



UNIVERSITY OF LEEDS

This is a repository copy of *On mean-field (GI/GI/1) queueing model: existence and uniqueness*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/149505/>

Version: Accepted Version

---

**Article:**

Veretennikov, AY (2020) On mean-field (GI/GI/1) queueing model: existence and uniqueness. *Queueing Systems*, 94 (3-4). pp. 243-255. ISSN 0257-0130

<https://doi.org/10.1007/s11134-019-09626-x>

---

© 2019, Springer Science+Business Media, LLC, part of Springer Nature. This is an author produced version of a journal article published in *Queueing Systems*. Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# On mean-field $GI/GI/1$ queueing model: existence and uniqueness <sup>\*</sup>

A.Yu. Veretennikov<sup>†</sup>

December 4, 2018

## Abstract

A mean-field extension of the queueing system  $GI/GI/1$  is considered. The process is constructed as a Markov solution of a martingale problem. Uniqueness in distribution is also established under a bit different sets of assumptions on intensities.

keywords:  $GI/GI/$ ; mean-field; existence; weak uniqueness; Skorokhod lemma  
MSC: 60-02; 60K25; 90B22

## 1 Introduction

Mean-field approach in the theory of queueing systems allows to take into consideration large interacting ensembles of queues by using the idea of replacing these interactions by a suitable “mean field”. This approach showed fruitful in systems with countable and more general state spaces, see, for example, [1], [2], [3], [5], [6], [11], and the references therein. However, to the best of the author’s knowledge, so far there was no general method of constructing mean-field extensions of a basic queueing model such as  $GI/GI/1$  – or, more precisely,  $GI/GI/1/\infty$  – in the literature. In this work we propose such a method under certain restrictions on intensities of arrivals and service, which intensities may both depend on the state as well as

---

<sup>\*</sup>This study has been funded by the Russian Academic Excellence Project ‘5-100’ (section 2.2) and by the RFBR grant 17-01-00633<sub>a</sub> (section 2.1).

<sup>†</sup>School of Mathematics, University of Leeds, Leeds, LS2 9JT, UK; email: a.veretennikov@leeds.ac.uk & National Research University Higher School of Economics, Moscow, Russian Federation, & Institute for Information Transmission Problems, Moscow, Russian Federation

on the marginal distribution of the process. This kind of dependence is natural in the study of limits of so-called multi-agent systems, or in other words, of a large number of weakly interacting queues (cf. again the references [1], [2], [3], [5], [6], [11]); in this paper we do not study such a setting because it seems reasonable to separate the problem of convergence from the problem of existence of the limiting process. Existence and weak uniqueness are discussed on the basis of compactness of measures, Skorokhod’s unique probability space Lemma, total variation metric and a Skorokhod–Girsanov’s density of measures theorem for jump processes. The basis for the study in the sections 2.1 and 2.2 is a technique similar to the one developed in the preprint on McKean-Vlasov stochastic equations [16]. Note that in some earlier papers and monographs intensities of transitions in queueing systems were assumed to depend only on the number of customers in the system. This means that the (conditional) distributions of the service times as well as the arrival times are exponential. In some situations this is not realistic. In particular, it does not allow heavy tails. Hence, motivation of our extension to a more general dependence is simple: it should relax the assumption of exponential arrival and service times. In the Theorem 1 below heavy tails of (conditional) service time distributions are allowed. In the Theorem 2 currently heavy tails are not possible, although, both service and arrival times still can be more general than exponential. Nevertheless, the author’s belief is that this is a technical matter to include the case of heavy tails in the conditions for uniqueness, too, which problem will be hopefully resolved in the near future. It is likely that the established results may be useful in the area of mathematical theory of reliability which is notably known to have the same basic formulae as queueing theory, see [8].

The paper consists of Introduction, Main section and t So, we can rigorously apply the principle of “complete probabilities” (or, better “complete expectation”) taking a summation as earlier in the intuitive version:wo Appendices. The Main section consists of two subsections related to the two topics shown in the title, with one theorem in each and with the proof of this theorem. The Appendix 1 contains the statement of Skorokhod’s Lemma about an equivalence of weak convergence of a sequence of processes to a convergence in probability of processes with the same distributions on a unique probability space, included for the reader’s convenience. The Appendix 2 offers a strict version of a non-rigorous calculus in the middle of the proof of the Theorem 1.

## 2 Main section

The state space of the process under consideration is the union

$$\mathcal{X} := (0, x) \cup \bigcup_{k=1}^{\infty} (k, x, y), \quad x, y \geq 0.$$

The meaning of  $k$  here is the number of “customers” in the system; the value  $x$  stands for the elapsed time from the last arrival, while  $y$  signifies the elapsed time of the current service. There is only one server which works without breaks (if there is at least one customer in the system) and it is always in a working state. All newly arrived customers stand in a queue of the infinite capacity, and for simplicity only we assume the FIFO discipline of service (First In First Out). It is assumed that at any time  $t$  at any state  $X = (k, x, y)$  (or  $X = (0, x)$  for  $k = 0$ ) there are *intensities* of service  $\Lambda^-[t, X_t, \mu_t]$  and arrivals  $\Lambda^+[t, X_t, \mu_t]$ , where  $\mu_t$  is the distribution of the random variable  $X_t$  itself. Note that occasionally we will be using notation  $(0, x, y)$  where  $y$  is a “false” variable, i.e., we identify all such triples with any  $y$  with a couple  $(0, x)$ . It will be sometimes convenient to denote  $k = k(X), x = x(X), y = y(X)$  for  $X = (k, x, y)$ . For technical reasons it is convenient to define the distance between two states  $X = (k, x, y)$  and  $X' = (k', x', y')$  as

$$\rho(X, Y) := |k - k'| + |x - x'| + |y - y'|.$$

The process is piecewise-linear Markov (PLMP, see [7]), which simply means that between any two subsequent jumps the continuous components  $(x, y)$  if  $k > 0$ , or just  $x$  if  $k = 0$  – grow linearly with rate 1, while the discrete component  $n$  remains unchanged.

### The assumptions:

(A1) There are Borel measurable, non-negative and bounded functions  $\lambda^+(t, X, Y)$  and  $\lambda^-(t, X, Y)$ .

(A2)

$$\Lambda^\pm[t, X, \mu] = \int \lambda^\pm(t, X, Y) \mu(dY)$$

(NB: Automatically, both  $\Lambda^\pm$  are Borel functions of  $(t, X)$ .)

(A3) The functions  $\lambda^\pm(t, X, Y)$  are continuous in all variables.

(A4) The functions  $\lambda^\pm(t, X, Y)$  are uniformly bounded away from zero except for  $\lambda^-(t, (0, x), (n, y, y')) = 0$ , for any  $x, y, y' \geq 0$  (no jump down from any state with zero customers).

Let us emphasize that neither Lipschitz nor any other regularity of the intensities  $\lambda^\pm$  is assumed, except for continuity in (A3). Probably continuity in  $t$  may be relaxed. Note that functions of time and two state variables will be denoted with round brackets, e.g., as  $\lambda^\pm(t, X, Y)$ , while functions where the third variable is a measure will be written with square brackets like  $\Lambda^\pm[t, X, \mu]$ .

