Coupling and Self-Stabilization

Laurent Fribourg, Stéphane Messika and Claudine Picaronny

LSV, CNRS & ENS de Cachan,
61 av. du Prés. Wilson
94235 Cachan cedex, France

Abstract. A randomized self-stabilizing algorithm $\mathcal{A}$ is an algorithm that, whatever the initial configuration is, reaches a set $\mathcal{L}$ of *legal configurations* in finite time with probability 1. The proof of convergence towards $\mathcal{L}$ is generally done by exhibiting a potential function φ, which measures the "vertical" distance of any configuration to $\mathcal{L}$, such that φ decreases with non-null probability at each step of $\mathcal{A}$. We propose here a method, based on the notion of coupling, which makes use of a "horizontal" distance δ between any pair of configurations, such that δ decreases in expectation at each step of $\mathcal{A}$. In contrast with classical methods, our coupling method does not require the knowledge of $\mathcal{L}$. In addition to the proof of convergence, the method allows us to assess the convergence rate according to two different measures. Proofs produced by the method are often simpler or give better upper bounds than their classical counterparts, as examplified here on Herman's mutual exclusion and Iterated Prisoner's Dilemma algorithms in the case of cyclic graphs.

1 Introduction

The notion of self-stabilization was introduced in computer science by Dijkstra [4]. A distributed algorithm is self-stabilizing if, whatever the initial configuration it starts from, it reaches within a finite time a set $\mathcal{L}$ of "legal" configurations, i.e, configurations satisfying a desired property. Self-stabilizing systems have notably received much attention because they propose an elegant way of solving the problem of fault-tolerance [14]. Randomization is often employed in self-stabilization to break the symmetry in anonymous systems (see [5]). With randomized self-stabilizing algorithms, the convergence towards $\mathcal{L}$ is guaranteed with probability 1.

We show here that we can use the notion of *coupling*, as used in the field of Applied Probability, to prove the self-stabilization property and at the same time, the rate of convergence to the set of legal configurations. Coupling is a method used for analysing the rate of convergence to equilibrium in Markov chain Monte Carlo experiments (see, e.g., [22]). The coupling time is the time that two faithful copies of a stochastic process coalesce together. Coupling time is generally used as an upper bound of the "mixing time" of a Markov chain $\mathcal{A}$, i.e., the time for the chain to be ε-close to its stationary distribution. Here we used the coupling mechanism in an original manner in order to simultaneously prove the self-stabilization of an algorithm $\mathcal{A}$ w.r.t. the set of legal configurations $\mathcal{L}$, and analyze the rate of convergence to $\mathcal{L}$, assuming this set is strongly connected. The coupling time will be used as an upper bound of two measures of the rates of convergence of $\mathcal{A}$: the expected time of reaching $\mathcal{L}$ ("hitting time") and the time after which $\mathcal{L}$ has been reached with high probability ("ε-absorption time").

Comparison with related work. Classically, self-stabilization is shown by finding an integer-valued potential function φ on the set Ω of configurations that decreases with non-null probability until $\mathcal{L}$ is reached. The expected time of hitting is calculated independently (see, e.g., [20, 5, 16]). There is an other classical method for both showing self-stabilization w.r.t. $\mathcal{L}$ and analysing the rate of convergence, as examplified in [7], which consists in finding an integer-valued potential function φ on Ω such that, basically:

$\varphi(X) = 0$ iff $X \in \mathcal{L}$ and $E[\varphi(X_{t+1})] \leq \beta\varphi(X_t)$ for some β $(0 \leq \beta < 1)$.

This function φ can be seen as a "vertical" distance that separates X from $\mathcal{L}$.

Our new method consists in finding a coupling (X_t, Y_t) and a "horizontal" distance δ on $\Omega \times \Omega$ such that, basically:

$E[\delta(X_{t+1}, Y_{t+1})] \leq \beta\delta(X_t, Y_t)$, for some β $(0 \leq \beta < 1)$.

The advantages of our coupling method are the following:

- it provides us not only with a proof of self-stabilization, but also with an upper bound for the hitting time and the ε-absorption time,
- it does not rely on the knowledge of $\mathcal{L}$,
- the evaluation of β can be greatly simplified by using various optimizations of coupling, such as path coupling (see [3]),
- on Herman's mutual exclusion and Iterated Prisoner's Dilemma algorithms, proofs produced by our method are simpler or give better upper bounds than their classical counterparts.

Our method is limited in its applicability in the context of self-stabilization, because we have to assume the set of "legal"states to be strongly connected and the scheduling to be fixed (e.g., synchronous or randomized central).

Plan of the paper. After some preliminaries on randomized distributed algorithms (Sec. 2), we define self-stabilization in terms of ergodicity (Sec. 3). We then relate the notion of coupling to that of self-stabilization (Sec. 4), which yields a new method for proving self-stabilization (Sec. 5). The method is refined via the technique of Path Coupling (Sec. 6). We conclude in Sec. 7.

2 Randomized Distributed Algorithms As Markov Chains

In a distributed system, the topology of the network of machines is generally given under the form of a graph $G = (V, E)$, where the set $V = \{1, \cdots, N\}$ of vertices corresponds to the locations of the machines. There is an edge between two vertices when the corresponding machines can communicate together. All the machines are here identical finite state machines. The space of states is Q. A *configuration* x of the network is the N-tuple of all the states of the machines. The set of configurations Q^N is denoted Ω. Given a configuration x of Ω, the state of the i-th machine is written $x(i)$. The communication between machines is done here through the reading of neighbors' states. Randomized distributed algorithms are characterized by a *scheduler* (or *adversary*), i.e., a mechanism which selects, at each step, a nonempty subset of machines, and a set of *actions* which applies simultaneously at each selected machine. In this paper, we suppose that the scheduler is fixed and memoryless (called "oblivious" in [21]): at each step, it selects a subset of machines depending on the current configuration only. For example, we will consider the case of a *synchronous scheduler* (resp. *randomized central scheduler*) which selects, at each step, all the machines (resp. a single machine randomly chosen). Once a machine is selected, its state (as well as possibly, the state of some of its neighbors) is changed by the action that applies. For a given memoryless scheduler, the randomized distributed algorithm can be seen as a Markov chain $\mathcal{A}$ on Ω (see [6]): the probability at each step to go from a configuration x to another one y is a constant, denoted $\mathbf{P}(x, y)$, that depends on x and y only. Suppose that all the configurations of Ω are of the form x_e with $e \in \{1, 2, \cdots, |\Omega|\}$. Then $\mathcal{A}$ is characterized by the matrix $\mathbf{P}$ on $\Omega \times \Omega$ which has $\mathbf{P}(x_e, x_f)$ as (e, f)-coordinate. The probability of going from x to y in t steps is $\mathbf{P}^t(x, y)$ (see, e.g., [12]).

