Sample Path Moderate Deviations for Shot Noise Processes in the High Intensity Regime

SUMITH REDDY ANUGU* AND GUODONG PANG*

Abstract. We study the sample-path moderate deviation principle (MDP) for shot noise processes in the high intensity regime. The shot noise processes have a renewal arrival process, non-stationary noises (with arrival-time dependent distributions) and a general shot response function of the noises. The rate function in the MDP exhibits a memory phenomenon in this asymptotic regime, which is in contrast with that in the conventional time-space scaling regime.

To prove the sample-path MDP, we first establish that this is equivalent to establishing the sample-path MDP of another process that is easier to study. We will then establish the sample-path MDP of this process combining the Gärtner-Ellis (to prove the finite dimensional MDP) and Dawson-Gärtner Theorem (to prove the sample-path MDP under the topology of pointwise convergence). Finally, we prove exponential tightness and strengthen the MDP to the Skorohod J_1 topology. In the proofs, because of the inherent non-stationarity of shot noise process, we establish a new maximal inequality and use it to prove exponential tightness and the aforementioned equivalence. The rate function is derived using the tools of reproducing kernel Hilbert space.

1. Introduction

Shot noise process can be viewed as a natural model for a system which experiences shocks that occur according to an arrival process and have an enduring effect on its dynamics. In particular, they have been found very useful in the areas of physics ([7, 39, 53]), queueing theory ([10, 27, 38]) and teletraffic theory ([37, 48]), insurance and risk theory ([35, 36, 41, 42, 44, 54]), storage processes ([11, 40]) and so on.

Various asymptotic properties and scaling limits have been established for shot noise processes. There are two asymptotic scaling regimes that have been studied in the literature. The first one is the conventional time-space scaling regime (speeding up time and scaling down space; see (2.23) and (2.31)). This is studied in different settings, including functional central limit theorems (FCLTs) in [29, 30, 31, 32, 34, 35, 36] and sample-path large deviation principles (LDPs) in [15, 19, 22]. The second one is referred to as the high intensity regime (the arrival rate is scaled up while time is not scaled in the shot response function, and space is scaled down; see (2.6) and (2.9)). FCLTs and relevant asymptotic properties in this regime have been studied in [4, 26, 28, 46, 47]. Infinite-server queues can be regarded as a shot noise process with a particular indicator response function, and heavy-traffic limits (that is, in the high intensity regime with no scaling on service times) have been established in the literature (see, e.g., [45] and references therein). Large deviation principles are also established for infinite-server queues in heavy traffic/high intensity regime, see [24] for results at a fixed time, and [6] for the sample-path LDP of a two-parameter process tracking elapsed/residual service times. However, to our best knowledge, no MDPs have been established for shot noise processes in either asymptotic regime in the literature.

---

† Department of Computational Applied Mathematics and Operations Research, George R. Brown School of Engineering, Rice University, Houston, TX 77005

E-mail addresses: anugu.reddy@rice.edu, gdpang@rice.edu.

Key words and phrases. Sample path moderate deviation principle, shot noise process, renewal arrival process, non-stationary noise distribution (arrival-time dependent distributions).

The main aim of this paper is to establish the sample-path MDP for shot noise processes in the high intensity regime. We consider the arrival process to be a renewal process satisfying a sample-path MDP, and assume that the noises are non-stationary, in particular, the noises are conditionally independent given the arrival times while the distribution of each noise depends on its own associated arrival time. The shot response function also satisfies rather general conditions (see Assumptions 2.1, 2.2, 2.3 and 2.4; see also the relevant remarks). The MDP-scaled process in this regime is given in (2.9), which has a centering term as in FCLTs, but has a scaling of arrival rate and space satisfying a certain condition (2.8). See further discussions on the scaling at the beginning of Section 2.2. The main result of the sample-path MDP is given in Theorem 2.2, in which the rate function consists of two components, one involving a "covariance" operator (corresponding to the covariance function in the Gaussian limit with memory of the FCLT in [46]) and the other being standard corresponding to that of a Brownian motion. We refer to Remark 2.12 on discussions about how the MDP rate functions are related to the Gaussian limit processes in the FCLT. As a corollary, we also state the sample-path MDP result for the $GI/G_t/\infty$ queueing model (Corollary 2.1).

To prove the sample-path MDP, we show that the scaled process of interest (see (2.9)) is equivalent (in an appropriate sense) to a new process that is easier to study. The main consequence of this is that the MDP of the process of interest is implied by the MDP of the new process with identical rates and rate functions. The important feature of this new process is that it can be written as the sum of two independent intermediate processes where one process (say P1) is independent of the arrival process and the other process (say P2) has randomness only through the arrival process. This feature helps us to study and establish the MDPs of these two intermediate processes separately and infer the MDP of their sum. The MDP of the P2 can be easily concluded by an application of the contraction principle. However, the MDP of the P1 is non-trivial. The proof involves the application of Gärtner-Ellis and Dawson-Gärtner Theorems to conclude the MDP under the topology of pointwise convergence. We then strengthen it to the Skorohod J_1 topology by establishing an appropriate version of tightness of P1. As already mentioned, due to the memory phenomenon, the sample path rate function is only expressed in a variational form which is in contrast to the memoryless case (for example, compound Poisson), where the rate function can be given in an explicit form.

As a comparison we also state the the sample-path MDP result in the conventional time-space scaling regime (Theorem 2.3 in Section 2.3). Although both the scalings give rise to the MDP with the same rates, the rate functions are dramatically different. In the case of the conventional time-space scaling, the rate function looks like the inverse of covariance of a Brownian motion (with a time-varying covariance function), whereas in the high intensity regime, the rate function looks like the inverse of covariance of a certain non-stationary Gaussian process with memory. See further discussions in Remarks 2.13 and 2.14.

As another comparison, we also state the sample-path LDPs for the shot noise processes in the two scaling regimes (Section 2.4). It is expected that the rate functions in the LDPs are very different in the two regimes (see (2.29) and (2.32)), since in the conventional time-space scaling regime, the 'lingering' effect of noises vanishes (equivalent to that the response function $H(t,x)$ is replaced by $H(\infty,x)$), while in the high intensity scaling regime, the effect of noises is indicated as 'memory'. On the other hand, the rate functions in the LDPs involve the LDP rate function of the renewal arrival process, which further involves the log moment generating function of the interarrival times (see (2.30)). However, the rate functions in the MDPs in both regimes only involve the mean and variance of the interarrival times of the renewal arrival process (Theorems 2.2 and 2.3). We highlight that some of the LDP results in both regimes are also new to the literature since we consider non-stationary noises (see further discussions in Section 2.4). The sample-path large deviation principles (LDPs) in [15, 19, 22] all assume Poisson arrival processes and stationary noises in the conventional time-space scaling regime. The results we present are more general involving both a renewal arrival process and non-stationary noises. The methodology we develop also goes beyond those in [15, 22] since the well developed LDP methods for Poisson random measures (see the recent monograph [13]) could be used, which is impossible in our setting.

To put our paper in the context of the vast literature of MDPs of stochastic systems, we give a partial overview of the different methods used to prove MDPs. The general approach to prove sample-path LDP and MDP is to use Dawson-Gärtner theorem [18, Theorem 4.6.1] in conjunction with Gärtner-Ellis theorem [18, Theorem 2.3.6] and exponential tightness in appropriate functional spaces. This has been used in the study of various Markov and non-Markov systems (see, e.g., [6, 19, 21, 22, 49, 50, 51]). Specific properties of processes of interest are often exploited to establish the required properties, such as convergence of the non-linear semi-groups of Markov processes [21], and semimartingale representations [50, 51]. For certain Markov processes driven by Brownian motion or Poisson random measure, a weak convergence approach using the variational representation of certain functionals of Brownian motion and/or Poisson random measure and the associated control problem formulation (see [13, Section 3.2, 3.3, 8.1 and 8.2]) has been used extensively (see, e.g., [3, 12, 14, 16, 17, 23, 43] for Markov models and [1, 9, 33] for non-Markov models). However, shot noise processes in our paper bring in new challenges with the non-Markovian and non-stationary characteristics. This leads us to great difficulties in proving the exponential tightness (Theorem 3.3) and exponential equivalence (Theorem 3.1), as well as identifying the rate function due to the memory property. We have developed new methods to tackle these challenges, which may turn to be useful to study LDP and MDP for other non-Markovian stochastic systems in future work.

1.1. **Organization of the paper.** In the rest of this section, we introduce notation that we will use throughout the paper. In Section 2.1, we introduce the model, and in Section 2.2, we state the main results on sample-path MDP in the high intensity regime. The assumptions on the model are given in Sections 2.1 and 2.2. In Sections 2.3 and 2.4, we compare with the MDP in the conventional time-space scaling regime, and the LDP results in the two scaling regimes. In Sections 3–5, we provide the proofs of the main result. In the Appendix, we prove a maximal equality that will be used in the proof of our main result. We also provide the sketch of the proof for the analogous LDP result in the high intensity regime.

1.2. **Notation.** $(\Omega, \mathcal{F}, \mathbb{P})$ denotes the abstract probability space. For a Polish space S, $\mathcal{B}(S)$ denotes the corresponding Borel σ–algebra. $\mathcal{P}(S)$ denotes the space of probability measures on S equipped with the topology of weak convergence. For a set $\mathtt{A} \in \mathcal{B}(S)$, $\bar{\mathtt{A}}$, $\mathtt{A}^c$, $\mathtt{A}^\circ$ denote the closure, the complement and the interior of $\mathtt{A}$, respectively. For a fixed $T > 0$, $(\mathcal{D}_T, J_1)$ denotes the space of real valued functions on $[0, T]$ that are right continuous with left limits equipped with the Skorohod topology ($d_{J_1}(\cdot,\cdot)$ is the corresponding metric). Let $\|x\|_T \doteq \sup_{0\le t\le T}|x(t)|$ for $x \in \mathcal{D}_T$, and $\langle\cdot,\cdot\rangle$ denote the Euclidean inner product on $\mathbb{R}^m$, for any $m \in \mathbb{N}$. Also, for any m, we write $\underline{x} = (x_1, x_2, \dots, x_m)$ for short. The set of continuous functions on $[0, T]$ equipped with the uniform norm is denoted by $(\mathcal{C}_T, \|\cdot\|_T)$ and the set of absolutely continuous functions $x : [0, T] \to \mathbb{R}$ such that $x(0) = 0$ by $\mathcal{AC}_0$. For $a, b \in \mathbb{R}$, we let $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. Let $\mathbb{R}_+ \doteq [0, \infty)$. For any two real-valued functions f and g, we write $f = \mathcal{O}(g)$ whenever $\limsup_{x\to\infty}\left|\frac{f(x)}{g(x)}\right| < \infty$. The indicator function associated to set A is denoted by $\mathbb{1}_A$. Finally, $V_T(h(\cdot))$ is the total variation of any function $h(\cdot)$ on $[0, T]$.

For a Polish space S and b_n ($b_n \uparrow \infty$), a family of S–valued random variables $\{\mathcal{Y}^n\}_{n\in\mathbb{N}}$ is said to satisfy a large deviation principle (LDP) with rate b_n and rate function $I : S \to [0, \infty]$, if the following holds:

(i) For $\mathtt{A} \in \mathcal{B}(S)$,
$$-\inf_{x\in\mathtt{A}^\circ} I(x) \le \liminf_{n\to\infty}\frac{1}{b_n}\log\mathbb{P}(\mathcal{Y}^n \in \mathtt{A}) \le \limsup_{n\to\infty}\frac{1}{b_n}\log\mathbb{P}(\mathcal{Y}^n \in \mathtt{A}) \le -\inf_{x\in\bar{\mathtt{A}}} I(x);$$

(ii) The set $\{x : I(x) \le l\}$ is compact in S, for every $l \ge 0$.

We say that $\{\mathcal{Y}^n\}_{n\in\mathbb{N}}$ is exponentially tight with rate b_n, if for every $l > 0$, there exists a compact set $\mathcal{K}_l$ such that

$$\limsup_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}(\mathcal{Y}^n \in \mathcal{K}_l^c) \le -l\,.$$

Two families of S–valued random variables $\{\mathcal{Y}_1^n\}_{n\in\mathbb{N}}$ and $\{\mathcal{Y}_2^n\}_{n\in\mathbb{N}}$ are said to be exponentially equivalent with respect to rate b_n, if the following holds: for every $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}\left(d(\mathcal{Y}_1^n, \mathcal{Y}_2^n) > \delta\right) = -\infty\,.$$

Here, $d(\cdot,\cdot)$ is the metric associated with S.

For stochastic processes $\{\mathcal{Y}^n\}$ whose paths are in $\mathcal{D}_T$, in the context of our paper, the processes $\mathcal{Y}^n$ are associated with appropriate scalings for LDP and MDP (n as the scaling parameter). The LDP and MDP will be given as the LDP described above, but with different rates b_n and rate functions I. See the corresponding statements in Theorem 2.2 for the MDP and Section 2.4.1 for the LDP for the shot noise process under different scalings.

2. Model and Results

2.1. The model.

We consider the following shot noise process

$$X(t) = \sum_{i=1}^{A(t)} H(t-\tau_i, \xi_i), \quad t \ge 0. \tag{2.1}$$

Here $H : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ is a Borel measurable function, with $H(t,\cdot) \equiv 0$ for $t < 0$. $\{A(t) : t \ge 0\}$ is a renewal process with almost surely positive i.i.d. interarrival times $\{\eta_i\}_{i\ge1}$, *i.e.,* with arrival times $\{\tau_i\}_{i\ge1}$ defined by $\tau_i = \sum_{k=1}^{i} \eta_k$ for $i \ge 1$,

$$A(t) = \sum_{k\ge1} \mathbb{1}_{\{\tau_k \le t\}}, \quad t \ge 0. \tag{2.2}$$

Let $\lambda \doteq (\mathbb{E}[\eta_1])^{-1}$ and $\sigma^2 \doteq \mathbb{E}[\eta_1^2] - (\mathbb{E}[\eta_1])^2$. The noises $\{\xi_i\}_{i\ge1}$ are given by

$$\xi_i = g(\tau_i, \vartheta_i), \tag{2.3}$$

where $g : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ is a measurable and deterministic function, and $\{\vartheta_i\}_{i\ge1}$ is a sequence of i.i.d. $\mathbb{R}^d$-valued random variables, independent of $\{\eta_i\}_{i\ge1}$. Let F be the distribution of ϑ_1.

Remark 2.1. The distribution of the noises $\xi_i = g(\tau_i, \vartheta_i)$ depends on the arrival time τ_i. Since $\vartheta_1, \vartheta_2, \ldots$ are i.i.d., the random variables $\xi_1, \xi_2, \ldots$ are independent given the arrival times $\{\tau_i\}_{i\in\mathbb{N}}$, and given that $\tau_i = s$, the distribution of ξ_i is

$$F_s(x) \doteq \mathbb{P}(\xi_i \le x | \tau_i = s) = \mathbb{P}(g(s,\vartheta_i) \le x) = \int_{\mathbb{R}^d} \mathbb{1}_{\{g(s,y)\le x\}} F(dy).$$

This is one approach to model the non-stationarity of noises. For instance, $F_s(x) = 1 - e^{-\mu(s)x}$ for a positive function $\mu(s)$ can be regarded as a non-stationary exponential distribution with rate $\mu(s) \ge 0$ for $s \ge 0$ and $x \ge 0$. In this case, if $\mu(s) > 0$ for all $s \ge 0$, one can take $g(s,\vartheta_i) = (\mu(s))^{-1}\vartheta_i$ with ϑ_i being an exponential random variable with mean 1. For another instance, if $\xi_i = g(s,\vartheta_i) = \int_s^{s+\vartheta_i} \phi(u)du$ for some $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ and $\vartheta_i \in \mathbb{R}_+$, that is, in the context of queueing systems, the realization of the service requirement ϑ_i is through a time-varying rate ϕ and ξ_i is the realized service time, and in this case, $F_s(x) = \int_{\mathbb{R}_+} \mathbb{1}_{\{\int_s^{s+y}\phi(u)du \le x\}} F(dy)$. A third example is that the functions are specified in the piecewise sense over t: for each fixed $x \in \mathbb{R}^d$, $g(s,x) = g_k(x)$ for $s \in [t_k, t_{k+1}]$ given $0 = t_1 < t_2 < \cdots < t_k < t_{k+1} < \cdots < t_K = T$ and $g_k : \mathbb{R}^d \to \mathbb{R}^d$ for each k, so that $F_s(x) = \sum_{k=1}^K \mathbb{1}_{\{s\in[t_k,t_{k+1}]\}} \int_{\mathbb{R}^d} \mathbb{1}_{\{g_k(y)\le x\}} F(dy)$ for $s \ge 0$.

Assumption 2.1. $H(\cdot, x)$ is right continuous with left limits for each $x \in \mathbb{R}^d$.

Remark 2.2. Assumption 2.1 will ensure that the process $\{X(t), t \in [0,T]\}$ is $\mathcal{D}_T$–valued. To see this, fix a sequence $\{\tau_i\}_{i\geq 1}$ of arrival times. Then $A(t) = k$, whenever $t \in [\tau_k, \tau_{k+1})$ for $k \geq 1$ and $A(t) = 0$, for $t \leq [0, \tau_1)$. For any $t \in [\tau_k, \tau_{k+1})$, since the interarrival times are almost surely positive, there is almost surely a $\delta > 0$ small enough such that $A(s) = k$ and

$$X(s) = \sum_{i=1}^{k} H(s - \tau_i, \xi_i),$$

for $s \in [t, t+\delta)$. Taking $s \downarrow t$, Assumption 2.1 gives us the following:

$$\lim_{s\downarrow t} X(s) = X(t).$$

This proves that for every t, $X(t)$ is right continuous and the finiteness of left limits trivially follows. Hence, $\{X(t), t \in [0,T]\}$ is a $\mathcal{D}_T$–valued process.

Remark 2.3. A common form of the function $H(t,x)$ is multiplicative, taking the form $H(t,x) = \tilde{H}(t)\varphi(x)$, where $\tilde{H} : \mathbb{R}_+ \to \mathbb{R}$ and $\varphi : \mathbb{R}^d \to \mathbb{R}$. The condition in Assumption 2.1 requires that $\tilde{H}$ is right continuous with left limits. The function $\tilde{H}(\cdot)$ in the expression of $X(t)$ represents the effect of the noises as time progresses. A typical example is the exponential decay effect, that is, $\tilde{H}(t) = e^{-\beta t}$ for some constant $\beta > 0$ and $t \in \mathbb{R}_+$. Another one is a power function effect, that is, $\tilde{H}(t) = t^\beta$ for some $\beta > 0$ and $t \in \mathbb{R}_+$. However, the function $H(t,x)$ can be also non-multiplicative, for example, in the $G/G/\infty$ queueing model, the queue length (number of customers/jobs) process has $H(t,x) = \mathbb{1}_{\{t<x\}}$ and the workload-input process has $H(t,x) = x\mathbb{1}_{\{t<x\}}$ for $x \geq 0$.

Define $H(t,s,x) \doteq H(t-s, g(s,x))$ for $t,s \in \mathbb{R}_+$ and $x \in \mathbb{R}^d$. The following functions are frequently used in what follows. For $t,s,u \in [0,T]$, if they exist, define

$$G_1(t,s) \doteq \int_{\mathbb{R}^d} H(t,s,x) F(dx)\,, \tag{2.4}$$

$$G_2(t,s,u) \doteq \int_{\mathbb{R}^d} H(t,u,x)H(s,u,x)F(dx)\,,$$

and

$$\Lambda(s,t) \doteq \lambda \int_0^{s\wedge t} \left(G_2(t,s,u) - G_1(t,u)G_1(s,u)\right) du. \tag{2.5}$$

In the special case of a multiplicative shot response function taking the form $H(t,x) = \tilde{H}(t)\varphi(x)$, where $\tilde{H} : \mathbb{R}_+ \to \mathbb{R}$ and $\varphi : \mathbb{R}^d \to \mathbb{R}$, we have the functions

$$G_1(t,s) = \tilde{H}(t-s) \int_{\mathbb{R}^d} \varphi(g(s,x))F(dx)\,,$$

$$G_2(t,s,u) = \tilde{H}(t-u)\tilde{H}(s-u) \int_{\mathbb{R}^d} \varphi(g(u,x))^2 F(dx)\,,$$

$$\Lambda(s,t) = \lambda \int_0^{s\wedge t} \tilde{H}(t-u)\tilde{H}(s-u)\left(\int_{\mathbb{R}^d} \varphi(g(u,x))^2 F(dx) - \left(\int_{\mathbb{R}^d} \varphi(g(u,x))F(dx)\right)^2 \right) du\,.$$

2.2. **Sample-path MDP in the high intensity regime.** In this section, we consider a scaled version of $X(t)$ in the high intensity regime and establish a sample-path MDP. Define the following scaled version of the shot noise process $X(\cdot)$ (defined in (2.1)):

$$\bar{X}^n(t) \doteq \frac{1}{n} \sum_{i=1}^{A^n(t)} H(t, \tau_i^n, \vartheta_i)\,, \tag{2.6}$$

where

$$A^n(t) \doteq A(nt), \quad \tau_i^n \doteq \frac{\tau_i}{n} = \frac{1}{n}\sum_{k=1}^{i} \eta_k\,. \tag{2.7}$$

We remark that the scaling in the process $\bar{X}^n(t)$ should be regarded as a high intensity scaling regime. The scaling of the renewal process $A^n(t) = A(nt)$ has the arrival times τ_i^n (and the interarrival times η_k) being scaled down by n, which is equivalent to the arrival rate λ being scaled up by n, that is, the arrival rate of $A^n(t)$ is $\lambda^n = n\lambda$. It can be thus regarded either as the usual scaling of (arrival) time, or as the scaling of the intensity (arrival rate). However, since the function $H(t, \tau_i^n, \vartheta_i)$ has no scaling in t, one should regard the scaling of $\bar{X}^n(t)$ in (2.6) as the high intensity scaling regime. See further discussions and comparison of the MDP results in the time-space scaling regime in Section 2.3 and the LDP results in the two different scaling regimes in Section 2.4.

