On mean-field (GI/GI/1) queueing model: existence and uniqueness

A. Yu. Veretennikov

Received: 22 February 2018 / Revised: 29 April 2019 / Published online: 4 September 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

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 slightly different set of assumptions on intensities in comparison with those required for existence.

Keywords
GI/GI/1 · Mean-field · Existence · Weak uniqueness · Skorokhod lemma

Mathematics Subject Classification 60-02 · 60K25 · 90B22

1 Introduction
The mean-field approach in the theory of queueing systems allows us to take into consideration large interacting ensembles of queues, or queueing networks, by using the idea of replacing these interactions by a suitable “mean field”. In particular, this approach proved rather fruitful in systems with countable state spaces. In this work, we propose a method of constructing a more general extension of the system GI/GI/1—or, more precisely, GI/GI/1/∞—under certain restrictions on intensities of arrivals and service, intensities of which may both depend on the state and on the marginal distribution of the process. Existence and weak uniqueness are discussed on the basis of

In memory and to the 90th anniversary of Alexander Dmitrievich Solovyev.

This study has been funded by the Russian Academic Excellence Project ‘5-100’ and by the RFBR Grant 17-01-00633_a.

A. Yu. Veretennikov
a.veretennikov@leeds.ac.uk

1 School of Mathematics, University of Leeds, Leeds LS2 9JT, UK
2 National Research University Higher School of Economics, Moscow, Russian Federation
3 Institute for Information Transmission Problems, Moscow, Russian Federation
tightness of measures, Skorokhod’s unique probability space lemma, the total variation metric and a Skorokhod–Girsanov density of measures theorem for jump processes. The basis for the study in Sects. 3.1 and 3.2 is a technique similar to the one developed in a recent preprint on stochastic McKean–Vlasov equations [14]. The main motivation is discussed in the next section. It is likely that the established results may also be useful in the area of mathematical theory of reliability; see [5]. Some earlier results on mean-field queueing models can be found, in particular, in [1,3,7,10], and in further references therein. However, both the model and especially the methods used in the present paper are different. In this paper, multiparticle systems and their convergence called “propagation of chaos” are not pursued, but it could be useful for motivation. Hence, a small section with one possible setting related to multiparticle approximations has been added.

The paper consists of an Introduction, the Motivation, the Main section, and an Appendix. In the Motivation, it is shown how a mean-field model arises as a limit in a network of queueing models with interaction; the statement in this section is not a rigorous result, hence, it is stated as a hypothesis. Yet, this hypothesis is not used in the calculus of the following section. 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 contains the statement of Skorokhod’s Lemma about an equivalence of weak convergence of a sequence of processes to convergence in probability of processes with the same distributions on a unique probability space.

2 Motivation: propagation of chaos hypothesis

Assume that, for each \( N > 1 \), there is a network consisting of \( N \) identical models of type GI/GI/1 weakly interacting with each other, so that the intensities of arrivals and service in model \( k \) (\( 1 \leq k \leq N \)) are given by the formulae

\[
\lambda_{t}^{k,N} := \frac{1}{N} \sum_{1 \leq j \leq N} \lambda(t, X_{t}^{k,N}, X_{t}^{j,N}),
\]

\[
h_{t}^{k,N} := \frac{1}{N} \sum_{1 \leq j \leq N} h(t, X_{t}^{k,N}, X_{t}^{j,N}),
\]