Example 1. We consider Herman's mutual exclusion algorithm [10]. The topology is a cyclic graph (ring) of N vertices, and the scheduler synchronous. The set of states is $Q = \{0, 1\}$, and the number of machines N is odd. At each step, the state of every machine $x(i)$ $(1 \leq i \leq N)$ is changed into $x'(i)$ as follows:

- if $x(i) \neq x(i-1)$ then $x'(i) = \neg x(i)$,
- if $x(i) = x(i-1)$ then $x'(i) = \begin{cases} 0 \text{ with probability } 1/2, \\ 1 \text{ with probability } 1/2. \end{cases}$

(When $i = 1$, $(i-1)$ stands for N. As usual, $\neg 0$ stands for 1, and $\neg 1$ for 0.)

Example 2. We consider the problem of the Iterated Prisoner's Dilemma, as modeled in [7]. The topology is a cyclic graph (ring) of N vertices, and the scheduler randomized central. The set of states is $Q = \{-, +\}$. At each each step, a vertex i $(1 \leq i \leq N)$ is chosen uniformly at random, and the values $x(i)$ and $x(i+1)$ are changed into $x'(i)$ and $x'(i+1)$ respectively as follows:

- if $x(i) = x(i+1)$, then $x'(i) = x'(i+1) = +$,
- if $x(i) = \neg x(i+1)$, then $x'(i) = x'(i+1) = -$.

(When $i = N$, $(i+1)$ stands here for 1. Here, $\neg +$ stands for $-$, and $\neg -$ for $+$.)

3 Self-Stabilization

3.1 Convergence

Let us consider a Markov chain $\mathcal{A}$. Two configurations x and y are in the same equivalence class if they are "strongly connected", i.e., if there exist t, u such that $\mathbf{P}^t(x, y) > 0$ and $\mathbf{P}^u(y, x) > 0$. Given two classes C and C', $C' \ll C$ means that $\mathbf{P}^t(x, y) > 0$ for some $x \in C, y \in C'$ and $t > 0$. The minimal classes for $\ll$ are called *ergodic sets*. More precisely:

Definition 1. *Let $\mathcal{M} \subset 2^{\Omega}$ be a set of configurations. $\mathcal{M}$ is an* ergodic set *if*

1. $\mathcal{M}$ is strongly connected, i.e.: $\forall x, y \in \mathcal{M}: \quad \mathbf{P}^t(x, y) > 0$ for some t, and

2. $\mathcal{M}$ is closed, i.e.: $(x \in \mathcal{M} \ \wedge \ \mathbf{P}(x, y) > 0) \ \Rightarrow \ y \in \mathcal{M}$.

Every finite Markov chain has always at least one ergodic set (since finite partial ordering $\ll$ must have at least one minimal element). Furthermore, two distinct ergodic sets are disjoint (since they are both strongly connected). We focus here on Markov chains with a single ergodic set. As explained below, they correspond to the notion of "self-stabilizing" algorithms, as originally defined by Dijkstra in the deterministic framework [4].

In [4], Dijkstra defined the set $\mathcal{L}$ of *legal configurations* of a distributed algorithm as a set of configurations meeting a global correctness criterion (e.g., the uniqueness of "token") with the constraints of (1) strong connectivity and (2) closure. Therefore, in our context of Markov chains, a set of legal configurations is necessarily an ergodic set. An algorithm $\mathcal{A}$ is *convergent w.r.t* a set $\mathcal{L}$ of legal configurations if, starting from any initial configuration, the system is guaranteed to reach a configuration of $\mathcal{L}$ within a finite number of transitions. For example, in mutual exclusion problems, a legal configuration is a configuration with a single token, which expresses the fact that only one machine can enjoy the resource. Following Dijkstra's definition, we have:

Definition 2. *Given a set $\mathcal{L}$ of configurations, $\mathcal{A}$ is* self-stabilizing w.r.t. *$\mathcal{L}$, in Dijkstra's sense*[1]*, if*

1. $\mathcal{L}$ is strongly connected,

2. $\mathcal{L}$ is closed, and

3. $\mathcal{A}$ is convergent w.r.t. $\mathcal{L}$.

In the probabilistic context of Markov chains, the convergence property (3) has to be guaranteed with probability 1. Formally, $\mathcal{A}$ is *convergent w.r.t. $\mathcal{L}$ (with probability 1)* if: $\forall x \ \sum_{y \in \mathcal{L}} \mathbf{P}^t(x, y) \to 1$ when $t \to \infty$. We have (see, e.g., [12]):

Proposition 1. *A finite Markov chain $\mathcal{A}$ converges with probability 1 to the union of the ergodic sets, whatever the initial configuration is.*

[1] The notion of self-stabilization has been relaxed since pioneering Dijkstra's work, and requirement (1) of strong connectivity for $\mathcal{L}$ is often dropped (see [20]). However, in this paper, we keep requirement (1) becauses it matches better with the notion of rapidly mixing Markov chains.

It follows:

Proposition 2. *A randomized distributed algorithm $\mathcal{A}$ is self-stabilizing (in Dijkstra's sense) w.r.t. a set $\mathcal{L}$ of configurations iff $\mathcal{L}$ is the unique ergodic set of $\mathcal{A}$.*

Proof:
($\Leftarrow$) Suppose that $\mathcal{L}$ is the unique ergodic set of $\mathcal{A}$. Then $\mathcal{L}$ satisfies the properties of closure and strong connectivity, and by Proposition 1, there is convergence with probability 1 to the union of the ergodic sets, viz. $\mathcal{L}$, whatever the initial configuration is. Hence $\mathcal{A}$ is self-stabilizing w.r.t. $\mathcal{L}$.
($\Rightarrow$) Suppose that $\mathcal{A}$ is self-stabilizing w.r.t. $\mathcal{L}$. Then $\mathcal{L}$ is closed and strongly connected, hence is an ergodic set. Let us show that $\mathcal{L}$ is the unique ergodic set by *reductio ad absurdum*: suppose that there is another (disjoint) ergodic set $\mathcal{L}'$, and let us show that $\mathcal{A}$ is not self-stabilizing w.r.t. $\mathcal{L}$. Consider an element $y \in \mathcal{L}'$. Every sequence of transitions starting from y stays in $\mathcal{L}'$ (since $\mathcal{L}'$ is closed). Hence no sequence of transitions starting from y can reach $\mathcal{L}$ (since $\mathcal{L}$ and $\mathcal{L}'$ are disjoint). So, for any starting configuration $y \in \mathcal{L}'$, the probability of reaching $\mathcal{L}$ is null. It follows that $\mathcal{A}$ is not convergent to $\mathcal{L}$, hence not self-stabilizing. □

It is generally easy to check that a given set $\mathcal{L}$ of configurations is ergodic for $\mathcal{A}$, as illustrated in Examples 1 and 2. What is difficult is to show the *uniqueness* of the ergodic set, i.e., the absence of any other ergodic set, besides $\mathcal{L}$: for example, for a mutual exclusion algorithm, the absence of any subset of "looping" configurations with two tokens.

