Stein's method for diffusive limit of queueing processes

Eustache Besançon, Laurent Decreusefond, Pascal Moyal

STEIN'S METHOD FOR DIFFUSIVE LIMITS OF QUEUEING PROCESSES

E. BESANÇON, L. DECREUSEFOND, AND P. MOYAL

Abstract. Donsker Theorem is perhaps the most famous invariance principle result for Markov processes. It states that when properly normalized, a random walk behaves asymptotically like a Brownian motion. This approach can be extended to general Markov processes whose driving parameters are taken to a limit, which can lead to insightful results in contexts like large distributed systems or queueing networks. The purpose of this paper is to assess the rate of convergence in these so-called diffusion approximations, in a queueing context. To this end, we extend the functional Stein method introduced for the Brownian approximation of Poisson processes, to two simple examples: the single-server queue and the infinite-server queue. By doing so, we complete the recent applications of Stein's method to queueing systems, with results concerning the whole trajectory of the considered process, rather than its stationary distribution.

Diffusion approximation, Queueing systems, Stein's method

1. Introduction

The Markovian analysis of queueing systems often leads to stochastic processes with an intricate evolution, for which the classical approach, which for instance requires the computation of the stationary distribution, is intractable. To gain some insights on the behavior of the process, it is then customary to push the parameters to their limit and analyze the limiting process which hopefully will reveal the inner structure of the model under analysis. Diffusion approximations, as they are called, have been and still are the subject of numerous papers (see [15, 19] and references therein to get a glimpse of the very rich literature on the subject). The most naive example which comes to mind is the convergence of a normalized Poisson process to a Brownian motion B: Letting N^λ be a Poisson process of intensity λ, we have that

$$\tilde{N}^\lambda \left(=: t \mapsto \frac{1}{\sqrt{\lambda}} (N^\lambda(t) - \lambda t) \right) \xrightarrow[\lambda \to \infty]{\text{dist. in } \mathbb{D}_T} B, \tag{1}$$

where the convergence holds in distribution on the Skorokhod space as λ goes to infinity. As the convergence in distribution is induced by a metric over the set of probability measures, Eqn. (1) just says that the distance between the distribution of $\tilde{N}^\lambda$ and the distribution of B over $\mathbb{D}_T$ tends to zero. The next step is to determine the rate at which this limit holds. The first study addressing this issue was due to Barbour in the 90's [1]. Since then, no papers on this subject appeared until [6, 13, 16]. These four papers share the same common ground, relying on the so-called Stein method (SM) [18], see Section 4.1 for a modern introduction. It is based on the fact that

the topology of convergence in distribution over a separable metric space χ can be defined through the distance

$$\mathrm{d}(\mu,\nu) = \sup_{f\in \mathrm{Lip}\,-1} \left(\int_\chi f \ \mathrm{d}\mu - \int_\chi f \ \mathrm{d}\nu \right),$$

where $\mathrm{Lip}\,-1$ is the set of Lipschitz continuous function: $f : \chi \to \mathbb{R}$ such that

$$|f(x) - f(y)| \leq \mathrm{d}_\chi(x,y), \ \forall x, y \in \chi.$$

An important avenue of literature has been dedicated to Stein's method in the case $\chi = \mathbb{R}$. Close to the class of models we have in mind, let us mention the fruitful recent applications of the SM, to assess the rate of convergence of the stationary distributions of various processes involved in queueing: Erlang-A and Erlang-C systems in [5]; a system with reneging and phase-type service time distributions (in which case the target distribution is the stationary distribution of a piecewise Ornstein-Uhlenbeck process) in [4], single-server queues in heavy-traffic in [12].

When χ is no longer $\mathbb{R}$, however, the development of the SM is much more involved. The main contribution of this work is to present applications of the Stein's method to estimate the rate of convergence in functional CLT's arising in queueing. Specifically, we complete existing functional Central Limit Theorems of classical queueing systems (namely the M/M/1 queue and the 'pure delay' M/M/∞ system) by assessing the rate of convergence to the diffusion limit, using Stein's method at the level of the whole stochastic process. These two examples thus provide good illustrations of how the SM can be fruitfully applied in a queueing context, at the process level. By completing two classical asymptotic results with a simple rate of convergence estimate, for classes of functions that have a practical meaning, the present work can thus constitute a promising starting point for similar development regarding a larger class of queueing systems.

This paper is organized as follows. In Section 2, we state our main results for the diffusion approximation of the M/M/1 and M/M/∞ queues. In Section 3, we introduce the intermediate processes, i.e. the affine interpolation of both the Markov process under study and the limit Brownian motion. Then, we estimate the error done by replacing the original processes by their affine interpolations. Section 3 is devoted to the functional Stein method with which we control the distance between the distributions of the interpolations defined above. The specific calculations for the M/M/1 queue are done in Section 5 and in Section 6 for the M/M/∞ system. The Appendix contains the proofs of two technical lemmas.

2. The results

In this section we present our main results. In Theorems 2.1 and 2.2 below, we provide bounds for the speed of convergence in the diffusion approximation of two standard queueing systems: the single-server and the infinite server queues, respectively. In what follows, for any $T > 0$, $\mathbb{D} := \mathbb{D}_T$ denotes the Skorokhod space of cÃădlÃăg functions from $[0,T]$ to $\mathbb{R}$. (We omit the dependance in T for notational simplicity.) The functional space

Σ, to be properly defined in Definition 4.2 below, is a subspace of the space of 1-Lipschitz continuous from $\mathbb{D}$ to $\mathbb{R}$. We denote for any U, V in $\mathbb{D}$,

$$\mathrm{d}_\Sigma(U,V) = \sup_{F\in\Sigma} |\mathbf{E}[F(U)] - \mathbf{E}[F(V)]|. \tag{2}$$

The distance d_Σ is then the appropriate tool to introduce the results to come.

2.1. **The M/M/1 queue.** We first consider the classical $\mathrm{M}_\lambda/\mathrm{M}_\mu/1$ queue, that is, a single server with infinite buffer, where the arrival is Poisson of intensity λ, and the service times are i.i.d. from the exponential law $\varepsilon(\mu)$. For all $t \geq 0$, we let $L^\dagger(t)$ denote the number of customers in the system (including the one in service, if any) at time t. The process $L^\dagger$ is clearly birth and death, and is ergodic if and only if $\lambda/\mu < 1$. If the initial size of the system is $x \in \mathbb{N}$, then $L^\dagger$ obeys the SDE

$$L^\dagger(t) = x + N_\lambda(t) - \int_0^t \mathbf{1}_{\{L^\dagger(s^-)>0\}} N_\mu(ds), \quad t \geq 0, \tag{3}$$

for two independent Poisson processes N_λ and N_μ. This process is rescaled by accelerating time by a factor n, while multiplying the initial value, and then dividing the number of customers in the system at any time by the same factor. For all $n \in \mathbb{N}^*$, the resulting normalized process $\overline{L^\dagger}$ then satisfies

$$\overline{L_n^\dagger}(t) = x + \frac{N_{n\lambda}(t)}{n} - \frac{N_{n\mu}(t)}{n} + \frac{1}{n}\int_0^t \mathbf{1}_{\{L_n^\dagger(ns^-)=0\}}\, \mathrm{d}N_{n\mu}(s), \quad t \geq 0.$$

It is a well established fact (see e.g. Proposition 5.16 in [15]) that the sequence $\left(\overline{L_n^\dagger} : n \geq 1\right)$ converges in probability and uniformly over compact sets, to the deterministic function

$$\overline{L^\dagger} : t \longmapsto (x + \lambda t - \mu t)^+,$$

and that the process

$$Z_n^\dagger : t \longmapsto \frac{\sqrt{n}}{\sqrt{\lambda+\mu}} \left(\overline{L_n^\dagger}(t) - \overline{L^\dagger}(t)\right)$$

converges in distribution in $\mathbb{D}$ to the standard Brownian motion.

We can control the speed of the latter convergence. For that purpose, we bound for any fixed n and any horizon T, the Σ-distance between these processes, defined by (2). We have the following result,

Theorem 2.1. *Suppose that $\lambda < \mu$ and let $T \leq \frac{x}{\mu-\lambda}$. Then, there exists a constant c_T such that for all $n \in \mathbb{N}$,*

$$\mathrm{d}_\Sigma\left(Z_n^\dagger, B\right) \leq \frac{c_T \log n}{\log\log n \sqrt{n}},$$

where B is a standard Brownian motion.

The proof of Theorem 2.1 is deferred to Section 5.

2.2. **The infinite server queue M/M/∞.** We now turn to the classical "infinite server" $\mathrm{M}_\lambda/\mathrm{M}_\mu/\infty$ queue: a potentially unlimited number of servers attend customers that enter the system following a Poisson process of intensity λ, requesting service times that are exponentially distributed of parameter μ (where $\lambda, \mu > 0$).

Assuming throughout that the system is initially empty, let $L^\sharp(t)$ denote the number of customers in the system at time t. The process $L^\sharp$ is a.s. an element of $\mathbb{D}$; this is an ergodic Markov process which obeys the SDE

$$L^\sharp(t) = N_\lambda(t) - \sum_{i=1}^{+\infty} \int_0^t \mathbf{1}_{\{L^\sharp(s^-)\geq i\}} N_\mu^i(ds), \quad t \geq 0,$$

where N_λ is a Poisson process of intensity λ and the N_μ^i's are independent Poisson processes of intensity μ. The classical scaling of the process $L^\sharp$ goes as follows; we accelerate time by a factor $n \in \mathbb{N}^*$ and divide the size of the system by n. The corresponding n-th rescaled process is then defined by

$$\overline{L_n^\sharp} : t \longmapsto \frac{N_{\lambda n}(t)}{n} - \frac{1}{n} \sum_{i=1}^{+\infty} \int_0^t \mathbf{1}_{\{\overline{L_n^\sharp}(s^-) > \frac{i}{n}\}} N_\mu^i(ds).$$

It is a well known fact (see e.g. Theorem 6.13 in [15]) that the sequence of processes $\left(\overline{L_n^\sharp}, n \geq 0\right)$ converges in L^1 and uniformly over compact sets to the deterministic function

$$\overline{L^\sharp} : t \longmapsto \rho - \rho e^{-\mu t}, \tag{4}$$

where $\rho = \lambda/\mu$. Moreover, if we define for all n the process

$$Z_n^\sharp : t \longmapsto \sqrt{n}\left(\overline{L_n^\sharp}(t) - \overline{L^\sharp}(t)\right), \tag{5}$$

then the sequence $\left(Z_n^\sharp : n \geq 0\right)$ converges in distribution to the process $Z^\sharp$ defined by

$$Z^\sharp : t \longmapsto Z^\sharp(t) = Z^\sharp(0)e^{-\mu t} + \int_0^t e^{-\mu(t-s)} \sqrt{h(s)}\ \mathrm{d}B(s), \tag{6}$$

where $h(t) = \lambda\left(2 - e^{-\mu t}\right)$ for all $t \geq 0$; see e.g. [2] or Theorem 6.14 in [15]. We have the following result,

Theorem 2.2. *For any $T > 0$, there exists a constant $c_T > 0$ such that for all $n \geq 1$,*

$$\mathrm{d}_\Sigma(Z_n^\sharp, Z^\sharp) \leq \frac{c_T \log n}{\log\log n\, \sqrt{n}}.$$

We defer the proof of Theorem 2.2 to Section 6.