Let $\{a_n\}_{n\in\mathbb{N}}$ be a positive real valued sequence such that

$$a_n \uparrow \infty \quad \text{and} \quad \frac{\sqrt{n}}{a_n} \uparrow \infty \quad \text{as} \quad n \to \infty. \tag{2.8}$$

We now define the following MDP-scaled process in the high intensity regime:

$$\begin{aligned} \widetilde{X}^n(t) &\doteq \frac{\sqrt{n}}{a_n}\left(\bar{X}^n(t) - \lambda\int_0^t G_1(t,s)ds\right) \\ &= \frac{\sqrt{n}}{a_n}\left(\frac{1}{n}\sum_{i=1}^{A^n(t)} H(t,\tau_i^n,\vartheta_i) - \lambda\int_0^t G_1(t,s)ds\right), \end{aligned} \tag{2.9}$$

where $G_1(t,s)$ is defined in (2.4).

Assumption 2.2. The interarrival times $\{\eta_i\}_{i\in\mathbb{N}}$ satisfy $\mathbb{E}[e^{\rho\eta_1}] < \infty$ for some $\rho > 0$.

To utilize the MDP results for renewal processes, we have imposed in Assumption 2.2 that the interarrival times have finite moment generating functions.

Remark 2.4. We remark that Assumption 2.2 is stronger than what is necessary to prove our main result which is Theorem 2.2. The reason behind choosing this assumption is as follows: One of the key ingredients of the proof of Theorem 2.2 is to invoke Theorem 2.1 given below. Under Assumption 2.2, the restrictions on a_n are the weakest where a_n satisfies (2.8) and Theorem 2.1 still holds. On the other hand, if we relax the assumption on moments of η_1, then we require a_n to satisfy stronger conditions than (2.8) for Theorem 2.1 to hold. See further discussions about how the conditions on a_n change as the conditions on moments of η_1 change in Remark 2.5.

Before presenting the main MDP result for $\widetilde{X}^n(t)$, we first present a version of the sample-path MDP for renewal processes.

Theorem 2.1. *[52, Theorem 6.2]*

Suppose Assumption 2.2 holds and a_n satisfies (2.8). Then, the family $\{\widetilde{A}^n\}_{n\in\mathbb{N}}$ defined by

$$\widetilde{A}^n(t) \doteq \frac{1}{a_n\sqrt{n}}\left(A^n(t) - n\lambda t\right), \quad \text{for } t\in[0,T] \tag{2.10}$$

where $A^n(t)$ is defined in (2.7), satisfies an MDP in $(\mathcal{D}_T, J_1)$ with rate a_n^2 and the following rate function $I_A^{MDP} : \mathcal{D}_T \to [0,\infty]$ given by

$$I_A^{MDP}(\phi) \doteq \begin{cases} \frac{1}{2\lambda^3\sigma^2}\int_0^T |\dot{\phi}(t)|^2 dt, & \text{whenever } \phi \in \mathcal{AC}_0, \\ \infty, & \text{otherwise.} \end{cases} \tag{2.11}$$

Remark 2.5. We remark that in the above theorem, the sample-path MDP for renewal processes is proved under either of the following two sets of conditions:

(i) $\frac{\log(n)}{a_n^2} \to \infty$ and $\mathbb{E}(\eta_1)^{2+\varepsilon} < \infty$, for some $\varepsilon > 0$;

(ii) For some $\beta \in (0,1]$, $\frac{n^{\frac{\beta}{2}}}{a_n^{2-\beta}} \to \infty$ and $\mathbb{E}\exp(b(\eta_1)^\beta) < \infty$, for some $b > 0$.

The proof is given in [50, Example 7.2]. As noted right after Theorem 6.2 in [52], the case $\beta = 1$ which is not included there is dealt with by the same argument. In our paper, to prove the MDP for $\widetilde{X}^n(t)$, we will decompose it into two components $\widetilde{X}_1^n(t)$ and $\widetilde{X}_2^n(t)$, the MDP for $\widetilde{X}_1^n(t)$ is proved under the conditions of a_n in (2.8), while the MDP for $\widetilde{X}_2^n(t)$ is proved using the contraction mapping theorem together with the known MDP result for renewal processes. Thus, to unify the conditions on a_n, we have imposed the conditions on the interarrival times $\{\eta_i\}_{i\in\mathbb{N}}$ in Assumption 2.2.

Let $M(x) \doteq \sup_{s,t\in[0,T]} |H(t,s,x)|$.

Assumption 2.3. The random variable ϑ_1 satisfies the following tail condition:

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log\Big[n\mathbb{P}\Big(M(\vartheta_1) \geq a_n\sqrt{n}\Big)\Big] = -\infty. \tag{2.12}$$

Remark 2.6. It is clear that if H is bounded uniformly in its arguments, then (2.12) necessarily holds. It is easy to show (upon direct application of Markov's inequality) that (2.12) also holds, whenever

$$\mathbb{E}\Big[e^{\tilde{\rho}M(\vartheta_1)}\Big] < \infty, \text{ for some } \tilde{\rho} > 0.$$

We also have the following result.

Lemma 2.1. *Suppose* (2.12) *holds. Then, we have the following:*

$$\mathbb{E}\Big[M(\vartheta_1)^2\Big] < \infty \tag{2.13}$$

$$\lim_{n\to\infty} \frac{a_n}{\sqrt{n}}\mathbb{E}\Big[M(\vartheta_1)^3\mathbb{1}_{\{M(\vartheta_1)\leq l\sqrt{n}a_n^{-1}\}}\Big] = 0, \textit{ for every } l > 0. \tag{2.14}$$

Proof. The proof of (2.13) follows directly from the arguments in the proof of [20, Lemma 2.5] and the proof of (2.14) follows from arguments in [20, Pg. 212]. □

Remark 2.7. Recall Remarks 2.4 and 2.5 on the conditions for a_n satisfying (2.8), which we have assumed throughout the paper. The following are two sufficient conditions under which (2.12) can also be shown to hold (upon direct application of Markov's inequality), but with a condition on a_n:

(i) $\frac{\log n}{a_n} \to \infty$ and $\mathbb{E}\big[M(\vartheta_1)^{2+\epsilon}\big] < \infty$, for some $\epsilon > 0$.

(ii) For some $\beta \in (0,1)$, $n^{\frac{\beta}{2}}/a_n^{2-\beta} \to \infty$ and $\mathbb{E}\big[e^{\widetilde{\rho}M(\vartheta_1)^\beta}\big] < \infty$, for some $\widetilde{\rho} > 0$.

Define, for $t,u \in [0,T]$ and $\delta,\delta' \in \mathbb{R}$,

$$\widehat{G}_2(t,u,\delta,\delta') \doteq \int_{\mathbb{R}^d} \Big(H\Big(t-u+\delta, g(u+\delta',x)\Big) - H\Big(t-u, g(u,x)\Big)\Big)^2 F(dx). \tag{2.15}$$

and for $t,u \in [0,T]$ and $\delta \in \mathbb{R}$, let $\widetilde{G}_2(t,u,\delta) = \widehat{G}_2(t,u,-\delta,0)$.

Assumption 2.4. The following hold.

(i) The total variation of $G_1(t,\cdot)$ is uniformly bounded in $t \in [0,T]$, *i.e.,*

$$\sup_{t\in[0,T]} V_T\Big(G_1(t,\cdot)\Big) < \infty.$$

(ii) Suppose $\delta^n : [0,T] \to \mathbb{R}$ such that $\delta^n(u) \to 0$, uniformly in $[0,T]$ and we have the following.

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \int_0^T \widehat{G}_2(t,u,\delta^n(u),\delta^n(u))du = 0.$$

(iii) For $t \in [0,T]$,

$$\lim_{\delta \downarrow 0} \int_0^T \breve{G}_2(t,u,\delta) du = 0. \tag{2.16}$$

Remark 2.8. Assumption 2.4(i) is only used in proving Proposition 3.2, where the contraction principle can be applied under Assumption 2.4(i). It handles the dependence of $\{\xi_i\}_{n\in\mathbb{N}}$ and $\{\tau_i\}_{n\in\mathbb{N}}$. It is a mild assumption and is satisfied in many practical applications. In particular, consider the case where $\{\xi_i\}$ are independent of $\{\tau_i\}_{i\in\mathbb{N}}$. In this case, the above assumption is satisfied as long as the shot shape function $H(\cdot,x)$ has bounded total variation on $[0,T]$, for every $x \in \mathbb{R}^d$. In particular, the examples discussed in Remark 2.3 satisfy Assumption 2.4(i).

Remark 2.9. Assumption 2.4(ii) is used crucially in the proofs of exponential equivalence (Theorem 3.1) and exponential tightness (Theorem 3.3). Assumption 2.4(iii) is used in the proof of exponential tightness. These are necessary to handle the non-stationarity of the process $\widetilde{X}^n$. These conditions can be easily verified for many examples that we come across in practice. In the very special case when the model has multiplicative response function $H(t,x) = \tilde{H}(t)\varphi(x)$ and when $\{\xi_i\}_{i\in\mathbb{N}}$ and $\{\tau_i\}_{i\in\mathbb{N}}$ are independent *i.e.,* $g(s,x) = \tilde{g}(x)$, Assumption 2.4(ii) becomes

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \int_0^T \Big(\tilde{H}\big(t-u+\delta^n(u)\big)\big) - \tilde{H}\big(t-u\big)\big)\Big)^2 du = 0\,,$$

since the noise component becomes a constant $\int_{\mathbb{R}^d} \tilde{g}(x)F(dx)$, and Assumption 2.4(iii) becomes

$$\lim_{\delta\downarrow 0} \int_0^T \Big(\tilde{H}\big(t-u-\delta\big) - \tilde{H}\big(t-u\big)\Big)^2 du = 0\,.$$

These conditions are reduced to only requiring the above conditions on $\tilde{H}$ (recalling Assumption 2.1 that $\tilde{H}$ is only assumed to be right continuous). In general, this is not the case, which will depend on both functions H and g.

We are now ready to state the main result of the paper.

Theorem 2.2. *Under Assumptions 2.1–2.4 and the conditions on a_n in (2.8), the family of random variables $\{\widetilde{X}^n\}_{n\in\mathbb{N}}$ satisfies the following.*

(i) For every Borel measurable set $\mathtt{A}$ in $(\mathcal{D}_T, J_1)$,

$$-\inf_{\phi\in\mathtt{A}^\circ} I^{MDP}(\phi) \le \liminf_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}(\widetilde{X}^n \in \mathtt{A}) \le \limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}(\widetilde{X}^n \in \mathtt{A}) \le -\inf_{\phi\in\bar{\mathtt{A}}} I^{MDP}(\phi). \tag{2.17}$$

(ii) For $l \ge 0$, $\{\phi : I^{MDP}(\phi) \le l\}$ is a compact set in $(\mathcal{D}_T, J_1)$.

Here,

$$I^{MDP}(\phi) = \inf_{(\phi_1,\phi_2)\in\mathcal{D}_T\times\mathcal{D}_T\colon \phi=\phi_1+\phi_2} \left\{ I_1^{MDP}(\phi_1) + I_2^{MDP}(\phi_2) \right\}, \tag{2.18}$$

with

$$I_1^{MDP}(\phi_1) \doteq \frac{1}{2}\int_0^T \int_0^T z(s)\Lambda(s,t)z(t)dsdt\,, \tag{2.19}$$

$$I_2^{MDP}(\phi_2) \doteq \frac{1}{2\lambda^3\sigma^2} \inf\left\{ \int_0^T |\dot{x}(t)|^2 dt \right\}, \tag{2.20}$$

where $\Lambda(s,t)$ is defined in (2.5) and z is a Lebesgue measurable function on $[0,T]$ such that $\phi_1(\cdot) = \int_0^T z(s)\Lambda(\cdot,s)ds$. If no such z exists, then we take $I_1^{MDP}(\phi_1) = \infty$. The infimum in (2.20) is over $x \in \mathcal{AC}_0$ such that

$$\phi_2(t) = \int_0^t G_1(t,u)dx(u) = x(t)G_1(t,t) - \int_0^t x(u-)d_uG_1(t,u), \quad \text{for} \ \ t\in[0,T].$$

Remark 2.10. From here on, if the family of $\mathcal{D}_T$–valued random variable $\{Z^n\}_{n\in\mathbb{N}}$ satisfies (i) and (ii) with a function I in place of I^{MDP}, then we say that $\{Z^n\}_{n\in\mathbb{N}}$ satisfies a moderate deviation principle (MDP) with rate function I and rate a_n^2.

We now apply Theorem 2.2 to a $GI/G_t/\infty$ queue which has a renewal arrival process and a time-varying service time. In this model, $H:\mathbb{R}_+\times\mathbb{R}_+\to\mathbb{R}_+$ with $H(t,x)=\mathbb{1}_{\{t<x\}}$ (non-multiplicative) and $X(t)$ is the number of customers/jobs in the system (in service) at time t. As a special case of Theorem 2.2, we obtain the following sample-path MDP for the scaled version of queueing process $X(t)$.

Corollary 2.1. *Under Assumptions 2.1–2.4, in the $GI/G_t/\infty$ queueing model, the MDP-scaled queueing process $\widetilde{X}^n$ satisfies a MDP in $(\mathcal{D}_T, J_1)$ with rate a_n^2 and rate function in (2.18), in which*

$$G_1(t,s)=1-F_s(t-s)\doteq F_s^c(t-s),\quad G_2(t,s,u)=1-F_u(t\vee s-u)\doteq F_u^c(t\vee s-u),$$

and

$$\Lambda(t,s)=\lambda\int_0^{t\wedge s}\Big(F_u^c(t\vee s-u)-F_u^c(t-u)F_u^c(s-u)\Big)du.$$

Remark 2.11. Puhalskii [51] recently established sample-path MDP for many-server queues with renewal arrival processes and i.i.d. service times in the so-called Halfin-Whitt regime (the arrival rate/intensity and number of servers are scaled up with fixed service rate in such a way that the system becomes critically loaded). For that model, the MDP for the sequential empirical process associated with the service times is proved and subsequently, the MDP for a process of the form in (2.1) with the arrival process being the entering service process of customers/jobs (τ_i's being the entering service times and hence the arrival process is no longer a renewal process due to the waiting times) and the functions $H(t,x)=\mathbb{1}_{\{t<x\}}$ and $g(t,x)=x$. The approach in that paper is different from ours. The approach in [51] makes use of a semi-martingale decomposition of the sequential empirical process along with the techniques of exponential martingales. That approach however cannot be adapted to general shot noise processes.

Remark 2.12. We remark that the function $\Lambda(t,s)$ is in fact the covariance function of the limit process for the CLT-scaled processes

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{A^n(t)}\left(H(t,\tau_i^n,\vartheta_i)-G_1(t,\tau_i^n)\right).$$

See Theorem 2.2 in [46] for shot noise processes and Theorem 3.2 in [45] for the $GI/G_t/\infty$ queue. The constant $\lambda^3\sigma^2$ is the variance coefficient of the Brownian limit process of the CLT-scaled renewal process

$$\hat{A}^n(t)\doteq\frac{1}{\sqrt{n}}(A^n(t)-\lambda nt)\,. \tag{2.21}$$

See Theorem 17.3 in [5]. These indicate how the rate functions in the MDP are related to the FCLT results.

2.3. **Comparison with the MDP for shot noise processes in a conventional time-space scaling regime.** In this section, we discuss the MDP results in the conventional time-space scaling regime. To that end, we suppose that Assumptions 2.1, 2.2, 2.3 and 2.4(i) still hold. Because of the structure of the scaling, it is necessary to specify how $H(t,x)$ behaves as $t\to\infty$. In particular, we assume that

$$H(\infty,x)\doteq\lim_{t\to\infty}H(t,x)\text{ exists uniformly in }x\in\mathbb{R}^d. \tag{2.22}$$

Consider the following MDP-scaled process with both time and space scalings:

$$\breve{X}^n(t) \doteq \frac{\sqrt{n}}{a_n}\left(\frac{1}{n}\sum_{i=1}^{A^n(t)} H(nt-\tau_i, g(\tau_i,\vartheta_i)) - \frac{\lambda}{n}\int_0^{nt} \mathbb{E}[H(nt-s, g(s,\vartheta_1))]ds\right). \tag{2.23}$$

For $u \in [0,T]$, define

$$G_1^\infty(u) \doteq \int_{\mathbb{R}^d} H(\infty, g(u,x))F(dx)\,,$$
$$G_2^\infty(u) \doteq \int_{\mathbb{R}^d} H(\infty, g(u,x))^2 F(dx)\,.$$

Theorem 2.3. *Under Assumptions mentioned above, the family of $\mathcal{D}_T$–valued random variables $\{\breve{X}^n\}_{n\in\mathbb{N}}$ satisfies an MDP in $(\mathcal{D}_T, J_1)$ with rate a_n^2 and rate function $I_\infty^{MDP}: \mathcal{D}_T \to [0,\infty]$ given by*

$$I_\infty^{MDP}(\phi) = \begin{cases} \frac{1}{2}\int_0^T \frac{|\dot\phi(t)|^2}{\sigma_\infty^2(t)}dt, & \text{whenever } \phi \in \mathcal{AC}_0,\\ \infty, & \text{otherwise,}\end{cases} \tag{2.24}$$

where

$$\sigma_\infty^2(u) \doteq \lambda\Big(G_2^\infty(u) - (G_1^\infty(u))^2\Big) + \lambda^3\sigma^2(G_1^\infty(u))^2\,. \tag{2.25}$$

Remark 2.13. The proof of this theorem can be carried out by taking a similar approach as the proof of Theorem 2.2 (details are omitted for brevity). To begin with, one can show that $\{\breve{X}^n\}_{n\in\mathbb{N}}$ is exponentially equivalent to $\{\breve{\mathcal{X}}^n\}_{n\in\mathbb{N}}$ defined as

$$\breve{\mathcal{X}}^n(t) \doteq \frac{\sqrt{n}}{a_n}\left(\frac{1}{n}\sum_{i=1}^{A^n(t)} H(\infty, g(\tau_i,\vartheta_i)) - \frac{\lambda}{n}\int_0^{nt} G_1^\infty(u)ds\right).$$

To prove this, one requires the uniform convergence condition on $H(t,x)$ as $t\to\infty$ in (2.22) (and its rate of convergence is irrelevant). This process can be decomposed into two processes:

$$\breve{\mathcal{X}}^n(t) = \frac{1}{a_n\sqrt{n}}\sum_{i=1}^{A^n(t)}\Big(H(\infty, g(\tau_i,\vartheta_i)) - G_1^\infty(\tau_i)\Big)$$
$$+ \int_0^t G_1^\infty(s)d\left(\frac{1}{a_n\sqrt{n}}(A(ns)-\lambda ns)\right).$$

These two processes are asymptotically independent, with the first capturing the variabilities in the noises $\{\vartheta_i\}$ and the second capturing the variabilities in the renewal arrival process $A(\cdot)$ (see the relevant discussions below (3.4) in the high intensity scaling regime). One can then prove the MDP for each component and obtain the MDP for $\{\breve{\mathcal{X}}^n\}_{n\in\mathbb{N}}$ with the rate function $I_\infty^{MDP}: \mathcal{D}_T \to [0,\infty]$ given by

$$I_\infty^{\mathrm{MDP}}(\phi) = \inf_{(\phi_1,\phi_2)\in\mathcal{D}_T\times\mathcal{D}_T:\,\phi=\phi_1+\phi_2}\left\{I_{1,\infty}^{\mathrm{MDP}}(\phi_1) + I_{2,\infty}^{\mathrm{MDP}}(\phi_2)\right\}\,, \tag{2.26}$$

where

$$I_{1,\infty}^{\mathrm{MDP}}(\phi_1) \doteq \frac{1}{2\lambda}\int_0^T \frac{|\dot\phi_1(t)|^2}{G_2^\infty(t) - (G_1^\infty(t))^2}dt\,,$$
$$I_{2,\infty}^{\mathrm{MDP}}(\phi_2) \doteq \frac{1}{2\lambda^3\sigma^2}\int_0^T \frac{|\dot\phi_2(t)|^2}{(G_1^\infty(t))^2}dt\,,$$

and the two components in the rate function correspond to the two processes in the decomposition above. Note that the infimum in (2.26) is over a convex functional of ϕ_1 and ϕ_2 subject to a constraint (*viz.*, $\phi = \phi_1 + \phi_2$). This can then be explicitly solved to obtain the desired result in the theorem.

Remark 2.14. We remark on how the rate function is related to the FCLT for the diffusion-scaled process

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{A^n(t)} H(nt-\tau_i,\xi_i) - \frac{\lambda}{n}\int_0^{nt} \mathbb{E}[H(nt-s,g(s,\vartheta_1))]ds\right).$$

Observe that it again be shown to be exponentially equivalent to a process that can be decomposed into two components:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{A^n(t)}\left(H(nt-\tau_i,\xi_i) - G_1^\infty(\tau_i)\right) + \int_0^t G_1^\infty(s)d\hat{A}^n(s)$$

where $\hat{A}^n(t)$ is defined in (2.21). It can be shown (for example, modifying the proofs in [29, 35]) that the diffusion-scaled processes converge in $(\mathcal{D}_T, J_1)$ to

$$B_1\left(\int_0^t \lambda \mathrm{Var}(H(\infty, g(s,\vartheta_1)))ds\right) + B_2\left(\int_0^t \lambda^3\sigma^2(G_1^\infty(s))^2 ds\right)$$

where B_1 and B_2 are two independent Brownian motions. Also, observe that the sum of the two independent Brownian limits is equivalent in distribution to a Brownian motion with the variance function $\int_0^t \sigma_\infty^2(u)du$, for $\sigma_\infty^2(u)$ given in (2.25). The function $\sigma_\infty^2(\cdot)$ is also exactly what appears in the rate function in (2.24).

