Averaging Principles for Markovian Models of Plasticity

Philippe Robert\textsuperscript{1} · Gaëtan Vignoud\textsuperscript{1,2}

Received: 5 January 2021 / Accepted: 26 May 2021 / Published online: 9 June 2021

Abstract

In this paper we consider a stochastic system with two connected nodes, whose unidirectional connection is variable and depends on point processes associated to each node. The input node is represented by an homogeneous Poisson process, whereas the output node jumps with an intensity that depends on the jumps of the input nodes and the connection intensity. We study a scaling regime when the rate of both point processes is large compared to the dynamics of the connection. In neuroscience, this system corresponds to a neural network composed by two neurons, connected by a single synapse. The strength of this synapse depends on the past activity of both neurons, the notion of synaptic plasticity refers to the associated mechanism. A general class of such stochastic models has been introduced in Robert and Vignoud (Stochastic models of synaptic plasticity in neural networks, 2020, arxiv: 2010.08195) to describe most of the models of long-term synaptic plasticity investigated in the literature. The scaling regime corresponds to a classical assumption in computational neuroscience that cellular processes evolve much more rapidly than the synaptic strength. The central result of the paper is an averaging principle for the time evolution of the connection intensity. Mathematically, the key variable is the point process, associated to the output node, whose intensity depends on the past activity of the system. The proof of the result involves a detailed analysis of several of its unbounded additive functionals in the slow-fast limit, and technical results on interacting shot-noise processes.

Communicated by Pablo A Ferrari.

Supported by PhD grant of École Normale Supérieure, ENS-PSL.

Philippe Robert
Philippe.Robert@inria.fr
http://www-rocq.inria.fr/who/Philippe.Robert

Gaëtan Vignoud
Gaetan.Vignoud@inria.fr

\textsuperscript{1} INRIA Paris, 2 rue Simone Iff, 75589 Paris Cedex 12, France
\textsuperscript{2} INRIA Paris, Center for Interdisciplinary Research in Biology (CIRB) - Collège de France (CNRS UMR 7241, INSERM U1050), 11 Place Marcelin Berthelot, 75005 Paris, France
1 Introduction

Neurons exchange electrical and chemical signals at specific spots, called synapses. The synaptic transmission between neural cells is unidirectional, in the sense that, the signal goes from the pre-synaptic neuron towards the post-synaptic neuron. This interaction is modulated over time, and particularly by the concomitant activity of both neurons. In Robert and Vignoud [28] we have introduced a general class of mathematical models to represent and study synaptic plasticity mechanisms.

A basic model to investigate such phenomenon consists of a pre-synaptic neuron connected through a synapse to a post-synaptic neuron. The associated stochastic process is described by two random variables \((X, W)\) and the spiking activity of each neuron is represented by a point process.

(a) Point process for pre-synaptic spikes: \(N_\lambda\).

This point process is associated to the instants when the pre-synaptic neuron is spiking, i.e. when it transmits a chemical/electrical signal to the post-synaptic neuron via the synapse. We assume that \(N_\lambda\) is an homogeneous Poisson process with rate \(\lambda\).

(b) Synaptic weight: \(W\).

It describes the strength of the connection from the pre-synaptic neuron to the post-synaptic neuron.

(c) Post-synaptic membrane potential: \(X\).

This variable is for the current activity of the post-synaptic neuron. At a jump of \(N_\lambda\), the membrane potential \(X\) is incremented by \(W\), where \(W\) is the current synaptic weight.

(d) Point process for post-synaptic spikes: \(N_{\beta,X}\).

In state \(X=x\), the post-synaptic neuron emits a spike at rate \(\beta(x)\), where \(\beta\) is the activation function of the neural cell. The point process associated to these instants is an inhomogeneous Poisson process denoted by \(N_{\beta,X}\). This is a key variable of the stochastic model. See Relation (6) for a formal definition.

As explained in Robert and Vignoud [28], for some synaptic mechanisms, the time evolution of \(W\) may depend, in a complex way, on past spiking times of both adjacent neurons. In our model it is a functional of the point processes \(N_\lambda\) and \(N_{\beta,X}\). The model relies on two clearly stated hypotheses: the effect of plasticity only depends on the relative timing of the activity of both neuronal cells and is seen over the synaptic strength on long timescales.

Accordingly, the purpose of the current paper is to prove limit theorems for a scaled version of the corresponding stochastic processes \((X(t), W(t))\).

