z_{kp} + b \qquad z'_{kq} = z_{kq} - b \qquad z'_{lp} = z_{lp} - b \qquad z'_{lq} = z_{lq} + b. \tag{67}$$

The alternation of signs ensures that the row and column sums are conserved (see figure 4). If (z'_{ij}) is a contingency table (that is to say if all coefficients are non-negative), then accept this elementary move; else reject it. This chain $\mathcal{M}_{DS}$ is reversible and converges towards the uniform stationary distribution [20]. The non-zero non-diagonal coefficients of its transition matrix $P_{DS}^{(m,n)}$ are all equal to $2/[m(m-1)n(n-1)]$.

The time τ to reach equilibrium has been calculated in the particular case of two-row contingency tables [21] and is polynomial in the number of columns n and in the table sum Σ. To extend this result to the general case seems to be a serious challenge [21].

We build a multi-decomposition by fixing the values of a single row k (or symmetrically of a single column p). Let $\mathcal{A}_k$ denote the set of all the possible configurations in $\Omega_{m,n}$ of the whole row k. For any such row-configuration $a \in \mathcal{A}_k$, $\Omega_{k;a}$ is the set of all contingency tables in $\Omega_{m,n}$ the kth row of which is identical to a. Then all the subsets $\Omega_{k;a}$ for $k = 1, \ldots, m$ and $a \in \mathcal{A}_k$ form a suitable multi-decomposition of $\mathcal{M}_{DS}$. Indeed, when a row is held fixed, the resulting chain on $\Omega_{k;a}$ is a chain of type $\mathcal{M}_{DS}$ on a set of smaller contingency tables of dimensions $(m-1) \times n$ and of sum $\Sigma' = \Sigma - r_k$. (In a similar way, should the pth *column* be held fixed, the resulting chain would be a chain of type $\mathcal{M}_{DS}$ on a set of smaller contingency tables of dimensions $m \times (n-1)$ and of sum $\Sigma'' = \Sigma - c_p$.) Moreover, this multi-decomposition is regular of redundancy $r = m - 2$ because each move of row indices k and l is forbidden in exactly two decompositions, $(\Omega_{k;a})$ and $(\Omega_{l;b})$.

Unfortunately, the calculation of the norm ν of the multi-projection Π is much more complicated in this case than in the previous examples because the matrices $\bar{\Pi}$ or $\hat{\Pi}$ lack regularity. However, thousands of numerical simulations lead us to the following conjecture:

Conjecture 2. *In the above 'contingency table problem', whatever the table dimensions m and n, whatever the table sum Σ, whatever the row and column sums r_k and c_p, the previous multi-decomposition is non-degenerate and the following inequality holds: $\nu \leqslant 2$.*

Note that we already know that conjecture 2 is satisfied in a particular case: up to transition rates, the Markov chain on the symmetric group of subsection 4.3 is of $\mathcal{M}_{DS}$-type when $k = p$ and the r_k and c_p are all set to 1. Then $\Omega_{m,m}$ is the group of permutation matrices of size m, which is isomorphic to S_m. We found $\nu = m/(m-1) \leqslant 2$.

If this conjecture is true, theorem 1 reads

$$g\big(P_{DS}^{(m,n)}\big) \geqslant \inf_k g(P_k). \tag{68}$$

Now P_k is the direct sum (see equation (14)) of matrices of Markov chains of type $\mathcal{M}_{DS}$ on $(m-1)\times n$ contingency tables, except that their transition rates are equal to $2/[m(m-1)n(n-1)]$ instead of $2/[(m-1)(m-2)n(n-1)]$. Thus

$$g\big(P_{DS}^{(m,n)}\big) \geqslant \frac{m-2}{m} \inf_k g\big(P_{k,DS}^{(m-1,n)}\big). \tag{69}$$

We keep track of the index k in this relation because the spectral gap of the restriction of P to the decomposition $(\Omega_{k;a})$ depends on k in general since r_k does. In a similar way, if one chooses to fix columns indexed by p,

$$g\big(P_{DS}^{(m,n)}\big) \geqslant \frac{n-2}{n} \inf_p g\big(P_{p,DS}^{(m,n-1)}\big). \tag{70}$$

Then we understand that inductively, by 'removing' successively rows and columns, one will end up with a Markov chain of type $\mathcal{M}_{DS}$ on 2×2 contingency tables and that the following inequality will hold:

$$g\big(P_{DS}^{(m,n)}\big) \geqslant \frac{2}{m(m-1)}\frac{2}{n(n-1)} \inf_{\substack{k_1<\cdots<k_{m-2}\\ p_1<\cdots<p_{n-2}}} g\big(P_{\{k_i;p_j\}}^{(2,2)}\big). \tag{71}$$

In this inequality, $P_{\{k_i;p_j\}}^{(2,2)}$ is the transition matrix of the Markov chain of type $\mathcal{M}_{DS}$ on 2×2 contingency tables obtained by fixing successively $m-2$ rows and $n-2$ columns of the initial $m\times n$ tables[3].

