

## EMBEDDED MARKOV CHAIN APPROXIMATIONS IN SKOROKHOD TOPOLOGIES

BY

BJÖRN BÖTTCHER (DRESDEN)

*Abstract.* We prove a  $J_1$ -tightness condition for embedded Markov chains and discuss four Skorokhod topologies in a unified manner.

To approximate a continuous time stochastic process by discrete time Markov chains, one has several options to embed the Markov chains into continuous time processes. On the one hand, there is a Markov embedding which uses exponential waiting times. On the other hand, each Skorokhod topology naturally suggests a certain embedding. These are the step function embedding for  $J_1$ , the linear interpolation embedding for  $M_1$ , the multistep embedding for  $J_2$  and a more general embedding for  $M_2$ . We show that the convergence of the step function embedding in  $J_1$  implies the convergence of the other embeddings in the corresponding topologies. For the converse statement, a  $J_1$ -tightness condition for embedded time-homogeneous Markov chains is given.

Additionally, it is shown that  $J_1$  convergence is equivalent to the joint convergence in  $M_1$  and  $J_2$ .

**2010 AMS Mathematics Subject Classification:** Primary: 60B10; Secondary: 60J75, 60J05.

**Key words and phrases:** Markov chain embedding, tightness, Skorokhod space, Skorokhod topologies, jump processes, Markov chain approximation.

### 1. INTRODUCTION

The space of right continuous functions with left limits plays a prominent role in the theory of stochastic processes. Skorokhod [15] was the first to consider this space with various metrics. He introduced four topologies:  $J_1$ ,  $J_2$ ,  $M_1$  and  $M_2$ . The main focus in the literature is on the  $J_1$  topology (e.g. [3], [7], [9]) and more recently on  $M_1$  (e.g. [2], [17]). We will be concerned with all four. But note that there are further topologies on the Skorokhod space: e.g. the sequential topology of Jakubowski [10] and the pseudo-path topology by Meyer and Zheng [12].

Given the relations of Skorokhod's topologies, for a fixed sequence the convergence in a stronger topology implies the convergence in a weaker topology,

i.e.,  $J_1$ -convergence implies  $M_1$ - and  $J_2$ -convergence, and either of these implies  $M_2$ -convergence. But when one starts with discrete time processes there are many ways to embed these into continuous time processes, and most embeddings do not converge in all four topologies. Actually, each of the four Skorokhod topologies suggests a particular embedding, the weaker the topology is the “wilder” the embedding can be (see Section 3). Thus a natural question arises: can we switch the topology and the corresponding embedding without losing convergence?

Consider Markov chains with time steps of size  $\frac{1}{n}$  and let  $n$  tend to infinity. In order to discuss a continuous time limit, it is necessary to embed the chains into continuous time processes. In our general setting the limit can be a process with jumps. For processes with continuous paths Sato [14] discussed a closely related problem: he showed that linearly interpolated Markov chains converge with respect to the uniform topology (in the space of continuous functions) if and only if the step function embedded Markov chains converge to a continuous process with respect to the  $J_1$  topology (in the Skorokhod space). Our result allows us in particular (cf. Example 4.1) to extend Markov chain approximations for Feller processes (cf. [5], [6]) to different embeddings. More general, we provide a  $J_1$ -tightness condition for Markov chains, see Theorem 4.6.

It turns out that, in the above setting, convergence is always preserved when switching from a topology to a weaker topology (and to the corresponding embedding), see Corollary 3.1 and Theorem 3.1. For the converse direction naturally some additional assumption is needed, see (Counter-)Examples 3.1 and Corollary 4.1.

In the next section we introduce the Skorokhod space and the topologies  $J_1$ ,  $J_2$ ,  $M_1$  and  $M_2$  in a unified framework, which consolidates the literature, e.g. [15], [17], [13]. In particular, we recall their relations and several representations. The relation between  $J_1$  and the combination of  $J_2$  and  $M_1$  (Lemma 2.1) seems to be neglected in the literature. It goes back to a remark without proof of Skorokhod [15], 2.2.10–13. In Section 3, the embeddings are introduced and their relations are discussed. In Section 4, a  $J_1$ -tightness condition (Theorem 4.6) for embedded Markov chains is presented, it enables us to switch from a weaker to a stronger topology (and to the corresponding embedding; see Corollary 4.1). The paper closes with the proof of Lemma 2.1.

## 2. THE SKOROKHOD SPACE AND ITS TOPOLOGIES

Throughout the paper segments between points  $x, y \in \mathbb{R}^d$  are denoted by

$$[[x, y]] := \{z \in \mathbb{R}^d \mid z = \alpha x + (1 - \alpha)y \text{ for some } \alpha \in [0, 1]\},$$

and  $\|\cdot\|_\infty$  denotes the supremum norm. Limits without superscript, e.g.  $f_n \rightarrow f$ , are meant in the Euclidean distance. Unless stated otherwise, limits are considered for the index tending to infinity, e.g.  $n \rightarrow \infty$ , and the dimension  $d \in \mathbb{N}$  is arbitrary.

DEFINITION 2.1. A function  $f : [0, 1] \rightarrow \mathbb{R}^d$  is *right continuous with left limits* (rcll) if

$$f(s+) := \lim_{t \downarrow s} f(t) = f(s) \text{ and } f(s-) := \lim_{t \uparrow s} f(t) \text{ exist in } \mathbb{R}^d.$$

The Skorokhod space is

$$D[0, 1] := D_d[0, 1] := \{f : [0, 1] \rightarrow \mathbb{R}^d \mid f \text{ is rcll and } f \text{ is left continuous at } 1\}.$$

On the Skorokhod space several metrics can be defined.

DEFINITION 2.2. Let  $f, f_1, f_2 \in D[0, 1]$ .

The *incomplete graph* of  $f$  is

$$\Gamma_f := \{(z, t) \in \mathbb{R}^d \times [0, 1] \mid z = f(t-) \text{ or } z = f(t)\}.$$

The *complete graph* of  $f$  is

$$\overline{\Gamma}_f := \{(z, t) \in \mathbb{R}^d \times [0, 1] \mid z \in [[f(t-), f(t)]]\}.$$

An *order* is defined on  $\overline{\Gamma}_f$  by

$$(z_1, t_1) \leq (z_2, t_2) \text{ if } \begin{cases} \text{either } t_1 < t_2 \\ \text{or } t_1 = t_2, |f(t_1-) - z_1| \leq |f(t_2-) - z_2|, \end{cases}$$

and the families of *parametric representations* of  $\overline{\Gamma}_f$  are given by

$$\begin{aligned} \pi(\overline{\Gamma}_f) &= \{(u, r) \mid u : [0, 1] \rightarrow \mathbb{R}^d \text{ continuous, } r : [0, 1] \rightarrow [0, 1] \text{ continuous,} \\ &\quad (u, r) \text{ is non-decreasing and } (u, r)[0, 1] = \overline{\Gamma}_f\}, \end{aligned}$$

$$\begin{aligned} \tilde{\pi}(\overline{\Gamma}_f) &= \{(u, r) \mid u : [0, 1] \rightarrow \mathbb{R}^d \text{ continuous, } r : [0, 1] \rightarrow [0, 1] \text{ continuous,} \\ &\quad r \text{ is non-decreasing and } (u, r)[0, 1] = \overline{\Gamma}_f\}. \end{aligned}$$

