



HAL
open science

Monotone coupling of Feller Markov processes on the real line

Jean Bérard, Brieuc Frénais

► **To cite this version:**

Jean Bérard, Brieuc Frénais. Monotone coupling of Feller Markov processes on the real line. 2023. hal-04309782v1

HAL Id: hal-04309782

<https://hal.science/hal-04309782v1>

Preprint submitted on 27 Nov 2023 (v1), last revised 23 Jan 2024 (v2)

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

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

Monotone coupling of Feller Markov processes on the real line

Jean Bérard and Briec Frénais

Abstract

We prove that, for any stochastically non-decreasing Feller Markov process on the real line, two versions of the process starting from distinct initial values always admit an order-preserving coupling, defined through a Feller Markov process on the set of ordered pairs.

1 Introduction

1.1 Statement of the main result

Consider a continuous-time homogeneous Markov process on the real line, characterized by a homogeneous family of Markov kernels¹ $(p_t)_{t \geq 0}$ on \mathbb{R} , and the corresponding Markov semi-group $(P_t)_{t \geq 0}$, which is defined on the space \mathcal{B}_b of bounded real-valued Borel functions on \mathbb{R} by²

$$\forall f \in \mathcal{B}_b \quad \forall t \geq 0 \quad \forall x \in \mathbb{R} \quad P_t f(x) = \int_{\mathbb{R}} f(y) p_t(x, dy).$$

If $(X_t)_{t \geq 0}$ is a family of random variables defined on a probability space $(\Omega, \mathcal{F}, (\mathbb{P}_x)_{x \in \mathbb{R}})$ providing a version of the process starting from $X_0 = x \in \mathbb{R}$, we thus have $p_t(x, \cdot) = \mathbb{P}_x(X_t \in \cdot)$, and $P_t f(x) = \mathbb{E}_x(f(X_t))$.

In addition to \mathcal{B}_b , let us denote by \mathcal{C}_b (resp. \mathcal{C}_0) the set of continuous bounded (resp. continuous vanishing at $\pm\infty$) real-valued functions on \mathbb{R} . On \mathcal{B}_b (and thus also \mathcal{C}_b and \mathcal{C}_0), we consider the supremum norm defined by $\|f\|_{\infty} = \sup_{x \in \mathbb{R}} |f(x)|$, with respect to which \mathcal{B}_b , \mathcal{C}_b and \mathcal{C}_0 are Banach spaces.

Recall that the semi-group $(P_t)_{t \geq 0}$ associated with $(p_t)_{t \geq 0}$ is said to enjoy the Feller property if:

- (a) $\forall t \geq 0, \forall f \in \mathcal{C}_0, P_t f \in \mathcal{C}_0$;
- (b) $\forall f \in \mathcal{C}_0, \lim_{t \rightarrow 0^+} \|P_t f - f\|_{\infty} = 0$.

Note that property (a) implies, but is not equivalent to

- (a') $\forall t \geq 0, \forall f \in \mathcal{C}_b, P_t f \in \mathcal{C}_b$,

with (a') often being termed "strong Feller property" in the literature (see e.g. [3]). Moreover, (b) may be replaced by the apparently weaker assumption of pointwise (instead of uniform) convergence

- (b') $\forall f \in \mathcal{C}_0, \forall x \in \mathbb{R}, \lim_{t \rightarrow 0^+} P_t f(x) = f(x)$,

¹Given a measurable space (E, \mathcal{E}) , a Markov kernel on E is a map $k : E \times \mathcal{E} \rightarrow \mathbb{R}$ such that (i) for all $x \in E$, $k(x, \cdot)$ is a probability measure on (E, \mathcal{E}) , and (ii) for all $B \in \mathcal{E}$, $k(\cdot, B)$ is a measurable real-valued function on E . Given two Markov kernels k, ℓ on E , the composition of the two kernels is yet another Markov kernel $k\ell$ defined by $(k\ell)(x, B) = \int_E k(x, dy)\ell(y, B)$. The composition is an associative (but in general non-commutative) operation on Markov kernels. We say that $(k_t)_{t \geq 0}$ is a homogeneous family of Markov kernels on E if (I) for all $x \in E$ and $B \in \mathcal{E}$, $k_0(x, B) = \delta_x(B)$, and (II) the Chapman-Kolmogorov equation holds: for all $s, t \geq 0$, $k_{s+t} = k_s k_t$.

²For all $f \in \mathcal{B}_b$, we denote by $P_t f$ the function $x \mapsto P_t f(x)$, and we have that $P_t f \in \mathcal{B}_b$. Moreover, P_t defines a linear operator from \mathcal{B}_b into itself, and satisfies (α) $\|P_t f\|_{\infty} \leq \|f\|_{\infty}$ (β) $P_t f \geq 0$ when $f \geq 0$ (γ) $P_t 1 = 1$. As a consequence of the Chapman-Kolmogorov equation for $(p_t)_{t \geq 0}$, the semi-group property holds: for all $s, t \geq 0$, $P_{s+t} = P_s P_t$.

but it turns out that a Markov semi-group satisfying (a)(b') also satisfies (b).

The Feller property has various useful consequences, among which the possibility of building versions of the corresponding Markov process enjoying "nice" properties: càdlàg trajectories, strong Markov property, quasi-left-continuity, etc. We refer e.g. to [4] (especially Chapter 19) for the key definitions and properties of Feller processes.

Then recall the definition of the stochastic dominance ordering between probability measures on \mathbb{R} . Denote by \mathcal{B}_b^\nearrow the set of non-decreasing bounded Borel functions on \mathbb{R} . Among several equivalent characterizations (see e.g. [7]), we say that $\mu \preceq \nu$ if and only if, for every $f \in \mathcal{B}_b^\nearrow$, one has that $\int f(x)\mu(dx) \leq \int f(x)\nu(dx)$. A celebrated result due to Strassen [9] states that the condition $\mu \preceq \nu$ is equivalent to the existence of a pair of random variables (Y, Z) such that $Y \sim \mu$, $Z \sim \nu$, and $Y \leq Z$ a.s.

Now we say that $(p_t)_{t \geq 0}$ is stochastically non-decreasing when, for all $t \geq 0$, and all $x, y \in \mathbb{R}$ such that $x \leq y$, one has $p_t(x, \cdot) \preceq p_t(y, \cdot)$. In terms of the semi-group, this condition rewrites as

$$(c) \quad \forall t \geq 0, \forall f \in \mathcal{B}_b^\nearrow, P_t f \in \mathcal{B}_b^\nearrow.$$

Thanks to Strassen's result, condition (c) also means that, starting from two initial positions $x^1 \leq x^2$, and for all $t \geq 0$, there exists a couple of random variables (X_t^1, X_t^2) such that X_t^1 and X_t^2 respectively follow the distribution of the Markov process at time t starting respectively from x^1 and from x^2 , and such that $X_t^1 \leq X_t^2$ a.s. The question we investigate in this paper is whether this statement can be strengthened to the effect that there exists a process of pairs $(X_t^1, X_t^2)_{t \geq 0}$ such that $(X_t^1)_{t \geq 0}$ and $(X_t^2)_{t \geq 0}$ respectively follow the distribution of the Markov process starting respectively from x^1 and from x^2 , and such that a.s. $\forall t \geq 0, X_t^1 \leq X_t^2$, while also enjoying "nice" properties. We shall consider the space of ordered pairs

$$S = \{(x^1, x^2) \in \mathbb{R}^2, x^1 \leq x^2\},$$

and show that it is indeed possible to build a Feller process of pairs on S , as we now explain.