Suppose these 2×2 contingency tables have row sums r and r' and column sums c and c' such that $r+r' = c+c' = Z$. These tables are entirely characterized by one of their coefficients, say z_{11}. This coefficient varies by ± 1 at each step of the Markov process. Therefore this chain is a one-dimensional random walker with non-diagonal transition rates equal to $1/4$ and 'von Neumann' boundary conditions; z_{11} has a minimal value $z_{\min}$ and a maximal one $z_{\max}$ which depend on the precise values of c, c', r and r'. Nevertheless, one is sure that $z_{\min} \geqslant 0$ and $z_{\max} \leqslant Z-1 \leqslant \Sigma - 1$, even if these bounds can be very loose. Therefore in all cases

$$g\big(P_{\{k_i;p_j\}}^{(2,2)}\big) \geqslant \frac{1}{2}\left(1-\cos\left(\frac{\pi}{\Sigma}\right)\right) \simeq \frac{\pi^2}{4}\frac{1}{\Sigma^2} \tag{72}$$

and inequality (71) reads

$$g\big(P_{DS}^{(m,n)}\big) \geqslant \frac{1}{2}\frac{2}{m(m-1)}\frac{2}{n(n-1)}\left(1-\cos\left(\frac{\pi}{\Sigma}\right)\right). \tag{73}$$

As a conclusion, $1/g\big(P_{DS}^{(m,n)}\big)$ is quadratic in m, n and Σ, provided conjecture 2 is true.

6.2. Dyer and Greenhill's chain [22] $\mathcal{M}_{DG}$

As compared to $\mathcal{M}_{DS}$, in this type of chain $\mathcal{M}_{DG}$, once the two rows and the two columns have been picked up at random, the new table is chosen uniformly among all the tables accessible for any value of $b \in \mathbb{Z}$. In practice, b can take all the integral values between a minimum value $b_{\min}$ and a maximum one $b_{\max}$, which depend on the table, the two chosen rows and

[3] It is possible that, during the previous process, after fixing a certain number of rows and columns, one obtains a problem on $m'\times n'$ tables with one (or even more) row (or column) sum equal to 0. In this case, it is useless to apply the multi-decomposition technique, because all the entries of that row (or column) are necessarily equal to 0. By fixing this column, one obtains only one subset $\Omega_{k;a}$ and inequality (68) is trivial in this case.

the chosen columns (see the discussion about z_{11} in the previous subsection). Apart from this (important) difference, all the previous development can be applied to the present chain. In particular, the multi-decomposition is adapted to $\mathcal{M}_{DG}$. Once again, one notices that the multi-decomposition technique does not depend on the details of the chain, provided the stationary distribution π and the subsets $\Omega_{k;a}$ are the same.

However, after removing $m-2$ rows and $n-2$ columns, the 'ultimate' Markov chain of type $\mathcal{M}_{DG}$ on 2×2 tables will differ from the $\mathcal{M}_{DS}$-type one. Indeed, at each step, the new 2×2 table is chosen uniformly at random among all the N accessible ones. Therefore all the coefficients of the transition matrix $P_{DG}^{(2,2)}$ associated with this chain are equal to $1/N$ and its gap is 1 whatever N, and inequality (71) becomes

$$g\bigl(P_{DG}^{(m,n)}\bigr) \geqslant \frac{2}{m(m-1)}\frac{2}{n(n-1)}. \tag{74}$$

Hence $1/g\bigl(P_{DG}^{(m,n)}\bigr)$ is also quadratic in m and n but not in Σ, provided conjecture 2 is true. Note that it was already proven rigorously that this chain is polynomial in n when m is held fixed [22, 23].