In the special case of i.i.d. noises with $g(t,x) = x$, the variance coefficient of the Brownian limit reduces to

$$\lambda \mathrm{Var}(H(\infty,\vartheta_1) + \lambda^3\sigma^2(\mathbb{E}[H(\infty,\vartheta_1)])^2.$$

This is the variance coefficient in the FCLT for the compound renewal process $\sum_{i=1}^{A(t)} H(\infty,\vartheta_i)$.

2.4. **Comparison with the LDPs for shot noise processes in the two scaling regimes.** In this section, we discuss the differences of the MDP results above with the LDP results in the two scaling regimes. We first recall the LDP result for renewal processes A^n defined in (2.7).

Theorem 2.4. *[52, Theorem 6.1(b)] Assume that* $\mathbb{E}[e^{\gamma\eta_1}] < \infty$*, for some* $\gamma > 0$*. Let* $\gamma^* \doteq \sup\{\gamma : \mathbb{E}[e^{\gamma\eta_1}] < \infty\}$*. Then* $\{n^{-1}A^n\}_{n\in\mathbb{N}}$ *satisfies LDP in* $(\mathcal{D}_T, J_1)$ *with rate* n *and rate function*

$$I_A^{LDP}(x) = \begin{cases} \int_0^T \varphi(\dot{x}(t))dt, & \textit{whenever } x \in \mathcal{AC}_0, \\ \infty, & \textit{otherwise,} \end{cases}$$

where

$$\varphi(x) \doteq \sup\left\{\gamma x - \log \mathbb{E}[e^{\gamma\eta_1}] : \gamma < \gamma^*\right\}.$$

Remark 2.15. Since $A^n(t) = A(nt)$ in both the conventional scaling and the high intensity regimes, the above LDP result is applicable in both the regimes.

2.4.1. *LDP in the high intensity regime.* Recall that $\bar{X}^n$ was defined in (2.6). We assume that Assumptions 2.1 and 2.2 hold. Moreover, we also assume that the following conditions hold.

$$\mathbb{E}\Big[\exp\Big(\rho \sup_{t,s\in[0,T]} |H(t,s,\vartheta_1)|\Big)\Big] < \infty, \text{ for every } \rho > 0. \tag{2.27}$$

$$\lim_{\delta\downarrow 0}\sup_{t\in[0,T]} \log \mathbb{E}\Big[\exp\Big(\rho \sup_{0\le u\le\delta}\sup_{t\in[0,T]} |H(t+u,s,\vartheta_i) - H(t,s,\vartheta_i)|\Big)\Big] = 0 \tag{2.28}$$

The family of $\mathcal{D}_T$–valued random variables $\{\bar{X}^n\}_{n\in\mathbb{N}}$ can be shown to satisfy the following:

(i) For every Borel measurable set $\mathtt{A}$ in $(\mathcal{D}_T, J_1)$,

$$-\inf_{\phi\in\mathtt{A}^\circ} I^{\mathrm{LDP}}[\bar{X}^n](\phi) \le \liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\bar{X}^n\in\mathtt{A}) \le \limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\bar{X}^n\in\mathtt{A}) \le -\inf_{\phi\in\bar{\mathtt{A}}} I^{\mathrm{LDP}}[\bar{X}^n](\phi).$$

(ii) For $l\ge 0$, $\{\phi : I^{\mathrm{LDP}}[\bar{X}^n](\phi)\le l\}$ is a compact set in $(\mathcal{D}_T, J_1)$.

Here,

$$I^{\mathrm{LDP}}[\bar{X}^n](\phi) = \begin{cases} \sup_{\rho\in\mathcal{C}_T}\int_0^T\Bigg(\dot{\phi}(t)\int_t^T\rho(u)du - \Psi_A\Bigg(\log\mathbb{E}\Bigg[\exp\Bigg(\int_t^T\rho(s)H(s,t,\vartheta_1)ds\Bigg)\Bigg]\Bigg)\Bigg)dt, \\ \qquad\qquad \text{if } \phi\in\mathcal{AC}_0, \\ \infty, \ \text{otherwise,} \end{cases} \tag{2.29}$$

where Ψ_A is defined as

$$\Psi_A(\rho) \doteq \lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[e^{\rho A^n(1)}].$$

From Theorem 1 in [25], we have

$$\Psi_A(\rho) = -\psi_\eta^{-1}(-\rho) \quad \text{with} \quad \psi_\eta(\rho)\doteq\log\mathbb{E}[e^{\rho\eta_1}]. \tag{2.30}$$

Remark 2.16. In what follows, whenever a family of $\mathcal{D}_T$–valued random variables $\{Z^n\}_{n\in\mathbb{N}}$ satisfy above conditions (i) and (ii) with function I in place of $I^{\mathrm{LDP}}[\bar{X}^n]$, we say that $\{Z^n\}_{n\in\mathbb{N}}$ satisfies a large deviation principle with rate function I and rate n.

In the special case of $GI/G_t/\infty$ queueing model, with $\xi_i = g(\tau_i,\vartheta_i)$ and $H(t,x) = \mathbb{1}_{\{t<x\}}$, the rate function in the LDP becomes

$$I^{\mathrm{LDP}}[\bar{X}^n](\phi) = \begin{cases} \sup_{\rho\in\mathcal{C}_T}\int_0^T\Bigg(\dot{\phi}(t)\int_t^T\rho(u)du - \Psi_A\Bigg(\log\mathbb{E}\Bigg[\exp\Bigg(\int_t^{T\wedge(t+g(t,\vartheta_1))}\rho(s)ds\Bigg)\Bigg]\Bigg)\Bigg)dt, \\ \qquad\qquad \text{if } \phi\in\mathcal{AC}_0, \\ \infty, \ \text{otherwise.} \end{cases}$$

Note that in [24], the LDP for the $GI/GI/\infty$ queue is established for the queueing process $X(t)$ at each fixed time $t=\hat{t}\in[0,T]$. (That result can be formally obtained by setting $\rho(\cdot)=\hat{\rho}\delta_{\hat{t}}(\cdot)$ for some constant $\hat{\rho}$.) Our result extends that to a model with non-stationary service times and the LDP result is in the sample-path sense (see also [6] for the sample-path LDP is established for a two-parameter process tracking the elapsed/residual service times in $GI/GI/\infty$ queues).

We also remark that the proof of the sample-path LDP for the scaled shot noise process $\bar{X}^n(t)$ differs from that of the MDP for the processes $\widetilde{X}^n(t)$ in (2.9). The main difference between LDP and MDP is similar to that of LLN and CLT *viz.,* centering is necessary for MDP and CLT. Although one can follow a similar approach in Section 3, the proof of finite-dimensional LDP would require appropriate limits of log moment generating of $A^n(\cdot)$ and ϑ_1, and the proof of exponential tightness could be modified appropriately without the centering terms. In Appendix B, we provide a sketch of the proofs of these results.

2.4.2. *LDP in the conventional time-space scaling regime.* Next, for the shot noise process $X(t)$ in (2.1), define the following scaled process

$$\bar{\bar{X}}^n(t) \doteq \frac{1}{n}\sum_{i=1}^{A(nt)} H(nt-\tau_i,\xi_i). \tag{2.31}$$

Here, we again assume that ξ_i is of the form $g(\tau_i,\vartheta_i)$, for a family of i.i.d random variables $\{\vartheta_i\}_{i\in\mathbb{N}}$ and that $H(t,x)\to H(\infty,x)$ as $t\to\infty$, uniformly in $x\in\mathbb{R}^d$. The LDP result gives the rate function:

$$I^{\mathrm{LDP}}[\bar{\bar{X}}^n](\phi) = \begin{cases} \sup_{\rho\in\mathcal{C}_T}\int_0^T\bigg(\dot\phi(t)\rho(t)-\Psi_A\bigg(\log\mathbb{E}\bigg[\exp\bigg(\rho(t)H(\infty,t,\vartheta_1)\bigg)\bigg]\bigg)\bigg)dt, \\ \qquad\qquad \text{if } \phi\in\mathcal{AC}_0, \\ \infty, \ \text{otherwise}. \end{cases} \tag{2.32}$$

This can be established by using an argument similar to the one that is sketched in Appendix B.

In the case of i.i.d. noises with $g(t,x)=x$, the LDP is like that for the compound renewal process as studied in [8]. However, with the general function $g(t,x)$, the LDP is an extension to compound renewal processes with non-stationary compound variables (arrival time dependent distributions).

In addition, in the special case of Poisson arrivals and i.i.d. noises, the LDP result coincides with that in [22] and also in [15] when their model does not have state-dependent component in H. In this case, the rate function is

$$I^{\mathrm{LDP}}[\bar{\bar{X}}^n](\phi) = \begin{cases} \int_0^T \Lambda^*(\dot\phi(t))dt, & \text{if } \phi\in\mathcal{AC}_0, \\ \infty, & \text{otherwise}, \end{cases} \tag{2.33}$$

where

$$\Lambda_{H(\infty,\xi)}(\theta)=\log\mathbb{E}[\exp(\theta H(\infty,\xi_1))], \quad \Lambda(\theta)=\lambda\Big(\exp\Big(\Lambda_{H(\infty,\xi)}(\theta)\Big)-1\Big)$$

and $\Lambda^*(x)$ is the Legendre transform of Λ, that is,

$$\Lambda^*(x)=\sup_{\theta\in\mathbb{R}}\{\theta x-\Lambda(\theta)\}.$$

One can easily see how the rate function in (2.32) reduces to that in (2.33) by noting $\Psi_A(\rho)=\lambda(e^\rho-1)$ for Poisson arrival process $A(t)$. However, even with Poisson arrivals, for non-stationary noises, one cannot simplify the rate function in (2.32) except using $\Psi_A(\rho)=\lambda(e^\rho-1)$.

3. Proof of Theorem 2.2

In this section we prove the sample-path MDP for $\widetilde{X}^n(t)$ in (2.9) as stated in Theorem 2.2. First, we observe that the process $\widetilde{X}^n$ can be decomposed into two processes:

$$\widetilde{X}^n(t)=\widetilde{X}_1^n(t)+\widetilde{X}_2^n(t), \tag{3.1}$$

where

$$\widetilde{X}_1^n(t)\doteq\frac{1}{a_n\sqrt{n}}\sum_{i=1}^{A^n(t)}\widetilde{H}(t,\tau_i^n,\vartheta_i)\,, \tag{3.2}$$

$$\widetilde{X}_2^n(t)\doteq\int_0^t G_1(t,s)d\widetilde{A}^n(s)\,, \tag{3.3}$$

with

$$\widetilde{H}(t,s,x)\doteq H(t,s,x)-G_1(t,s), \tag{3.4}$$

and $\widetilde{A}^n$ is as defined in (2.10).

Before proceeding to the proof, we provide an overview of the proof strategy. From the statement of the rate function in Theorem 2.2, as a sum of the two rate functions associated with the rate functions of $\widetilde{X}_1^n(t)$ and $\widetilde{X}_2^n(t)$, it may appear that the rate function comes from two independent processes. However, the two processes $\widetilde{X}_1^n(t)$ and $\widetilde{X}_2^n(t)$ are not independent since both depend on the arrival process $A^n(t)$. On the other hand, because of the independence of $\eta_1, \eta_2, \dots$ and $\vartheta_1, \vartheta_2, \dots$, the two processes $\widetilde{X}_1^n(t)$ and $\widetilde{X}_2^n(t)$ are asymptotically independent as $n \to \infty$, in the sense that the variability of $\widetilde{X}_1^n(t)$ (as given in $\Lambda(t,s)$ in (2.5)) only depends on the arrival rate λ of the renewal process $A^n(t)$, and comes rather from the variability of the noises, while the variability of $\widetilde{X}_2^n(t)$ depends on the variability of the interarrival times of renewal process $A^n(t)$ (as in the associated FCLT). (This is also the case in the FCLT result for the shot noise processes in the high intensity regime [46].) Therefore, we construct another process $\{\widehat{X}_1^n\}$ (see (3.5) below), which is exponentially equivalent (see Theorem 3.1) to $\{\widetilde{X}_1^n\}$, and more importantly, independent of $\{\widetilde{X}_2^n\}$. Thus, by [18, Theorem 4.2.13], instead of establishing the MDP of $\{(\widetilde{X}_1^n, \widetilde{X}_2^n)\}_{n\in\mathbb{N}}$, it then suffices to establish the MDPs of $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ and $\{\widetilde{X}_2^n\}$ separately.

The proof of the sample-path MDP of $\{\widehat{X}_1^n\}$ in $\mathcal{D}_T$ uses Gärtner–Ellis theorem [18, Theorem 2.3.6] and Dawson–Gärtner theorem [18, Theorem 4.6.1]. We first establish the MDP in the topology of pointwise convergence and then strengthen it to be in the Skorohod J_1 topology. To be more elaborate, we first prove the MDP for the finite dimensional distributions (see Lemma 3.2) by considering appropriate limits of the log-moment generating function (see Proposition 3.1). Using this, we arrive at the sample-path MDP of $\widehat{X}_1^n$ in topology of pointwise convergence by invoking Dawson–Gärtner theorem. We next establish exponential tightness to arrive at the desired MDP in $(\mathcal{D}_T, J_1)$. The MDP of $\{\widetilde{X}_2^n\}_{n\in\mathbb{N}}$ is derived using the contraction principle and the existing MDP for renewal processes (Theorem 2.1).

3.1. **Exponential equivalence of $\widetilde{X}_1^n$ and $\widehat{X}_1^n$.** To prove this, we introduce a new process that involves an appropriate truncation. This truncation is in such a way that as $n \to \infty$, truncation parameter also goes to infinity. This truncation technique is borrowed from [20] (see Theorem 2.2 of that paper). Fix $l > 0$ and define

$$\widehat{X}_1^n(t) \doteq \frac{1}{a_n\sqrt{n}} \sum_{i=1}^{\lfloor \lambda nt \rfloor} \underline{H}^n(t, s_i^n, \vartheta_i), \tag{3.5}$$

where $\underline{H}^n(t,s,x) \doteq \widetilde{H}^n(t,s,x) - \mathbb{E}\Big[\widetilde{H}^n(t,s,\vartheta_1)\Big]$ with

$$\widetilde{H}^n(t,s,x) \doteq H(t,s,x)\mathbb{1}_{\{|H(t,s,x)| \le l\sqrt{n}a_n^{-1}\}}.$$

In the above, $s_i^n \doteq \frac{i}{\lambda n}$. Note that comparing with $\widetilde{X}_1^n$, in the definition of $\widehat{X}_1^n$, we have replaced the arrival times $\tau_i^n = \frac{\tau_i}{n}$ by $s_i^n = \frac{i}{\lambda n}$ and $A^n(t)$ by $\lfloor \lambda nt \rfloor$, and thus removed the randomness from the arrival process A^n. We prove that $\widehat{X}_1^n$ is exponentially equivalent to $\widetilde{X}_1^n$ in $(\mathcal{D}_T, J_1)$ below.

We give an important auxiliary result which is often used in the rest of the paper.

Lemma 3.1. *The following hold. For $t_1, t_2, s_1, s_2 \in [0,T]$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{\lfloor \lambda nt_1 \rfloor} \int_{\mathbb{R}^d} \underline{H}^n(t_1, s_i^n, x)^2 F(dx) = \lambda \int_0^{t_1} \int_{\mathbb{R}^d} \widetilde{H}(t_1, s, x)^2 F(dx) ds\,, \tag{3.6}$$

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}\Bigg[\sum_{i=1}^{\lfloor \lambda nt_1 \rfloor} \sum_{j=1}^{\lfloor \lambda nt_2 \rfloor} \underline{H}^n(t_1, s_i^n, \vartheta_i) \underline{H}^n(t_2, s_j^n, \vartheta_j) \Bigg] = \Lambda(t_1, t_2)\,, \tag{3.7}$$

$$\limsup_{n\to\infty} \frac{1}{n} \mathbb{E}\Bigg[\sum_{i=1}^{\lfloor \lambda nT \rfloor} \int_{\mathbb{R}^d} \Big(\underline{H}^n(t_1, s_1, x) - \underline{H}^n(t_2, s_2, x)\Big)^2 F(dx) \Bigg]$$

$$= \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}\bigg[\sum_{i=1}^{\lfloor \lambda nT\rfloor} \int_{\mathbb{R}^d} \Big(\widetilde{H}(t_1,s_1,x) - \widetilde{H}(t_2,s_2,x)\Big)^2 F(dx)\bigg]. \tag{3.8}$$

Proof. We begin by recalling that

$$\underline{H}^n(t,s,x) = \widetilde{H}^n(t,s,x) - \mathbb{E}[\widetilde{H}^n(t,s,\vartheta_1)].$$

Clearly, $\widetilde{H}^n(t,s,x)$ and $\mathbb{E}[\widetilde{H}^n(t,s,\vartheta_1)]$ increase to $H(t,s,x)$ and $\mathbb{E}[H(t,s,\vartheta_1)]$, respectively as $n\to\infty$. Consider

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t,s\in[0,T]} |\underline{H}^n(t,s,\vartheta_1) - \widetilde{H}(t,s,\vartheta_1)|^2\Big]\\
&= \mathbb{E}\Big[\sup_{t,s\in[0,T]} \Big|H(t,s,\vartheta_1)\mathbb{1}_{\{|H(t,s,x)|>l\sqrt{n}a_n^{-1}\}} - \mathbb{E}\Big[H(t,s,\vartheta_1)\mathbb{1}_{\{|H(t,s,x)|>l\sqrt{n}a_n^{-1}\}}\Big]\Big|^2\Big]\\
&\le 2\mathbb{E}\Big[\sup_{t,s\in[0,T]} \Big|H(t,s,\vartheta_1)\mathbb{1}_{\{|H(t,s,x)|>l\sqrt{n}a_n^{-1}\}}\Big|^2\Big] + \sup_{t,s\in[0,T]} \mathbb{E}\Big[H(t,s,\vartheta_1)\mathbb{1}_{\{|H(t,s,x)|>l\sqrt{n}a_n^{-1}\}}\Big]\Big|^2.
\end{aligned}\tag{3.9}$$

Since using Lemma 2.1 we know that

$$\mathbb{E}\Big[\sup_{t,s\in[0,T]} |H(t,s,\vartheta_1)|^2\Big] < \infty,$$

taking $n\to\infty$, terms in (3.9) goes to zero which in turn, implies

$$\lim_{n\to\infty} \mathbb{E}\Big[\sup_{t,s\in[0,T]} |\underline{H}^n(t,s,\vartheta_1) - \widetilde{H}(t,s,\vartheta_1)|^2\Big] = 0$$

Below, we only prove (3.6) as the proofs of (3.7) and (3.8) follow on the similar lines. From the above display, for every $\epsilon > 0$, we can choose n large enough (uniformly in $t\in[0,T]$ and $1\le i\le \lfloor\lambda nt_1\rfloor$) such that

$$\Big|\int_{\mathbb{R}^d} \underline{H}^n(t_1,s_i^n,x)^2F(dx) - \int_{\mathbb{R}^d} \widetilde{H}(t_1,s_i^n,x)^2F(dx)\Big| < \epsilon.$$

Therefore, using the Lebesgue measurability of $\widetilde{H}$, we have

$$\begin{aligned}
&\Big|\frac{1}{n}\sum_{i=1}^{\lfloor\lambda nt_1\rfloor}\int_{\mathbb{R}^d} \underline{H}^n(t_1,s_i^n,x)^2F(dx) - \lambda\int_0^{t_1}\int_{\mathbb{R}^d}\widetilde{H}(t_1,s,x)^2F(dx)ds\Big|\\
&\le \Big|\frac{1}{n}\sum_{i=1}^{\lfloor\lambda nt_1\rfloor}\int_{\mathbb{R}^d} \underline{H}^n(t_1,s_i^n,x)^2F(dx) - \frac{1}{n}\sum_{i=1}^{\lfloor\lambda nt_1\rfloor}\int_{\mathbb{R}^d}\widetilde{H}(t_1,s_i^n,x)^2F(dx)\Big|\\
&\quad + \Big|\frac{1}{n}\sum_{i=1}^{\lfloor\lambda nt_1\rfloor}\int_{\mathbb{R}^d}\widetilde{H}(t_1,s_i^n,x)^2F(dx) - \lambda\int_0^{t_1}\int_{\mathbb{R}^d}\widetilde{H}(t_1,s,x)^2F(dx)ds\Big|
\end{aligned}$$

Now taking $n\to\infty$ and then taking $\epsilon\downarrow 0$, (3.6) follows. □

Theorem 3.1. *Under Assumptions 2.1, 2.2, 2.3 and 2.4(ii), for any $\delta>0$,*

$$\lim_{n\to\infty}\frac{1}{a_n^2}\log\mathbb{P}\left(\|\widetilde{X}_1^n - \widehat{X}_1^n\|_T > \delta\right) = -\infty.$$

Remark 3.1. Using Theorem 3.1 and Theorem 4.2.13 of [18], we can conclude the MDP of $\{\widetilde{X}_1^n\}_{n\in\mathbb{N}}$ in $(\mathcal{D}_T, J_1)$ from the MDP of $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ in $(\mathcal{D}_T, J_1)$.

The proof is given in Section 4.

3.2. **MDP for the finite dimensional distributions of $\widehat{X}_1^n$.**

Proposition 3.1. *Under Assumption 2.3, for $N \geq 1$ and $0 < t_1 < t_2 < t_3 < \ldots < t_N \leq T$, we have*

$$\lim_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{E}\left[\exp\left(a_n^2 \sum_{m=1}^{N} \rho_m \widehat{X}_1^n(t_m)\right)\right] = \frac{1}{2}\sum_{i,j=1}^{N} \rho_i\rho_j \Lambda(t_i,t_j) \tag{3.10}$$

for every $\{\rho_m\}_{m=1}^N \subset \mathbb{R}$, where $\Lambda(\cdot,\cdot)$ is defined in (2.5). In particular, for $t \in [0,T]$ and $\rho \in \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{E}\left[\exp\left(a_n^2 \rho \widehat{X}_1^n(t)\right)\right] = \frac{1}{2}\rho^2 \Lambda(t,t).$$

Proof. Recall the definition of $\widehat{X}_1^n$:

$$\widehat{X}_1^n(t) = \frac{1}{a_n\sqrt{n}} \sum_{i=1}^{\lfloor \lambda nt \rfloor} \underline{H}^n(t, s_i^n, \vartheta_i), \text{ for } t \in [0,T],$$

where $\{\vartheta_i\}_{i\in\mathbb{N}}$ is a family of $\mathbb{R}^d$–valued i.i.d. random variables distributed according to F.