respectively, where \( X_{t}^{k,N} \) stands for the process representing model \( k \). Note that the notation \( X_{t}^{k,N} \)—or just \( X \) if the values \( k, N, t \) are fixed—is as follows: informally, \( X \) is a triple \( X = (n, x, y) \) consisting of the number of customers \( n \) in the system or server (with number \( k \)), of the elapsed time since the last arrival to this system, and of \( y \), the elapsed time since the beginning of the current service; for the mean-field model, this will be repeated in the beginning of the Main section. The initial distribution of \( X_{0}^{k,N} \) is assumed to be the same for all \( k \), and, for each \( N \), the random variables \( X_{0}^{k,N}, 1 \leq k \leq N \), are independent. Note that each term \( \lambda(t, X_{t}^{k,N}, X_{t}^{j,N}) \) or \( h(t, X_{t}^{k,N}, X_{t}^{j,N}) \) may include an additive summand which does not depend on other states, that is, \( \lambda(t, X_{t}^{k,N}) \) and \( h(t, X_{t}^{k,N}) \), respectively.
Such a dependence of the intensities of the elapsed times for other particles could take into account various situations. If \( N \) is large and many of the values \( X^j_{t,N} \) are also large, it may have a repelling effect on the new arrivals. This can be modelled by functions \( \lambda(t, x, x') \) decreasing in the last and possibly second variables. At the same time, if there are lots of customers with a long service, the system may switch on some additional resources so as to speed up the remaining service; this could be modelled by the function \( h(t, x, x') \) increasing in the last and possibly second variables.

The following generic statement, formulated here as a hypothesis, stands behind the mean-field model under consideration in this paper. Since there is no immediate goal to prove it, we do not try to specify any conditions here. This task is postponed for further studies.

**Hypothesis 1** (propagation of chaos) Under “appropriate assumptions” on the functions \( \lambda \) and \( h \), for each \( N \), let all initial distributions \( \mu_0 \) for all “particles” \( X^k_0 \) be i.i.d. \((1 \leq k \leq N)\). Then, the processes \((X^k_{t,N}, t \geq 0)\) all converge weakly as \( N \to \infty \) to a process \((\bar{X}_t, t \geq 0)\) with the “limiting” intensities of jumping “up” and, respectively, “down” given by

\[
\Lambda[t, X_t, \mu_t] = \int \lambda(t, X_t, Y) \mu_t(dY), \tag{1}
\]

and

\[
H[t, X_t, \mu_t] = \int h(t, X_t, Y) \mu_t(dY), \tag{2}
\]

where \( \mu_t \) is the marginal distribution of \( X_t \) for each \( t \geq 0 \). Also, all components \((X^1_{t,N}, X^2_{t,N}, \ldots, X^m_{t,N})\) for any finite fixed \( m \) are asymptotically independent as \( N \to \infty \).

The intuitive reasoning for this statement is an asymptotic independence of the “particles” as \( N \to \infty \); if so, the claim about such a convergence becomes a law of large numbers type result. At a physical level, this idea of a replacement of forces due to a large number of particles by some “mean field” was introduced by Vlasov in 1938 [18], then explored in probability by Kac [9] and McKean [13], followed by lots of further investigations in various areas of probability and applications. Since then such limiting models are often called McKean–Vlasov (or, Vlasov–McKean), in particular, in stochastic analysis. Note that the idea of such limiting approximations as the number \( N \) becomes large is especially in demand in the epoch of the “big data” challenge.

As was explained, we neither pursue the goal to prove this statement in this paper nor use this hypothesis in the sequel. It was added only for motivation: before trying to establish such a convergence, it seems rather natural to study the limiting process, which is precisely the main goal of the paper. For similar results in some other multiparticle models with a weak interaction in queueing, see the cited papers [1,3,7,10] and further references therein. One of the important studies in the area of interacting queueing networks in the last years is [15], where some other recent references may be found.
Let us repeat that the goal of this paper is to prepare the basis for further studies, i.e., to show existence and uniqueness of the limiting process with intensities given by Eqs. (1) and (2). Clearly, if there is any hope of establishing convergence towards a limiting process in any reasonable probability sense, then this limiting process must exist and has to be unique in distribution. In the sequel, it is convenient to re-denote $\Lambda$ by $\Lambda^+$ and $H$ by $\Lambda^-$. Generally, any dependence of intensities of the marginal distribution for the model is regarded as an “environment”.