The sets of *time transformations* are  $\Lambda := \{\lambda : [0, 1] \rightarrow [0, 1] \mid \lambda \text{ is bijective}\}$  and  $\Lambda_c := \Lambda \cap C[0, 1]$ . Let  $\text{id}$  denote the identity function on  $[0, 1]$ ; then *metrics on Skorokhod space* are given by

$$d_{J_1}(f_1, f_2) := \inf_{\lambda \in \Lambda_c} \{\|f_1 \circ \lambda - f_2\|_\infty \vee \|\lambda - \text{id}\|_\infty\},$$

$$\tilde{d}_{J_2}(f_1, f_2) := \inf_{\lambda \in \Lambda} \{\|f_1 \circ \lambda - f_2\|_\infty \vee \|\lambda - \text{id}\|_\infty\},$$

$$d_{M_1}(f_1, f_2) := \inf_{\substack{(u_j, r_j) \in \pi(\overline{\Gamma}_{f_j}) \\ j=1,2}} \{\|u_1 - u_2\|_\infty \vee \|r_1 - r_2\|_\infty\},$$

$$\tilde{d}_{M_2}(f_1, f_2) := \inf_{\substack{(u_j, r_j) \in \tilde{\pi}(\overline{\Gamma}_{f_j}) \\ j=1,2}} \{\|u_1 - u_2\|_\infty \vee \|r_1 - r_2\|_\infty\},$$

$$d_{J_2}(f_1, f_2) := m_H(\Gamma_{f_1}, \Gamma_{f_2}),$$

$$d_{M_2}(f_1, f_2) := m_H(\overline{\Gamma}_{f_1}, \overline{\Gamma}_{f_2}),$$

where  $m_H(A, B) = \inf\{\varepsilon > 0 : A \subset B^\varepsilon, B \subset A^\varepsilon\}$  is the Hausdorff distance with  $B^\varepsilon := \{(x, t) \in \mathbb{R}^d \times [0, 1] \mid \exists(y, s) \in B : |x - y| \vee |t - s| \leq \varepsilon\}$ .

In the following,  $\mathcal{T}$  will always denote one of  $J_1, J_2, M_1, M_2$ . Note that  $d_{\mathcal{T}}$  is a metric on  $D[0, 1]$ . This is clear by the definition for  $J_2$  and  $M_2$ , for  $J_1$  see [3], and for  $M_1$  see [17]. For  $f, f_n \in D[0, 1]$  ( $n \in \mathbb{N}$ ) we say that  $f_n$  converges to  $f$  in  $\mathcal{T}$  (in  $\mathcal{T}$ -topology) if  $d_{\mathcal{T}}(f_n, f) \rightarrow 0$ . The convergence is denoted by

$$f_n \xrightarrow{\mathcal{T}} f.$$

REMARK 2.1. Skorokhod [15] used  $\tilde{d}_{J_2}, \tilde{d}_{M_2}$  to introduce the topologies  $J_2$  and  $M_2$ , respectively. But  $d_{J_2}, d_{M_2}$  yield the same topologies ([13], II.4.1, p. 82, II.4.2, p. 83). Note that on  $\mathbb{R}^d$  with  $d > 1$  one could also define a complete graph by including for each coordinate the whole interval between the start and endpoints of the jumps, i.e.  $\times_{i=1}^d [[f_i(t-), f_i(t)]]$ , where  $f_i$  denotes the  $i$ th component of  $f$ . The definition above only uses the linear interpolation, i.e.,  $[[f(t-), f(t)]]$ , thus here the  $M$  topologies are strong in the sense of Whitt ([17], Sections 12.3 and 12.10).

Regarding completeness of the corresponding metric spaces see Remark 4.1(1).

The convergence in these topologies can also be characterized by oscillation functions.

DEFINITION 2.3 (Oscillation functions). Define for  $x, x_1, x_2 \in \mathbb{R}^d$

$$(2.1) \quad \begin{aligned} J(x, x_1, x_2) &:= |x - x_1| \wedge |x - x_2|, \\ M(x, x_1, x_2) &:= |x - [[x_1, x_2]]| := \inf_{y \in [[x_1, x_2]]} |x - y|, \end{aligned}$$

and for  $\delta > 0$

$$\begin{aligned} T_1(\delta) &:= \{(t, t_1, t_2) \mid (t - \delta) \vee 0 \leq t_1 < t < t_2 \leq (t + \delta) \wedge 1\}, \\ T_2(\delta) &:= \{(t, t_1, t_2) \mid t \in [0, 1], t_1 \in [(t - \delta) \vee 0, (t - \delta) \vee 0 + \delta/2], \\ &\quad t_2 \in [(t + \delta) \wedge 1 - \delta/2, (t + \delta) \wedge 1]\}. \end{aligned}$$

The oscillation functions for  $f : [0, 1] \rightarrow \mathbb{R}^d$ ,  $\delta > 0$  and  $i = 1, 2$  are

$$(2.2) \quad \begin{aligned} \Delta_{J_i}(\delta, f) &:= \sup_{(t, t_1, t_2) \in T_i(\delta)} J(f(t), f(t_1), f(t_2)), \\ \Delta_{M_i}(\delta, f) &:= \sup_{(t, t_1, t_2) \in T_i(\delta)} M(f(t), f(t_1), f(t_2)), \\ \Delta_U^{\{0,1\}}(\delta, f) &:= \sup_{0 < t < \delta} |f(0) - f(t)| + \sup_{1 - \delta < t < 1} |f(1) - f(t)|. \end{aligned}$$

The following theorem states the fundamental relation of the oscillation functions and the metrics.

**THEOREM 2.1.** *Let  $f_n, f \in D[0, 1]$ . Then*

$$f_n \xrightarrow{\mathcal{T}} f \Leftrightarrow \begin{cases} \text{(i) } f_n(t) \rightarrow f(t), \forall t \in B, \text{ where } 0, 1 \in B \text{ and} \\ \hspace{10em} B \text{ is a dense subset of } [0, 1], \\ \text{(ii) } \lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \Delta_{\mathcal{T}}(\delta, f_n) = 0. \end{cases}$$

**Proof.** The proofs can be found for  $M_2$  in [15], 2.3.4, for  $M_1$  in [15], 2.4.1, for  $J_2$  in [13], II.4.4, [15], 2.5.3, and for  $J_1$  in [15], 2.6.1. ■

Note that the oscillation functions satisfy the following relations.

**LEMMA 2.1.** *Let  $\delta > 0$  and  $f : [0, 1] \rightarrow \mathbb{R}^d$ . Then*

$$(2.3) \quad \begin{array}{ccccc} & & \Delta_{J_2}(\delta, f) & \leq & \\ \Delta_{M_2}(\delta, f) & \leq & & \leq & \Delta_{J_1}(\delta, f) \leq \Delta_{M_1}(\delta, f) + \Delta_{J_2}(\delta, f). \\ & \leq & \Delta_{M_1}(\delta, f) & \leq & \end{array}$$

**Proof.** The first four inequalities follow directly from the definition of the oscillation functions, since  $M(x, x_1, x_2) \leq J(x, x_1, x_2)$  and  $T_2(\delta) \subset T_1(\delta)$ . The last inequality is proved in Section 5. ■

Thus we have the following relations of the convergences:

$$(2.4) \quad \begin{array}{ccccc} & & J_2 & \leq & \\ M_2 & \leq & & \leq & J_1 \Leftrightarrow M_1 + J_2. \\ & \leq & M_1 & \leq & \end{array}$$

As remarked by Skorokhod [15], 2.2.10–13, there are further equivalent characterizations of the convergence in these topologies for functions in  $\mathbb{R}^d$  with  $d = 1$ .

**THEOREM 2.2.** *Let  $f_n, f \in D_1[0, 1]$ .*

(1)  *$M_2$  is characterized by the convergence of the local extrema:*

$$f_n \xrightarrow{M_2} f \Leftrightarrow \inf_{t \in [t_1, t_2]} f_n(t) \rightarrow \inf_{t \in [t_1, t_2]} f(t) \text{ and } \sup_{t \in [t_1, t_2]} f_n(t) \rightarrow \sup_{t \in [t_1, t_2]} f(f)$$

for all  $t_1, t_2$  being points of continuity of  $f$ .