In particular, intensities  $\Lambda^\pm$  may include additional (non-negative) terms not depending on the measure, say,  $\lambda_0^\pm(t, X)$ ; this may be helpful so as to justify the assumption (A4), as the terms  $\lambda_0^\pm(t, X)$  can be reasonably assumed uniformly bounded away from zero. Emphasize that (A4) will only be used in the Theorem 2, and as was mentioned earlier, there is a plausible hypothesis that this assumption even in this theorem may be relaxed; however, we postpone this issue till further investigations.

For  $X \in \mathcal{X}$  let us denote

$$\begin{aligned} X^+ &:= (k + 1, 0, y), & \text{for } X = (k, x, y), \ k \geq 0, \\ X^- &:= (k - 1, x, 0), & \text{for } X = (k, x, y), \ k \geq 1, \\ X + \delta &:= (k, x + \delta, y + \delta), & \text{for } X = (k, x, y), \ k \geq 0, \end{aligned}$$

Naturally,  $X^-$  is not defined for  $X = (0, x)$ .

## 2.1 Existence

The initial value  $X_0$  of the process may be distributed, which distribution is denoted by  $\mu_0$  (in particular,  $\mu_0$  may be a delta-measure concentrated at one point).

**Theorem 1** *Let the assumptions (A1)–(A3) be satisfied. Then for any initial distribution  $\mu_0$  on  $\mathcal{X}$ , on some probability space there exists a Markov process  $(X_t, t \geq 0)$  with marginal distributions  $\mu_t$  and intensities  $\Lambda[t, X_t, \mu_t]$ ,  $H[t, X_t, \mu_t]$ ; in other words, such that for any bounded continuous function  $g(X)$  with bounded continuous derivatives in  $(x, y)$ , the expression*

$$M_t := g(X_t) - g(X_0) - \int_0^t L[s, X_s, \mu_s]g(X_s) ds \tag{1}$$

is a martingale, where for  $X = (k, x, y)$ ,  $X' = (k', x', y')$ ,  $k \geq 0$ ,  $t \geq 0$ ,

$$\begin{aligned} L[t, X', \mu]g(X) &:= \Lambda^+[t, X', \mu](g(X^+) - g(X)) \\ &+ 1(n > 0)\Lambda^-[t, X', \mu](g(X^-) - g(X)) \\ &+ \frac{\partial}{\partial x}g(k, x, y) + 1(k > 0)\frac{\partial}{\partial y}g(k, x, y). \end{aligned}$$

Moreover, for any given measure-valued function  $(\mu_s, s \geq 0)$  in  $L[s, X_s, \mu_s]$ , the martingale problem (see [10]) (1) has a (weakly) unique solution.

The processes  $(X_t, t \geq 0)$ , or later in the proof of the Theorem 1  $(X_t^n, t \geq 0)$  for  $n \geq 1$  being constructed, let us introduce on some probability space *independent* equivalent processes  $(\xi_t, t \geq 0)$ , or, respectively,  $(\xi_t^n, t \geq 0)$ ; let  $\mathbb{E}'$  stand in all cases for the integration with respect to the *third variable*, e.g.,

$$\mathbb{E}'\lambda^\pm(t, X_t, \xi_t) := \int \lambda^\pm(t, X_t, Y)\mu_t(dY),$$

or

$$\mathbb{E}'\lambda^\pm(t, X_t^n, \xi_t^n) := \int \lambda^\pm(t, X_t^n, Y)\mu_t^n(dY),$$

where  $\mu_t^n$  is the distribution of  $X_t^n$ ; this will be repeated in the proof of the Theorem.

Recall that

$$\Lambda^\pm[t, X', \mu] = \int \lambda^\pm(t, X', y)\mu(dy) = \mathbb{E}'\lambda^\pm(t, X', \xi),$$

where  $\xi$  has distribution  $\mu$ . So, the operator  $L$  can be also presented in the form,

$$L[t, X', \mu]g(X) = \mathbb{E}'L(t, X', \xi)g(X),$$

for  $X = (k, x, y)$ , where

$$\begin{aligned} L(t, X', y)g(X) &:= \lambda^+(t, X', y)(g(X^+) - g(X)) \\ &+ 1(k > 0)\lambda^-(t, X', y)(g(X^-) - g(X)) \\ &+ \frac{\partial}{\partial x}g(k, x, y) + 1(k > 0)\frac{\partial}{\partial y}g(k, x, y). \end{aligned}$$

Note that the (possibly extended) *generator* of the Markov process  $X_t$  (cf., for example, [10]) is, of course,  $L[t, X, \mu_t]$ ; different variables  $X$  and  $X'$  in the definition above are needed only for the convenience of the proof. Equivalently, *Dynkin's identity* holds true for any function  $g(X)$  from the same class,

$$\mathbb{E}_{0, X_0} g(X_t) = g(X_0) + \mathbb{E}_{0, X_0} \int_0^t L[s, X_s, \mu_s] g(X_s) ds. \quad (2)$$

Moreover, equivalently, for any  $0 \leq t_1 < t_2 \dots < t_{m+1}$ , and for any Borel bounded functions  $\phi_k(X)$ ,  $X \in \mathcal{X}$ ,

$$\mathbb{E}_{0, X_0} \left( g(X_{t_{m+1}}) - g(X_{t_m}) - \int_{t_m}^{t_{m+1}} L[s, X_s, \mu_s] g(X_s) ds \right) \prod_{k=1}^m \phi_k(X_{t_k}) = 0. \quad (3)$$

Also note that for the validity of the equation (3) for any Borel bounded continuous functions  $\phi_k(X)$ ,  $X \in \mathcal{X}$  it suffices to verify it for any bounded continuous  $\phi_k(X)$ ,  $X \in \mathcal{X}$ , due to the property of measures on  $\mathbb{R}^d$  which are uniquely determined by the values of their integrals with continuous bounded functions (see, e.g., [13, Theorem 1.2.4]). The latter formula (3) may be called one more version of Dynkin's identity; it will be the basis for establishing existence. With a bit of abuse of the standard terminology, (3) may also be called a martingale problem. Note, however, that weak uniqueness (= uniqueness in distribution) in this Theorem *given*  $(\mu_s, s \geq 0)$  does not mean a total uniqueness in distribution of the process under construction because there is no claim of uniqueness of  $(\mu_s, s \geq 0)$ , not even talking about a distribution in the space of trajectories.

*Proof of Theorem 1.* For any  $n \geq 1$  consider a process  $(X_t^n)$ , with initial data  $X_0^n = X_0$  and intensities of jumps up and down, respectively,

$$\Lambda^+[t, X_{(t-1/n)_+}^n, \mu_{(t-1/n)_+}^n], \quad \Lambda^-[t, X_{(t-1/n)_+}^n, \mu_{(t-1/n)_+}^n].$$

The process  $(X_t^n)$  for each  $n$  is constructed by induction successfully on the intervals  $[0, 1/n], [1/n, 2/n]$ , etc. Due to the boundedness assumption on both intensities, there is no blow up and the processes for any  $n$  are defined for any  $t \geq 0$  as càdlàg processes without any point of jump accumulating. Moreover, for any  $t$  probability of jump exactly at time  $t$  for any  $X^n$  equals zero.