### 3 Main section

The state space of the process under consideration is the union

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

The meaning of $n$ here is the number of “customers” in the system; the value $x$ stands for time elapsed since the last arrival, while $y$ denotes time elapsed since the beginning of the current service. There is only one server which works without breaks (provided 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 infinite capacity, and, for simplicity, we assume the FIFO discipline of service. It is assumed that at any time $t$ and at $X = (n, x, y)$ (or $X = (0, x)$ for $n = 0$) 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 the notation $(0, x, y)$, where $y$ is a “false” variable, i.e., we identify all such triples with any $y$ with a couple $(0, x)$. The process is piecewise linear Markov (PLMP, see [4]), which simply means that the continuous components—$(x, y)$ if $n > 0$, or just $x$ if $n = 0$—grow linearly with rate 1 between any jumps, while the discrete component $n$ remains unchanged.

**Assumptions:**

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

(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).
Note that both intensities $\Lambda^\pm$ may include additional (non-negative) terms not depending on the measure, say, $\lambda^\pm_0(t, X)$; in particular, this may be helpful for justification of assumption (A4), as the terms $\lambda^\pm_0(t, X)$ can be easily assumed bounded away from zero uniformly and independently of $N$.

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

$$X^+ := (n + 1, 0, y), \quad \text{for } X = (n, x, y),$$

$$X^- := (n - 1, x, 0), \quad \text{for } X = (n, x, y), \quad n \geq 1.$$ 

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

### 3.1 Existence

The initial value $X_0$ of the process may have a distribution $\mu_0$. (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 $\Lambda[t, X_t, \mu_t], H[t, X_t, \mu_t]$: namely, for any bounded continuous function $g(X)$ with bounded continuous derivatives in $(x, y)$, the expression

$$M_t := g(X_t) - g(X_0) - \int_0^t L(s, X_s, \mu_s)g(X_s) \, ds$$

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

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

Moreover, for any given measure-valued function $(\mu_s, s \geq 0)$ in $L(s, X_s, \mu_s)$, the martingale problem (see [8]) (3) has a weakly unique solution.

Note that the generator of the Markov process is, of course, $L(t, X, \mu)$; different variables $X$ and $X'$ are needed only for the convenience of the proof. Equivalently, Dynkin’s identity holds true for any function $g(X)$ from the same class:

$$\mathbb{E}_{0, X_0} g(X_t) = g(X_0) + \mathbb{E}_{0, X_0} \int_0^t L(s, X_s, \mu_s)g(X_s) \, ds.$$
Moreover, equivalently, for any $0 \leq t_1 < t_2 \ldots < t_{m+1}$, and for any Borel bounded functions $\phi_k(X)$, $X \in \mathcal{X}$,

$$
\mathbb{E}_{0, X_0} \left( g(X_{t_{m+1}}) - g(X_{t_{m}}) - \int_{t_{m}}^{t_{m+1}} L(s, X_s, \mu_s) g(X_s) \, ds \right) \prod_{k=1}^{m} \phi_k(X_{t_k}) = 0. \quad (5)
$$

This latter formula may be called another version of Dynkin’s identity and will be the basis for establishing existence. With a bit of abuse of the standard terminology, (5) 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.

**Proof of Theorem 1** For any $n \geq 1$, consider a process $(X^n_t)$, with initial data $X^n_{0, \delta} = X_0$ and intensities of jumps up and down, respectively, given by

$$
\Lambda^+[t, X^n_{t-1/n}, \mu^n_{t-1/n}], \quad \Lambda^-[t, X^n_{t-1/n}, \mu^n_{t-1/n}],
$$

where $X^n_t$ with $t < 0$ is understood as $X^n_0$, and similarly for $\mu^n_t$. The processes $(X^n_t)$ for each $n$ are constructed by induction successfully on the intervals $[0, 1/n], [1/n, 2/n], \ldots$. Due to the boundedness assumption on both intensities, the processes for any $n$ are defined for any $t \geq 0$ as càdlàg pure jump processes. Moreover, for any $t$, the probability of having a jump exactly at time $t$ for any $X^n$ equals zero. Now that the processes $(X^n_t, t \geq 0)$ for $n \geq 1$ have been constructed, let us introduce independent equivalent processes $(\xi^n_t, t \geq 0)$ on some probability space; let $\mathbb{E}'$ stand in all cases for the integration with respect to the third variable, for example,

$$
\mathbb{E}' \lambda^\pm(t, X^n_t, \xi^n_t) := \int \lambda^\pm(t, X^n_t, Y) \mu^n_t(dY).
$$