1.1 A Simple Model

We begin by the description of a simplified model to highlight the role of the different components in these stochastic models. We consider the following set of Stochastic Differential Equations (SDEs),

\[
\begin{align*}
\mathrm{d}X(t) & = -X(t)\mathrm{d}t + W(t-)N_\lambda(\mathrm{d}t), \\
\mathrm{d}Z(t) & = -\gamma Z(t)\mathrm{d}t + B_1N_\lambda(\mathrm{d}t) + B_2N_{\beta,X}(\mathrm{d}t), \\
\mathrm{d}W(t) & = Z(t-)N_{\beta,X}(\mathrm{d}t),
\end{align*}
\]

where \(h(t-)\) is the left-limit of the function \(h\) at \(t>0\) and, for \(i=1,2\), \(B_i\in\mathbb{R}_+\).
In this model, the time evolution of \((W(t))\) is overly simplified, plasticity processes are modeled by an increase of the synaptic weight \(W\) at each jump of \(N_{\beta,X}\) by the value of \((Z(t))\).

The process \((Z(t))\) encodes the past spiking activity of both neurons through an additive functional of \(N_x\) and \(N_{\beta,X}\) with an exponential decay factor \(\gamma>0\). The process \((Z(t))\) is associated to a cellular process, in the general model this is a multi-dimensional process.

The main process of interest is the strength of the synaptic connection \((W(t))\). It has been extensively studied both in experimental neuroscience and in physics, there are nevertheless few rigorous mathematical results on the dynamical evolution of \(W\).

From a mathematical perspective, the variables \((X(t) , Z(t) , W(t))\), solutions of SDE (1) are central to the model. The point process \(N_{\beta,X}\) is nevertheless the key component of the system since it drives the time evolution of \((Z(t))\) and \((W(t))\) and, consequently, of \((X(t))\). Most of the mathematical difficulties of our paper are related to asymptotic estimates of linear functionals of \(N_{\beta,X}\).

The scaling approach of this paper follows from the fact that the model can be expressed as a slow-fast system. An important property of this system is that neuronal processes, associated to a cellular process, in the general model this is a multi-dimensional process.

1.2 Averaging Principles

From a mathematical point of view, there exist for a discussion on this topic. (See Kempter et al. [18] for example.)

The main process of interest is the strength of the synaptic connection \((W(t))\). It has been extensively studied both in experimental neuroscience and in physics, there are nevertheless few rigorous mathematical results on the dynamical evolution of \(W\).

The main goal of the present paper is to establish a limit result, or averaging principle, for \((W(t))\) when \(\varepsilon \to 0\) for a general class of synaptic plasticity models.

We denote by \((X^w, Z^w)\) the solution \((X(t) , Z(t))\) of Relation (2) when the process \((W(t))\) is constant and equal to \(w\). Under appropriate conditions, it has a unique equilibrium distribution \(\Pi_w\). The averaging principle for the simple model can be expressed as follows.

There exists \(S_0 \in (0, +\infty]\), such that the processes \((W^{\varepsilon}(t), 0 \leq t < S_0)\) is tight for the convergence in distribution when \(\varepsilon\) goes to 0, and any limiting point \((w(t), 0 \leq t < S_0)\) satisfies the following integral equation,

\[
W(t) = w(0) + \int_{0}^{t} \int_{\mathbb{R}_+^2} z \beta(x) \Pi_w(ds) dx dz, \quad t \in [0, S_0].
\]
See Sect. 5 of Chapter 1 of Billingsley [3] for general results on tightness properties and convergence in distribution.

We discuss now some of the technical difficulty to derive such results for our model. The integration of SDE (2) gives the relation

$$W_\varepsilon(t) = W_\varepsilon(0) + \int_0^t Z_\varepsilon(s) \varepsilon N_{\beta/\varepsilon, X_\varepsilon}(ds).$$

The tightness of the family of processes $(W_\varepsilon(t))$ is equivalent to the tightness of

$$\left( \int_0^t Z_\varepsilon(s) \varepsilon N_{\beta/\varepsilon, X_\varepsilon}(ds) \right)$$

(4)

A general approach to prove averaging principles is presented in Kurtz [21] for jump processes. See Chapter 7 of Freidlin and Wentzell [8] and especially Papanicolalou et al. [24] for an introduction to averaging principles. If we had an expression of the type

$$\left( \int_0^t F(X_\varepsilon(s), Z_\varepsilon(s))ds \right),$$

where $s \rightarrow F(X_\varepsilon(s), Z_\varepsilon(s))$ is a bounded continuous function on $\mathbb{R}_+$, a direct use of the results of Kurtz [21], Lemma 1.3 and 1.5, would give the desired tightness. This is the case of Helson [14] for the time-elapsed model of plasticity for which this representation holds, one of the few rigorous results in this domain.