Recall that the processes  $(X_t^n, t \geq 0)$  for  $n \geq 1$  being constructed, we introduce on some probability space *independent* equivalent processes  $(\xi_t^n, t \geq 0)$ , and that  $\mathbb{E}'$

stands in all cases for the integration with respect to the *third variable*, e.g.,

$$\mathbb{E}'\lambda^\pm(t, X_t^n, \xi_t^n) := \int \lambda^\pm(t, X_t^n, Y)\mu_t^n(dY).$$

It can be checked that the assumptions of the Lemma 1 from the Appendix are satisfied.

Indeed, given  $\epsilon > 0$ , let us firstly choose  $c_0 > 0$  so that

$$\mathbb{P}_{0, \mu_0}(|X_0| > c_0) < \epsilon/2.$$

On the event  $(|X_0| \leq c_0)$  we have for any  $n$ ,

$$\sup_{0 \leq t \leq T} (|x(X_t^n)| + |y(X_t^n)|) \leq c_0 + 2T.$$

Further, since the intensity of jump up  $\Lambda^+$  is bounded, say,  $\Lambda^+ \leq \bar{\lambda}$ , then the number of jumps up on  $[0, T]$  is bounded in probability, that is, uniformly with respect to  $n$  (recall that  $X_0^n = X_0$ ),

$$\mathbb{P}(\sup_{0 \leq t \leq T} k(X_t^n) - k(X_0) > c) \rightarrow 0, \quad c \rightarrow \infty.$$

Hence, the first condition (9) of the Lemma 1 for the family of processes  $(X^n)$  follows.

To check the second condition (10), note that

$$\begin{aligned} \mathbb{P}(|k(X_t^n) - k(X_s^n)| > 0) &\leq \mathbb{P}(\text{at least one jump on } [s, t]) \\ &\leq 1 - \exp(-\bar{\lambda}|t - s|) \leq \bar{\lambda}|t - s|. \end{aligned}$$

Next, for  $\epsilon > 2h \geq 2|t - s|$ ,

$$\mathbb{P}(|x(X_t^n) - x(X_s^n)| + |y(X_t^n) - y(X_s^n)| > \epsilon; \text{ no jumps on } [s, t]) = 0.$$

So, the equality (10) for the family of processes  $(X^n)$  follows as required.

Hence, on some new probability space there exist equivalent processes  $(\tilde{X}_t^n, \tilde{\xi}_t^n)$ , and a limiting pair  $(\tilde{X}_t, \tilde{\xi}_t)$  such that for some subsequence  $(\tilde{X}_t^{n'}, \tilde{\xi}_t^{n'}) \xrightarrow{\mathbb{P}} (\tilde{X}_t, \tilde{\xi}_t)$ ,  $n' \rightarrow \infty$ , for each  $t$ . It follows due to the boundedness of all intensities that the limiting process  $(\tilde{X}_t, \tilde{\xi}_t)$  is also stochastically continuous. More than that, with probability



one the pair  $(\tilde{X}_t, \tilde{\xi}_t)$  is a process with a finite number of jumps on any bounded interval. Moreover, the property  $\lim_{h \downarrow 0} \sup_n \sup_{t, s \leq T; |t-s| \leq h} \mathbb{P}(|\tilde{X}_t^n - \tilde{X}_s^n| > \epsilon) = 0$  implies that for any  $\epsilon > 0$  there is a following convergence in probability,

$$\tilde{X}_{(t-1/n')_+}^{n'} \xrightarrow{\mathbb{P}} \tilde{X}_t, \quad n' \rightarrow \infty.$$

For the sequel, denote by  $\tilde{\mathcal{F}}_t^n$  the sigma-algebra  $\sigma(\tilde{X}_s^n : 0 \leq s \leq t)$ , and again slightly abusing notations we will drop the upper index  $n$  here.

The analogue of Dynkin's formula (3) for the pair  $(\tilde{X}_t^{n'}, \tilde{\xi}_t^{n'})$  reads,

$$\begin{aligned} \mathbb{E}_{0, X_0} \left[ \left( g(\tilde{X}_{t_{m+1}}^{n'}) - g(\tilde{X}_{t_m}^{n'}) - \int_{t_m}^{t_{m+1}} \mathbb{E}' L(s, \tilde{X}_{(s-1/n')_+}^{n'}, \tilde{\xi}_{(s-1/n')_+}^{n'}) g(\tilde{X}_s^{n'}) ds \right) \right. \\ \left. \times \prod_{k=1}^m \phi_k(\tilde{X}_{t_k}^{n'}) \right] = 0, \quad t_1 < \dots < t_m < t_{m+1}. \end{aligned} \quad (4)$$

The formula (4) follows straightforward from the “complete expectation” arguments (cf., for example, [18]) and from the definition of intensities.

Indeed, irrespectively on whether or not the intensities depend on the current state of the process  $(\tilde{X}_t)$ , or on some past values with a delay, they intuitively mean that for  $\delta > 0$  we have,

$$\begin{aligned} \mathbb{P} \left( \tilde{X}_{t+\delta}^n = (k+1, x', y + \delta), 0 \leq x' \leq \delta \mid \tilde{\mathcal{F}}_t^n; \tilde{X}_t^n = (k, x, y), \tilde{X}_{(t-1/n)_+}^n = (k', x', y') \right) \\ = \mathbb{E}' \Lambda^+(t, (k', x', y'), \tilde{\xi}_{(t-1/n)_+}^n) \delta + o(\delta), \end{aligned}$$

and if  $k(\tilde{X}_{(t-1/n)_+}^n) > 0$ ,

$$\begin{aligned} \mathbb{P} \left( \tilde{X}_{t+\delta}^n = (k-1, x + \delta, y'), 0 \leq y' \leq \delta \mid \tilde{\mathcal{F}}_t^n; \tilde{X}_t^n = (k, x, y), \tilde{X}_{(t-1/n)_+}^n = (k', x', y') \right) \\ = \mathbb{E}' \Lambda^-(t, (k', x', y'), \tilde{\xi}_{(t-1/n)_+}^n) \delta + o(\delta), \end{aligned}$$

and finally,

$$\begin{aligned} \mathbb{P} \left( \tilde{X}_{t+\delta}^n = (k, x + \delta, y + \delta) \mid \tilde{\mathcal{F}}_t^n; \tilde{X}_t^n = (k, x, y), \tilde{X}_{(t-1/n)_+}^n = (k', x', y') \right) \\ = 1 - \mathbb{E}' (\Lambda^+ + \Lambda^-)(t, (k', x', y'), \tilde{\xi}_{(t-1/n)_+}^n) \delta + o(\delta). \end{aligned}$$

Continuity of the intensities were implicitly used here; later this assumption will be dropped.

