# Greedy Coupling of 3-State Markov Chains

Iddo Ben-Ari, Mason DiCicco

### Abstract

In probability theory, *coupling* is a valuable proof technique often used to produce upper bounds on the total variation norm between distribution of two Markov Chains with the same transition function but different initial distributions by creating one joint process whose marginals are the respective processes. Without additional restrictions, there is a large diversity of coupling methods one can use. The desirability of a particular method is determined by various attributes, such as whether the joint process is itself a Markov chain, or how fast the marginal processes meet, or *couple*. In this work we review the theory, present some known and new qualitative criteria which help determine whether a given coupling provides sharp bounds. We then discuss the notion of *greedy* couplings, those maximizing the probability of meeting in the next step, and use the criteria presented to provide an exhaustive study of greedy couplings for the simple yet not entirely trivial case of 3-state Markov chains.

## 1 Background

### 1.1 Markov Chains

Suppose that $S$ is a finite or countable nonempty set. A (time-homogeneous) *Markov Chain* (MC) with *state space* $S$ is a sequence of random variables $\mathbf{X} = (X_0, X_1, \dots)$ taking values in $S$, such that the conditional distributions $\mathbb{P}(X_{n+1} = \cdot | X_0, X_1, \dots, X_n), n \in \mathbb{Z}_+$ is only a function of $X_n$. More precisely, there exists a function $p : S \times S \to [0, 1]$, the *transition function* (TF), representing these conditional distributions in the following sense:

$$\mathbb{P}(X_{n+1} = j | X_0, X_1, \dots, X_n) = p(X_n, j). \tag{1}$$

As the distribution of "the future" depends on the past only through the present state, this is often known as the *memoryless property* or the *Markov property*. Any transition function satisfies the following:

1. $p(i, j) \geq 0$ (non-negativity)

2. $\sum_j p(i, j) = 1$ (stochasticity)

The number $p(i, j)$ represents the transition probability from state $i$ to state $j$.

By iterating (1), it follows that the distribution of the MC $\mathbf{X}$ is determined by its transition function $p$, and the distribution of $X_0$, known as the *initial distribution of the process*.

Given any distribution on $S$ and a function $p$ satisfying 1. and 2. above, there always exists a MC $\mathbf{X}$ whose initial distribution is the one prescribed and whose transition function is $p$.

As there is no loss of generality assuming that $S$ takes the form $\{1, \ldots, n\}$, or $\{1, 2, \ldots\}$, for the remainder of this section we will make this assumption.

Transition functions can be identified with square matrices, with rows and columns indexed by the states. More precisely, if $p$ is a transition function on the state space $\{1, \ldots, n\}$, the corresponding *transition matrix* is the $n \times n$ matrix whose entry in the $i$-th row and $j$-th column is $p(i, j)$, the probability of a transition from state $i$ to state $j$. The matrix representation of $p$ is particularly useful because, by direct calculation, it follows from (1) that

$$\mathbb{P}(X_{n+m} = j | X_0, X_1, \ldots, X_n) = p^m(X_n, j), \ n, m \in \mathbb{Z}_+.$$

where $p^m$ is the $m$-th power of transition matrix. To ease notation, in the sequel we will write $p^0$ for the *identity matrix* on the state space, that is $p^0(i, i) = 1$ for all $i$ and $p^0(i, j) = 0$ when $j \neq i$.

If $\mathbf{X}$ is a MC with transition function $p$ and initial distribution $\mu$, we write $\mathbb{P}_\mu$ for the distribution of $\mathbf{X}$. The Markov property then gives us that for any $n \in \mathbb{Z}_+, i_0, \ldots, i_n \in S$

$$\mathbb{P}_\mu(X_0 = i_0, \ldots, X_n = i_n) = \mu(i_0)p(i_0, i_1) \cdots p(i_{n-1}, i_n).$$

In particular, the distribution of $X_n$ is given by the matrix product $\mu p^n = \sum_i \mu(i) p^n(i, \cdot)$, where here and henceforth a probability measure on $S$ will be considered as a row vector. When the initial distribution $\mu$ is a delta measure at state $x$, we write $\mathbb{P}_x$ for the corresponding distribution.

**Example 1.** A general 2-state Markov chain has the following transition matrix.

$$p = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix}.$$

A transition function corresponds to a weighted directed graph whose vertex set is the state space, and an edge from $i$ to $j$ exists iff $p(i, j) > 0$ with corresponding weight given by $p(i, j)$, leading to the following graphical representation of the chain from Example 1, assuming $a, b \in (0, 1)$ :
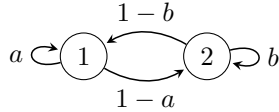


A *stationary distribution* for the TF $p$ is a probability distribution $\pi$ satisfying

$$\sum_i \pi(i)p(i, j) = \pi(j).$$

Thus, $\pi$ can be though of as a row vector, a left-eigenvector for the transition matrix $p$ with eigenvalue 1, all of whose entries are nonnegative and normalized to have sum 1. From its definition, we see that $\pi p^n = \pi$ for all $n \in \mathbb{Z}_+$. In terms of the corresponding MC, this means

$$\mathbb{P}_\pi(X_n = \cdot) = \pi, \ n \in \mathbb{Z}_+.$$

A TF (or the corresponding MC) is called *irreducible* if there is a path between every two states. More precisely, $p$ is irreducible if for any pair of distinct states $i, j$, there exists $m \in \mathbb{Z}_+$ such that $p^m(i, j) > 0$. A transition function is called *aperiodic* if for any state $i$, the greatest common divisor of the set of integers $I_i = \{m \in \mathbb{N} : p^m(i, i)\}$ is 1.

2

## 1.2 The Fundamental Theorem for Markov Chains

Let $S$ be a finite or countable set. The *total variation norm* between probability measures $\mu_1, \mu_2$ (or more generally two functions) on $S$ is defined as half the $\ell^1$-norm of $\mu_1 - \mu_2$:

$$
\begin{aligned}
\|\mu_1 - \mu_2\|_{TV} &= \frac{1}{2} \sum_{s \in S} |\mu_1(s) - \mu_2(s)| \\
&= \sum_{s \in S, \mu_1(s) > \mu_2(s)} \mu_1(s) - \mu_2(s) \\
&= \max_{A \subset S} (\mu_1(A) - \mu_2(A))
\end{aligned}
\tag{2}
$$

We also define

$$
d_t(\mu_1, \mu_2) = \|\mathbb{P}_{\mu_1}(X_t \in \cdot) - \mathbb{P}_{\mu_2}(X_t \in \cdot)\|_{TV}.
\tag{3}
$$

The triangle inequality for the $\ell^1$-norm then gives

$$
d_t(\mu_1, \mu_2) \leq \sup_{\mu_1(x) > 0, \mu_2(y) > 0} d_t(x, y).
$$

Another expressions that follows immediately from (2) is

$$
d_t(\mu_1, \mu_2) = 1 - \sum_{s \in S} \min(\mathbb{P}_{\mu_1}(X_t = s), \mathbb{P}_{\mu_2}(X_t = s)).
\tag{4}
$$

The following result is often known as the Fundamental Theorem of Markov Chains.

**Theorem 1.** *[LPW06, Theorem 4.9] Suppose that $p$ is an irreducible transition function on the finite state space $S$. Then*

1. *$p$ has a unique stationary distribution $\pi$.*

2. *If, in addition, $p$ is aperiodic, then there exists a constant $c > 0$ and $\rho < 1$ such that*

$$
d_t(x, \pi) \leq \max_x d_t(x, y) \leq c\rho^n
\tag{5}
$$

   *for all $t \in \mathbb{Z}_+$.*

**Theorem 2.** *Suppose that $p$ is a transition function on the finite state space $S$. Then*

1. *[HJ12, Theorem 8.3.4] All eigenvalues of $p$ are in the closed unit disk, and $1$ is an eigenvalue.*

2. *[HJ12, Theorem 8.4.4] If, in addition, $p$ is irreducible and aperiodic, then*

   (a) *$1$ is a simple eigenvalue.*

   (b) *All other eigenvalues are in the open unit disk.*

Let $\lambda_2$ denote the maximum among the norms of all eigenvalues of $p$ different than $1$ in Theorem 2. Then the Jordan decomposition for $p$ gives

**Proposition 1.** *Suppose that $p$ is an irreducible transition function on the finite state space $S$. Then*

1. *If $\rho$ satisfies (5), then $\rho \geq \lambda_2$.*

2. *Choosing $\rho = \lambda_2$ satisfies (5).*

Therefore the algebraic quantity $\lambda_2$ is the geometric rate of convergence in the fundamental theorem. In this work we study a purely probabilistic approach for estimating it, known as *coupling*.

# 2 Couplings

## 2.1 Couplings

Throughout this section we will fix a transition function $p$ on a finite or countable state space $S$.

**Definition 1** (Coupling). A *coupling* for $p$ is an $S \times S$-valued process $(X, Y)$ such that $X$ and $Y$ are each MC with transition function $p$. That is, for every $n$ and $j \in S$

$$\begin{cases} \mathbb{P}(X_{n+1} = j | X_n, X_{n-1}, \ldots, X_0) = p(X_n, j) \\ \mathbb{P}(Y_{n+1} = j | Y_n, Y_{n-1}, \ldots, Y_0) = p(Y_n, j) \end{cases}$$

The simplest example of a coupling is when we take $X = Y$. The second simplest may be $Y_n = X_{n+m}$ for some $m \in \mathbb{Z}_+$, and another "generic" example is when $Y$ is independent of $X$. Note that in the definition of a coupling we do not require the process $(X, Y)$ to be a Markov chain itself.

