HEAVY TRAFFIC LIMIT WITH DISCONTINUOUS COEFFICIENTS VIA A NON-STANDARD SEMIMARTINGALE DECOMPOSITION

RAMI ATAR AND MASAKIYO MIYAZAWA

Abstract. This paper studies a single server queue in heavy traffic, with general inter-arrival and service time distributions, where arrival and service rates vary discontinuously as a function of the (diffusively scaled) queue length. It is proved that the weak limit is given by the unique-in-law solution to a stochastic differential equation in $[0, \infty)$ with discontinuous drift and diffusion coefficients. The main tool is a semimartingale decomposition for point processes introduced in [6], which is distinct from the Doob-Meyer decomposition of a counting process. Whereas the use of this tool is demonstrated here for a particular model, we believe it may be useful for investigating the scaling limits of queueing models very broadly.

1. Introduction

The goal of this paper is twofold: To continue a line of research on the multi-level $GI/G/1$ queue, at the diffusion scale [14,15,16], and to demonstrate the applicability of a semimartingale decomposition for general point processes, introduced in [6], to the study of queueing models under scaling limits.

We consider a version of the multi-level $GI/G/1$ queueing model, which will be referred to in what follows as the *multi-level queue*. In this model, the half line $\mathbb{R}_+ = [0, \infty)$ is partitioned into a finite number of disjoint intervals, called levels, and arrivals and services, driven by two independent renewal processes, undergo time change depending on the level to which the queue length belongs. In other words, the arrival and service rates are determined by a piecewise constant function of the queue length. Diffusion limits of models where arrival and service rates vary with the queue length have been studied before under continuous dependence; see [18] and references therein. However, there are very few limit results for models where the dependence is discontinuous, because such discontinuities present some technical challenges. Our main result is the weak convergence of the queue length, normalized at the diffusion scale, to a reflected diffusion with discontinuous drift and diffusion coefficients.

The proof of the result is based on a semi-martingale decomposition for general point processes derived in [6], referred to as *the Daley-Miyazawa decomposition*. This tool has recently been used in a crucial way in [4] to study a different queueing model in a different scaling regime, specifically, a load balancing model under a mean-field many-server scaling. We believe that this decomposition, along with some of its consequences developed in this paper in

Date: February 23, 2025.

2010 *Mathematics Subject Classification.* 60K25; 60H10; 60H30.

Key words and phrases. semimartingale decomposition; single server queue; level dependent arrival and service rates; heavy traffic; diffusion approximation.

§3, may be useful in far greater generality when it comes to proving scaling limits of queueing models. From this viewpoint, this papers serves to provide an example illustrating the applicability of the approach.

The Daley-Miyazawa decomposition for a counting process was introduced in [6] in order to study extensions of Blackwell's renewal theorem. This decomposition is distinct from the well-known Doob-Meyer decomposition of counting processes (e.g., see Theorem 3.17 in Chapter I of [9]) , a special case of the Doob-Meyer decomposition of submartingales (note that, as a nondecreasing process, a counting process is always a submartingale). In the latter, the finite variation component is predictable, and called the *predictable compensator*. In the special case where this component is absolutely continuous, its density is called the *stochastic intensity* of the point process, a notion that has been used very broadly in the study of point processes. However, when it comes to queueing models, the stochastic intensity and, more generally, the Doob-Meyer decomposition of counting processes, have rarely been used beyond cases where the driving processes are Poisson or compound Poisson. It seems that the reason for this is that the stochastic intensity is not analytically tractable beyond these cases. Contrary to this, the finite variation component is not predictable in the Daley-Miyazawa decomposition. However, its first order term is equal to the law-of-large-numbers limit of the point process, a fact that makes it particularly attractive for scaling limit analysis. Under diffusion scaling, the convergence of its martingale component to a stochastic integral is based on the general theory of martingales, and is quite different from the traditional approach to heavy traffic limits which is based on the central limit theorem.

1.1. **Related work.** *On martingales associated with renewal processes.* Martingales associated with renewal processes have been constructed before and used for analyzing non-Markovian queues in heavy traffic [11, 12]. To contrast these contributions with [6], we point out several differences. First, the Daley-Miyazawa decomposition addresses general point processes, not only renewals. This is important in the setting of this paper as seen in §3 where we apply the decomposition to complicated point processes denoted there by A^n and D^n, representing arrival and departure counts. Further, it is not only the martingale component, but the entire structure of the decomposition, including the bounded variation component and the filtration, that comes into play in the analysis. We have already mentioned a useful property of the bounded variation component. As for the filtration, it turns out that one can construct the decomposition of the various processes on a common filtration (specifically, this is true for A^n and D^n, as we will show in Lemma 3.1), a fact crucially used in our treatment. These aspects are intrinsic to the decomposition proposed in [6] and are important elements of the approach.

On convergence to diffusion with discontinuous coefficients. The paper [13] proposed an approach to establishing weak limits given by diffusion with discontinuous coefficients. This approach gives general conditions in any dimension, based on predictable characteristics. Although potentially applicable here, the method seems less convenient to apply to queueing systems driven by renewals. Indeed, the queueing model provided in [13] to demonstrate their approach is a Markovian one. The reader is referred to [13] for some other works proving convergence of queueing models to diffusions with discontinuous coefficients.

One setting where such discontinuities arise naturally is that of control problem formulations, where a cost is to be minimized in the heavy traffic limit. This was the case with [3], where a controlled queueing problem gave rise to a diffusion with discontinuous drift, and in [1, 2], where, under various different settings, the limit process under asymptotically optimal control was characterized as a diffusion where both coefficients are discontinuous.

On the multi-level queue. A model closely related to the one studied here has appeared in [16], which is called the 2-level $GI/G/1$ queue. The two models have the same level structure, but the dynamics for arrivals is different. In the model from [16], arrivals are subject to different renewal processes according to the level to which the queue length belongs, while service speed is changed in the same way as in our model. Our model simplifies the arrival structure by using a single renewal process for arrivals. This simplification may be meaningful in application because it may well reflect the influence due to congestion through random time change. Theoretically, it enables us to derive a reflected diffusion with discontinuous coefficients as a process limit, whereas only the weak convergence of the stationary distribution is studied in [16]. Furthermore, as discussed there, its process limit may not be characterized by the form of (2.14) except for a special case of its system parameters, so it would be much harder to derive the process limit for the model of [16]. Thus, from theoretical viewpoint, the present result may be considered as the first step to solve this harder problem.

1.2. **Paper organization and notation.** This paper is structured as follows. In the rest of this section, we introduce notation used throughout this paper. In §2, the model is defined and the main result is stated. A martingale toolbox based on the Daley-Miyazawa decomposition is provided in §3. Finally, the proof of the main result is given in §4.

Let $\iota : [0, \infty) \to [0, \infty)$ denote the identity map. For $(\mathbf{X}, d_{\mathbf{X}})$ a Polish space, let $C_{\mathbf{X}}[0, \infty)$ and $D_{\mathbf{X}}[0, \infty)$ denote the space of continuous and, respectively, càdlàg paths $[0, \infty) \to \mathbf{X}$, endowed with the topology of uniform convergence on compacts and, respectively, the J_1 topology. Denote by $C^{\uparrow}_{[0,\infty)}[0, \infty)$ (resp., $D^{\uparrow}_{[0,\infty)}[0, \infty)$) the members of $C_{[0,\infty)}[0, \infty)$ (resp., $D_{[0,\infty)}[0, \infty)$) that are nondecreasing and start at 0. Let $\mathbb{R}^N$ be the N-dimensional real vector space, and denote the Euclidean norm of $x \in \mathbb{R}^N$ by $\|x\|$. For $\xi \in D_{\mathbb{R}^N}[0, \infty)$ and $T \in (0, \infty)$, denote

$$\begin{aligned} \varDelta\xi(t) &= \xi(t) - \xi(t-), \qquad t > 0, \qquad \varDelta\xi(0) = \xi(0), \\ w_T(\xi, \delta) &= \sup\{\|\xi(t) - \xi(s)\| : s, t \in [0, T], |s - t| \le \delta\}, \\ \|\xi\|^*_T &= \sup\{\|\xi(t)\| : t \in [0, T]\}. \end{aligned}$$

Denote by W is a standard Brownian motion in dimension 1. Denote by $\Rightarrow$ convergence in distribution. A sequence of probability measures on $D_{\mathbf{X}}[0, \infty)$ is said to be C-tight if it is tight with the usual topology on $D_{\mathbf{X}}[0, \infty)$ and every weak limit point is supported on $C_{\mathbf{X}}[0, \infty)$. With a slight abuse of terminology, a sequence of random elements (random variables or processes) is referred to as tight when their probability laws form a tight sequence of probability measures, and a similar use is made for the term C-tight.