2.3. **Consequences.** Let us quote a few functionals which are often encountered in queueing analysis, and which are regular enough to be elements of Σ (see Definition 4.2 below). This is the case, first, for the function F_f, that is defined for any mild enough function f and $T > 0$, by

$$F_f : \begin{cases} \mathbb{D} & \longrightarrow \mathbb{R} \\ x = \left(x_t,\, t \in [0,T]\right) & \longmapsto \dfrac{1}{T}\displaystyle\int_0^T f(x_s)\ \mathrm{d}s, \end{cases}$$

observing that $F_f(X)$ goes to $\mathbb{E}_\pi[f]$ for large T whenever the Markov process X is ergodic of invariant probability π. The proof is deferred to Remark 2 below. Similarly, for $M \geq 0$ and $p \geq 2$,

$$F_{M,p} : \begin{cases} \mathbb{D} & \longrightarrow \mathbb{R} \\ x & \longmapsto \left(\int_0^T |x_s \wedge M|^p \, \mathrm{d}s \right)^{1/p} \end{cases}.$$

also belongs to the set of admissible test functions. Observe that for M and p large enough, $F_{M,p}(x)$ can be considered as an ersatz to $\sup_{s \leq T} |x_s|$.

For any of these functionals F, if $\mathrm{d}(\mathbf{P}_{Xn}, \mathbf{P}_X)$ tends to 0 as $n^{-\alpha}$, then the distribution of the random variables $(F(X_n),\, n \geq 1)$ converges in the sense of a damped Kantorovith-Rubinstein distance at a rate $n^{-\alpha}$:

$$\sup_{\varphi \in \mathcal{C}_b^3} \left| \mathbf{E}\left[\varphi\Big(F(X_n)\Big)\right] - \mathbf{E}\left[\varphi\Big(F(X)\Big)\right] \right| \leq c\, n^{-\alpha},$$

where $\mathcal{C}_b^3$ is the set of three times differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ with bounded derivatives of any order. Note that this kind of result is inaccessible via the standard Stein's method in dimension 1, since we usually cannot achieve the first step of the SM, which consists in devising a functional characterization of the distribution of $F(X)$.

3. Interpolation of Markov processes

To prove Theorems 2.1 and 2.2, we will be led to bound the distance between the affine interpolation of the Markov process under consideration ($Z^\dagger$ in the first case, $Z^\sharp$ in the second), and that of a (time-changed) Brownian motion, on a finite horizon $T > 0$.

For fixed $T > 0$ and $n \in \mathbb{N}^*$, let us denote throughout this paper, by t_i^n, $i = 0, ..., n$, the points of the discretization of $[0, T]$ of constant mesh T/n, namely $t_i^n = iT/n$ for all $i = 1, ..., n$. For a function $f \in \mathbb{D}$, denote by $\Pi_n f$ its affine interpolation on the latter grid, that is, for all $t \in [0, T]$, for $k = 1, ..., n$ such that $t \in \left[t_{k-1}^n, t_k^n\right]$,

$$\Pi_n f(t) = \frac{n}{T}\left(f\left(t_k^n\right) - f\left(t_{k-1}^n\right)\right)\left(t - t_{k-1}^n\right) + f\left(t_{k-1}^n\right). \tag{7}$$

An immediate computation then shows that for all $t \leq T$ and for k as above, we have that

$$\begin{aligned} \Pi_n f(t) &= \frac{n}{T}\left(\left(f\left(t_k^n\right) - f\left(t_{k-1}^n\right)\right)\left(t - t_{k-1}^n\right) + \sum_{i=1}^{k-1} \left(f\left(t_i^n\right) - f\left(t_{i-1}^n\right)\right)\frac{T}{n} \right) + f(0) \\ &= \frac{n}{T}\left(\sum_{i=1}^{n} \left(f\left(t_i^n\right) - f\left(t_{i-1}^n\right)\right) \int_0^t \mathbf{1}_{\left[t_{i-1}^n, t_i^n\right)}(s)\, ds \right) + f(0) \\ &= \sqrt{\frac{n}{T}}\left(\sum_{i=1}^{n} \left(f\left(t_i^n\right) - f\left(t_{i-1}^n\right)\right) h_i^n(t) \right) + f(0), \end{aligned} \tag{8}$$

where

$$h_i^n : t \longmapsto \sqrt{\frac{n}{T}} \int_0^t \mathbf{1}_{\left[t_{i-1}^n, t_i^n\right)}(s)\, ds, \quad i = 1, ..., n. \tag{9}$$

In what follows, B denotes a standard one dimensional Brownian motion and observe that $\Pi_n B$ and B_n defined by (16) below, are equal in law. Let us define the space

$$W := W_T = \{\text{continuous mappings from } [0,T] \text{ to } \mathbb{R}\}, \tag{10}$$

which, furnished with the sup norm $\| \, . \, \|_W$ defined for all $f \in W$ by $\| f \| = \sup_{x\in[0,T]} |f(x)|$, is a Banach space.

The Proposition 13.20 in [11] states that for all $T > 0$, for some $c > 0$, we have that

$$\mathbf{E}\left[\| \Pi_n B - B \|_W\right] \leq c\, n^{-1/2}, \text{ for all } n \in \mathbb{N}^*. \tag{11}$$

We now estimate the distance between the sample-paths of Birth-and-Death processes and their interpolation. Specifically,

Lemma 3.1. *Let $T > 0$, $n \in \mathbb{N}^*$, and let X be a $\mathbb{N}$-valued Markov jump process on $[0,T]$ of infinitesimal generator $\mathscr{A}$. Suppose that there exist two constants $J \in \mathbb{N}$ and $\alpha > 0$ such that*

- *the magnitude of the jumps of X is bounded by $J > 0$, i.e. for all $i, j \in \mathbb{N}$, $\mathscr{A}(i,j) = 0$ whenever $|j - i| > J$;*
- *the intensities of the jumps of X are bounded by $n\alpha$, i.e. for all $i, j \in \mathbb{N}$, $i \neq j$, $\mathscr{A}(i,j) \leq n\alpha$.*

Then,

$$\mathbf{E}\left[\| X - \Pi_n X \|_W\right] \leq 2J \, \frac{\log n}{\log\log n}.$$

Proof. Fix $n \in \mathbb{N}$ and within this proof, set for $t_i^n = \frac{iT}{n}$ for $i = 0, ..., n$. For any $t \in [0,T]$, for $i \leq n$ such that $t \in \left[t_{i-1}^n, t_i^n\right]$ we have that

$$\begin{aligned} |X(t) - \Pi_n X(t)| &= \left| X(t) - X\left(t_{i-1}^n\right) - \frac{n}{T}\left(t - t_{i-1}^n\right)\left(X\left(t_i^n\right) - X\left(t_{i-1}^n\right)\right)\right| \\ &\leq 2 \sup_{t\in\left[t_{i-1}^n, t_i^n\right]} \left|X(t) - X\left(t_{i-1}^n\right)\right|, \end{aligned}$$

so that

$$\mathbf{E}\left[\| X - \Pi_n X \|_W\right] \leq 2\mathbf{E}\left[\max_{i\in[0,n-1]} \sup_{t\in\left[t_{i-1}^n; t_i^n\right]} \left|X(t) - X\left(t_{i-1}^n\right)\right|\right]. \tag{12}$$

But for any i and any $t \in \left[t_{i-1}^n, t_i^n\right]$, we have that

$$\left|X(t) - X\left(t_{i-1}^n\right)\right| \leq J\left(A_n^i + D_n^i\right),$$

where A_n^i and D_n^i denote respectively the number of up and down jumps of the process X within the interval $\left[t_{i-1}^n, t_i^n\right]$. In turn, by assumption $A_n^i + D_n^i$ is stochastically dominated by a Poisson r.v., say P^i, of parameter $\alpha n \frac{T}{n} = \alpha T$. All in all, we obtain with (12) that

$$\mathbf{E}\left[\| X - \Pi_n X \|_W\right] \leq 2J\, \mathbf{E}\left[\max_{i\in[1,n]} P^i\right],$$

and we conclude using Proposition A.1. □ □

4. A functional Stein method

4.1. **Stein's method in a nutshell.** Say that we want to compare a distribution ν on $\mathbb{R}^n$, $q \geq 1$, to the standard Gaussian distribution on $\mathbb{R}^n$, denoted by μ_n. Consider the processes

$$t \mapsto X(x,t) = e^{-t}x + \sqrt{2}\int_0^t e^{-(t-s)}\ \mathrm{d}B^n(s), \quad x \in \mathbb{R}^n, \tag{13}$$

where B^n is an ordinary Brownian motion in $\mathbb{R}^n$. For all x, it is a Gaussian process whose distribution at time t is a Gaussian law of mean $e^{-t}x$ and covariance matrix $(1-e^{-2t})\mathrm{Id}_n$. For $t \geq 0$, $x \in \mathbb{R}^n$, let

$$Q_t^n f(x) = \mathbf{E}\left[f(X(x,t))\right] = \int_{\mathbb{R}^n} f(e^{-t}x + \beta_t y)\ \mathrm{d}\mu_n(y),$$

where $\beta_t = \sqrt{1-e^{-2t}}$. The dominated convergence theorem entails that

$$Q_t^n f(x) \xrightarrow{t\to\infty} \int_{\mathbb{R}^n} f\ \mathrm{d}\mu_n, \quad x \in \mathbb{R}^n.$$

Moreover, the Dynkin Lemma and the ItÃt formula entail (see [10]) that

$$Q_t^n f(x) - f(x) = \int_0^t A^n Q_s^n f(x)\ \mathrm{d}s, \quad x \in \mathbb{R}^n,\ t \geq 0, \tag{14}$$

where for f regular enough

$$A^n f(x) =: \frac{d}{dt}(Q_t^n f)(x)\Big|_{t=0} = \langle x,\, d_n f(x)\rangle_{\mathbb{R}^n} - \Delta_n f(x).$$

The notation $d_n f$ represents the usual gradient of $f : \mathbb{R}^n \to \mathbb{R}$ and $\Delta_n f$ is its Laplacian. Integrate both sides of (14) with respect to ν to obtain the so-called Stein-Dirichlet representation: for any f in a well chosen functional space $\mathcal{F}$ (i.e. we must at least require that the previous limits do exist and that $A^n Q^n f$ is well defined and integrable for $f \in \mathcal{F}$),

$$\mathrm{d}_{\mathcal{F}}(\nu,\, \mu_n) = \sup_{f\in\mathcal{F}} \int_{\mathbb{R}^n}\int_0^\infty A^n Q_s^n f(x)\ \mathrm{d}s\ \mathrm{d}\nu(x). \tag{15}$$

This formula is the first step of the modern approach to the Stein's method, see [8].