(2)  *$M_1$  is characterized by the convergence of the number of oscillations:*

$$f_n \xrightarrow{M_1} f \Leftrightarrow \nu_{[t_1, t_2]}^{[a, b]} f_n \rightarrow \nu_{[t_1, t_2]}^{[a, b]} f$$

for all  $t_1, t_2$  being points of continuity of  $f$  and almost all  $a < b$ . Here  $\nu_{[t_1, t_2]}^{[a, b]} f$  is the largest  $k$  such that there exist  $t^{(0)} < \dots < t^{(k)}$  in  $[t_1, t_2]$  with  $f(t^{(0)}) \leq a$ ,  $f(t^{(1)}) \geq b$ ,  $f(t^{(2)}) \leq a, \dots$

(3)  $J_2$  is characterized by the convergence of the first overshoots:

$$f_n \xrightarrow{J_2} f \Leftrightarrow \gamma_{[t_1, t_2], a}^+ f_n \rightarrow \gamma_{[t_1, t_2], a}^+ f$$

for all  $t_1, t_2$  being points of continuity of  $f$  and almost all  $a$ . Here, using the convention  $\inf \emptyset = 1$ , define  $\tau_{a, f} := \inf\{t \in [0, 1] \mid f(t) \geq a\}$  and

$$\gamma_{[0, 1], a}^+ f := \begin{cases} f(\tau_{a, f}) - a, & \tau_{a, f} < 1, \\ -1, & \text{otherwise,} \end{cases}$$

and, in general, use

$$\gamma_{[t_1, t_2], a}^+ f := \gamma_{[0, 1], a}^+ \tilde{f} \quad \text{with } \tilde{f}(t) := \begin{cases} f(t_1+), & t \leq t_1, \\ f(t), & t \in (t_1, t_2), \\ f(t_2-), & t \geq t_2. \end{cases}$$

(4)  $J_1$  is characterized by the convergence of the first overshoots and the number of oscillations:

$$f_n \xrightarrow{J_1} f \Leftrightarrow \gamma_{[t_1, t_2], a}^+ f_n \rightarrow \gamma_{[t_1, t_2], a}^+ f \text{ and } \nu_{[t_1, t_2]}^{[a, b]} f_n \rightarrow \nu_{[t_1, t_2]}^{[a, b]} f$$

for all  $t_1, t_2$  being points of continuity of  $f$  and almost all  $a < b$ . For the definition of  $\gamma$  and  $\nu$  see (2) and (3).

**Proof.** Statements (1) and (3) are a consequence of the definition of these metrics via the Hausdorff metric. Statement (2) can be found in Whitt [17], Theorem 12.7.4, p. 412. Statement (4) is due to the equivalence of the convergences (Lemma 2.1):  $J_1 \Leftrightarrow M_1 + J_2$ . ■

**REMARK 2.2.** (1) The characterizations in Theorem 2.2 are tailored to  $d=1$ . For higher dimensions Whitt [17], Theorem 12.7.2, showed, for example, that

$$(2.5) \quad f_n \xrightarrow{M_1} f \Leftrightarrow \eta \cdot f_n \xrightarrow{M_1} \eta \cdot f \text{ as functions in } D_1[0, 1] \text{ for all } \eta \in \mathbb{R}^d$$

for  $f_n, f \in D_d[0, 1]$ .

(2) Throughout this section we only considered  $D[0, 1]$ . By replacing 1 by  $T \in (0, \infty)$  we have an obvious extension to  $D[0, T]$ . An approach to define convergence for  $f_n, f \in D[0, \infty)$  is

$$(2.6) \quad f_n \xrightarrow{\mathcal{T}} f \Leftrightarrow f_n|_{[0, T]} \xrightarrow{\mathcal{T}} f|_{[0, T]} \text{ in } D[0, T] \text{ for all } T \in \{t \mid f \text{ is continuous in } t\}.$$

For further details on the extension to  $D[0, \infty)$  see Lindvall [11].

### 3. EMBEDDINGS AND APPROXIMATIONS

Let  $n \in \mathbb{N}$  and  $y^{(n)}$  be a sequence  $(y_0^{(n)}, y_1^{(n)}, y_2^{(n)}, \dots)$  in  $\mathbb{R}^d$ . Define for each topology  $\mathcal{T}$  the embeddings  $x^{n,\mathcal{T}}$  to be functions in  $D[0, 1]$  such that for  $k < n$

$$(3.1) \quad x^{n,\mathcal{T}}\left(\frac{k}{n}\right) = y_k^{(n)}$$

and for all  $t \in \left(\frac{k}{n}, \frac{k+1}{n}\right)$

$$(3.2) \quad \begin{aligned} x^{n,J_1}(t) &= y_k^{(n)} && \text{(step functions for } J_1), \\ x^{n,M_1}(t) &= y_k^{(n)} + \left(t - \frac{k}{n}\right)(y_{k+1}^{(n)} - y_k^{(n)}) && \text{(linear interpolation for } M_1), \\ x^{n,J_2}(t) &\in \{y_k^{(n)}, y_{k+1}^{(n)}\} && \text{(multiple steps for } J_2), \\ x^{n,M_2}(t) &\in [[y_k^{(n)}, y_{k+1}^{(n)}]] && \text{(any rcll function for } M_2). \end{aligned}$$

Note that the requirement  $x^{n,\mathcal{T}} \in D[0, 1]$  ensures that  $x^{n,\mathcal{T}}(1) = \lim_{t \nearrow 1} x^{n,\mathcal{T}}(t)$ . Clearly, in the above definition only those  $k$  with  $k < n$  are used, but in the next section it will be convenient that each  $y^{(n)}$  is a countable sequence.

Lemma 2.1 implies the following result.

**COROLLARY 3.1.**  $x^{n,J_1}$  converges in  $J_1$  implies that  $x^{n,J_1}$  also converges in  $\mathcal{T}$ .

Moreover, in a given topology we can always switch between its embedding and the  $J_1$  embedding.

**THEOREM 3.1.**  $x^{n,\mathcal{T}}$  converges in  $\mathcal{T}$  if and only if  $x^{n,J_1}$  converges in  $\mathcal{T}$ .

**Proof.** By the definition of the metrics and the embeddings we obtain

$$(3.3) \quad d_{J_2}(x^{n,J_2}, x^{n,J_1}) \leq \frac{1}{n} \quad \text{and} \quad d_{M_2}(x^{n,M_2}, x^{n,J_1}) \leq \frac{1}{n}.$$

For  $M_1$  we use (2.5) and Theorem 2.2. Note that for all  $\eta \in \mathbb{R}^d$  and all  $t_1, t_2$  which are points of continuity of the limit and almost all  $a < b$

$$(3.4) \quad \left| \nu_{[t_1, t_2]}^{[a, b]}(\eta \cdot x^{n,M_1}) - \nu_{[t_1, t_2]}^{[a, b]}(\eta \cdot x^{n,J_1}) \right| \xrightarrow{n \rightarrow \infty} 0,$$

since for  $\frac{k-1}{n} < t_1 \leq \frac{k}{n}$  and  $\frac{l}{n} \leq t_2 < \frac{l+1}{n}$  the number of oscillations coincides for the segment from  $\frac{k}{n}$  to  $\frac{l}{n}$ . In the limit, no overshoot appears at the two boundary segments since  $t_1$  and  $t_2$  are points of continuity. Thus, if the limit is in  $D[0, 1]$ , the statement follows by the triangle inequality. ■

We close this section with basic counterexamples which show that the converse implication of Corollary 3.1 fails.

EXAMPLE 3.1. Let  $n \geq 4$  and  $k \in \mathbb{N} \cup \{0\}$ .

(1) Let  $y_k^{(n)} = 0$  for  $\frac{k}{n} < \frac{1}{2}$ ,  $y_k^{(n)} = \frac{1}{2}$  for  $\frac{k-1}{n} < \frac{1}{2} \leq \frac{k}{n}$  and  $y_k^{(n)} = 1$  otherwise. Then  $x^{n,J_1}$  converges to  $\mathbf{1}_{[1/2,1]}$  in  $M_1$ , but not in  $J_1$  and not in  $J_2$ .