Some notation related to the Skorohod map on the half line is as follows (see e.g. [5, Section 8] for details). Given $\psi \in D_{\mathbb{R}}[0, \infty)$ with $\psi(0) \ge 0$, there exists a unique pair $(\phi, \eta) \in D_{[0,\infty)}[0, \infty) \times D^{\uparrow}_{[0,\infty)}[0, \infty)$ such that $\phi = \psi + \eta$ and $\int_{[0,\infty)} \phi(t) d\eta(t) = 0$, and this pair is given by

$$\eta(t) = \sup_{s\in[0,t]} \psi(s)^-, \qquad \phi(t) = \psi(t) + \eta(t), \qquad t \geq 0, \tag{1.1}$$

where $x^- = \max(0, -x)$ for $x \in \mathbb{R}$. The solution map from ψ to (ϕ, η) is denoted by Γ, i.e., $(\phi, \eta) = \Gamma(\psi)$.

2. Model and main results

In this section we introduce a sequence of multi-level queues, and state heavy traffic conditions for its diffusion approximation. Then, we present a main theorem, whose proof will be given in the subsequent sections.

2.1. **Sequence of the queueing models.** We index the sequence of multi-level queues by $n \in \mathbb{N}$. The number of threshold levels, denoted by $K \geq 2$, does not depend on n. For the n-th system, threshold levels are nonnegative numbers denoted by $0 = \ell_0^n < \ell_1^n < \cdots < \ell_{K-1}^n$, which partition $[0, \infty)$ into $K + 1$ disjoint sets given as

$$\mathbf{S}_0^n = \{0\}, \qquad \mathbf{S}_i^n = (\ell_{i-1}^n, \ell_i^n], \qquad 1 \leq i \leq K-1, \qquad \mathbf{S}_K^n = (\ell_{K-1}^n, \infty).$$

We are given sequences of positive numbers

$$\lambda_0^n, \qquad \lambda_i^n, \quad \mu_i^n, \quad 1 \leq i \leq K, \qquad n \in \mathbb{N}, \tag{2.1}$$

and $\mu_0^n = 0$. These numbers specify the inter-arrival rate and the service rate, respectively, at times when the queue length belongs to $\mathbf{S}_i^n$.

Denote the queue length of the n-th system at time t by $X^n(t)$, and let $X^n \equiv \{X^n(t); t \geq 0\}$ be the queue length process. Let

$$I^n(t) = \int_0^t 1_{\{X^n(s)=0\}} ds, \tag{2.2}$$

which is the cumulative idleness process. To define the arrival and service processes, we are given the two sequences of positive *i.i.d.* random variables,

$$Z_A(j), \qquad Z_S(j), \qquad j = 0, 1, \ldots \tag{2.3}$$

having unit means and finite and positive variances σ_A^2 and σ_S^2, respectively. Then, define renewal processes A and S by

$$A(t) = \inf\Big\{i \geq 0; t < \sum_{j=0}^{i} Z_A(j)\Big\}, \quad S(t) = \inf\Big\{i \geq 0; t < \sum_{j=0}^{i} Z_S(j)\Big\}, \quad t \geq 0. \tag{2.4}$$

These notations without the index n are commonly used for all the indexed systems.

We return to the n-th system. Let

$$H_i^n(t) = \int_0^t 1_{\{X^n(s)\in\mathbf{S}_i^n\}} ds, \qquad 0 \leq i \leq K, \tag{2.5}$$

and note that $H_0^n(t) = I^n(t)$. Further, let

$$U^n(t) = \sum_{i=0}^{K} \lambda_i^n H_i^n(t), \qquad V^n(t) = \sum_{i=1}^{K} \mu_i^n H_i^n(t), \qquad t \geq 0. \tag{2.6}$$

Using $U^n(t)$ and $V^n(t)$, define the arrival and, respectively, departure counting processes of the n-th system by

$$A^n(t) = A(U^n(t)), \qquad D^n(t) = S(V^n(t)), \qquad t \geq 0. \tag{2.7}$$

Then we have

$$X^n(t) = X^n(0) + A^n(t) - D^n(t), \qquad t \geq 0. \tag{2.8}$$

For simplicity it will be assumed that $X^n(0) = 0$. We regard the processes A and S as the stochastic primitives, and the tuple $(X^n, I^n, H_i^n, U^n, V^n, A^n, D^n)$, which represents the queueing dynamics, to be determined from the primitives via the system of equations (2.2), (2.5)–(2.8). This system of equations has a unique solution given the primitives, as can be proved by induction on the jump times. The inductive argument is standard, and we omit the details. Note that, by construction, A and S are right-continuous, and as a result so are X^n, A^n, D^n.

2.2. **Heavy traffic conditions and main theorem.** We now introduce assumptions on the scaling of the various parameters. The first assumption is about the levels. Assume that, for given constants $\ell_0 = 0 < \ell_1 < \ldots < \ell_{K-1}$, the levels are given by

$$\ell_i^n = n^{1/2} \ell_i, \qquad 0 \leq i \leq K-1. \tag{2.9}$$

Using these ℓ_i, define a partition of $\mathbb{R}_+$ as

$$\mathbf{S}_0 = \{0\}, \qquad \mathbf{S}_i = (\ell_{i-1}, \ell_i], \qquad 1 \leq i \leq K-1, \qquad \mathbf{S}_K = (\ell_{K-1}, \infty).$$

The second assumption is about the state dependent arrival and service rates given by (2.1). Let constants

$$\lambda_i \in (0, \infty), \quad \hat{\lambda}_i \in \mathbb{R}, \qquad 0 \leq i \leq K, \qquad \mu_i \in (0, \infty), \quad \hat{\mu}_i \in \mathbb{R}, \qquad 1 \leq i \leq K,$$

be given. Assume that

$$\lambda_i^n = n\lambda_i + n^{1/2}(\hat{\lambda}_i + o(1)), \tag{2.10}$$

$$\mu_i^n = n\mu_i + n^{1/2}(\hat{\mu}_i + o(1)). \tag{2.11}$$

Note that, according to these assumptions, the time scaling by n is incorporated in the n-th system parameters. From these assumptions, we have

$$\hat{\lambda}_i^n := n^{-1/2}(\lambda_i^n - n\lambda_i) \to \hat{\lambda}_i, \quad 0 \leq i \leq K, \quad \hat{\mu}_i^n := n^{-1/2}(\mu_i^n - n\mu_i) \to \hat{\mu}_i, \quad 1 \leq i \leq K,$$

as $n \to \infty$. We also use the notation

$$\bar{\lambda}_i^n := n^{-1}\lambda_i^n, \qquad 0 \leq i \leq K, \qquad \bar{\mu}_i^n := n^{-1}\mu_i^n, \qquad 1 \leq i \leq K.$$