It does not seem to be possible to handle functionals of the form (4) in this way. The process $(Z_\varepsilon(s))$ is clearly not bounded and neither is the intensity of the point process $N_{\beta/\varepsilon, X_\varepsilon}$, since $(\beta(X_\varepsilon(s))$ is also not bounded. Remember that fast processes are on a rapid time scale so visit their state space “quickly”, the values of the integrals of (4) can be large and therefore must be controlled in an appropriate way.

Two other interesting properties emerge from the stochastic averaging result from Sect. 4.

(a) **UNIQUENESS.**

If Relation (3) has a unique solution for a given initial state, a result for the convergence in distribution of $(W_\varepsilon(t))$ when $\varepsilon$ goes to 0 is therefore obtained. Uniqueness holds if the integrand, with respect to $s$, of the right-hand side of Relation (3) is locally Lipschitz as a function of $w(s)$. Regularity properties of the invariant distribution $\Pi_w$ as a function of $w$ need to be verified and this is not a concern in the case of our simple model. We will consider in fact much more general models for $(X^w, Z^w)$, when $Z^w$ a multi-dimensional process in particular. We did not try to state a set of conditions that can ensure the desired regularity properties of the corresponding $\Pi_w$. The proof of the Harris ergodicity of $(X^w, Z^w)$ for a fixed $w$ of Sect. C of Appendix, though not really difficult, is already cumbersome.

The proof of Proposition 20 for the simple model gives an example of how this property can be established. In a general context, this kind of result is generally proved via the use of a common Lyapounov function for $(X^w, Z^w)$ for all $w$ is in the neighborhood of some $w_0>0$. See Has’minski˘ı [11], for example. Uniqueness results have already been obtained in Sects. 5 and 6 of Robert and Vignoud [28] for several important practical cases. In Sect. 8 we investigate these questions for our simple model.

(b) **BLOW-UP PHENOMENON.**

The convergence properties are stated on a fixed time interval $[0, S_0)$. For some models, the variable $S_0$ cannot be taken as $+\infty$, see the example of Sect. 4 and Proposition 22. More specifically, the limit in distribution of $(W_\varepsilon(t))$ as $\varepsilon$ goes to 0 blows-up, i.e. hits
infinity in finite time. An analogue property holds for some mathematical models of large populations of neural cells with fixed synaptic strengths. See Cáceres et al. [4] for example, where the blow-up phenomenon is the result of mutually exciting dynamics of populations of neural cells. In our case, the strengthening of the connection may grow without bounds when the activation function $\beta$ has a linear growth. See Proposition 22 of Sect. 8.

1.3 A Brief Description of the General Model of Synaptic Plasticity

We shortly describe the general setting of the models investigated in this paper. See Sect. 2 for a detailed presentation.

(a) The process $(X(t))$. The output neuron follows leaky-integrate dynamics as in Equation (1). In addition, the influence of a post-synaptic spike $N_{\beta,X}$ at time $t>0$ is represented as a drop $-g(X(t-))$ of the post-synaptic potential after the spike;

(b) The process $(Z(t)) = (Z_i(t))$ is a multi-dimensional process satisfying the same type of ODE as in our simple case but with the constants $B_1$ and $B_2$ being replaced by functions $k_1$ and $k_2$ of $Z(t)$. A constant drift term $k_0$ is also added to the dynamics. The $i$th component $(Z_i(t))$ satisfies an SDE of the type

$$dZ_i(t) = (-\gamma_i Z_i(t)+k_{0,i})dt+k_{1,i}(Z(t-))N_\lambda(dt)+k_{2,i}(Z(t-))N_{\beta,X}(dt).$$

(c) Evolution of $(W(t))$. The dependence is more sophisticated since it involves two additional processes $(\Omega_p, \Omega_d)$. The first one, $(\Omega_p(t))$ integrates, with an exponential decay $\alpha$ a linear combination of the processes leading to potentiation, i.e. to increase the synaptic weight. The process $(\Omega_d(t))$ has a similar role for depression, i.e. to decrease the synaptic weight. They are expressed as, for $a \in \{p, d\}$,

$$\Omega_a(t) = \int_0^t e^{-\alpha(t-s)} \left[ n_{0,a}(Z(s))ds + n_{1,a}(Z(s-))N_\lambda(ds) + n_{2,a}(Z(s-))N_{\beta,X}(ds) \right].$$