**Definition 2** (Coupling Time). Given a coupling $(X, Y)$ for $p$ define

1. The *coupling time* $\tau$ as
$$\tau = \inf\{t \in \mathbb{Z}_+ : X_t = Y_t\}, \tag{6}$$
with the infimum over the empty set defined as $+\infty$.

2. The coupling is called *sticky* if for all $t \in \mathbb{Z}_+$
$$X_t = Y_t \text{ on } \{\tau \geq t\}.$$

That is, a coupling is sticky if the two copies coalesce at the coupling time. We will return to these notions later and we will continue listing a number of important classes of couplings.

**Definition 3** (Markovian Coupling). A coupling for $p$ is *Markovian* if each of the marginal processes is a Markov chain with respect to the joint history. That is, for every $n$ and $j \in S$

$$\begin{cases} \mathbb{P}(X_{n+1} = j | (X_n, Y_n), (X_{n-1}, Y_{n-1}), \ldots, (X_0, Y_0)) = p(X_n, j) \\ \mathbb{P}(Y_{n+1} = j | (X_n, Y_n), (X_{n-1}, Y_{n-1}), \ldots, (X_0, Y_0)) = p(Y_n, j) \end{cases} \tag{7}$$

A Markovian coupling may not be itself a Markov chain, yet it it often that authors make this additional assumption. Conversely, a coupling which is a Markov chain may not be Markovian.

**Proposition 2.** *A coupling $(X, Y)$ for $p$ is Markovian if and only if for every $s \in \mathbb{Z}_+$, the distribution of the process $t \to (X_{s+t}, Y_{s+t})$, conditioned on $((X_0, Y_0), \ldots, (X_s, Y_s))$ is a.s. a coupling for $p$ with initial distribution $(X_s, Y_s)$*

*Proof.* Assume that the coupling is Markovian. Then the distribution of each of the Marginals, $t \to X_{s+t}$ and $t \to Y_{s+t}$, conditioned on $((X_0, Y_0), \ldots, (X_s, Y_s))$ is a MC with transition function $p$ and initial distribution $p$, and therefore the conditional process is a coupling.

Conversely, if the conditional process is a coupling, then (7) trivially holds. $\square$

**Definition 4.** (Faithful coupling [Ros97, DDB17]) A coupling for $p$ is *faithful* if it is both a Markov chain and Markovian. That is,

a)

$$\mathbb{P}((X_{n+1}, Y_{n+1}) = (x, y)|(X_n, Y_n), (X_{n-1}, Y_n), \ldots, (X_0, Y_0)) = \mathfrak{p}((X_n, Y_n), (x, y)) \quad (8)$$

from some transition function $\mathfrak{p}$ on $S \times S$.

b) Equation (7) holds. Under condition a), this is equivalent to

$$\begin{cases} \sum_{y \in S} \mathfrak{p}((x', y'), (x, y)) = p(x', x), \ x, x', y' \in S \\ \sum_{x \in S} \mathfrak{p}((x', y'), (x, y)) = p(y, y'), \ y, x', y' \in S. \end{cases} \quad (9)$$

We observe that every coupling where the two copies are independent is automatically faithful. Here are examples for couplings where one of the conditions in the definition fails.

**Example 2.** Suppose $X$ is already given. We define $Y_n = X_{n+1}$. Then $(X, Y)$ is a coupling for $p$. Observe that

$$\mathbb{P}((X_{n+1}, Y_{n+1}) = (x, y)|(X_n, Y_n), \ldots, (X_0, Y_0)) = \mathfrak{p}((X_n, Y_n), (x, y)),$$

where $\mathfrak{p}$ is the transition function on $S \times S$

$$\mathfrak{p}((x', y'), (x, y)) = \delta_{x, y'} p(y', y).$$

Therefore this coupling is a Markov chain, that is (8) holds. Summing over $y$, we have

$$\mathbb{P}(X_{n+1} = x|(X_n, Y_n), \ldots, (X_0, Y_0)) = \sum_{y_{n+1}} \delta_{x_{n+1}, Y_n} p(Y_n, y) = \delta_{x, Y_n}.$$

Thus, except for trivialities, the coupling is not faithful because (9) fails.

**Example 3.** Let $X_0 = 0$ and $Y_0 = 1$. Let $(U_n : n \in \mathbb{Z}_+)$ be IID $\mathrm{Bin}(\frac{1}{2})$. We continue according to the following algorithm. For $n \in \mathbb{Z}_+$, let $\hat{p}_n$ denote $\frac{1}{2(n+1)} \sum_{j \le n} (X_j + Y_j)$. Given $\hat{p}_n$, let $B_n$ be $\mathrm{Bin}(\hat{p}_n)$, and define

$$X_{n+1} = U_{2(n+1)} B_n + (1 - U_{2(n+1)})(1 - B_n) \text{ and}$$
$$Y_{n+1} = U_{2(n+1)} B_n + U_{2n+3}(1 - B_n)$$

Observe that since $\hat{p}_n$ is a function of $(X_0, Y_0), \ldots, (X_n, Y_n)$, it follows that

$$\mathbb{P}(X_{n+1} = 1|(X_n, Y_n), \ldots, (X_0, Y_0)) = \hat{p}_n \frac{1}{2} + (1 - \hat{p}_n)\frac{1}{2} = \frac{1}{2},$$

and that the same holds for $Y$. Therefore, (9) holds. Note, however, that $(X, Y)$ is not a Markov chain. Indeed,

$$\mathbb{P}(X_{n+1} = Y_{n+1} = 1|(X_n, Y_n), \ldots, (X_0, Y_0)) = \hat{p}_n \frac{1}{2} + (1 - \hat{p}_n)\frac{1}{4} = \frac{1 + \hat{p}_n}{4},$$

and $\hat{p}_n$ is not a function of $(X_n, Y_n)$ because the latter assumes at most 4 distinct values, while for $n$ large enough $\hat{p}_n$ can take an arbitrarily large number of values.

Here is a restatement of (9):

**Proposition 3.** *Suppose $(X, Y)$ is a coupling for $p$ which is also a Markov chain. Then the coupling is faithful if and only if for every $n$,*

- *$X_{n+1}$ and $Y_n$ are independent, conditioned on $X_n$; and*

- *$Y_{n+1}$ and $X_n$ are independent, conditioned on $Y_n$.*

*Proof.* We need to show that the two identities in (9) are equivalent to the the two independence statements. We will only prove this for the first pair, as the other is identical with the appropriate changes. The first equality in (9) can be rewritten as

$$p(x', x) = \sum_y \mathfrak{p}((x', y'), (x, y)) = \mathbb{P}(X_{n+1} = x | X_n = x', Y_n = y').$$

This is equivalent to

$$\mathbb{P}(X_{n+1} = x, Y_n = y' | X_n = x') = \frac{p(x', x) \mathbb{P}(X_n = x', Y_n = y')}{\mathbb{P}_{x_0}(X_n = x')}$$
$$= \mathbb{P}(X_{n+1} = x | X_n = x') \mathbb{P}(Y_n = y' | X_n = x'),$$

proving the equivalence. $\qquad\square$

**Definition 5** (Component Exchangeability)**.** A coupling which is a Markov chain with transition function $\mathfrak{p}$ is *component exchangeable* if for all $x, y, x', y' \in S$

$$\mathfrak{p}((x', y'), (x, y)) = \mathfrak{p}((y', x'), (y, x)).$$

In other words, the dynamics are invariant under swapping the components. Example 2 shows a coupling which is a Markov chain yet not component exchangeable, with the exception of trivialities.

**Proposition 4.** *Let $\mathfrak{p}$ be the transition function of a faithful coupling for $p$. Let*

$$\mathfrak{q}((x', y'), (x, y)) = \begin{cases} \mathfrak{p}((x', y'), (x, y)) & x' \neq y' \\ \delta_{x,y} p(x', x) & x' = y'. \end{cases} \tag{10}$$

*Then $\mathfrak{q}$ is the transition function of a sticky faithful coupling for $p$. If $\mathfrak{p}$ is component exchangeable, then so is $\mathfrak{q}$.*

This is an extremely important property of faithful couplings, and when working with couplings which are not faithful or are not sticky, some interesting things may happen, see for example [HM18].

As the proof follows directly from the definition of the faithful condition, we will omit it. We also note that often the procedure is performed path-wise. Suppose that $(X, Y)$ is a coupling corresponding to $\mathfrak{p}$, then letting

$$Y'_t = \begin{cases} Y_t & t \leq \tau \\ X_t & t > \tau \end{cases}$$

we obtain that $(X', Y')$ is a coupling corresponding to $\mathfrak{q}$.

6

## 2.2 Greedy couplings

**Definition 6.** A coupling $(X, Y)$ for $p$ is called greedy if at any step the probability of coupling is maximized. That is, for every $t \in \mathbb{Z}_+$ and $s \in S$

$$\mathbb{P}(X_{t+1} = Y_{t+1} | (X_t, Y_t), \ldots, (X_0, Y_0)) = \sum_{s \in S} \min\{p(X_t, s), p(Y_t, s)\}.$$

Observe that greedy couplings are always sticky, but not necessarily faithful nor are uniquely determined by the condition.