The critical load condition is assumed:

$$\lambda_i = \mu_i, \qquad 1 \leq i \leq K.$$

Let

$$\hat{X}^n(t) = n^{-1/2} X^n(t), \qquad \hat{I}^n(t) = n^{1/2} \bar{\lambda}_0^n I^n(t), \tag{2.12}$$

and let $b_i^n = \hat{\lambda}_i^n - \hat{\mu}_i^n$, $b_i = \hat{\lambda}_i - \hat{\mu}_i$, $\sigma_i = (\lambda_i \sigma_A^2 + \mu_i \sigma_S^2)^{1/2}$ for $1 \le i \le K$, and $\sigma_0 = \lambda_0^{1/2} \sigma_A$. Define

$$b(x) = \sum_{i=1}^K b_i 1_{\mathbf{S}_i}, \qquad \sigma(x) = \sum_{i=0}^K \sigma_i 1_{\mathbf{S}_i}(x), \qquad x \in \mathbb{R}_+. \tag{2.13}$$

Consider the SDE for $(X, L) \in C_{[0,\infty)}[0,\infty) \times C^{\uparrow}_{[0,\infty)}[0,\infty)$

$$\begin{aligned} X(t) &= x_0 + \int_0^t b(X(s))ds + \int_0^t \sigma(X(s))dW(s) + L(t), \\ \int_{[0,\infty)} & X(t)dL(t) = 0. \end{aligned} \tag{2.14}$$

All assumptions made thus far are in force throughout the paper. The main result of this paper is the following.

Theorem 2.1. *There exists a weak solution to* (2.14)*, and it is unique in law. Moreover, denoting by* (X, L, W) *a solution corresponding to* $x_0 = 0$*, and by* $(\hat{X}^n, \hat{I}^n)$ *the processes associated with the mutli-level queue as in* (2.12)*, one has* $(\hat{X}^n, \hat{I}^n) \Rightarrow (X, L)$ *in* $D_{[0,\infty)}[0,\infty) \times C^{\uparrow}_{[0,\infty)}[0,\infty)$.

Theorem 2.1 will be proved in §4. For this proof, we prepare our key tool in the next section.

3. Daley-Miyazawa decompositions

Recall that A and S were constructed based on $Z_A(j), Z_S(j)$, $j = 0, 1, \ldots$, denoting the times between successive counting events. Let

$$\zeta_A(j) = 1 - Z_A(j), \qquad \zeta_S(j) = 1 - Z_S(j), \qquad j = 0, 1, \ldots$$

Let

$$R_A(t) = \inf\{s > 0 : A(t+s) > A(t)\}, \qquad t \ge 0$$

denote the residual time to the next counting event. Define similarly $R_S(t)$. By definition,

$$t + R_A(t) = \sum_{j=0}^{A(t)} Z_A(j), \qquad t + R_S(t) = \sum_{j=0}^{S(t)} Z_S(j). \tag{3.1}$$

Denote

$$M_A(t) = \sum_{j=1}^{A(t)} \zeta_A(j), \qquad M_S(t) = \sum_{j=1}^{S(t)} \zeta_S(j). \tag{3.2}$$

We then have the Daley-Miyazawa semimartingale representation [6]

$$A(t) = t + R_A(t) - Z_A(0) + M_A(t), \qquad S(t) = t + R_S(t) - Z_S(0) + M_S(t), \tag{3.3}$$

with M_A and M_S, which are martingales with respect to the filtrations generated, respectively, by $\{(A(t), R_A(t)); t \geq 0\}$ and $\{(S(t), R_S(t)); t \geq 0\}$. Our main interest will be in martingales that can be viewed as time changed versions of the above ones. These are

$$\hat{M}_A^n(t) = n^{-1/2} \sum_{j=1}^{A^n(t)} \zeta_A(j), \qquad \hat{M}_S^n(t) = n^{-1/2} \sum_{j=1}^{D^n(t)} \zeta_S(j), \qquad \hat{M}^n = \hat{M}_A^n - \hat{M}_S^n. \tag{3.4}$$

Note, by (3.3), that

$$\begin{aligned} A^n(t) &= A(U^n(t)) = U^n(t) + R_A(U^n(t)) - Z_A(0) + n^{1/2}\hat{M}_A^n(t), \\ D^n(t) &= S(V^n(t)) = V^n(t) + R_S(V^n(t)) - Z_S(0) + n^{1/2}\hat{M}_S^n(t). \end{aligned} \tag{3.5}$$

Of crucial importance to the analysis will be the fact, proved next, that (3.5) gives a semi-martingale decomposition of the arrival and departure processes, A^n and D^n, on a *single* filtration. To make this precise, let the 'state' of the system be given by

$$\mathscr{S}^n(t) = (X^n(t), A^n(t), D^n(t), R_A(U^n(t)), R_S(V^n(t)), \tilde{R}_A^n(t), \tilde{R}_D^n(t)),$$

where

$$\tilde{R}_A^n(t) = \inf\{s > 0 : A^n(t+s) > A^n(t)\}, \qquad \tilde{R}_D^n(t) = \inf\{s > 0 : D^n(t+s) > D^n(t)\}$$

denote the residual times to the next counting events of A^n and D^n. Let

$$\mathscr{F}_t^n = \sigma\{\mathscr{S}^n(s) : 0 \leq s \leq t\}.$$

Because of the inclusion of $\tilde{R}_A^n$ and $\tilde{R}_D^n$, the (right-continuous) counting processes A^n and D^n are $\{\mathscr{F}_t^n\}$-predictable.

Lemma 3.1. *The processes $\hat{M}_A^n$ and $\hat{M}_S^n$ are $\{\mathscr{F}_t^n\}$-square integrable martingales, with optional quadratic variations*

$$[\hat{M}_A^n](t) = n^{-1} \sum_{j=1}^{A^n(t)} \zeta_A(j)^2 \qquad [\hat{M}_S^n](t) = n^{-1} \sum_{j=1}^{D^n(t)} \zeta_S(j)^2, \tag{3.6}$$

and the optional cross variation $[\hat{M}_A^n, \hat{M}_S^n](t)$, is an $\{\mathscr{F}_t^n\}$-martingale. Moreover, the predictable quadratic variations are given by

$$\langle \hat{M}_A^n \rangle(t) = n^{-1}\sigma_A^2 A^n(t) \qquad \langle \hat{M}_S^n \rangle(t) = n^{-1}\sigma_S^2 D^n(t), \qquad \langle \hat{M}_A^n, \hat{M}_S^n \rangle(t) = 0. \tag{3.7}$$

Proof. Because X^n is, by definition, $\{\mathscr{F}_t^n\}$-adapted, so are H_i^n, U^n and V^n. Because $R_A(U^n(t))$ is $\{\mathscr{F}_t^n\}$-adapted and $R_A(U^n(0)) = R_A(0) = Z_A(0)$, one has that $Z_A(0)$ is $\mathscr{F}_0^n$-measurable. The same is true for $Z_S(0)$. In view of the identities (3.5), $\hat{M}_A^n$ and $\hat{M}_S^n$ are also $\{\mathscr{F}_t^n\}$-adapted.