It can be checked that the assumptions of Lemma 1 from the Appendix are satisfied. Hence, on some new probability space, there are equivalent—and, hence, *Markov with the same generators*—processes $(\tilde{X}^n_t, \tilde{\xi}^n_t)$, and a limiting pair $(\tilde{X}_t, \tilde{\xi}_t)$ such that, for some subsequence $\{n'\} \subset \{n\}$, we have $(\tilde{X}^n_t, \tilde{\xi}^n_t) \xrightarrow{p} (\tilde{X}_t, \tilde{\xi}_t)$, $n' \to \infty$, 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 pure jump process with a finite number of jumps on any bounded interval. Moreover, the property $\lim_{k \to 0} \sup_{s \leq T, |t-s| \leq \epsilon} \mathbb{P}(|\tilde{X}^n_{t^\delta} - \tilde{X}^n_{s}^T| > \epsilon) = 0$ implies that for any $\epsilon > 0$ there is the following convergence in probability:

$$
\tilde{X}^{n'}_{t-1/n'} \xrightarrow{p} \tilde{X}_t, \quad n' \to \infty.
$$
The analogue of Dynkin’s formula (5) for the pair \((\tilde{X}^n_t, \tilde{\xi}^n_t)\) reads
\[
\mathbb{E}_{0, X_0} \left[ \left( g(\tilde{X}^n_{t_m+1}) - g(\tilde{X}^n_{t_m}) - \int_{t_m}^{t_{m+1}} \mathbb{E}' L(s, \tilde{X}_s, \tilde{\xi}_s) g(\tilde{X}_s) \, ds \right) \prod_{k=1}^{m} \phi_k(\tilde{X}_{t_k}) \right] = 0.
\] (6)

This formula (6) follows straightforwardly from the “complete expectation” arguments (cf., for example, [17]).

By continuity of the functions \(\lambda, h\), 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 (6) in the limit with continuous bounded functions \((\phi_k)\),
\[
\mathbb{E}_{0, X_0} \left( g(\tilde{X}_{t_{m+1}}) - g(\tilde{X}_{t_m}) - \int_{t_m}^{t_{m+1}} \mathbb{E}' L(s, \tilde{X}_s, \tilde{\xi}_s) g(\tilde{X}_s) \, ds \right) \prod_{k=1}^{m} \phi_k(\tilde{X}_{t_k}) = 0. \] (7)

Since the distribution of the random variable \(\tilde{\xi}_t\) is the same as that of \(\tilde{X}_t\)—let us denote it by \(\tilde{\mu}_t\)—Eq. (7) can be equivalently written as
\[
\mathbb{E}_{0, X_0} \left( g(\tilde{X}_{t_{m+1}}) - g(\tilde{X}_{t_m}) - \int_{t_m}^{t_{m+1}} Lg(s, \tilde{X}_s, \tilde{\mu}_s) \, ds \right) \prod_{k=1}^{m} \phi_k(\tilde{X}_{t_k}) = 0. \] (8)

Due to the properties of measures on \(\mathbb{R}^d\), formula (8) holds true for any Borel bounded functions \((\phi_k)\), too. Due to [6], the solution of the “martingale problem” (8)—or, more precisely, of the martingale problem
\[
M_t := g(\tilde{X}_{t_{m+1}}) - g(\tilde{X}_{t_m}) - \int_{t_m}^{t_{m+1}} Lg(s, \tilde{X}_s, \tilde{\mu}_s) \, ds \quad \text{is a martingale},
\] (9)

with a given family of marginal measures \((\tilde{\mu}_s, s \geq 0)\) is unique. Hence, according to [11], or [8, Theorem 4.4.2], the limiting process \(\tilde{X}\) is Markov. The form of its generator with the required intensities \(\Lambda^\pm\) follows from (8). This finishes the proof of Theorem 1.