**Theorem 3.** *Let $p$ be a transition function on the state space $S$. Then there exists a faithful, component exchangeable coupling $(X, Y)$ for $p$ such that*

1. *The coupling is greedy.*

2. *Conditional on $\{\tau > n\}$, the component processes $(X_0, \ldots, X_n), (Y_0, \ldots, Y_n)$ are independent.*

*Furthermore, conditions 1,2 uniquely determine the transition function for the coupling.*

*Proof.* Let $(x_0, y_0) \in S \times S$ we will construct a coupling satisfying all the requirements with $(X_0, Y_0) = (x_0, y_0)$. For each $x, y, j \in S$, let $p_{x,y,j} = \min\{p(x, j), p(y, j)\}$, and let $p_{x,y} = \sum_j p(x, y, j)$. Set $(X_0, Y_0) = (x_0, y_0)$. Continue inductively. Suppose $(X_j, Y_j)$, $j \leq n$ have been defined. Let $U_{n+1} \sim \mathrm{U}[0, 1]$ be independent of them. Conditioning on $(X_n, Y_n) = (x, y)$, $(X_{n-1}, Y_{n-1}), \ldots, (X_0, Y_0)$ and $U_{n+1}$:

- let $X_{n+1} = Y_{n+1} = j$ if $U_{n+1} \in [0, p(x, y, j)) + \sum_{k < j} p(x, y, k)$,

- let $X_{n+1}$ and $Y_{n+1}$ move independently otherwise, that is if $U_{n+1} > p_{x,y}$, with conditional probabilities as follows.

$$\mathbb{P}(X_{n+1} = i, Y_{n+1} = j | (X_n, Y_n) = (x, y), (X_{n-1}, Y_{n-1}), \ldots, (X_0, Y_0), U_{n+1} > p(x, y))$$
$$= \frac{p(x, i) - p(x, y, i)}{1 - p(x, y)} \times \frac{p(x, j) - p(x, y, j)}{1 - p(x, y)}.$$
(11)

Then the construction guarantees that $(X, Y)$ is a Markov chain, and furthermore, as

$$\mathbb{P}(X_{n+1} = i | X_n = x, Y_n = y) = p(x, y, i) + \frac{p(x, i) - p(x, y, i)}{1 - p(x, y)}(1 - p(x, y)) = p(x, i),$$

with a similar expression for $Y_{n+1}$, the coupling is faithful.

Independence and component exchangeability are built into our construction. The uniqueness follows from induction whose details we omit. □

Here is why we are interested in greedy couplings.

**Theorem 4.** *Suppose that $(X, Y)$ is a faithful coupling for $p$ with coupling time $\tau$. Then there exists a greedy faithful coupling for $p$ with the same initial distributions as $(X, Y)$ and coupling time $\sigma$ which is stochastically dominated by $\tau$. That is,*

$$\mathbb{P}(\sigma > t) \leq \mathbb{P}(\tau > t).$$

*Furthermore, if $(X, Y)$ is component exchangeable then so is the greedy coupling.*

*Proof.* Without loss of generality, we assume that

- The probability space can support an IID sequence $(U_t : t \in \mathbb{N})$ of $U[0,1]$, independent of $(X,Y)$. We will construct a greedy coupling $(X',Y')$ based on $(X,Y)$ and which will have a coupling time less than or equal to the coupling time of $(X,Y)$, $\tau$.

- $(X_0, Y_0) = (x_0, y_0)$ with $x_0 \neq y_0$.

We now construct a greedy coupling $(X',Y')$ for $p$ satisfying the requirements. Let $(X'_0, Y'_0) = (x_0, y_0)$ and continue by induction as follows, with the induction assumption being:

1. $((X_s, Y_s) : s \leq t)$ is a greedy faithful coupling for $p$ with $(X_0, Y_0) = (x_0, y_0)$.

2. On the event $\{\tau > t\}$ either $X'_t = Y'_t$ or $(X'_s, Y'_s) = (X_s, Y_s)$ for all $s \leq t$.

3. On $\{\tau \leq t\}$, $X'_t = Y'_t$.

The induction hypothesis trivially holds for $t = 0$. Continue according to the following alternatives:

1. On $\{X'_t = Y'_t\}$ continue the coupling according to the sticky condition.

2. The remaining event is $\{\tau > t\} \cap \{(X'_t, Y'_t) = (X_t, Y_t)\}$. Here let $U_{t+1} \sim U[0,1]$, and for $i = 1, 2, \ldots$, let

$$p_i = p(X_t, i) \wedge p(Y_t, i) - p((X_t, Y_t), (i,i)),$$

independently of the past. Let

$$I_1 = [0, p_1], \ I_2 = (p_1, p_1 + p_2], \ \ldots, \ I_k = (p_1 + \cdots + p_{n-1}, p_1 + \cdots + p_n], \ldots$$

If $X_{t+1} = Y_{t+1}$, set $(X'_{t+1}, Y'_{t+1}) = (X_{t+1} = Y_{t+1})$. Otherwise, if

  (a) if $U_{t+1} \in I_i$ for some $i$, then set $X'_{t+1} = Y'_{t+1} = i$;
  (b) and if $U_{t+1} \notin \cup_{i=1}^{\infty} I_i$, then set $(X'_{t+1}, Y'_{t+1}) = (X_{t+1}, Y_{t+1})$.

Note that in either alternative, the conditional probability that $X'_{t+1} = i$ conditioned on the entire past of the process is $p(X'_t, i)$, with the analogous statement for $Y'_{t+1}$. Furthermore, the coupling is greedy by construction, so that the first statement in the induction hypothesis holds for $t + 1$. The last two also follow directly from the construction. This completes the induction.

As our algorithm is invariant under swapping the components, if $(X, Y)$ is component exchangeable, then so is $(X', Y')$. □

## 2.3 Maximality and Efficiency

Recall the definition of $d_t(\mu_1, \mu_2)$ from (3). Then for any coupling $(X, Y)$ with initial distributions $(\mu_1, \mu_2)$

$$
\begin{aligned}
d_t(\mu_1, \mu_2) &= \max_{A \subseteq S} \mathbb{E}[\mathbf{1}_A(X_t) - \mathbf{1}_A(Y_t)] \\
&= \sup_{f:S \to [0,1]} \mathbb{E}[f(X_t) - f(Y_t)] \\
&= \frac{1}{2} \sup_{f:S \to [-1,1]} \mathbb{E}[f(X_t) - f(Y_t)]
\end{aligned}
\tag{12}
$$

**Proposition 5.** *Let $(X, Y)$ be a coupling of two $S$-valued Markov chains with initial distributions $(\mu_1, \mu_2)$. Then*

1. *(Aldous' inequality) For every $t \in \mathbb{Z}_+$,*

$$d_t(\mu_1, \mu_2) \leq \mathbb{P}(X_t \neq Y_t) \tag{13}$$

2. *An equality in Aldous' inequality holds if and only if*

$$\mathbb{P}(Y_t \in A_t^X, X_t \neq Y_t) = 0,$$

   *where*

$$A_t^X = \{j : \mathbb{P}(X_t = j) \geq \mathbb{P}(Y_t = j)\}. \tag{14}$$

   *The analogous statement holds when replacing the roles of $X$ and $Y$.*

We comment that Proposition 5 does not make any assumption on the coupling.

*Proof.* From the definition,

$$d_t(\mu_1, \mu_2) = \mathbb{E}\left[\mathbf{1}_{A_t^X}(X_t) - \mathbf{1}_{A_t^X}(Y_t)\right].$$

In addition,

$$\mathbf{1}_{A_t^X}(X_t) - \mathbf{1}_{A_t^X}(Y_t) = (\mathbf{1}_{A_t^X}(X_t) - \mathbf{1}_{A_t^X}(Y_t))\mathbf{1}_{\{X_t \neq Y_t\}} \leq \mathbf{1}_{\{X_t \neq Y_t\}}.$$

This proves the first claim. As for the second claim. To prove the second claim, observe that the inequality is strict if and only if $X_t \neq Y_t$ and one of the following holds:

- $X_t \in A_t^X$ and $Y_t \in A_t^X$.

- $X_t \notin A_t^X$ and $Y_t \in A_t^X$.

Therefore the inequality is strict on the event $\{X_t \neq Y_t\} \cap \{Y_t \in A_t^X\}$, and in particular, the difference between the expectation of the two sides is strictly positive if and only if that event has positive probability. $\qquad\square$

**Definition 7** (Efficiency)**.** Let $(X, Y)$ be a coupling for $p$ with initial distributions $(\mu_1, \mu_2)$.

- The coupling is called *maximal* at time $t$ if $d_t(\mu_1, \mu_2) = \mathbb{P}(X_t \neq Y_t)$.

- The coupling is called *efficient* if there exists a constant $c > 0$ satisfying

$$\mathbb{P}(X_t \neq Y_t) \leq c\, d_t(\mu_1, \mu_2) \text{ for all } t \in \mathbb{Z}_+ \tag{15}$$

Maximal couplings always exist [Gri75, Lin92], but may be pretty complex and not tractable. The notion of efficient coupling was introduced in [BK00, Definition 2.2] for continuous-time chains.