We turn to proving square integrability. To this end, note first that, as a renewal process, $\mathbb{E}A(t) < \infty$ for all t. Since $U^n(t) \leq cnt$ for some constant c, it follows that $\mathbb{E}A^n(t) < \infty$ as well, and by a similar argument, $\mathbb{E}D^n(t) < \infty$ for all t. Next, for an $\{\mathscr{F}_t^n\}$-stopping time τ, denote

$$\mathscr{F}_{\tau-}^n = \mathscr{F}_0^n \vee \sigma\{C \cap \{t < \tau\} : C \in \mathscr{F}_t^n, t \geq 0\}. \tag{3.8}$$

Then it is a standard fact that $\tau \in \mathscr{F}_{\tau-}^n$ [9, I.1.11, I.1.14]. Consider

$$s^n(k) = \inf\{t \geq 0 : A^n(t) \geq k\}, \qquad t^n(k) = \inf\{t \geq 0 : D^n(t) \geq k\}, \qquad k = 1, 2, \ldots$$

These are $\{\mathscr{F}_t^n\}$-stopping times, and as a consequence

$$s^n(k) \in \mathscr{G}_k^n := \mathscr{F}^n_{s^n(k)-}, \qquad t^n(k) \in \mathscr{H}_k^n := \mathscr{F}^n_{t^n(k)-}, \qquad k = 1, 2, \dots \tag{3.9}$$

We now argue that

$$\zeta_A(j) \in \mathscr{G}_k^n \text{ for } j \le k-1, \text{ and } \{\zeta_A(l), l \ge k\} \text{ is independent of } \mathscr{G}_k^n, \qquad k = 1, 2, \dots \tag{3.10}$$

and

$$\text{for every } t,\ A^n(t) \text{ is a stopping time on the discrete parameter filtration } \{\mathscr{G}_k^n\}. \tag{3.11}$$

To this end, let $k \ge 1$. The first assertion in (3.10) follows from the fact that $\{Z_A(j), j \le k-1\}$ can be determined from $\{s^n(j), j \le k\}$ and $\{U^n(s), s < s^n(k)\}$. We next consider the tuple

$$\Theta_k = \{Z_A(j),\ j = 0, 1, \dots, k-1, Z_S(j),\ j = 0, 1, 2, \dots\}.$$

By the construction of the model, all the information about the state of the system up to $s^n(k)-$, namely $\{\mathscr{S}^n(t), t < s^n(k)\}$, is determined by $\Theta_k \in \{\mathscr{G}_k^n\}$, whereas $\{Z_A(l), l \ge k\}$ (hence $\{\zeta_A(l), l \ge k\}$) is independent of Θ_k. This shows the second assertion in (3.10). Thus, (3.10) is proved. Next, by (3.9),

$$\{A^n(t) \le k\} = \{s^n(k) \ge t\} \in \mathscr{G}_k^n,$$

proving (3.11).

The structure (3.9)–(3.10) makes it possible to apply Wald's equations. In particular, by Wald's second equation [7, Theorem 4.1.6],

$$\mathbb{E}\Big[\Big(\sum_{j=1}^{A^n(t)} \zeta_A(j)\Big)^2\Big] = \mathbb{E}[A^n(t)]\mathbb{E}[\zeta_A(0)^2] < \infty.$$

This shows that $\mathbb{E}[\hat{M}_A^n(t)^2] < \infty$ for all t. The argument for $\hat{M}_S^n$ is completely analogous.

For the martingale property of $\hat{M}_A^n$, note first by (3.10) that

$$\mathbb{E}[\zeta_A(k) | \mathscr{G}_k^n] = 0. \tag{3.12}$$

Next, as in the proof of [6, Lemma 2.1], write

$$\hat{M}_A^n(t) = n^{-1/2} \sum_{k=1}^{A^n(t)} \zeta_A(k) = n^{-1/2} \sum_{k \ge 1} \zeta_A(k) 1_{\{s^n(k) \le t\}},$$

and for $s < t$,

$$\mathbb{E}[\hat{M}_A^n(t) | \mathscr{F}_s^n] - \hat{M}^n(s) = n^{-1/2} \sum_{k \ge 1} \mathbb{E}[\zeta_A(k) 1_{\{s < s^n(k) \le t\}} | \mathscr{F}_s^n]. \tag{3.13}$$

By (3.8), if $C \in \mathscr{F}_s^n$ then $C \cap \{s < s^n(k)\} \in \mathscr{G}_k^n$. Moreover, by (3.9), $\{s^n(k) \le t\} \in \mathscr{G}_k^n$. As a result, by (3.12),

$$\begin{aligned}\mathbb{E}[\zeta_A(k) 1_{\{s < s^n(k) \le t\}} 1_C] &= \mathbb{E}[\mathbb{E}[\zeta_A(k) 1_{\{s < s^n(k) \le t\}} 1_C | \mathscr{G}_k^n]] \\ &= \mathbb{E}[\mathbb{E}[\zeta_A(k) | \mathscr{G}_k^n] 1_{\{s < s^n(k) \le t\}} 1_C] = 0, \qquad \forall C \in \mathscr{F}_s^n.\end{aligned}$$

This gives $\mathbb{E}[\zeta_A(k)1_{\{s<s^n(k)\le t\}}|\mathscr{F}^n_s]=0$. Hence, going back to (3.13), we have

$$\mathbb{E}[\hat{M}^n_A(t)|\mathscr{F}^n_s]-\hat{M}^n(s)=0.$$

A similar argument holds for $\hat{M}^n_S$. This completes the proof that $\hat{M}^n_A$ and $\hat{M}^n_S$ are square integrable $\{\mathscr{F}^n_t\}$-martingale.

The expression stated for $[\hat{M}^n_A](t)$ in (3.6) (similarly for $[\hat{M}^n_S](t)$) is immediate from (3.4), which can be written as $\hat{M}^n_A(t)=\sum_{s\le t}\varDelta\hat{M}^n_A(s)$, namely, it is piecewise constant with finitely many jumps on finite intervals.

As for the cross variation,

$$\begin{aligned}m^n(t):=[\hat{M}^n_A,\hat{M}^n_S](t)&=\sum_{s\le t}\varDelta\hat{M}^n_A(s)\varDelta\hat{M}^n_S(s)\\&=n^{-1}\sum_{k=1}^{A^n(t)}\sum_{l=1}^{D^n(t)}\zeta_A(k)\zeta_S(l)1_{\{s^n(k)=t^n(l)\}}.\end{aligned}$$

$\{\mathscr{F}^n_t\}$-adaptedness of m^n is clear from that of $\hat{M}^n_A$ and $\hat{M}^n_S$, and absolute integrability of the former follows from square integrability of the latter. Let now $u^n(k,l)=\min(s^n(k),t^n(l))$. For $k,l\ge 1$, let

$$\Theta_{k,l}=\{Z_A(k'),\,k'<k,\,,Z_S(l'),\,l'<l\}.$$

Then, by construction, all information about $\{\mathscr{S}(s),s<u^n(k,l)\}$ can be determined by $\Theta_{k,l}$, whereas $Z_A(k)$, $Z_S(l)$ and $\Theta_{k,l}$ are mutually independent. This gives

$$\mathbb{E}[\zeta_A(k)\zeta_S(l)|\mathscr{F}^n_{u^n(k,l)-}]=0,\qquad k,l\ge 1. \tag{3.14}$$

Hence for $0\le s\le t$,

$$\begin{aligned}\mathbb{E}[m^n(t)|\mathscr{F}^n_s]-m^n(s)&=n^{-1}\sum_{k\ge1}\sum_{l\ge1}\mathbb{E}[\zeta_A(k)\zeta_S(l)1_{\{s<s^n(k)=t^n(l)\le t\}}|\mathscr{F}^n_s]\\&=n^{-1}\sum_{k\ge1}\sum_{l\ge1}\mathbb{E}[\mathbb{E}[\zeta_A(k)\zeta_S(l)1_{\{s<s^n(k)=t^n(l)\le t\}}|\mathscr{F}^n_{u^n(k,l)-}]|\mathscr{F}^n_s],\end{aligned}$$

where the argument appearing below (3.13) is used again. It follows from (3.9) now that $\{s<s^n(k)=t^n(l)\le t\}\in\mathscr{F}^n_{u^n(k,l)-}$. Hence (3.14) gives

$$\mathbb{E}[\zeta_A(k)\zeta_S(l)1_{\{s<s^n(k)=t^n(l)\le t\}}|\mathscr{F}^n_{u^n(k,l)-}]=0,$$

completing the proof of the martingale property of m^n.