Hence, we can write for any bounded function  $g$  in the domain of the operator  $L$ ,

$$\begin{aligned} \mathbb{E} \left( g(\tilde{X}_{t+\delta}^n) | \tilde{\mathcal{F}}_t^n \right) &= g(\tilde{X}_t^{n,+}) \mathbb{E}' \Lambda^+(t, \tilde{X}_{(t-1/n)_+}^n, \tilde{\xi}_{(t-1/n)_+}^n) \delta \\ &\quad + g(\tilde{X}_{(t-1/n)_+}^{n,-}) \mathbb{E}' \Lambda^-(t, \tilde{X}_t^n, \tilde{\xi}_{(t-1/n)_+}^n) \delta \\ &\quad + g(\tilde{X}_t^n + \delta) \left( 1 - \mathbb{E}'(\Lambda^+ + \Lambda^-)(t, \tilde{X}_{(t-1/n)_+}^n, \tilde{\xi}_{(t-1/n)_+}^n) \delta \right) + o(\delta), \end{aligned}$$

as  $\delta \downarrow 0$ . Therefore,

$$\begin{aligned} &\mathbb{E} \left( g(\tilde{X}_{t+\delta}^n) - g(\tilde{X}_t^n) | \tilde{\mathcal{F}}_t^n \right) \\ &= g(\tilde{X}_t^{n,+}) \mathbb{E}' \Lambda^+(t, \tilde{X}_{(t-1/n)_+}^n, \tilde{\xi}_{(t-1/n)_+}^n) \delta \\ &\quad + g(\tilde{X}_t^{n,-}) \mathbb{E}' \Lambda^-(t, \tilde{X}_{(t-1/n)_+}^n, \tilde{\xi}_{(t-1/n)_+}^n) \delta \\ &\quad + g(\tilde{X}_t^n + \delta) \left( 1 - \mathbb{E}'(\Lambda^+ + \Lambda^-)(t, \tilde{X}_{(t-1/n)_+}^n, \tilde{\xi}_{(t-1/n)_+}^n) \delta \right) - g(\tilde{X}_t^n) + o(\delta) \\ &= \mathbb{E}' L(t, \tilde{X}_{(t-1/n)_+}^n, \tilde{\xi}_{(t-1/n)_+}^n) g(\tilde{X}_t^n) \delta + o(\delta). \end{aligned}$$

Applying now still intuitively the principle of “complete probabilities” (here more accurately it could be called “complete expectation”), that is, taking a summation we can obtain the equation (4). Indeed, let us split the interval  $[t_m, t_{m+1}]$  into  $N$  equal small sub-intervals,  $t_m = s_0 < \dots < s_N = t_{m+1}$  so that  $(s_{i+1} - s_i) =: \delta$  and

write down,

$$\begin{aligned}
\mathbb{E} \left( g(\tilde{X}_{t_{m+1}}^n) - g(\tilde{X}_{t_m}^n) | \tilde{\mathcal{F}}_{t_m}^n \right) &= \sum_{i=0}^{N-1} \mathbb{E} \left( \mathbb{E} \left( g(\tilde{X}_{s_{i+1}}^n) - g(\tilde{X}_{s_i}^n) | \tilde{\mathcal{F}}_{s_i}^n \right) | \tilde{\mathcal{F}}_{t_m}^n \right) \\
&= \sum_{i=0}^{N-1} \left[ \mathbb{E} \left( \mathbb{E}' L(s_i, \tilde{X}_{(s_i-1/n)_+}^n, \tilde{\xi}_{(s_i-1/n)_+}^n) g(\tilde{X}_{s_i}^n) | \tilde{\mathcal{F}}_{t_m}^n \right) \delta + o(\delta) \right] \\
&= \mathbb{E} \left( \int_{t_m}^{t_{m+1}} \mathbb{E}' L(s, \tilde{X}_{(s-1/n)_+}^n, \tilde{\xi}_{(s-1/n)_+}^n) g(\tilde{X}_s^n) ds | \tilde{\mathcal{F}}_{t_m}^n \right) + o(1), \quad (5)
\end{aligned}$$

which implies the equation (4) (recall that we dropped prime at  $n$  to simplify notations). This intuitive calculus may be made strict; for the convenience of the reader we provide such a rigorous version in the Appendix 2.

However, another easier way is just to recall the definition of intensity via the martingale property. In our particular case – with jumps of the first component of  $X$  just up or down – the term intensity is applied to the random variable  $\Lambda = \Lambda^\pm(t, \omega)$  if and only if for any bounded measurable function  $g(X)$  ( $X = (n, x, y)$ ) with bounded derivatives with respect to  $x$  and  $y$  and for any  $t_0 \geq 0$ , the process defined by the expression

$$M_t^n := g(\tilde{X}_t^n) - g(\tilde{X}_{t_0}^n) - \int_{t_0}^t \mathbb{E}' L(s, \tilde{X}_{(s-1/n)_+}^n, \tilde{\xi}_{(s-1/n)_+}^n) g(\tilde{X}_s^n) ds, \quad t \geq t_0,$$

is a martingale: see, e.g., [15, Sec.3.III.5.5] for pure jump processes and for indicator functions, which extends straightforward to our case and to Borel measurable functions; the intuition behind this definition has been offered in the little calculus above (see also the Appendix 2). So,

$$\mathbb{E} \left[ g(\tilde{X}_{t_{m+1}}^n) - g(\tilde{X}_{t_m}^n) - \int_{t_m}^{t_{m+1}} \mathbb{E}' L(s, \tilde{X}_{(s-1/n)_+}^n, \tilde{\xi}_{(s-1/n)_+}^n) g(\tilde{X}_s^n) ds | \tilde{\mathcal{F}}_{t_m}^n \right] = 0 \text{ (a.s.)}.$$

Therefore, it follows that

$$\begin{aligned} \mathbb{E}_{0, X_0} \left( g(\tilde{X}_{t_{m+1}}^{n'}) - g(\tilde{X}_{t_m}^{n'}) - \int_{t_m}^{t_{m+1}} \mathbb{E}' L(s, \tilde{X}_{(s-1/n')_+}^{n'}, \tilde{\xi}_{(s-1/n')_+}^{n'}) g(\tilde{X}_s^{n'}) ds \right) \prod_{k=1}^m \phi_k(\tilde{X}_{t_k}^{n'}) \\ = \mathbb{E}_{0, X_0} \left( \prod_{k=1}^m \phi_k(\tilde{X}_{t_k}^{n'}) \right) \mathbb{E} \left[ g(\tilde{X}_{t_{m+1}}^n) \right. \\ \left. - g(\tilde{X}_{t_m}^n) - \int_{t_m}^{t_{m+1}} \mathbb{E}' L(s, \tilde{X}_{(s-1/n')_+}^{n'}, \tilde{\xi}_{(s-1/n')_+}^{n'}) g(\tilde{X}_s^n) ds \middle| \tilde{\mathcal{F}}_{t_m}^n \right] = 0, \end{aligned}$$

as required. This justifies the equation (4).