Next, let

$$\begin{cases} B_t^X = (A_t^Y)^c = \{j : \mathbb{P}(X_t = j) > \mathbb{P}(Y_t = j)\} \\ B_t^Y = (A_t^X)^c = \{j : \mathbb{P}(Y_t = j) > \mathbb{P}(X_t = j)\} \end{cases}$$

9

From the Proposition, we have an equality in Aldous inequality if and only if

$$\mathbb{P}\left(\left(\{Y_t \in A_t^X\} \cup \{X_t \in A_t^Y\}\right) \cap \{X_t \neq Y_t\}\right) = 0.$$

Or,

$$\mathbb{P}\left(\left(\{Y_t \in B_t^Y\} \cap \{X_t \in B_t^X\}\right) \cup \{X_t = Y_t\}\right) = 1.$$

Since $\{Y_t \in B_t^Y\} \cap \{X_t \in B_t^X\} \subseteq \{Y_t \neq X_t\}$, this equality leads to

$$\{Y_t \in B_t^Y\} \cap \{X_t \in B_t^X\} = \{X_t \neq Y_t\} \ \mathbb{P}\text{-a.s.} \tag{16}$$

Now this implies that the distributions of $X_t$ and of $Y_t$ conditioned on $X_t \neq Y_t$ are supported on disjoint sets. Conversely, if the distributions of $X_t$ and $Y_t$ are supported on disjoint sets, we have equality in Aldous inequality. This leads to the following:

**Corollary 1.** *Let $(X, Y)$ be as in proposition 5. Then the following are equivalent:*

1. *The coupling is maximal at time $t$.*

2.
$$\{(X_t, Y_t) \in B_t^X \times B_t^Y\} = \{X_t \neq Y_t\}, \quad \mathbb{P}\text{-a.s.}$$

3. *The distributions of $X_t$ and $Y_t$, both conditioned on $\{X_t \neq Y_t\}$ are supported on disjoint sets.*

4.
$$\|\mathbb{P}(X_t \in \cdot | X_t \neq Y_t) - \mathbb{P}(Y_t \in \cdot | X_t \neq Y_t)\|_{TV} = 1.$$

The definition of efficiency of a coupling immediately leads to

**Corollary 2.** *The coupling is efficient if and only if*

$$\liminf_{t \to \infty} \|\mathbb{P}(X_t \in \cdot | X_t \neq Y_t) - \mathbb{P}(Y_t \in \cdot | X_t \neq Y_t)\|_{TV} > 0.$$

Finally, suppose that $(X, Y)$ is a sticky maximal coupling. In this case $\{X_t \neq Y_t\} = \{\tau > t\}$. Let

$$\tau_R = \inf\{t \in \mathbb{Z}_+ : (X_t, Y_t) \notin B_t^X \times B_t^Y\}. \tag{17}$$

Then (16) gives

$$\{\tau > t\} = \bigcap_{s=0}^{t} \{\tau > s\} = \bigcap_{s=0}^{t} \{(X_s, Y_s) \in B_s^X \times B_s^Y\} = \{\tau_R > t\},$$

leading to the following corollary:

**Corollary 3.** *Let $(X, Y)$ be a sticky coupling for p. Then the coupling is maximal for all $t \in \mathbb{Z}_+$, if and only if*

$$\tau = \tau_R, \ \mathbb{P}\text{-a.s.},$$

*where $\tau_R$ was defined in* (17).

## 2.4 Tests for Maximality and Efficiency

In this section we focus on tests for maximality and efficiency of a coupling based on coarse properties of the coupling. These tests are not new, and some have appeared in slightly different contexts before. The first result we present follows immediately from Theorem 4.

**Corollary 4.** *If $(X, Y)$ is a faithful coupling for $p$ which is maximal, then it is greedy.*

**Corollary 5.** *If for some $i \neq j$ and for some $t \in \mathbb{Z}_+$,*

$$\mathbb{P}(X_t = i, Y_t = j)\mathbb{P}(X_t = j, Y_t = i) > 0,$$

*then the coupling is not maximal at time $t$.*

Indeed, the stated condition violates condition 3 of Corollary 1. The converse of Corollary 5 is not true. The following example provides two couplings for a Markov chain with both being faithful and sticky, with the first being maximal and the other being efficient but not maximal, yet not satisfying the condition.

**Example 4.** Let $S = \{-1, 0, 1\}$, with

$$p = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

and where the rows and columns are listed in increasing order of the states. Let $X_0 = -1$ and $Y_0 = 1$. Here are two couplings, both satisfying $X_t < Y_t$ for all $t < \tau$, and therefore $\mathbb{P}(X_t = i, Y_t = j)\mathbb{P}(X_t = j, Y_t = i) = 0$ for all $i < j$.

1. A standard mirror coupling. Define $X_t$ for all $t$, and set $Y_t = -X_t$ until $X$ transitions to 0, and $Y_t = X_t$ afterwords. The coupling time $\mathbb{P}(\tau = t) \sim \text{Geom}(1/2)$, so its expectation is 2. Furthermore, this coupling is maximal because

$$\{X_t = -1, Y_t = 1\} = \{X_t \neq Y_t\} \ \mathbb{P}\text{-a.s.}$$

2. Copies are independent until they coupled or one unit apart. In the latter case, one copy is at 0 and another is at $\pm 1$, which we denote by $(0, \pm 1)$. We transition to $(\pm 1, \pm 1)$ with probability $1/2$ and to $(\mp 1, 0)$ with probability 1. That is couple with probability $1/2$, or "shift" the system, with probability $1/2$. The coupling time is the sum of $\text{Geom}(3/4)$, time until coupled or one apart, plus, with probability $2/3$, another $\text{Geom}(1/2)$. Therefore the expected coupling time is $8/3$. In light of the above, this cannot be maximal. Yet, this coupling is efficient, because the geometric tail for the coupling time is the same as for the maximal coupling constructed above.

A weak version of Corollary 1 for maximality of a coupling leads to the following sufficient condition for efficiency, which is a straightforward generalization of [BK00, Thereom 2.6-(ii)].

**Proposition 6.** *Suppose $p$ is a transition function on the finite state space $S$. If $(X, Y)$ is a sticky coupling for $p$ with the property that for each $t$ there exists some function $f_t : S \rightarrow \mathbb{R}$ such that $\{f_t(Y_t) > f_t(X_t)\} = \{\tau > t\}$, a.s. then the coupling is efficient.*

*Proof.* Since $S$ is finite, we may assume that for all $t$, the range of $f_t$ is contained in $[0,1]$, and that there exists some $\epsilon > 0$, such that on the event $\{\tau > t\}$, $f_t(Y_t) - f_t(X_t) > \epsilon$ for all $t$, a.s. Let $(\mu_1, \mu_2)$ denote the initial distributions of $(X, Y)$, respectively. Then

$$\mathbb{P}(\tau > t) \leq \frac{1}{\epsilon} E[f_t(Y_t) - f_t(X_t), \tau > t] \overset{(12)}{\leq} \frac{1}{\epsilon} d_t(\mu_1, \mu_2).$$

$\square$

Here is a condition that guarantees when a coupling is not efficient. This is a version of [BK00, Theorem 2.6-(i)] which was proven for continuous-time chains, where aperiodicity is not an issue. Here we bring a different proof based on an $h$-transform and the fundamental theorem for Markov chains.

**Theorem 5.** *Let $p$ be a transition function on the finite state space $S$, and let $\mathfrak{p}$ be the transition function of a component exchangeable coupling for $p$ which is a Markov chain. Suppose that the restriction of $\mathfrak{p}$ to $(S \times S) - D$ where $D = \{(s,s) : s \in S\}$, is aperiodic and irreducible. Then any coupling $(X,Y)$ with transition function $\mathfrak{p}$ and $(X_0, Y_0) = (x_0, y_0) \in (S \times S) - D$ is not efficient.*

*Proof.* Let $\mathfrak{p}_0$ be the restriction of $\mathfrak{p}$ to $(S \times S) - D$. Then by assumption $\mathfrak{p}_0$ is irreducible and aperiodic and therefore has a Perron root $\lambda \in (0,1)$ and a corresponding eigenvector $\phi$. Define

$$\mathfrak{q}(i,j) = \frac{1}{\phi(i)\lambda}\mathfrak{p}_0(i,j)\phi(j),$$

where here we use $i, j$ for generic elements in $(S \times S) - D$. Then $\mathfrak{q}$ is a bonafide irreducible and aperiodic transition function on $(S \times S) - D$.

Now let $(X, Y)$ be a coupling with transition function $\mathfrak{p}$, with $(X_0, Y_0) = (x_0, y_0)$, where $x_0 \neq y_0$. Then for $j \in (S \times S) - D$

$$\mathbb{P}((X_t, Y_t) = j, \tau > t) = \mathfrak{p}_0^t((x_0, y_0), j) = \lambda^t \phi((x_0, y_0))\mathfrak{q}^t((x_0, y_0), j)\frac{1}{\phi(j)}. \tag{18}$$

Let $\tilde{\phi}((y,x)) = \phi((x,y))$. Then using the fact that $\phi$ is a Perron eigenvector and the component exchageability, we have

$$\begin{aligned}
\lambda\tilde{\phi}((y_0, x_0)) &= \lambda\phi((x_0, y_0)) \\
&= \sum_{(x,y)} \mathfrak{p}_0((x_0, y_0), (x,y))\phi((x,y)) \\
&= \sum_{(y,x)} \mathfrak{p}_0((y_0, x_0), (y,x))\phi((x,y)) \\
&= \mathfrak{p}_0\tilde{\phi}((y_0, x_0)).
\end{aligned}$$

Therefore $\tilde{\phi}$ is also a Perron eigenvector. By uniqueness up to a constant factor, it therefore follows that $\tilde{\phi}((a,b)) = \phi((a,b))$. But $\tilde{\phi}((a,b)) = \phi((b,a))$, and we have have shown that $\phi((a,b)) = \phi((b,a))$. That is, $\mathfrak{q}$ is also component exchangeable. Let $\pi^{\mathfrak{q}}$ be the stationary distribution for $\mathfrak{q}$. A similar argument shows that it is also symmetric, that is $\pi^{\mathfrak{q}}((x,y)) = \pi^{\mathfrak{q}}((y,x))$. Now, from (18),

$$\mathbb{E}[f((X_t, Y_t)), \tau > t] = \lambda^t\phi((x_0, y_0))\mathbb{E}^{\mathfrak{q}}\big[\frac{f((X_t, Y_t))}{\phi((X_t, Y_t))}\big], \tag{19}$$

where $\mathbb{E}^{\mathfrak{q}}$ is the distribution of $(X, Y)$ under the transition function $\mathfrak{q}$. By the fundamental theorem for Markov chains,

$$\lim_{t \to \infty} \sup_{f, |f| \le 1} |\mathbb{E}^{\mathfrak{q}}[\frac{f((X_t, Y_t))}{\phi((X_t, Y_t))}] - \pi^{\mathfrak{q}}(f/\phi)| \to 0.$$