The changes of $(Z(t))$ are thus integrated “smoothly” in the evolution of $(W(t))$ in agreement with measurements of the biological literature. See Appendix A of [28]. Finally, $(W(t)) \in K_W$ verifies

$$dW(t) = M(\Omega_p(t), \Omega_d(t), W(t)) \, dt,$$

where $K_W \subset \mathbb{R}$ represents the synaptic weight domain, and the functional $M$ is such that $W(t)$ stays in $K_W$ for all $t \geq 0$.

It has been shown in Sect. 3 of [28] that these models encompass most classical STDP models from statistical physics. The multiple coordinates of $(Z(t))$ can be interpreted as the concentrations of chemical components implicated in plasticity, that are created/suppressed by spiking mechanisms.

1.4 Links to Non-linear Hawkes Point Processes

The spiking instants of a neuron can also be seen as a self-exciting point process since its instantaneous jump rate depends on past instants of its jumps. This corresponds to the class of Hawkes point process $\mathcal{M}$ on $\mathbb{R}_+$ associated to a function $\phi$ and exponential decay $\gamma$. More
precisely, it is a non-homogeneous Poisson point process $\mathcal{M}$ whose intensity function $(\lambda(t))$ is given by
\[
(\lambda(t)) = \left( \phi \left( \int_0^t e^{-\gamma(t-s)} \mathcal{M}(ds) \right) \right).
\]
These processes have received a lot of attention from the mathematical literature, for some time now. They are mainly used in models of mathematical finance, but also in neurosciences. See the pioneering works of Hawkes and Oakes [13] and Kerstan [19].

A special case of the first equation of Relation (1) is, for $w \geq 0$,
\[
dX(t) = -X(t)dt + wN_\lambda(dt) - N_\beta,X(dt),
\]
if $X(0)=0$, Lemma 5 below gives the representation
\[
X(t) = w \int_0^t e^{-(t-s)}N_\lambda(ds) - \int_0^t e^{-(t-s)}N_\beta,X(ds), \forall t \geq 0.
\]
Hence, $N_\beta,X$ can be seen as an extended Hawkes process with activation function $\beta$ and exponential decay.

In the system of equations (1), $(X(t))$ and $(Z(t))$ can also be represented as a multi-dimensional Hawkes processes. See Hawkes [12]. However in our model, an important feature not present in studies of Hawkes processes has been added: the synaptic weight process $(W(t))$ is not constant.

### 1.5 Extensions

The stochastic model with plastic connections presented in this paper may also be used in other contexts than neuroscience. Auto-exciting processes, Hawkes process, used in finance Embrechts et al. [6], genomics Gusto and Schbath [10] and Reynaud-Bouret and Schbath [25], sociology Crane and Sornette [5] suppose, in general, that the influence of each point process on the others is constant over time. For example in Crane and Sornette [5], a Hawkes process is defined to describe the cascade of influences that exist in a social network, taking the example of Youtube videos views. The coefficients that model interactions between different individuals are constant. One could extend this model by taking into account the fact that individuals who watch repeatedly videos at the same time, may develop a stronger interaction. It should be therefore possible to extend these classes of models by adding a dependence of the connections on the past activity of the Hawkes process as for our models. Using classical results on stationary distributions of Hawkes processes with fixed connectivity, a slow-fast analysis similar to the one developed here should then be possible for these models.

### 1.6 Organization of the Paper

In Sect. 2, the main processes and definitions are introduced as well as assumptions to prove an averaging principle. The scaling is presented in Sect. 3 and the averaging principle in Sect. 4. In this section the general strategy for the proof of the main theorem is detailed. Section 5 investigates monotonicity properties and a coupling result, crucial in the proof of tightness, is proved. Section 6 is devoted to the proof of tightness results when the process $(W_\varepsilon(t))$ is assumed to be bounded. Finally, the proof of the main theorem is completed in Sect. 7. In Sect. B of Appendix, several useful tightness results are proved for interacting shot-noise processes. The ergodicity properties of fast processes are analyzed in Sect. C of
Appendix. Sect. D of the Appendix discusses averaging principles for related discrete models of synaptic plasticity.