Since for every n, $\underline{H}^n(t,s,x)$ is uniformly bounded in $(t,s,x) \in \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^d$, we infer that $\mathbb{E}\Big[\exp\left(a_n^2 \rho \widehat{X}_1^n(t)\right)\Big]$ is finite for every $\rho \in \mathbb{R}$, $t \in [0,T]$ and $n \in \mathbb{N}$. Therefore, the following series expansion holds:

$$\begin{aligned}\mathbb{E}\left[\exp\left(a_n^2 \rho \widehat{X}_1^n(t)\right)\right] &= \mathbb{E}\left[\exp\left(\rho \frac{a_n}{\sqrt{n}} \sum_{i=1}^{\lfloor \lambda nt \rfloor} \underline{H}^n(t, s_i^n, \vartheta_i)\right)\right] \\ &= \sum_{k=0}^{\infty} \frac{\rho^k}{k!}\left(\frac{a_n}{\sqrt{n}}\right)^k \mathbb{E}\left[\left(\sum_{i=1}^{\lfloor \lambda nt \rfloor} \underline{H}^n(t, s_i^n, \vartheta_i)\right)^k\right].\end{aligned}$$

Note that the linear term in the above equation is zero as $\mathbb{E}[\underline{H}^n(t, s_i^n, \vartheta_i)] = 0$. Since $\{\vartheta_i\}_{i\in\mathbb{N}}$ is an i.i.d. sequence, using the definition of $\underline{H}^n(t, s_i^n, \vartheta_i)$ we have

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{i=1}^{\lfloor \lambda nt \rfloor} \underline{H}^n(t, s_i^n, \vartheta_i)\right)^2\right] &= \sum_{i=1}^{\lfloor \lambda nt \rfloor} \int_{\mathbb{R}^d} \underline{H}^n(t, s_i^n, x)^2 F(dx) \\ \mathbb{E}\left[\left(\sum_{i=1}^{\lfloor \lambda nt \rfloor} \underline{H}^n(t, s_i^n, \vartheta_i)\right)^3\right] &= \sum_{i=1}^{\lfloor \lambda nt \rfloor} \int_{\mathbb{R}^d} \underline{H}^n(t, s_i^n, x)^3 F(dx).\end{aligned} \tag{3.11}$$

Taking the logarithm on both sides of the above and then applying the Taylor's theorem to the function $\log(1+x)$, we get

$$\begin{aligned}&\frac{1}{a_n^2} \log\mathbb{E}\left[\exp\left(a_n^2 \rho \widehat{X}_1^n(t)\right)\right] \\ &\quad = \frac{1}{a_n^2}\frac{\rho^2 a_n^2}{2n}\mathbb{E}\left[\sum_{i=1}^{\lfloor \lambda nt \rfloor} \underline{H}^n(t, s_i^n, \vartheta_i)^2\right] + \mathcal{O}\Big(|\mathbb{E}[\sup_{t,s\in[0,T]} \underline{H}^n(t,s,\vartheta_1)^3]| a_n n^{-\frac{1}{2}}\Big).\end{aligned} \tag{3.12}$$

Recall (2.14). Now from (3.12) and Lemma 2.1, taking $n \to \infty$, we obtain

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{E}\left[\exp\left(a_n^2 \rho \widehat{X}_1^n(t)\right)\right] &= \lim_{n\to\infty} \frac{1}{n}\frac{\rho^2}{2} \sum_{i=1}^{\lfloor \lambda nt \rfloor} \int_{\mathbb{R}^d} \underline{H}^n(t, s_i^n, x)^2 F(dx) \\ &= \frac{\rho^2 \lambda}{2} \int_0^t \int_{\mathbb{R}^d} \widetilde{H}(t,s,x)^2 ds F(dx) \\ &= \frac{\rho^2}{2}\Lambda(t,t).\end{aligned}$$

Here the second equality follows from (3.6) of Lemma 3.1 and the last equation follows from the definition of $\Lambda(\cdot,\cdot)$.

To prove (3.10) in the finite-dimensional case, we use a similar argument. Using the fact that $\underline{H}^n$, for every n uniformly bounded in all the arguments and applying Taylor's theorem to the function $\exp(x)$, we have

$$\begin{aligned}
&\mathbb{E}\Bigg[\exp\Bigg(\frac{a_n}{\sqrt{n}}\sum_{m=1}^{N}\rho_m\sum_{i=1}^{\lfloor\lambda n t_m\rfloor}\underline{H}^n(t_m,s_i^n,\vartheta_i)\Bigg)\Bigg]\\
&\quad = 1+\mathbb{E}\Bigg[\Bigg(\frac{a_n}{\sqrt{n}}\sum_{m=1}^{N}\rho_m\sum_{i=1}^{\lfloor\lambda n t_m\rfloor}\underline{H}^n(t_m,s_i^n,\vartheta_i)\Bigg)^2\Bigg]+\mathcal{O}\Big(\mathbb{E}\Big[\sup_{t,s\in[0,T]}|\underline{H}^n(t,s,\vartheta_1)|^3\Big]a_n n^{-\frac{1}{2}}\Big), \qquad (3.13)\\
&\quad = 1+\frac{a_n^2}{n}\sum_{k,m=1}^{N}\rho_m\rho_k\mathbb{E}\Bigg[\sum_{i=1}^{\lfloor\lambda n t_m\rfloor}\sum_{j=1}^{\lfloor\lambda n t_k\rfloor}\underline{H}^n(t_m,s_i^n,\vartheta_i)\underline{H}^n(t_k,s_j^n,\vartheta_j)\Bigg]\\
&\qquad +\mathcal{O}\Big(\mathbb{E}\Big[\sup_{t,s\in[0,T]}|\underline{H}^n(t,s,\vartheta_1)|^3\Big]a_n n^{-\frac{1}{2}}\Big). \qquad (3.14)
\end{aligned}$$

In the above, we again use the fact that $\mathbb{E}[\underline{H}^n(t_m,s_i^n,\vartheta_i)]=0$ and $\{\vartheta_i\}_{i\in\mathbb{N}}$ is an i.i.d. sequence. Taking the limit as $n\to\infty$, we get

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{a_n^2}\log\mathbb{E}\Bigg[\exp\Bigg(\frac{a_n}{\sqrt{n}}\sum_{m=1}^{N}\rho_m\sum_{i=1}^{\lfloor\lambda n t_m\rfloor}\underline{H}^n(t_m,s_i^n,\vartheta_i)\Bigg)\Bigg]\\
&\qquad = \lim_{n\to\infty}\frac{1}{n}\sum_{k,m=1}^{N}\rho_m\rho_k\sum_{i=1}^{\lfloor\lambda n t_m\rfloor\wedge\lfloor\lambda n t_k\rfloor}\big(G_2(t_m,t_k,s_i^n)-G_1(t_m,s_i^n)G_1(t_k,s_i^n)\big) \qquad (3.15)\\
&\qquad = \frac{1}{2}\sum_{k,m=1}^{N}\rho_m\rho_k\Lambda(t_m,t_k),\ \text{from the definition of } \Lambda(\cdot,\cdot). \qquad (3.16)
\end{aligned}$$

In the above, we get (3.15) after using (3.7) of Lemma 3.1 and also using the fact that the term in (3.14) goes to zero as $n\to\infty$, using Lemma 2.1. This completes the proof of the lemma. $\square$

The following lemma gives us the MDP of $\{\widehat{X}^n_{1,N}\}_{n\in\mathbb{N}}\doteq\{\widehat{X}^n_1(t_1),\widehat{X}^n_1(t_2),\widehat{X}^n_1(t_3),\ldots,\widehat{X}^n_1(t_N)\}_{n\in\mathbb{N}}$, for every $N\in\mathbb{N}$ and $0<t_1<t_2<t_2<t_3<\ldots<t_N\leq T$.

Lemma 3.2. *Under Assumption 2.3, the family of $\mathbb{R}^N$–valued random variables $\{\widehat{X}^n_{1,N}\}_{n\in\mathbb{N}}$ satisfies the following:*

(i) For every Borel measurable set $\mathtt{A}$ in $\mathbb{R}^N$,

$$-\inf_{x\in\mathtt{A}^\circ}I_f^N(x)\leq\liminf_{n\to\infty}\frac{1}{a_n^2}\log\mathbb{P}(\widehat{X}^n_{1,N}\in\mathtt{A})\leq\limsup_{n\to\infty}\frac{1}{a_n^2}\log\mathbb{P}(\widehat{X}^n_{1,N}\in\mathtt{A})\leq-\inf_{x\in\bar{\mathtt{A}}}I_f^N(x). \qquad (3.17)$$

(ii) For $l\geq 0$, $\{x: I_f^N(x)\leq l\}$ is a compact set in $\mathbb{R}^N$.

Here, $I_f^N:\mathbb{R}^N\to[0,\infty]$ is given by

$$I_f^N(\underline{x})=\frac{1}{2}\langle\underline{x},\hat{\Lambda}^{-1}\underline{x}\rangle, \qquad (3.18)$$

where $\{\hat{\Lambda}_{ij}\}\doteq\{\Lambda(t_i,t_j)\}$ for $1\leq i,j\leq N$ and $\Lambda(\cdot,\cdot)$ is as defined in (2.5).

Proof. From Proposition 3.1, it is clear that

$$\chi(\underline{\rho})\doteq\lim_{n\to\infty}\frac{1}{a_n^2}\log\mathbb{E}\left[\exp\left(a_n^2\sum_{m=1}^{N}\rho_m\widehat{X}^n_1(t_m)\right)\right]=\frac{1}{2}\langle\underline{\rho},\hat{\Lambda}\underline{\rho}\rangle$$

for every $\underline{\rho} \in \mathbb{R}^N$. The Fenchel-Legendre transform of $\chi(\cdot)$ is given by

$$\chi^*(\underline{x}) \doteq \sup_{\underline{\rho} \in \mathbb{R}^N} \big(\langle \underline{x}, \underline{\rho} \rangle - \chi(\underline{\rho}) \big) = \frac{1}{2} \langle \underline{x}, \hat{\Lambda}^{-1} \underline{x} \rangle. \tag{3.19}$$

The second equality follows from a simple calculation. Now applying Theorem 2.3.6 in [18], we have the result with $I_f^N = \chi^*$ after noting that statement in (ii) is clearly true. □

3.3. **Sample-path MDP of $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ in $\mathcal{D}_T$ with the topology of pointwise convergence.** We next extend the above finite-dimensional MDP result to the sample-path MDP of $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ in $\mathcal{D}_T$ endowed with the topology of pointwise convergence. The form of I_f^N makes it difficult in taking $N \to \infty$. To overcome this difficulty, we simplify (3.18) in the following way: Instead of solving the optimization problem

$$\sup_{\underline{\rho} \in \mathbb{R}^N} \big(\langle \underline{x}, \underline{\rho} \rangle - \chi(\underline{\rho}) \big)$$

explicitly in x using calculus (say) which can be difficult to perform in the case of infinite dimensions, we solve it using the property of Euclidean inner product $\langle \cdot, \cdot \rangle$. The advantage in doing so is that this method works even when $\langle \cdot, \cdot \rangle$ denotes a general inner product. With this observation in mind, we first note that $\hat{\Lambda}$ is positive definite. Therefore, for every $\underline{x}$, there is a unique $\underline{v} \in \mathbb{R}^N$ such that $\underline{x} = \hat{\Lambda}\underline{v}$. We now solve the aforementioned optimization problem using just the properties of Euclidean inner product $\langle \cdot, \cdot \rangle$:

$$\begin{aligned}
\langle \underline{x}, \underline{\rho} \rangle - \frac{1}{2} \langle \underline{\rho}, \hat{\Lambda} \underline{\rho} \rangle &= \langle \hat{\Lambda} \underline{v}, \underline{\rho} \rangle - \frac{1}{2} \langle \underline{\rho}, \hat{\Lambda} \underline{\rho} \rangle \\
&= \langle \underline{v}, \hat{\Lambda} \underline{\rho} \rangle - \frac{1}{2} \langle \underline{\rho}, \hat{\Lambda} \underline{\rho} \rangle \\
&= \frac{1}{2} \langle \underline{v}, \hat{\Lambda} \underline{v} \rangle - \frac{1}{2} \langle (\underline{\rho} - \underline{v}), \hat{\Lambda} (\underline{\rho} - \underline{v}) \rangle.
\end{aligned}$$

where we have used the fact that $\hat{\Lambda}$ is symmetric to get the second equality. From above, it is clear that the supremum in (3.19) occurs when $\underline{\rho} = \underline{v}$ and the value is

$$\frac{1}{2} \langle \underline{v}, \hat{\Lambda} \underline{v} \rangle.$$

Clearly, as mentioned above, this method is robust enough to be applied for an infinite dimensional Hilbert space. This is the motivation for the construction of a reproducing kernel Hilbert space in the proof of the Theorem 3.2 below.

We are now in a position to state the sample-path MDP of $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ in $\mathcal{D}_T$ with the topology of pointwise convergence. Recall that $\{f_n\}_{n\in\mathbb{N}} \subset \mathcal{D}_T$ converges to $f \in \mathcal{D}_T$ as $n \to \infty$ in this topology if $f_n(t) \to f(t)$, as $n \to \infty$, for $t \in [0, T]$.

Theorem 3.2. *Suppose Assumptions 2.1, 2.3 hold. Then the family of $\mathcal{D}_T$–valued random variables $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ satisfies an MDP in the topology of pointwise convergence with rate a_n^2 and rate function $I_1^{MDP} : \mathcal{D}_T \to [0, \infty]$ given by (2.19).*

Proof. Using Lemma 3.2 and Theorem 4.6.1 in [18], we have the MDP of $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ with rate a_n^2 and rate function $I^{\mathrm{MDP}} : \mathcal{D}_T \to [0, \infty]$ given by

$$\begin{aligned}
I_1^{\mathrm{MDP}}(x) &\doteq \sup_{0<t_1<t_2<\ldots<t_N \le T} I_f^N(x(t_1), x(t_2), \ldots, x(t_N)) \\
&= \frac{1}{2} \sup_{0<t_1<t_2<\ldots<t_N \le T} \sum_{i,j=1}^{N} v_i^N \Lambda(t_i, t_j) v_j^N,
\end{aligned} \tag{3.20}$$

where v_j^N's are such that

$$x(t_i) = \sum_{j=1}^{N} \Lambda(t_i, t_j) v_j^N, \text{ for } 1 \le i \le N.$$

The rest of the proof is to show that the right hand sides of (2.19) and (3.20) are equal, when they are finite. To that end, we interpolate $(x(t_1), x(t_2), \ldots, x(t_N))$ in the following way:

$$\tilde{x}^N(t) \doteq \sum_{j=1}^{N} \Lambda(t, t_j) v_j^N. \tag{3.21}$$

The reason for doing this is that it lies in the Hilbert space defined below (in particular, a reproducing kernel Hilbert space with norm denoted by $\|\cdot\|_\Lambda$),

$$\|\tilde{x}^N\|_\Lambda = \sum_{i,j=1}^{N} v_i^N \Lambda(t_i, t_j) v_j^N.$$

We now construct the aforementioned Hilbert space. This is a standard construction [2, Sections 1.2 and 1.3], but we give it nonetheless for completeness. From the definition of $\Lambda(\cdot,\cdot)$, it is clear that it is positive definite, *i.e.,* for any $v \in \mathbb{R}^k$, $k \in \mathbb{N}$,

$$\sum_{i,j=1}^{k} v_i v_j \Lambda(t_i, t_j) \ge 0,$$

for any $(t_1, t_2, \ldots, t_k) \in [0,T]^k$. Now consider the span (say $\mathcal{H}_0$) of all functions of the form $f : [0,T] \to \mathbb{R}$,

$$f(t) = \sum_{i=1}^{K} a_j \Lambda(t_j, t),$$

where $K \in \mathbb{N}$, $(a_1, a_2, a_3 \ldots, a_K) \in \mathbb{R}^K$ and $(t_1, t_2, \ldots, t_K) \in [0,T]^K$. Now we define an inner product on $\mathcal{H}_0$ in the following way: for $f, g \in \mathcal{H}_0$ and $(s_1, s_2, \ldots, s_J) \in [0,T]^J$ with

$$f = \sum_{i=1}^{K} a_i \Lambda(t_i, t) \quad \text{and} \quad g = \sum_{j=1}^{J} b_j \Lambda(s_j, t),$$

$$\langle f, g\rangle_\Lambda \doteq \sum_{i=1}^{K} \sum_{j=1}^{J} a_i b_j \Lambda(t_i, s_j).$$

Finally, we get the reproducing kernel Hilbert space (say $\mathcal{H}$) by completing $\mathcal{H}_0$ under $\langle\cdot,\cdot\rangle_\Lambda$ [2, Section 1.2 and 1.3].

Suppose $x \in \mathcal{H}_0$. Then from the definition of $\mathcal{H}_0$, we know

$$x(t) = \sum_{i=1}^{N} v_j^N \Lambda(t, t_i), \quad \text{for some } N \text{ and } (v_1^N, v_2^N, \ldots, v_N^N) \in \mathbb{R}^N.$$

Therefore,

$$\|x\|_\Lambda = \sum_{i,j=1}^{N} v_i^N \Lambda(t_i, t_j) v_j^N = \sum_{i,j=1}^{\tilde{N}} \tilde{v}_i^{\tilde{N}} \Lambda(t_i, t_j) \tilde{v}_j^{\tilde{N}},$$

for any other representation of $x(\cdot)$ such that

$$x(t) = \sum_{i=1}^{\tilde{N}} \tilde{v}_j^{\tilde{N}} \Lambda(t, t_i).$$

This in turn means that for $x \in \mathcal{H}_0$,

$$I_1^{\text{MDP}}(x) = \frac{1}{2} \sup_{0<t_1<t_2<\ldots<t_N \le T} \sum_{i,j=1}^{N} v_i^N \Lambda(t_i, t_j) v_j^N$$
$$= \frac{1}{2} \sup_{0<t_1<t_2<\ldots<t_N \le T} \|x\|_\Lambda^2 = \frac{1}{2}\|x\|_\Lambda^2 .$$

Since any $x \in \mathcal{H}$ is a limit point of some sequence $\{x_n\}_{n\in\mathbb{N}} \subset \mathcal{H}_0$, we can conclude that

$$I_1^{\text{MDP}}(x) = \frac{1}{2}\|x\|_\Lambda^2, \quad \text{for } x \in \mathcal{H}.$$

In the following, we explicitly give the expression for $\|x\|_\Lambda^2$. To that end, we rewrite

$$x_n(t) = \sum_{i=1}^{N_n} (w_j^n - w_{j-1}^n)\Lambda(t, t_j), \quad \text{for some } (w_1^n, w_2^n, \ldots, w_{N_n}^n) \in \mathbb{R}^{N_n}.$$

Then, we have

$$\|x_n\|_\Lambda^2 = \sum_{i,j=1}^{N_n} \frac{w_i^n - w_{i-1}^n}{t_i - t_{i-1}} \Lambda(t_i, t_j) \frac{w_j^n - w_{j-1}^n}{t_j - t_{j-1}} (t_i - t_{i-1})(t_j - t_{j-1}).$$

Since $\Lambda(\cdot,\cdot)$ is Lebesgue measurable on $[0,T] \times [0,T]$, as $n \to \infty$, we have

$$\|x\|_\Lambda^2 = \lim_{n\to\infty} \|x_n\|_\Lambda^2 = \int_0^T \int_0^T \dot{w}(s)\Lambda(s,t)\dot{w}(t)dsdt, \quad \text{where } x(\cdot) = \int_0^T \Lambda(\cdot, s)\dot{w}(s)ds.$$

Since only $\dot{w}$ is involved in the above equation (and it is sufficient for w to lie in $\mathcal{AC}_0$ to have the above integrals well-defined), we replace $\dot{w}(\cdot) = z(\cdot)$, where z is a Lebesgue measurable function on $[0,T]$. This gives us

$$\|x\|_\Lambda^2 = \lim_{n\to\infty} \|x_n\|_\Lambda^2 = \int_0^T \int_0^T z(s)\Lambda(s,t)z(t)dsdt, \quad \text{where } x(\cdot) = \int_0^T \Lambda(\cdot, s)z(s)ds.$$

This proves the result. □

3.4. **Exponential tightness of $\widehat{X}_1^n$ in $(\mathcal{D}_T, J_1)$.** In order to prove the MDP $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ in $(\mathcal{D}_T, J_1)$, we need to first establish the exponential tightness (see Section 1.2 for the definition) of $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ in $(\mathcal{D}_T, J_1)$.

Theorem 3.3. *Suppose Assumptions 2.1, 2.3, 2.4(ii) and 2.4(iii) hold. Then the following holds*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \sup_{t\in[0,T]} \frac{1}{a_n^2} \log \mathbb{P}\left(\sup_{0\le s\le\delta} |\widehat{X}_1^n(t+s) - \widehat{X}_1^n(t)| > \epsilon \right) = -\infty, \tag{3.22}$$

for every $\epsilon > 0$. As a consequence, $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ is exponentially tight in $(\mathcal{D}_T, J_1)$.

The proof is given in Section 5. The space $\mathcal{D}_T$ under the topology of pointwise covergence is not Hausdorff. Hence, [18, Corollary 4.2.6] cannot be applied, and thus, one cannot directly conclude the sample-path MDP of $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ in $(\mathcal{D}_T, J_1)$ in Theorems 3.2 and 3.3.