Now let $f((x, y)) = \mathbf{1}_A(x)$. By the above claimed component exchangeability, it follows that

$$\lim_{t \to \infty} \sup_A |\mathbb{E}^{\mathfrak{q}}[\frac{\mathbf{1}_A(X_t) - \mathbf{1}_A(Y_t)}{\phi(X_t, Y_t)}]| = 0. \tag{20}$$

Putting (20) into (19), we obtain

$$d_t(\mu_1, \mu_2) = \sup_A |\mathbb{P}(X_t \in A, \tau > t) - \mathbb{P}(Y_t \in A, \tau > t)| = \lambda^t o(1).$$

On the other hand, letting $f = \mathbf{1}$ in (19), we have

$$\mathbb{P}(\tau > t) \sim \lambda^t \phi(x_0, y_0) \pi^{\mathfrak{q}}(1/\phi),$$

and therefore $d_t(x_0, y_0) = \mathbb{P}(\tau > t) o(1)$, completing the proof. $\qquad\square$

The proof can be repeated verbatim to give the following result.

**Corollary 6.** *Let $p$ be a transition function on the finite state space $S$, and let $\mathfrak{p}$ be the transition function of a component exchangeable coupling for $p$ which is a Markov chain, and let $\mathfrak{p}_0$ be the restriction of $\mathfrak{p}$ to $(S \times S) - D$. suppose that $\mathcal{A} \subseteq (S \times S) - D$ is irreducible and aperiodic for $\mathfrak{p}_0$. Then any coupling $(X, Y)$ with transition function $\mathfrak{p}$ and $(X_0, Y_0) = (x_0, y_0) \in \mathcal{A}$ is not efficient.*

# 3 Efficiency for 3-state chains

## 3.1 Various Conditions for the Existence of Efficient Couplings

In this section we focus on the relatively simple case of 3-state chains. As a warmup, we begin with two-state chains, where a greedy coupling is always unique and maximal.

**Example 5.** Let $p$ be a transition function on $S = \{0, 1\}$, and let $(X, Y)$ be a greedy coupling with $(X_0, Y_0) = (i, 1-i)$. We show that the greedy coupling is unique and maximal. This is obvious if the two rows of $p$ are identical. Assume they are not. Let $j$ be such that $p(j, 0) > p(1 - j, 0)$, then necessarily $p(1 - j, 1) > p(j, 1)$, and therefore conditioning on not coupling by time $t + 1$, then the copy that was in $j$ at time $t$ will be in 0 at time $t + 1$ and the other copy will be in 1 at time $t + 1$. From Corollary 1, this guarantees that the coupling is maximal. As this event will occur with conditional probability

$$p(j, 0) - p(1-j, 0) = p(1-j, 1) - p(j, 1) = 1 - p(j, 0) \wedge p(1-j, 0) - p(j, 1) \wedge p(1-j, 1),$$

we also have

$$d_t(0, 1) = |p(1, 0) - p(0, 0)|^t = |1 - p_{1,1} - p_{0,0}|^t.$$

Next, we show that in some cases, no efficient Markovian coupling exists. The following is version of a result that appeared in [Swe18].

**Theorem 6.** *Suppose that $p$ is a symmetric transition function on the state space $S = \{1, 2, 3\}$ with $p_{1,3} = p_{2,2} = p_{3,1}$. Then*

1. *$p$ has an efficient Markovian coupling if and only if at least two entries in the first row of $p$ are identical, and in this case the greedy coupling is efficient.*

2. *Any greedy Markovian coupling for $p^2$ is maximal.*

We comment that the maximal coupling for $p^2$ from part 2 of the theorem can be lifted to obtain a maximal coupling for $p$ which is a Markov chain but not Markovian. We also comment that nonexistence results of this type have appeared for continuous-time processes in [BK00] and [CM00].

*Proof.* Assume $p$ is of the following form:

$$p = \begin{pmatrix} \alpha & \beta & \gamma \\ \beta & \gamma & \alpha \\ \gamma & \alpha & \beta \end{pmatrix}$$

Thus, the trace of $p$ is 1, and since 1 is an eigenvalue for $p$, it follows that the two remaining eigenvalues are $\lambda$ and $-\lambda$, and the determinant of $p$ is equal to $-\lambda^2$. Therefore the trace of $p^2$ is equal to $1 + 2\lambda^2$.

By the symmetry of $p$ and since the second and the third rows are the first row shifted, it follows that the three diagonal elements of $p^2$ are all equal to $\alpha^2 + \beta^2 + (1 - \alpha - \beta)^2$, and so we have two representations for the trace of $p^2$:

$$1 + 2\lambda^2 = 3(\alpha^2 + \beta^2 + (1 - \alpha - \beta)^2). \tag{21}$$

Suppose $(X, Y)$ is a Markovian coupling for $p$ with $(X_0, Y_0) = (x, y)$. Let $\tau$ be the coupling time. Then

$$\mathbb{P}(\tau > t + 1) = \sum_{x', y'} \mathbb{P}(\tau > t, X_t = x', Y_t = y', X_{t+1} \neq Y_{t+1})$$

$$= \sum_{x', y'} \mathbb{P}(X_{t+1} \neq Y_{t+1} | \tau > t, X_t = x', Y_t = y') \mathbb{P}(\tau > t, X_t = x', Y_t = y'), \tag{22}$$

where the summation is over $(x', y')$ such that $\mathbb{P}(\tau > t, X_t = x', Y_t = y') > 0$. Clearly,

$$\mathbb{P}(X_{t+1} \neq Y_{t+1} | \tau > t, X_t = x', Y_t = y') = 1 - \sum_{\ell} \mathbb{P}_{x,y}(X_{t+1} = Y_{t+1} = \ell | \tau > t, X_t = x', Y_t = y').$$

The event $\{X_{t+1} = Y_{t+1} = \ell\}$ is the intersection of the events $\{X_{t+1} = \ell\}$ and $\{Y_{t+1} = \ell\}$. By the Markovian property of the coupling, the probabilities of the latter two events conditioned on $\{\tau > t, X_t = x', Y_t = y'\}$, are $p(x', \ell)$ and $p(y', \ell)$, respectively. Note that we do not assume $(X, Y)$ to be a Markov chain: we only assume each component is a Markov chain with respect to the joint filtration. Therefore

$$\mathbb{P}(X_{t+1} = Y_{t+1} = \ell | \tau > t, X_t = x', Y_t = y') \leq \min(p(x', \ell), p(y', \ell)),$$

with equality if we choose a greedy coupling. From this we obtain

$$\mathbb{P}_{x,y}(X_{t+1} \neq Y_{t+1} | \tau > t, X_t = x', Y_t = y') \geq 1 - \sum_{\ell} \min\{p(x', \ell), p(y', \ell)\}.$$

14

Now let $a$ be the minimal element in the first row of $p$, and let $b$ be the maximal element in the first row and $c$ be the remaining element (note that any two may be equal).

In our case, for every choice of distinct $x', y'$, $a$ will appear twice in the sum on the righthand side and $c = 1 - a - b$ will appear once. Thus, the righthand side is equal to $2a + (1 - a - b) = a - b + 1$: $\mathbb{P}(\tau > t+1 | \tau > t, X_t = x', Y_t = y') = b - a$. Plugging this into (22), summing over $x', y'$, and induction give

$$\mathbb{P}(\tau > t+1) \geq (b-a)^{t+1}, \quad \text{for all } t \in \mathbb{Z}_+, \tag{23}$$

with equality if the coupling is greedy. Since by Aldous' inequality $\lambda \leq b - a$ with equality if and only if the coupling is efficient, it follows from (21) that letting

$$f(a, b) = 1 + 2(b-a)^2 - 3(a^2 + b^2 + (1 - a - b)^2),$$

then

$$f(a, b) \geq 0,$$

with equality if and only if the coupling is efficient. Now

$$
\begin{aligned}
0 &\leq 1 + (2a^2 + 2b^2 - 4ab) - 3a^2 - 3b^2 - 3c^2 \\
&= (1 - (a+b)^2) - 2ab - 3c^2 \\
&= (1 - a - b)(1 + a + b) - 2ab - 3c^2 \\
&= c(2 - c) - 2ab - 3c^2 \\
&= c(2 - 4c) - 2ab \\
&= 2c(1 - 2c) - 2ab \\
&= 2c(a + b - c) - 2ab \\
&= -2(c^2 - c(a+b) + ab) \\
&= -2(c - a)(c - b) \\
&= 2(c - a)(b - c)
\end{aligned}
$$

Therefore $f$ is equal to zero if and only if $c = a$ or $c = b$. Summarizing: a greedy Markovian coupling is efficient if and only if at least two of the entries $\alpha, \beta, \gamma$ are equal. This completes the proof of the first statement.