(2) Let  $y_k^{(n)} = 0$  for  $\frac{k}{n} < \frac{1}{2}$ ,  $y_k^{(n)} = 1$  for  $\frac{k-1}{n} < \frac{1}{2} \leq \frac{k}{n}$ ,  $y_k^{(n)} = 0$  for  $\frac{k-2}{n} < \frac{1}{2} \leq \frac{k-1}{n}$  and  $y_k^{(n)} = 1$  otherwise. Then  $x^{n,J_1}$  converges to  $\mathbf{1}_{[1/2,1]}$  in  $J_2$ , but not in  $J_1$  and not in  $M_1$ .

(3) Let  $y_k^{(n)} = 0$  for  $\frac{k}{n} < \frac{1}{2}$ ,  $y_k^{(n)} = \frac{1}{2}$  for  $\frac{k-1}{n} < \frac{1}{2} \leq \frac{k}{n}$ ,  $y_k^{(n)} = 1$  for  $\frac{k-2}{n} < \frac{1}{2} \leq \frac{k-1}{n}$ ,  $y_k^{(n)} = 0$  for  $\frac{k-3}{n} < \frac{1}{2} \leq \frac{k-2}{n}$  and  $y_k^{(n)} = 1$  otherwise. Then  $x^{n,J_1}$  converges to  $\mathbf{1}_{[1/2,1]}$  in  $M_2$ , but not in  $J_2$  and not in  $M_1$ , and thus not in  $J_1$ .

#### 4. CONVERGENCE OF PROCESSES AND MARKOV CHAINS

Let  $X, X^{(n)}$  ( $n \in \mathbb{N}$ ) be  $D[0, 1]$ -valued random variables on some probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . To fix notation, we recall the following standard definitions.

DEFINITION 4.1. (1)  $X^{(n)} \xrightarrow{d} X$  w.r.t.  $\mathcal{T}: \mathbb{E}(G(X^{(n)})) \rightarrow \mathbb{E}(G(X))$  for all bounded and  $\mathcal{T}$ -continuous functions  $G: D[0, 1] \rightarrow \mathbb{R}$ .

(2)  $X_t^{(n)} \xrightarrow{d} X_t: \mathbb{E}(g(X_t^{(n)})) \rightarrow \mathbb{E}(g(X_t))$  for all bounded and continuous functions  $g: \mathbb{R}^d \rightarrow \mathbb{R}^d$ .

(3)  $X^{(n)} \xrightarrow{\mathbb{P}} X$  w.r.t.  $\mathcal{T}: \lim_{n \rightarrow \infty} \mathbb{P}(d_{\mathcal{T}}(X^{(n)}, X) > \varepsilon) = 0$  for all  $\varepsilon > 0$ .

(4)  $X^{(n)} \xrightarrow{\text{fdd}} X$  on  $I: (X_{t_1}^{(n)}, \dots, X_{t_k}^{(n)}) \xrightarrow{d} (X_{t_1}, \dots, X_{t_k})$  for all  $t_i \in I$ .

(5)  $(X^{(n)})_{n \in \mathbb{N}}$  is  $\mathcal{T}$ -tight: for all  $\varepsilon > 0$  there exists a  $\mathcal{T}$ -compact set  $K \in \mathcal{B}(D[0, 1])$  such that  $\sup_n \mathbb{P}(X^{(n)} \in K^c) \leq \varepsilon$ .

(6)  $(X^{(n)})_{n \in \mathbb{N}}$  is relative  $\mathcal{T}$ -compact: for every subsequence  $(X^{(n_k)})_{k \in \mathbb{N}}$  there exists a further subsequence  $(X^{(n_{k_l})})_{l \in \mathbb{N}}$  and a  $D[0, 1]$ -valued random variable  $Y$  such that  $X^{(n_{k_l})} \xrightarrow{d} Y$  w.r.t.  $\mathcal{T}$ .

The following result is the standard tool to handle convergence on  $D[0, 1]$ . We include a sketch of the proof since we are going to point out a particular detail later.

THEOREM 4.1. We have

$$X^{(n)} \xrightarrow{d} X \text{ w.r.t. } \mathcal{T} \quad \Leftrightarrow \quad \begin{cases} \text{(i)} & X^{(n)} \xrightarrow{\text{fdd}} X \text{ on a dense subset of } [0, 1], \\ \text{(ii)} & (X^{(n)})_{n \in \mathbb{N}} \text{ is relatively } \mathcal{T}\text{-compact.} \end{cases}$$

PROOF.  $\Leftarrow$ . By (ii) every subsequence of  $X^{(n)}$  has a converging subsequence whose limit has, by (i), the same finite-dimensional distributions as  $X$ . The finite-dimensional distributions define uniquely the distribution of a process in  $D[0, 1]$ , thus the limit is  $X$ .



$\Rightarrow$ . The projection  $\pi_t : D[0, 1] \rightarrow \mathbb{R}^d$ ,  $\pi_t(f) := f(t)$  is measurable,  $X^{(n)} \xrightarrow{d} X$  and the set  $T := \{t : \mathbb{P}(|X_t - X_{t-}|) > 0\}$  is countable. Thus, for all  $t \in T^c$

$$\mathbb{P}(X \in \{f \in D[0, 1] : \pi_t(f) \text{ is discontinuous at } t\}) = \mathbb{P}(|X_t - X_{t-}| > 0) = 0$$

and the statement follows by a continuous mapping theorem, see e.g. [17], Theorem 3.4.3. ■

REMARK 4.1. (1) A sufficient condition for relative  $\mathcal{T}$ -compactness is given by Prokhorov's theorem:

$$\mathcal{T}\text{-tightness} \quad \Rightarrow \quad \text{relative } \mathcal{T}\text{-compactness.}$$

The converse holds if  $D[0, 1]$  is, with the topology induced by  $d_{\mathcal{T}}$ , a complete and separable space. This is the case for  $J_1$  and  $M_1$ , but for  $J_2$  and  $M_2$  it is still an open problem. To avoid confusion, note that the metric spaces  $(D[0, 1], d_{J_1})$  and  $(D[0, 1], d_{M_1})$  are not complete – but there exist complete metrics which generate the same topologies (one approach to construct these complete metrics is to add to the given metrics the Lévy distance of distributions obtained via the oscillation functions; see Section 12.8 in [17]).

(2) Note that we assumed that  $X$  is  $D[0, 1]$ -valued. For condition (i) in Theorem 4.1 this can be relaxed, at least if  $J_1$  is considered. Topsøe [16], Theorem 2, showed that  $J_1$ -tightness and the mere convergence of the finite-dimensional distributions on a dense subset are sufficient to identify a process in  $D[0, 1]$  which is the  $J_1$  limit.

(3) Looking at the proof of Theorem 4.1 (see also [9], Theorem 3.14), note that if the process  $X$  is stochastically continuous, i.e.,

$$(4.1) \quad \forall s \in [0, 1] \forall \varepsilon > 0 : \lim_{t \rightarrow s} \mathbb{P}(|X_t - X_s| > \varepsilon) = 0,$$

then the dense subset of  $[0, 1]$  can be taken to be the whole set  $[0, 1]$ . See also Proposition 4.1 below.

(4) A necessary condition for  $X^n \xrightarrow{d} X$  w.r.t.  $\mathcal{T}$  is

$$(4.2) \quad \forall \varepsilon > 0 \exists R > 0 : \sup_n \mathbb{P}(\|X^{(n)}\|_{\infty} \geq R) < \varepsilon,$$

since otherwise some mass would dissipate and hence  $X$  would have, with positive probability, values not in  $D[0, 1]$ . Also note that  $\{f \in D[0, 1] \mid \|f\|_{\infty} < R\}$  is not  $\mathcal{T}$ -compact, since e.g.  $(\mathbf{1}_{[1/2, 1/2+1/n]})_{n \geq 2}$  has no converging subsequence.

For  $J_1$  there are several conditions for tightness, we will start with a standard result (see e.g. [9], Theorem 3.21).