3.5. **Completing the proof of Theorem 2.2.**

Theorem 3.4. *Suppose Assumptions 2.1, 2.3, 2.4(ii) and 2.4(iii) hold. Then the family of $\mathcal{D}_T$-valued random variables $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ satisfies an MDP in $(\mathcal{D}_T, J_1)$ with rate a_n^2 and rate function I_1^{MDP} given by (2.19).*

Proof. By the exponential tightness property in Theorem 3.3, from Theorem P in [49] and Theorem A.3 in [51], we have the following consequence: For every subsequence n_k, there exists a further subsequence (still denoted by n_k) such that $\{\widehat{X}_1^{n_k}\}_{k\in\mathbb{N}}$ satisfies MDP in $\mathcal{D}_T$ under the Skorohod J_1 topology with some rate function $I_{(n_k)}$. Moreover, $I_{(n_k)}(x) = \infty$ whenever $x \in \mathcal{D}_T \setminus \mathcal{C}_T$. This means that for any $L > 0$, the set $\widehat{K}_L \doteq \{I_{(n_k)} \leq L\}$ is a compact set of $\mathcal{D}_T$ in the J_1 topology and contains only continuous functions. This means that $\widehat{K}_L$ is also compact in the uniform topology (*i.e.,* the topology that is induced by $\|\cdot\|_T$). This follows from the fact that the J_1 topology restricted to $\mathcal{C}_T$ is the same as the uniform topology.

Now consider a closed set $\mathfrak{C} \subset \mathcal{D}_T$ in the J_1 topology. We have

$$
\begin{aligned}
\limsup_{k\to\infty} \frac{1}{a_{n_k}^2} \log \mathbb{P}\Big(\widehat{X}_1^{n_k} \in \mathfrak{C}\Big) &\leq \limsup_{k\to\infty} \frac{1}{a_{n_k}^2} \log \Big(\mathbb{P}\Big(\widehat{X}_1^{n_k} \in \mathfrak{C} \cap \widehat{K}_L\Big) + \mathbb{P}\Big(\widehat{X}_1^{n_k} \in \mathfrak{C} \cap \widehat{K}_L^c\Big)\Big) \\
&\leq \max\Bigg\{ \limsup_{k\to\infty} \frac{1}{a_{n_k}^2} \log \mathbb{P}\Big(\widehat{X}_1^{n_k} \in \mathfrak{C} \cap \widehat{K}_L\Big), \\
&\qquad \limsup_{k\to\infty} \frac{1}{a_{n_k}^2} \log \mathbb{P}\Big(\widehat{X}_1^{n_k} \in \mathfrak{C} \cap \widehat{K}_L^c\Big) \Bigg\} \\
&\leq \max\Bigg\{ \limsup_{k\to\infty} \frac{1}{a_{n_k}^2} \log \mathbb{P}\Big(\widehat{X}_1^{n_k} \in \mathfrak{C} \cap \widehat{K}_L\Big), \\
&\qquad \limsup_{k\to\infty} \frac{1}{a_{n_k}^2} \log \mathbb{P}\Big(\widehat{X}_1^{n_k} \in \widehat{K}_L^c\Big) \Bigg\} \\
&\leq \max\Bigg\{ \limsup_{k\to\infty} \frac{1}{a_{n_k}^2} \log \mathbb{P}\Big(\widehat{X}_1^{n_k} \in \mathfrak{C} \cap \widehat{K}_L\Big), -L \Bigg\}.
\end{aligned}
$$

We have obtained the last inequality after using the fact that $\{\widehat{X}_1^{n_k}\}_{k\in\mathbb{N}}$ satisfies MDP in $(\mathcal{D}_T, J_1)$ with rate function $I_{(n_k)}$ and the upper bound in the definition of MDP. Indeed,

$$
\limsup_{k\to\infty} \frac{1}{a_{n_k}^2} \log \mathbb{P}\Big(\widehat{X}_1^{n_k} \in \widehat{K}_L^c\Big) \leq - \inf_{x\in\widehat{K}_L} I_{(n_k)}(x) \leq -L\,.
$$

Now observe that $\mathfrak{C} \cap \widehat{K}_L$ is closed in $\mathcal{D}_T$ under uniform topology, and then it is also closed under the topology of pointwise convergence. Therefore, using Theorem 3.2 and the MDP upper bound, we get

$$
\begin{aligned}
\limsup_{k\to\infty} \frac{1}{a_{n_k}^2} \log \mathbb{P}\Big(\widehat{X}_1^{n_k} \in \mathfrak{C}\Big) &\leq \max\Big\{ - \inf_{x\in\mathfrak{C}\cap\widehat{K}_L} I_1^{\mathrm{MDP}}(x), -L \Big\} \\
&\leq \max\Big\{ - \inf_{x\in\mathfrak{C}} I_1^{\mathrm{MDP}}(x), -L \Big\} \\
&\leq - \inf_{x\in\mathfrak{C}} I_1^{\mathrm{MDP}}(x), \qquad \text{after taking } L \uparrow \infty.
\end{aligned}
$$

Since the right hand above is independent of the subsequence n_k, we have

$$
\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\Big(\widehat{X}_1^{n} \in \mathfrak{C}\Big) \leq - \inf_{x\in\mathfrak{C}} I_1^{\mathrm{MDP}}(x).
$$

We now move on to proving the lower bound in the definition of MDP. To do that, we remark that it only suffices to prove the following: for any $x \in \mathcal{D}_T$ such that $I_1^{\mathrm{MDP}}(x) < \infty$, for any subsequence

n_k and any open ball (in the J_1 topology) of radius $\delta > 0$ around x (denoted by $O_x(\delta)$), we have

$$\liminf_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\Big(\widehat{X}_1^n \in O_x(\delta)\Big) \geq -I_1^{\mathrm{MDP}}(x)\,.$$

Without loss of generality, we can take $x \in \mathcal{C}_T$. Since x is continuous on $[0,T]$, it is also uniformly continuous and let $w_x(\cdot)$ be its modulus of continuity. Then from the property of the J_1 topology, there is an open r-ball $B_x(\|\cdot\|_T, r)$ around x in uniform topology such that $B_x(\|\cdot\|_T, r) \subset O_x(\delta)$. To see this, let $e(\cdot)$ denote the identity map on $[0,T]$ and $a(\cdot) : [0,T] \to [0,T]$ be a non-decreasing and onto function. Using the definition of the J_1 topology, we have

$$\max\{\|y - x \circ a\|_T, \|e - a\|_T\} \leq \max\Big\{\|y - x\|_T + w_x(\|e - a\|_T), \|e - a\|_T\Big\}$$

with $f \circ g$ denoting the composition of functions f and g. Therefore, choosing $r = \frac{\delta}{2}$ and $a \equiv e$, we can ensure that

$$B_x(\|\cdot\|_T, r) \subset O_x(\delta).$$

Again fix a subsequence n_k, along which $\{\widehat{X}_1^{n_k}\}_{k\in\mathbb{N}}$ satisfies the MDP in $(\mathcal{D}_T, J_1)$. For $L > 0$, choose $\widehat{K}_L$ corresponding to this subsequence as earlier. From the above discussion, we have

$$\liminf_{k\to\infty} \frac{1}{a_{n_k}^2} \log \mathbb{P}\Big(\widehat{X}_1^{n_k} \in B_x(\|\cdot\|_T, \delta)\Big) \leq \liminf_{k\to\infty} \frac{1}{a_{n_k}^2} \log \mathbb{P}\Big(\widehat{X}_1^{n_k} \in O_x(\delta)\Big).$$

Following the proof of the lower bound in Theorem 4.2.4 in [18] (whose proof involves choosing a continuous function g mapping $(\mathcal{D}_T, \|\cdot\|_T)$ to $\mathcal{D}_T$ with the topology of pointwise convergence and a compact set in $(\mathcal{D}_T, \|\cdot\|_T)$; to that end, we choose $\widehat{K}_L$ as the compact set and the function g as the continuous injection of $(\mathcal{C}_T, \|\cdot\|_T)$ into $\mathcal{C}_T$ with topology of pointwise convergence), we get

$$\liminf_{k\to\infty} \frac{1}{a_{n_k}^2} \log \mathbb{P}\Big(\widehat{X}_1^{n_k} \in O_x(\delta)\Big) \geq \liminf_{k\to\infty} \frac{1}{a_{n_k}^2} \log \mathbb{P}\Big(\widehat{X}_1^{n_k} \in B_x(\|\cdot\|_T, \delta)\Big) \geq -I_1^{\mathrm{MDP}}(x).$$

From the arbitrariness of the sequence n_k, we have

$$\liminf_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\Big(\widehat{X}_1^n \in O_x(\delta)\Big) \geq -I_1^{\mathrm{MDP}}(x).$$

To show that $\{x \in \mathcal{D}_T : I_1^{\mathrm{MDP}}(x) \leq l\}$ is a compact set (in the J_1 topology) of $\mathcal{D}_T$ for every $l \geq 0$, we do the following: Using the same argument as earlier, over any subsequence, $\{\widehat{X}_1^n\}_{n\in\mathbb{N}}$ satisfies an MDP in $(\mathcal{D}_T, J_1)$ with some rate function $I_{(n)}$ which is such that $\{x \in \mathcal{D}_T : I_{(n)}(x) \leq l\}$ is a compact set (in the J_1 topology) of $\mathcal{D}_T$. From Lemma 4.1.4 in [18], the rate function $I_{(n)}$ and I_1^{MDP} are identical. Hence, I_1^{MDP} also satisfies the desired property. This completes the proof. ☐

Next we prove the MDP of $\{\widetilde{X}_2^n\}_{n\in\mathbb{N}}$ by using the MDP for renewal processes in Theorem 2.1.

Proposition 3.2. *Suppose Assumptions 2.1, 2.2 and 2.4(i) hold. Then the family of $\mathcal{D}_T$–valued random variables $\{\widetilde{X}_2^n\}_{n\in\mathbb{N}}$ satisfies an MDP in $(\mathcal{D}_T, J_1)$ with rate a_n^2 and rate function I_2^{MDP} given by* (2.20).

Proof. From Theorem 2.1, we know that $\{\widetilde{A}^n\}_{n\in\mathbb{N}}$ satisfies MDP with rate a_n^2 and rate function I_A^{MDP}. By integration by parts, we can write

$$\begin{aligned}\widetilde{X}_2^n(t) &= \int_0^t G_1(t,s)d\widetilde{A}^n(s)\\ &= \widetilde{A}^n(t)G_1(t,t) - \int_0^t \widetilde{A}^n(u-)dG_1(t,u)\,.\end{aligned}$$

From Lemma 6.1 in [46], Assumptions 2.1 and 2.4(i), we know that the mapping $\phi : (\mathcal{D}_T, J_1) \to (\mathcal{D}_T, J_1)$ defined by

$$\phi(f) \doteq f(t)G_1(t,t) - \int_0^t f(u-)dG_1(t,u)$$

is continuous. Therefore, using the contraction principle [18, Theorem 4.2.1], we can conclude that $\{\widetilde{X}_2^n\}_{n\in\mathbb{N}}$ satisfies MDP in $(\mathcal{D}_T, J_1)$ with rate a_n^2 and rate function given by (2.20). □

Proof of Theorem 2.2. Let

$$\widehat{X}^n(t) \doteq \widehat{X}_1^n(t) + \widetilde{X}_2^n(t), \text{ for } t \in [0,T].$$

From Theorem 3.1 and Theorem 4.2.13 in [18], the MDP of $\{\widetilde{X}^n\}_{n\in\mathbb{N}}$ in $(\mathcal{D}_T, J_1)$ is implied by the MDP of $\{\widehat{X}^n\}_{n\in\mathbb{N}}$ in $(\mathcal{D}_T, J_1)$ with the same rate and rate function. To get the MDP of $\{\widehat{X}^n\}_{n\in\mathbb{N}}$, we use Proposition 3.2 and Theorem 3.4. This concludes the proof of the theorem. □

4. Proof of Exponential Equivalence (Theorem 3.1)

Firstly from Theorem 2.1, it is evident that for large n, $A^n(t)$ satisfies the following with large probability: for any $k \ge 0$, there is $n(k)$ such that for $n \ge n(k)$,

$$\lfloor \lambda nt \rfloor - \lfloor k a_n \sqrt{n} \rfloor \le A^n(t)(\omega) \le \lfloor \lambda nt \rfloor + \lfloor k a_n \sqrt{n} \rfloor, \ \forall\, t \in [0,T]. \tag{P}$$

This suggests that we split the probability

$$\mathbb{P}\left(\|\widetilde{X}_1^n - \widehat{X}_1^n\|_T > \delta\right)$$

into probabilities over two sets *viz.*, over a set where the above property (P) holds (say $P_1^n(k)$) and over a set where the above property (P) does not hold (say $P_2^n(k)$).

We will show that $P_2^n(k)$ can be ignored for large n and hence, the only relevant term for large n is $P_1^n(k)$. To that end, define the following set:

$$\mathcal{W}(k) \doteq \Big\{\omega : \Lambda^-(t,k) < A^n(t)(\omega) < \Lambda^+(t,k), \text{ for } n \ge n(k) \text{ and } t \in [0,T]\Big\} \tag{4.1}$$

where $\Lambda^\pm(s,k) \doteq \lfloor \lambda sn \rfloor \pm \lfloor k a_n \sqrt{n} \rfloor$. As mentioned already, the reason behind defining and considering the above sets is as follows: from Theorem 2.1, it is clear that for any $k \ge 0$, the event

$$\Big\{\omega : |A^n(t) - \lambda nt| \ge \lfloor k a_n \sqrt{n} \rfloor \Big|, \text{ for } t \in [0,T]\}$$

occurs with probability of the order of $e^{-k a_n^2}$, for large n. Therefore, defining the set $\mathcal{W}(k)$ as above allows us to study events that are very probable. Note that on $\mathcal{W}(k)$, we have the following:

$$s_i^n - \frac{1 + \lfloor k a_n \sqrt{n} \rfloor}{\lfloor \lambda n \rfloor} \le \tau_i^n \le s_i^n + \frac{\lfloor k a_n \sqrt{n} \rfloor}{\lfloor \lambda n \rfloor}, \text{ for } 1 \le i \le \lfloor \lambda nT \rfloor + \lfloor k a_n \sqrt{n} \rfloor. \tag{4.2}$$

To see this, we note that

$$|A^n(t) - \lfloor \lambda nt \rfloor| < \lfloor k a_n \sqrt{n} \rfloor, \text{ for } t \in [0,T]$$

on $\mathcal{W}(k)$. Since $A^n(\tau_i^n) = i$, we write

$$\begin{aligned} &\lfloor \lambda n \tau_i^n \rfloor - \lfloor k a_n \sqrt{n} \rfloor < i < \lfloor \lambda n \tau_i^n \rfloor + \lfloor k a_n \sqrt{n} \rfloor \\ \implies\ & \lambda n \tau_i^n - \lfloor k a_n \sqrt{n} \rfloor < i < \lambda n \tau_i^n + 1 + \lfloor k a_n \sqrt{n} \rfloor \\ \implies\ & s_i^n - \frac{1 + \lfloor k a_n \sqrt{n} \rfloor}{\lambda n} < \tau_i^n < s_i^n + \frac{\lfloor k a_n \sqrt{n} \rfloor}{\lambda n}. \end{aligned}$$

We use the following representation in what follows. For a bounded Borel measurable function $f : [0, T] \to \mathbb{R}$,

$$\frac{1}{n}\sum_{i=1}^{\lfloor \lambda nT \rfloor} f(s_i^n) = \lambda \int_0^T f\Big(s^n_{\lfloor \lambda ns \rfloor}\Big) ds + \frac{f(s_0^n)}{n} + \lambda f(s^n_{\lfloor \lambda nT \rfloor})\Big(T - \frac{\lfloor \lambda nT \rfloor}{\lambda n}\Big)$$
$$= \lambda \int_0^T f\Big(s^n_{\lfloor \lambda ns \rfloor}\Big) ds + \mathcal{O}\Big(n^{-1}\Big) \tag{4.3}$$

Recall that $s_i^n = \frac{i}{\lambda n}$.

Proof of Theorem 3.1. Fix $\delta > 0$ and $k > 0$. Consider the event $\mathcal{W}(k)$ defined in (4.1). From now on, we write $\mathbb{P}_k(A)$ for $\mathbb{P}(A \cap \mathcal{W}(k))$ and $\mathbb{E}_k[\cdot]$ for $\mathbb{E}[\cdot \mathbb{1}_{\mathcal{W}(k)}]$, for short. Suppose we have

$$\lim_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k\left(\|\widetilde{X}_1^n - \widehat{X}_1^n\|_T > \delta\right) = -\infty, \text{ for every } \delta > 0 \tag{4.4}$$

Then consider $\mathcal{W}(k)^c$ which is a closed set. From Theorem 2.1 and the definition of MDP (as $\{\widetilde{A}^n\}_{n\in\mathbb{N}}$ satisfies an MDP from Theorem 2.1), we have

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\Big(\Big\{\|\widetilde{X}_1^n - \widehat{X}_1^n\|_T > \delta\Big\} \cap \mathcal{W}(k)^c\Big) \le \limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\Big(\mathcal{W}(k)^c\Big)$$
$$\le -\inf_{x\in\mathcal{W}(k)^c} I_A^{\text{MDP}}(x).$$

Combining (4.4) and the above display, we have

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\left(\|\widetilde{X}_1^n - \widehat{X}_1^n\|_T > \delta\right) \le \max\Bigg\{ \limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\Big(\Big\{\|\widetilde{X}_1^n - \widehat{X}_1^n\|_T > \delta\Big\} \cap \mathcal{W}(k)^c\Big),$$
$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\Big(\Big\{\|\widetilde{X}_1^n - \widehat{X}_1^n\|_T > \delta\Big\} \cap \mathcal{W}(k)\Big)\Bigg\}$$
$$\le -\inf_{x\in\mathcal{W}(k)^c} I_A^{\text{MDP}}(x).$$

Now letting $k \uparrow \infty$, we have the result. Therefore, it suffices to show that (4.4) holds for every $k > 0$.

To do this, we now define an intermediate process

$$Y_1^n(t) \doteq \frac{1}{a_n\sqrt{n}} \sum_{i=1}^{A^n(t)} \widetilde{H}(t, \tau_i^n, \vartheta_i) \mathbb{1}_{\{|H(t,\tau_i^n,\vartheta_i)| \le \frac{\delta}{2} a_n\sqrt{n}\}}.$$

We now show that $\widetilde{X}_1^n$ and Y_1^n satisfy

$$\lim_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k\left(\|\widetilde{X}_1^n - Y_1^n\|_T > \delta\right) = -\infty, \text{ for every } \delta > 0 \tag{4.5}$$

and then proceed to show that Y_1^n and the process Y_2^n defined as

$$Y_2^n(t) \doteq \frac{1}{a_n\sqrt{n}} \sum_{i=1}^{A^n(t)} \Big(\underline{H}^n(t, \tau_i^n, \vartheta_i) - \mathbb{E}\Big[\underline{H}^n(t, \tau_i^n, \vartheta_i)\Big]\Big),$$

satisfy

$$\lim_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k\left(\|Y_1^n - Y_2^n\|_T > \delta\right) = -\infty, \text{ for every } \delta > 0. \tag{4.6}$$

Finally, proving that

$$\lim_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k \left(\|Y_2^n - \widehat{X}_1^n\|_T > \delta \right) = -\infty, \text{ for every } \delta > 0 \tag{4.7}$$

and combining (4.5), (4.6), (4.7), we have (4.4).

Proof of (4.5)**:** We have

$$|\widetilde{X}_1^n(t) - Y_1^n(t)| \le \frac{1}{a_n\sqrt{n}} \sum_{i=1}^{A^n(t)} \left| \widetilde{H}(t, \tau_i^n, \vartheta_i) \mathbb{1}_{\{|H(t,\tau_i^n,\vartheta_i)| > \frac{\delta}{2} a_n\sqrt{n}\}} \right| \tag{4.8}$$

Therefore,

$$\left\{ \|\widetilde{X}_1^n - Y_1^n\|_T > \frac{\delta}{2} \right\} \subset \left\{ \exists t \in [0,T],\ 1 \le i \le A^n(t) : |\widetilde{H}(t, \tau_i^n, \vartheta_i)| > \frac{\delta}{2} a_n\sqrt{n} \right\}.$$

This implies

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k \left(\|\widetilde{X}_1^n - Y_1^n\|_T > \frac{\delta}{2} \right) \\
&\quad \le \limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k \left(\exists\, 1 \le i \le A^n(T) : \sup_{t,s\in[0,T]} |H(t,s,\vartheta_i)| > \frac{\delta}{2} a_n\sqrt{n} \right) &(4.9)\\
&\quad \le \limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k \left(\exists\, 1 \le i \le \lfloor \lambda n T \rfloor + \lfloor k a_n \sqrt{n} \rfloor : \sup_{t,s\in[0,T]} |H(t,s,\vartheta_i)| > \frac{\delta}{2} a_n\sqrt{n} \right) &(4.10)\\
&\quad \le \limsup_{n\to\infty} \frac{1}{a_n^2} \log \left(\left(\lfloor \lambda n T \rfloor + \lfloor k a_n \sqrt{n} \rfloor \right) \mathbb{P} \left(\sup_{t,s\in[0,T]} |H(t,s,\vartheta_1)| > \frac{\delta}{2} a_n\sqrt{n} \right) \right). &(4.11)
\end{aligned}$$

In the above, to get (4.9), we use the fact that $A^n(t) \le A^n(T)$, for $t \in [0,T]$; to get (4.10), we use the fact that on $\mathcal{W}(k)$, $A^n(T) \le \lfloor \lambda n T \rfloor + \lfloor k a_n \sqrt{n} \rfloor$; to get (4.11), we use the union bound and the fact that $\mathbb{P}_k(\cdot) \le \mathbb{P}(\cdot)$.

Now for $\epsilon > 0$, choose n large such that $\lfloor k a_n \sqrt{n} \rfloor \le \epsilon n$. Therefore, using (2.12) and (4.11), we have (4.5).