Example 3. Consider Herman's algorithm in the case where N is odd. In a configuration, a "token" at position i ($1 \leq i \leq N$) corresponds to the presence of two contiguous states of the same value (00 or 11) at position $i-1$ and i. Since N is odd, any configuration contains always at least one token. It is easy to see that such a set is ergodic .

Example 4. In the Iterated Prisoner's Dilemma, the set $\mathcal{L}$ of legal configurations is the singleton made of the configuration $x^* = (+)^N$. Obviously, any action transforms x^* to itself. Hence, $\{x^*\}$ is trivially an ergodic set.

In the following, we assume that we are given a Markov chain $\mathcal{A}$ and an ergodic set $\mathcal{L}$, and we focus on the problem of proving the self-stabilization property of $\mathcal{A}$ w.r.t. $\mathcal{L}$. We are also interested in evaluating the rate of convergence of $\mathcal{A}$ to $\mathcal{L}$. We will use two different measures of convergence: the "expected hitting time" and the "ε-absorption time".

The expected hitting time is the standard rate of convergence used in the self-stabilization community (see, e.g., [5], p. 118). It is the expected time for $\mathcal{A}$ to reach $\mathcal{L}$, starting from the "worst" configuration, i.e.:

Definition 3. *Given a Markov chain $\mathcal{A}$ and a set $\mathcal{L}$, the* expected hitting time *of $\mathcal{L}$ (or more simply the* hitting time*) is:*

$$\mathbf{H}_{\mathcal{L}} = max_{x \in \Omega}\ E(H_{x\mathcal{L}}),$$
where $E(.)$ *denotes expectation and* $H_{x\mathcal{L}} = min\{t:\ X_t \in \mathcal{L} \mid X_0 = x\}$.

We will also use a different rate of convergence, called here "ε-absorption time", that gives the time after which $\mathcal{L}$ has been reached with high probability.

Definition 4. *Given a Markov chain* $\mathcal{A}$ *and an ergodic set* $\mathcal{L}$*, the* time of ε-absorption by $\mathcal{L}$ *(or simply the* ε-absorption time*) is:*

$$\mathbf{\Theta}_{\mathcal{L}}(\varepsilon) = max_{x \in \Omega}\ \Theta_{x\mathcal{L}}(\varepsilon),$$
where $\Theta_{x\mathcal{L}}(\varepsilon) = min\{t:\ Pr(X_t \in \mathcal{L}) \geq 1 - \varepsilon \mid X_0 = x\}$.

So $\Theta_{\mathcal{L}}(\varepsilon)$ is the minimal number of steps in which $\mathcal{A}$ reaches $\mathcal{L}$ with probability at least $1 - \varepsilon$. This notion is, for example, used in [7], for measuring the rate of convergence of the Iterated Prisoner's Dilemma. The notion is closed to the notion of "mixing time", that measures the number of steps after which the chain is ε-close of the "stationary distribution" of $\mathcal{A}$. [2] An upper bound on the mixing time is often computed by finding the "coupling time" (see, e.g., [19, 22]), that is defined henceforth.

Remark. Various notions of convergence rates are compared together in [1] and [15], but these studies concern only the case of "irreducible" chains where $\mathcal{L}$ and Ω coincide (all the configurations are legal and inter-connected) while, here, we are concerned with "reducible" chains where $\mathcal{L}$ is a strict subset of Ω.

4 Coupling

Let us come back to $\mathcal{A}$ viewed as a Markov chain. It will be characterized by a sequence of random variables (X_t) taking their values on the space Ω of configurations. The method of "coupling" is an elementary probabilistic method for measuring the "agreement" time between the components of a stochastic process (see, e.g., [22, 19]).

Definition 5. *A* coupling *is a Markov chain on* $\Omega \times \Omega$ *defining a stochastic process* $(X_t, Y_t)_{t=1}^{\infty}$ *with the properties*

1. Each of the processes (X_t) *and* (Y_t) *is a faithful copy of* $\mathcal{A}$ *(given initial configurations* $X_0 = x$ *and* $Y_0 = y$*).*

2. If $X_t = Y_t$*, then* $X_{t+1} = Y_{t+1}$.

Condition 1 ensures that each process, viewed in isolation, is just simulating the original chain – yet the coupling updates them simultaneously so that they tend to move closer together, according to some notion of distance. Once the pair of configurations agree, condition 2 guarantees they agree from that time forward.

Definition 6. *Given a coupling* (X_t, Y_t)*, the* (expected) coupling time *is:*

$$\mathbf{T} = max_{x \in \Omega, y \in \Omega}\ E(T_{x,y}),$$
where $T_{x,y} = min\{t : X_t = Y_t \mid X_0 = x, Y_0 = y\}$.

[2] Actually, the two notions coincide when the set $\mathcal{L}$ is reduced to a single configuration, as in the example of Iterated Prisoner's Dilemma.

The coupling time is often computed as un upper bound on the mixing time, in order to show the property of "rapid mixing" for $\mathcal{A}$ (i.e, the fact that the mixing time is bounded above by a polynomial in N and $\ln(\frac{1}{\varepsilon})$). We show hereafter that the coupling time gives also an upper bound on the hitting time.

Theorem 1. *Given a Markov chain $\mathcal{A}$ and an ergodic set $\mathcal{L}$, if there exists a coupling of finite expected time $\mathbf{T}$, then:*

1. $\mathcal{A}$ is self-stabilizing w.r.t. $\mathcal{L}$.

2. The hitting time $\mathbf{H}_{\mathcal{L}}$ is less than or equal to $\mathbf{T}$: $\quad \mathbf{H}_{\mathcal{L}} \leq \mathbf{T}$.

Proof:
1. By *reductio ad absurdum*: Suppose that there are two non-empty ergodic sets $\mathcal{L}_1$ and $\mathcal{L}_2$ with two elements $X_0 = x \in \mathcal{L}_1$ and $Y_0 = y \in \mathcal{L}_2$. Then, for all $t > 0$, $X_t \in \mathcal{L}_1$ and $Y_t \in \mathcal{L}_2$ since $\mathcal{L}_1$ and $\mathcal{L}_2$ are closed. Therefore for all $t > 0$, $X_t \neq Y_t$ since $\mathcal{L}_1$ and $\mathcal{L}_2$ are disjoint. Hence $T_{x,y}$ is infinite. So is $\mathbf{T}$, which contradicts the assumption.