6.3. A remark about mixing times $\tau_{\mathrm{var}}(\varepsilon)$

We use relation (13) to calculate an upper bound of mixing times $\tau_{\mathrm{var}}(\varepsilon)$ for both Markov chains. Since π is uniform, $1/\ln(\pi(x))$ is constant and equal to $\ln N_{m,n}$, where $N_{m,n}$ denotes the cardinality of $\Omega_{m,n}$. Now each integral coefficient z_{kp} of the table is non-negative and is bounded above by the table sum Σ. Hence $N_{m,n}\leqslant \Sigma^{mn}$ and $\ln N_{m,n}\leqslant mn\ln\Sigma$. As a consequence,

$$\begin{aligned}\tau_{\mathrm{var}}(\varepsilon) &\leqslant \frac{1}{2}m(m-1)n(n-1)\frac{1}{1-\cos(\pi/\Sigma)}\left[mn\ln\Sigma+\ln\left(\frac{1}{\varepsilon}\right)\right]\\ &\simeq \frac{1}{\pi^2}m^3n^3\Sigma^2\ln\Sigma+\frac{1}{\pi^2}m^2n^2\Sigma^2\ln\left(\frac{1}{\varepsilon}\right)\end{aligned} \tag{75}$$

for the Diaconis and Sallof-Coste's Markov chain, and

$$\tau'_{\mathrm{var}}(\varepsilon) \leqslant \frac{1}{4}m^2(m-1)n^2(n-1)\ln\Sigma+\frac{1}{4}m(m-1)n(n-1)\ln\left(\frac{1}{\varepsilon}\right) \tag{76}$$

for the Dyer and Greenhill's one. As discussed in [22, 23], the first chain $\mathcal{M}_{DS}$ is polynomial in m, n and Σ, whereas the second one $\mathcal{M}_{DG}$ is polynomial in m and n but only logarithmic in Σ (if conjecture 2 is true). The present method provides a good understanding of this phenomenon by naturally relating the dynamics on $m\times n$ tables to the dynamics on 2×2 ones.

To close this section, let us mention the existence of generalizations in higher dimensions of the previous Markov chains on contingency tables [24]. It would be interesting to test whether this technique can be applied to them.

7. Conclusion

The multi-decomposition technique is a generalization of the 'chain decomposition method' of D Randall and co-authors. Instead of a single decomposition, the method uses several intricate ones. It relates the spectral gap of a complex Markov chain to the spectral gaps of Markov chains on smaller subsets of the configuration space. In particular, a key quantity intervenes in this relation, the so-called 'norm of the multi-projection' of the multi-decomposition

(see section 3). It is calculated *at equilibrium*, whereas the spectral gap is a dynamical quantity, thus relating dynamical properties to equilibrium ones.

We have illustrated the potentialities of the multi-decomposition technique on several examples. The first examples (random walkers and 'card shuffling') are elementary, in the sense that their spectral gaps can be calculated by other means. However, they are a good illustration of the method efficiency. We have also seen in these examples that the same multi-decomposition can be applied to different dynamics on the same system.

The last examples are more interesting since they deal with complex Markov chains of theoretical interest. The 'Backgammon model' and the 'Monkey model' belong to the wider class of 'urn models' and are related to the physics of glassy systems. We have recovered in a rigorous way that the former is slow whereas the latter is rapid. This fundamental difference is only due to the fact that particles are distinguishable in the first model and indistinguishable in the second one. These two examples have also been the occasion to introduce temperature in the method. The 'contingency table problem' is a notoriously difficult problem from probability theory. We have not been able to solve completely this problem. However, the computational advantage of the method is illuminating in this case: the matrix $\hat{\Pi}$ is much smaller than the original matrix P and can be numerically diagonalized for much larger systems. These numerical calculations support a conjecture, the veracity of which would imply that the Markov chains under interest are rapid. In this case also, the same multi-decomposition can be applied to two different dynamics.

In addition, we have remarked that for the first four examples examined in this paper, the gap relation (18) of theorem 1 turned out to be an equality, since we obtained the exact gaps at the end of the calculations. It is legitimate to wonder whether this is always true. Indeed, even if the present version of the theorem is sufficient to establish that some Markov chains are rapidly mixing, this would be a stronger result than theorem 1. To answer this question, we mention that we know an example on a 20-state Markov chain where (18) is only an inequality and not an equality. Hence the question becomes: is it possible to establish a criterion to decide whether relation (18) is an equality or not? It is difficult to tackle this question in the frame of our calculations, because inequalities appear in the demonstration of theorem 1.