Proof of (4.6)**:** First choose n large such that $|\mathbb{E}[\widetilde{H}^n(t,s,\vartheta_i)]| < \frac{\delta}{2}$ for every $t, s \in [0,T]$. This can be done as $H(t,s,\vartheta_1)$ is uniformly integrable in $t, s \in [0,T]$ (see Lemma 2.1). Now observe that with

$$K^n(t,s,x) \doteq \widetilde{H}(t,s,x) \mathbb{1}_{\{|H(t,s,x)| \le \frac{\delta}{2} a_n\sqrt{n}\}} - \left(\underline{H}^n(t,s,x) - \mathbb{E}\left[\underline{H}^n(t,s,\vartheta_1) \right] \right),$$

we have $K^n(t,s,x) \le \frac{\delta}{2} a_n \sqrt{n}$, for every $t, s \in [0,T]$ and $x \in \mathbb{R}^d$. For $\rho > 0$, consider

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k \left(\|Y_1^n - Y_2^n\|_T > \delta \right) \\
&\quad \le \limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k \left(\frac{1}{a_n\sqrt{n}} \left\| \sum_{i=1}^{A^n(t)} K^n(t, \tau_i^n, \vartheta_i) \right\|_T > \delta \right) \\
&\quad \le \limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k \left(\frac{1}{a_n\sqrt{n}} \sum_{i=1}^{A^n(T)} \sup_{t,s\in[0,T]} |K^n(t,s,\vartheta_i)| > \delta \right) &(4.12)\\
&\quad \le \limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P} \left(\frac{1}{a_n\sqrt{n}} \sum_{i=1}^{\lfloor \lambda n T \rfloor + \lfloor k a_n \sqrt{n} \rfloor} \sup_{t,s\in[0,T]} |K^n(t,s,\vartheta_i)| > \delta \right) &(4.13)\\
&\quad \le -\rho\delta + \limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{E} \left[\exp \left(\frac{\rho a_n}{n} \sum_{i=1}^{\lfloor \lambda n T \rfloor + \lfloor k a_n \sqrt{n} \rfloor} |K^n(t,s,\vartheta_i)| \right) \right] &(4.14)
\end{aligned}$$

$$\leq -\rho\delta + \limsup_{n\to\infty} \frac{\lfloor \lambda nT\rfloor + \lfloor ka_n\sqrt{n}\rfloor}{a_n^2} \log \mathbb{E}\left[\exp\left(\frac{\rho a_n}{\sqrt{n}}|K^n(t,s,\vartheta_i)|\right)\right]. \tag{4.15}$$

In the above, to get (4.12), we use the fact that $A^n(t) \leq A^n(T)$; to get (4.13), we use the fact that on $\mathcal{W}(k)$, $A^n(T) \leq \lfloor \lambda nT\rfloor + \lfloor ka_n\sqrt{n}\rfloor$; to get (4.14), we use Markov's inequality; to get (4.15), we use the fact that $\{\vartheta_i\}$ is a family of i.i.d. random variables. Now suppose

$$\sup_{\rho>0} \limsup_{n\to\infty} \frac{\lfloor \lambda nT\rfloor + \lfloor ka_n\sqrt{n}\rfloor}{a_n^2} \log \mathbb{E}\left[\exp\left(\frac{\rho a_n}{\sqrt{n}}|K^n(t,s,\vartheta_i)|\right)\right] < \infty.$$

Then taking $\rho \uparrow \infty$ gives us (4.6). The proof of the above display follows using the arguments in [20, Pg. 213-214] with one small change, *i.e.,* we do not take $l \to \infty$ (the authors in [20] take τ in that paper to ∞; this does not change the applicability of the arguments in our case).

Proof of (4.7): On $\mathcal{W}(k)$, we consider

$$\begin{aligned}
Y_2^n(t) - \widehat{X}_1^n(t) &= \frac{1}{a_n\sqrt{n}} \sum_{i=1}^{A^n(t)} \underline{H}^n(t,\tau_i^n,\vartheta_i) - \frac{1}{a_n\sqrt{n}} \sum_{i=1}^{\lfloor \lambda nt\rfloor} \underline{H}^n(t,s_i^n,\vartheta_i) \\
&= \frac{1}{a_n\sqrt{n}} \sum_{i=1}^{A^n(t)\wedge\lfloor \lambda nt\rfloor} \left(\underline{H}^n(t,\tau_i^n,\vartheta_i) - \underline{H}^n(t,s_i^n,\vartheta_i)\right) \\
&\quad + \frac{1}{a_n\sqrt{n}} \sum_{j=A^n(t)\wedge\lfloor \lambda nt\rfloor+1}^{A^n(t)\vee\lfloor \lambda nt\rfloor} \left(\mathbb{1}_{\{A^n(t)>\lfloor \lambda nt\rfloor\}}\underline{H}^n(t,\tau_j^n,\vartheta_j) - \mathbb{1}_{\{A^n(t)\leq\lfloor \lambda nt\rfloor\}}\underline{H}^n(t,s_i^n,\vartheta_i)\right) \\
&\doteq \mathcal{J}_1^n(t) + \mathcal{J}_2^n(t)\,.
\end{aligned}$$

We next estimate

$$\mathbb{P}_k\Big(\|\mathcal{J}_1^n\|_T > \frac{\delta}{2}\Big) \text{ and } \mathbb{P}_k\Big(\|\mathcal{J}_2^n\|_T > \frac{\delta}{2}\Big)$$

as we know that the sum of these probabilities bound

$$\mathbb{P}_k\left(\|Y_2^n - \widehat{X}_1^n\|_T > \delta\right)$$

from above. This will be the content of the next lemma.

Lemma 4.1. *The following holds:*

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k\Big(\|\mathcal{J}_1^n\|_T \vee \|\mathcal{J}_2^n\|_T > \frac{\delta}{2}\Big) = -\infty. \tag{4.16}$$

Proof. Since

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k\Big(\|\mathcal{J}_1^n\|_T \vee \|\mathcal{J}_2^n\|_T > \frac{\delta}{2}\Big) \leq \max_{i=1,2}\left\{\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k\Big(\|\mathcal{J}_i^n\|_T > \frac{\delta}{2}\Big)\right\},$$

we show that the right hand side is $-\infty$.

We first show that

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k\Big(\|\mathcal{J}_2^n\|_T > \frac{\delta}{2}\Big) = -\infty. \tag{4.17}$$

We start by observing that on $\mathcal{W}(k)$, $\sup_{t\in[0,T]} |A^n(t) - \lfloor \lambda nt\rfloor| \leq \lfloor ka_n\sqrt{n}\rfloor$. Therefore, conditioned on A^n, $\mathcal{J}_2^n(t)$ is a sum of at most $\lfloor ka_n\sqrt{n}\rfloor$ independent terms, for every $t \in [0,T]$. So, by conditioning on A^n and using the fact that $\{\vartheta_i\}_{i\in\mathbb{N}}$ i.i.d given A^n, we can infer that the process $\mathcal{J}_2^n$ is exactly of the form $\mathcal{Z}^n$ defined in (A.1) in the Appendix A, if i started from $A^n(t) \wedge \lfloor \lambda nt\rfloor + 1$ instead of 1. Therefore, from Corollary A.1, we have (4.17).

To complete the proof, all that remains is to show that

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}_k\Big(\|\mathcal{J}_1^n\|_T > \frac{\delta}{2}\Big) = -\infty. \tag{4.18}$$

Define $\widehat{\tau}_\delta^n \doteq \inf\{t > 0 : \mathcal{J}_2^n(t) > \frac{\delta}{2}\}$. To begin with, observing again that upon conditioning on A^n and using the fact that $\{\vartheta_i\}_{i\in\mathbb{N}}$ is i.i.d. given A^n, we can infer that the process $\mathcal{J}_1^n$ is exactly of the form $\mathcal{Z}^n$ defined in (A.1) in the Appendix A. Therefore, from Theorem A.1, we can show that (4.18) holds if we can show that

$$\begin{aligned}\Theta_1^J &\doteq \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}_k\Bigg[\sum_{i=1}^{A^n(T)\wedge\lfloor\lambda nT\rfloor} \int_{\mathbb{R}^d} \Big(\underline{H}^n(\widehat{\tau}_\delta^n \wedge T, \tau_i^n, x) - \underline{H}^n(\widehat{\tau}_\delta^n \wedge T, s_i^n, x)\Big)^2 F(dx)\Bigg] \\ &= \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}_k\Bigg[\sum_{i=1}^{A^n(T)\wedge\lfloor\lambda nT\rfloor} \int_{\mathbb{R}^d} \Big(\widetilde{H}(\widehat{\tau}_\delta^n \wedge T, \tau_i^n, x) - \widetilde{H}(\widehat{\tau}_\delta^n \wedge T, s_i^n, x)\Big)^2 F(dx)\Bigg] = 0.\end{aligned}$$

The first equality above follows from (3.8) of Lemma 3.1. To prove (4.18), we follow the same arguments of the proof of Theorem A.1 with two minor changes after choosing $N^n(t) = A^n(t) \wedge \lfloor\lambda nt\rfloor$, $Z_i(t,s,x) = \widetilde{H}(t,s,x) - \widetilde{H}(t,s_i^n,x)$. Since given the event $\mathcal{W}(k)$, $\{\vartheta_i\}_{n\in\mathbb{N}}$ is i.i.d., the arguments work even if we replace $\mathbb{P}(\cdot)$ and $\mathbb{E}[\cdot]$ with $\mathbb{P}_k(\cdot)$ and $\mathbb{E}_k[\cdot]$, respectively. Then, we use the tower property of conditional expectation and write all the expectations $\mathbb{E}_k[\cdot]$ as $\mathbb{E}_k[\mathbb{E}[\cdot|A^n_{[0,T]}]]$. This is allowed as the event $\mathcal{W}(k)$ lies in the σ–algebra generated by the process A^n on $[0,T]$. Due to these minor changes, we omit the proof to avoid repetition.

To prove the desired result, it now suffices for us to show that $\Theta_1^J = 0$. To that end,

$$\Theta_1^J \le \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}_k\Bigg[\sum_{i=1}^{\lfloor\lambda nT\rfloor+\lfloor ka_n\sqrt{n}\rfloor} \int_{\mathbb{R}^d} \Big(\widetilde{H}(\widehat{\tau}_\delta^n \wedge T, \tau_i^n, x) - \widetilde{H}(\widehat{\tau}_\delta^n \wedge T, s_i^n, x)\Big)^2 F(dx)\Bigg] \tag{4.19}$$

$$= \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}_k\Bigg[\sum_{i=1}^{\lfloor\lambda nT\rfloor} \int_{\mathbb{R}^d} \Big(\widetilde{H}(\widehat{\tau}_\delta^n \wedge T, \tau_i^n, x) - \widetilde{H}(\widehat{\tau}_\delta^n \wedge T, s_i^n, x)\Big)^2 F(dx)\Bigg] \doteq \widetilde{\Theta}_1^J. \tag{4.20}$$

To arrive at (4.19), we used the fact that $A^n(T)$ can be at most $\lfloor\lambda nT\rfloor + \lfloor ka_n\sqrt{n}\rfloor$ as $A^n \in \mathcal{W}(k)$ and to arrive at (4.20), we used the fact that $\frac{\lfloor ka_n\sqrt{n}\rfloor}{n} \to 0$ as $n\to\infty$ and the boundedness of the integral (from Lemma 2.1). We will now show that the expression in (4.20) is zero. Recall that $s_i^n = \frac{i}{\lambda n}$ and τ_i^n is ith arrival time defined in (2.7). It also clear that as $n\to\infty$, $s^n_{\lfloor\lambda ns\rfloor} \to s$, uniformly in $s\in[0,T]$ and since we are conditioning on $\mathcal{W}(k)$, $\tau^n_{\lfloor\lambda ns\rfloor} \to s$, uniformly in $s\in[0,T]$ (see (4.2)). Let $\delta_1^n(s) \doteq s^n_{\lfloor\lambda ns\rfloor} - s$ and $\delta_2^n(s) \doteq \tau^n_{\lfloor\lambda ns\rfloor} - s$. Write

$$\widetilde{H}(\widehat{\tau}_\delta^n \wedge T, \tau^n_{\lfloor\lambda ns\rfloor}, x) - \widetilde{H}(\widehat{\tau}_\delta^n \wedge T, s^n_{\lfloor\lambda ns\rfloor}, x) = \widetilde{H}\Big(\widehat{\tau}_\delta^n \wedge T, s+\delta_2^n(s), x\Big) - \widetilde{H}\Big(\widehat{\tau}_\delta^n \wedge T, s+\delta_1^n(s), x\Big)$$

Then, using (4.3), we have

$$\begin{aligned}\widetilde{\Theta}_1^J &= \lambda \limsup_{n\to\infty} \mathbb{E}_k\Bigg[\int_0^T \int_{\mathbb{R}^d} \Big(\widetilde{H}\Big(\widehat{\tau}_\delta^n \wedge T, s+\delta_2^n(s), x\Big) - \widetilde{H}\Big(\widehat{\tau}_\delta^n \wedge T, s+\delta_1^n(s), x\Big)\Big)^2 F(dx)ds\Bigg] \\ &\le 2\lambda \limsup_{n\to\infty} \mathbb{E}_k\Bigg[\int_0^T \int_{\mathbb{R}^d} \Big(\widetilde{H}\Big(\widehat{\tau}_\delta^n \wedge T, s+\delta_2^n(s), x\Big) - \widetilde{H}\Big(\widehat{\tau}_\delta^n \wedge T, s, x\Big)\Big)^2 F(dx)ds\Bigg] \\ &\quad + 2\lambda \limsup_{n\to\infty} \mathbb{E}_k\Bigg[\int_0^T \int_{\mathbb{R}^d} \Big(\widetilde{H}\Big(\widehat{\tau}_\delta^n \wedge T, s, x\Big) - \widetilde{H}\Big(\widehat{\tau}_\delta^n \wedge T, s+\delta_1^n(s), x\Big)\Big)^2 F(dx)ds\Bigg] \\ &\le 2\lambda \limsup_{n\to\infty} \sup_{t\in[0,T]} \int_0^T \int_{\mathbb{R}^d} \Big(\widetilde{H}\Big(t, s+\delta_2^n(s), x\Big) - \widetilde{H}\Big(t, s, x\Big)\Big)^2 F(dx)ds\end{aligned}$$

$$+ 2\lambda \limsup_{n\to\infty} \sup_{t\in[0,T]} \int_0^T \int_{\mathbb{R}^d} \Big(\widetilde{H}\big(t,s,x\big) - \widetilde{H}\big(t, s+\delta_1^n(s), x\big)\Big)^2 F(dx)ds$$
$$= 0\,.$$

To arrive at the last equation, we used Assumption 2.4(ii) in conjunction with the fact that $\delta_1^n(\cdot), \delta_2^n(\cdot) \to 0$, uniformly on $[0,T]$ as $n\to\infty$. This proves Lemma 4.1. □

Therefore, we have proved that (4.4) holds and hence, this completes the proof of Theorem 3.1. □

5. Proof of Exponential Tightness (Theorem 3.3)

Before we give the proof of Theorem 3.3, we first give the following lemma.

Lemma 5.1. *Under Assumptions 2.1, 2.3 and 2.4(iii), the following hold:*

$$\lim_{\delta\to 0} \sup_{t\in[0,T]} \int_0^T \breve{G}_2(t,u,\delta)du = 0\,. \tag{5.1}$$

Proof. To begin with, we first observe that under Assumption 2.3, using Lemma 2.1, we can conclude that $\breve{G}_2$ is uniformly bounded in all the arguments. Now define $\mathcal{G} : [0,T]\times[0,T] \to \mathbb{R}^+$ as

$$\mathcal{G}(t,s) \doteq \int_0^T \breve{G}_2(t,u,s)du.$$

Suppose that $\mathcal{G}$ is continuous on $[0,T]\times[0,T]$. Then $\mathcal{G}$ is uniformly continuous on $[0,T]\times[0,T]$. This in turn, implies the existence of a modulus of continuity ϖ such that $\varpi : \mathbb{R}^+ \to \mathbb{R}^+$ and $\lim_{\delta\to 0}\varpi(\delta) = 0$ associated to $\mathcal{G}$ such that for $(t,s),(t',s') \in [0,T]$,

$$|\mathcal{G}(t,s) - \mathcal{G}(t',s')| \le \varpi(|t-t'| + |s-s'|).$$

For $\delta > 0$, now choosing $s = t' = s' = t+\delta$ and noting that $\mathcal{G}(s,s) = 0$, we have $\sup_{t\in[0,T]} |\mathcal{G}(t,t+\delta)| \le \varpi(\delta)$. This immediately implies (5.1) after taking $\delta\to 0$.

Therefore, it remains for us to prove that $\mathcal{G}$ is continuous on $[0,T]\times[0,T]$. To that end, it is clearly sufficient to show that functions $\mathcal{G}(\cdot,t)$ and $\mathcal{G}(t,\cdot)$ are continuous on $[0,T]$, for every $t\in[0,T]$. This is what we will do. Since $\mathcal{G}(t,s) = \mathcal{G}(s,t)$, we only consider $\mathcal{G}(\cdot,t)$. Fix $t, t_0 \in [0,T]$ and ϵ small and consider

$$\begin{aligned}
&|\mathcal{G}(t_0,t) - \mathcal{G}(t_0+\epsilon,t)| \\
&= \Big| \int_0^T \Big(\breve{G}_2(t_0,u,t) - \breve{G}_2(t_0+\epsilon,u,t)\Big)du \Big| \\
&= \Big| \int_0^T \int_{\mathbb{R}^d} \Big(H\big(t_0-u, g(u,x)\big) - H\big(t_0+\epsilon-u, g(u,x)\big)\Big) \\
&\qquad \times \Big(H\big(t_0-u, g(u,x)\big) + H\big(t_0+\epsilon-u, g(u,x)\big) - 2H\big(t-u, g(u,x)\big)\Big)F(dx)du \Big| \qquad (5.2)\\
&\le \int_0^T \Big(\int_{\mathbb{R}^d} \Big(H\big(t_0-u, g(u,x)\big) - H\big(t_0+\epsilon-u, g(u,x)\big)\Big)^2 F(dx)\Big)^{\frac{1}{2}} \\
&\qquad \times \Big(\int_{\mathbb{R}^d} \Big(H\big(t_0-u, g(u,x)\big) + H\big(t_0+\epsilon-u, g(u,x)\big) - 2H\big(t-u, g(u,x)\big)\Big)^2 F(dx)\Big)^{\frac{1}{2}} du \qquad (5.3)\\
&\le C \int_0^T \Big(\int_{\mathbb{R}^d} \Big(H\big(t_0-u, g(u,x)\big) - H\big(t_0+\epsilon-u, g(u,x)\big)\Big)^2 F(dx)\Big)^{\frac{1}{2}} du \qquad (5.4)
\end{aligned}$$

$$\leq CT\Big(\int_0^T \int_{\mathbb{R}^d} \Big(H\big(t_0-u,g(u,x)\big)-H\big(t_0+\epsilon-u,g(u,x)\big)\Big)^2 F(dx)du\Big)^{\frac{1}{2}}. \tag{5.5}$$

In the above, to get (5.2), we use the identity $x^2-y^2=(x+y)(x-y)$; to get (5.3), we apply Cauchy-Schwartz inequality; to get (5.4), we use the fact $\mathbb{E}[\sup_{t,s\in[0,T]}|H(t,s,\vartheta_1)|^2]$ from Lemma 2.1; to get (5.5), we again apply Cauchy-Schwartz inequality.

Now taking $\epsilon\to 0$, (5.5) goes to 0 for the following reason: When $\epsilon\to 0^+$, the right continuity of H (from Assumption 2.1) in conjunction with the dominated convergence theorem implies that (5.5) goes to 0. Finally, when $\epsilon\to 0^-$, Assumption 2.4(iii) implies that (5.5) goes to 0. This proves the continuity of $\mathcal{G}$ on $[0,T]\times[0,T]$ and also the result. □

Proof of Theorem 3.3. Fix $0\leq t\leq T$ and $\delta>0$. For $0\leq s\leq\delta$, consider

$$\begin{aligned}\widehat{X}_1^n(t+s)-\widehat{X}_1^n(t) &= \frac{1}{a_n\sqrt{n}}\sum_{i=1}^{\lfloor\lambda n(t+s)\rfloor}\underline{H}^n(t+s,s_i^n,\vartheta_i)-\frac{1}{a_n\sqrt{n}}\sum_{j=1}^{\lfloor\lambda nt\rfloor}\underline{H}^n(t,s_j^n,\vartheta_j)\\ &= \frac{1}{a_n\sqrt{n}}\sum_{i=1}^{\lfloor\lambda nt\rfloor}[\underline{H}^n(t+s,s_i^n,\vartheta_i)-\underline{H}^n(t,s_i^n,\vartheta_i)]+\frac{1}{a_n\sqrt{n}}\sum_{j=\lfloor\lambda nt\rfloor+1}^{\lfloor\lambda n(t+s)\rfloor}\underline{H}^n(t+s,s_j^n,\vartheta_j).\end{aligned}$$

For $\epsilon>0$, to estimate the probability of the event

$$\Big\{\sup_{0\leq s\leq\delta}|\widehat{X}_1^n(t+s)-\widehat{X}_1^n(t)|>\epsilon\Big\},$$

we again make use of Theorem A.1. To that end, notice that the process $\widehat{Z}_t^n(\cdot)\doteq\widehat{X}_1^n(t+\cdot)-\widehat{X}_1^n(\cdot)$ is of the form $\mathcal{Z}^n$ defined in (A.1) with $N^n(u)=\lfloor\lambda n(t+u)\rfloor$ and

$$Z_i(u,r,x)=\begin{cases}\underline{H}^n(t+u,s_i^n,x)-\underline{H}^n(t,s_i^n,x), & \text{if } i\leq\lfloor\lambda nt\rfloor,\\ \underline{H}^n(t+u,s_i^n,x), & \text{otherwise.}\end{cases}$$

Define $\breve{\tau}_{\epsilon,t}^n\doteq\inf\{s>0:\widehat{Z}_t^n(s)>\epsilon\}$ and $\breve{\tau}_{\epsilon,t}^n=\delta+1$, $\widehat{Z}_t^n(s)\leq\epsilon$, for every $s\in[0,\delta]$.