### 3.2 Weak uniqueness

We highlight that boundedness of all intensities and the condition that they are (uniformly) bounded away from zero will be used in the sequel. While it is clear that boundedness from above may be relaxed for the purpose of establishing existence—for example, under Lyapunov-type conditions, or under linear growth, or otherwise—and that boundedness away from zero is not required for existence at all, yet for uniqueness, boundedness both from above and from below seem essential (although also
could be, apparently, slightly relaxed). On the other hand, continuity of the intensities in this section is not necessary and it is 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 the required intensities $\Lambda[t, x, \mu_t]$ and $H[t, x, \mu_t]$.

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

**Proof of Theorem 2** This is based on the Skorokhod–Girsanov change of measure formula for jump processes (see, for example, [12]). Suppose there are two solutions, $(X^1_t, \mu^1_t)$ and $(X^2_t, \mu^2_t)$. Denote by $\Omega_n$ the event that the trajectory $X$ has precisely $n$ jumps on $[0, T]$. Recall—see, for example, [12], [16]—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 $\omega = (t^+_1, \ldots, t^+_n)$ with overall $n \geq 0$ jumps up $(t^+_i)$ or down $(t^-_j)$ reads

$$
\rho_T := \frac{d\mathbb{P}^{\mu^2}}{d\mathbb{P}^{\mu^1}}(\omega)_{|\Omega_n} = \prod_{i=1}^{n} \left( \frac{\Lambda^+[t^+_i, X_{t_i}, \mu^2_{t_i}]}{\Lambda^+[t^+_i, X_{t_i}, \mu^1_{t_i}]} \exp\left(-\int_0^T (\tilde{\Lambda}[t, X, \mu^2] - \tilde{\Lambda}[t, X, \mu^1]) dt\right) \right),
$$

where $X = (X_s, 0 \leq s \leq T)$ and $(t^+_i)$ are the moments of jumps of the trajectory $X$, up or down, respectively; we keep the same sign at $\Lambda$, too, i.e., $\Lambda^+[t^+, \ldots]$ or, respectively, $\Lambda^-[t^-, \ldots]$. The usual convention $\prod_{i=1}^{n} = 1$ 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 intensities. Note also that the above expression $\rho_T$ is a probability density. We have

$$
\mathbb{E}^{\mu^1} \prod_{i=1}^{n} \frac{\Lambda^+[t^+_i, X_{t_i}, \mu^2_{t_i}]}{\Lambda^+[t^+_i, X_{t_i}, \mu^1_{t_i}]} \exp\left(-\int_0^T (\tilde{\Lambda}[t, X, \mu^2] - \tilde{\Lambda}[t, X, \mu^1]) dt\right)
$$

$$
= \sum_{n=0}^{\infty} \mathbb{E}^{\mu^1}(\Omega_n) \prod_{i=1}^{n} \frac{\Lambda^+[t^+_i, X_{t_i}, \mu^2_{t_i}]}{\Lambda^+[t^+_i, X_{t_i}, \mu^1_{t_i}]} \exp\left(-\int_0^T (\tilde{\Lambda}[t, X, \mu^2] - \tilde{\Lambda}[t, X, \mu^1]) dt\right)
$$

$$
= \sum_{n=0}^{\infty} \mathbb{E}^{\mu^2} \int_{0 < t_1 < \cdots < t_n < T} \prod_{i=1}^{n} \Lambda^+[t^+_i, X_{t_i}, \mu^2_{t_i}] \exp\left(-\int_0^T \tilde{\Lambda}[t, X, \mu^2] dt\right) \prod_{i=1}^{n} dt_i
$$

$$
= \sum_{n=0}^{\infty} \mathbb{E}^{\mu^2}(\Omega_n) = 1.
$$
Note that given the initial state $X_0$, the value without expectation $\mathbb{E} \mu^2$ here equals, actually, 

$$\sum_{n=0}^{\infty} \int \cdots \int \prod_{i=1}^{n} A^{\pm}[t_i^+, X_{t_i}, \mu_i^2] \exp \left( -\int_0^T \tilde{A}^{\pm}[t, X_t, \mu_t^2] \, dt \right) \prod_{i=1}^{n} dt_i,$$

which itself equals identically one, while the expectation $\mathbb{E} \mu^2$ relates to integration of each term over $X_0$ if it is random. It is worthwhile recalling that the evolution rule for 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, $\mu^1_{[0,T]}$ and $\mu^2_{[0,T]}$, and then to use the inequality that the distance between the marginals of any two measures does not exceed the distance between the measures themselves:

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

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 [6], as mentioned already in the proof of Theorem 1.