**THEOREM 4.2.** *The sequence of processes  $(X^{(n)})_{n \in \mathbb{N}}$  is  $J_1$ -tight if and only if (4.2) holds and*

$$(4.3) \quad \forall \varepsilon > 0 : \limsup_{\delta \downarrow 0} \limsup_n \mathbb{P}(\Delta_{J_1}(\delta, X^{(n)}) + \Delta_U^{\{0,1\}}(\delta, X^{(n)}) > \varepsilon) = 0.$$

**REMARK 4.2.** *For  $M_1$  a result analogous to Theorem 4.2 holds (cf. [17], Theorem 12.12.3). For  $M_2$  and  $J_2$  the corresponding version of condition (4.3) is not necessary for compactness (cf. [15], 2.7.2–4).*

A well-known sufficient  $J_1$ -tightness condition is due to Aldous [1].

**THEOREM 4.3** (Aldous [1]). *The sequence  $(X^{(n)})_{n \in \mathbb{N}}$  is  $J_1$ -tight if*

$$(4.4) \quad \forall \varepsilon > 0 : \lim_{n \rightarrow \infty} \mathbb{P}(|X_{(\tau_n + t_n) \wedge 1}^{(n)} - X_{\tau_n}^{(n)}| > \varepsilon) = 0$$

*for all sequences  $(\tau_n)_{n \in \mathbb{N}}$ , with  $\tau_n$  being a stopping time for  $X^{(n)}$ , and all sequences  $(t_n)_{n \in \mathbb{N}}$  with  $t_n \geq 0$ ,  $t_n \rightarrow 0$ .*

As a motivation we also recall a closely related result by Gikhman and Skorokhod [8], Theorem 4, p. 431, formulated here for the time-homogeneous setting:

**THEOREM 4.4** (Gikhman and Skorokhod [8]). *Let  $X, X^{(n)}$  be time-homogeneous Markov processes with  $X^{(n)} \xrightarrow{fdd} X$  on some dense subset of  $[0, 1]$  and*

$$(4.5) \quad \forall \varepsilon > 0 : \lim_{h \downarrow 0} \limsup_{n \rightarrow \infty} \sup_{\substack{x \in \mathbb{R}^d \\ t \leq h}} \mathbb{P}(|X_t^{(n)} - x| > \varepsilon \mid X_0^{(n)} = x) = 0.$$

*Then  $X^{(n)} \xrightarrow{d} X$  w.r.t.  $J_1$ .*

Thus (4.5) is a  $J_1$ -tightness condition, actually ensuring that the limit is spatial-uniformly stochastically continuous from the right. The Aldous tightness condition (4.4) and condition (4.5) are both not necessary for convergence, a counterexample is a process with a fixed jump, e.g. consider the deterministic time-homogeneous Markov process whose transition probabilities for  $t > 0$  and  $x \in \mathbb{R}$  are

$$\begin{aligned} \mathbb{P}(X_t = x \mid X_0 = x) &= 1 && \text{for all } x \in [0, 1), \\ \mathbb{P}(X_t = x + t \mid X_0 = x) &= 1 && \text{for all } x \in [1, \infty) \text{ or } x \in (-\infty, 0 \wedge (-t)), \\ \mathbb{P}(X_t = x + t + 1 \mid X_0 = x) &= 1 && \text{for all } x \in (-\infty, 0) \text{ and } t + x \geq 0. \end{aligned}$$

Incidentally, this counterexample also shows that for time-homogeneous Markov processes stochastic continuity (4.1) is stronger than stochastic continuity from the right, i.e.,  $\lim_{t \downarrow 0} \mathbb{P}(|X_t - x| > \varepsilon \mid X_0 = x) = 0$  for all  $\varepsilon > 0$  and  $x \in \mathbb{R}^d$ . In fact, the following holds.

**PROPOSITION 4.1.** *Let  $X$  be a  $D[0, 1]$ -valued process. Then the following are equivalent:*

(1)  $X$  is stochastically continuous:

$$\forall s \in [0, 1] \forall \varepsilon > 0 : \lim_{t \rightarrow s} \mathbb{P}(|X_t - X_s| > \varepsilon) = 0.$$

(2)  $X$  has no fixed discontinuities:  $\forall s \in [0, 1] : \mathbb{P}(|X_s - X_{s-}| > 0) = 0$ .

If additionally  $X$  is a time-homogeneous Markov process, then properties (1) and (2) are implied by

(3)  $X$  is locally spatial-uniformly stochastically continuous from the right:

$$(4.6) \quad \forall \varepsilon > 0 \forall R > 0 : \lim_{t \downarrow 0} \sup_{|x| < R} \mathbb{P}(|X_t - x| > \varepsilon | X_0 = x) = 0.$$

**Proof.** Let  $X$  be a  $D[0, 1]$ -valued process. Then it has a right continuous path with left limits and therefore (1) and (2) are equivalent.

Moreover, for each  $\varepsilon' > 0$  there exists an  $R > 0$  such that  $\mathbb{P}(\|X\|_\infty \geq R) < \varepsilon'$ . For  $X$  being a time-homogeneous Markov process and  $0 \leq h \leq t$  we find

$$(4.7) \quad \begin{aligned} \mathbb{P}(|X_{t-h} - X_t| > \varepsilon) &= \int \mathbb{P}(|x - X_h| > \varepsilon | X_0 = x) \mathbb{P}(X_{t-h} \in dx) \\ &\leq \sup_{|x| < R} \mathbb{P}(|X_h - x| > \varepsilon | X_0 = x) + \varepsilon', \end{aligned}$$

which implies the result. ■

In some sense Proposition 4.1 suggests that it might be possible to localize condition (4.5). In fact, the following is a simple consequence of Aldous' result.

**THEOREM 4.5.** Let  $X^{(n)}$  be a time-homogeneous strong Markov process satisfying (4.2) and

$$(4.8) \quad \forall \varepsilon > 0 \forall R > 0 : \lim_{h \downarrow 0} \limsup_{n \rightarrow \infty} \sup_{\substack{|x| < R \\ t \leq h}} \mathbb{P}(|X_t^{(n)} - x| > \varepsilon | X_0 = x) = 0.$$

Then  $(X^{(n)})_{n \in \mathbb{N}}$  is  $J_1$ -tight.

**Proof.** Assume that (4.2) for  $\varepsilon' > 0$  and (4.8) hold and let  $(\tau_n)_{n \in \mathbb{N}}$  be such that  $\tau_n$  is a stopping time for  $X^{(n)}$ . Furthermore, let  $\varepsilon > 0$  and  $(t_n)_{n \in \mathbb{N}}$  be a sequence in  $[0, 1]$  with  $t_n \rightarrow 0$ . Hence

$$(4.9) \quad \mathbb{P}(|X_{(\tau_n + t_n) \wedge 1}^{(n)} - X_{\tau_n}^{(n)}| > \varepsilon) \leq \varepsilon' + \sup_{\substack{|x| < R \\ t \leq t_n}} \mathbb{P}(|X_t^{(n)} - x| > \varepsilon | X_0 = x)$$

and Theorem 4.3 implies the result. ■

So far we have discussed conditions for Markov processes. In the following we will adapt these conditions to the Markov chain setting.

Let  $Y^{(n)}$  be a time-homogeneous Markov chain  $(Y_0^{(n)}, Y_1^{(n)}, Y_2^{(n)}, \dots)$  on  $(\Omega, \mathcal{A}, \mathbb{P})$  and define the embeddings  $X^{n,T}(\omega)$  analogous to those in the previous section for each  $\omega$ . Then each  $X^{n,T}$  is a  $D[0, 1]$ -valued random variable. But in general,  $X^{n,T}$  is not a Markov process!

Starting with a Markov chain, a Markov process can be constructed by subordination: let  $(N_t)_{t \geq 0}$  be a Poisson process with intensity one, which is independent of the Markov chain. Then one can embed the Markov chain  $Y^{(n)}$  into a continuous time Markov process  $(Z_t^{(n)})_{t \geq 0}$  by setting

$$(4.10) \quad Z_t^{(n)} := Y_{N_{nt}}^{(n)} \text{ for } t \in [0, 1) \quad \text{and} \quad Z_1^{(n)} := Z_{1-}^{(n)}.$$

The  $J_1$  embedding and the Markov embedding are closely related as the following (technical) result shows.