Note that the definition and key properties of Feller processes that we have quoted above, are valid not just for \mathbb{R} -valued Markov processes, but in the general context of a locally compact separable metric space, which is clearly the case of S (note that S is also a complete metric space, as a closed subset of \mathbb{R}^2). One just has to replace the spaces $\mathcal{B}_b, \mathcal{C}_b, \mathcal{C}_0$ of real-valued functions defined on \mathbb{R} , by the corresponding spaces of real-valued functions defined on S , hereafter denoted by $\mathcal{B}_b(S), \mathcal{C}_b(S), \mathcal{C}_0(S)$.

Theorem 1 *If $(P_t)_{t \geq 0}$ enjoys the Feller property (a)(b) and the stochastic monotonicity property (c), there exists a homogeneous family of Markov kernels $(r_t)_{t \geq 0}$ on S , with a corresponding Markov semi-group $(R_t)_{t \geq 0}$, such that:*

- for all $t \geq 0$, all $x^1 \leq x^2$, and every Borel subset B of \mathbb{R} ,

$$\begin{cases} r_t((x_1, x_2), (B \times \mathbb{R}) \cap S) = p_t(x_1, B) \\ r_t((x_1, x_2), (\mathbb{R} \times B) \cap S) = p_t(x_2, B) \end{cases} ;$$

- the semi-group $(R_t)_{t \geq 0}$ enjoys the Feller property.

In terms of trajectories, Theorem 1 shows that, starting from any two initial positions $x^1 \leq x^2$, we can define a Markov process of pairs $(X_t^1, X_t^2)_{t \geq 0}$ on \mathbb{R}^2 such that $(X_t^1)_{t \geq 0}$ and $(X_t^2)_{t \geq 0}$ respectively follow the distribution of the Markov process starting respectively from x^1 and from x^2 , and such that a.s. $\forall t \geq 0, X_t^1 \leq X_t^2$. Moreover, the Feller property of $(R_t)_{t \geq 0}$ entails that the process of pairs can be constructed so as to have càdlàg paths, the strong Markov property, and quasi-left-continuity, these properties being inherited by the coordinate processes $(X_t^1)_{t \geq 0}$ and $(X_t^2)_{t \geq 0}$.

1.2 Discussion

Note the following two classical settings in which Assumption (c) is satisfied, and for which a process of pairs can be defined in a rather direct and explicit way:

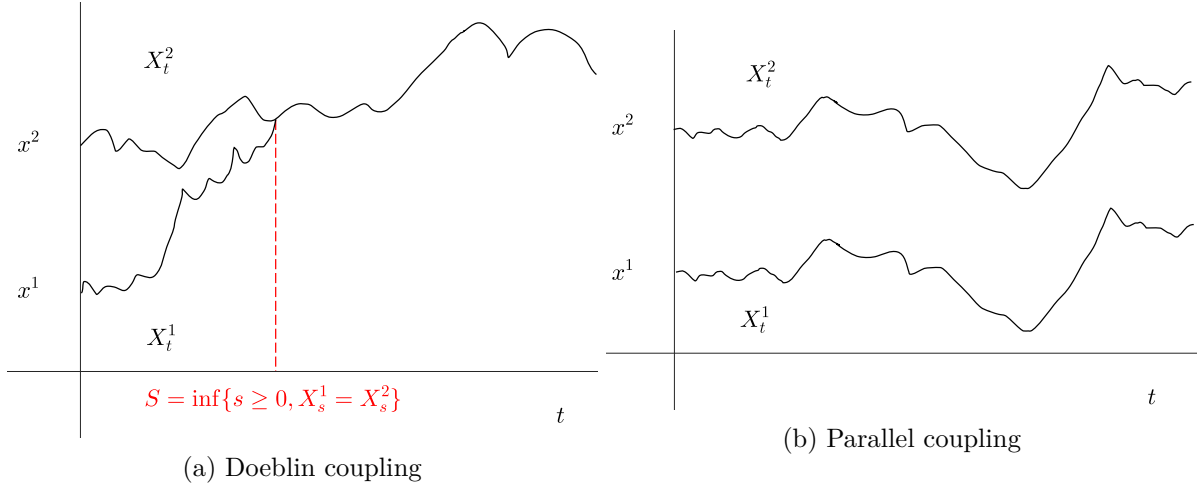


Figure 1: Special cases of Theorem 1 with easier constructions

- strong Markov processes with continuous paths; one can use the so-called *Doebelin* or *classical coupling*, in which the two processes evolve independently until they meet, and then stick together (see Figure 1a);
- space-homogeneous Markov processes (e.g. Lévy processes); one can use the *parallel coupling*, *i.e.* always keeping the same distance between both versions of the process (see Figure 1b).

In our view, the interest of Theorem 1 is that it holds in a completely general setting: the existence of a Feller process of ordered pairs on \mathbb{R}^2 is established under optimal assumptions. One may wonder whether weaker assumptions on the transition function $(p_t)_{t \geq 0}$ may lead to a (weaker) version of the conclusion of Theorem 1, but we do not pursue such a question here.

Note that a similar question has been considered by Kamae et al. ([5]), in the broader context of partially ordered Polish spaces, and with two possibly distinct Markov processes $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ being compared (not just two versions of the same process started at two distinct initial values) through the appropriate extension of condition (c). Specifically, Theorem 5 in [5] states that, provided that individual càdlàg versions of the two processes exist, it is possible to define càdlàg versions of both processes on the same probability space, in such a way that a.s. $\forall t \geq 0 X_t \leq Y_t$ (Feller-type properties are not considered in [5]). Unfortunately, we believe that a part of the proof of Theorem 5 in [5] is problematic, as explained in Section 3.

Let us also mention [10], which deals with the case of jump processes on partially ordered Polish spaces (relying on [1]), and [6], which provides a rather general sufficient criterion for a (suitably regular) one-dimensional Feller process to enjoy property (c).

Finally, note that, although we focused on a coupling between two versions of a Markov process, the arguments used to prove Theorem 1 can be immediately extended to deal with an ordered n -tuple of processes, with $n \geq 2$, instead of just a pair.

2 Proof of Theorem 1

Step 1. For any fixed dyadic time t , a sequence of Markov kernels $(q_t^{(n)})_{n \geq n_0}$.