For the first part of (3.7), it suffices to show that $n^{-1}\sigma_A^2A^n(t)$ is the compensator of $[\hat{M}^n_A](t)$ (see [9, Proposition I.4.50 p. 53]). This is an immediate consequence of A^n being predictable and $\mathbb{E}[\zeta_A(j)^2]=\sigma_A^2$. A similar proof holds for the expression $\langle\hat{M}^n_S\rangle$. Finally, $\langle\hat{M}^n_A,\hat{M}^n_S\rangle$ is the compensator of $[\hat{M}^n_A,\hat{M}^n_S]$. Since the latter is a zero mean martingale, it follows that the former is zero. $\square$

4. Proof of main result

The proof of Theorem 2.1 proceeds by showing the following. Recall that $\hat{X}^n$ and $\hat{I}^n$ are defined by (2.12).

Claim 1. Uniqueness in law holds for (2.14).

Claim 2. The sequence $(\hat{X}^n, \hat{I}^n)$ is C-tight.

Claim 3. Every subsequential limit satisfies the SDE (2.14) (with $x_0 = 0$).

Theorem 2.1 will be proved once the three claims are established.

For Claim 1, denote by $\mathcal{A}$ the operator

$$\mathcal{A}f(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x),$$

with b, σ as in (2.13), on the domain $\mathscr{D}(\mathcal{A}) = \{f \in C_0^\infty[0,\infty) : f'(0) = 0\}$, where $C_0^\infty[0,\infty)$ is the set of compactly supported members of $C^\infty[0,\infty)$.

Lemma 4.1. *Let ν be a probability measure on $[0,\infty)$. Then there exists a unique solution to the martingale problem for $(\mathcal{A}, \nu)$. Equivalently, there exists a unique in law weak solution (X, L, W) to SDE* (2.14)*, with $x_0 \sim \nu$, and the law of X is equal to the unique solution to the martingale problem.*

This lemma is closely related to Lemma 2.2 of [14], which, however, is not concerned with the martingale problem.

Proof. For the case of a single discontinuity point ($K = 2$), uniqueness of solutions to the martingale problem was proved in [1, Lemma 4.1(3)]. The proof carries over to the case of multiple discontinuity points with obvious adaptations. Next, by Ito's lemma, if (X, L, W) is a weak solution to the SDE and $f \in \mathscr{D}(\mathcal{A})$ then by the boundary property of L and the condition $f'(0) = 0$, $f(X(t)) - \int_0^t \mathcal{A}f(X(s))ds$ is a martingale. Therefore whenever (X, L, W) is a solution to the SDE, the law of X is a solution to the martingale problem. Moreover, the law of (X, L) is determined by that of X since L is the local time of X at zero. Thus uniqueness in law for weak solutions to the SDE holds.

As for existence of weak solutions to the SDE (hence to the martingale problem), let us first extend b and σ to $\mathbb{R}$ via $b(x) = \mathrm{sgn}(x)b(|x|)$ and $\sigma(x) = \mathrm{sgn}(x)\sigma(|x|)$, where $\mathrm{sgn}(x) = 1$ if $x > 0$ and -1 if $x \le 0$. By [10, Theorem 5.15 p. 341], there exists a weak solution (Q, W) to the SDE without reflection, namely

$$dQ(t) = b(Q(t))dt + \sigma(Q(t))dW(t), \qquad Q(0) \sim \nu.$$

Taking $X(t) = |Q(t)|$, it follows by Tanaka's formula [17, Theorem IV.1.12, p. 222] that (X, L, W) is a weak solution to the SDE, with $L = L_0^Q$, the local time of Q at 0. □

In the rest of this section we prove Claims 2 and 3. The decompositions introduced in §2 now come into play. From (2.8) and (3.5),

$$\hat{X}^n(t) = n^{-1/2}X^n(t) = \hat{M}^n(t) + n^{-1/2}(U^n(t) - V^n(t)) + e^n(t),$$

where

$$e^n(t) = n^{-1/2}\Big[R_A(U^n(t)) - R_S(V^n(t)) - Z_A(0) + Z_S(0)\Big]. \tag{4.1}$$

For $1 \le i \le K$, $n^{1/2}(\bar\lambda_i^n - \bar\mu_i^n) = b_i^n$. Therefore

$$n^{-1/2}(U^n(t) - V^n(t)) = \sum_{i=1}^K b_i^n H_i^n(t) + \hat I^n(t),$$

and we arrive at the relation

$$\hat X^n(t) = \hat M^n(t) + \sum_{i=1}^K b_i^n H_i^n(t) + \hat I^n(t) + e^n(t). \tag{4.2}$$

We now proceed in several steps.

Step 1. Denote $\bar A^n = n^{-1}A^n$ and $\bar D^n = n^{-1}D^n$. Then the sequence $(\bar A^n, \bar D^n)$ is C-tight, and, for each t, the sequences $\bar A^n(t)$ and $\bar D^n(t)$ are uniformly integrable.

To prove the first claim, recall that A is a renewal process for which the first moment of inter-event distribution has mean 1. Therefore $n^{-1}A(n\cdot) \to \iota$ in probability. Writing

$$\bar A^n(t) = n^{-1}A\Big(n\sum_{i=0}^K \bar\lambda_i^n H_i^n(t)\Big), \tag{4.3}$$

and noting that $\bar\lambda_i^n$ are bounded and H_i^n are 1-Lipschitz, gives C-tightness of $\bar A^n$ (see e.g. [9, Proposition VI.3.26] for a necessary and sufficient condition for C-tightness). A similar conclusion holds for $\bar D^n$.

As for the second claim, in view of (4.3), for fixed t, $\bar A^n(t) \le n^{-1}A(cn)$ for some $c = c(t)$. Uniform integrability $n^{-1}A(cn)$ is well known (see [8, (5.6), Chapter 2, page 58]), proving the claim for $\bar A^n(t)$, and similarly for $\bar D^n(t)$.

Step 2. Let $Z(j)$, $j = 1, 2, \dots$ be a collection of i.i.d. random variables such that $\mathbb{E}[Z(1)^2] < \infty$. Then

$$n^{-1/2}\max_{j\le n} Z(j) \to 0 \text{ in probability, as } n \to \infty. \tag{4.4}$$

This is a standard fact, but for completeness we provide a proof. Fix $\varepsilon > 0$. Then, denoting $Z_\varepsilon = \varepsilon^{-2}Z(1)^2$,

$$\begin{aligned}\mathbb{P}(n^{-1/2}\max_{j\le n} Z(j) > \varepsilon) &\le n\mathbb{P}(Z(1) > \varepsilon n^{1/2}) \le n\mathbb{P}(Z(1)^2 > \varepsilon^2 n)\\ &= 2\frac{n}{2}\mathbb{P}(Z_\varepsilon > n) \le 2\sum_{[n/2]\le k\le n}\mathbb{P}(Z_\varepsilon > k) \le 2\sum_{k\ge[n/2]}\mathbb{P}(Z_\varepsilon > k).\end{aligned}$$

The last expression converges to 0 as $n \to \infty$, because, by assumption, $\mathbb{E}[Z_\varepsilon] < \infty$. This proves (4.4).