2 A Stochastic Model for Plasticity

We define the stochastic model associated to Markovian plasticity kernels introduced in Robert and Vignoud [28]. The probabilistic setting of these models along with formal definitions are detailed in the following section.

2.1 Definitions and Notations

The space of Borelian subsets of a topological space $H$, is denoted as $\mathcal{B}(H)$. Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space. We assume that two independent Poisson processes, $\mathcal{P}_1$ and $\mathcal{P}_2$ on $\mathbb{R}_+^2$, with intensity $dx \times dy$ are defined on $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. See Kingman [20] for example. For $\mathcal{P} \in \mathcal{P}_1, \mathcal{P}_2$ and $A, B \in \mathcal{B}(\mathbb{R}_+)$ and a Borelian function $f$ on $\mathbb{R}_+$,

$$\mathcal{P} (A \times B) \stackrel{\text{def}}{=} \int_{A \times B} \mathcal{P}(dx, dy), \quad \int_{\mathbb{R}^+} f(y) \mathcal{P}(A, dy) \stackrel{\text{def}}{=} \int_{A \times \mathbb{R}^+} f(y) \mathcal{P}(dx, dy).$$

For $t \geq 0$, the $\sigma$-field $\mathcal{F}_t$ of the filtration $(\mathcal{F}_t)_{t \geq 0}$ is assumed to contain all events before time $t$ for both point processes, i.e.

$$\sigma (\mathcal{P}_1 (A \times (s, t]), \mathcal{P}_2 (A \times (s, t]), A \in \mathcal{B}(\mathbb{R}_+), s \leq t) \subset \mathcal{F}_t. \quad (5)$$

A stochastic process $(H(t))$ is adapted if, for all $t \geq 0$, $H(t)$ is $\mathcal{F}_t$-measurable. It is a càdlàg-process if, almost surely, it is right continuous and has a left limit at every point $t > 0$, $H(t−)$ denotes the left limit of $(H(t))$ at $t$. The Skorohod space of càdlàgfunctions from $[0, T]$ to $S$ is denoted as $\mathcal{D}([0, T], S)$. See Billingsley [3] and Ethier and Kurtz [7]. The mention of adapted stochastic processes, or of martingale, will be implicitly associated to the filtration $(\mathcal{F}_t)_{t \geq 0}$.

The set of real continuous bounded functions on the metric space $S \subset \mathbb{R}^d$ is denoted by $C_b(S)$. $C^k_b(S) \subset C_b(S)$ is the set of bounded, $k$-differentiable functions on $S$ with respect to each coordinate, with all derivatives bounded and continuous. The multi-dimensional extensions to $S$ are denoted by $C^k_b(S, S)$.

INHOMOGENEOUS POISSON PROCESS FOR THE OUTPUT NODE. We introduce an important point process $\mathcal{N}_{\phi, H}$, that represents the jumps of a node whose activity process is $(H(t))$, with activation function $\phi$. $\phi$ is a non-negative càdlàgfunction on $\mathbb{R}_+$, it is defined by

$$\int_{\mathbb{R}_+} f(u) \mathcal{N}_{\phi, H}(du) \stackrel{\text{def}}{=} \int_{\mathbb{R}_+} f(u) \mathcal{P}_2(\{(0, \phi(H(u−)))\}, du) \quad (6)$$

for any Borelian function $f$ on $\mathbb{R}_+.$

2.2 The Plasticity Process

In the rest of the paper, we will consider a more generic connected system with plasticity, inspired from the neuronal example given in the Introduction. An input node will be represented by an homogeneous Poisson process with intensity $\lambda$, taking up the role of the
pre-synaptic neuron. This input node will interact with an output node, whose activity \( X \) (i.e. membrane potential) integrates over time the jumps of the input node, with an amplitude \( W \) and an exponential decay taken with time constant 1. The output node jumps with an inhomogeneous rate \( \beta(X) \) that depends on the output node activity \( X \). The connection intensity \( W \) is plastic and depends on previous interactions between both nodes, through a Markovian multi-dimensional variable \( Z \), in the same way that the synaptic weight undergoes synaptic plasticity in the neuronal model.