Fix $t > 0$ and, for every $x \in \mathbb{R}$, denote by $F_{x,t}^{[-1]}$ the quantile function of X_t starting from x , *i.e.* for $u \in (0, 1)$

$$F_{x,t}^{[-1]}(u) = \inf\{y \in \mathbb{R} \text{ such that } u \leq \mathbb{P}_x(X_t \leq y)\}.$$

Then consider a random variable U with uniform distribution on $(0, 1)$. For every $x \in \mathbb{R}$, the distribution of the random variable $F_{x,t}^{[-1]}(U)$ is the distribution of X_t starting from x , i.e. $p_t(x, \cdot)$. Moreover, the stochastic monotonicity property (c) implies that, for every $u \in (0, 1)$, and every $x^1 \leq x^2$, one has $F_{x^1,t}^{[-1]}(u) \leq F_{x^2,t}^{[-1]}(u)$, so that $(F_{x^1,t}^{[-1]}(U), F_{x^2,t}^{[-1]}(U))$ is an S -valued random pair. Now, for every $(x^1, x^2) \in S$, we define $q_t((x^1, x^2), \cdot)$ as the probability distribution of $(F_{x^1,t}^{[-1]}(U), F_{x^2,t}^{[-1]}(U))$ on S , and we observe that its marginals coincide respectively with $p_t(x^1, \cdot)$ and $p_t(x^2, \cdot)$. We also define $Q_t f(x^1, x^2) = \int_S f(x'^1, x'^2) q_t((x^1, x^2), (dx'^1, dx'^2))$, for every $(x^1, x^2) \in S$ and every bounded real-valued Borel function f on S .

Our goal is now to check that for any Borel set B of S , the map $(x^1, x^2) \mapsto q_t((x^1, x^2), B)$ is Borel, so that q_t defines a Markov kernel on S . We start by checking that, if $f \in \mathcal{C}_b(S)$, then $Q_t f$ is a continuous function.

By the Feller property (a), if we take a sequence of real numbers $(x_k)_{k \geq 1}$ converging to $x \in \mathbb{R}$, $p_t(x_k, \cdot)$ converges weakly³ to $p_t(x, \cdot)$. Hence $F_{x_k,t}^{[-1]}(U)$ converges a.s. to $F_{x,t}^{[-1]}(U)$ (see for instance [2], Chapter 14, Proposition 5). As a consequence, if we take a sequence $(x_k^1, x_k^2)_{k \geq 1}$ converging to (x^1, x^2) in S , we have the a.s. convergence of $f(F_{x_k^1,t}^{[-1]}(U), F_{x_k^2,t}^{[-1]}(U))$ to $f(F_{x^1,t}^{[-1]}(U), F_{x^2,t}^{[-1]}(U))$, for every $f \in \mathcal{C}_b(S)$. By the dominated convergence theorem, this implies that $Q_t f(x_k^1, x_k^2)$ converges to $Q_t f(x^1, x^2)$, so we have proved that $Q_t f$ is a continuous function.

Now we apply Theorem 0.2.2 in [8], which is a functional version of the monotone class theorem: the set \mathcal{H} of bounded real-valued Borel functions on S such that $Q_t f$ is Borel contains $\mathcal{C}_b(S)$ – which plays the role of \mathcal{C} in [8] – and the assumptions of the theorem are met⁴, so that \mathcal{H} contains all $\sigma(\mathcal{C}_b(S))$ -measurable functions, and thus all bounded real-valued Borel functions on S . In particular $Q_t \mathbf{1}_B$ is a Borel function for any Borel set B in S , so q_t is indeed a Markov kernel on S .

Now we call D_+ the set of positive dyadic rational numbers. Given $t \in D_+$, we write $t = k2^{-n_0} \in D_+$ where $k \geq 1$ and $n_0 \geq 0$ are integers, and n_0 has the minimum possible value in such an expression. Then, for every integer $n \geq n_0$, we let

$$q_t^{(n)} = (q_{2^{-n}})^{k2^{n-n_0}},$$

i.e. $q_t^{(n)}$ is the repeated composition of kernels $q_{2^{-n}} \cdots q_{2^{-n}}$ with a total of $k2^{n-n_0}$ kernels in the composition. As a result, $(q_t^{(n)})_{n \geq n_0}$ is a sequence of Markov kernels on S . Moreover, thanks to the fact that, for every $(x^1, x^2) \in S$, the marginals of $q_{2^{-n}}((x^1, x^2), \cdot)$ coincide respectively with $p_{2^{-n}}(x^1, \cdot)$ and $p_{2^{-n}}(x^2, \cdot)$, and to the fact that, thanks to the Chapman-Kolmogorov equation, $p_t = (p_{2^{-n}})^{k2^{n-n_0}}$, we have that, for every $(x^1, x^2) \in S$ and $n \geq n_0$, the marginals of $q_t^{(n)}((x^1, x^2), \cdot)$ coincide respectively with $p_t(x^1, \cdot)$ and $p_t(x^2, \cdot)$.

In the sequel, we denote by $Q_t^{(n)} f$ the action of $q_t^{(n)}$ on the bounded real-valued Borel function f on S .

Step 2. Equicontinuity of $(Q_t^{(n)} f)_{n \geq 1}$.

For every $K > 0$, we define a function $\phi_K : \mathbb{R} \rightarrow \mathbb{R}$ by setting, for every $x \in \mathbb{R}$,

$$\phi_K(x) = -K \mathbf{1}_{\{x < -K\}} + x \mathbf{1}_{\{-K \leq x \leq K\}} + K \mathbf{1}_{\{x > K\}},$$

and we note that ϕ_K is non-decreasing, bounded (by K) and continuous. Now, for every bounded real-valued Lipschitz function f on S , we define a function f_K on S by setting, for every $(x^1, x^2) \in S$,

$$f_K(x^1, x^2) = f(\phi_K(x^1), \phi_K(x^2)).$$

³We call weak convergence the convergence in distribution for probability measures.

⁴The set \mathcal{H} is a vector space that contains the constant functions and the supremum of any bounded non-decreasing sequence of its non-negative elements (thanks to the monotone convergence theorem); the set \mathcal{C} is stable under pointwise multiplication.

We claim that, for fixed t , K and f , the sequence $(Q_t^{(n)} f_K)_{n \geq n_0}$ is equicontinuous: given $(x^1, x^2) \in S$, for any $\varepsilon > 0$, there exists $\eta > 0$ such that for all $n \geq n_0$ and $(x'^1, x'^2) \in S$ such that $|x'^1 - x^1| + |x'^2 - x^2| \leq \eta$, we have

$$|Q_t^{(n)} f_K(x'^1, x'^2) - Q_t^{(n)} f_K(x^1, x^2)| \leq \varepsilon. \quad (1)$$