Step 3. The processes e^n, e_A^n and e_S^n converge to 0 in probability in $D_{\mathbb{R}}[0,\infty)$, where

$$e_A^n(t) = [\hat M_A^n](t) - \langle \hat M_A^n\rangle(t), \qquad e_S^n(t) = [\hat M_S^n](t) - \langle \hat M_S^n\rangle(t).$$

We prove the claim regarding e^n_A (a similar proof holds for e^n_S). By Lemma 3.1,

$$e^n_A(t) = n^{-1} \sum_{j=1}^{A^n(t)} \beta(j), \qquad \beta(j) = \zeta_A(j)^2 - \sigma_A^2.$$

Denoting $\gamma(N) = \sum_{j=1}^N \beta(j)$, $N \in \mathbb{Z}_+$, we have $e^n_A(t) = n^{-1}\gamma(A^n(t))$. Fix t. Given ε, using Step 1, we can find c such that $\mathbb{P}(\bar{A}^n(t) > c) < \varepsilon$. On the event $\bar{A}^n(t) \le c$,

$$\|n^{-1}\gamma(A^n(\cdot))\|^*_t \le n^{-1}\|\gamma\|^*_{[cn]}.$$

Now, $\beta(j)$ are i.i.d. with mean 0. By the functional LLN, $N^{-1}\|\gamma\|^*_N \to 0$ in probability as $N \to \infty$. Thus for any $\delta > 0$

$$\mathbb{P}(\bar{A}^n(t) \le c, \|e^n_A\|^*_t > \delta) \le \mathbb{P}(n^{-1}\|\gamma\|^*_{[cn]} > \delta) \to 0,$$

as $n \to \infty$. Hence $\limsup_n \mathbb{P}(\|e^n_A\|^*_t > \delta) \le \varepsilon$. Since $\varepsilon > 0$ is arbitrary, this shows that $e^n_A \to 0$ in probability.

We now show that $\|e^n\|^*_t \to 0$ in probability as $n \to \infty$ for each fixed $t \ge 0$. It suffices to show

$$\sup_{0\le s\le t} n^{-1/2} R_A\left(\sum_{i=0}^K \lambda^n H^n_i(s)\right) \to 0 \text{ in probability as } n \to \infty,$$

and a similar statement for the term involving R_S in (4.1). Note that $\sum_i H^n_i(t) = t$, while $\bar{\lambda}^n_i$ are bounded. Hence it suffices to show that

$$n^{-1/2}\|R_A\|^*_{cn} \to 0 \quad \text{in probability,} \tag{4.5}$$

for any constant c. To this end, using (3.1), we can write

$$\sup_{0\le s\le T} R_A(s) = \max_{0\le j\le A(T)} Z_A(j).$$

Fix $c > 0$, $\varepsilon > 0$ and $\delta > 0$. Let n_0 and $c_1 < \infty$ be such that $\mathbb{P}(A(cn) > c_1 n) < \delta$ for all $n > n_0$ (such constants exist again by the fact that $n^{-1}A(n\cdot) \to \iota$ in probability). Then

$$\mathbb{P}(n^{-1/2}\|R_A\|^*_{cn} > \varepsilon) \le \mathbb{P}(A(cn) > c_1 n) + \mathbb{P}(n^{-1/2} \max_{0\le j\le c_1 n} Z_A(j) > \varepsilon).$$

The first term on the right is bounded by δ for $n > n_0$, and the second term converges to 0 as $n \to \infty$, by Step 2. Sending $\delta \to 0$, it follows that $\lim_{n\to\infty} \mathbb{P}(n^{-1/2}\|R_A\|^*_{cn} > \varepsilon) = 0$, proving (4.5).

A similar argument holds for the term involving R_S, and it follows that $\|e^n\|^*_t \to 0$ for each fixed $t \ge 0$.

Step 4. The sequences $\langle \hat{M}^n \rangle$ and $[\hat{M}^n]$ are C-tight in $D_{\mathbb{R}}[0,\infty)$.

In view of Lemma 3.1, the C-tightness of $\langle \hat{M}^n_A \rangle$ and $\langle \hat{M}^n_S \rangle$ follows from that of $\bar{A}^n$ and $\bar{D}^n$, proved in Step 1. In view of (3.7),

$$\langle \hat{M}^n \rangle = \langle \hat{M}^n_A \rangle - 2\langle \hat{M}^n_A, \hat{M}^n_S \rangle + \langle \hat{M}^n_S \rangle = \langle \hat{M}^n_A \rangle + \langle \hat{M}^n_S \rangle,$$

and the C-tightness of $\langle \hat{M}^n \rangle$ follows.

By Steps 3 and 4, $[\hat{M}^n_A]$ and $[\hat{M}^n_S]$ are C-tight. To prove C-tightness of $[\hat{M}^n]$ it thus suffices to prove C-tightness of $m^n = [\hat{M}^n_A, \hat{M}^n_S]$. We argue as follows. Given T and δ,

$$\begin{aligned} w_T(m^n,\delta) &= \sup\Big\{\Big|\sum_{t_1<s\le t_2} \Delta\hat{M}^n_A(s)\Delta\hat{M}^n_S(s)\Big| : 0\le t_1<t_2\le T,\, t_2-t_1\le\delta\Big\} \\ &\le \sup\Big\{\sum_{t_1<s\le t_2}\frac{1}{2}(\Delta\hat{M}^n_A(s)^2+\Delta\hat{M}^n_S(s)^2) : 0\le t_1<t_2\le T,\, t_2-t_1\le\delta\Big\} \\ &\le \frac{1}{2}(w_T([\hat{M}^n_A],\delta)+w_T([\hat{M}^n_S],\delta)). \end{aligned}$$

Similarly, $\|m^n\|^*_T \le ([\hat{M}^n](T)+[\hat{M}^n_S](T))/2$. As a result, C-tightness of m^n follows from that of $[\hat{M}^n]$ and $[\hat{M}^n_S]$ (see e.g. [9, Proposition VI.3.26 p. 351]).

Step 5. The sequence $\hat{M}^n$ is tight in $D_{\mathbb{R}}[0,\infty)$. This is immediate from Step 4, using [9, Theorem VI.4.13, p. 358].

Step 6. The sequence $(\hat{M}^n,[\hat{M}^n])$ is C-tight, and if, along a subsequence, $\hat{M}^n \Rightarrow M$, then $(\hat{M}^n,[\hat{M}^n]) \Rightarrow (M,[M])$ along this subsequence.

To prove this statement, we first verify the condition

$$\sup_n \mathbb{E}(\sup_{s\le t}|\Delta\hat{M}^n(s)|)<\infty \qquad \text{for all } t. \tag{4.6}$$

To this end, write

$$\mathbb{E}\sup_{s\le t}|\Delta\hat{M}^n_A(s)| \le 1+\int_1^\infty \mathbb{P}(n^{-1/2}\sup_{j\le A^n(t)}|\zeta_A(j)|>r)dr \le 1+\int_1^\infty \frac{\mathbb{E}\sup_{j\le A^n(t)}\zeta_A(j)^2}{nr^2}dr.$$

For $t\ge 1$, we have $A^n(t)\le A(cnt)$ for suitable constant c, and so

$$\mathbb{E}\sup_{j\le A^n(t)}\zeta_A(j)^2 \le \mathbb{E}\sup_{j\le A(cnt)}\zeta_A(j)^2 \le \mathbb{E}\sum_{j=1}^{A(cnt)}\zeta_A(j)^2 = \mathbb{E}(A(cnt))\mathbb{E}(\zeta_A(1)^2).$$

Since $\mathbb{E}A(cnt)\le c_1nt$ for some constant c_1, we obtain that $\mathbb{E}\sup_{s\le t}|\Delta\hat{M}^n_A(s)|$ is bounded uniformly in n. A similar statement holds for $\hat{M}^n_S$, and, since $|\Delta\hat{M}^n|\le|\Delta\hat{M}^n_A|+|\Delta\hat{M}^n_S|$, also for $\hat{M}^n$. This verifies (4.6).

Consider a subsequence along which $\hat{M}^n$ converges in distribution in $D_{\mathbb{R}}[0,\infty)$ and denote its limit by M. We can now apply [9, Corollary VI.6.30, p. 385], by which, under condition (4.6), $(\hat{M}^n,[\hat{M}^n])\Rightarrow(M,[M])$ holds along the same sequence. Next, in view of Step 4, $[M]$ is continuous. This implies that M is also continuous (see e.g. [9, Theorem I.4.47(c), p. 52]). Therefore $(\hat{M}^n,[\hat{M}^n])$ is C-tight.