With the choices of N^n and Z_i, using the arguments of the proof of Theorem A.1 until we take $n\to\infty$, we obtain

$$\begin{aligned}&\frac{1}{a_n^2}\log\mathbb{P}\left(\sup_{0\leq s\leq\delta}|\widehat{X}_1^n(t+s)-\widehat{X}_1^n(t)|>\epsilon\right)\\ &\quad\leq-\rho\epsilon+\frac{\rho^2}{2n}\sum_{i=1}^{\lfloor\lambda nt\rfloor}\mathbb{E}\Big[\Big(\underline{H}^n(t+(\breve{\tau}_{\epsilon,t}^n\wedge\delta),s_i^n,\vartheta_i)-\underline{H}^n(t,s_i^n,\vartheta_i)\Big)^2\Big]\\ &\qquad+\frac{\rho^2}{2n}\sum_{i=\lfloor\lambda nt\rfloor}^{\lfloor\lambda n(t+\delta)\rfloor}\mathbb{E}\Big[\Big(\underline{H}^n(t+(\breve{\tau}_{\epsilon,t}^n\wedge\delta),s_i^n,\vartheta_i)\Big)^2\Big]+\mathcal{O}\Big(\mathbb{E}\Big[\sup_{t,s\in[0,T]}|\underline{H}^n(t,s,\vartheta_1)|^3\Big]a_n n^{-\frac{1}{2}}\Big).\end{aligned}$$

Taking supremum over $t\in[0,T]$ in the above display and taking $n\to\infty$, we have

$$\begin{aligned}&\limsup_{n\to\infty}\sup_{t\in[0,T]}\frac{1}{a_n^2}\log\mathbb{P}\left(\sup_{0\leq s\leq\delta}|\widehat{X}_1^n(t+s)-\widehat{X}_1^n(t)|>\epsilon\right)\\ &\leq-\rho\epsilon+\frac{\rho^2}{2}\limsup_{n\to\infty}\sup_{t\in[0,T]}\frac{1}{n}\sum_{i=1}^{\lfloor\lambda nt\rfloor}\mathbb{E}\Big[\Big(\widetilde{H}(t+(\breve{\tau}_{\epsilon,t}^n\wedge\delta),s_i^n,\vartheta_i)-\widetilde{H}(t,s_i^n,\vartheta_i)\Big)^2\Big]+\frac{1}{2}\rho^2\lambda\delta\mathbb{E}\Big[M(\vartheta_1)^2\Big].\end{aligned} \tag{5.6}$$

In the above, we have arrived at the second term from (3.8) of Lemma 3.1, and the third term on the right hand side after invoking Lemma 2.1 and using the fact that the numbers of terms in the summation is $\lfloor \lambda n(t+\delta)\rfloor - \lfloor \lambda nt\rfloor + 1$. To prove the result, it suffices to show that

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\sup_{t\in[0,T]}\frac{1}{n}\sum_{i=1}^{\lfloor \lambda nt\rfloor}\left(\mathbb{E}\Big[\Big(\widetilde{H}(t+(\breve{\tau}^n_{\epsilon,t}\wedge\delta),s^n_i,\vartheta_i)-\widetilde{H}(t,s^n_i,\vartheta_i)\Big)^2\Big]\right)=0. \tag{5.7}$$

To that end, define $\delta^n(u)\doteq s^n_{\lfloor \lambda nu\rfloor}-u$. Using (4.3), we write

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{\lfloor \lambda nt\rfloor}\mathbb{E}\Big[\Big(\widetilde{H}(t+(\breve{\tau}^n_{\epsilon,t}\wedge\delta),s^n_i,\vartheta_i)-\widetilde{H}(t,s^n_i,\vartheta_i)\Big)^2\Big]\\
&\quad=\lambda\mathbb{E}\Big[\int_0^T\int_{\mathbb{R}^d}\Big(\widetilde{H}(t+(\breve{\tau}^n_{\epsilon,t}\wedge\delta),u+\delta^n(u),x)-\widetilde{H}(t,u+\delta^n(u),x)\Big)^2F(dx)du\Big]+\mathcal{O}(n^{-1})\\
&\quad\le 3\lambda\mathbb{E}\Big[\int_0^T\int_{\mathbb{R}^d}\Big(\widetilde{H}(t+(\breve{\tau}^n_{\epsilon,t}\wedge\delta),u+\delta^n(u),x)-\widetilde{H}(t+(\breve{\tau}^n_{\epsilon,t}\wedge\delta),u,x)\Big)^2F(dx)du\Big]\\
&\qquad+3\lambda\mathbb{E}\Big[\int_0^T\int_{\mathbb{R}^d}\Big(\widetilde{H}(t,u,x)-\widetilde{H}(t,u+\delta^n(u),x)\Big)^2F(dx)du\Big]\\
&\qquad+3\lambda\mathbb{E}\Big[\int_0^T\int_{\mathbb{R}^d}\Big(\widetilde{H}(t+(\breve{\tau}^n_{\epsilon,t}\wedge\delta),u,x)-\widetilde{H}(t,u,x)\Big)^2F(dx)du\Big]+\mathcal{O}(n^{-1})\\
&\quad\doteq V^n_1(\delta,t)+V^n_2(\delta,t)+V^n_3(\delta,t)+\mathcal{O}(n^{-1}).
\end{aligned}$$

Now we have

$$\begin{aligned}
&\limsup_{\delta\downarrow 0}\limsup_{n\to\infty}\sup_{t\in[0,T]}V^n_1(\delta,t)\\
&=3\lambda\limsup_{\delta\downarrow 0}\limsup_{n\to\infty}\sup_{t\in[0,T]}3\lambda\mathbb{E}\Big[\int_0^T\int_{\mathbb{R}^d}\Big(\widetilde{H}(t,u+\delta^n(u),x)-\widetilde{H}(t,u,x)\Big)^2F(dx)du\Big]=0.
\end{aligned}$$

In the above, we have used Assumption 2.4(ii) and the fact that the random variable $\tau^n_{\epsilon,t}\wedge\delta$ is uniformly bounded by δ and also the fact that $\delta^n(\cdot)\to 0$, uniformly as $n\to\infty$. Similarly, we can conclude that

$$\limsup_{\delta\downarrow 0}\limsup_{n\to\infty}\sup_{t\in[0,T]}V^n_2(\delta,t)=0$$

To show that

$$\limsup_{\delta\downarrow 0}\limsup_{n\to\infty}\sup_{t\in[0,T]}V^n_3(\delta,t)=0,$$

we use Assumption 2.4(iii), in conjunction with Lemma 5.1. This proves (5.7). By taking $\delta\downarrow 0$ and $\rho\uparrow\infty$ in (5.6), we obtain our desired result in (3.22). Therefore, exponential tightness follows from Theorem A.3 in [51]. □

Appendix A. A Maximal inequality

As usual, we choose $a_n\uparrow\infty$ to satisfy (2.8) throughout the Appendix. In the following, we prove a new maximal inequality. The proof of this inequality is based on stopping times. We state and prove the inequality in a form that is convenient for us to work with. Let $\{T^n_i\}_{i\in\mathbb{N}}$ be a sequence of deterministic positive constants that are strictly increasing with $T^n_0=0$, and let $N^n(t)$ be a counting process with $\{T^n_i\}_{i\in\mathbb{N}}$ being the corresponding sequence of arrival times, that is, $N^n(t)=\max\{k\ge 0: T^n_k\le t\}$ for $t\ge 0$. Let $\mathcal{Z}^n(t)$ be a process defined by

$$\mathcal{Z}^n(t)\doteq\frac{1}{a_n\sqrt{n}}\sum_{i=1}^{N^n(t)}Z^n_i(t,T^n_i,\vartheta_i), \tag{A.1}$$

where for every $i, n \in \mathbb{N}$, $Z_i^n : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$ is a Borel measurable function. We assume that $\mathbb{E}[Z_i^n(t,s,\vartheta_1)] = 0$, for every $s, t \in \mathbb{R}^+$ and $i, n \in \mathbb{N}$. Let $\mathcal{F}_0$ be the trivial σ–algebra and $\mathcal{F}_k \doteq \sigma\{\vartheta_1, \vartheta_2, \ldots, \vartheta_k\}$. In contrast to the case where $Z_i^n(t,s,x)$ independent of t, $\mathcal{Z}^n(t)$ cannot be treated as a martingale and hence the martingale based proofs cannot be used directly to the above case. Define $\tau_\epsilon^n \doteq \inf\{t > 0 : \mathcal{Z}^n(t) > \epsilon\}$ with $\tau_\epsilon^n = T+1$, if $\sup_{t\in[0,T]} \mathcal{Z}^n(t) \le \epsilon$ and

$$\Theta_\epsilon \doteq \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{N^n(T)} \mathbb{E}\left[Z_i^n(\tau_\epsilon^n \wedge T, T_i^n, \vartheta_1)^2\right].$$

Theorem A.1. *Assume that the following condition holds: for every $n \in \mathbb{N}$*

$$\sup_{i\in\mathbb{N},\, t,s\in[0,T]} |Z_i^n(t,s,\vartheta_1)| \le l\frac{\sqrt{n}}{a_n}, \text{ for some } l > 0. \tag{A.2}$$

Suppose for some $C > 0$, $N^n(t) \le Cnt$, for $t \in [0,T]$. Then for every $\epsilon > 0$, the following holds:

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\left(\sup_{0\le t\le T} |\mathcal{Z}^n(t)| > \epsilon\right) \begin{cases} \le -\frac{\epsilon^2}{2\Theta_\epsilon}, & \text{if } \Theta_\epsilon > 0, \\ = -\infty, & \text{if } \Theta_\epsilon = 0. \end{cases} \tag{A.3}$$

Proof. We set $\tau = \tau_\epsilon^n$. Clearly, the event $\{\tau \le t\} \subset \mathcal{F}_{N^n(t)}$ which implies $\{N^n(\tau) \le i\} \subset \mathcal{F}_i$. Moreover, $\{N^n(\tau) \ge i\}$ is only dependent on $\vartheta_1, \vartheta_2, \ldots, \vartheta_{i-1}$ and independent of $\vartheta_i, \vartheta_{i+1}, \ldots, \vartheta_{N^n(T)}$.

For $\rho > 0$, consider

$$\mathbb{P}\left(\sup_{t\in[0,T]} \mathcal{Z}^n(t) > \epsilon\right) = \mathbb{P}(\tau \le T) \le e^{-\rho a_n^2 \epsilon} \mathbb{E}\left[e^{\rho a_n^2 \mathcal{Z}^n(\tau)} \mathbb{1}_{\{\tau \le T\}}\right] \le e^{-\rho a_n^2 \epsilon} \mathbb{E}\left[e^{\rho a_n^2 \mathcal{Z}^n(\tau\wedge T)}\right].$$

This implies that

$$\begin{aligned} &\frac{1}{a_n^2} \log \mathbb{P}\left(\sup_{t\in[0,T]} \mathcal{Z}^n(t) > \epsilon\right) \\ &\quad \le -\rho\epsilon + \frac{1}{a_n^2} \log \mathbb{E}\left[e^{\rho a_n^2 \mathcal{Z}^n(\tau\wedge T)}\right] \\ &\quad \le -\rho\epsilon + \rho \mathbb{E}\Big[\mathcal{Z}^n(\tau\wedge T)\Big] + \frac{\rho^2 a_n^2}{2} \mathbb{E}\Big[\mathcal{Z}^n(\tau\wedge T)^2\Big] + \mathcal{O}\Big(a_n^4 \mathbb{E}\Big[\mathcal{Z}^n(\tau\wedge T)^3\Big]\Big) \end{aligned} \tag{A.4}$$

Now let us evaluate the expectations in the above equation. We have

$$\begin{aligned} a_n\sqrt{n}\mathbb{E}\Big[\mathcal{Z}^n(\tau\wedge T)\Big] &= \mathbb{E}\Big[\sum_{i=1}^{N^n(\tau\wedge T)} Z_i^n(\tau\wedge T, T_i^n, \vartheta_i)\Big] \\ &= \sum_{i=1}^{N^n(T)} \mathbb{E}\Big[Z_i^n(\tau\wedge T, T_i^n, \vartheta_i) \mathbb{1}_{\{N^n(\tau)\ge i\}}\Big] \\ &= \sum_{i=1}^{N^n(T)} \mathbb{E}\Big[Z_i^n(\tau\wedge T, T_i^n, \vartheta_i)\Big] \mathbb{P}\Big(N^n(\tau) \ge i\Big) \qquad \text{(A.5)} \\ &= \sum_{i=1}^{N^n(T)} \int_{N^n(u)\ge i} \int_{\mathbb{R}^d} Z_i^n(u\wedge T, T_i^n, x) F(dx) P_\tau(du) \qquad \text{(A.6)} \\ &= 0\,. \end{aligned}$$

Here, P_τ is the distribution of τ. To arrive at (A.5), we have used the fact that $\{N^n(\tau) \ge i\}$ is independent of ϑ_i and to arrive at (A.6), we have used the fact that $\mathbb{E}[Z_i^n(t,s,\vartheta_1)] = 0$, for

$s, t \in [0, T]$ and $i \in \mathbb{N}$. Now consider

$$\begin{aligned}
a_n^2 n \mathbb{E}\Big[\mathcal{Z}^n(\tau \wedge T)^2\Big] &= \mathbb{E}\Big[\Big(\sum_{i=1}^{N^n(\tau\wedge T)} Z_i^n(\tau \wedge T, T_i^n, \vartheta_i)\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big(\sum_{i=1}^{N^n(T)} Z_i^n(\tau \wedge T, T_i^n, \vartheta_i)\mathbb{1}_{\{N^n(\tau)\geq i\}}\Big)^2\Big] \\
&= \sum_{i=1}^{N^n(T)} \mathbb{E}\Big[Z_i^n(\tau \wedge T, T_i^n, \vartheta_i)^2 \mathbb{1}_{\{N^n(\tau)\geq i\}}\Big] \\
&\quad + 2\sum_{\substack{i,j=1\\ i<j}}^{N^n(T)} \mathbb{E}\Big[Z_i^n(\tau \wedge T, T_i^n, \vartheta_i)\mathbb{1}_{\{N^n(\tau)\geq i\}} Z_j^n(\tau \wedge T, T_j^n, \vartheta_j)\mathbb{1}_{\{N^n(\tau)\geq j\}}\Big] \\
&= \sum_{i=1}^{N^n(T)} \mathbb{E}\Big[Z_i^n(\tau \wedge T, T_i^n, \vartheta_i)^2 \mathbb{1}_{\{N^n(\tau)\geq i\}}\Big] \\
&\quad + 2\sum_{\substack{i,j=1\\ i<j}}^{N^n(T)} \mathbb{E}\Big[Z_j^n(\tau \wedge T, T_j^n, \vartheta_j)\Big]\mathbb{E}\Big[Z_i^n(\tau \wedge T, T_i^n, \vartheta_i)\mathbb{1}_{\{N^n(\tau)\geq i\}}\mathbb{1}_{\{N^n(\tau)\geq j\}}\Big] \qquad \text{(A.7)} \\
&= \sum_{i=1}^{N^n(T)} \mathbb{E}\Big[Z_i^n(\tau \wedge T, T_i^n, \vartheta_i)^2 \mathbb{1}_{\{N^n(\tau)\geq i\}}\Big] \leq \sum_{i=1}^{N^n(T)} \mathbb{E}\Big[Z_i^n(\tau \wedge T, T_i^n, \vartheta_i)^2\Big]. \qquad \text{(A.8)}
\end{aligned}$$

To arrive at (A.7), we have used the fact that $i < j$ and independence of $Z_j^n(\tau \wedge T, T_j^n, \vartheta_j)$ and $Z_i^n(\tau \wedge T, T_i^n, \vartheta_i)\mathbb{1}_{\{N^n(\tau)\geq i\}}\mathbb{1}_{\{N^n(\tau)\geq j\}}$. Finally, to arrive at (A.8), we have used the fact that $\mathbb{E}[Z_i^n(t, s, \vartheta_1)] = 0$, for $s, t \in [0, T]$.

We now analyse the expectation in the fourth term in (A.4).

$$\begin{aligned}
a_n^4 \mathbb{E}\Big[\mathcal{Z}^n(\tau \wedge T)^3\Big] &= \frac{a_n^4}{a_n^3 n\sqrt{n}} \mathbb{E}\Big[\Big(\sum_{i=1}^{N^n(T)} Z_i^n(\tau \wedge T, T_i^n, \vartheta_i)\mathbb{1}_{\{N^n(\tau)\geq i\}}\Big)^3\Big] \\
&= \frac{a_n}{n\sqrt{n}} \sum_{i=1}^{N^n(T)} \mathbb{E}\Big[Z_i^n(\tau \wedge T, T_i^n, \vartheta_i)^3 \mathbb{1}_{\{N^n(\tau)\geq i\}}\Big] \qquad \text{(A.9)} \\
&\leq \frac{Ca_n}{\sqrt{n}} \mathbb{E}\Big[\sup_{i\in\mathbb{N},\, t,s\in[0,T]} Z_i^n(t, s, \vartheta_i)^3\Big] \qquad \text{(A.10)}
\end{aligned}$$

To arrive at (A.9), we argue exactly in the same way as we did to arrive at (A.7); To arrive at (A.10), we used the fact that $N^n(t) \leq Cnt$ and $\mathbb{1}_{\{N^n(\tau)\geq i\}} \leq 1$. From (A.2), following the arguments of [20, Pg. 212], we can conclude that

$$\lim_{n\to\infty} \frac{a_n}{\sqrt{n}} \mathbb{E}\Big[\sup_{i\in\mathbb{N},\, t,s\in[0,T]} Z_i^n(t, s, \vartheta_i)^3\Big] = 0.$$

We now take $n \to \infty$, to get

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\left(\sup_{t\in[0,T]} \mathcal{Z}^n(t) > \epsilon\right) &\leq -\rho\epsilon + \frac{\rho^2}{2} \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{N^n(T)} \mathbb{E}\Big[Z_i^n(\tau \wedge T, T_i^n, \vartheta_i)^2\Big] \\
&\leq -\frac{\epsilon^2}{2\Theta_\epsilon}, \text{ if } \Theta_\epsilon > 0.
\end{aligned}$$

To arrive at the last inequality, we have chosen the optimal value of $\rho = \frac{\epsilon}{\Theta_\epsilon}$. If $\Theta_\epsilon = 0$, then simply take $\rho \uparrow \infty$ to get $-\infty$. Therefore, we have shown that

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\left(\sup_{t\in[0,T]} \mathcal{Z}^n(t) > \epsilon\right) \le -\frac{\epsilon^2}{2\Theta_\epsilon}, \text{ if } \Theta_\epsilon > 0$$

and is equal to $-\infty$, when $\Theta_\epsilon = 0$. Using exactly the same analysis as above, we can show that

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\left(\sup_{t\in[0,T]} -\mathcal{Z}^n(t) > \epsilon\right) \le -\frac{\epsilon^2}{2\Theta_\epsilon}, \text{ if } \Theta_\epsilon > 0$$

and is equal to $-\infty$, when $\Theta_\epsilon = 0$. Combining the above two displays, we have

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\left(\sup_{t\in[0,T]} |\mathcal{Z}^n(t)| > \epsilon\right) \le -\frac{\epsilon^2}{2\Theta_\epsilon}, \text{ if } \Theta_\epsilon > 0$$

and is equal to $-\infty$, when $\Theta_\epsilon = 0$. This proves the result. □

The following is a simple corollary of the above theorem.

Corollary A.1. *Suppose* (A.2) *holds and* $\lim_{n\to\infty} \frac{N^n(T)}{n} = 0$, *then*

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbb{P}\left(\sup_{0\le t\le T} |\mathcal{Z}^n(t)| > \epsilon\right) = -\infty\,.$$

Proof. From the arguments of the proof of [20, Lemma 2.2] and (A.2), we can conclude that

$$\mathcal{M} \doteq \mathbb{E}\Big[\sup_{i\in\mathbb{N},\, t,s\in[0,T]} Z_i^n(t,s,\vartheta_i)^2\Big] < \infty.$$

From the hypothesis of the corollary and the boundedness of $\mathcal{M}$,

$$\Theta_\epsilon \le \mathcal{M} \limsup_{n\to\infty} \frac{N^n(T)}{n} = 0.$$

Therefore, from Theorem A.1, we have the desired result. □

Appendix B. Sketch proof for the LDP results in Section 2.4

In this section, we give a sketch proof for the LDP result in the high intensity regime, and the proof for that in the conventional time-space scaling regime can be done in similar steps and hence is omitted for brevity. The sketch will be divided into two parts: (i) proving the LDP in $\mathcal{D}_T$ under the topology of pointwise convergence and identifying the corresponding rate function using Gärtner-Ellis theorem [18, Theorem 2.3.6] and Dawson-Gärtner theorem [18, Theorem 4.6.1]; and (ii) proving the exponential tightness in $(\mathcal{D}_T, J_1)$.

We now proceed with part (i). Recall $\bar{X}^n(t)$ in (2.6) and $A^n(t)$ in (2.7). Also recall that $\Psi_A(\rho) \doteq \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{\rho A^n(1)}]$.