4.2. **Generalization to infinite dimension.** As we mentioned above, the proofs of Theorems 2.1 and 2.2 critically rely on bounding the distance between the affine interpolations of the Markov processes under consideration and their diffusion approximations. For this, we need to go to a functional setup, that is, to bound a similar expression to (15) when the target measure is that of a Gaussian process, instead of a d-dimensional Gaussian random variable. This is done in the main result of this section, Theorem 4.5.

Fix $T > 0$ and an integer $n \geq 1$. Recall (9), and define the following subspace of W,

$$W_n = \mathrm{span}\{h_j^n, j = 1, \cdots, n\},$$

equipped with the sup-norm $\|\ .\ \|_W$. Now define the process

$$B_n = \sum_{j=1}^n Y_j\, h_j^n, \tag{16}$$

where $(Y_j,\, j = 1, \cdots, n)$ is a Gaussian vector of distribution μ_n. Clearly, B_n belongs to W_n with probability 1, thereby defining a Gaussian distribution, denoted by π_n, on W_n. We also need a space to define the gradients. For this, we now consider the space

$$H_n = \text{span}\{h_j^n, j = 1, \cdots, n\}$$

equipped with the scalar product

$$\langle h,\, g\rangle_{H_n} = \int_0^T h'(s)g'(s)\ \mathrm{d}s,\ h, g, \in H_n.$$

Remark 1. *Distinguishing between the spaces H_n and W_n may seem spurious, as these are algebraically the same set, and only differ by their norms. Actually, W_n (respectively, H_n) is the image by the map Π_n defined by (7), of the set W defined by (10) (resp., of the Banach set H - dense in W - that is defined by* (23) *below). An intuitive explanation of our need to introduce the space H is as follows: As mentioned above, the control of the properties of the solution of the Stein equation requires dealing with the derivative of this function. In functional spaces, the usual notion of derivative is replaced by that of FrÃľchet differential: A function F from W into $\mathbb{R}$ is FrÃľchet differentiable whenever for any $w, w' \in W$, the function*

$$\varepsilon \longmapsto F(w + \varepsilon\, w')$$

is differentiable with respect to ε in a neighbor of 0*. For technical reasons, which are detailed in* [7]*, assuming that F is FrÃľchet differentiable in a probabilistic context is too stringent a condition. It turns out that the notion of weak differentiability, i.e. the function*

$$\varepsilon \longmapsto F(w + \varepsilon\, h)$$

is differentiable with respect to ε in a neighbor of 0 *for any $w \in W$ and $h \in H$ is sufficient for what we aim to do, and do not put too strong a constraint on F. Hence the necessity of considering W (the space into which the sample-paths of our processes take place) and H (the set of the admissible directions of differentiation), and thus to distinguish between the spaces W_n and H_n at the level of the interpolated processes.*

The space $H_n^{\otimes(2)}$ is then the vector space

$$H_n^{\otimes(2)} = \text{span}\left\{h_j^n \otimes h_k^n = \Big((s_1, s_2) \longmapsto h_j^n(s_1)h_k^n(s_2)\Big),\, j, k = 1, \cdots, n\right\},$$

equipped with the scalar product: For any $h, g \in H_n^{\otimes(2)}$,

$$\langle h, g\rangle_{H_n^{\otimes(2)}} = \int_0^T \int_0^T \frac{\partial^2 h}{\partial s_1 \partial s_2}(s_1, s_2)\ \frac{\partial^2 g}{\partial s_1 \partial s_2}(s_1, s_2)\ \mathrm{d}s_1\ \mathrm{d}s_2.$$

For a regular enough function $f : W_n \to \mathbb{R}$, we denote by $D_n f$ its differential, i.e. for any $w \in W_n$, for any $h \in H_n$,

$$\langle D_n f(w),\, h\rangle_{H_n} = \frac{d}{d\varepsilon} f(w + \varepsilon h)\Big|_{\varepsilon=0}. \tag{17}$$

We even need to iterate this definition and consider the second order differential, for any $w \in W_n$, for any $h_1, h_2 \in H_n$,

$$\left\langle D_n^{(2)} f(w),\, (h_1, h_2) \right\rangle_{H_n^{\otimes(2)}} = \frac{\partial^2}{\partial\varepsilon_1 \partial\varepsilon_2} f(w + \varepsilon_1 h_1 + \varepsilon_2 h_2)\Big|_{\varepsilon_1=\varepsilon_2=0}. \tag{18}$$

The map

$$T_n : \begin{cases} \mathbb{R}^n & \longrightarrow W_n \\ (y_1, \cdots, y_n) & \longmapsto \sum_{j=1}^{n} y_j\, h_j^n, \end{cases}$$

is a morphism of probability spaces, i.e. it is linear, continuous and preserves the probability measures: the image measure of μ_n by T_n is actually π_n. Then, we can generalize the construction we just followed on $\mathbb{R}^n$ to the finite dimensional space W_n. The family of maps $(P_t^n,\, t \geq 0)$ is defined as follows: $P_0^n = \mathrm{Id}$ and for all $t > 0$,

$$P_t^n : \begin{cases} L^1(\pi_n) & \longrightarrow L^1(\pi_n) \\ f & \longmapsto \left(w \longmapsto P_t^n f(w) = \displaystyle\int_{W_n} f(e^{-t} w + \beta_t \zeta)\, \mathrm{d}\pi_n(\zeta) \right) \end{cases}$$

Since T_n is linear and since π_n is the image of μ_n by T_n, we easily see that for any $t \geq 0$

$$Q_t^n (f \circ T_n) = (P_t^n f) \circ T_n,$$

which can be written

$$P_t^n f = Q_t^n (f \circ T_n) \circ T_n^{-1}. \tag{19}$$

Thus, $(P_t^n,\, t \geq 0)$ is a semi-group such that

$$P_t^n f(w) \xrightarrow{t\to\infty} \int_{W_n} f\, \mathrm{d}\pi_n, \quad w \in W_n.$$

From (19), we also infer that for $f : W_n \to \mathbb{R}$ twice differentiable,

$$L^n f =: \frac{d}{dt}(P_t^n f)\Big|_{t=0} = A^n (f \circ T_n) \circ T_n^{-1}.$$

Hence, we have that

$$P_t^n f(w) - f^n(w) = \int_0^t L^n P_s^n f(w)\, \mathrm{d}s, \quad w \in W_n,\, t \geq 0, \tag{20}$$

where for f regular enough

$$L^n f(w) = \langle D_n f(w),\, w \rangle_{H_n} - \sum_{j=1}^{n} \left\langle D^{(2)} f(w),\, h_j^n \otimes h_j^n \right\rangle_{H_n^{\otimes(2)}}.$$

Thus, for any measure ν_n on W_n,

$$\mathrm{d}_{\mathcal{F}_n}(\nu_n,\, \pi_n) = \sup_{f \in \mathcal{F}_n} \int_{W_n} \int_0^\infty L^n P_s^n f(w)\, \mathrm{d}s\, \mathrm{d}\nu_n(w), \tag{21}$$

where $\mathcal{F}_n$ is a space of regular enough test functions from W_n into $\mathbb{R}$. We can now precise which kind of test functions we are going to consider. In view of (21), it must contains twice differentiable functions but for technical reasons, we need more than that.

Definition 4.1. *A function $f : W_n \to \mathbb{R}$ is said to belong to the class Σ_n whenever it is 1-Lipschitz continuous, twice differentiable in the sense of* (18)*, and we have*

(22)
$$\sup_{w\in W_n}\left|\left\langle D_n^{(2)}f_n(w) - D_n^{(2)}f_n(w+g),\, h\otimes k\right\rangle_{H_n^{\otimes(2)}}\right| \leq \|g\|_W\,\|h\|_{L^2}\|k\|_{L^2},$$

for any $g \in W_n,\ h,k \in H_n$.

Actually, in the definition of the distance between distributions of processes, the test functions are defined on the whole space W. Hence, we must find a class of functions whose restriction to W_n belong to Σ_n for any $n \geq 1$. This involves the notion of H-differential on W. Let

$$H = \left\{h,\ \exists! h' \in L^2([0,T]) \text{ such that } h(t) = \int_0^t h'(s)\ \mathrm{d}s\right\}. \tag{23}$$

It is an Hilbert space when equipped with the scalar product

$$\langle h, g\rangle_H = \int_0^T h'(s)g'(s)\ \mathrm{d}s.$$

A function $f : W \to \mathbb{R}$ is said to be twice H-differentiable whenever for any $w \in W$, for any $h \in H$, the function

$$\left\{\begin{array}{ll}\mathbb{R} & \longrightarrow \mathbb{R}\\ \varepsilon & \longmapsto f(w+\varepsilon h)\end{array}\right.$$

is twice differentiable in a neighbor of 0. We denote by Df and $D^{(2)}f$ its first and second order gradient, defined by

$$\begin{aligned}\langle Df(x),\, h\rangle_H &= \frac{d}{d\varepsilon}f(x+\varepsilon h)\Big|_{\varepsilon=0},\\ \left\langle D^{(2)}f(x),\, h_1\otimes h_2\right\rangle_{H^{\otimes(2)}} &= \frac{\partial^2}{\partial\varepsilon_1\partial\varepsilon_2}f(w+\varepsilon_1 h_1+\varepsilon_2 h_2)\Big|_{\varepsilon_1=\varepsilon_2=0}.\end{aligned}$$

Definition 4.2. *The class Σ is the set of 1-Lipschitz continuous, twice H-differentiable functions such that*

$$\sup_{w\in W}\left|\left\langle D^{(2)}f(w) - D^{(2)}f(w+g),\, h\otimes k\right\rangle_{H^{\otimes(2)}}\right| \leq \|g\|_W\,\|h\|_{L^2}\|k\|_{L^2},$$

for any $g \in W,\ h,k \in H$.