To begin with, since f is a Lipschitz function, there exists a real number $L > 0$ such that for all $(x^1, x^2), (x'^1, x'^2) \in S$ we have

$$|f(x'^1, x'^2) - f(x^1, x^2)| \leq L(|x'^1 - x^1| + |x'^2 - x^2|).$$

Now, using a sequence $(U_1, \dots, U_{k2^{n-n_0}})$ of i.i.d. random variables with uniform distribution on $(0, 1)$, we build two pairs of random variables (X^1, X^2) and (X'^1, X'^2) whose distributions are respectively $q_t^{(n)}((x^1, x^2), \cdot)$ and $q_t^{(n)}((x'^1, x'^2), \cdot)$, and such that the ordering between x^1, x'^1, x^2, x'^2 is the same as the ordering between X^1, X'^1, X^2, X'^2 . To do so, define inductively

$$\begin{cases} (X_0^1, X_0^2) = (x^1, x^2) \\ (X_i^1, X_i^2) = (F_{X_{i-1}^1, 2^{-n}}^{[-1]}(U_i), F_{X_{i-1}^2, 2^{-n}}^{[-1]}(U_i)) \quad 1 \leq i \leq k2^{n-n_0} \end{cases}$$

and set $(X^1, X^2) = (X_{k2^{n-n_0}}^1, X_{k2^{n-n_0}}^2)$. Then make the same construction starting from $(X_0'^1, X_0'^2) = (x'^1, x'^2)$ to build $(X'^1, X'^2) = (X_{k2^{n-n_0}}'^1, X_{k2^{n-n_0}}'^2)$, using the same random variables U_i . It is apparent from the definition that both (X^1, X^2) and (X'^1, X'^2) have the desired distributions, since each step in the inductive construction above amounts to one Markov transition step according to the kernel $q_{2^{-n}}$. Moreover, we have already observed in Step 1 that, thanks to the stochastic monotonicity property (c), the map $x \mapsto F_{x,t}^{[-1]}(u)$ is non-decreasing for every given $u \in (0, 1)$. As a consequence, since we used the same U_i 's for both (X^1, X^2) and (X'^1, X'^2) , the initial ordering between $(X_0^1, X_0'^1, X_0^2, X_0'^2) = (x^1, x'^1, x^2, x'^2)$ is preserved at each step of the construction, and is thus identical to the ordering between (X^1, X'^1, X^2, X'^2) .

We now write

$$|Q_t^{(n)} f_K(x'^1, x'^2) - Q_t^{(n)} f_K(x^1, x^2)| = |\mathbb{E}(f_K(X'^1, X'^2) - f_K(X^1, X^2))|.$$

Using the triangle inequality and the Lipschitz property of f , we obtain

$$|Q_t^{(n)} f_K(x'^1, x'^2) - Q_t^{(n)} f_K(x^1, x^2)| \leq L\mathbb{E}(|\phi_K(X'^1) - \phi_K(X^1)| + |\phi_K(X'^2) - \phi_K(X^2)|). \quad (2)$$

Now by construction the sign of $X'^1 - X^1$, and hence of $\phi_K(X'^1) - \phi_K(X^1)$ since ϕ_K is non-decreasing, is the same as that of $x'^1 - x^1$, regardless of the randomness, so we can get the absolute value out of the expectation:

$$\mathbb{E}(|\phi_K(X'^1) - \phi_K(X^1)|) = |\mathbb{E}(\phi_K(X'^1) - \phi_K(X^1))| = |P_t \phi_K(x'^1) - P_t \phi_K(x^1)|,$$

and the same goes for X^2 and X'^2 . Plugging these identities into (2), we obtain that

$$|Q_t^{(n)} f_K(x'^1, x'^2) - Q_t^{(n)} f_K(x^1, x^2)| \leq L(|P_t \phi_K(x'^1) - P_t \phi_K(x^1)| + |P_t \phi_K(x'^2) - P_t \phi_K(x^2)|). \quad (3)$$

By the Feller property (a') for $(P_t)_{t \geq 0}$, since ϕ_K is in $\mathcal{C}_b(\mathbb{R})$, $P_t \phi_K$ is in $\mathcal{C}_b(\mathbb{R})$ too, so that, invoking continuity,

$$|P_t \phi_K(x'^1) - P_t \phi_K(x^1)| \leq \frac{\varepsilon}{2L}$$

as soon as x'^1 is close enough to x^1 . Doing the same with x'^2 and x^2 , we can bound the the r.h.s. in (3) independently from n by $\varepsilon > 0$, as soon as x'^1 and x'^2 are close enough to x^1 and x^2 (*i.e.* when the distance between them is smaller than a certain $\eta > 0$ given by the continuity of $P_t \phi_K$ at x^1 and x^2), which gives (1).

Step 3. Construction of a limiting transition kernel by diagonal extraction.

We have already observed that, given $t \in D_+$ and $(x^1, x^2) \in S$, the marginals of $q_t^{(n)}((x^1, x^2), \cdot)$ are respectively $p_t(x^1, \cdot)$ and $p_t(x^2, \cdot)$ for every $n \geq n_0$. From this observation, we deduce that the sequence $(q_t^{(n)}((x^1, x^2), \cdot))_{n \geq n_0}$ is tight⁵, allowing us to extract a weakly converging subsequence.

Now fix a countable dense subset D_S of S . By diagonal extraction, there exists an increasing sequence of integers $\varphi(n)$ such that for all $t \in D_+$ and $(x^1, x^2) \in D_S$, as $n \rightarrow +\infty$, the sequence $q_t^{(\varphi(n))}((x^1, x^2), \cdot)$ converges weakly to some probability measure $r_t((x^1, x^2), \cdot)$ on S . We shall prove that r_t can be extended to a limiting Markov kernel on the whole of S .

For any bounded real-valued Borel function f on S , and every $(x^1, x^2) \in D_S$ let

$$R_t f(x^1, x^2) = \int_S f(x'^1, x'^2) r_t((x^1, x^2), (dx'^1, dx'^2)).$$

Now take $t \in D_+$ and $(x^1, x^2) \in S$, fix $K > 0$ and a certain bounded Lipschitz function f as in Step 2, and let $\varepsilon > 0$. Since D_S is dense in S , we can find $(x'^1, x'^2) \in D_S$ such that $|x'^1 - x^1| + |x'^2 - x^2| \leq \eta$, where η is the modulus of equicontinuity at (x^1, x^2) given in Step 2 for the sequence $(Q_t^{(n)} f_K)_{n \geq n_0}$. Thus we have, for all $n \geq n_0$,

$$|Q_t^{(n)} f_K(x'^1, x'^2) - Q_t^{(n)} f_K(x^1, x^2)| \leq \varepsilon.$$

Now since (x'^1, x'^2) is in D_S , $Q_t^{(\varphi(n))} f_K(x'^1, x'^2)$ converges to $R_t f_K(x'^1, x'^2)$ as n goes to infinity, so for all large enough n we have that

$$|Q_t^{(\varphi(n))} f_K(x'^1, x'^2) - R_t f_K(x'^1, x'^2)| \leq \varepsilon,$$

so that, combining the two inequalities,

$$|Q_t^{(\varphi(n))} f_K(x^1, x^2) - R_t f_K(x^1, x^2)| \leq 2\varepsilon. \quad (4)$$

Now if we also take $m \geq n$, we have, using (4) twice and the triangle inequality

$$|Q_t^{(\varphi(n))} f_K(x^1, x^2) - Q_t^{(\varphi(m))} f_K(x^1, x^2)| \leq 4\varepsilon.$$

We have thus proved that the sequence $(Q_t^{(\varphi(n))} f_K(x^1, x^2))_{n \geq n_0}$ is Cauchy, and thus converges to a certain limit as n goes to infinity. We denote this limit by $R_t f_K(x^1, x^2)$, thus extending the definition of $R_t f_K$ to every $(x^1, x^2) \in S$ (when $(x^1, x^2) \in D_S$, we already know that the limit exists and is equal to $R_t f_K(x^1, x^2)$). As a consequence, any weak limit ν along a subsequence of $(q_t^{(\varphi(n))}((x^1, x^2), \cdot))_{n \geq n_0}$ has to satisfy $\int_S f_K(y^1, y^2) \nu(dy^1, dy^2) = R_t f_K(x^1, x^2)$. Since this is true for every $K > 0$ and bounded Lipschitz f , we deduce⁶ that there can be at most one such weak limit. Moreover, as we have already noted, the sequence $(q_t^{(\varphi(n))}((x^1, x^2), \cdot))_{n \geq n_0}$ is tight, so it indeed has a weak limit, which we denote by $r_t((x^1, x^2), \cdot)$, thus extending the definition of r_t to every $(x^1, x^2) \in S$. We extend the definition of R_t accordingly, so that $R_t f(x^1, x^2)$ is defined for every $(x^1, x^2) \in S$ and every bounded Borel real-valued function f on S .