Further, by continuity of  $\lambda$  and  $h$ , and due to the stochastic continuity of the processes  $\tilde{X}$  and  $\tilde{\xi}$ , and since all integrand expressions are bounded, and by virtue of Lebesgue's bounded convergence Theorem, we obtain from (4) in the limit with *continuous bounded* functions  $(\phi_k)$ ,

$$\mathbb{E}_{0, X_0} \left( g(\tilde{X}_{t_{m+1}}) - g(\tilde{X}_{t_m}) - \int_{t_m}^{t_{m+1}} \mathbb{E}' L(s, \tilde{X}_s, \tilde{\xi}_s) g(\tilde{X}_s) ds \right) \prod_{k=1}^m \phi_k(\tilde{X}_{t_k}) = 0. \quad (6)$$

Since the distribution of the random variable  $\tilde{\xi}_t$  is the same as the one of  $\tilde{X}_t$  – let us denote it by  $\tilde{\mu}_t$  – then (6) can be equivalently written as

$$\mathbb{E}_{0, X_0} \left( g(\tilde{X}_{t_{m+1}}) - g(\tilde{X}_{t_m}) - \int_{t_m}^{t_{m+1}} Lg(s, \tilde{X}_s, \tilde{\mu}_s) ds \right) \prod_{k=1}^m \phi_k(\tilde{X}_{t_k}) = 0. \quad (7)$$

As was mentioned earlier, due to the properties of measures on  $\mathbb{R}^d$  the formula (7) holds true for any Borel bounded functions  $(\phi_k)$ , too. Due to [9], solution of the “martingale problem” (7) – or, more precisely, of the martingale problem

$$M_t := g(\tilde{X}_t) - g(\tilde{X}_0) - \int_0^t Lg(s, \tilde{X}_s, \tilde{\mu}_s) ds, \quad t \geq 0, \quad \text{is a martingale,} \quad (8)$$

with a given family of marginal measures  $(\tilde{\mu}_s, s \geq 0)$  is unique. Hence, according to [12], or [10, Theorem 4.4.2] the limiting process  $\tilde{X}$  is Markov. The form of its generator with the required intensities  $\Lambda^\pm$  follows from (7). This finishes the proof of the Theorem 1.

## 2.2 Weak uniqueness

Emphasize that we will use essentially boundedness of all intensities and the condition that they are (uniformly) bounded away from zero. While it is clear that the boundedness from above may be relaxed for the purpose of establishing existence – e.g., under Lyapunov type conditions, or under a linear growth, or otherwise, – and that boundedness away from zero is not required for the existence at all, yet for the uniqueness both boundedness from above and from below seems essential (although also could be, apparently, slightly relaxed). On the other hand, continuity of the intensities in this section is not necessary and they are not assumed.

**Theorem 2** *Let the assumptions (A1)–(A2) and (A4) be satisfied. Then, for any fixed distribution  $\mathcal{L}(X_0)$ , there exists no more than one distribution of the process  $(X_t, t \geq 0)$  with required intensities  $\Lambda^+[t, X, \mu_t]$  and  $\Lambda^-[t, X, \mu_t]$ .*

Recall that no Lipschitz assumptions on the intensities are assumed. In the calculus the total variation metric will be used.

Let  $\bar{\Lambda}[t, X, \mu] := \Lambda^+[t, X, \mu] + \Lambda^-[t, X, \mu]$ .

*Proof of Theorem 2* is based on Skorokhod–Girsanov’s change of measure formula for jump processes (see, e.g., [14]). Suppose there are two solutions,  $(X_t^1, \mu_t^1)$  and  $(X_t^2, \mu_t^2)$ . Denote by  $\Omega_n$  the event that the trajectory  $X$  has precisely  $n$  jumps on  $[0, T]$ . Recall – see, e.g., [14], [17] – that on the interval of time  $[0, T]$  the density of one distribution with respect to the other – we denote them by  $\mathbb{P}^{\mu^i}$ ,  $i = 1, 2$  – on a typical trajectory  $\omega = (t_1^\pm, \dots, t_n^\pm)$  with overall  $n \geq 0$  jumps up  $(t_i^+)$  or down  $(t_j^-)$  reads,

$$\begin{aligned} \rho_T &:= \frac{d\mathbb{P}^{\mu^2}}{d\mathbb{P}^{\mu^1}}(\omega)|_{\Omega_n} \\ &= \prod_{i=1}^n \frac{\Lambda^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^2]}{\Lambda^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^1]} \exp\left(-\int_0^T (\bar{\Lambda}[t, X_t, \mu_t^2] - \bar{\Lambda}[t, X_t, \mu_t^1]) dt\right), \end{aligned}$$

where  $X = (X_s, 0 \leq s \leq T)$  and  $(t_i^\pm)$  are the moments of jumps of the trajectory  $X$ , up or down, respectively; we keep the same sign at  $\Lambda$ , too, i.e.,  $\Lambda^+[t^+, \dots]$  or, respectively,  $\Lambda^-[t^-, \dots]$ . The usual convention  $\prod_{i=1}^0 \dots = 1$  is assumed. Note that, of course, the number of jumps  $n$  is random – i.e., it is a function of the trajectory – but in any case it is almost surely finite due to the boundedness of the intensities. Note also that the expression  $\rho_T$  above is a probability density. We have,

$$\begin{aligned}
& \mathbb{E}^{\mu^1} \prod_{i=1}^n \frac{\Lambda^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^2]}{\Lambda^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^1]} \exp\left(-\int_0^T (\bar{\Lambda}[t, X_t, \mu_t^2] - \bar{\Lambda}[t, X_t, \mu_t^1]) dt\right) \\
&= \sum_{n=0}^{\infty} \mathbb{E}^{\mu^1} 1(\Omega_n) \prod_{i=1}^n \frac{\Lambda^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^2]}{\Lambda^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^1]} \exp\left(-\int_0^T (\bar{\Lambda}[t, X_t, \mu_t^2] - \bar{\Lambda}[t, X_t, \mu_t^1]) dt\right) \\
&= \sum_{n=0}^{\infty} \mathbb{E}^{\mu^2} \int \dots \int_{0 < t_1 < \dots < t_n < T} \prod_{i=1}^n \Lambda^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^2] \exp\left(-\int_0^T \bar{\Lambda}[t, X_t, \mu_t^2] dt\right) \prod_{i=1}^n dt_i \\
&= \sum_{n=0}^{\infty} \mathbb{P}^{\mu^2}(\Omega_n) = 1.
\end{aligned}$$

Note that given the initial state  $X_0$ , the value without expectation  $\mathbb{E}^{\mu^2}$  here equals, actually,

$$\sum_{n=0}^{\infty} \int \dots \int_{0 < t_1 < \dots < t_n < T} \prod_{i=1}^n \Lambda^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^2] \exp\left(-\int_0^T \bar{\Lambda}^\pm[t, X_t, \mu_t^2] dt\right) \prod_{i=1}^n dt_i,$$

which itself equals identically one, while expectation  $\mathbb{E}^{\mu^2}$  relates to integration of each term over  $X_0$  if it is random. It is worthwhile to recall that the rule of the evolution of the trajectory  $X$  between the moments of jumps  $t_i$  is deterministic and linear with rate one for the continuous components, and the discrete component does not change between any two consequent jumps.