For $f : W \to \mathbb{R}$, let $f_n = f_{|W_n}$. If f is once H-differentiable, then, we have that for any $w_n \in W_n$, any $j \in \{0,\cdots,n-1\}$,

$$\begin{aligned}\left\langle Df(\mathfrak{e}(w_n)),\, h_n^j\right\rangle_H = \frac{d}{dt}f(\mathfrak{e}(w_n+\varepsilon h_n^j))\Big|_{\varepsilon=0} &= \frac{d}{dt}f_n(w_n+\varepsilon h_n^j)\Big|_{\varepsilon=0}\\ &= \left\langle D_n f_n(w_n),\, h_n^j\right\rangle_{H_n}.\end{aligned} \tag{24}$$

Thus, it is straightforward that if f belongs to Σ then f_n belongs to Σ_n for any $n \geq 1$.

Remark 2. *We can now show how to prove that the functionals mentioned in the introduction do belong to Σ. Consider the first one :*

$$F_f(x) = \frac{1}{T}\int_0^T f(x_s)\ ds.$$

Then, for any $x, y \in W$,

$$|F_f(x) - F_f(y)| \leq \|x - y\|_W$$

provided that f is Lipschitz continuous. Moreover, a classical computation shows that

$$\left\langle D^{(2)}F_f(x+g) - D^{(2)}f(x),\, h\otimes k\right\rangle_{H^{\otimes(2)}} = \frac{1}{T}\int_0^T (f''(x_s+g(s)) - f''(x_s))h_s k_s\ ds.$$

Hence F_f belongs to Σ as long as f'' does exist and is Lipschitz continuous. The other cases are handled similarly.

4.3. **Functionals of Poisson marked point processes.** Let N_ν be a marked point process on $E = [0,T]\times\mathbb{R}^+$ whose jump times are denoted by $(T_n, n \geq 1)$, and jumps magnitude by $(Z_n, n \geq 1)$. It is said to be a Poisson marked point process of (diffuse) control measure ν whenever for any function $u = (u(s,z),\ s\in[0,T],\ z\in\mathbb{R}^+)$ in $L^2(\nu)$, the process

$$t \longmapsto (\nabla_\nu^* u)(t) = \sum_{T_n\leq t} u(T_n, Z_n) - \int_0^t\int_{\mathbb{R}^+} u(s,z)\ \mathrm{d}\nu(s,z)$$

is a square integrable martingale. We set

$$\nabla_\nu^* u = (\nabla_\nu^* u)(T). \tag{25}$$

Consider the so-called discrete gradient [9, 14],

$$\nabla_{s,z} f(N_\nu) = f(N_\nu + \varepsilon_{s,z}) - f(N_\nu),\ s\in[0,T],\ z\in\mathbb{R}^+,$$

where $N_\nu + \varepsilon_{s,z}$ represents the sample-path N_ν to which we add an atom at time s of size z. Since ν is diffuse, there is a zero probability that an atom at time s already exists in N_ν. Similarly, we denote by $N_\nu - \varepsilon_{s,z}$ the sample-path N_ν to which we remove the atom $\varepsilon_{s,z}$ provided it is present in N_ν, otherwise N_ν remains unchanged.

Definition 4.3. *We define the domain of ∇ as*

$$\mathrm{dom}\,\nabla = \left\{ f,\ \mathbf{E}\left[\int_{[0,T]\times\mathbb{R}^+} |\nabla_{s,z} f(N_\nu)|^2\ d\nu(s,z)\right] < \infty, \right\}.$$

We then have the integration by parts formula [9]:

Lemma 4.4. *For $u \in L^2(\nu)$, for $f \in \mathrm{dom}\,\nabla$, we have that*

$$\mathbf{E}\left[f(N_\nu)\ \nabla_\nu^* u\right] = \mathbf{E}\left[\int_{[0,T]\times\mathbb{R}^+} \nabla_{s,z} f(N_\nu)\, u(s,z)\ d\nu(s,z)\right]. \tag{26}$$

For the sake of completeness, we reproduce the proof of this identity, which is a mere rewriting of the Campbell-Mecke formula for Poisson processes.

Proof. By the very definition of ∇,

$$\begin{aligned}(27)\quad &\mathbf{E}\left[\int_E \nabla_{s,z} f(N_\nu)\, u(s,z)\ \mathrm{d}\nu(s,\, z)\right] \\ &= \mathbf{E}\left[\int_E f(N_\nu + \varepsilon_{s,z}) u(s,z)\ \mathrm{d}\nu(s,\, z)\right] - \mathbf{E}\left[\int_E f(N_\nu) u(s,z)\ \mathrm{d}\nu(s,\, z)\right].\end{aligned}$$

The Campbell-Mecke formula for Poisson processes says that

$$(28)\quad \mathbf{E}\left[\int_E f(N_\nu + \varepsilon_{s,z}) u(s,z)\ \mathrm{d}\nu(s,\, z)\right] \;=\; \mathbf{E}\left[f(N_\nu) \sum_{T_n \le T} u(T_n, Z_n)\right].$$

Plug (28) into the right-hand-side of (27) to obtain (26). □ □

Remark 3. *If we have an unmarked Poisson process of intensity $d\nu(s) = \nu\ ds$, then (26) still holds by suppressing all occurrences of the z variable.*

We are now equipped to prove the cornerstone theorem of our paper. For u_j^n, $j = 1, \cdots, n$ a family of elements of $L^2([0,T] \times \mathbb{R}^+, \nu)$, set

$$u^n(s,z,t) = \sum_{j=1}^n u_j^n(s,z)\, h_j^n(t) \text{ and } \nabla_\nu^* u^n(t) = \sum_{j=1}^n \nabla_\nu^*(u_j^n)\ h_j^n(t).$$

For any $j \in \{1, \cdots, n\}$, let

$$(29)\qquad \xi_{j,n}^2 = \int_0^T \int_{\mathbb{R}^+} u_j^n(s,z)^2\ \mathrm{d}\nu(s,z)$$

and consider

$$\Gamma_{\xi_n} = \mathrm{diag}(\xi_{j,n}^2,\, j = 1, \cdots, n).$$

Furthermore, take $Y = (Y_j,\, j \ge 1)$ a family of independent standard Gaussian random variables and let

$$(30)\qquad B_{\xi_n}(t) = \sum_{j=1}^n \xi_{j,n}\ Y_j\, h_j^n(t).$$

Theorem 4.5. *Assume that $(u_j^n,\, j = 1, \cdots, n)$ is an orthogonal family of elements of $L^2(\nu)$. Then, for any $f_n \in \Sigma_n$,*

$$\begin{aligned}&\left|\mathbf{E}\left[f_n(B_{\xi_n})\right] - \mathbf{E}\left[f_n(\nabla_\nu^* u^n)\right]\right| \\ &\le \frac{n^{-3/2} T^2}{4} \sum_{j,k,l=1}^n \xi_{j,n}\xi_{k,n}\xi_{l,n} \int_{[0,T]\times\mathbb{R}^+} \left|u_j^n(s,z) u_k^n(s,z)\right| \left|u_l^n(s,z)\right|\ d\nu(s,z),\end{aligned}$$

where ∇_ν^* is defined by (25).

Proof. For the sake of notational simpicity, we remove the suffix n as it is fixed along the proof. Note that in view of (24), there is no ambiguity to denote D_n as D since they coincide on W_n. To shorten the equations, E stands for $[0,T] \times \mathbb{R}^+$ and $x = (s,z)$ is a generic point of E.

Dividing each u_j^n by $\xi_{j,n}$, $j \geq 1$, it is sufficient to prove the result for $\xi_{j,n} = 1$, $j \geq 1$. Now recall (16). First, in view of (20),

$$\mathbf{E}\left[f(B_n)\right] - \mathbf{E}\left[f(\nabla_\nu^* u)\right] = -\sum_{j=1}^n \int_0^\infty \mathbf{E}\left[\nabla_\nu^* u_j \left\langle DP_t f(\nabla_\nu^* u)\,,\, h_j\right\rangle_H\right]\, \mathrm{d}t \tag{31}$$
$$+ \sum_{j=1}^n \int_0^\infty \mathbf{E}\left[\left\langle D^{(2)} P_t f(\nabla_\nu^* u)\,,\, h_j \otimes h_j\right\rangle_{H^{\otimes(2)}}\right]\, \mathrm{d}t.$$

According to the integration by parts formula (26) and to the fundamental theorem of calculus, we get that

$$\sum_{j=1}^n \mathbf{E}\left[\nabla_\nu^* u_j \left\langle h_j,\ DP_t f(\nabla_\nu^* u)\right\rangle_H\right]$$
$$= \sum_{j=1}^n \mathbf{E}\left[\int_E u_j(x)\, \left\langle DP_t f(\nabla_\nu^* u + u(x)) - DP_t f(\nabla_\nu^* u)\,,\, h_j\right\rangle_H\, \mathrm{d}\nu(x)\right]$$
$$= \sum_{j,k=1}^n \mathbf{E}\left[\int_E \int_0^1 u_j(x) u_k(x) \left\langle D^{(2)} P_t f(\nabla_\nu^* u + r\, u(x)\,,\, h_j \otimes h_k)\right\rangle_{H^{\otimes(2)}} \mathrm{d}r\ \mathrm{d}\nu(x)\right].$$

But as the u_k's are orthonormal,

$$\mathbf{E}\left[\sum_{j=1}^n \left\langle D^{(2)} P_t f(\nabla_\nu^* u)\,,\, h_j \otimes h_j\right\rangle_{H^{\otimes(2)}}\right]$$
$$= \mathbf{E}\left[\sum_{j,k=1}^n \int_E \int_0^1 u_j(x) u_k(x) \left\langle D^{(2)} P_t f(\nabla_\nu^* u)\,,\, h_j \otimes h_k\right\rangle_H\, \mathrm{d}r\ \mathrm{d}\nu(x)\right].$$

Since f belongs to Σ_n, the right-hand-side of (31) becomes

$$\sum_{j,k=1}^n \int_0^\infty \int_E \int_0^1 \mathbf{E}\left[\left\langle D^{(2)} P_t f(\nabla_\nu^* u + ru(x)) - D^{(2)} P_t f(\nabla_\nu^* u)\,,\, h_j \otimes h_k\right\rangle_{H^{\otimes(2)}}\right]$$
$$\times\, u_j(x) u_k(x)\, \mathrm{d}r\ \mathrm{d}\nu(x)\ \mathrm{d}t$$
$$\leq \sum_{j,k=1}^n \|h_j\|_{L^2} \|h_k\|_{L^2} \int_E \|u(x)\|_W |u_j(x)\, u_k(x)|\ \mathrm{d}\nu(x) \left(\int_0^T r\ \mathrm{d}r\right)\left(\int_0^\infty e^{-2t}\ \mathrm{d}t\right).$$