To conclude that r_t defines a Markov kernel on S , we prove that the class of functions f such that $R_t f$ is Borel contains $\mathcal{C}_b(S)$. Given $f \in \mathcal{C}_b(S)$, weak convergence implies that $R_t f$ is the pointwise limit of the sequence of functions $Q_t^{(\varphi(n))} f$ as $n \rightarrow +\infty$. Each function in the sequence is Borel since f is bounded and Borel and $Q_t^{(\varphi(n))}$ is a Markov kernel, so the pointwise limit of the sequence is Borel. We then argue exactly as in Step 1.

⁵Let H_1, H_2 be compact subsets of \mathbb{R} such that $p_t(x^1, H_1^c) \leq \frac{\varepsilon}{2}$ and $p_t(x^2, H_2^c) \leq \frac{\varepsilon}{2}$ for a certain $\varepsilon > 0$. Then $H = (H_1 \times H_2) \cap S$ is compact and, by the union bound, $q_t^{(n)}((x^1, x^2), H^c) \leq p_t(x^1, H_1^c) + p_t(x^2, H_2^c) \leq \varepsilon$ for all $n \geq n_0$.

⁶Any bounded Lipschitz function with compact support can be written $f = f_K$ for a sufficiently large K . And the class of bounded Lipschitz functions with compact support separates Borel probability measures on S .

Step 4. $(R_t)_{t \in D_+}$ has the Feller property.

We have built for all $t \in D_+$ a Markov kernel r_t , and now we prove that $(R_t)_{t \in D_+}$ enjoys the characteristic properties of a Feller semi-group, restricted to indices $t \in D_+$.

- First we prove that, if $f \in \mathcal{C}_b(S)$, then $R_t f \in \mathcal{C}_b(S)$. Fix $\varepsilon > 0$ and $(x^1, x^2) \in S$. As in Steps 2 and 3, fix a bounded Lipschitz function f and $K > 0$, and take $\eta > 0$ a modulus of equicontinuity as in Step 2, meaning that for every $(x'^1, x'^2) \in S$ such that $|x'^1 - x^1| + |x'^2 - x^2| \leq \eta$, we have for all n such that $\varphi(n) \geq n_0$

$$|Q_t^{(\varphi(n))} f_K(x'^1, x'^2) - Q_t^{(\varphi(n))} f_K(x^1, x^2)| \leq \varepsilon.$$

Now, thanks to the weak convergence, we also have for n large enough (depending on x^1, x^2, t, f and K but not on x'^1, x'^2)

$$|Q_t^{(\varphi(n))} f_K(x^1, x^2) - R_t f_K(x^1, x^2)| \leq \varepsilon,$$

so that

$$|Q_t^{(\varphi(n))} f_K(x'^1, x'^2) - R_t f_K(x^1, x^2)| \leq 2\varepsilon, \quad (5)$$

and taking the limit $n \rightarrow +\infty$, we obtain

$$|R_t f_K(x'^1, x'^2) - R_t f_K(x^1, x^2)| \leq 2\varepsilon. \quad (6)$$

Thus, we have proved that $R_t f_K(x'^1, x'^2)$ converges to $R_t f_K(x^1, x^2)$ as (x'^1, x'^2) goes to (x^1, x^2) . Moreover, the class of functions of the form f_K , where $K > 0$ and f is a bounded Lipschitz function, characterizes weak convergence towards a probability measure⁷, so that $r_t((x'^1, x'^2), \cdot)$ converges weakly to $r_t((x^1, x^2), \cdot)$ as (x'^1, x'^2) goes to (x^1, x^2) . In particular, for $g \in \mathcal{C}_b(S)$, $R_t g(x'^1, x'^2)$ converges to $R_t g(x^1, x^2)$ as (x'^1, x'^2) goes to (x^1, x^2) . We have proved that $R_t g$ is continuous. It is also bounded by $\|g\|_\infty$, so it is in $\mathcal{C}_b(S)$.

- Now let us prove that if g is in $\mathcal{C}_0(S)$, then $R_t g$ is also in $\mathcal{C}_0(S)$. We already know it is continuous, so we just have to prove that it vanishes at infinity. First, fix $\varepsilon > 0$ and let $M > 0$ such that $|g|$ is bounded by ε on $S \setminus [-M, M]^2$ (M exists since g vanishes at $\pm\infty$). Then we have

$$|R_t g(x^1, x^2)| \leq \|g\|_\infty r_t((x^1, x^2), S \cap [-M, M]^2) + \varepsilon.$$

Now, since $(P_t)_t$ maps \mathcal{C}_0 into \mathcal{C}_0 , and r_t 's marginals are respectively $p_t(x^1, \cdot)$ and $p_t(x^2, \cdot)$, $r_t((x^1, x^2), S \cap [-M, M]^2)$ goes to zero as x^1 and x^2 go to infinity⁸, so we can make $|R_t g(x^1, x^2)|$ arbitrarily small by taking (x^1, x^2) far enough from the origin, and thus $R_t g$ is in $\mathcal{C}_0(S)$.

- Let us prove the semi-group property. First, recall $t = k_0 2^{-n_0}$ and also take $s = l_0 2^{-m_0}$ another dyadic time. If n is such that $\varphi(n) \geq \max(n_0, m_0)$, we have by construction

$$Q_s^{(\varphi(n))} Q_t^{(\varphi(n))} = Q_{s+t}^{(\varphi(n))}.$$

To prove that $R_s R_t = R_{s+t}$ we will prove that the corresponding operators coincide on the functions f_K , with f Lipschitz and $K > 0$. Fix $M > 0$. We have for every $g, h \in \mathcal{C}_b(S)$ bounded by a same constant $C > 0$

$$\begin{aligned} & |Q_s^{(\varphi(n))} g(x^1, x^2) - Q_s^{(\varphi(n))} h(x^1, x^2)| \\ & \leq \|g - h\|_{\infty, S \cap [-M, M]^2} + 2C q_s^{(\varphi(n))}((x^1, x^2), S \setminus [-M, M]^2), \end{aligned} \quad (7)$$

⁷Any bounded Lipschitz function with compact support can be written $f = f_K$ for a sufficiently large K . And the class of bounded Lipschitz functions with compact support characterizes weak convergence towards a Borel probability measure on S .