LEMMA 4.1. *Let*

$$(4.11) \quad \sup_{k < |n-1-N_{n-}|} |Y_{((n-1) \wedge N_{n-})+k}^{(n)} - Y_{(n-1) \wedge N_{n-}}^{(n)}| \xrightarrow{\mathbb{P}} 0.$$

*Then  $X^{n,J_1}$  converges in distribution w.r.t.  $J_1$  if and only if  $Z^{(n)}$  converges in distribution w.r.t.  $J_1$ .*

**Proof.** Recall that  $X_t^{n,J_1} = Y_{[nt]}^{(n)}$  and  $Z_t^{(n)} = Y_{N_{nt}}^{(n)}$  for  $t < 1$ . The first  $(n-1) \wedge N_{n-}$  steps of these processes coincide by definition, they just appear at different times ( $\frac{k}{n}$  vs.  $k$ -th jump time of  $N_{nt}$ ). By a time change with a piecewise linear function  $\lambda \in \Lambda_c$  both paths (up to the waiting time after the  $((n-1) \wedge N_{n-})$ -th jump) can be made to coincide. The value of  $\|\lambda - \text{id}\|_\infty$  is attained at one of the jump times, thus (since  $N_{n-} = \lfloor n \frac{N_{n-}}{n} \rfloor$ ) one can show that

$$(4.12) \quad \|\lambda - \text{id}\|_\infty = \sup_{s \in [0,1)} \left| s - \frac{N_{ns}}{n} \right|.$$

The steps after  $(n-1) \wedge N_{n-}$  cannot be compensated by a time transformation. They have to be estimated explicitly. Therefore,

$$(4.13) \quad d_{J_1}(X^{n,J_1}, Z^{(n)}) \leq \sup_{s \in [0,1)} \left| s - \frac{N_{ns}}{n} \right| + \sup_{k < |n-1-N_{n-}|} |Y_{((n-1) \wedge N_{n-})+k}^{(n)} - Y_{(n-1) \wedge N_{n-}}^{(n)}|.$$

Since  $s - N_{ns}/n$  is a martingale, we find with Doob's maximal inequality that

$$(4.14) \quad \mathbb{E} \left( \sup_{s \in [0,1]} \left| s - \frac{N_{ns}}{n} \right| \right) \leq \sqrt{\mathbb{E} \left( \sup_{s \in [0,1]} \left| s - \frac{N_{ns}}{n} \right|^2 \right)} \\ \leq 2 \sqrt{\mathbb{E} \left( 1 - \frac{N_n}{n} \right)^2} = \frac{2}{\sqrt{n}}.$$

This implies with (4.11) that the  $J_1$  distance of  $X^{n,J_1}$  and  $Z^{(n)}$  converges in probability to zero. Thus the convergence in distribution of either  $X^{n,J_1}$  or  $Z^{(n)}$  implies also the convergence in distribution of the other, e.g. by [3], Theorem 4.1, p. 25. ■

Before analyzing condition (4.11) consider the question we have asked at the beginning: when does the converse of Corollary 3.1 hold? Suppose a step embedded (i.e., using the  $J_1$ -embedding of (3.2)) Markov chain converges, for example, in  $J_2$  but not in  $J_1$ . Then the limit (before identifying it with a  $D[0, 1]$ -function) has to have some states which it reaches by a jump and leaves instantaneously by another jump. The following condition is sufficient to ensure that such limit points do not exist:

$$(4.15) \quad \forall \varepsilon > 0 \forall R > 0 : \lim_{h \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{\substack{|x| < R \\ t \leq h}} \mathbb{P}(|Y_{[tn]}^{(n)} - x| > \varepsilon \mid Y_0^{(n)} = x) = 0.$$

Note that this is the Markov chain version of (4.8). It ensures, as (4.8), that the limit process is locally spatial-uniformly stochastically continuous from the right and together with the Markov chain version of (4.2), i.e.,

$$(4.16) \quad \forall \varepsilon > 0 \forall m \in \mathbb{N} \exists R > 0 : \sup_n \mathbb{P}(\|Y_{[\cdot nm]}^{(n)}\|_\infty > R) < \varepsilon,$$

we will get a  $J_1$ -tightness condition, see Theorem 4.6. Note that in (4.16) the extra  $m$  is needed since the subordinated chain might have more than  $n$  steps. In the context of  $D[0, 1]$  this might seem surprising, but for processes in  $D[0, \infty)$  the condition remains unchanged and becomes natural (cf. the last paragraph of this section). Now we can relate (4.11) to these conditions.

**LEMMA 4.2.** *Let (4.15) and (4.16) hold. Then for any sequence  $(l(n))_{n \in \mathbb{N}} \subset [0, \infty)$  with  $\frac{l(n)}{n} \rightarrow 0$*

$$(4.17) \quad \lim_{n \rightarrow \infty} \sup_{|x| < R} \mathbb{P} \left( \sup_{k \leq l(n)} |Y_k^{(n)} - x| > \varepsilon \mid Y_0^{(n)} = x \right) = 0$$

and (4.11) hold.

**Proof.** Let  $\tau_{B_\varepsilon(x)}^{(n)}$  denote the time of the first exit of  $X^{n,J_1}$  from the ball with center  $x$  and radius  $\varepsilon$ . Then (4.17) becomes

$$(4.18) \quad \lim_{n \rightarrow \infty} \sup_{|x| < R} \mathbb{P} \left( \tau_{B_\varepsilon(x)}^{(n)} \leq \frac{l(n)}{n} \mid X_0^{n,J_1} = x \right) = 0.$$

Suppose that the limit in (4.18) is not zero. Then the limiting process (if it exists) would not be locally spatial-uniformly stochastically continuous from the right, and this contradicts (4.15). Hence (4.17) must hold. Alternatively, for a direct proof note that analogously to [8], Lemma 2, p. 420, one gets for  $n$  large

$$(4.19) \quad \mathbb{P} \left( \sup_{k \leq l(n)} |Y_k^{(n)} - x| > \varepsilon \mid Y_0^{(n)} = x \right) \\ \leq \frac{\mathbb{P}(|Y_{l(n)}^{(n)} - x| > \frac{\varepsilon}{2} \mid Y_0^{(n)} = x)}{1 - \sup_{|y| < R} \mathbb{P}(|Y_k^{(n)} - y| > \frac{\varepsilon}{2} \mid Y_0^{(n)} = y) - \varepsilon_R},$$

where  $\varepsilon_R \in [0, 1)$  is some constant depending on  $R$ . Hence the statement follows by (4.15) and the estimate

$$(4.20) \quad \limsup_{n \rightarrow \infty} \sup_{|x| < R} \mathbb{P}(|Y_{l(n)}^{(n)} - x| > \frac{\varepsilon}{2} \mid Y_0^{(n)} = x) \\ \leq \limsup_{n \rightarrow \infty} \sup_{\substack{|x| < R \\ t \leq l(n)/n}} \mathbb{P}(|Y_{[tn]}^{(n)} - x| > \frac{\varepsilon}{2} \mid Y_0^{(n)} = x) \\ \leq \limsup_{n \rightarrow \infty} \sup_{\substack{|x| < R \\ t \leq h}} \mathbb{P}(|Y_{[tn]}^{(n)} - x| > \frac{\varepsilon}{2} \mid Y_0^{(n)} = x),$$

which holds for any  $h \in (0, 1]$ .