Now, we want to estimate the distance *in total variation* between two probability measures in the space of trajectories,  $\mu_{[0,T]}^1$  and  $\mu_{[0,T]}^2$  and then to use the inequality that the distance between the marginals of any two measures does not exceed the distance of the measures themselves,

$$\varphi_T := \|\mu_T^1 - \mu_T^2\|_{TV} \leq \|\mu_{[0,T]}^1 - \mu_{[0,T]}^2\|_{TV} = 2 - 2\mathbb{E}^{\mu^1}(\rho_T \wedge 1) =: \psi_T.$$

Now, the idea is to estimate the right hand side in the last term via  $\varphi$  and, hence, to show that, at least, for small values of  $T > 0$  this value equals zero. If this is

realized, then the claim that  $\varphi_t = 0$  for  $t \leq T$ ,  $t \leq 2T$ , etc., and, eventually, for all  $t \geq 0$  would follow by induction. In fact, we will be able to estimate the right hand side via another expression with  $\psi_T$  itself. Note, by the way, that although normally marginal distributions of any process may not determine the distribution in the space of trajectories, in our case with intensities it is, of course, the case which follows from [9], as mentioned already in the proof of the Theorem 1.

The first goal is to find a suitable lower bound for the value  $\mathbb{E}^{\mu^1}(\rho_T \wedge 1)$  from below. Let us split it as follows:

$$\mathbb{E}^{\mu^1}(\rho_T \wedge 1) = \sum_{n=0}^{\infty} \mathbb{E}^{\mu^1} \mathbf{1}(\Omega_n) (\rho_T \wedge 1).$$

Further, we have for  $n = 0$ ,

$$\begin{aligned} & \mathbb{E}^{\mu^1} \mathbf{1}(\Omega_0) (\rho_T \wedge 1) \\ &= \mathbb{E}^{\mu^1} \mathbf{1}(\Omega_0) \exp \left( - \int_0^T (\bar{\Lambda}[t, X_t, \mu_t^2] - \bar{\Lambda}[t, X_t, \mu_t^1]) dt \right) \wedge 1 \\ &\geq \exp \left( - \int_0^T \|\lambda\| \|\mu_t^2 - \mu_t^1\|_{TV} dt \right) \mathbb{E}^{\mu^1} \mathbf{1}(\Omega_0) \\ &\geq \exp(-\|\lambda\| T \psi_T) \mathbb{E}^{\mu^1} \mathbf{1}(\Omega_0). \end{aligned}$$

All norms like  $\|\lambda\|$  are sup-norms (except for the total variation norm, which is always shown explicitly). We used the fact that  $|\bar{\Lambda}[t, X, \mu]| \leq \|\lambda\|$ , and that

$$|\bar{\Lambda}[t, X_t, \mu_t^2] - \bar{\Lambda}[t, X_t, \mu_t^1]| \leq \|\lambda\| \|\mu_{[0,t]}^1 - \mu_{[0,t]}^2\|_{TV} \leq \|\lambda\| \|\mu_{[0,T]}^1 - \mu_{[0,T]}^2\|_{TV}, \quad 0 \leq t \leq T.$$

Similarly for  $n \geq 1$ , with a notation  $\tilde{\Lambda}^\pm[t^\pm, \dots] := \ln \Lambda^\pm[t^\pm, \dots]$ ,

$$\begin{aligned}
& \mathbb{E}^{\mu^1} 1(\Omega_n) (\rho_T \wedge 1) \\
&= \mathbb{E}^{\mu^1} 1(\Omega_n) \left\{ \prod_{i=1}^n \frac{\Lambda^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^2]}{\Lambda^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^1]} \times \right. \\
&\quad \left. \times \exp \left( - \int_0^T (\bar{\Lambda}[t, X_t, \mu_t^2] - \bar{\Lambda}[t, X_t, \mu_t^1]) dt \right) \right\} \wedge 1 \\
&\geq \mathbb{E}^{\mu^1} 1(\Omega_n) \exp \left( - \sum_{i=1}^n |\tilde{\Lambda}[t_i^\pm, X_{t_i}, \mu_{t_i}^2] - \tilde{\Lambda}[t_i^\pm, X_{t_i}, \mu_{t_i}^1]| \right) \\
&\quad \times \exp \left( - \int_0^T |\bar{\Lambda}[t, X_t, \mu_t^2] - \bar{\Lambda}[t, X_t, \mu_t^1]| dt \right).
\end{aligned}$$

Minimum with 1 here was dropped after all multipliers were estimated from below by the values less than one. Further, since the derivative of  $\ln x$  is bounded on any interval  $0 < a \leq x \leq b$ , say, by a constant  $K$ , we have with  $a = \inf \lambda(\dots) =: \underline{\lambda}$  and  $b = \|\lambda\|$ ,

$$\begin{aligned}
|\tilde{\Lambda}^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^2] - \tilde{\Lambda}[t_i^\pm, X_{t_i}, \mu_{t_i}^1]| &\leq K |\Lambda^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^2] - \Lambda^\pm[t_i^\pm, X_{t_i}, \mu_{t_i}^1]| \\
&\leq K \|\Lambda\| \|\mu_{t_i}^2 - \mu_{t_i}^1\|_{TV}.
\end{aligned}$$

Hence,

$$\begin{aligned}
& \mathbb{E}^{\mu^1} 1(\Omega_n) (\rho_T \wedge 1) \\
&\geq \mathbb{E}^{\mu^1} 1(\Omega_n) \exp \left( - \sum_{i=1}^n K \|\Lambda\| \|\mu_{t_i}^2 - \mu_{t_i}^1\|_{TV} \right) \\
&\quad \times \exp \left( - \int_0^T \|\lambda\| \|\mu_t^2 - \mu_t^1\|_{TV} dt \right).
\end{aligned}$$

(Here by definition  $t_0 = 0$ .) Here the infimum  $\underline{\lambda}$  is positive by the assumption. Thus, using the bound  $1 - \exp(-a) \leq a$  and estimates  $\mathbb{E}^{\mu^1} 1(\Omega_0) \leq \exp(-\underline{\lambda}T)$  and



$\mathbb{E}^{\mu^1} 1(\Omega_n) \leq \frac{(\|\lambda\|T)^n}{n!} \exp(-\underline{\lambda}T)$ ,  $n \geq 1$ , we get,

$$\begin{aligned}
\frac{1}{2}\psi_T &= 1 - \sum_n \mathbb{E}^{\lambda^1} 1(\Omega_n) (\rho_T \wedge 1) \\
&\leq (1 - \exp(-\|\lambda\| T \psi_T)) \mathbb{E}^{\lambda^1} 1(\Omega_0) \\
&+ \sum_{n=1}^{\infty} \mathbb{E}^{\lambda^1} 1(\Omega_n) \left( 1 - \exp \left( -K \sum_{i=1}^n \|\Lambda\| \|\mu_{t_i}^2 - \mu_{t_i}^1\|_{TV} \right) \right) \times \\
&\quad \times \exp \left( - \int_0^T \|\lambda\| \|\mu_t^2 - \mu_t^1\|_{TV} dt \right) \\
&\leq \|\lambda\| T \psi_T \mathbb{E}^{\mu^1} 1(\Omega_0) \\
&+ \sum_{n=1}^{\infty} \mathbb{E}^{\mu^1} 1(\Omega_n) (1 - \exp(-nK\|\Lambda\| \psi_T - T\|\lambda\| \psi_T)) \\
&\leq \psi_T \exp(-\underline{\lambda}T) \left( \|\lambda\|T + \sum_{n=1}^{\infty} (nK\|\Lambda\| + T\|\lambda\|) \frac{(\|\lambda\|T)^n}{n!} \right) \\
&= T\psi_T \exp(-\underline{\lambda}T) \left( \|\lambda\| + \sum_{n=0}^{\infty} ((n+1)K\|\Lambda\| + T\|\lambda\|) \frac{(\|\lambda\|T)^{n+1}}{(n+1)!} \right).
\end{aligned}$$

The series in the right hand side here converges and does not exceed some constant, say,  $C > 0$ , if  $T \leq 1$ . Hence, overall, we obtain,

$$0 \leq \frac{1}{2}\psi_T \leq CT\psi_T, \quad T \leq 1.$$

This implies that

$$\psi_T = 0, \quad T < (2C)^{-1} \wedge 1,$$

and, therefore, also

$$\varphi_T = 0, \quad T < (2C)^{-1} \wedge 1,$$

as required. In other words, we have shown that the two marginal measures  $\mu_t^1$  and  $\mu_t^2$  coincide for all  $t < (2C)^{-1} \wedge 1$ .