⁸We have $r_t((x^1, x^2), S \cap [-M, M]^2) \leq p_t(x^1, [-M, M])$, and, given a function $f \in \mathcal{C}_0(\mathbb{R})$ such that $f \geq \mathbf{1}_{[-M, M]}$, $P_t f$ vanishes at infinity.

where $\|\cdot\|_{\infty, B}$ is the supremum norm on functions restricted to a domain B . Using the fact that the sequence of probability measures $(q_s^{(\varphi(n))}((x^1, x^2), \cdot))_{n \in \mathbb{N}}$ is tight since it converges weakly, we can make the second term in the r.h.s. of (7) arbitrarily small by taking M large enough. Now if we take $g = Q_t^{(\varphi(n))} f_K$ and $h = R_t f_K$, both bounded by $\|f\|_{\infty}$, we will see that we can also make the first term as small as we want. Indeed, looking back at (5) and (6), we have that, for all $(y^1, y^2) \in S$, and all $\varepsilon > 0$, there exists an $\eta > 0$ such that, for all large enough n and (y^1, y^2) such that $|y^1 - y^1| + |y^2 - y^2| \leq \eta$,

$$|Q_t^{(\varphi(n))} f_K(y^1, y^2) - R_t f_K(y^1, y^2)| \leq 4\varepsilon,$$

which means that the convergence of $Q_t^{(\varphi(n))} f_K$ to $R_t f_K$ is locally uniform around (y^1, y^2) , and thus uniform on any compact subset of S . Taking $S \cap [-M, M]^2$ for this compact, we can make (7) arbitrarily small, namely

$$|Q_s^{(\varphi(n))} Q_t^{(\varphi(n))} f_K(x^1, x^2) - Q_s^{(\varphi(n))} R_t f_K(x^1, x^2)| \leq \varepsilon \quad (8)$$

for n large enough.

Now since $R_t f_K$ is in $\mathcal{C}_b(S)$ (because f_K is) and $r_s((x^1, x^2), \cdot)$ is the weak limit of $q_s^{(\varphi(n))}$, $Q_s^{(\varphi(n))} R_t f_K(x^1, x^2)$ converges to $R_s R_t f_K(x^1, x^2)$, so that $Q_s^{(\varphi(n))} Q_t^{(\varphi(n))} f_K(x^1, x^2)$ also converges to $R_s R_t f_K(x^1, x^2)$ thanks to (8). Furthermore, since $Q_s^{(\varphi(n))} Q_t^{(\varphi(n))} = Q_{s+t}^{(\varphi(n))}$, and since $s+t$ is also dyadic, the weak convergence of $q_{s+t}^{(\varphi(n))}((x^1, x^2), \cdot)$ ensures that $Q_s^{(\varphi(n))} Q_t^{(\varphi(n))} f_K(x^1, x^2)$ converges to $R_{s+t} f_K(x^1, x^2)$, hence $R_s R_t f_K(x^1, x^2) = R_{s+t} f_K(x^1, x^2)$. Using again the fact that the functions f_K separate probability measures on S , we deduce that $r_s r_t = r_{s+t}$ and that $R_s R_t = R_{s+t}$.

- Finally we prove the uniform convergence of $R_t f$ to f , for any $f \in \mathcal{C}_0(S)$, as t goes to 0 (for $t \in D_+$). Fix $\varepsilon > 0$ and η a modulus of uniform continuity for f , *i.e.* for all $(x^1, x^2), (x^1, x^2)$ in S such that $|x^1 - x^1| + |x^2 - x^2| \leq \eta$, we have $|f(x^1, x^2) - f(x^1, x^2)| \leq \varepsilon$. Also fix $M > 0$ such that $|f| < \varepsilon$ on $S \setminus [-M, M]^2$. We want to dominate $|R_t f(x^1, x^2) - f(x^1, x^2)|$ uniformly in (x^1, x^2) .
 - Assume that (x^1, x^2) is outside $[-(M+1), M+1]^2$. Since $|f| \leq \varepsilon$ outside $[-M, M]^2$,

$$\begin{aligned} |R_t f(x^1, x^2) - f(x^1, x^2)| &\leq \varepsilon + |R_t f(x^1, x^2)| \\ &\leq 2\varepsilon + R_t(|f| \mathbf{1}_{[-M, M]^2})(x^1, x^2) \\ &\leq 2\varepsilon + \|f\|_{\infty} p_t(x^1, [-M, M]). \end{aligned}$$

Now observe that the Feller property for P_t used on the function h depicted in Figure 2a proves that $p_t(x, [-M, M])$ converges to zero as $t \rightarrow 0$, uniformly for x outside $[-(M+1), M+1]$: indeed, $p_t(x, [-M, M])$ is smaller than $P_t h(x)$, which converges uniformly to $h(x) = 0$. We deduce that $|R_t f(x^1, x^2) - f(x^1, x^2)|$ is smaller than 3ε for t small enough, uniformly for $(x^1, x^2) \notin [-(M+1), M+1]^2$.

- Assume that (x^1, x^2) is in $[-(M+1), M+1]^2$. Denote (X_t^1, X_t^2) a couple of random variables whose distribution is $r_t((x^1, x^2), \cdot)$, and, for $x \in \mathbb{R}$ and $\delta > 0$, denote by $I_\delta(x)$ the interval $[x - \delta, x + \delta]$. We have

$$\begin{aligned} |R_t f(x^1, x^2) - f(x^1, x^2)| &\leq \mathbb{E}(|f(X_t^1, X_t^2) - f(x^1, x^2)|) \\ &\leq \varepsilon + 2\|f\|_{\infty} \mathbb{P}(|X_t^1 - x^1| + |X_t^2 - x^2| > \eta) \\ &\leq \varepsilon + 2\|f\|_{\infty} \left(\mathbb{P}\left(|X_t^1 - x^1| > \frac{\eta}{2}\right) + \mathbb{P}\left(|X_t^2 - x^2| > \frac{\eta}{2}\right) \right) \\ &= \varepsilon + 2\|f\|_{\infty} \left(p_t(x^1, I_{\eta/2}^c(x^1)) + p_t(x^2, I_{\eta/2}^c(x^2)) \right). \end{aligned}$$

Now, we only need to prove that $p_t(x, I_{\eta/2}^c(x))$ converges to zero uniformly on the interval $[-(M+1), M+1]$. To do so, we prove that this convergence is locally

uniform around every x in \mathbb{R} , so that it is uniform on every compact subset of \mathbb{R} . To this end, we apply the Feller property for P_t to the function h depicted in Figure 2b. Since $h \leq \mathbf{1}_{I_{2\delta}(x)}$, we have for $x' \in I_\delta(x)$, $P_t h(x') \leq p_t(x', I_{2\delta}(x)) \leq p_t(x', I_{3\delta}(x'))$. Since $P_t h$ converges uniformly to h as t goes to zero, $P_t h(x')$ converges to 1 uniformly for $x' \in I_\delta(x)$, hence so does $p_t(x', I_{3\delta}(x'))$. Taking $\delta = \frac{\eta}{6}$ allows us to complete the proof of uniform convergence of $R_t f$ to f as t goes to zero.

Step 5. Extension of $(R_t)_{t \in D_+}$ to all times.