For the second part of the statement let  $\varepsilon, \varepsilon' > 0$ ,  $m \in \mathbb{N}$  and, using (4.16),  $R$  be such that  $\mathbb{P}(\|Y_{\lfloor \cdot n(m+2) \rfloor}\|_\infty > R) < \varepsilon'$ . Note that  $\mathbb{P}(|n-1-N_{n-}| \geq \sqrt{nm}) \leq \frac{1}{m}$ . Thus

$$(4.21) \quad \mathbb{P} \left( \sup_{k < |n-1-N_{n-}|} |Y_{((n-1) \wedge N_{n-})+k}^{(n)} - Y_{(n-1) \wedge N_{n-}}^{(n)}| > \varepsilon \right) \\ \leq \varepsilon' + \frac{1}{m} + \sup_{|x| < R} \mathbb{P} \left( \sup_{k \leq \sqrt{nm}} |Y_k^{(n)} - x| > \varepsilon \mid Y_0^{(n)} = x \right)$$

and (4.17) implies (4.11), since  $\frac{\sqrt{nm}}{n} \rightarrow 0$ . ■

Now we can prove a  $J_1$ -tightness condition for embedded Markov chains, i.e., conditions (2) and (3) in the following theorem.

**THEOREM 4.6.** *Let  $Y^{(n)}$ ,  $X^{n,J_1}$  and  $X$  be as above. Suppose the following conditions hold:*

- (1)  $X^{n,J_1} \xrightarrow{fdd} X$  on a dense subset of  $[0, 1]$ ;
- (2) condition (4.16), i.e.,

$$\forall \varepsilon > 0 \forall m \in \mathbb{N} \exists R > 0 : \sup_n \mathbb{P}(\|Y_{[\cdot, nm]}^{(n)}\|_\infty > R) < \varepsilon;$$

- (3) condition (4.15), i.e.,

$$\forall \varepsilon > 0 \forall R > 0 : \lim_{h \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{\substack{|x| < R \\ t \leq h}} \mathbb{P}(|Y_{[tn]}^{(n)} - x| > \varepsilon \mid Y_0^{(n)} = x) = 0.$$

Then

$$X^{n,J_1} \xrightarrow{d} X \text{ w.r.t. } J_1.$$

**Proof.** Assume that the conditions hold and let  $Z^{(n)}$  be the Markov embedding of  $Y^{(n)}$  as defined in (4.10). By Lemmas 4.1 and 4.2,

$$(4.22) \quad X^{n,J_1} \xrightarrow{d} X \text{ w.r.t. } J_1 \quad \Leftrightarrow \quad Z^{(n)} \xrightarrow{d} X \text{ w.r.t. } J_1.$$

Next note that

$$\begin{aligned} (4.23) \quad \mathbb{P}(\|Y_{N_n}^{(n)}\|_\infty > R) &= \mathbb{P}(\sup_{k \leq N_n} |Y_k^{(n)}| > R) \\ &= \sum_{l=1}^{\infty} \mathbb{P}(\sup_{k \leq l} |Y_k^{(n)}| > R) \mathbb{P}(N_n = l) \\ &\leq \mathbb{P}(\sup_{k \leq nm} |Y_k^{(n)}| > R) + \mathbb{P}(N_n \geq mn) \end{aligned}$$

with  $\mathbb{P}(N_n \geq mn) \leq \frac{1}{m}$  and condition (2) implies that  $Z^{(n)}$  satisfies (4.2). Furthermore, let  $\varepsilon > 0$ . Then, as in (4.23),

$$\begin{aligned} (4.24) \quad \mathbb{P}(|Y_{N_{tn}}^{(n)} - x| > \varepsilon \mid Y_0^{(n)} = x) \\ &= \sum_{l=0}^{\infty} \mathbb{P}(|Y_l^{(n)} - x| > \varepsilon \mid Y_0^{(n)} = x) \mathbb{P}(N_{tn} = l) \\ &\leq \sup_{l \leq [tnm]} \mathbb{P}(|Y_l^{(n)} - x| > \varepsilon \mid Y_0^{(n)} = x) + \frac{1}{m}. \end{aligned}$$

Condition (3) and the arbitrary choice of  $m$  imply that  $Z^{(n)}$  satisfies (4.8), letting therein  $h < \frac{1}{m}$ . Thus, by Theorem 4.5, the family  $(Z^{(n)})_{n \in \mathbb{N}}$  is  $J_1$ -tight.

Hence, for every sequence  $n_k \nearrow \infty$  there is a subsequence  $n_{k_l}$  such that  $Z^{(n_{k_l})}$  converges in distribution w.r.t.  $J_1$  to some limit, and  $X^{(n_{k_l})}$  must have the same limit in distribution. But, by (1), the limit of  $X^{(n_{k_l})}$  is  $X$  and it is independent of the sequence. Thus  $X^{n,J_1} \xrightarrow{d} X$  w.r.t.  $J_1$ . ■

Furthermore, (4.15) also yields a statement about the convergence of finite-dimensional distributions when switching the embedding.

LEMMA 4.3. *Let (4.15) hold and  $X^{n,\mathcal{T}} \xrightarrow{d} X$  w.r.t.  $\mathcal{T}$ , hence (by Theorem 4.1) also  $X^{n,\mathcal{T}} \xrightarrow{fdd} X$  on some  $I$  which is a dense subset of  $[0, 1]$ . Then*

$$X^{n,J_1} \xrightarrow{fdd} X \text{ on } I.$$

PROOF. Let the assumptions be satisfied and note that, by the definition of the embeddings,  $X_t^{n,\mathcal{T}} \in [[Y_{[tn]}^{(n)}, Y_{[tn]+1}^{(n)}]]$ . Fix  $\varepsilon > 0$ ; then, by Remark 4.1(4), there exists an  $R > 0$  such that

$$(4.25) \quad \sup_n \mathbb{P}(\|Y_{[\cdot n]}^{(n)}\|_\infty \geq R) < \varepsilon.$$

Thus

$$(4.26) \quad \begin{aligned} \mathbb{P}(|X_t^{n,J_1} - X_t^{n,\mathcal{T}}| > \varepsilon) \\ &= \mathbb{P}(|Y_{[tn]}^{(n)} - X_t^{n,\mathcal{T}}| > \varepsilon) \leq \mathbb{P}(|Y_{[tn]}^{(n)} - Y_{[tn]+1}^{(n)}| > \varepsilon) \\ &\leq \sup_{x < R} \mathbb{P}(|Y_1^n - x| > \varepsilon \mid Y_0 = x) + \sup_n \mathbb{P}(\|Y_{[\cdot n]}^{(n)}\|_\infty \geq R) \end{aligned}$$

and the first summand converges, by (4.15), to zero as  $n \rightarrow \infty$ . Since  $\varepsilon$  is arbitrary, the result follows by Slutsky's theorem as in Lemma 4.1. ■

Finally, we get the following extension to Corollary 3.1.

COROLLARY 4.1. *Let (4.15) and (4.16) hold. Then  $X^{n,\mathcal{T}}$  converges in distribution w.r.t.  $\mathcal{T}$  if and only if  $X^{n,J_1}$  converges in distribution w.r.t.  $J_1$ .*

PROOF. The direction from  $J_1$  to the other topologies is just Theorem 3.1 and Corollary 3.1.