Observing that

$$\|u(x)\|_W \leq \sum_{l=1}^n |u_l(x)|\, \|h_l\|_W = n^{-1/2} \sum_{l=1}^n |u_l(x)|,$$

the result follows by recalling that $\|h_j\|_{L^2} \leq n^{-1/2}$ for all $j \in \{1, \cdots, n\}$. □ □

5. Proof of Theorem 2.1

We now turn to the proof of Theorem 2.1. Fix $T \leq \frac{x}{\mu - \lambda}$. Then for all $n \in \mathbb{N}^*$ we readily have that

$$\mathrm{d}_\Sigma\left(Z_n^\dagger, B\right) \leq \mathrm{d}_\Sigma\left(Z_n^\dagger, \Pi_n Z_n^\dagger\right) + \mathrm{d}_\Sigma(\Pi_n Z_n^\dagger, \Pi_n B) + \mathrm{d}_\Sigma(\Pi_n B, B). \tag{32}$$

First observe that the function $\overline{L^\dagger}$ is affine, and hence coincides with $\Pi_n\overline{L^\dagger}$ on $[0,T]$. Moreover, the operator Π_n is linear and the elements of Σ are 1-Lipschitz-continuous, thus we have that for all n,

$$\begin{aligned}\mathrm{d}_\Sigma(Z_n^\dagger, \Pi_n Z_n^\dagger) \leq \mathbf{E}\left[\parallel Z_n^\dagger - \Pi_n Z_n^\dagger \parallel_W\right] &\leq \frac{1}{\sqrt{n(\lambda+\mu)}}\mathbf{E}\left[\parallel L_n^\dagger - \Pi_n L_n^\dagger \parallel_W\right] \\ &\leq \frac{c\log n}{\log\log n\,\sqrt{n}}, \end{aligned} \tag{33}$$

where the last inequality follows from applying Lemma 3.1 to the Markov processes $\left(L_n^\dagger : n \geq 1\right)$ for $J \equiv 1$ and $\alpha \equiv \lambda \vee \mu$. Now, for any $n \in \mathbb{N}^*$, if we let $\tau_0^n = \inf\{t > 0,\, L_n^\dagger(t) = 0\}$, for any $F \in \Sigma$ we have that

$$\begin{aligned}&\mathbf{E}\left[\left|F\left(\Pi_n Z_n^\dagger\right) - F\left(\Pi_n B\right)\right|\right] \\ &\quad= \mathbf{E}\left[\left|F\left(\Pi_n Z_n^\dagger\right) - F\left(\Pi_n B\right)\right| \mathbf{1}_{\{T<\tau_0^n\}}\right] \\ &\qquad\qquad + \mathbf{E}\left[\left|F\left(\Pi_n Z_n^\dagger\right) - F\left(\Pi_n B\right)\right| \mathbf{1}_{\{T\geq\tau_0^n\}}\right].\end{aligned} \tag{34}$$

We first prove that for some $c > 0$,

$$\mathbf{E}\left[\left|F\left(\Pi_n Z_n^\dagger\right) - F\left(\Pi_n B\right)\right| \mathbf{1}_{\{T<\tau_0^n\}}\right] \leq \frac{c}{\sqrt{n}},\ n \in \mathbb{N}^*. \tag{35}$$

Fix $n \in \mathbb{N}^*$. On the event $\{T < \tau_0^n\}$, for any $t \in [0,T)$ we have that

$$\begin{aligned}Z_n^\dagger(t) &= \frac{1}{\sqrt{\lambda+\mu}}\left(\sqrt{\lambda}\left(\frac{N_{n\lambda}(t)}{\sqrt{\lambda n}} - \sqrt{\lambda n}t\right) - \sqrt{\mu}\left(\frac{N_{n\mu}(t)}{\sqrt{\mu n}} - \sqrt{\mu n}t\right)\right) \\ &=: \frac{1}{\sqrt{\lambda+\mu}}\left(Z_{\lambda,n}^\dagger(t) - Z_{\mu,n}^\dagger(t)\right).\end{aligned}$$

To apply Theorem 4.5, it is useful to represent the processes $Z_n^\dagger$, $n \geq 1$ as marked Poisson processes. For this, we fix $n \in \mathbb{N}^*$, and let $N_{n(\lambda+\mu)}^\dagger$ be the marked Poisson point process on $[0,T] \times \{-1,1\}$ of control measure

$$\mathrm{d}\nu_n^\dagger(s,r) = n(\lambda+\mu)\ \mathrm{d}s \otimes \Big(\frac{\lambda}{\lambda+\mu}\,\varepsilon_1(\,\mathrm{d}r) + \frac{\mu}{\lambda+\mu}\,\varepsilon_{-1}(\,\mathrm{d}r)\Big),$$

that is, an ordinary Poisson process on the positive half-line with intensity $n(\lambda+\mu)$, such that each atom is assigned a mark $+1$ or -1, independently of everything else, with respective probability $\lambda(\lambda+\mu)^{-1}$ and $\mu(\lambda+\mu)^{-1}$. By the thinning property of Poisson processes, the point process counting the atoms of $N_{n(\lambda+\mu)}^\dagger$ with mark $+1$ (respectively -1) is Poisson of intensity $n\lambda$ (respectively $n\mu$). For any $t \in [0,T]$, let

$$v_t\ :\ \begin{cases}[0,T]\times\{-1,1\} & \longrightarrow \mathbb{R} \\ (s,r) & \longmapsto \frac{1}{\sqrt{n(\lambda+\mu)}}\ r\,\mathbf{1}_{[0,t)}(s),\end{cases}$$

and define for all $i = 1, \cdots, n$,

$$u_i^\dagger(s,r) = \sqrt{\frac{n}{T}}\Big(v_{t_i^n}(s,r) - v_{t_{i-1}^n}(s,r)\Big) = \frac{1}{\sqrt{T(\lambda+\mu)}}\ r\,\mathbf{1}_{\left[t_{i-1}^n, t_i^n\right)}(s).$$

Then, it is easily checked that

$$Z_n^\dagger(t) \overset{\text{dist}}{=} \Delta^*_{\nu_n^\dagger} v_t,\ t \leq T,$$

which, recalling (8) and (9), yields to

$$\Pi_n Z^\dagger_{\lambda,n} \overset{\text{dist}}{=} \sum_{i=1}^n \Delta^*_{\nu_n^\dagger} u_i^\dagger \; h_i^n.$$

It is then clear that for all $i, j \leq n$,

$$\int_{[0,T]\times\{-1,1\}} u_i^\dagger(s,r) u_j^\dagger(s,r) \; \mathrm{d}\nu_n^\dagger(s,r) = \delta_{ij},$$

so $\{u_i^\dagger,\, i = 1, \cdots, n\}$ is an orthogonal family. Moreover, comparing (8) to (30), we readily obtain that $\Pi_n B \overset{\text{dist}}{=} B_{\xi^\dagger}$ when letting $\xi^\dagger_{j,n} = 1$ for all $j = 1, \cdots, n$. Consequently, (35) follows from Theorem 4.5 and the fact that

$$\sum_{j,k,l=1}^n \int_E |u_j^\dagger u_k^\dagger u_l^\dagger| \; \mathrm{d}\nu_n^\dagger = \frac{1}{T^{3/2}\sqrt{\lambda+\mu}} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} n \; \mathrm{d}s = \frac{n}{\sqrt{T(\lambda+\mu)}}.$$

Regarding the second term on the right-hand side of (34), observe that F is in particular bounded, so there exists a constant c' such that for all $n \in \mathbb{N}^*$,

$$\mathbf{E}\left[\left|F(\Pi_n Z_n^\dagger) - F(\Pi_n B)\right| \mathbf{1}_{\{T>\tau_0^n\}}\right] \leq c\, \mathbf{P}\left[T > \tau_0^n\right].$$

But $\mathbf{P}\left[T > \tau_0^n\right]$ tends to 0 with exponential speed from Theorem 11.9 of [17]: if $\rho < 1$, for any $x > 0$ and any $y < 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left[\tau_0^n \leq \frac{x}{\lambda-\mu} + y\right] = -f(y),$$

where f is strictly positive on $(0,\infty)$. This shows that for some c'',

$$\mathbf{E}\left[\left|F(Z_n^\dagger) - F(B)\right| \mathbf{1}_{\{T>\tau_0^n\}}\right] \leq c'' e^{-n}$$

for all n which, together with (35) in (34), shows that for some constant c, for all $n \in \mathbb{N}^*$,

$$\mathrm{d}_\Sigma(\Pi_n Z_n^\dagger, \Pi_n B) \leq \frac{c}{\sqrt{n}}.$$

This, together with (33) and (11) in (32), concludes the proof.

6. Proof of Theorem 2.2

We now turn to the speed of convergence in the diffusion approximation of the infinite server queue. Fix $T > 0$ throughout this section.

6.1. **An integral transformation.** We know from eq. (6.23) of [15] that the sequence of processes $\left(Y_n^\sharp :\, n \geq 1\right)$ defined for all $n \geq 1$ by

$$t \mapsto Y_n^\sharp(t) := Z_n^\sharp(t) - Z_n^\sharp(0) + \mu \int_0^t Z_n^\sharp(s) \; \mathrm{d}s \tag{36}$$

converges in distribution to the time-changed standard brownian motion $B \circ \gamma$, where

$$\gamma(t) = 2\lambda t - \frac{\lambda}{\mu}(1 - e^{-\mu t}), \quad t \geq 0. \tag{37}$$

This integral transformation of the processes $Z_n^\sharp$, $n \geq 1$ will turn out to be useful to bound the rate of convergence of $\{Z_n^\sharp\}$ to the Ornstein-Uhlenbeck process $Z^\sharp$ defined by (6). Specifically, as will be shown below, the latter

rate of convergence is in fact bounded by that of $\{Y_n^\sharp\}$ to the time-changed brownian motion $B \circ \gamma$. First observe that

Proposition 6.1. *The mapping*

$$\Theta : \begin{cases} \mathbb{D} & \longrightarrow \mathbb{R} \times \mathbb{D}_T^0 \\ f & \longmapsto \left(f(0)\,,\, f(.) - f(0) + \mu \int_0^. f(s)ds \right) \end{cases}$$

is linear, continuous (for the Skorohod topology on $\mathbb{D}$), and one to one.