**Definition 1** (Time Evolution) The càdlàg process

\[
(U(t)) = (X(t), Z(t), \Omega_p(t), \Omega_d(t), W(t)) \in \mathbb{R} \times \mathbb{R}_+^\ell \times \mathbb{R}_+^2 \times K_W,
\]

is solution of the following Stochastic Differential Equations (SDE), starting from some initial state \( U(0) = U_0 = (x_0, z_0, \omega_{0,p}, \omega_{0,d}, w_0) \).

\[
\begin{align*}
\frac{dX(t)}{dt} &= -X(t)dt + W(t)\mathcal{N}_z(dt) - g(X(t-))\mathcal{N}_{\beta,X}(dt), \\
\frac{dZ(t)}{dt} &= -(\gamma \circ Z(t) + \kappa_0)dt + k_1(Z(t-))\mathcal{N}_z(dt) + k_2(Z(t-))\mathcal{N}_{\beta,X}(dt), \\
\frac{d\Omega_a(t)}{dt} &= -a\Omega_a(t)dt + n_{a,0}(Z(t))dt \\
&\quad + n_{a,1}(Z(t-))\mathcal{N}_z(dt) + n_{a,2}(Z(t-))\mathcal{N}_{\beta,X}(dt), \\
\frac{dW(t)}{dt} &= M(\Omega_p(t), \Omega_d(t), W(t))dt,
\end{align*}
\]

with the notation \( a \circ b = (a_k b_k) \) for the Hadamard product, for \( a = (a_k), b = (b_k) \in \mathbb{R}_+^\ell \).

Recall that, see Sect. 1.3, \( K_W \) is an interval of \( \mathbb{R} \) which contains the range of values for the connection intensity.

We now state the assumptions used for the proof of Theorem 4.

### 2.2.1 Inputs Jumps

The jumps of the input node are given by a Poisson process with rate \( \lambda > 0 \),

\[
\mathcal{N}_z(dt) \overset{\text{def}}{=} \mathcal{P}_1((0, \lambda], dt),
\]

where \( \mathcal{P}_1 \) is the Poisson point process introduced in Sect. 2.1.

### 2.2.2 Output Jumps

When the output node activity is \( x \), a jump of the output node occurs at rate \( \beta(x) \) and leads to a decrease of output activity \( x - g(x) \).

- It is assumed that \( \beta \) is a non-negative, continuous function on \( \mathbb{R} \) and that \( \beta(x) = 0 \) for \( x \leq -c_\beta \leq 0 \). Additionally, there exists a constant \( C_\beta \geq 0 \) such that

\[
\beta(x) \leq C_\beta(1 + |x|), \quad \forall x \in \mathbb{R}.
\]

- The function \( g \) is continuous on \( \mathbb{R} \) and \( 0 \leq g(x) \leq \max(c_g, x) \) holds for all \( x \in \mathbb{R} \), for some \( c_g \geq 0 \).

The jumps of the output node are represented by the point process \( \mathcal{N}_{\beta,X} \). Recall that

\[
\mathcal{N}_{\beta,X}(dt) = \mathcal{P}_2((0, \beta(X(t-)]), dt).
\]
2.2.3 The Process \((Z(t))\)

The process \((Z(t))\) is a multi-dimensional process, with values in \(\mathbb{R}^\ell_+\), it is driven by the general spiking activity of the system, and therefore, depends only on the point processes \(N_\lambda\) and \(N_{\beta,X}\). For some models it describes the time evolution of chemical components within the synapse. \((Z(t))\) is a càdlàg function with values in \(\mathbb{R}^\ell_+\), solution of the stochastic differential equation

\[
\text{d}Z(t) = (-\gamma \odot Z(t) + k_0)\text{d}t + k_1(Z(t-))N_\lambda(\text{d}t) + k_2(Z(t-))N_{\beta,X}(\text{d}t),
\]

where, as before, \(\odot\) is for the Hadamard product, \(k_0 \in \mathbb{R}^\ell_+\) is a constant and \(k_1\) and \(k_2\) are measurable functions from \(\mathbb{R}^\ell_+\) to \(\mathbb{R}^\ell\). Furthermore, the \((k_i)\) are chosen such that \((z(t))\) has values in \(\mathbb{R}^\ell_+\) whenever \(z(0) \in \mathbb{R}^\ell_+\).

It is assumed that

(a) All coordinates of the vector \(\gamma\) are positive;
(b) The non-negative functions \(k_i, i=[0,1,2]\), are \(C^1\) \((\mathbb{R}^\ell_+, \mathbb{R}^\ell_+)\) and bounded by \(C_k \geq 0\).