2. Let us show: $\mathbf{H}_{\mathcal{L}} \leq \mathbf{T}$. Recall that: $H_{x\mathcal{L}} = min\{t : \ X_t \in \mathcal{L} \mid X^0 = x\}$, and $T_{xy} = min\{t : \ X_t = Y_t \mid X^0 = x, Y^0 = y\}$. Suppose now that $y \in \mathcal{L}$. Then $Y_t \in \mathcal{L}$ since $\mathcal{L}$ is closed. Hence: $H_{x\mathcal{L}} \leq T_{xy}$ for all $x \in \Omega, y \in \mathcal{L}$. And by taking the expectations, then the maxima of the two sides: $\mathbf{H}_{\mathcal{L}} \leq \mathbf{T}$. □

5 Two Sufficient Criteria of Self-Stabilization

By Theorem 1, finding an upper bound on the time of coupling T allows us to prove simultaneously the self-stabilization and to obtain an upper bound on the hitting time. Following classical results on mixing time (see e.g. [8]), we give hereafter two sufficient conditions for bounding the coupling time. In each case, this provides us additionally with an upper bound not only for the hitting time, but also for the ε-absorption time.

Theorem 2. *Given a Markov chain $\mathcal{A}$ and an ergodic set $\mathcal{L}$, suppose there exists a coupling (X_t, Y_t) and an integer-valued function δ on $\Omega \times \Omega$ which takes values in $\{0, 1, \cdots, B\}$ such that $\delta(X_t, Y_t) = 0$ iff $X_t = Y_t$, and:*

$$\exists \beta < 1 \quad \forall (X_t, Y_t) \quad E(\delta(X_{t+1}, Y_{t+1})) \leq \beta \delta(X_t, Y_t). \qquad (*)$$

Then:

1. $\mathcal{A}$ is self-stabilizing w.r.t. $\mathcal{L}$.

2. The expected hitting time satisfies: $\mathbf{H}_{\mathcal{L}} \leq \frac{B}{1-\beta}$.

3. The ε-absorption time satisfies: $\Theta_{\mathcal{L}}(\varepsilon) \leq \frac{\ln(B/\varepsilon)}{1-\beta}$.

Proof: See Appendix 1.

A similar theorem exists even when $\beta = 1$, i.e.: $E(\delta(X_{t+1}, Y_{t+1})) \leq \delta(X_t, Y_t)$, provided that the probability of $(X_{t+1}, Y_{t+1}) \neq (X_t, Y_t)$ can be bounded below.

Theorem 3. *Given a Markov chain $\mathcal{A}$ and an ergodic set $\mathcal{L}$, suppose there exists a coupling (X_t, Y_t) and an integer-valued function δ on $\Omega \times \Omega$ which takes values* *in* $\{0, 1 \cdots, B\}$ *such that* $\delta(X_t = Y_t) = 0$ *iff* $X_t = Y_t$*, and such that, there exists* $\alpha > 0$ *such that, for all* (X_t, Y_t) *with* $X_t \neq Y_t$*:*

$$E(\delta(X_{t+1}, Y_{t+1})) \leq \delta(X_t, Y_t) \ \wedge \ Pr(\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)) \geq \alpha. \qquad (**)$$

Then:

1. $\mathcal{A}$ *is self-stabilizing w.r.t.* $\mathcal{L}$*.*
2. *The expected hitting time satisfies:* $\mathbf{H}_{\mathcal{L}} \leq B^2/\alpha$*.*
3. *The* ε*-absorption time satisfies:* $\Theta_{\mathcal{L}}(\varepsilon) \leq \lceil e\frac{B^2}{\alpha}\rceil \lceil \ln(\frac{1}{\varepsilon})\rceil$*.*

Proof: See Appendix 2.

Therefore finding a coupling (X_t, Y_t) and a function δ such that $(*)$ (resp. $(**)$) holds allows us to prove that $\mathcal{A}$ is self-stabilizing and gives an upper bound on two different rates of convergence towards the (unique) ergodic set.

6 Refinement of Coupling

6.1 Path Coupling

As pointed out in [19], it is often cumbersome to measure the expected change in distance between two arbitrary configurations. The method of *path coupling*, introduced by Bubley and Dyer [3], simplifies the approach by showing that only pairs of configurations that are "close" need to be considered. Path coupling involves defining a coupling (X_t, Y_t) by considering a *path*, or sequence $X_t = Z_0, Z_1, \cdots, Z_r = Y_t$ between X_t and Y_t where the Z_i satisfy certain conditions. The following version of the path coupling method is convenient:

Lemma 1. (Dyer and Greenhill [8]) *Let* δ *be an integer valued metric defined on* $\Omega \times \Omega$ *which takes value in* $\{0, \cdots, B\}$*. Let* U *be a subset of* $\Omega \times \Omega$ *s.t., for all* $(X_t, Y_t) \in \Omega \times \Omega$*, there exists a path* $X_t = Z_0, Z_1, \cdots, Z_r = Y_t$ *between* X_t *and* Y_t *such that* $(Z_i, Z_{i+1}) \in U$ *for* $0 \leq i < r$ *and* $\sum_{i=0}^{r-1} \delta(Z_i, Z_{i+1}) = \delta(X_t, Y_t)$*.*

Suppose there exists a coupling $(X, Y) \mapsto (X', Y')$ *of the Markov chain* $\mathcal{A}$ *on all pairs* $(X, Y) \in U$*, and a constant* $\beta \leq 1$ *such that, for all* $(X, Y) \in U$*:*

$$E[\delta(X', Y')] \leq \beta\delta(X, Y). \qquad (* * *)$$

Then this coupling can be extended to a coupling of $\mathcal{A}$ *on all pairs* $(X, Y) \in \Omega \times \Omega$ *which also satisfies* $(* * *)$*.*

Two configurations X and Y are said to be *adjacent* if $(X, Y) \in U$. The advantage of this lemma is that it allows to check the crucial property $(* * *)$ only on the set U of adjacent pairs instead of the entire space $\Omega \times \Omega$. Lemma 1 combined with Theorem 2 (resp. Theorem 3) allows us to enhance our coupling method for proving self-stabilization.

6.2 Application to Herman

Let us come back to Herman's algorithm (see Example 1).

Theorem 4. *For Herman's algorithm and N odd, there exists a subset U of $\Omega \times \Omega$, an integer valued metric δ on $\Omega \times \Omega$ taking value in $\{0, \cdots, N\}$, and a coupling defined on U s.t.:* $\forall (X_t, Y_t) \in U \quad E[\delta(X_{t+1}, Y_{t+1})] \leq \delta(X_t, Y_t)$, *and* $\forall (X_t, Y_t) \in \Omega \times \Omega$ *(with $X_t \neq Y_t$)* : $Pr[\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)] \geq 1/2$.