Proof. Let us fix $\eta \in \mathbb{D}_T^0$ and consider the following integral equation of unknown function z,

$$z(t) - z(0) = -\mu \int_0^t z(s)\ \mathrm{d}s + \eta(t).$$

We clearly have for all $t \geq 0$,

$$z(t) = z(0)e^{-\mu t} + \eta(t) - \mu \int_0^t e^{-\tau(t-s)} \eta(s)\ \mathrm{d}s,$$

hence Θ is bijective and for all $(x, \eta) \in \mathbb{R} \times \mathbb{D}_T^0$,

$$\Theta^{-1}(x,\ \eta) = \left(t \longmapsto x e^{-\mu t} + \eta(t) - \mu \int_0^t e^{-\mu(t-s)} \eta(s)\ \mathrm{d}s \right). \tag{38}$$

Linearity and continuity are then straightforward. □ □

Also,

Lemma 6.2. *On the subset of $\{0\} \times \Theta(\mathbb{D})$ whose image by Θ^{-1} is in $\mathbb{D}$, Θ^{-1} is linear and continuous.*

Proof. For all η, $\omega \in \Theta(\mathbb{D})$ and all $t \leq T$, we have that

$$\Theta^{-1}(0,\eta)(t) - \Theta^{-1}(0,\omega)(t) = \eta(t) - \omega(t) - \mu \int_0^t e^{-\mu(t-s)} (\eta(s) - \omega(s))\ \mathrm{d}s.$$

Hence, by an immediate change of variable we get that

$$\| \Theta^{-1}(0,\eta) - \Theta^{-1}(0,\omega) \|_W < \| \eta - \omega \|_W + \mu \| \eta - \omega \|_W \sqrt{\int_0^T e^{-2\mu s}\ \mathrm{d}s},$$

so that for some positive constant k,

$$\| \Theta^{-1}(x,\eta) - \Theta^{-1}(y,\omega) \|_W < k \| \eta - \omega \|_W .$$

This completes the proof. □ □

We obtain the following,

Corollary 6.3. *These exists a positive constant c such that that for all $n \in \mathbb{N}^*$,*

$$\mathrm{d}_\Sigma(\Pi_n Z_n^\sharp, Z^\sharp) \leq c\ \mathrm{d}_\Sigma(\Pi_n Y_n^\sharp, B \circ \gamma).$$

Proof. In view of the weak convergence $Z_n^\sharp \Rightarrow B \circ \gamma$, the linearity and continuity of Θ and the Continuous Mapping Theorem, we have the weak convergence

$$\Theta(Z_n^\sharp) = (0, Y_n^\sharp) \Rightarrow (0, B \circ \gamma)\,.$$

However, expression (6.34) in [15] shows that for all t, $\Theta\left(Z^\sharp\right) = (0, B \circ \gamma)$ which, together with the linearity of Θ and of the operator Π_n for all n, concludes the proof. □

6.2. **Alternative representation.** With Corollary 6.3 in hand, we are rendered to assess the rate of convergence of $\left(Y_n^\sharp : n \geq 0\right)$ to the time-changed brownian motion $B \circ \gamma$. For that purpose, we aim at applying again Theorem 4.5 and, as above, it is useful for this to view the processes $L_n^\sharp$, $n \geq 1$ as simple functions of marked Poisson processes.

Specifically, following Section 7.2 of [15], we have the following alternative representation of the process $L^\sharp$: A point (x, z) represents a customer arriving at time x and requiring a service of duration z, and we let $N_{\lambda,\mu}$ be a Poisson process on $\mathbb{R}^+ \times \mathbb{R}^+$ of control measure $\lambda\, \mathrm{d}x \otimes \mu e^{-\mu z}\, \mathrm{d}z$. At any time $t \geq 0$, the number of busy servers at t equals the number of points located in the shaded trapeze bounded by the axes of equation $x = 0$ and $x = t$, and above the line $z = t - x$: in other words,

$$L^\sharp(t) = \int_{C_t} \mathrm{d}N_{\lambda,\mu}(x, z),\ t \geq 0,$$

where

$$C_t = \{(x, z), 0 \leq x \leq t, z \geq t - x\}. \tag{39}$$

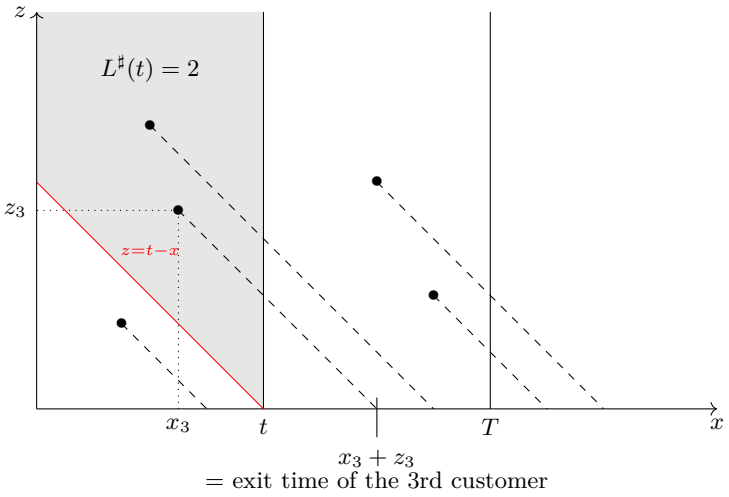

FIGURE 1. Representation of the $M/M/\infty$ queue

Fix a positive integer n throughout this section. After scaling, for all $t \geq 0$ we get that

$$\overline{L_n^\sharp}(t) = \frac{1}{n} N_{\lambda n,\mu}(C_t).$$

Let us denote for all (x,z) in the positive orthant by

$$\mathrm{d}\nu_n^\sharp(x,z) := \lambda n \ \mathrm{d}x \otimes \mu e^{-\mu z} \ \mathrm{d}z,$$

the control measure of $N_{n\lambda,\mu}$. As readily follows from (4), the fluid limit $\overline{L^\sharp}$ can be written as

$$\overline{L^\sharp}(t) = \frac{1}{n} \int \mathbf{1}_{C_t}(x,z) \ \mathrm{d}\nu_n^\sharp(x,z), \ t \geq 0,$$

in a way that

$$Z_n^\sharp(t) = \frac{1}{\sqrt{n}} \int \mathbf{1}_{C_t} \left(\mathrm{d}N_{\lambda n,\mu} - \ \mathrm{d}\nu_n^\sharp \right), \ t \geq 0, \tag{40}$$

for C_t defined by (39). We deduce that for all $t \geq 0$,

(41)

$$\begin{aligned} Y_n^\sharp(t) = \ & \frac{1}{\sqrt{n}} \int \mathbf{1}_{C_t} \left(\mathrm{d}N_{\lambda n,\mu} - \ \mathrm{d}\nu_n^\sharp \right) + \mu \int_0^t \frac{1}{\sqrt{n}} \int \mathbf{1}_{C_s} \left(\mathrm{d}N_{\lambda n,\mu} - \ \mathrm{d}\nu_n^\sharp \right) \mathrm{d}u \\ & = \frac{1}{\sqrt{n}} \nabla^*_{\lambda n,\mu}(\mathbf{1}_{C_t}) + \mu \int_0^t \frac{1}{\sqrt{n}} \nabla^*_{\lambda n,\mu}(\mathbf{1}_{C_u}) \ \mathrm{d}u, \end{aligned}$$

where $\nabla^*_{\lambda n,\mu}$ is defined by (25).

6.3. **Reduction to the finite dimension.** Fix $n \in \mathbb{N}^*$ and recall (8). It follows from (41) that

$$\begin{aligned} \Pi_n Y_n^\sharp &= \sum_{i=1}^n \frac{1}{\sqrt{T}} \left(\nabla^*_{\lambda n,\mu} \left(\mathbf{1}_{C_{t_i^n}} - \mathbf{1}_{C_{t_{i-1}^n}} \right) + \mu \int_{t_{i-1}^n}^{t_i^n} \nabla^*_{\lambda n,\mu}(\mathbf{1}_{C_u}) \ \mathrm{d}u \right) h_i^n \\ &= \sum_{i=1}^n \nabla^*_{\lambda n,\mu}(u_i^\sharp) \, h_i^n, \end{aligned}$$

where for all $i = 1, \cdots, n$ and all $(x,z) \in \mathbb{R}^2$,

$$u_i^\sharp(x,z) = \frac{1}{\sqrt{T}} \left(\mathbf{1}_{C_{t_i^n}}(x,z) - \mathbf{1}_{C_{t_{i-1}^n}}(x,z) + \mu \int_{t_{i-1}^n}^{t_i^n} \mathbf{1}_{C_u}(x,z) \ \mathrm{d}u \right). \tag{42}$$

Let us denote for any $i = 1, \cdots, n$,

$$\xi_{i,n}^\sharp := \sqrt{\gamma\left(t_i^n\right) - \gamma\left(t_{i-1}^n\right)}.$$

The following result is proven in appendix B,

Proposition 6.4. *For any n, the family $\left(u_i^\sharp, \ i = 1, \cdots, n\right)$ has the following properties:*

(i) It is orthogonal in $L^2\left(\nu_n^\sharp\right)$;

(ii) For some constant c independent of n,

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \int_E |u_i^\sharp u_j^\sharp u_k^\sharp| \ d\nu_n^\sharp \leq nc.$$

(iii) For any $i \in \{1, \cdots, n\}$,

$$\int \int u_i^\sharp u_i^\sharp \ d\nu_n^\sharp = \frac{n}{T} \left(\xi_{i,n}^\sharp\right)^2.$$

Notice that for a large enough n, for all $t \geq 0$,

$$\frac{n}{t}\left(\xi_{i,n}^{\sharp}\right)^2 \xrightarrow{i/n \to t} \gamma'(t) \text{ and for a fixed } i,\ \frac{n}{t}\left(\xi_{i,n}^{\sharp}\right)^2 \xrightarrow{n\to\infty} \gamma'(0).$$

We thus have the following result,

Proposition 6.5. *For some c, for all positive integer n, the respective interpolations of $Y_n^{\sharp}$ and $B \circ \gamma$ satisfy*

$$\mathrm{d}_{\Sigma}(\Pi_n Y_n^{\sharp}, \Pi_n(B \circ \gamma)) \leq \frac{c}{\sqrt{n}}.$$

Proof. Fix $n \in \mathbb{N}^*$. It is an immediate consequence of (8) and (30) that

$$\pi_n(B \circ \gamma) \overset{\text{dist}}{=} \sum_{j=1}^{n} Y_j^{\sharp}\, h_j^n = B_{\xi^{\sharp}},$$

where $\left(Y_k^{\sharp},\, k = 1, \cdots, n\right)$ is a family of independent centered Gaussian random variables such that $\mathrm{var}(Y_k^{\sharp}) = (\xi_k^{\sharp})^2$ for all k. From assertion (i) of Proposition 6.4, we can apply Theorem 4.5 : for any $f \in \Sigma$,

$$\left|\mathbf{E}\left[f(B_{\xi^{\sharp}})\right] - \mathbf{E}\left[f(\nabla^*_{\lambda n,\mu}(u^{\sharp}))\right]\right| \leq \frac{n^{-3/2}T^2}{4} \sum_{j,k,l=1}^{n} \int_E |u_j^{\sharp} u_k^{\sharp}|\, |u_l^{\sharp}|\, \mathrm{d}\nu_n^{\sharp}.$$