We turn to the second statement. Let $r_1, r_2, r_3$ be the rows of $p$. Then by symmetry and since all rows of $p^2$ are cyclic permutations of $r_1$, it follows that

$$p^2(i, j) = \begin{cases} \alpha' = r_1 \cdot r_2 = ab + bc + ca & i \neq j \\ \beta' = r_1 \cdot r_1 = a^2 + b^2 + c^2 & i = j \end{cases}$$

Furthermore, from Cauchy-Schwarz, $\alpha' \leq \beta'$. Consider any greedy Markovian coupling for $p^2$ with initial states $(x, y)$, $x \neq y$. Then from the structure of $p^2$ and the greedy condition, on $\{\tau > t\}$, $X_t = x$ and $Y_t = y$. It follows from Corollary 1 that the coupling is maximal. $\square$

**Example 6.** Let

$$p = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{pmatrix}$$

15

Then $p$ satisfies the conditions of Theorem 6. Consider any Markovian coupling $(X, Y)$ with $(X_0, Y_0) = (x_0, y_0)$, $x_0 \neq y_0$. Then if $X_t \neq Y_t$, the probability of coupling in time $t+1$ is bounded above by $a + a + c = \frac{1}{6} + \frac{1}{6} + \frac{1}{3} = \frac{2}{3}$. Therefore $\mathbb{P}(\tau > t) \geq (\frac{1}{3})^t$, with equality if and only if the coupling is greedy. As the determinant of $p$ is equal to $-\frac{1}{12}$, the eigenvalues of $p$ are $1, \pm\frac{1}{2\sqrt{3}}$, and since $\frac{1}{2\sqrt{3}} < \frac{1}{3}$, we have shown that

1. There does not exist an efficient Markovian coupling for $p$.

2. The coupling time of any Markovian coupling dominates a Geom(2/3) random variable, with equality if and only if the coupling is greedy.

## 3.2 Greedy and Faithful Couplings for 3-state Chains

We begin with a definition:

**Definition 8.** For distinct states $1 \leq i, j \leq 3$ define the set

$$\mathcal{M}_{ij} := \{k : p_{ik} > p_{jk}\} \subseteq \{1, 2, 3\},$$

and declare state $i$ as being dominated by state $j$, written as $i \lessdot j$, if $|\mathcal{M}_{ij}| = 1$.

Note that $|\mathcal{M}_{ij}| + |\mathcal{M}_{ji}| \in \{0, 2, 3\}$, where the sum is 0 if and only if $p_{ik} = p_{jk}$ for all $k$, equal to 2 if and only if there exists exactly one $k$ such that $p_{ik} = p_{jk}$, and is 3 when $p_{ik} \neq p_{jk}$ for all $k$. In particular, given any two rows $i, j$, then either they are equal, $i \lessdot j$ or $j \lessdot i$.

**Proposition 7.** *Let $p$ be a transition function on the state space $S = \{1, 2, 3\}$. Then all greedy and faithful couplings for $p$ share the same transition function given by the following formula. Let $(x_0, x_1)$ be a state.*

$$\mathfrak{p}((x_0, x_1), (i_0, i_1)) = \begin{cases} p(x_0, i) \wedge p(x_1, i) & i_0 = i_1 \\ (p(x_{1-n}, i_{1-n}) - p(x_n, i_{1-n}))_+ & x_n \lessdot x_{1-n} \text{ and } i_n \neq i_{n-1} \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

*Furthermore, $\mathfrak{p}$ is component exchangeable and sticky.*

We comment that if both $x_0 \lessdot x_1$ and $x_1 \lessdot x_0$, then the rows have exactly one $k$ such that $p(x_0, k) = p(x_1, k)$. In this case

$$(p(x_1, i_0) - p(x_0, i_0))_+ = (p(x_0, i_1) - p(x_0, i_1))_+,$$

and therefore the formula is well-defined.

In words, the formula means the following:

1. If, $x_0 \lessdot x_1$ or $x_1 \lessdot x_0$, say $x_0 \lessdot x_1$, with $p(x_0, a) > (x_1, a)$, then in the next step

   (a) the two copies will be coupled in the next step with probability $\sum_{i=1}^{3} p(x_0, i) \wedge p(x_1, i)$; or

   (b) with the remaining probability, $p(x_0, a) - p(x_1, a)$, the first component will move to $a$ and the second component will move to one of the remaining states $b, c$ with the leftover probabilities $p(x_1, b) - p(x_0, b)$ and $p(x_1, c) - p(x_0, c)$, respectively.

2. Otherwise (which can only happen if the rows $x_0$ and $x_1$ are identical), the two copies are necessarily coupled in the next step.

See Figure 1 for an illustration.

*Proof.* Let $\mathfrak{p}$ be the transition function of such a coupling.

Let's look at the $x$-th and the $y$-th row of $p$. Then

- A state $a$ such that $p(x,a) \geq p(y,a)$ and $p(x,j) \leq p(y,j)$ for $j \neq a$; or

- A state $b$ such that $p(y,b) \geq p(x,b)$ and $p(y,j) \leq p(x,j)$ for $j \neq b$.

It is possible that both hold. Let's continue assuming the former, and find all possible transitions from $(x,y)$. The other case is treated *mutatis mutandis*. Let $b, c$ be the two states different from $a$. Then

$$\mathfrak{p}((x,y),(b,a)) = 0, \ \mathfrak{p}((x,y),(b,b)) = p(x,b), \ \mathfrak{p}((x,y),(b,c)) = 0. \qquad (25)$$

Similarly,

$$\mathfrak{p}((x,y),(c,a)) = 0, \ \mathfrak{p}((x,y),(c,b)) = 0, \ \mathfrak{p}((x,y),(c,c)) = p(x,c). \qquad (26)$$

Now

$$p(y,c) = \mathfrak{p}((x,y),(a,c)) + \mathfrak{p}((x,y),(b,c)) + \mathfrak{p}((x,y),(c,c)) = \mathfrak{p}((x,y),(a,c)) + p(x,c),$$

and so

$$\mathfrak{p}((x,y),(a,c)) = p(y,c) - p(x,c).$$

Repeating for $b$ instead of $c$, we have

$$\mathfrak{p}((x,y),(a,b)) = p(y,b) - p(x,b).$$

That is

$$\begin{aligned}
\mathfrak{p}((x,y),(a,a)) &= p(y,a), \\
\mathfrak{p}((x,y),(a,b)) &= p(y,b) - p(x,b), \\
\mathfrak{p}((x,y),(a,c)) &= p(y,c) - p(x,c).
\end{aligned} \qquad (27)$$

Bottom line, if the first alternative holds, then

$$\mathfrak{p}((x,y),(i,j)) = \begin{cases} p(x,i) \wedge p(y,i) & i = j \\ (p(y,j) - p(x,j))_+ & j \neq i \end{cases}$$

$\square$

**Example 7.** Suppose that $p$ is a transition function on $S = \{1, 2, 3\}$ with rows $2, 3$ being identical. Let $\mathfrak{p}$ be the transition function from Proposition 7. Clearly, any coupling corresponding to $\mathfrak{p}$ starting from $(2,3)$ or $(3,2)$ has coupling time equal to 1, and is therefore maximal.

Suppose the coupling starts from $(1,2)$. We will show it is efficient. Observe that once both copies are in $\{2,3\}$, the process necessarily couples in one step. Each step, the probability of coupling or a transition to a state of the form $(2,3), (3,2)$ is therefore at least

17

$$\sum_{j=1,2,3} \mathfrak{p}((1,2),(j,j)) + \mathfrak{p}((1,2),(2,3)) + \mathfrak{p}((1,2),(3,2))$$

$$= p(3,3) - \mathfrak{p}((1,2),(1,3)) + p(2,2) - \mathfrak{p}((1,2),(1,2)) + \mathfrak{p}((1,2),(1,1))$$
$$= p_{3,3} + p_{2,2} - p_{1,1} + 2\mathfrak{p}((1,2),(1,1)).$$

But $\mathfrak{p}((1,2),(1,1)) = p(1,1) \wedge p(2,1)$. If $p(1,1) < p(2,1)$, then the expression we obtain is $\mathrm{Tr}(p)$. Otherwise, since $p(2,1) = 1 - p(2,2) - p(2,3)$, the expression is equal to $2 - p(2,2) - p(3,3) - p_{1,1} = 2 - \mathrm{Tr}(p)$.

Now the probability of either coupling or transitioning to a state of the form $(1,2),(2,1)$ again is in either case at most $|\mathrm{Tr}(p) - 1|$. As a result the probability of not leaving the set $\{(1,2),(2,1)\}$ by time $t$ is bounded above by $|\mathrm{Tr}(p) - 1|^{t-1}$. Since the coupling time is at less than or equal to the time the process leaves this set $+1$, we have that

$$\mathbb{P}(\tau > t) \le |\mathrm{Tr}(p) - 1|^{t-1}.$$

Since the eigenvalues of $p$ are 1,0 and the $\mathrm{Tr}(p) - 1$, this shows that the coupling is efficient.

We close this section with analysis of "almost all" 3-state Markov chains. We will work under the following assumptions:

**Assumption 1.** *1. $p$ is a transition function on the state space $\{1,2,3\}$.*

*2. all entries of $p$ are strictly positive.*

*3. the three entries on each column of $p$ are distinct.*

Since $p$ is stochastic and by Assumption 1 no two rows are identical, we have that for $i \ne j$, $\mathcal{M}_{ij}$ has 1 or 2 elements. As a result, $i \lessdot j$ or $j \lessdot i$ (both hold if and only if $p_{i,k} = p_{j,k}$ for exactly one $k$).

Note that the relation "$\lessdot$" is invariant under relabeling (permuting) the states and that it is not transitive.