Step 7. Let

$$\hat{Y}^n = \hat{M}^n + \sum_{i=1}^K b^n_i H^n_i + e^n.$$

We show that

$$(\bar{A}^n, \bar{D}^n, \hat{X}^n, \hat{I}^n, \hat{M}^n, [\hat{M}^n], \hat{Y}^n, (H_i^n)_{0\le i\le K}, e^n, e_A^n, e_S^n) \tag{4.7}$$

is C-tight, and every subsequential weak limit is given by

$$(\bar{A}, \bar{D}, X, L, M, [M], Y, (H_i)_{0\le i\le K}, 0, 0, 0), \tag{4.8}$$

where M is a martingale (on its own filtration), and

$$(X, L) = \Gamma(Y), \qquad H_0 = 0, \qquad Y = M + \sum_{i=1}^K b_i H_i,$$

$$\bar{A} = \sum_{i=1}^K \lambda_i H_i, \qquad \bar{D} = \sum_{i=1}^K \mu_i H_i, \qquad [M] = \sum_{i=1}^K \sigma_i^2 H_i. \tag{4.9}$$

To prove this statement, note that $\hat{X}^n = \hat{Y}^n + \hat{I}^n \ge 0$ and, by definition,

$$\int_{[0,\infty)} \hat{X}^n(t) d\hat{I}^n(t) = 0.$$

As a result, we can rewrite (4.2) as

$$(\hat{X}^n, \hat{I}^n) = \Gamma(\hat{Y}^n). \tag{4.10}$$

Now, H_i^n are uniformly Lipschitz and are therefore C-tight. Also, b_i^n are bounded, therefore it follows from Steps 1 and 6 that $\hat{Y}^n$ are C-tight. The Lipschitz continuity of the map $\Gamma : D_{\mathbb{R}}[0,\infty) \to (D_{\mathbb{R}}[0,\infty))^2$ with respect to the uniform topology on compacts, implies that the sequence (4.7) C-tight, and that the relation (4.10) is preserved under the limit. Moreover, the tightness of $\hat{I}^n$ implies that $H_0 = 0$.

Fix a convergent subsequence of (4.7) and denote its weak limit by (4.8). Because M is the weak limit of the martingales $\hat{M}^n$, to show that M is a martingale, it suffices to show that for fixed t, $\hat{M}^n(t)$ are uniformly integrable. This is true by $\sup_n \mathbb{E}(\hat{M}^n(t)^2) < \infty$, which follows from Lemma 3.1 and $\sup_n n^{-1}\mathbb{E}A^n(t) < \infty$. Hence M is a martingale.

The claim that the limit $\bar{A}$ of $\bar{A}^n$ is given by the expression in (4.9) follows from (4.3) and the convergence $n^{-1}A(n\cdot) \to \iota$. The form (4.9) of $\bar{D}$ follows similarly.

By Lemma 3.1 and the convergence of $(\bar{A}^n, \bar{D}^n)$, we have $(\langle \hat{M}_A^n\rangle, \langle \hat{M}_S^n\rangle) \Rightarrow (\sigma_A^2 \bar{A}, \sigma_S^2 \bar{D})$. Because $\langle \hat{M}_A^n, \hat{M}_S^n\rangle = 0$, it follows that $\langle \hat{M}^n\rangle \Rightarrow \sigma_A^2\bar{A} + \sigma_S^2\bar{D}$, which is $\sum_{i=1}^K \sigma_i^2 H_i$, in view of $\sigma_i^2 = \lambda_i\sigma_A^2 + \mu_i\sigma_S^2$, $1 \le i \le K$.

By Step 6, $(\hat{M}^n, [\hat{M}^n]) \Rightarrow (M, [M])$ along the convergent subsequence. Suppose we also show that, along the subsequence,

$$\langle M\rangle = \tilde{H} := \sum_{i=1}^K \sigma_i^2 H_i. \tag{4.11}$$

Since M is a continuous martingale, $[M] = \langle M\rangle$, and thus the expression for $[M]$ in (4.9) would be proved.

Thus, to complete the proof of Step 7, it remains to show (4.11). To this end, note that we have shown that $\mathit{\Lambda}^n := (\hat{M}^n)^2 - \langle \hat{M}^n \rangle \Rightarrow \mathit{\Lambda} := M^2 - \tilde{H}$. Let us show that $\mathit{\Lambda}^n(t)$ are uniformly integrable for each t. Recall from Step 1 that $\bar{A}^n(t)$ and $\bar{D}^n(t)$ are uniformly integrable for each fixed t. In view of (3.7), this implies that $\langle \hat{M}^n \rangle$ is uniformly integrable. By the definition (3.4), we can apply [8, Theorem 6.2, Chapter 1, page 32] to $\hat{M}^n_A$ and $\hat{M}^n_A$, which gives the uniform integrability of $(\hat{M}^n_A)^2$ and $(\hat{M}^n_S)^2$. This proves the uniform integrability of $(\hat{M}^n)^2$ because $(\hat{M}^n)^2(t) \le 2[(\hat{M}^n_A)^2(t) + (\hat{M}^n_S)^2(t)]$. Hence, $\mathit{\Lambda}^n$ is uniformly integrable. Next, let $\tilde{\mathscr{F}}_t = \mathscr{F}^{\tilde{H}}_t \vee \mathscr{F}^M_t$. We will now show that $\mathit{\Lambda}$ is an $\{\tilde{\mathscr{F}}_t\}$-martingale. Since this process is adapted and integrable, it remain to show that for $0 \le s \le t$, $\mathbb{E}[\mathit{\Lambda}(t) - \mathit{\Lambda}(s)|\tilde{\mathscr{F}}_s] = 0$ a.s. For this, it suffices that for any bounded continuous $h : D_{\mathbb{R}}[0,\infty) \times C_{\mathbb{R}}[0,\infty) \to \mathbb{R}$,

$$\mathbb{E}[h(M(\cdot \wedge s), \tilde{H}(\cdot \wedge s))(\mathit{\Lambda}(t) - \mathit{\Lambda}(s))] = 0. \tag{4.12}$$

Since $\mathit{\Lambda}^n$ is an $\{\mathscr{F}^n_t\}$-martingale for each $n \ge 1$, we have

$$\mathbb{E}[h(\hat{M}^n(\cdot \wedge s), \langle \hat{M}^n \rangle(\cdot \wedge s))(\mathit{\Lambda}^n(t) - \mathit{\Lambda}^n(s))] = 0.$$

The convergence $(\hat{M}^n, \langle \hat{M}^n \rangle, \mathit{\Lambda}^n) \Rightarrow (M, \tilde{H}, \mathit{\Lambda})$, the boundedness of h and uniform integrability of $\mathit{\Lambda}^n(s)$ and $\mathit{\Lambda}^n(t)$ imply (4.12). This shows that $\mathit{\Lambda}$ is an $\{\tilde{\mathscr{F}}_t\}$-martingale. Since $\tilde{H}$ is adapted to this filtration and continuous, it is predictable on it. Hence by the uniqueness of the Doob-Meyer decomposition for the submartingale M^2, we conclude that $\langle M \rangle = \sum_{i=1}^K \sigma_i^2 H_i$. This proves (4.11) and completes Step 7.

This also completes the proof of Claim 2.

It remains to show that the limit of every convergent subsequence forms a solution to the SDE. Fix a convergent subsequence of (4.7), and denote its limit by (4.8). Invoking Skorohod's representation, assume without loss of generality that this convergence is a.s.