Assertions (ii) and (iii) of Proposition 6.4 allow us to conclude. □

6.4. **Proof of Theorem 2.2.** We are now in position to prove Theorem 2.2. For all $n \in \mathbb{N}^*$,. We have that

$$\begin{aligned}(43)\quad & \mathrm{d}_{\Sigma}(Z_n^{\sharp}, Z^{\sharp}) \\ & \leq \mathrm{d}_{\Sigma}(Z_n^{\sharp}, \Pi_n Z_n^{\sharp}) + \mathrm{d}_{\Sigma}(\Pi_n Z_n^{\sharp}, Z^{\sharp}) \\ & \leq \mathrm{d}_{\Sigma}(Z_n^{\sharp}, \Pi_n Z_n^{\sharp}) + c\,\mathrm{d}_{\Sigma}(\Pi_n Y_n^{\sharp}, B \circ \gamma) \\ & \leq \mathrm{d}_{\Sigma}(Z_n^{\sharp}, \Pi_n Z_n^{\sharp}) + c\,\mathrm{d}_{\Sigma}(\Pi_n Y_n^{\sharp}, \Pi_n B \circ \gamma) + c\,\mathrm{d}_{\Sigma}(\Pi_n B \circ \gamma, B \circ \gamma),\end{aligned}$$

where we applied Corollary 6.3 in the second inequality. Now define the stopping times

$$\tau_n^{\sharp} = \inf\left\{t \geq 0 \,:\, N_{n\lambda}(t) \geq 2\lambda n T\right\},\ n \in \mathbb{N}^*.$$

Then, as all functions of Σ are bounded and Lipschitz continuous we obtain that for all n,

$$\begin{aligned}(44)\quad & \mathrm{d}_{\Sigma}\left(Z_n^{\sharp}, \Pi_n Z_n^{\sharp}\right) \\ & \leq \sup_{F \in \Sigma} \mathbf{E}\left[\left|F\left(Z_n^{\sharp}\right) - F\left(\Pi_n Z_n^{\sharp}\right)\right| \mathbf{1}_{\left\{T < \tau_n^{\sharp}\right\}}\right] + c\,\mathbf{P}\left[T \geq \tau_n^{\sharp}\right] \\ & \leq \mathbf{E}\left[\parallel Z_n^{\sharp} - \Pi_n Z_n^{\sharp} \parallel_W \mathbf{1}_{\left\{T < \tau_n^{\sharp}\right\}}\right] + c\,\mathbf{P}\left[T \geq \tau_n^{\sharp}\right] \\ & \leq \mathbf{E}\left[\parallel Z_n^{\sharp}\left(. \wedge \tau_n^{\sharp}\right) - \Pi_n\left(Z_n^{\sharp}\left(. \wedge \tau_n^{\sharp}\right)\right) \parallel_W \mathbf{1}_{\left\{T < \tau_n^{\sharp}\right\}}\right] + c\,\mathbf{P}\left[T \geq \tau_n^{\sharp}\right].\end{aligned}$$

On the one hand, from Tchebychev inequality we have that for all n,

$$(45)\qquad \mathbf{P}\left[T \geq \tau_n^{\sharp}\right] = \mathbf{P}\left[N_{n\lambda}(T) \geq 2\lambda n T\right] \leq \frac{\mathrm{Var}\left(N_{n\lambda}(T)\right)}{(\lambda n T)^2} \leq \frac{c}{n}.$$

Also, for any n, on $\{T < \tau_n^\sharp\}$ we have that

$$L_n^\sharp\left(t \wedge \tau_n^\sharp\right) \leq N_{n\lambda}(t) \leq 2\lambda n T,$$

therefore the Markov process $L_n^\sharp\left(. \wedge \tau_n^\sharp\right)$ satisfies to the Assumptions of Lemma 3.1 for $J \equiv 1$ and $\alpha \equiv \lambda \vee (\mu T)$. Thus we obtain as in (33) that for all n,

$$\begin{aligned}(46)\quad \mathbf{E}\left[\parallel Z_n^\sharp\left(. \wedge \tau_n^\sharp\right) - \Pi_n\left(Z_n^\sharp\left(. \wedge \tau_n^\sharp\right)\right) \parallel_W \mathbf{1}_{\left\{T<\tau_n^\sharp\right\}}\right] \\ \leq \frac{1}{\sqrt{n}}\mathbf{E}\left[\parallel L_n^\sharp - \Pi_n L_n^\sharp \parallel_W\right] + \sqrt{n} \parallel \overline{L^\sharp} - \Pi_n\overline{L^\sharp} \parallel_W \\ \leq \frac{c\log n}{\log\log n\ \sqrt{n}},.\end{aligned}$$

where, recalling (4), we use the fact that

$$\begin{aligned}\sqrt{n} \parallel \overline{L^\sharp} - \Pi_n\overline{L^\sharp} \parallel_W &\leq 2\sqrt{n} \max_{i\in[0,n-1]} \sup_{t\in\left[t_i^n;\frac{(i+1)T}{n}\right]} \left|e^{-\mu t} - e^{-\mu\frac{iT}{n}}\right| \\ &\leq 2\sqrt{n}\left(e^{-\frac{\mu}{n}} - 1\right) \leq \frac{c}{\sqrt{n}}.\end{aligned}$$

Finally, gathering (46) with (45) in (44) entails that for all n,

$$\mathrm{d}_\Sigma(Z_n^\sharp, \Pi_n Z_n^\sharp) \leq \frac{c\log\ n}{\sqrt{n}}$$

which, together with with Proposition 6.5 and (11) in (43), concludes the proof.

Appendix A. Moment bound for Poisson variables

By following closely Chapter 2 in [3], we show hereafter a moment bound for the maximum of n Poisson variables. (Notice that, contrary to Exercise 2.18 in [3] we do not assume here that the Poisson variables are independent.)

Proposition A.1. *Let $n \in \mathbb{N}$ and let X_i, $i = 1, \cdots, n$ be Poisson random variables of parameter ν. Then for some c depending only on ν we have that*

$$(47)\qquad \mathbf{E}\left[\max_{i=1,\cdots,n} X_i\right] \leq c\,\frac{\log n}{\log\log n}.$$

Proof. Denote for all i, $Z_i = X_i - \nu$, and by Ψ_{Z_i} the moment generating function of Z_i. By Jensen's inequality and the monotonicity of $\exp(.)$ we get that

$$\exp\left(u\mathbf{E}\left[\max_{i=1,\cdots,n} Z_i\right]\right) \leq \mathbf{E}\left[\max_{i=1,\cdots,n} \exp(uZ_i)\right] \leq \sum_{i=1}^n \mathbf{E}\left[\exp(uZ_1)\right] \leq n\exp\left(\Psi_{Z_i}(u)\right).$$

After a quick algebra, this readily implies that

$$\mathbf{E}\left[\max_{i=1,\cdots,n} Z_i\right] \leq \inf_{u\in\mathbb{R}}\left(\frac{\log n + \nu\left(e^u - u - 1\right)}{u}\right) = \frac{\log n + \nu\left(e^{\frac{a}{W(a)}} - 1 - W(a) - 1\right)}{1 + W(a)},$$

where W is the so-called Lambert function, solving the equation $W(x)e^{W(x)} = x$ over $[-1/e, \infty]$, and $a = \frac{\log(n/e^\nu)}{e^\nu}$. This entails in turn that

$$\mathbf{E}\left[\max_{i=1,\cdots,n} X_i\right] \leq \nu e \frac{a}{W(a)} - \nu + \nu = \frac{\log(n/e^\nu)}{W(\log(n/e^\nu)/e^\nu)}.$$

We conclude by observing that $W(z) \geq \log(z) - \log\log(z)$ for all $z > e$. Therefore there exists $c > 0$ such that for $n \geq \exp\left(e^{\nu+1} + \nu\right)$,

$$\mathbf{E}\left[\max_{i=1,\cdots,n} X_i\right] \leq \frac{\log(n/e^\nu)}{\log(\log(n/e^\nu)/e^\nu) - \log\log(\log(n/e^\nu)/e^\nu)} \leq c\,\frac{\log n}{\log\log n},$$

which completes the proof. □

Appendix B. Proof of Proposition 6.4

Fix n throughout this section, and denote for all $i = 0, ..., n-1$ and $(x, z) \in \mathbb{R}^2$,

$$\alpha_i(x,z) = \mathbf{1}_{C_{t_i^n}}(x,z), \qquad \beta_i(x,z) = \int_{t_i^n}^{t_{i+1}^n} \mathbf{1}_{C_u}(x,z)\,\mathrm{d}u.$$

Proof of (i). Recall (42), and fix two indexes $0 \leq i < j \leq n-1$. We have that

$$\begin{aligned}(48)\quad \int\int u_i^\sharp u_j^\sharp \,\mathrm{d}\nu_n^\sharp &= \int\int (\alpha_{i+1} - \alpha_i)(\alpha_{j+1} - \alpha_j)\,\mathrm{d}\nu_n^\sharp \\ &+\mu\int\int \beta_i(\alpha_{j+1} - \alpha_j)\,\mathrm{d}\nu_n^\sharp + \mu\int\int \beta_j(\alpha_{i+1} - \alpha_i)\,\mathrm{d}\nu_n^\sharp + \mu^2\int\int \beta_j\beta_j\,\mathrm{d}\nu_n^\sharp \\ &=: I_1 + I_2 + I_3 + I_4,\end{aligned}$$

where straightforward computations show that

$$\begin{aligned}
I_1 &= \lambda n\left(2e^{-\mu(t_j^n - t_i^n)} - e^{-\mu(t_j^n - t_{i+1}^n)} - e^{-\mu(t_j^n - t_{i-1}^n)}\right);\\
I_2 &= \frac{\lambda n}{\mu}\left(2e^{-\mu(t_j^n - t_i^n)} - e^{-\mu(t_j^n - t_{i+1}^n)} - e^{-\mu(t_j^n - t_{i-1}^n)}\right) - \lambda\left(e^{-\mu t_{j+1}^n} - e^{-\mu t_j^n}\right);\\
I_3 &= \frac{\lambda n}{\mu}\left(-2e^{-\mu(t_j^n - t_i^n)} + e^{-\mu(t_j^n - t_{i+1}^n)} + e^{-\mu(t_j^n - t_{i-1}^n)}\right);\\
I_4 &= \frac{\lambda n}{\mu}\left(-2e^{-\mu(t_j^n - t_i^n)} + e^{-\mu(t_j^n - t_{i+1}^n)} + e^{-\mu(t_j^n - t_{i-1}^n)}\right) + \lambda\left(e^{-\mu t_{j+1}^n} - e^{-\mu t_j^n}\right).
\end{aligned}$$