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

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

Further, we have, for $n = 0$,

$$\mathbb{E} \mu^1_{1}(\Omega_0) (\rho_T \land 1) = \mathbb{E} \mu^1_{1}(\Omega_0) \exp \left( -\int_0^T (\tilde{A}[t, X_t, \mu_t^2] - \tilde{A}[t, X_t, \mu_t^1]) \, dt \right) \land 1 \geq \exp \left( -\int_0^T \| \lambda \| \| \mu_t^2 - \mu_t^1 \|_{TV} \, dt \right) \mathbb{E} \mu^1_{1}(\Omega_0) \geq \exp(-\| \lambda \| T \psi_T) \mathbb{E} \mu^1_{1}(\Omega_0).$$

All norms like $\| \lambda \|$ are sup-norms (except for the total variation norm, which is always shown explicitly). We used the fact that $|\tilde{A}[t, X, \mu]| \leq \| \lambda \|$, and that
Similarly, for $n \geq 1$, with the notation $\tilde{\Lambda}^\pm [t^\pm, \ldots] := \ln \Lambda^\pm [t^\pm, \ldots]$, 

$$
\mathbb{E}^{\mu^1 1}(\Omega_n) (\rho_T \land 1) = \mathbb{E}^{\mu^1 1}(\Omega_n) \left( \prod_{i=1}^{n} \frac{\Lambda^\pm [t_i^\pm, X_{t_i}, \mu_{t_i}^2]}{\Lambda^\pm [t_i^\pm, X_{t_i}, \mu_{t_i}^1]} \times \exp \left( - \int_0^T (\tilde{\Lambda}[t, X_t, \mu_t^2] - \tilde{\Lambda}[t, X_t, \mu_t^1]) \, dt \right) \right) \land 1
$$

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

The minimum with 1 here was dropped after all multipliers were estimated from below by 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 $a = \inf \lambda(\ldots) =: \underline{\lambda}$ and $b = \|\lambda\|$, 

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

Hence,

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

(Let by definition, $t_0 = 0$.) Here, $\underline{\lambda} > 0$ stands for the infimum of all intensities, and this infimum is positive by 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

$$
\frac{1}{2} \psi_T = 1 - \sum_{n=1}^{\infty} \mathbb{E}^{\mu^1 1}(\Omega_n) (\rho_T \land 1) 
\leq (1 - \exp(-\|\lambda\|T \psi_T)) \mathbb{E}^{\mu^1 1}(\Omega_0) 
+ \sum_{n=1}^{\infty} \mathbb{E}^{\mu^1 1}(\Omega_n) \left( 1 - \exp \left( -K \sum_{i=1}^{n} \|\Lambda\| \|\mu_{t_i}^2 - \mu_{t_i}^1\|_{TV} \right) \right)
$$

\( \odot \) Springer
\[
\times \exp \left( - \int_0^T \|\lambda\| \|\mu_2^t - \mu_1^t\|_{TV} \, dt \right) \\
\leq \|\lambda\| T \psi_T \mathbb{E}^{\mu_1^1} 1(\Omega_0) \\
+ \sum_{n=1}^{\infty} \mathbb{E}^{\mu_1^1} 1(\Omega_n) \left( 1 - \exp \left( -n K \|\Lambda\| \psi_T - T \|\lambda\| \psi_T \right) \right) \\
\leq \psi_T \exp(-\lambda T) \left( \|\lambda\| T + \sum_{n=1}^{\infty} (n K \|\Lambda\| + T \|\lambda\|) \frac{(\|\lambda\| T)^n}{n!} \right) \\
= T \psi_T \exp(-\lambda T) \left( \|\lambda\| + \sum_{n=0}^{\infty} ((n+1) K \|\Lambda\| + T \|\lambda\|) \frac{(\|\lambda\|)^{n+1} T^n}{(n + 1)!} \right). 
\]

Here, the series on the right-hand side converges and does not exceed some constant, say, \(C > 0\), if \(T \leq 1\). Hence, overall, we obtain,

\[
0 \leq \frac{1}{2} \psi_T \leq C T \psi_T, \quad T \leq 1.
\]

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 \(\mu_1^t\) and \(\mu_2^t\) 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. Theorem 2 is proved.