Step 8. Denote $\mathscr{F}_t = \mathscr{F}^X_t \vee \mathscr{F}^M_t$. We show that M is an $\{\mathscr{F}_t\}$-martingale, by an argument similar to the one in the previous step. Clearly, it is adapted and integrable, so it remains to show that for $0 \le s \le t$, $\mathbb{E}[M(t) - M(s)|\mathscr{F}_s] = 0$ a.s. For this, it suffices to show that for any bounded continuous $h : (D_{\mathbb{R}}[0,\infty))^2 \to \mathbb{R}$,

$$\mathbb{E}[h(X(\cdot \wedge s), M(\cdot \wedge s))(M(t) - M(s))] = 0. \tag{4.13}$$

Since $\hat{M}^n$ is an $\{\mathscr{F}^n_t\}$-martingale, we have

$$\mathbb{E}[h(\hat{X}^n(\cdot \wedge s), \hat{M}^n(\cdot \wedge s))(\hat{M}^n(t) - \hat{M}^n(s))] = 0.$$

Moreover, $(\hat{X}^n, \hat{M}^n) \to (X, M)$ a.s. in $(D_{\mathbb{R}}[0,\infty))^2$, and so (4.13) will follow provided that $h(\hat{X}^n(\cdot \wedge s), \hat{M}^n(\cdot \wedge s))(\hat{M}^n(t) - \hat{M}^n(s))$ are uniformly integrable. This last condition holds by the boundedness of h and the fact that $\sup_n \mathbb{E}(\hat{M}^n(t)^2) < \infty$, which follows from Lemma 3.1 and $\sup_n n^{-1}\mathbb{E}A^n(t) < \infty$.

Step 9. We show that

$$\int_0^\infty 1_{\{X_t = \ell\}} dt = 0, \qquad \ell = \ell_0, \ell_1, \ldots, \ell_{K-1}, \qquad a.s. \tag{4.14}$$

To this end, note first that, a.s., $d\langle M\rangle(t) = d[M](t) \geq c_1 dt$ as measures, where $c_1 = \min_{1\leq i\leq K} \sigma_i^2 > 0$. This follows from the fact that $\sum_{i=1}^K H_i(t) = t$ (recall that $H_0 = 0$), and the expression for $[M]$ in (4.9).

Fix $0 \leq i \leq K-1$ and set $\ell = \ell_i$. Then by the occupation time formula [17, Corollary VI.1.6], with $L_a^X(t)$ the local time of X at $a \in [0,\infty)$ by time t,

$$\int_0^t 1_{\{X_s=\ell\}} d\langle X\rangle(s) = \int_0^\infty L_a^X(t) 1_{\{a=\ell\}} da = 0.$$

But $d\langle X\rangle(t) = d\langle M\rangle(t) \geq c_1 dt$, and (4.14) follows.

Step 10. Finally, we show that for any subsequential limit (4.8) of (4.7), the law of X is a solution to the martingale problem. In view of Lemma 4.1, this proves Claim 3.

The convergence $\hat{X}^n \to X$ along with Step 9 imply that H_i, $1 \leq i \leq K$, which are the limits of $H_i^n = \int_0^\cdot 1_{\{\hat{X}^n(s)\in\mathbf{S}_i\}} ds$, are given by $H_i = \int_0^\cdot 1_{\{X(s)\in\mathbf{S}_i\}} ds$. As a result,

$$\sum_{i=1}^K b_i H_i(t) = \int_0^t b(X(s))ds,$$

where the expression (2.13) for b is used. Hence, from the limit (4.8), we have

$$X(t) = \int_0^t b(X(s))ds + M(t) + L(t). \tag{4.15}$$

As for $[M]$, the expression in (4.9) and the fact that $H_0 = 0$ shows that

$$[M](t) = \sum_{i=1}^K \sigma_i^2 H_i = \sum_{i=0}^K \sigma_i^2 H_i = \int_0^t \sigma^2(X_s)ds,$$

where the expression in (2.13) for σ is used. Apply now Ito's lemma to the semimartingale (4.15) for $f \in \mathscr{D}(\mathcal{A})$, that is, for smooth f satisfying $f'(0) = 0$, to get

$$\begin{aligned} f(X(t)) - \int_0^t \mathcal{A}f(X(s))ds &= f(0) + \int_0^t f'(X_t)dL(t) + \int_0^t f'(X_t)dM(t) \\ &= f(0) + \int_0^t f'(X_t)dM(t). \end{aligned}$$

Since M is an $\{\mathscr{F}_t\}$-martingale by Step 8, so is the last term. This shows that the law of X is a solution to the martingale problem for $(\delta_0, \mathcal{A})$. By Lemma 4.1, this proves Claim 3, and the proof of Theorem 2.1 is complete. □

Acknowledgment. The first author is supported by ISF grant 1035/20.

References

[1] R. Atar, E. Castiel, and M. I. Reiman. Asymptotic optimality of switched control policies in a simple parallel server system under an extended heavy traffic condition. *Stochastic Systems*, 14(4), 2024.

[2] R. Atar, E. Castiel, and M. I. Reiman. Parallel server systems under an extended heavy traffic condition: A lower bound. *The Annals of Applied Probability*, 34(1B):1029–1071, 2024.

[3] R. Atar and A. Lev-Ari. Workload-dependent dynamic priority for the multiclass queue with reneging. *Mathematics of Operations Research*, 43(2):494–515, 2018.

[4] R. Atar and G. Wolansky. Invariance principle and McKean-Vlasov limit for randomized load balancing in heavy traffic. *arXiv preprint arXiv:2402.18143*, 2024.

[5] K. L. Chung and R. J. Williams. *Introduction to Stochastic Integration*, volume 2. Springer, 1990.

[6] D. J. Daley and M. Miyazawa. A martingale view of Blackwell's renewal theorem and its extensions to a general counting process. *Journal of Applied Probability*, 56(2):602–623, 2019.

[7] R. Durrett. *Probability: Theory and Examples.* Cambridge University Press, 2019.

[8] A. Gut. *Stopped Random Walks.* Springer, 2009.

[9] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 1987. xviii+601 pp.

[10] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, second edition, 1991. xxiv+470 pp. URL `http://dx.doi.org/10.1007/978-1-4612-0949-2`.

[11] E. Krichagina. Diffusion-approximation for a queue in a multiserver system with multistage service. *Automation and Remote Control*, 50(3):346–354, 1989.

[12] E. V. Krichagina and M. I. Taksar. Diffusion approximation for GI/G/1 controlled queues. *Queueing systems*, 12:333–367, 1992.

[13] N. Krylov and R. Liptser. On diffusion approximation with discontinuous coefficients. *Stochastic Processes and Their Applications*, 102(2):235–264, 2002.

[14] M. Miyazawa. Multi-level reflecting Brownian motion on the half line and its stationary distribution. *Journal of the Indian Society for Probability and Statistics*, 25:543–574, 2024. URL `https://arxiv.org/abs/2405.16764`.

[15] M. Miyazawa. The stationary distributions of state-dependent diffusions reflected at one and two sides. Technical report, 2024. URL `https://arxiv.org/abs/2410.11496`.

[16] M. Miyazawa. Diffusion approximation of the stationary distribution of a two-level single server queue. *Advances in Applied Probability (to appear)*, 57(2), June 2025. URL `https://arxiv.org/abs/2312.11284`.

[17] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, third edition, 1999. xiv+602 pp.

[18] A. Weerasinghe. Diffusion approximations for G/M/n+ GI queues with state-dependent service rates. *Mathematics of Operations Research*, 39(1):207–228, 2014.

Viterbi Faculty of Electrical and Computer Engineering, Technion – Israel Institute of Technology, Haifa, Israel

Email address: `rami@technion.ac.il`

Department of Information Sciences, Tokyo University of Science, Noda, Chiba, Japan

Email address: `miyazawa@rs.tus.ac.jp`