Proof:

- *Metric δ.* We define δ as the Hamming distance, i.e.: $\delta(X_t, Y_t)$ is the number of positions at which X_t and Y_t differ. The couple (X_t, Y_t) belongs to U iff $\delta(X_t, Y_t) = 1$.
- *Coupling.* Coupling is such that if, for all i ($1 \leq i \leq N$), $X_t(i)$ and $Y_t(i)$ both perform *randomized* actions (i.e., when $X_t(i) = X_t(i-1)$ and $Y_t(i) = Y_t(i-1)$), the i-th machines of X_t and Y_t are forced to do the same probabilistic choice so that $X_{t+1}(i)$ and $Y_{t+1}(i)$ always coincide:

$$X_{t+1}(i) = Y_{t+1}(i) = \begin{cases} 0 \text{ with probability } 1/2, \\ 1 \text{ with probability } 1/2. \end{cases}$$

- *Proof of $E[\delta(X_{t+1}, Y_{t+1})] = \delta(X_t, Y_t)$ on U, and $Pr(\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)) \geq 1/2$ for all $(X_t, Y_t) \in \Omega \times \Omega$ with $X_t \neq Y_t$.*
  At each step, the state of all the machines at position $1, \cdots, N$ are updated. Let ℓ be the position of disagreement. In order to fix the ideas consider the following vector

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \nu_1 \; \nu_2 \; \cdots \; \nu_{\ell-2} \; 0 \; \mathbf{0} \; 0 \; \nu_{\ell+2} \; \cdots \; \nu_N \\ \nu_1 \; \nu_2 \; \cdots \; \nu_{\ell-2} \; 0 \; \mathbf{1} \; 0 \; \nu_{\ell+2} \; \cdots \; \nu_N \end{pmatrix}$$

  where all the ν_i are in $\{0, 1\}$, the figures in bold font correspond to positions ℓ. (The other cases are similar.) After one step, we have:

$$\begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix} = \begin{pmatrix} \nu'_1 \; \nu'_2 \; \cdots \; \nu'_{\ell-2} \; \nu'_{\ell-1} \; \mathbf{0/1} \; 0/1 \; \nu'_{\ell+2} \; \cdots \; \nu'_N \\ \nu'_1 \; \nu'_2 \; \cdots \; \nu'_{\ell-2} \; \nu'_{\ell-1} \;\; \mathbf{0} \;\;\; 1 \;\; \nu'_{\ell+2} \; \cdots \; \nu'_N \end{pmatrix}$$

  where '0/1' means "0 with prob. 1/2 and 1 with prob. 1/2". Note that, for $1 \leq i \leq \ell - 1$ and $\ell + 2 \leq i \leq N$, $X_{t+1}(i) = Y_{t+1}(i) = \nu'_i$ thanks to our coupling. So X_{t+1} and Y_{t+1} coincide everywhere except, perhaps, at positions ℓ or $\ell + 1$. We have:

$$\delta(X_{t+1}, Y_{t+1}) = \begin{cases} \delta(X_t, Y_t) = 1 \qquad \text{with probability } 1/2, \\ \delta(X_t, Y_t) - 1 = 0 \text{ with probability } 1/4, \\ \delta(X_t, Y_t) + 1 = 2 \text{ with probability } 1/4. \end{cases}$$

  Hence $E(\delta(X_{t+1}, Y_{t+1})) = \delta(X_t, Y_t)$, for all $(X_t, Y_t) \in U$. Furthermore, it is easy to show that $Pr(\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)) \geq 1/2$, for all $(X_t, Y_t) \in \Omega \times \Omega$ such that $X_t \neq Y_t$. □

Since $\delta(X_t, Y_t)$ takes its values in $\{0, 1, \cdots, N\}$, it then follows from Theorem 3, Lemma 1 and Theorem 4:

Corollary 1. *For N odd, Herman's algorithm $\mathcal{A}$ is such that:*

1. *$\mathcal{A}$ is self-stabilizing w.r.t. the set $\mathcal{L}$ of configurations with a single token.*
2. *The hitting time satisfies: $\mathbf{H}_{\mathcal{L}} \leq 2N^2$.*
3. *The ε-absorption time satisfies: $\Theta_{\mathcal{L}}(\varepsilon) \leq 2eN^2 \lceil \ln(\frac{1}{\varepsilon}) \rceil$.*

Note that the metric δ on $\Omega \times \Omega$ found here (Hamming distance) is much simpler than the function φ on Ω used by Herman, which involves the number of tokens of a configuration x together with the minimal distance between two tokens of x. Our method gives also directly an upper bound for the hitting time with no need for a separate analysis as done in Herman's work [10]. Besides, it gives a quadratic bound for the ε-absorption time (not considered by Herman).

The method can be applied in the same manner to several other self-stabilizing algorithms on cyclic graphs (e.g., mutual exclusion Flatebo-Datta's algorithm [9] with central randomized scheduler, Mayer-Ostrovsky-Yung's binary clock algorithm with synchronous scheduler [17]).

6.3 Application to Iterated Prisoner's Dilemma

Let us come back to Iterated Prisoner's Dilemma algorithm (Example 2). Recall that, in this case, the set $\mathcal{L}$ made of the unique configuration x^*, with $x^*(i) = +$ for all $1 \leq i \leq N$, is ergodic. Let us show that the algorithm is self-stabilizing.

Theorem 5. *For the Prisoner's Dilemma algorithm, there exist a subset U of $\Omega \times \Omega$, an integer valued metric δ on $\Omega \times \Omega$ taking value in $\{0, \cdots, 11N\}$, and a coupling defined on U such that, for all $(X_t, Y_t) \in U$:*

$$E[\delta(X_{t+1}, Y_{t+1})] \leq (1 - \tfrac{1}{18N})\delta(X_t, Y_t).$$

Proof:

- *Adjacent pairs.* A pair (X, Y) belongs to U iff X and Y coincide everywhere except on k contiguous positions, with $k = 1, 2, 3, 4$ or 5, where they disagree.
- *Metric δ.* Consider a pair $(X, Y) \in U$ which disagrees exactly at k contiguous positions ($1 \leq k \leq 5$). Let $\delta(X, Y) = a_k$ where a_k is a positive constant that will be determined later. By convention, we let $a_0 = 0$. The function δ on U extends to the entire space $\Omega \times \Omega$ as explained hereafter. Consider $(X, Y) \in \Omega \times \Omega$ such that X and Y differ only on ℓ contiguous positions. We have: $\ell = 5m + r$ for some $m \geq 0$ and $0 \leq r \leq 4$. The function δ is then defined by: $\delta(X, Y) = ma_5 + a_r$. Suppose that X and Y disagree on n separated zones of contiguous positions W_p $(1 \leq p \leq n)$. Let m_p and r_p be the quotient and the remainder of the length of W_p divided by 5 ($|W_p| = 5m_p + r_p$ with $0 \leq r_p \leq 4$). Then, for all $(X, Y) \in \Omega \times \Omega$, δ is defined by: $\delta(X, Y) = \sum_{p=1}^{n} m_p a_5 + a_{r_p}$. We will show later that, for appropriate values of a_k $(1 \leq k \leq 5)$, function δ is a metric which satisfies the conditions required by Lemma 1 (i.e., for all path $X = Z_0, Z_1, \cdots, Z_r = Y$ where $(Z_j, Z_{j+1}) \in U$, $\sum_{j=0}^{r-1} \delta(Z_j, Z_{j+1}) = \delta(X, Y)$).