Further, note the constant  $C$  in this calculus does not depend on the initial distribution of the process. Hence, using the Markov property of the process and repeating the same arguments on  $[T, 2T]$ ,  $[2T, 3T]$ , etc., by induction we conclude that

$$\psi_t = 0, \quad t \geq 0,$$

and, therefore, also

$$\varphi_t = 0, \quad t \geq 0,$$

as required. So, the two measures  $\mu^1$  and  $\mu^2$  on the space of trajectories are equal. The Theorem 2 is proved.

## Appendix 1

The following celebrated Lemma is stated for the convenience of the reader.

**Lemma 1 (Skorokhod [17, Ch.1, §6])** *Let  $\xi_t^n$  ( $t \geq 0, n = 0, 1, \dots$ ) be some  $d$ -dimensional stochastic processes defined on some probability space and let for any  $T > 0, \epsilon > 0$  the following hold true:*

$$\lim_{c \rightarrow \infty} \sup_n \sup_{t \leq T} \mathbb{P}(|\xi_t^n| > c) = 0, \quad (9)$$

$$\lim_{h \downarrow 0} \sup_n \sup_{t, s \leq T; |t-s| \leq h} \mathbb{P}(|\xi_t^n - \xi_s^n| > \epsilon) = 0. \quad (10)$$

*Then there exists a subsequence  $n' \rightarrow \infty$  and a new probability can be constructed with processes  $\tilde{\xi}_t^{n'}, t \geq 0$  and  $\tilde{\xi}_t, t \geq 0$ , such that all finite-dimensional distributions of  $\tilde{\xi}^{n'}$  coincide with those of  $\xi^{n'}$  and such that for any  $\epsilon > 0$  and all  $t \geq 0$ ,*

$$\mathbb{P}(|\tilde{\xi}_t^{n'} - \tilde{\xi}_t| > \epsilon) \rightarrow 0, \quad n' \rightarrow \infty.$$

## Appendix 2

Let us show how the intuitive calculus leading to (5) may be performed more rigorously. We have,

$$\begin{aligned} & \mathbb{P} \left( \tilde{X}_{t+\delta}^n = (k+1, x', y+\delta), 0 \leq x' \leq \delta \mid \tilde{\mathcal{F}}_t^n; \tilde{X}_t^n, \tilde{X}_{(t-1/n)_+}^n \right) \Big|_{\tilde{X}_t^n = (k, x, y)} \\ &= \mathbb{P} \left( \tilde{X}_{t+\delta}^n = (k+1, x', y+\delta), 0 \leq x' \leq \delta \right); \end{aligned}$$

precisely one jump up on  $[t, t+\delta] \mid \tilde{\mathcal{F}}_t^n; \tilde{X}_t^n, \tilde{X}_{(t-1/n)_+}^n$ )  $+ o(\delta)$

$$\begin{aligned} &= \int_t^{t+\delta} \mathbb{E}' \Lambda^+(s, \tilde{X}_{(s-1/n)_+}^n, \tilde{\xi}_{(s-1/n)_+}^n) \\ &\quad \times \exp \left( - \int_t^{t+\delta} \mathbb{E}' \bar{\Lambda}(r, \tilde{X}_{(r-1/n)_+}^n, \tilde{\xi}_{(r-1/n)_+}^n) dr \right) ds + o(\delta) \\ &= \int_t^{t+\delta} \mathbb{E}' \Lambda^+(s, \tilde{X}_{(s-1/n)_+}^n, \tilde{\xi}_{(s-1/n)_+}^n) ds + o(\delta); \end{aligned}$$

if  $k(\tilde{X}_{(t-1/n)_+}^n) > 0$ ,

$$\begin{aligned} & \mathbb{P}\left(\tilde{X}_{t+\delta}^n = (k-1, x+\delta, y'), 0 \leq y' \leq \delta \mid \tilde{\mathcal{F}}_t^n; \tilde{X}_t^n, \tilde{X}_{(t-1/n)_+}^n\right) \Big|_{\tilde{X}_t^n=(k,x,y)} \\ &= \mathbb{P}\left(\tilde{X}_{t+\delta}^n = (k-1, x+\delta, y'), 0 \leq y' \leq \delta; \right. \end{aligned}$$

precisely one jump down on  $[t, t+\delta] \mid \tilde{\mathcal{F}}_t^n; \tilde{X}_t^n, \tilde{X}_{(t-1/n)_+}^n\Big) + o(\delta)$

$$\begin{aligned} &= \int_t^{t+\delta} \mathbb{E}' \Lambda^-(s, \tilde{X}_{(s-1/n)_+}^n, \tilde{\xi}_{(s-1/n)_+}^n) \\ &\quad \times \exp\left(-\int_t^{t+\delta} \mathbb{E}' \bar{\Lambda}(r, \tilde{X}_{(r-1/n)_+}^n, \tilde{\xi}_{(r-1/n)_+}^n) dr\right) ds + o(\delta) \\ &= \int_t^{t+\delta} \mathbb{E}' \Lambda^-(s, \tilde{X}_{(s-1/n)_+}^n, \tilde{\xi}_{(s-1/n)_+}^n) ds + o(\delta); \end{aligned}$$

and finally,

$$\begin{aligned} & \mathbb{P}\left(\tilde{X}_{t+\delta}^n = (k, x+\delta, y+\delta) \mid \tilde{\mathcal{F}}_t^n; \tilde{X}_t^n, \tilde{X}_{(t-1/n)_+}^n\right) \Big|_{\tilde{X}_t^n=(k,x,y)} \\ &= \mathbb{P}\left(\tilde{X}_{t+\delta}^n = (k, x+\delta, y+\delta); \text{no jumps on } [t, t+\delta] \mid \tilde{\mathcal{F}}_t^n; \tilde{X}_t^n, \tilde{X}_{(t-1/n)_+}^n\right) + o(\delta) \\ &= \exp\left(-\int_t^{t+\delta} \mathbb{E}' \bar{\Lambda}(r, \tilde{X}_{(r-1/n)_+}^n, \tilde{\xi}_{(r-1/n)_+}^n) dr\right) + o(\delta) \\ &= 1 - \int_t^{t+\delta} \mathbb{E}' \bar{\Lambda}(r, \tilde{X}_{(r-1/n)_+}^n, \tilde{\xi}_{(r-1/n)_+}^n) dr + o(\delta). \end{aligned}$$