Acknowledgments

I wish to express my gratitude to Dana Randall who pointed out to me the fact that beyond random tilings, the present method might be of interest in the 'contingency table problem'. I am also grateful to Vianney Desoutter for constructive discussions about urn models.

References

[1] Grimmett G and Stirzaker D 1982 *Probability and Random Processes* (Oxford: Clarendon)
[2] Feller W 1968 *An Introduction to Probability Theory and Its Applications vol I 3rd edn* (New York: Wiley)
[3] Sinclair A 1993 *Algorithms for Random Generation and Counting: a Markov Chain Approach* (Boston: Birkhäuser)
[4] Jerrum M and Sinclair A 1996 The Markov chain Monte Carlo method: an approach to approximate counting and integration *Approximation Algorithms for NP-hard Problems* ed D S Hochbaum (Boston: PWS Publishing)
[5] Newman M E J and Barkema G T 1999 *Monte Carlo Methods in Statistical Mechanics* (Oxford: Clarendon)
[6] Ritort F 1995 Glassiness in a model without energy barriers *Phys. Rev. Lett.* **75** 1190
[7] Diaconis P and Efron B 1985 Testing for independence in a two-way table *Ann. Statistics* **13** 845
[8] Barlow R 1989 *Statistics* (Chichester: Wiley)
[9] Aldous D 1983 Random walks on finite groups and rapidly mixing Markov chains *Séminaire de probabilités XVII (Springer Lecture Notes in Mathematics vol 986)* (Berlin: Springer) p 243

[10] Martin R A and Randall D 2000 Sampling adsorbing staircase walks using a new Markov chain decomposition method *Proc. 41st Symp. Foundations of Computer Science (FOCS)*
Madras N and Randall D 2002 Markov chain decomposition for convergence rate analysis *Ann. Appl. Prob.* **12** 581
[11] Destainville N 2002 Flip dynamics in octagonal rhombus tiling sets *Phys. Rev. Lett.* **88** 30601
[12] Lipowski A 1997 Absorption time in certain urn models *J. Phys. A: Math. Gen.* **30** L91
[13] Murthy K P N and Kehr K W 1997 Relaxation at late stages in an entropy barrier model for glassy systems *J. Phys. A: Math. Gen.* **30** 6671
[14] Mehta M L 1989 *Matrix Theory, Selected Topics and Useful Results* (Les Éditions de Physique, Les Ulis, France)
[15] Wilson D B Mixing times of lozenge tiling and card shuffling Markov chains *The Annals of Applied Probability* (*Preprint* math.PR/0102193) at press
[16] Diaconis P and Strook D 1991 Geometric bounds for eigenvalues of Markov chains *Ann. Appl. Prob.* **1** 36
[17] Diaconis P and Shahshahani M 1981 Generating a random permutation with random transpositions *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **57** 159
[18] Ehrenfest P and Ehrenfest T 1907 Über zwei bekannte Einwände gegen das Boltzmannsche *H*-Theorem *Phys. Z.* **8** 311
Godrèche C, Bouchaud J P and Mézard M 1995 Entropy barriers and slow relaxation in some random walk models *J. Phys. A: Math. Gen.* **28** L603
Godrèche C and Luck J-M 1996 Long-time regime and scaling of correlations in a simple model with glassy behavior *J. Phys. A: Math. Gen.* **29** 1915
[19] Godrèche C and Luck J-M 2002 Nonequilibrium dynamics of urn models *J. Phys.: Condens. Matter* **14** 1601
[20] Diaconis P and Saloff-Coste L 1995 Random walk on contingency tables with fixed row and column sums Tech. Report, Department of Mathematics, Harvard University
[21] Hernek D 1998 Random generation of $2 \times n$ contingency tables *Random Struct. Algorithms* **13** 71
[22] Dyer M and Greenhill C 2000 Polynomial-time counting and sampling of two-rowed contingency tables *Theor. Comput. Sci.* **246** 265
[23] Cryan M, Dyer M, Goldberg L A, Jerrum M and Martin R 2002 Rapidly mixing Markov chains for sampling contingency tables with a constant number of rows *Proc. 43rd Annual IEEE Symp. on Foundations of Computer Science*
[24] Aoki S and Takemura A 2003 Minimal basis for connected Markov chains over $3 \times 3 \times K$ contingency tables with fixed two-dimensional marginals *Aust. J. Stat.* **45** (*Preprint* METR Technical Report, 02-02) accepted