- *Coupling.* The coupling $(X, Y) \mapsto (X', Y')$ is defined such that, at each step, the position chosen uniformly at random coincides for X and Y. (So, at each step, the state of the machine of the selected position, say j, and the state of the $j+1$-th machine are updated simultaneously in X and Y.)
- *Proof of $E[\delta(X', Y')] \leq \beta\delta(X, Y)$*: Consider a vector $(X_t, Y_t) \in U$ with k contiguous disagreeing positions. Let i be the first disagreeing position. The vector is of the form

  $$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \gamma_1 \cdots \gamma_{i-2}\, \gamma_{i-1}\;\; \gamma_i \;\;\cdots\;\; \gamma_{i+k-1}\;\; \gamma_{i+k} \cdots \gamma_N \\ \gamma_1 \cdots \gamma_{i-2}\, \gamma_{i-1}\, \neg\gamma_i \cdots \neg\gamma_{i+k-1}\, \gamma_{i+k} \cdots \gamma_N \end{pmatrix}$$

  where the γ_ℓ are in $\{-, +\}$. Suppose that the selected position j is such that $1 \leq j \leq i-2$ or $i+k \leq j \leq N$. Then $X_t(j) = Y_t(j)$ and $X_t(j+1) = Y_t(j+1)$, so $X_{t+1}(j) = Y_{t+1}(j)$ and $X_{t+1}(j+1) = Y_{t+1}(j+1)$, and the disagreement zone is not modified. Suppose now that the selected position j is equal to $i-1$. Then, after one step, we have:

  $$\begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix} = \begin{pmatrix} \gamma_1 \cdots \gamma_{i-2}\;\; \gamma'_{i-1}\;\; \gamma'_i \;\;\cdots\;\; \gamma_{i+k-1}\;\; \gamma_{i+k} \cdots \gamma_N \\ \gamma_1 \cdots \gamma_{i-2}\, \neg\gamma'_{i-1}\, \neg\gamma'_i \cdots \neg\gamma_{i+k-1}\, \gamma_{i+k} \cdots \gamma_N \end{pmatrix}$$

  where $\gamma'_{i-1} = \gamma'_i = +$ if $\gamma_{i-1} = \gamma_i$, and $\gamma'_{i-1} = \gamma'_i = -$ otherwise. This means that the disagreement zone has progressed on position at the left. A symmetrical case exists for $j = i+k-1$. We say that j is an "outer rim position". All the other possible cases for j are studied in Appendix 3. A simple case analysis shows that, for appropriate values of a_k $(1 \leq k \leq 5)$, there exists β with $\beta \leq 1 - \frac{1}{18N}$ such that $E[\delta(X', Y')] \leq \beta\delta(X, Y)$. Furthermore, for these values of a_k, the maximal value B of δ on $\Omega \times \Omega$ is such that $B \leq 11N$. See Appendix 3. □

Therefore, from Theorem 2, Lemma 1 and Theorem 5, it follows:

Corollary 2. *For Iterated Prisoner's Dilemma algorithm $\mathcal{A}$, we have:*
1. *$\mathcal{A}$ is self-stabilizing w.r.t. the set $\mathcal{L} = \{(+)^N\}$.*
2. *The hitting time satisfies: $\mathbf{H}_{\mathcal{L}} \leq 198N^2$.*
3. *The ε-absorption time satisfies: $\Theta_{\mathcal{L}}(\varepsilon) \leq 18N \ln(\frac{11N}{\varepsilon})$.*

Thus the quasi-linear bound on the ε-time of absorption is obtained, as found in [7]. Note that the linearity factor is better here (18 vs. 49/2). We retrieve also the quadratic bound on the hitting time found empirically in [13].

The proof presented here bears some resemblance with the proof by Dyer et al. in [7]: A function δ has been found here on $\Omega \times \Omega$ which satisfies $E\delta(X', Y') \leq \beta\delta(X, Y)$ (with $\beta < 1$), while they found a function φ on Ω satisfying $E\varphi(X') \leq \beta'\varphi(X)$ (with $\beta' < 1$) and $\varphi(\mathcal{L}) = 0$. Note that their function φ is somewhat simpler than δ (φ mainly involves isolated singletons $(-)$ and doublets $(--)$ while δ involves isolated sequences of disagreement of length up to 5). However, thanks to the path coupling method, it is easier to show the decrease in expectation for δ than for φ. Furthermore, we obtain here a better ε-absorption time $(\beta = 18 < \beta' = 49/2)$.

7 Final Remarks

We have shown that the method of coupling, which is classically used to evaluate the rate of convergence to equilibrium of Monte Carlo Markov chains, can be used to prove self-stabilization of distributed algorithms in an original manner. It allows us also to analyse the rate of convergence of these algorithms according to two different measures. The method has been enhanced by using the refinement of coupling, called "path coupling". This suggests to explore applications of the method using other refinements of coupling, such as Huber's bounding chain method [11]. We believe that our method still applies when the set $\mathcal{L}$ of legal states is not strongly connected, in the case where the various ergodic sets of $\mathcal{L}$ can be abstracted together using symmetries (e.g., in the case of randomized consensus protocols [2]). Finally, we plan to adapt our method on algorithms working with arbitrary schedulers (modelled by Markov decision proceses [18]) using, for example, the technique of scheduler-luck games (see [5]).

Acknowledgment: We are grateful to Alistair Sinclair for helpful discussions on an earlier draft of this paper.

References

1. A. Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graph.* draft at http:/www.stat.Berkeley.EDU/users/aldous/book.html, To appear.
2. J. Aspnes and M. Herlihy. Fast randomized consensus using shared memory. *Journal of Algorithms*, 11(3):441–461, 1990.
3. R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *Proc. of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS'97)*, pages 223–231, 1997.
4. E.W. Dijkstra. Self-stabilizing systems in spite of distributed control. *Communications of the ACM*, 17(11):643–644, Nov. 1974.
5. S. Dolev, A. Israeli, and S. Moran. Analyzing expected time by scheduler luck games. *IEEE transactions on Software Engineering*, 21(5):429–439, May 1995.
6. M. Duflot, L. Fribourg, and C. Picaronny. Randomized distributed algorithms as Markov chains. In *Proc. 15th Int. Conf. on Distributed Computing (DISC 2001), LNCS 2180*, pages 240–254. Springer, 2001.
7. M. Dyer, L.A. Goldberg, C. Greenhill, G. Istrate, and M. Jerrum. Convergence of the Iterated Prisoner's Dilemma Game. *Combinatorics, Probability and Computing*, 11(2), 2002.
8. M.E. Dyer and C. Greenhill. A more rapidly mixing Markov chain for graph colorings. *Random Structures and Algorithms*, 13:285–317, 1998.
9. M. Flatebo and A.K. Datta. Two-state self-stabilizing algorithms for token rings. *IEEE Transactions on Software Engineering*, 20(6):500–504, June 1994.
10. T. Herman. Probabilistic self-stabilization. *IPL*, 35(2):63–67, June 1990.
11. M. Huber. Exact sampling and approximate counting techniques. In *Proc. of the 30th Annual ACM Symp. on Theory of Computing (STOC'98)*, pages 31–40, 1998.
12. J.G. Kemeny and J.L. Snell. *Finite Markov Chains.* D. van Nostrand Co., 1969.