2.2.4 The Process \((\Omega_p(t), \Omega_d(t))\)

These variables, in \(\mathbb{R}^p_+\) encode, with an exponential decay, the total memory of instantaneous plasticity processes represented by the process \((Z(t))\). The process \((\Omega_p(t))\) is driving potentiation of the connection, i.e. the derivative of the connection intensity is an increasing function of this variable. In an analogous way, \((\Omega_d(t))\) is associated to depression, i.e. the derivative of the connection intensity is a decreasing function of this variable. The system of equations for \((\Omega_p(t), \Omega_d(t))\) is a set of two one-dimensional SDEs, for \(a \in \{p, d\}\),

\[
\text{d}\Omega_a(t) = -\alpha\Omega_a(t)\text{d}t + n_{a,0}(Z(t))\text{d}t \\
+ n_{a,1}(Z(t-))N_\lambda(\text{d}t) + n_{a,2}(Z(t-))N_{\beta,X}(\text{d}t).
\]

We suppose that there exists a constant \(C_n\) such that, for \(j \in \{0,1,2\}, a \in \{p, d\}\), \(n_{a,j}\) verifies,

\[
n_{a,j}(z) \leq C_n(1 + \|z\|),
\]

where, for \(z \in \mathbb{R}^\ell_+\), \(\|z\| = z_1 + \cdots + z_\ell\).

For any \(w \in K_W\) and \(a \in \{p, d\}\), the discontinuity points of

\[
(x,z) \mapsto (n_{a,0}(z), n_{a,1}(z), \beta(x)n_{a,2}(z))
\]

are negligible for the invariant probability distribution \(\Pi_w\) of \((X(t), Z(t))\) when \((W(t))\) is constant equal to \(w\). See Sect. C.

When \((W(t))\) is constant, the process \((X(t), Z(t))\) can be seen as generalized shot-noise processes, see Sect. B. It is well-known that the invariant distribution of the classical, one-dimensional, shot-noise process is absolutely continuous w.r.t Lebesgue’s measure. See examples of Sects. 5 and 6 of Robert and Vignoud [28] and the reference Beznea et al. [2] for criteria in this domain.

2.2.5 Dynamics of the Connection Intensity

The functional \(M\) drives the dynamics of the connection intensity, the corresponding equation is given by Relation (7). In particular, for any \(w \in K_W\) and any càdlàg piecewise-continuous
functions $h_1$ and $h_2$ on $\mathbb{R}_+$, the ODE

$$\frac{dw}{dt}(t)=M(h_1(t), h_2(t), w(t)) \text{ with } w(0)=w,$$  \hspace{1cm} (12)

for all points of continuity of $h_1$ and $h_2$, has a unique continuous solution denoted by $(S[h_1, h_2](w,t))$ in $K_w$. We assume that $M$ can be decomposed as $M(\omega_p, \omega_d, w)=M_p(\omega_p, w)-M_d(\omega_d, w)-\mu w$, where $M_a(\omega_a, w)$ is non-negative continuous function, non-decreasing on the first coordinate for a fixed $w\in K_w$, and,

$$M_a(\omega_a, w) \leq C M(1+\omega_a),$$

for all $w\in K_w$, for $a\in\{p, d\}$.

### 2.3 Discrete Models of Plasticity

A model of plasticity with discrete state space has been introduced in Robert and Vignonud [28]. The proof of the associated averaging principles for the continuous case can be adapted to such systems. Relevant parts of the proof are briefly presented in Sect. D of the Appendix.

### 3 The Scaled Process

The SDEs of Definition 1 are difficult to study without any additional hypothesis. Existence and uniqueness of solutions to this system are guaranteed by Proposition 1 of Robert and Vignonud [28]. It can be seen as an intricate fixed point equation for the processes $(X(t), W(t))$ involving functionals of these processes like $N_{\beta, X}$ defined by Relation (6).

As explained in Sect. 4 of [28], $(X(t), Z(t))$ are associated to fast dynamics at the cellular level while the process $(W(t))$ evolves on a much longer timescale. For this reason, a scaling parameter $\varepsilon>0$ is introduced so that $(X(t), Z(t))$ evolves on the timescale $t\mapsto t/\varepsilon$. More precisely,

- **Fast Processes**: $(X(t))$ and $(Z(t))$.
  
  The point processes associated to input and output jumps driving the time evolution of $(X(t))$ and $(Z(t))$ are sped-up by a factor $1/\varepsilon$: $N_{\lambda} \rightarrow N_{\lambda/\varepsilon}$ and $N_{\beta, X} \rightarrow N_{\beta/\varepsilon, X}$. The deterministic part of the evolution is changed accordingly $dt \rightarrow dt/\varepsilon$.