Call our states $i, j, k$. Suppose $i \lessdot j$. Then either

- $i \lessdot j$ and $j \lessdot k$; or

- $i \lessdot j$ and $k \lessdot j$, in which case, either

  - $i \lessdot k$ and $k \lessdot j$; or
  - $k \lessdot i$ and $i \lessdot j$.

The first implies $i \lessdot k$ and $k \lessdot j$, while the latter implies $k \lessdot i$ and $i \lessdot j$.

Bottom line, we can label the states as $a, b, c$ so that

$$a \lessdot b \text{ and } b \lessdot c. \tag{28}$$

Under a labeling satisfying (28), there are two alternatives:

1. $a \lessdot b$, $b \lessdot c$, $c \lessdot a$. We call this a "**cycle**".

18

2. $a \ll b$, $b \ll c$, $a \ll c$, or "$c$ **is a dominating state**".

The main result of this section is the following:

**Theorem 7.** *Let $p$ be as in Assumption 1, and label the states so that (28) holds. Then the greedy, sticky faithful coupling for for $p$ is efficient for some initial points $(x_0, y_0)$ $x_0 \neq y_0$ if and only if*

1. *$\mathcal{M}_{ab} = b, \mathcal{M}_{bc} = a, \mathcal{M}_{ac} = b$, and the coupling is efficient but not maximal for all such $(x_0, y_0)$; or*

2. *If $\mathcal{M}_{bc} = \mathcal{M}_{ac} = \{c\}$, and the coupling is maximal whenever $\{x_0, y_0\} \neq \{a, b\}$.*

We will now prove the theorem. Unfortunately, our proof will be based on analysis of cases rather than an elegant argument.

*Proof.* We begin with an observation that greatly simplifies the structure of greedy couplings. Let $(X, Y)$ be a greedy coupling for $p$. Conditioning on $(X_t = i, Y_t = j)$ and $\{\tau > t + 1\}$, then necessarily $X_{t+1} \in \mathcal{M}_{i,j}$ and $Y_{t+1} \in \mathcal{M}_{j,i}$. In other words, for a greedy coupling,

$$\{\tau > t + 1\} \subseteq \{(X_{t+1}, Y_{t+1}) \in \mathcal{M}_{X_t, Y_t} \times \mathcal{M}_{Y_t, X_t}\}. \tag{29}$$

If the process is such that at time $t$, one copy is at $i$ and the other is at $j$, with $i \ll j$, the conditioning on not coupling at time $t + 1$, the copy in $i$ will move to $\mathcal{M}_{ij}$ and the copy in $j$ will move to one of the two remaining states $M_{ji}$.
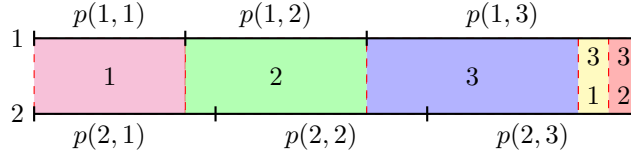


Figure 1: The graphical representation of the greedy Markovian coupling on three states starting from $(1, 2)$. Here, (29) follows from Assumption 1, which guarantees the "overlap", or leftover probability with which the process does not couple.

The diagrams below represent possible transitions for the couplings off the diagonal. Any arrow from $(i, j)$ to $(i', j')$ represents a nonzero transition probability. A solid arrow represents a transition from $(i, j)$ to $(i', j')$ and a dotted arrow represents a transition from $(i, j)$ to $(j', i')$. We chose to use the dotted arrows in order to keep the diagrams compact. We only listed states of the form $(i, j)$, with $i \ll j$. Note that we are not missing any transitions due to the assumed component-exchangeability.

In what follows, $\mathfrak{p}_0$ is the restriction of $\mathfrak{p}$, the transition function of the coupling, off the diagonal, that is to the set $\{1, 2, 3\}^2 - \{(1, 1), (2, 2), (3, 3)\}$.

**Dominating state**

Without loss of generality, we assume $1 \ll 2, 2 \ll 3$ and $1 \ll 3$.

Each such case corresponds to an ordered sequence $(s_1, s_2, s_3)$, representing the unique element in $\mathcal{M}_{12}, \mathcal{M}_{23}, \mathcal{M}_{13}$, respectively. Observe that

- $\mathcal{M}_{12} = \mathcal{M}_{23} = a \implies \mathcal{M}_{13} = a$. Indeed, $\mathcal{M}_{13}$ has a single element, and $p_{1,a} > p_{2,a} > p_{3,a}$.

- If $a \neq b$ and $\mathcal{M}_{12} = a$, $\mathcal{M}_{23} = b$, then $\mathcal{M}_{13} \in \{a, b\}$. Indeed, let $c$ denote the third state, then $p_{1,c} < p_{2,c} < p_{3,c}$. Therefore, $c \notin M_{13}$.

This leaves a total of $15 = 3 + 6 * 2$ cases, we will now review.

All diagrams in red correspond to $\mathfrak{p}_0$ being irreducible and aperiodic, and therefore the coupling is not efficient for any initial states $(x_0, y_0)$, $x_0 \neq y_0$, per Theorem 5.

As for the remaining four diagrams, let's consider first the three corresponding to $\mathcal{M}_{13} = \mathcal{M}_{23} = 3$. In each of the diagrams, the set of states $\mathcal{A} = \{(2,3), (3,2), (1,3), (3,1)\}$ forms an absorbing component for $\mathfrak{p}_0$. Furthermore, $\mathcal{A}$ has period 2. At each unit of time, one copy is at 3, alternating from being in first component to the second component. It therefore follows from Corollary 1 that for any of the initial states in $\mathcal{A}$ the coupling is maximal. When the initial state is not in $\mathcal{A}$, then maximality fails because as can be seen from the graphs, the support of $X_t$ and $Y_t$, conditioned on $\{\tau > t\}$ are not disjoint for $t \geq 3$.

ase is $\mathcal{M}_{12} = 2, \mathcal{M}_{23} = 1, \mathcal{M}_{13} = 2$. Here $\mathfrak{p}_0$ is irreduciblebut not aperiodic. Suppose that the initial state is $(1, 2)$. Then the transitions with strictly positive probability under $\mathfrak{p}_0$ are

1. In one step to $(2, 1)$ and to $(2, 3)$.

2. In two steps, to $(1, 2), (3, 2)$ and $(1, 3)$.

3. In three steps, to $(3, 1), (2, 3), (2, 1)$.

We observe that the supports of $X_t$ and $Y_t$, conditioned on $\{\tau > t\}$ are not disjoint (except for the triviality at times $t = 0$ and $t = 1$), and therefore the coupling fails to be maximal except at times 0 and 1 (where total variation distance between the distribution of $X_t$ and $Y_t$ is trivially equal to 1).

Nevertheless, letting

- $f_t(1) = 3, f_t(3) = 2, f_t(2) = 1$ for $t$ even; and

- $f_t(2) = 3, f_t(2) = 2, f_t(1) = 1$ for $t$ odd,

the condition on Proposition 6 holds for all $t$, which proves that the coupling is efficient. Due to irreducibility, similar treatment holds for all remaining initial pairs.

$(2,3)$

$(1,2)$

$(3,1)$

$\mathcal{M}_{12} = 1$
$\mathcal{M}_{23} = 3$
$\mathcal{M}_{13} = 3$

$(2,3)$

$(1,2)$

$(3,1)$

$\mathcal{M}_{12} = 2$
$\mathcal{M}_{23} = 1$
$\mathcal{M}_{13} = 1$

$(2,3)$

$(1,2)$

$(3,1)$

$\mathcal{M}_{12} = 2$
$\mathcal{M}_{23} = 1$
$\mathcal{M}_{13} = 2$

$(2,3)$

$(1,2)$

$(3,1)$

$\mathcal{M}_{12} = 2$
$\mathcal{M}_{23} = 2$
$\mathcal{M}_{13} = 2$

$(2,3)$

$(1,2)$

$(3,1)$

$\mathcal{M}_{12} = 2$
$\mathcal{M}_{23} = 3$
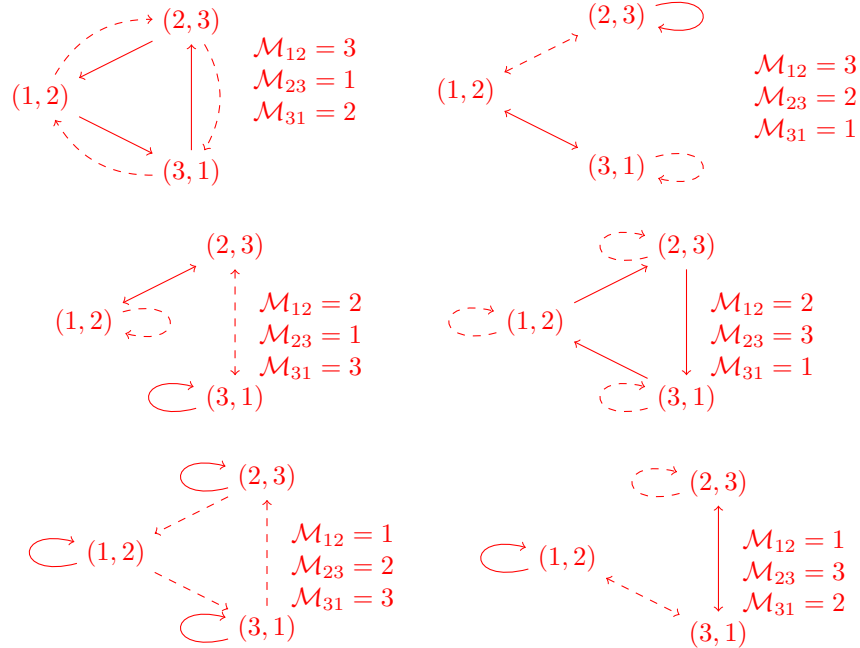$\mathcal{M}_{13} = 2$