Note that unlike in the intuitive calculus earlier, we did not use any regularity conditions on the intensities here and that up to  $o(\delta)$  the formulae above are all exact, not approximate as in the earlier intuitive version. Hence, we can write for any bounded

function  $g$  in the domain of the operator  $L$  (in particular, continuous in the second and third components of the state variable),

$$\begin{aligned} \mathbb{E} \left( g(\tilde{X}_{t+\delta}^n) | \tilde{\mathcal{F}}_t^n \right) &= g(\tilde{X}_t^{n,+}) \int_t^{t+\delta} \mathbb{E}' \Lambda^+(s, \tilde{X}_{(s-1/n)_+}^n, \tilde{\xi}_{(s-1/n)_+}^n) ds \\ &\quad + g(\tilde{X}_{(t-1/n)_+}^{n,-}) \int_t^{t+\delta} \mathbb{E}' \Lambda^-(s, \tilde{X}_{(s-1/n)_+}^n, \tilde{\xi}_{(s-1/n)_+}^n) ds \\ &\quad + g(\tilde{X}_t^n + \delta) \left( 1 - \int_t^{t+\delta} \mathbb{E}' \bar{\Lambda}(r, \tilde{X}_{(r-1/n)_+}^n, \tilde{\xi}_{(r-1/n)_+}^n) dr \right) + o(\delta), \end{aligned}$$

as  $\delta \downarrow 0$ . Therefore, it follows rigorously that

$$\begin{aligned} &\mathbb{E} \left( g(\tilde{X}_{t+\delta}^n) - g(\tilde{X}_t^n) | \tilde{\mathcal{F}}_t^n \right) \\ &= \int_t^{t+\delta} \mathbb{E}' L(s, \tilde{X}_{(s-1/n)_+}^n, \tilde{\xi}_{(s-1/n)_+}^n) g(\tilde{X}_s^n) ds + o(\delta). \end{aligned}$$

So, we can rigorously apply the principle of “complete probabilities” (or, better “complete expectation”) taking a summation as earlier in the intuitive version: let us split the interval  $[t_m, t_{m+1}]$  into  $N$  equal small sub-intervals,  $t_m = s_0 < \dots < s_N = t_{m+1}$  so that  $(s_{i+1} - s_i) =: \delta$ ; then we get similarly to (5),

$$\begin{aligned} \mathbb{E} \left( g(\tilde{X}_{t_{m+1}}^n) - g(\tilde{X}_{t_m}^n) | \tilde{\mathcal{F}}_{t_m}^n \right) &= \sum_{i=0}^{N-1} \mathbb{E} \left( \mathbb{E} \left( g(\tilde{X}_{s_{i+1}}^n) - g(\tilde{X}_{s_i}^n) | \tilde{\mathcal{F}}_{s_i}^n \right) | \tilde{\mathcal{F}}_{t_m}^n \right) \\ &= \mathbb{E} \left( \int_{t_m}^{t_{m+1}} \mathbb{E}' L(s, \tilde{X}_{(s-1/n)_+}^n, \tilde{\xi}_{(s-1/n)_+}^n) g(\tilde{X}_s^n) ds | \tilde{\mathcal{F}}_{t_m}^n \right) + o(1), \end{aligned}$$

which rigorously implies the equation (4).

## Acknowledgements

The techniques used in this paper were stimulated by the methods developed in a long-term joint work on formally quite different McKean-Vlasov SDE equations with

Yu. Mishura, as well as in fruitful discussions of the author on the same subject with D. Šiska, and L. Szpruch. S. Pirogov, A. Rybko, and G. Zverkina helped to find some (quite a few) technicalities to be corrected in the earlier versions of the text. The author is sincerely thankful to all these colleagues and to two referees for very useful advice. The deepest gratitude is to Professor Alexander Dmitrievich Solovyev (06.09.1927 – 06.04.2001) who was the author’s supervisor at BSc and MSc programmes at Moscow State University.

## References

- [1] Aghajani, R., Li, X., Ramanan, K.: Mean-field Dynamics of Load-Balancing Networks with General Service Distributions, <https://arxiv.org/abs/1512.05056>
- [2] Baccelli, F., Karpelevich, F.I., Kelbert, M.Ya. et al.: A Mean-Field Limit for a Class of Queueing Networks, *Journal of Statistical Physics*, 66(3/4), (1992)
- [3] Baccelli, F., Rybko, A.N., Shlosman, S.B.: Queueing networks with mobile servers: the mean-field approach, *Problems Inform. Transmission*, 52(2), 178-199 (2016)
- [4] Billingsley, P.: *Convergence of Probability Measures*. New York, NY: John Wiley & Sons (1999)
- [5] Borovkov, K.A.: Propagation of chaos for queueing networks, *Theory of Probability and its Applications*, 42:3, 385-394 (1998)
- [6] Dawson, D.A., Tang, J., Zhao, Y.Q.: Balancing Queues by Mean Field Interaction, *Queueing Systems* 49, 335-361 (2005)
- [7] Gnedenko, B.V., Kovalenko, I.N.: *Introduction to Queueing Theory*, 2nd ed., Birkhäuser, Boston et al. (1989)
- [8] Gnedenko, B.V., Belyaev, Yu.K., Solovyev, A.D.: *Mathematical Methods in Reliability Theory*, Academic Press, New York (1969)
- [9] Davis, M.H.A.: Piecewise–Deterministic Markov Processes: A General Class of Non–Diffusion Stochastic Models. *J. Royal Stat. Soc. Ser. B (Methodological)*. 46(3), 353–388 (1984)
- [10] Ethier, S.N., Kurtz, T.G.: *Markov processes: Characterisation and Convergence*. Wiley, Hoboken, N.J. (2005)

- [11] Karpelevich, F.I., Rybko, A.N.: Asymptotic Behavior of the Thermodynamical Limit for Symmetric Closed Queueing Networks, *Problems Inform. Transmission*, 36(2), 154-179 (2000)
- [12] Krylov, N.V.: On the selection of a Markov process from a system of processes and the construction of quasi-diffusion processes, *Mathematics of the USSR - Izvestija*, 7(3), 691-709 (1973)
- [13] Krylov, N.V.: Introduction to the theory of random processes, AMS, Providence (2002)
- [14] Liptser, R.Sh., Shiryaev, A.N.: Theory of martingales, Kluwer Academic Publishers Group, Dordrecht (1989)
- [15] Liptser, R.Sh., Shiryaev, A.N.: Stochastic Calculus on Filtered Probability Spaces, In: *Probability Theory III. Encyclopaedia of Mathematical Sciences*, vol 45, Springer, Berlin, Heidelberg, 111-157 (1998)
- [16] Mishura, Yu.S., Veretennikov, A.Yu.: Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations, <https://arxiv.org/abs/1603.02212>
- [17] Skorokhod, A.V.: *Studies in the theory of random processes*. Addison-Wesley, Reading, Mass. (1965)
- [18] Veretennikov, A.Yu., Zverkina, G.A.: Simple Proof of Dynkin’s Formula for Single-Server Systems and Polynomial Convergence Rates, *Markov Processes Relat. Fields*. 20, 479-504 (2014)