Adding up the above in (48) yields the result. □ □

Proof of (ii). For all $0 \le i,j,k \le n-1$ we write

(49)

$$I_{i,j,k} := \int_{\mathbb{R}^2} |u_i^\sharp u_j^\sharp u_k^\sharp| \, \mathrm{d}\nu_n^\sharp \le \int |(\alpha_{i+1}-\alpha_i)(\alpha_{j+1}-\alpha_j)(\alpha_{k+1}-\alpha_k)| \, \mathrm{d}\nu_n^\sharp$$
$$+ \int |(\alpha_{i+1}-\alpha_i)(\alpha_{j+1}-\alpha_j)\mu\beta_k| \, \mathrm{d}\nu_n^\sharp + \int |(\alpha_{j+1}-\alpha_j)(\alpha_{k+1}-\alpha_k)\mu\beta_i| \, \mathrm{d}\nu_n^\sharp$$
$$+ \int |(\alpha_{i+1}-\alpha_i)(\alpha_{k+1}-\alpha_k)\mu\beta_j| \, \mathrm{d}\nu_n^\sharp + \int \left|(\alpha_{i+1}-\alpha_i)\,\mu^2\beta_j\beta_k\right| \, \mathrm{d}\nu_n^\sharp$$
$$+ \int \left|(\alpha_{j+1}-\alpha_j)\,\mu^2\beta_i\beta_k\right| \, \mathrm{d}\nu_n^\sharp + \int \left|(\alpha_{k+1}-\alpha_k)\,\mu^2\beta_i\beta_j\right| \, \mathrm{d}\nu_n^\sharp$$
$$+ \int \left|\mu^3\beta_i\beta_j\beta_k\right| \, \mathrm{d}\nu_n^\sharp =: \sum_{l=1}^{8} I_{i,j,k}^l.$$

It can be easily retrieved that

$$I_{i,i,i}^1 = n\left(\frac{\lambda}{n} - \frac{\lambda}{\mu}\left(1-e^{-\frac{\mu T}{n}}\right)\left(1-e^{\mu T\frac{i+1}{n}}\right)\right) \le \frac{\lambda}{\mu};$$
$$I_{i,j,k}^1 = 0, \quad 1 \le i < j < k \le n;$$
$$I_{i,i,k}^1 = \frac{\lambda n}{\mu}\left(e^{\mu t_{i+1}^n} - e^{\mu t_i^n}\right)\left(e^{-\mu t_k^n} - e^{-\mu t_{k+1}^n}\right) \le \frac{\lambda T^2}{\mu n}, \quad i = j < k,$$

and the other cases can be treated similarly. Also, simple computations show that if $i < j$,

$$\mu \int |(\alpha_{i+1}-\alpha_i)(\alpha_{j+1}-\alpha_j)\beta_k| \, \mathrm{d}\nu_n^\sharp \le \lambda\left(e^{\mu t_{i+1}^n} - e^{\mu t_i^n}\right)\left(e^{-\mu t_j^n} - e^{-\mu t_{j+1}^n}\right) \le \frac{\lambda T^2}{n^2},$$

whereas if $i = j$, the above integral is upper bounded by

$$2\lambda T\left(2 + e^{-\mu t_{i+1}^n} - e^{-\mu t_i^n} - 2e^{-\frac{\mu T}{n}}\right) \le \frac{2\lambda\mu T^2}{n} T^2.$$

It readily follows that in all cases, $I_{i,j,k}^2, I_{i,j,k}^3$ and $I_{i,j,k}^4$ are less than $c\,n^{-1}$ for some constant c. Reasoning similarly, we also obtain that for all i,j,k,

$$\mu^2 \int \left|(\alpha_{i+1}-\alpha_i)\mu^2\beta_j\beta_k\right| \, \mathrm{d}\nu_n^\sharp \le \frac{\mu^2}{n}\left(\frac{\lambda}{n} - \frac{\lambda}{\mu}\left(1-e^{-\frac{\mu T}{n}}\right)\left(1-e^{\mu T\frac{i+1}{n}}\right)\right) \le \frac{\lambda}{\mu n^2} T,$$

so that in all cases the $I_{i,j,k}^5, I_{i,j,k}^6$ and $I_{i,j,k}^7$'s are less than $c\,n^{-2}$ for some c. Finally, observing that for all u,v,w,

$$\int\int \mathbf{1}_{C_u}\mathbf{1}_{C_v}\mathbf{1}_{C_w}\lambda\mu e^{-\mu y} \, \mathrm{d}x \, \mathrm{d}y = \frac{\lambda}{\mu}(e^{-\mu(\max(u,v,w)-\min(u,v,w))} - e^{-\mu\max(u,v,w)})$$

we can similarly bound $I_{i,j,k}^8$ by a $c\,n^{-2}$ for all i,j,k. To summarize, all the $I_{i,j,k}$'s are less than $c\,n^{-2}$ for some c, except for the $I_{i,i,i}^1$'s, $i = 1,...,n$, which are bounded by a constant but are only n in number, and all terms where one index appears twice, which are less than $c\,n^{-1}$ for some c, but are only n^2 in number. Hence (ii).

□ □

Proof of (iii). We have for all $0 \leq i \leq n-1$,

$$\begin{aligned}(50)\quad \int\int u_i^\sharp u_i^\sharp \,\mathrm{d}\nu_n^\sharp &= \int\int \alpha_{i+1}\,\mathrm{d}\nu_n^\sharp + \int\int \alpha_i\,\mathrm{d}\nu_n^\sharp - 2\int\int \alpha_{i+1}\alpha_i\,\mathrm{d}\nu_n^\sharp \\ &+ 2\mu\int\int \beta_i\alpha_{i+1}\,\mathrm{d}\nu_n^\sharp - 2\mu\int\int \beta_i\alpha_i\,\mathrm{d}\nu_n^\sharp + \mu^2\int\int \beta_i\beta_i\,\mathrm{d}\nu_n^\sharp \\ &= J_1+J_2+J_3+J_4+J_5+J_6,\end{aligned}$$

where straightforward calculations show that

$$\begin{aligned}J_1 &= \frac{\lambda n}{\mu}\left(1-e^{-\mu t_{i+1}^n}\right); \quad J_2 = \frac{\lambda n}{\mu}\left(1-e^{-\mu t_i^n}\right);\\ J_3 &= -2\frac{\lambda n}{\mu}\left(e^{-\frac{\mu T}{n}} - e^{-\mu t_{i+1}^n}\right); \quad J_4 = 2\frac{\lambda n}{\mu}(1-e^{-\frac{\mu T}{n}}) - 2\lambda e^{-\mu t_{i+1}^n};\\ J_5 &= -2\frac{\lambda n}{\mu}(1-e^{-\frac{\mu T}{n}}) - 2\frac{\lambda n}{\mu}(e^{-\mu t_{i+1}^n} - e^{-\mu t_i^n});\\ J_6 &= \lambda\left(2+2e^{-\mu t_{i+1}^n} + \frac{2n}{\mu}(e^{-\mu t_{i+1}^n} - e^{-\mu t_i^n} + e^{\frac{-\mu T}{n}} - 1)\right).\end{aligned}$$

Recalling (37), adding up the J_j's, $j = 1, ..., 6$, concludes the proof. □ □

References

[1] Barbour, A. D. (1990). Stein's method for diffusion approximations. *Probability Theory and Related Fields* **84,** 297–322.

[2] Borovkov, A. A. (1967). Limit Laws for queueing processes in multichannel systems. *Sibirsk Mat* **8,** 983–1004.

[3] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration inequalities.* Oxford University Press.

[4] Braverman, A. and Dai, J. G. (2017). Stein's method for steady-state diffusion approximations of $M/Ph/n+M$ systems. *The Annals of Applied Probability* **27,** 550–581.

[5] Braverman, A., Dai, J. G. and Feng, J. (2016). Stein's method for steady-state diffusion approximations: an introduction through the Erlang-A and Erlang-C models. *Stochastic Systems* **6,** 301–366.

[6] Coutin, L. and Decreusefond, L. (2013). Stein's method for Brownian approximations. *Communications in Stochastic Analysis,***7,** 349–372.

[7] Coutin, L. and Decreusefond, L. (2019). Stein's method for rough paths. *Potential Analysis.* DOI = 10.1007/s11118-019-09773-z.

[8] Decreusefond, L. (2015). The Stein-Dirichlet-Malliavin method. *ESAIM: Proceedings* 11.

[9] Decreusefond, L. and Moyal, P. (2012). *Stochastic Modeling and Analysis of Telecom Networks.* ISTE Ltd and John Wiley & Sons Inc.

[10] Ethier, S. and Kurtz, T. (1986). *Markov Processes : Characterizations and Convergence.* Wiley.

[11] Friz, P. K. and Victoir, N. B. (2010). *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications.* Cambridge University Press.

[12] Gaunt, R. E. and Walton, N. (2020). SteinâĂŹs method for the single server queue in heavy traffic. *Statistics & Probability Letters* **156,** 108566.

[13] Kasprzak, M. J. (2017). Diffusion approximations via Stein's method and time changes. *arXiv:1701.07633.*

[14] Privault, N. (2009). *Stochastic Analysis in Discrete and Continuous Settings with Normal Martingales* vol. 1982 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin.

[15] Robert, P. (2003). *Stochastic Networks and Queues* Springer-Verlag, Berlin.

[16] Shih, H.-H. (2011). On Stein's method for infinite-dimensional Gaussian approximation in abstract Wiener spaces. *Journal of Functional Analysis* **261,** 1236–1283.
[17] Shwartz, A. and Weiss, A. (1995). *Large Deviations For Performance Analysis: Queues, Communication and Computing*. Chapman and Hall/CRC, London.
[18] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. The Regents of the University of California.
[19] Whitt, W. (2002). *Stochastic-process limits: an introduction to stochastic-process limits and their application to queues*. Springer.

LTCI, Telecom Paris, I.P. Paris, France
E-mail address: eustache.besancon@mines-telecom.fr
E-mail address: coutin@math.univ-toulouse.fr

LTCI, Telecom Paris, I.P. Paris, France
E-mail address: laurent.decreusefond@mines-telecom.fr

UniversitÃŁ de Lorraine, France
E-mail address: pascal.moyal@univ-lorraine.fr