$(2,3)$

$(1,2)$

$(3,1)$

$\mathcal{M}_{12} = 2$
$\mathcal{M}_{23} = 3$
$\mathcal{M}_{13} = 3$

$(2,3)$

$(1,2)$

$(3,1)$

$\mathcal{M}_{12} = 3$
$\mathcal{M}_{23} = 1$
$\mathcal{M}_{13} = 1$

$(2,3)$

$(1,2)$

$(3,1)$

$\mathcal{M}_{12} = 3$
$\mathcal{M}_{23} = 1$
$\mathcal{M}_{13} = 3$

$(2,3)$

$(1,2)$

$(3,1)$

$\mathcal{M}_{12} = 3$
$\mathcal{M}_{23} = 2$
$\mathcal{M}_{13} = 2$

$(2,3)$

$(1,2)$

$(3,1)$

$\mathcal{M}_{12} = 3$
$\mathcal{M}_{23} = 2$
$\mathcal{M}_{13} = 3$

$(2,3)$

$(1,2)$

$(3,1)$

$\mathcal{M}_{12} = 3$
$\mathcal{M}_{23} = 3$
$\mathcal{M}_{13} = 3$
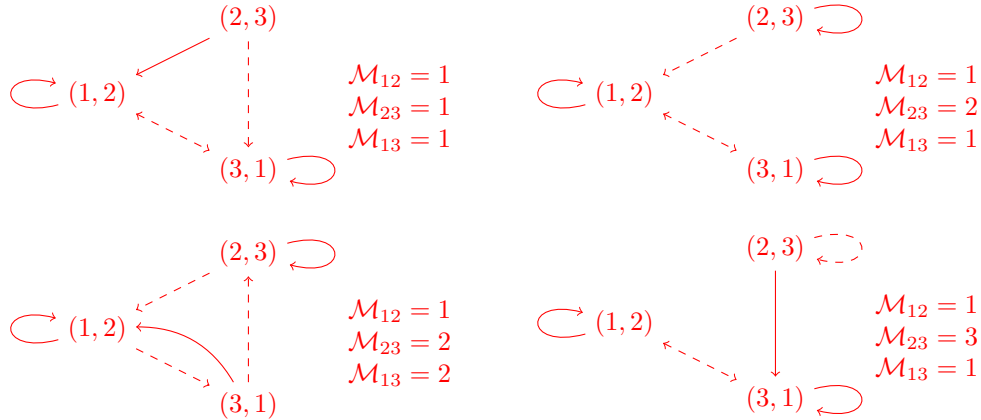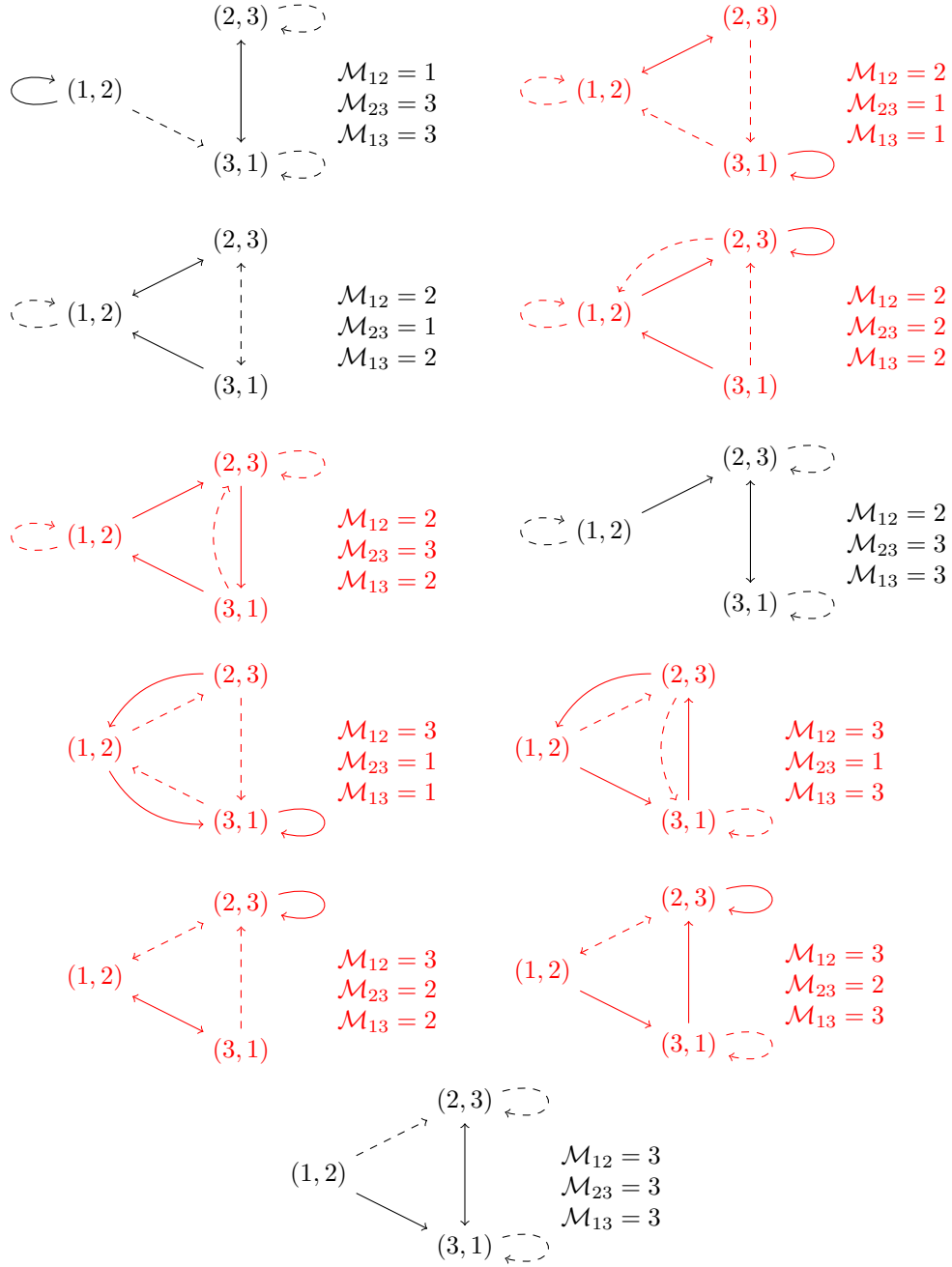
**Cycle**

Without loss of generality, $1 \lessdot 2, 2 \lessdot 3, 3 \lessdot 1$. Now if $\mathcal{M}_{12} = \mathcal{M}_{23} = i$, then $p_{1j} < p_{2j} < p_{3j}$, for $j \in \{1, 2, 3\} - \{i\}$. Therefore $1 \lessdot 3$, violating the assumption. Due to the cyclic structure, we conclude that $\mathcal{M}_{12}, \mathcal{M}_{23}, \mathcal{M}_{31}$ are distinct, reducing to the 6 permutations of

21

$\{1, 2, 3\}$. Below are the corresponding diagrams, all of which are irreducible and aperiodic, and therefore we conclude from Theorem 5 that the couplings are not efficient.

$$\mathcal{M}_{12} = 3 \quad \mathcal{M}_{23} = 1 \quad \mathcal{M}_{31} = 2$$

$$\mathcal{M}_{12} = 3 \quad \mathcal{M}_{23} = 2 \quad \mathcal{M}_{31} = 1$$

$$\mathcal{M}_{12} = 2 \quad \mathcal{M}_{23} = 1 \quad \mathcal{M}_{31} = 3$$

$$\mathcal{M}_{12} = 2 \quad \mathcal{M}_{23} = 3 \quad \mathcal{M}_{31} = 1$$

$$\mathcal{M}_{12} = 1 \quad \mathcal{M}_{23} = 2 \quad \mathcal{M}_{31} = 3$$

$$\mathcal{M}_{12} = 1 \quad \mathcal{M}_{23} = 3 \quad \mathcal{M}_{31} = 2$$

□

# 4 Acknowledgements

# References

[BK00]  Krzysztof Burdzy and Wilfrid S. Kendall. Efficient markovian couplings: Examples and counterexamples. *The Annals of Applied Probability*, 10(2):362–409, 2000.

[CM00]  M Cranston and T. S. Mountford. Efficient coupling on the circle. In *Game theory, optimal stopping, probability and statistics*, volume 35 of *IMS Lecture Notes Monogr. Ser.*, pages 191–203. Inst. Math. Statist., Beachwood, OH, 2000.

[DDB17] Debojyoti Dey, Pranjal Dutta, and Somenath Biswas. A note on faithful coupling of markov chains, 2017.

[Gri75] David Griffeath. A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:95–106, 1974/75.

[HJ12]     Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, USA, 2nd edition, 2012.

[HM18]     Timo Hirscher and Anders Martinsson. Segregating Markov chains. *J. Theoret. Probab.*, 31(3):1512–1538, 2018.

[Lin92]    Torgny Lindvall. Lectures on the coupling method. *New York: Wiley*, 1992.

[LPW06]  David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006.

[Ros97]    Jeffrey S. Rosenthal. Faithful couplings of Markov chains: now equals forever. *Adv. in Appl. Math.*, 18(3):372–381, 1997.

[Swe18]    Joseph Sweeney. Nonexistence of efficient markovian coupling for finite state markov chains. *Senior Thesis*, University of Connecticut, 2018.