Let $t > 0$ be a non-dyadic time, and $s, s' \in D_+$ such that $t < s < s'$. Also fix $f \in \mathcal{C}_0(S)$. Thanks to the semi-group property for dyadic times, we have

$$\|R_{s'}f - R_s f\|_\infty \leq \|R_s(R_{s'-s}f - f)\|_\infty \leq \|R_{s'-s}f - f\|_\infty,$$

so $(R_s f)_{s>t}$ satisfies the Cauchy condition with respect to the $\|\cdot\|_\infty$ norm as s goes to t with $s > t$, thanks to the uniform convergence of $R_{s'-s}f$ to f as $s' - s$ goes to 0. Then we can define $R_t f = \lim_{s \rightarrow t^+} R_s f \in \mathcal{C}_0(S)$. Consider a sequence $(s_n)_{n \geq 1}$ converging to t such that, for all $n \geq 1$, $s_n \in D_+$, and $s_n > t$. For all $x^1 \leq x^2$, the marginals of $r_{s_n}((x^1, x^2), \cdot)$ are given by $p_{s_n}(x^1, \cdot)$ and $p_{s_n}(x^2, \cdot)$, which, by the Feller property (b), converge weakly to $p_t(x^1, \cdot)$ and $p_t(x^2, \cdot)$. As a consequence, the sequence $(r_{s_n}((x^1, x^2), \cdot))_{n \geq 1}$ is tight, and the convergence of $R_{s_n} f$ to $R_t f$ for every $f \in \mathcal{C}_0(S)$ shows that there indeed is a probability distribution $r_t((x^1, x^2), \cdot)$ which is the weak limit of $r_{s_n}((x^1, x^2), \cdot)$ as $n \rightarrow +\infty$, with $R_t f(x^1, x^2) = \int_S f(x^1, x^2) r_t((x^1, x^2), (dx^1, dx^2))$ for every $f \in \mathcal{C}_0(S)$, and whose marginals are $p_t(x^1, \cdot)$ and $p_t(x^2, \cdot)$. As in Step 3, we check that r_t is a Markov kernel as a (pointwise) weak limit of Markov kernels.

Note that, by construction, R_t maps $\mathcal{C}_0(S)$ into $\mathcal{C}_0(S)$, and, composing uniform limits, we also obtain that $R_t f$ goes uniformly to f as t goes to zero, so properties (a) and (b) are satisfied. Finally, the semi-group property comes from the uniform convergence on the dyadics: let $s, t > 0$ and consider two sequences of elements of D_+ denoted by $(s_n)_{n \geq 1}$, $(t_n)_{n \geq 1}$, with $s_n \geq s$ and $t_n \geq t$ for all n , and $s_n \rightarrow s, t_n \rightarrow t$. Then we have, for all $f \in \mathcal{C}_0(S)$,

$$\begin{aligned} \|R_{s+t}f - R_s R_t f\|_\infty &\leq \|R_{s+t}f - R_{s_n} R_{t_n} f\|_\infty + \|R_{s_n} R_{t_n} f - R_{s_n} R_t f\|_\infty \\ &\quad + \|R_{s_n} R_t f - R_s R_t f\|_\infty. \end{aligned}$$

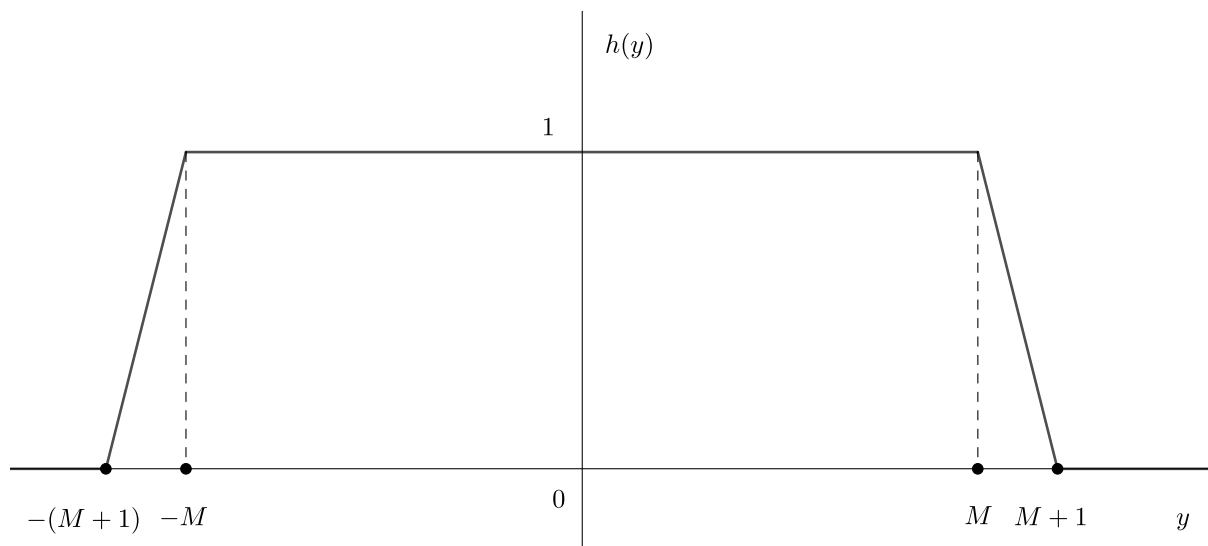
Now, thanks to the semi-group property on D_+ , the first term is $\|R_{s+t}f - R_{s_n+t_n}f\|_\infty$ and can be made arbitrarily small since $s_n + t_n$ converges to $s + t$. The third term is also small, because $R_{s_n} g$ converges to $R_s g$ uniformly for all $g \in \mathcal{C}_0(S)$. Finally, the second term is smaller than $\|R_{t_n} f - R_t f\|_\infty$ and thus also converges to zero. Hence we obtain the semi-group property of $(R_t)_{t \geq 0}$ on $\mathcal{C}_0(S)$, and the corresponding Chapman-Kolmogorov equation for $(r_t)_{t \geq 0}$.