**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, as well as to the anonymous referee. 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 University.

4 Appendix

The following celebrated Lemma is stated for the convenience of the reader [2, Theorem 25.6].

Lemma 1 (Skorokhod [16, Ch.1, §6]) Let \( \xi^n_t \) \((t \geq 0, n = 0, 1, \ldots)\) be some \( d \)-dimensional stochastic processes defined on some probability space, and let the following hold true for any \( T > 0, \epsilon > 0 \):

\[
\lim_{c \to \infty} \sup_n \sup_{t \leq T} \mathbb{P}(|\xi^n_t| > c) = 0,
\]

\[
\lim_{h \downarrow 0} \sup_n \sup_{t,s \leq T, |t-s| \leq h} \mathbb{P}(|\xi^n_t - \xi^n_s| > \epsilon) = 0.
\]

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

\[
\mathbb{P}(|\tilde{\xi}'_t - \tilde{\xi}_t| > \epsilon) \to 0, \quad n' \to \infty.
\]

References

1. Baccelli, F., Rybko, A.N., Shlosman, S.B.: Queueing networks with mobile servers: the mean-field approach. Probl. Inf. Transm. 52(2), 178–199 (2016)
2. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1999)
3. Borovkov, K.A.: Propagation of chaos for queueing networks. Theory Probab. Appl. 42(3), 385–394 (1998)
4. Gnedenko, B.V., Kovalenko, I.N.: Introduction to Queueing Theory, 2nd edn. Birkhäuser, Boston (1989)
5. Gnedenko, B.V., Belyaev, Yu.K., Solovyev, A.D.: Mathematical Methods in Reliability Theory. Academic Press, New York (1969)
6. Davis, M.H.A.: Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models. J. R. Stat. Soc. Ser. B (Methodological) 46(3), 353–388 (1984)
7. Dawson, D.A., Tang, J.Q., Zhao, Y.Q.: Balancing queues by mean field interaction. Queueing Syst. 49(3), 335–361 (2005)
8. Ethier, S.N., Kurtz, T.G.: Markov processes: characterisation and convergence. Wiley, Hoboken (2005)
9. Kac, M.: Foundations of kinetic theory. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. 3, pp. 171–197. University of California Press, California (1956)
10. Karpelevich, F.I., Rybko, A.N.: Asymptotic behavior of the thermodynamical limit for symmetric closed queueing networks. Probl. Inf. Transm. 36(2), 154–179 (2000)
11. Krylov, N.V.: On the selection of a Markov process from a system of processes and the construction of quasi-diffusion processes. Math. USSR—Izvestija 7(3), 691–709 (1973)
12. Liptser, R.Sh., Shiryaev, A.N.: Theory of Martingales. Kluwer Academic Publishers Group, Dordrecht (1989)
13. McKean, H.P.: A class of Markov processes associated with nonlinear parabolic equations. Proc. Natl. Acad. Sci. USA 56(6), 1907–1911 (1966)
14. Mishura, Yu.S., Veretennikov, A.Yu.: Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations, arXiv:1603.02212
15. Rybko, A., Shlosman, S., Vladimirov, A.: Poisson hypothesis for open networks at low load. Mosc. Math. J. 17(1), 145–160 (2017)
16. Skorokhod, A.V.: Studies in the Theory of Random Processes. Addison-Wesley, Reading (1965)
17. Veretennikov, A.Yu., Zverkina, G.A.: Simple proof of Dynkin’s formula for single-server systems and polynomial convergence rates. Markov Process. Relat. Fields 20, 479–504 (2014)
18. Vlasov, A.A.: On Vibration properties of electron gas. J. Exp. Theory Phys. 8(3), 291–318 (1938). (in Russian)

Publisher’s Note  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.