13. J.E. Kittock. Emergent conventions and the structure of multi-agent systems. In L. Nadel and D. Stein, editors, *Proc. of the 1993 Complex systems summer school.* Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley, 1995.
14. L. Lamport. Solved problems, unsolved problems and non-problems in concurrency. In *Proc. of the 3rd ACM Symp. on Principles of Distributed Computing (PODC'84)*, pages 1–11, 1984.
15. L. Lovász and P. Winkler. Mixing Times. *Microsurveys in Discrete Probability*, pages 85–134, 1998.
16. N.A. Lynch. *Distributed Algorithms.* Morgan Kaufmann Publishers, Inc., 1997.
17. A. Mayer, R. Ostrovsky, and M. Yung. Self-Stabilizing Algorithms for Synchronous Unidirectional Rings. In *Proc. of the 7th ACM-SIAM Symp. on Discrete Algorithms (SODA-96)*, 1996.
18. Martin L. Puterman. *Markov Decision Processes : Discrete Stochastic Dynamic Programming.* Wiley-Interscience, 1994.
19. D. Randall. Mixing. In *Proc. of the 44th Annual IEEE Symp. on Foundations of Computer Science (FOCS'03)*, 2003.
20. M. Schneider. Self-stabilization. *ACM Computing Surveys*, 25:45–67, 1993.
21. R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems.* PhD thesis, Massachusetts Institue of Technology, Jun. 1995.
22. A. Sinclair. Convergence rates for Monte Carlo experiments. In *Numerical Methods for Polymeric Systems*, pages 1–18. IMA Volumes in Mathematics & Its application, 1997.
23. D. Williams. *Probability with Martingales.* Cambridge University Press, 1991.

Appendix 1: Proof of Theorem 2

We will use the following Proposition that is easily proven by supermartingale theory, applying Doob's optional stopping Theorem (see for example [23]):

Proposition 3. *Suppose that* $D = (D_0, D_1, \cdots)$ *is a nonnegative stochastic process on* $\{0, 1, \cdots, B\}$ *such that* $E[D_{t+1}] \leq \beta D_t$ *(with* $0 < \beta < 1$*) when* $D_t > 0$. *Then if* τ *is the first time that* $D_t = 0$*, we have:* $E[\tau] \leq B/(1-\beta)$.

Proof of Theorem 2

1. By *reductio ad absurdum*: Suppose that $\mathcal{A}$ is not self-stabilizing. Then there are two non-empty ergodic sets $\mathcal{L}_1$ and $\mathcal{L}_2$ with two elements $x \in \mathcal{L}_1$ and $y \in \mathcal{L}_2$. Then, for all $t > 0$, $X_t \in \mathcal{L}_1$ and $Y_t \in \mathcal{L}_2$ since $\mathcal{L}_1$ and $\mathcal{L}_2$ are closed. Therefore for all $t > 0$, $X_t \neq Y_t$ since $\mathcal{L}_1$ and $\mathcal{L}_2$ are disjoint. Hence, for all $t > 0$, $\delta(X_t, Y_t) \geq 1$. Therefore $\forall t \; E(\delta(X_t, Y_t)) \geq 1$. On the other hand, we have: $\forall t > 0 \;\; E(\delta(X_t, Y_t) \leq \beta^t \delta(x, y)) \leq \beta^t B$. This leads to:
   $\forall t > 0 \; \beta^t B \geq 1$, which is false (e.g., for $t > \frac{\ln(B)}{\ln(1/\beta)}$).
2. Consider two elements $x, y \in \Omega$, and the coupling (X_t, Y_t) starting from $(X_0, Y_0) = (x, y)$. Let D_t be the process defined by $D_t = \delta(X_t, Y_t)$ for $t \geq 0$. Since $\delta(X_t, Y_t) = 0$ iff $X_t = Y_t$, the quantity $T_{x,y}$ is the time required for D_t to reach 0. Consider the coupling (X_t, Y_t) which starts from $(X_0, Y_0) = (x, y)$. Therefore by Proposition 3, we have, for all $x, y \in \Omega$, $E(T_{x,y}) \leq B/(1-\beta)$. Now, by Theorem 1 (statement 2), $\mathbf{H}_{\mathcal{L}} \leq max_{x,y} \; E(T_{x,y})$. Hence $\mathbf{H}_{\mathcal{L}} \leq B/(1-\beta)$.

3. Since $E(\delta(X_{t+1}, Y_{t+1})) \leq \beta\delta(X_t, Y_t)$, we have $E(\delta(X_t, Y_t)) \leq \beta^t\delta(X_0, Y_0) \leq \beta^t B$. But, by Markov's inequality ($P(X \geq a) \leq E[X]/a$): $Pr(\delta(X_t, Y_t) \geq 1) \leq E(\delta(X_t, Y_t))$. Hence, for all $X_0, Y_0 \in \Omega$ and all $t > 0$: $Pr(X_t \neq Y_t) = Pr(\delta(X_t, Y_t) > 0) = Pr(\delta(X_t, Y_t) \geq 1) \leq E(\delta(X_t, Y_t)) \leq \beta^t B$. Therefore, for all $X_0, Y_0 \in \Omega$ and all $t > 0$: $Pr(X_t = Y_t) \geq 1 - \beta^t B$. Suppose that $Y_0 \in \mathcal{L}$. Then $Y_t \in \mathcal{L}$ (because $\mathcal{L}$ closed), and $X_t = Y_t$ implies $X_t \in \mathcal{L}$. So, for all $X_0 \in \Omega$ and all $t > 0$: $Pr(X_t \in \mathcal{L}) \geq 1 - \beta^t B$. It follows that $Pr(X_t \in \mathcal{L}) \geq 1 - \varepsilon$, as soon as $\beta^t B \geq \varepsilon$, i.e., $t \geq \frac{\ln(B/\varepsilon)}{\ln(1/\beta)}$. Hence $Pr(X_t \in \mathcal{L}) \geq 1 - \varepsilon$, as soon as $t \geq \frac{\ln(B/\varepsilon)}{1-\beta}$ (because $1 - \beta \leq \ln(\frac{1}{\beta})$). □

Appendix 2: Proof of Theorem 3

The proof of Theorem 3 is analogous to that of Theorem 2, but relies on the following proposition (see for example [23]):

Proposition 4. *Suppose that $D = (D_0, D_1, \cdots)$ is a nonnegative stochastic process on $\{0, 1, \cdots, B\}$ such that $E[D_{t+1}] \leq D_t$ when $D_t > 0$. Furthermore suppose that $Pr(D_{t+1} \neq D_t) \geq \alpha$ (with $\alpha > 0$). Then if τ is the first time that $D_t = 0$, we have: $E[\tau] \leq B^2/\alpha$.*