3 About the proof of Kamae et al.

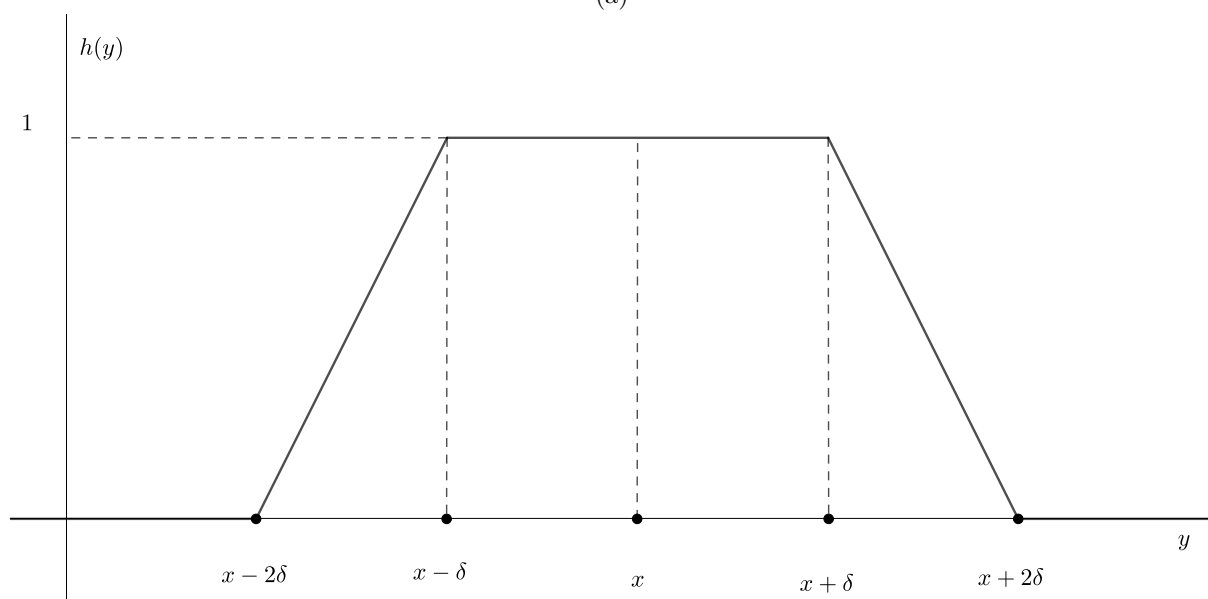
The argument of Kamae et al. in [5] to deal with the case of continuous-time Markov processes with càdlàg paths (Theorem 5) is by reduction to the discrete-time case (Theorem 2 via Theorem 4), invoking a step-by-step construction over rational time indices which is then extended to time-indices in \mathbb{R}_+ .

We now explain the issue with the step-by-step construction, in the context of two versions of the same process Markov X starting from distinct initial values being compared. The construction relies on an enumeration $(t_n)_{n \geq 1}$ of non-negative rational numbers, and the Markov kernels

$$\tilde{p}_{t_1, \dots, t_n}((x_1, \dots, x_{n-1}), \cdot) = \mathbb{P}(X_{t_n} \in \cdot | X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}).$$



(a)



(b)

Figure 2: Auxiliary functions used to obtain Feller property

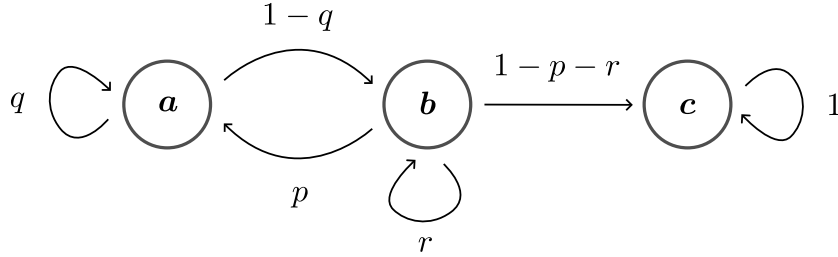


Figure 3: A non-decreasing Markov chain whose post-conditioned kernels are not non-decreasing

with the key assumption that, for all $n \geq 1$ and all $x^1, x^2 \in \mathbb{R}^{n-1}$ such that $x_i^1 \leq x_i^2$ for $1 \leq i \leq n-1$

$$\tilde{p}_{t_1, \dots, t_n}((x_1^1, \dots, x_{n-1}^1), \cdot) \preceq \tilde{p}_{t_1, \dots, t_n}((x_1^2, \dots, x_{n-1}^2), \cdot), \quad (9)$$

where \preceq denotes the stochastic dominance ordering between probability measures.

From property (c) and the Markov property, (9) is certainly true when $t_n \geq t_i$ for all $1 \leq i \leq n-1$. However, since it is certainly not possible to enumerate a dense subset of \mathbb{R}_+ using an increasing sequence, the sequence $(t_n)_{n \geq 1}$ cannot be increasing, and the success of the step-by-step construction relies on (9) being true also when there exist $1 \leq i, j \leq n-1$ such that $t_i < t_n < t_j$. Unfortunately, property (c) is not sufficient to ensure that (9) is true in such situations, as we now show using a counter-example.

Consider a Markov chain X on a set of three states $a < b < c$, whose transition kernel involves parameters p, q, r , as depicted on Figure 3. For $x_0, x_1, x_2 \in \{a, b, c\}$, denote by F_{x_0} the c.d.f. of X_1 under the condition $X_0 = x_0$, $F_{x_0 x_2}$ that of X_1 under the condition $X_0 = x_0, X_2 = x_2$, and $p_{x_0 x_1}$ the probability of transition from x_0 to x_1 (in one jump). One checks that, as soon as $p \leq q$, we have $F_a \geq F_b \geq F_c$, so that the Markov chain is stochastically non-decreasing. Now let us compute $F_{aa}(a)$ and $F_{bb}(a)$. We have

$$\begin{cases} F_{aa}(a) = \frac{p_{aa} p_{aa}}{p_{aa} p_{aa} + p_{ab} p_{ba} + p_{ac} p_{ca}} = \frac{q^2}{q^2 + (1-q)p}, \\ F_{bb}(a) = \frac{p_{ba} p_{ab}}{p_{ba} p_{ab} + p_{bb} p_{bb} + p_{bc} p_{cb}} = \frac{p(1-q)}{p(1-q) + r^2}, \end{cases}$$

As a consequence, if we find $p, q, r \in [0, 1]$ such that

$$\frac{p(1-q)}{p(1-q) + r^2} - \frac{q^2}{q^2 + (1-q)p} > 0$$

with conditions $p \leq q$ and $p + r \leq 1$, we have $F_{bb}(a) > F_{aa}(a)$, which shows that we do not have the ordering $\mathbb{P}(X_1 \in \cdot | X_0 = a, X_2 = a) \preceq \mathbb{P}(X_1 \in \cdot | X_0 = b, X_2 = b)$, despite the fact that $a < b$.

For instance⁹, taking $p = \frac{1}{2}, q = \frac{3}{5}$ and $r = \frac{1}{4}$ works.

⁹Notice that if we had $r = 1 - p$, the third state c would be useless. However it is not, precisely because the inequality has no solution in that case.

References

- [1] M.-F. Chen. *From Markov chains to non-equilibrium particle systems*. World Scientific Publishing Co., Inc., River Edge, NJ, second edition, 2004.
- [2] B. Fristedt and L. Gray. *A modern approach to probability theory*. Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA, 1997.
- [3] M. Hairer and J. Mattingly. The strong Feller property for singular stochastic PDEs. *Ann. Inst. Henri Poincaré Probab. Stat.*, 54(3):1314–1340, 2018.
- [4] O. Kallenberg. *Foundations of Modern Probability*. Springer, 2nd edition, 2002.
- [5] T. Kamae, U. Krengel, and G. L. O’Brien. Stochastic inequalities on partially ordered spaces. *Ann. Probability*, 5(6):899–912, 1977.
- [6] V. N. Kolokoltsov. Stochastic monotonicity and duality of one-dimensional Markov processes. *Mat. Zametki*, 89(5):694–704, 2011.
- [7] A. Müller and D. Stoyan. *Comparison methods for stochastic models and risks*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2002.
- [8] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer, 3rd edition, 1999.
- [9] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.
- [10] S. Zhang. Existence and application of optimal Markovian coupling with respect to non-negative lower semi-continuous functions. *Acta Math. Sin. (Engl. Ser.)*, 16(2):261–270, 2000.