We first show the LDP for the finite-dimensional distributions of $\bar{X}^n$. For every $N \ge 1$ and $0 = t_0 < t_1 < t_2 < t_3 < \ldots < t_N = T$, define $\bar{X}_N^n = \left(\bar{X}^n(t_1), \bar{X}^n(t_2), \bar{X}^n(t_3), \ldots, \bar{X}^n(t_N)\right)$. We show that the family of $\mathbb{R}^N$–valued random variables $\{\bar{X}_N^n\}_{n\in\mathbb{N}}$ satisfies LDP with rate n and rate function $I_N^{\mathrm{LDP}} : \mathbb{R}^N \to [0,\infty]$ given by

$$I_N^{\mathrm{LDP}}(\underline{x}) \doteq \sup_{\underline{\varrho}\in\mathbb{R}^N} \left\{ \sum_{i=1}^N (x_{i+1} - x_i) \sum_{j=i}^N \varrho_j - \sum_{i=1}^N \int_{t_i}^{t_{i+1}} \Psi_A\left(\log \mathbb{E}\left[\exp\Big(\sum_{j=i}^N \varrho_j H(t_j, s, \vartheta_1)\Big)\right]\right) ds \right\}.$$

Since (2.27) and Assumption 2.2 hold, a direct calculation of the log moment generating function of $\bar{X}^n_N$ by conditioning, leads to

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\Big[\exp\Big(n\sum_{i=1}^N \varrho_i \bar{X}^n(t_i)\Big)\Big] = \sum_{i=1}^N \int_{t_i}^{t_{i+1}} \Psi_A\Bigg(\log \mathbb{E}\Bigg[\exp\Big(\sum_{j=i}^N \varrho_j H(t_j, s, \vartheta_1)\Big)\Bigg]\Bigg) ds$$

for $\{\varrho_i\}_{i=1}^N \subset \mathbb{R}$. From Gärtner-Ellis theorem, we know that $\{\bar{X}^n_N\}_{n\in\mathbb{N}}$ now satisfies LDP with rate n and rate function given by

$$\begin{aligned} I_N^{\mathrm{LDP}}(\underline{x}) &= \sup_{\varrho\in\mathbb{R}^N}\Bigg(\langle \underline{\varrho}, \underline{x}\rangle - \sum_{i=1}^N \int_{t_i}^{t_{i+1}} \Psi_A\Bigg(\log \mathbb{E}\Bigg[\exp\Big(\sum_{j=i}^N \varrho_j H(t_j, s, \vartheta_1)\Big)\Bigg]\Bigg) ds\Bigg) \\ &= \sup_{\varrho\in\mathbb{R}^N}\Bigg(\sum_{i=1}^N (x_{i+1}-x_i)\sum_{j=i}^N \varrho_j - \sum_{i=1}^N \int_{t_i}^{t_{i+1}} \Psi_A\Bigg(\log \mathbb{E}\Bigg[\exp\Big(\sum_{j=i}^N \varrho_j H(t_j, s, \vartheta_1)\Big)\Bigg]\Bigg) ds\Bigg). \end{aligned}$$

Using the Dawson-Gaẗner theorem, we can now conclude that $\{\bar{X}^n\}_{n\in\mathbb{N}}$ satisfies LDP in $\mathcal{D}_T$ under the topology of pointwise convergence with rate n and the rate function $I^{\mathrm{LDP}}[\bar{X}] : \mathcal{D}_T \to [0,\infty]$ given by

$$\begin{aligned} I^{\mathrm{LDP}}[\bar{X}](\phi) &\doteq \sup_{0=t_0<t_1<t_2\ldots<t_N=T} I_f^N(\phi(t_1), \phi(t_2), \phi(t_3), \ldots, \phi(t_N)) \\ &= \begin{cases} \sup_{\rho\in\mathcal{C}_T} \int_0^T \Bigg(\dot{\phi}(t)\int_t^T \rho(u)du - \Psi_A\Bigg(\log \mathbb{E}\Bigg[\exp\Bigg(\int_t^T \rho(s)H(s,t,\vartheta_1)ds\Bigg)\Bigg]\Bigg)\Bigg) dt, \\ \qquad\qquad \text{if } \phi \in \mathcal{AC}_0, \\ \infty, \ \text{otherwise.} \end{cases} \end{aligned}$$

We next prove part (ii), that is, to show that $\{\bar{X}^n\}_{n\in\mathbb{N}}$ is exponentially tight. By [51, Theorem A.3] (which also holds in the case of LDP), it suffices for us to prove that for $\epsilon > 0$,

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \sup_{t\in[0,T]} \frac{1}{n} \log \mathbb{P}\Bigg(\sup_{0\le s\le\delta} |\bar{X}^n(t+s) - \bar{X}^n(t)| > \epsilon\Bigg) = -\infty.$$

Fix $\epsilon, \delta > 0$ and $0 \le s \le \delta$. We can write

$$\bar{X}^n(t+s) - \bar{X}^n(t) = \mathcal{J}_1^n(t, t+s) + \mathcal{J}_2^n(t, t+s)\,,$$

where

$$\begin{aligned} \mathcal{J}_1^n(t, t+s) &:= \frac{1}{n}\sum_{i=1}^{A^n(t)} \big(H(t+s, \tau_i^n, \vartheta_i) - H(t, \tau_i^n, \vartheta_i)\big)\,, \\ \mathcal{J}_2^n(t, t+s) &:= \frac{1}{n}\sum_{i=A^n(t)+1}^{A^n(t+s)} H(t+s, \tau_i^n, \vartheta_i)\,. \end{aligned}$$

It then suffices to show that for $i = 1, 2$,

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \sup_{t\in[0,T]} \frac{1}{n} \log \mathbb{P}\Bigg(\sup_{0\le s\le\delta} |\mathcal{J}_i^n(t, t+s)| > \epsilon/2\Bigg) = -\infty. \tag{B.1}$$

For $\mathcal{J}_1^n(t, t+s)$, we can use an analogous result to Theorem A.1 in the context of LDP. Let $\widetilde{\tau}^n_{t,\delta} \doteq \inf\{s > 0 : |\mathcal{J}_1^n(t, t+s)| > \frac{\epsilon}{2}\}$ and $\widetilde{\tau}^n_{t,\delta} = \delta + 1$, whenever $|\mathcal{J}_1(t, t+s)| < \frac{\epsilon}{2}$, for $0 \le s \le \delta$. We have

$$\mathbb{P}\Bigg(\sup_{0\le s\le\delta} |\mathcal{J}_1^n(t, t+s)| > \frac{\epsilon}{2}\Bigg) = \mathbb{E}[\widetilde{\tau}^n_{t,\delta} \le \delta] = e^{\rho n \frac{\epsilon}{2}} \mathbb{E}\Big[\exp\Big(\rho n |\mathcal{J}_1^n(t, t+\widetilde{\tau}^n_{t,\delta})|\Big) \mathbb{1}_{\{\widetilde{\tau}^n_{t,\delta}\le\delta\}}\Big]$$

$$\leq e^{\rho n \frac{\epsilon}{2}} \mathbb{E}\Big[\exp\Big(\rho n |\mathcal{J}_1^n(t, t + (\widetilde{\tau}_{t,\delta}^n \wedge \delta))|\Big)\Big].$$

From the above, we have

$$\frac{1}{n} \log \mathbb{P}\left(\sup_{0\leq s\leq \delta} |\mathcal{J}_1^n(t,t+s)| > \frac{\epsilon}{2}\right) \leq -\frac{\rho\epsilon}{2} + \frac{1}{n}\log \mathbb{E}\Big[\exp\Big(\rho n|\mathcal{J}_1^n(t,t+(\widetilde{\tau}_{t,\delta}^n \wedge \delta))|\Big)\Big].$$

If we can show that

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \sup_{t\in[0,T]} \frac{1}{n} \log \mathbb{E}\Big[\exp\Big(\rho n|\mathcal{J}_1^n(t,t+(\widetilde{\tau}_{t,\delta}^n \wedge \delta))|\Big)\Big] = 0, \tag{B.2}$$

then taking $\rho \uparrow \infty$ gives (B.1) for $i = 1$. This is what we will do now. Consider

$$\begin{aligned}
&\mathbb{E}\Big[\mathbb{E}\Big[\exp\Big(\rho n|\mathcal{J}_1^n(t,t+(\widetilde{\tau}_{t,\delta}^n \wedge \delta))|\Big)\Big| A^n\Big]\Big] \\
&\quad \leq \mathbb{E}\Big[\prod_{i=1}^{A^n(t)} \mathbb{E}\Big[\exp\Big(\rho|H(t+(\widetilde{\tau}_{t,\delta}^n \wedge \delta), \tau_i^n, \vartheta_i) - H(t,\tau_i^n,\vartheta_i)|\Big)\Big| A^n\Big]\Big] \\
&\quad \leq \mathbb{E}\Big[\exp\Big(\sum_{i=1}^{A^n(t)} \Phi_\delta(\tau_i^n)\Big)\Big],
\end{aligned}$$

where

$$\Phi_\delta(s) \doteq \log \mathbb{E}\left[\exp\left(\rho \sup_{0\leq u\leq\delta} \sup_{t\in[0,T]} |H(t+u,s,\vartheta_i) - H(t,s,\vartheta_i)|\right)\right].$$

Again using [24, Theorem 2.4], we have

$$\limsup_{\delta\downarrow 0} \limsup_{n\to\infty} \sup_{t\in[0,T]} \frac{1}{n}\log\mathbb{E}\Big[\exp\Big(\rho n|\mathcal{J}_1^n(t,t+(\widehat{\tau}_{t,\delta}^n\wedge\delta))|\Big)\Big] \leq \limsup_{\delta\downarrow 0} \int_0^T \Psi_A(\Phi_\delta(s))ds = 0.$$

In the above, we use the fact that (2.28) holds.

We now move on to $\mathcal{J}_2^n(t,t+s)$. We can again use an analogous result to Theorem A.1 in the context of LDP. Let $\widehat{\tau}_{t,\delta}^n \doteq \inf\{s > 0 : |\mathcal{J}_2^n(t,t+s)| > \frac{\epsilon}{2}\}$ and $\widehat{\tau}_{t,\delta}^n = \delta + 1$, whenever $|\mathcal{J}_2(t,t+s)| < \frac{\epsilon}{2}$, for $0 \leq s \leq \delta$. We have

$$\begin{aligned}
\mathbb{P}\left(\sup_{0\leq s\leq\delta} |\mathcal{J}_2^n(t,t+s)| > \frac{\epsilon}{2}\right) &= \mathbb{E}[\widehat{\tau}_{t,\delta}^n \leq \delta] \\
&\leq e^{\rho n\frac{\epsilon}{2}} \mathbb{E}\Big[\exp\Big(\rho n|\mathcal{J}_2^n(t,t+\widehat{\tau}_{t,\delta}^n)|\Big)\mathbb{1}_{\{\widehat{\tau}_{t,\delta}^n\leq\delta\}}\Big] \\
&\leq e^{\rho n\frac{\epsilon}{2}} \mathbb{E}\Big[\exp\Big(\rho n|\mathcal{J}_2^n(t,t+(\widehat{\tau}_{t,\delta}^n\wedge\delta))|\Big)\Big].
\end{aligned}$$

From the above, we have

$$\frac{1}{n}\log\mathbb{P}\left(\sup_{0\leq s\leq\delta}|\mathcal{J}_2^n(t,t+s)| > \frac{\epsilon}{2}\right) \leq -\frac{\rho\epsilon}{2} + \frac{1}{n}\log\mathbb{E}\Big[\exp\Big(\rho n|\mathcal{J}_2^n(t,t+(\widehat{\tau}_{t,\delta}^n\wedge\delta))|\Big)\Big].$$

Next, we obtain

$$\begin{aligned}
\mathbb{E}\Big[\mathbb{E}\Big[\exp\Big(\rho n|\mathcal{J}_2^n(t,t+(\widehat{\tau}_{t,\delta}^n\wedge\delta))|\Big)\Big|A^n\Big]\Big] &\leq \mathbb{E}\Big[\prod_{i=A^n(t)+1}^{A^n(t+\delta)} \mathbb{E}\Big[\exp\Big(\rho|H(t+(\widehat{\tau}_{t,\delta}^n\wedge\delta),\tau_i^n,\vartheta_i)|\Big)\Big|A^n\Big]\Big] \\
&\leq \mathbb{E}\Big[\exp\Big(\sum_{i=A^n(t)+1}^{A^n(t+\delta)} \widehat{\Phi}(\tau_i^n)\Big)\Big],
\end{aligned}$$

where

$$\widehat{\Phi}(u) \doteq \log\mathbb{E}\Big[\exp\Big(\rho\sup_{t\in[0,T]}|H(t,u,\vartheta_i)|\Big)\Big].$$

Again using [24, Theorem 2.4], we have

$$\limsup_{\delta\downarrow 0}\limsup_{n\to\infty}\sup_{t\in[0,T]}\frac{1}{n}\log\mathbb{E}\Big[\exp\Big(\rho n|\mathcal{J}_2^n(t,t+(\widehat{\tau}_{t,\delta}^n\wedge\delta))|\Big)\Big]\leq\limsup_{\delta\downarrow 0}\int_t^{t+\delta}\Psi_A(\widehat{\Phi}(u))du=0.$$

In the above, we use the fact that (2.27) holds. Then taking $\rho\uparrow\infty$ gives our desired result. This proves (B.1) for $i=2$ and also the desired exponential tightness.

Acknowledgement

This work is partly funded by the NSF Grant DMS 2216765.

References

[1] S. R. Anugu and G. Pang. Large deviations and long-time behavior of stochastic fluid queues with generalized fractional brownian motion input. *Queueing Systems*, 105(1):47–98, 2023.
[2] A. Berlinet and C. Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.
[3] Z. Bezemek and K. Spiliopoulos. Moderate deviations for fully coupled multiscale weakly interacting particle systems. *arXiv:2202.08403*, 2022.
[4] H. Biermé and A. Desolneux. Crossings of smooth shot noise processes. *The Annals of Applied Probability*, 22(6):2240 – 2281, 2012.
[5] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 1968.
[6] J. Blanchet, X. Chen, and H. Lam. Two-parameter sample path large deviations for infinite-server queues. *Stochastic Systems*, 4(1):206–249, 2014.
[7] L. Bondesson. Shot-noise processes and distributions. *Encyclopedia of statistical sciences*, 2004.
[8] A. A. Borovkov. On large deviation principles for compound renewal processes. *Mathematical Notes*, 106, 2019.
[9] S. Bourguin, T. Dang, and K. Spiliopoulos. Moderate deviation principle for multiscale systems driven by fractional Brownian motion. *Journal of Theoretical Probability*, 36(1):1, 2023.
[10] O. Boxma and M. Mandjes. Shot-noise queueing models. *Queueing Systems*, 99(1-2):121–159, 2021.
[11] P. J. Brockwell, S. I. Resnick, and R. L. Tweedie. Storage processes with general release rule and additive inputs. *Advances in Applied Probability*, 14(2):392–433, 1982.
[12] A. Budhiraja, J. Chen, and P. Dupuis. Large deviations for stochastic partial differential equations driven by a Poisson random measure. *Stochastic Processes and their Applications*, 123(2):523–560, 2013.
[13] A. Budhiraja and P. Dupuis. *Analysis and approximation of rare events*. Springer, 2019.
[14] A. Budhiraja, P. Dupuis, and A. Ganguly. Moderate deviation principles for stochastic differential equations with jumps. *Annals of Probability*, 44(3):1723–1775, 2016.
[15] A. Budhiraja and P. Nyquist. Large deviations for multidimensional state-dependent shot-noise processes. *Journal of Applied Probability*, 52(4):1097–1114, 2015.
[16] A. Budhiraja and R. Wu. Moderate deviation principles for weakly interacting particle systems. *Probability Theory and Related Fields*, 168(3-4):721–771, 2017.
[17] A. Chiarini and M. Fischer. On large deviations for small noise Itô processes. *Advances in Applied Probability*, 46(4):1126–1147, 2014.
[18] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*. Springer Science & Business Media, 2009. Stochastic Modelling and Applied Probability (SMAP, Volume 38).
[19] K. R. Duffy and G. L. Torrisi. Sample path large deviations of Poisson shot noise with heavy-tailed semiexponential distributions. *Journal of Applied Probability*, 48(3):688–698, 2011.

[20] P. Eichelsbacher and M. Löwe. Moderate deviations for I.I.D. random variables. *ESAIM: Probability and Statistics*, 7:209–218, 2003.

[21] J. Feng and T. G. Kurtz. *Large deviations for stochastic processes.* American Mathematical Society, 2006.

[22] A. Ganesh, C. Macci, and G. Torrisi. Sample path large deviations principles for Poisson shot noise processes and applications. *Electronic Journal of Probability*, 10:1026–1043, 2005.

[23] I. Gasteratos, M. Salins, and K. Spiliopoulos. Moderate deviations for systems of slow–fast stochastic reaction–diffusion equations. *Stochastics and Partial Differential Equations: Analysis and Computations*, 11(2):503–598, 2023.

[24] P. W. Glynn. Large deviations for the infinite server queue in heavy traffic. *Stochastic Networks*, 71:387–395, 1995.

[25] P. W. Glynn and W. Whitt. Large deviations behavior of counting processes and their inverses. *Queueing Systems*, 17(1):107–128, 1994.

[26] L. Heinrich and V. Schmidt. Normal convergence of multidimensional shot noise and rates of this convergence. *Advances in Applied Probability*, 17(4):709–730, 1985.

[27] F. W. Huffer. Inequalities for the $M/G/\infty$ queue and related shot noise processes. *Journal of Applied Probability*, 24(4):978–989, 1987.

[28] D. L. Iglehart. Weak convergence of compound stochastic process, I. *Stochastic Processes and their Applications*, 1(1):11–31, 1973.

[29] A. Iksanov. Functional limit theorems for renewal shot noise processes with increasing response functions. *Stochastic Processes and their Applications*, 123(6):1987–2010, 2013.

[30] A. Iksanov, A. Marynych, and M. Meiners. Limit theorems for renewal shot noise processes with eventually decreasing response functions. *Stochastic Processes and their Applications*, 124(6):2132–2170, 2014.

[31] A. Iksanov, A. Marynych, and M. Meiners. Asymptotics of random processes with immigration I: Scaling limits. *Bernoulli*, 23(2):1233 – 1278, 2017.

[32] A. Iksanov, A. Marynych, and M. Meiners. Asymptotics of random processes with immigration II: Convergence to stationarity. *Bernoulli*, 23(2):1279 – 1298, 2017.

[33] A. Jacquier and A. Pannier. Large and moderate deviations for stochastic Volterra systems. *Stochastic Processes and their Applications*, 149:142–187, 2022.

[34] C. Klüppelberg and C. Kühn. Fractional Brownian motion as a weak limit of Poisson shot noise processes–with applications to finance. *Stochastic Processes and their Applications*, 113(2):333–351, 2004.

[35] C. Klüppelberg and T. Mikosch. Explosive Poisson shot noise processes with applications to risk reserves. *Bernoulli*, 1(1-2):125 – 147, 1995.

[36] C. Klüppelberg, T. Mikosch, and A. Schärf. Regular variation in the mean and stable limits for Poisson shot noise. *Bernoulli*, 9(3):467–496, 2003.

[37] T. Konstantopoulos and S.-J. Lin. Macroscopic models for long-range dependent network traffic. *Queueing systems*, 28(1):215–243, 1998.

[38] D. T. Koops, O. J. Boxma, and M. Mandjes. Networks of $\cdot/G/\infty$ queues with shot-noise-driven arrival intensities. *Queueing Systems*, 86:301–325, 2017.

[39] S. Lowen and M. Teich. Power-law shot noise. *IEEE Transactions on Information Theory*, 36(6):1302–1318, 1990.

[40] R. B. Lund. The stability of storage models with shot noise input. *Journal of Applied Probability*, 33(3):830–839, 1996.

[41] C. Macci, G. Stabile, and G. L. Torrisi. Lundberg parameters for non standard risk processes. *Scandinavian Actuarial Journal*, 2005(6):417–432, 2005.

[42] C. Macci and G. L. Torrisi. Asymptotic results for perturbed risk processes with delayed claims. *Insurance: Mathematics and Economics*, 34(2):307–320, 2004.

[43] M. R. Morse and K. Spiliopoulos. Moderate deviations for systems of slow-fast diffusions. *Asymptotic Analysis*, 105(3-4):97–135, 2017.

[44] G. Pang and M. S. Taqqu. Nonstationary self-similar Gaussian processes as scaling limits of power-law shot noise processes and generalizations of fractional Brownian motion. *High Frequency*, 2(2):95–112, 2019.

[45] G. Pang and Y. Zhou. Two-parameter process limits for an infinite-server queue with arrival dependent service times. *Stochastic Processes and their Applications*, 127(5):1375–1416, 2017.

[46] G. Pang and Y. Zhou. Functional limit theorems for a new class of non-stationary shot noise processes. *Stochastic Processes and their Applications*, 128(2):505–544, 2018.

[47] A. Papoulis. High density shot noise and Gaussianity. *Journal of Applied Probability*, 8(1):118–127, 1971.

[48] M. Parulekar and A. M. Makowski. $M/G/\infty$ input processes: A versatile class of models for network traffic. *Proceedings of INFOCOM '97*, 2:419–426 vol.2, 1997.

[49] A. A. Puhalskii. On functional principle of large deviations. *Vol. 1 Proceedings of the Bakuriani Colloquium in Honour of Yu.V. Prohorov, Bakuriani, Georgia, USSR, 24 February–4 March, 1990. Editors V. V. Sazonov and T. Shervashidze*, pages 198–218, 1991.

[50] A. A. Puhalskii. Large deviations of semimartingales via convergence of the predictable characteristics. *Stochastics: An International Journal of Probability and Stochastic Processes*, 49(1-2):27–85, 1994.

[51] A. A. Puhalskii. Moderate deviations of many-server queues in the Halfin-Whitt regime and weak convergence methods. *arXiv:2305.01612*, 2023.

[52] A. A. Puhalskii and W. Whitt. Functional large deviation principles for first-passage-time processes. *The Annals of Applied Probability*, 7(2):362 – 381, 1997.

[53] S. O. Rice. Mathematical analysis of random noise. *The Bell System Technical Journal*, 23(3):282–332, 1944.

[54] G. Torrisi. Simulating the ruin probability of risk processes with delay in claim settlement. *Stochastic Processes and their Applications*, 112(2):225–244, 2004.