For the converse, let (4.15) and (4.16) hold and  $X^{n,\mathcal{T}} \xrightarrow{d} X$  w.r.t.  $\mathcal{T}$ . By Lemma 4.3 we get  $X^{n,J_1} \xrightarrow{fdd} X$  on some dense subset of  $[0, 1]$  and hence (with Remark 4.1(4)) all conditions of Theorem 4.6 are satisfied. Therefore,  $X^{n,J_1} \xrightarrow{d} X$  w.r.t.  $J_1$ . ■

EXAMPLE 4.1. If  $(Y^{(n)})_{n \in \mathbb{N}}$  is the Markov chain approximation to a Feller process with symbol  $(x, \xi) \mapsto q(x, \xi)$  (see [5], [6] for the definitions and further details), then  $(Y_{Nnt}^{(n)})_{t \geq 0}$  is a Feller process with symbol  $(x, \xi) \mapsto n(e^{-\frac{1}{n}q(x, \xi)} - 1)$ . Thus, by [6], Corollary 5.2, p. 114, there exists a constant  $c$  such that

$$(4.27) \quad \mathbb{P}\left(\sup_{s \leq t} |Y_{Nns}^{(n)} - x| > r \mid Y_0^{(n)} = x\right) \leq ct \sup_{|y-x| \leq r} \sup_{|\xi| \leq 1/r} |n(e^{-\frac{1}{n}q(x, \xi)} - 1)|$$



for all  $x \in \mathbb{R}^d$ ,  $r, t > 0$ . Hence for any  $\varepsilon > 0$

$$\begin{aligned}
 (4.28) \quad & \limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{s \leq t} |Y_{[ns]}^{(n)} - x| > r \mid Y_0^{(n)} = x \right) \\
 & \leq \limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{s \leq t+\varepsilon} |Y_{N_{ns}}^{(n)} - x| > r, N_{n(t+\varepsilon)} \geq nt \mid Y_0^{(n)} = x \right) \\
 & \quad + \limsup_{n \rightarrow \infty} \mathbb{P}(N_{n(t+\varepsilon)} < nt) \\
 & \leq c(t + \varepsilon) \sup_{|y-x| \leq r} \sup_{|\xi| \leq 1/r} |q(y, \xi)|.
 \end{aligned}$$

Now (4.15) is satisfied if the supremum is finite, e.g. for  $q$  continuous. Assuming

$$(4.29) \quad \lim_{r \rightarrow \infty} \sup_{|y-x| \leq r} \sup_{|\xi| \leq 1/r} |q(y, \xi)| = 0 \quad \text{for all } x \in \mathbb{R}^d$$

and using (4.28) and (5.1) from [4], one finds an  $R$  such that condition (4.16) holds for large values of  $n$ , say  $n \geq N$ . Taking the maximum of this and the finite number of  $R$ 's corresponding to  $n < N$  yields (4.16).

Thus, we obtained a new proof of the convergence of the Markov chain approximation of Feller processes in  $J_1$ . Moreover, using the introduced embeddings, the approximations converge in the four Skorokhod topologies.

Finally, we want to emphasize that condition (4.15) ensures the local spatial-uniform stochastic continuity from the right for the limit. Hence, by Proposition 4.1, the limit is stochastically continuous and therefore, by Remark 4.1(3), we can consider the convergence of the finite-dimensional distributions on the whole interval  $[0, 1]$  (no exceptional times!). Hence, if  $X^{(n)}, X$  are  $D[0, \infty)$ -valued processes and  $X$  is stochastically continuous, then the extension from  $D[0, T]$  to  $D[0, \infty)$  (cf. Remark 2.2(2)) does not need a restriction of the time set, i.e., we have (see [11], Theorem 3')

$$\begin{aligned}
 (4.30) \quad & X^{(n)} \xrightarrow{d} X \text{ w.r.t. } J_1 \\
 & \Leftrightarrow X^{(n)}|_{[0, T]} \xrightarrow{d} X|_{[0, T]} \text{ w.r.t. } J_1 \text{ in } D[0, T] \text{ for all } T \in (0, \infty).
 \end{aligned}$$

Additionally, in this setting,  $X^{(n)} \xrightarrow{d} X$  w.r.t.  $\mathcal{T}$  implies (cf. Remark 4.1(4)) that (4.16) is satisfied.

## 5. PROOF OF LEMMA 2.1

We start with two elementary inequalities. Let  $a, b, c, d \in \mathbb{R}$ . Then

$$(5.1) \quad (a + b) \wedge (c + d) \leq (a + b \vee d) \wedge (c + b \vee d) = a \wedge c + b \vee d.$$

For  $x, y, z \in \mathbb{R}^d$  let  $\gamma \in [[x, z]]$  such that  $|y - \gamma| = |y - [[x, z]]|$ . Then

$$(5.2) \quad |x - y| \leq |x - \gamma| + |y - \gamma| \leq |x - z| + |y - [[x, z]]|.$$

Now, let  $\delta > 0$  and  $f : [0, 1] \rightarrow \mathbb{R}^d$ . Furthermore, let  $(t, t_1, t_2) \in T_1(\delta)$  and  $t_1^* \leq t_1, t_2^* \geq t_2$  such that  $(t, t_1^*, t_2^*) \in T_2(\delta)$ . Applying (5.2) and (5.1) yields

$$(5.3) \quad \begin{aligned} & |f(t) - f(t_1)| \wedge |f(t) - f(t_2)| \\ & \leq (|f(t) - f(t_1^*)| + |f(t_1) - [[f(t), f(t_1^*)]]|) \\ & \quad \wedge (|f(t) - f(t_2^*)| + |f(t_2) - [[f(t), f(t_2^*)]]|) \\ & \leq (|f(t) - f(t_1^*)| \wedge |f(t) - f(t_2^*)|) \\ & \quad + (|f(t_1) - [[f(t), f(t_1^*)]]| \vee |f(t_2) - [[f(t), f(t_2^*)]]|) \\ & \leq \Delta_{J_2}(\delta, f) + \Delta_{M_1}(\delta, f). \end{aligned}$$

Thus

$$(5.4) \quad \Delta_{J_1}(\delta, f) \leq \Delta_{J_2}(\delta, f) + \Delta_{M_1}(\delta, f),$$

since  $(t, t_1, t_2) \in T_1(\delta)$  was arbitrary. ■

**Acknowledgments.** The author is grateful for the comments of an anonymous referee who helped to improve the paper.

## REFERENCES

- [1] D. Aldous, *Stopping times and tightness*, Ann. Probab. 6 (2) (1978), pp. 335–340.
- [2] F. Avram and M. S. Taqqu, *Probability bounds for M-Skorohod oscillations*, Stochastic Process. Appl. 33 (1) (1989), pp. 63–72.
- [3] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York–London 1968.
- [4] B. Böttcher, *On the construction of Feller processes with unbounded coefficients*, Electron. Commun. Probab. 16 (2011), pp. 545–555.
- [5] B. Böttcher and R. L. Schilling, *Approximation of Feller processes by Markov chains with Lévy increments*, Stoch. Dyn. 9 (1) (2009), pp. 71–80.
- [6] B. Böttcher, R. Schilling, and J. Wang, *Lévy-Type Processes: Construction, Approximation and Sample Path Properties*, Lecture Notes in Math., Vol. 2099, Springer, Cham 2013.
- [7] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley, New York 1986.
- [8] I. I. Gikhman and A. V. Skorokhod, *The Theory of Stochastic Processes*, Springer, Berlin 2007.
- [9] J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, second edition, Springer, Berlin 2003.
- [10] A. Jakubowski, *A non-Skorohod topology on the Skorohod space*, Electron. J. Probab. 2 (1997), pp. 1–21.
- [11] T. Lindvall, *Weak convergence of probability measures and random functions in the function space  $D(0, \infty)$* , J. Appl. Probab. 10 (1973), pp. 109–121.
- [12] P. Meyer and W. A. Zheng, *Tightness criteria for laws of semimartingales*, Ann. Inst. Henri Poincaré Probab. Stat. 20 (4) (1984), pp. 353–372.

- [13] J. L. Pomarede, *A Unified Approach via Graphs to Skorohod's Topologies on the Function Space  $D$* , Ph.D. Thesis, Yale University, ProQuest LLC, Ann Arbor, MI, 1976.
- [14] K. Sato, *A note on convergence of probability measures on  $C$  and  $D$* , Ann. Sci. Kanazawa Univ. 14 (1977), pp. 1–5.
- [15] A. V. Skorokhod, *Limit theorems for stochastic processes*, Teor. Veroyatn. Primen. 1 (3) (1956), pp. 289–319.
- [16] F. Topsøe, *A criterion for weak convergence of measures with an application to convergence of measures on  $D[0, 1]$* , Math. Scand. 25 (1969), pp. 97–104.
- [17] W. Whitt, *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*, Springer, New York 2002.

Björn Böttcher  
TU Dresden, Fachrichtung Mathematik  
Institut für Mathematische Stochastik  
01062 Dresden, Germany  
E-mail: bjoern.boettcher@tu-dresden.de

*Received on 26.11.2016;  
revised version on 2.2.2018*

---