Appendix 3: Proof of Theorem 5 ($E\delta(X', Y') \leq \beta\delta(X, Y)$)

Consider $(X, Y) \in U$. Let $[i, i + k - 1]$ be the interval of contiguous disagreeing positions between X and Y (with $1 \leq k \leq 5$). The random choice of the selected machine j modifies the zone of disagreement iff j corresponds to:

- *Outer rim position:* This means that $j = i - 1$ or $j = i + k - 1$. There are two outer rim positions for every $1 \leq k \leq 5$. Choosing an outer rim position extends the disagreement zone by one. This happens with probability $2/N$, and contributes to modify $E(\delta)$ by: $(a_{k+1} - a_k)(2/N)$. (For $k = 5$, $a_{k+1} = a_6$ stands for $a_5 + a_1$.)
- *Inner rim position:* This means that $j = i$ or $j = i + k - 2$. There are no inner rim position if $k = 1$, one inner rim position if $k = 2$, and two inner rim positions if $k = 3, 4, 5$. Choosing an inner rim position decreases the disagreement zone by two. This happens with probability $1/N$ (resp. $2/N$) when $k = 2$ (resp. $k = 3, 4, 5$). It contributes to modify $E(\delta)$ by $(a_0 - a_2)(1/N) = (-a_2)(1/N)$ when $k = 2$, and by $(a_{k-2} - a_k)(2/N)$ when $k = 3, 4, 5$.
- *Internal position:* This means that $j = i + 1$ or $j = i + k - 3$. There are no internal position if $k = 1, 2$ or 3, one internal position if $k = 4$, and two internal positions if $k = 5$. For $k = 4$, choosing an internal position ($j = i+1$) transforms the disagreement zone into two separated disagreement zones of length 1. This happens with probability $1/N$, and contributes to modify $E(\delta)$ by: $(2a_1 - a_4)(1/N)$. For $k = 5$, choosing an internal position ($j = i + 1$ or $j = i+2$) transforms the disagreement zone into two separated disagreement

zones of length 1 and 2. This happens with probability $2/N$, and contributes to modify $E(\delta)$ by: $(a_1 + a_2 - a_5)(2/N)$.

Accordingly, we have the following cases:

1. *Case k=1.* Then:
   $E(\delta(X', Y')) - \delta(X, Y) = (a_2 - a_1)(2/N)$.
   Hence $E(\delta(X', Y')) = \beta_1 \delta(X, Y)$ with $\beta_1 = (1 - \frac{2}{N}\frac{a_1 - a_2}{a_1})$
   (using the fact that $\delta(X, Y)$ is equal here to a_1).
2. *Case k=2.* Then:
   $E(\delta(X', Y')) - \delta(X, Y) = (2(a_3 - a_2) + (a_0 - a_2))(1/N) = (2a_3 - 3a_2)(1/N)$.
   Hence $E(\delta(X', Y')) = \beta_2 \delta(X, Y)$ with $\beta_2 = (1 - \frac{1}{N}\frac{3a_2 - 2a_3}{a_2})$
   (using the fact that $\delta(X, Y)$ is equal here to a_2).
3. *Case k=3.* Then:
   $E(\delta(X', Y')) - \delta(X, Y) = ((a_4 - a_3) + (a_1 - a_3))(2/N) = (a_4 - 2a_3 + a_1)(2/N)$.
   Hence $E(\delta(X', Y')) = \beta_3 \delta(X, Y)$ with $\beta_3 = (1 - \frac{2}{N}\frac{2a_3 - a_4 - a_1}{a_3})$
   (using the fact that $\delta(X, Y)$ is equal here to a_3).
4. *Case k=4.* Then:
   $$\begin{aligned} E(\delta(X', Y')) - \delta(X, Y) &= (2(a_5 - a_4) + 2(a_2 - a_4) + (2a_1 - a_4))(1/N) \\ &= (2a_5 - 5a_4 + 2a_2 + 2a_1)(1/N). \end{aligned}$$
   Hence $E(\delta(X', Y')) = \beta_4 \delta(X, Y)$ with $\beta_4 = (1 - \frac{1}{N}\frac{5a_4 - 2a_5 - 2a_2 - 2a_1}{a_4})$
   (using the fact that $\delta(X, Y)$ is equal here to a_4).
5. *Case k=5.* Then:
   $$\begin{aligned} E(\delta(X', Y')) - \delta(X, Y) &= ((a_5 + a_1 - a_5) + (a_3 - a_5) + (a_1 + a_2 - a_5))(2/N) \\ &= (-2a_5 + a_3 + a_2 + 2a_1)(2/N). \end{aligned}$$
   Hence $E(\delta(X', Y')) = \beta_5 \delta(X, Y)$ with $\beta_5 = (1 - \frac{2}{N}\frac{2a_5 - a_3 - a_2 - 2a_1}{a_5})$
   (using the fact that $\delta(X, Y)$ is equal here to a_5).

Therefore, for all $(X, Y) \in U$, $E(\delta(X', Y')) \leq \beta\delta(X, Y)$, with $\beta = max\{\beta_k\}_{1 \leq k \leq 5}$.
We have now to find $a_1, \cdots, a_5$ such that β satisfies $0 < \beta < 1$. A possible solution is: $a_1 = 21, a_2 = 20, a_3 = 29, a_4 = 36, a_5 = 48$.
It follows $\beta \leq 1 - (1/18N)$, hence $\frac{1}{1-\beta} \leq 18N$.
It remains to check that δ is a metric on $\Omega \times \Omega$, i.e.:

1. $\delta(X, Y) = 0$ iff $X = Y$.
2. $\forall x, y, z \ \delta(x, z) \leq \delta(x, y) + \delta(y, z)$.

The first item holds because all the coefficients a_i are positive.
The proof of the second item relies on the following fact: For all $k = 1, 2, 3, 4, 5$ and all partition $i_1, .., i_\ell$ of k (i.e: $i_1 + .. + i_\ell = k$), we have: $a_k < a_{i_1} + a_{i_2} + ... + a_{i_\ell}$.
Finally, it remains to check that, for all $x, y \in \Omega \times \Omega$, there exists a path $x = z_0, z_1, \cdots, z_r = y$ where $(z_j, z_{j+1}) \in U$, $\sum_{j=0}^{r-1} \delta(z_j, z_{j+1}) = \delta(x, y)$. For all x, y in Ω, we consider the path from x to y as follows: We first eliminate all the disagreement zones of length 5 (beginning at the leftmost site of disagreement of each zone), then all the disagreement zones of length 4, 3, 2 and finally 1. It comes from the definition of our metric that for this path $x = z_0, z_1, \cdots, z_k = y$, we have: $\sum_{j=0}^{k-1} \delta(z_j, z_{j+1}) = \delta(x, y)$.
Finally, let us note that the maximal value B of δ on $\Omega \times \Omega$ is at most
$a_1\lceil \frac{N}{2} \rceil \leq 11N$. □