- **Slow Processes**: $(W(t))$ and $(\Omega_p(t), \Omega_d(t))$.
  
  Update of $(\Omega_p(t))$ and $(\Omega_d(t))$ due to fast jump processes have a small amplitude, $N_{\lambda} \rightarrow \varepsilon N_{\lambda/\varepsilon}$ and $N_{\beta, X} \rightarrow \varepsilon N_{\beta/\varepsilon, X}$.

Formally, we define the scaled process $(U_{\varepsilon}(t))=(X_{\varepsilon}(t), Z_{\varepsilon}(t), \Omega_{\varepsilon, p}(t), \Omega_{\varepsilon, d}(t), W_{\varepsilon}(t))$, the evolution equations of Definition 1 become

$$dX_{\varepsilon}(t) = -X_{\varepsilon}(t)dt/\varepsilon + W_{\varepsilon}(t)N_{\lambda/\varepsilon}(dt) - g(X_{\varepsilon}(t-))N_{\beta/\varepsilon, X_{\varepsilon}}(dt),$$  \hspace{1cm} (13)

$$dZ_{\varepsilon}(t) = (-\gamma \circ Z_{\varepsilon}(t) + k_0)dt/\varepsilon + k_1(Z_{\varepsilon}(t-))N_{\lambda/\varepsilon}(dt) + k_2(Z_{\varepsilon}(t-))N_{\beta/\varepsilon, X_{\varepsilon}}(dt),$$  \hspace{1cm} (14)

$$d\Omega_{\varepsilon, a}(t) = -a \Omega_{\varepsilon, a}(t)dt + n_{a,0}(Z_{\varepsilon}(t))dt + n_{a,1}(Z_{\varepsilon}(t-))\varepsilon N_{\lambda/\varepsilon}(dt) + n_{a,2}(Z_{\varepsilon}(t-))\varepsilon N_{\beta/\varepsilon, X_{\varepsilon}}(dt), \quad a\in\{p, d\}$$  \hspace{1cm} (15)

$$dW_{\varepsilon}(t) = M(\Omega_{\varepsilon, p}(t), \Omega_{\varepsilon, d}(t), W(t))dt.$$  \hspace{1cm} (16)
For simplicity, the initial condition of \((U_\varepsilon(t))\) is assumed to be constant,

\[
U_\varepsilon(0) = U_0 = (x_0, z_0, \omega_{p,0}, \omega_{d,0}, w_0).
\]

Some simplifications of this (heavy) mathematical framework can be expected when \(\varepsilon\) goes to 0. We first introduce the notion of fast variables which correspond to the processes \((X(t), Z(t))\) with the connection intensity process \((W(t))\) taken as constant.

### 3.1 Fast Processes

**Definition 2** For \(w \in K_w\), \((X^w(t), Z^w(t))\) is the Markov process in \(\mathbb{R} \times \mathbb{R}_+^\ell\) defined by the SDEs

\[
\begin{align*}
\frac{dX^w(t)}{dt} &= -X^w(t) + w \mathcal{N}(\lambda_x, N_{x, w}(dt)), \\
\frac{dZ^w(t)}{dt} &= -\gamma \circ Z^w(t) + k_0 dt + k_1 (Z^w(t-)) N_{x, w}(dt) + k_2 (Z^w(t-)) \mathcal{N}_{x, w}(dt).
\end{align*}
\]

Let \(f \in C^1_b(\mathbb{R} \times \mathbb{R}_+^\ell)\), then, with Equations (13) and (14), we have that

\[
(M_{f, \varepsilon}^F(t)) \overset{\text{def}}{=} \left( f(X_\varepsilon(t), Z_\varepsilon(t)) - f(x_0, z_0) - \frac{1}{\varepsilon} \int_0^t B_{W_\varepsilon(s)}^F(f)(X_\varepsilon(s), Z_\varepsilon(s)) ds \right)
\]

is a local martingale, where, for \(v = (x, z) \in \mathbb{R} \times \mathbb{R}_+^\ell\) and,

\[
B_w^F(f)(v) \overset{\text{def}}{=} \frac{\partial f}{\partial x}(x, z) + \left( -\gamma \circ z + k_0, \frac{\partial f}{\partial z}(x, z) \right) + \lambda \left( f(x + w, z + k_1(z)) - f(u) \right) + \beta(x) \left( f(x - g(x), z + k_2(z)) - f(u) \right),
\]

with

\[
\frac{\partial f}{\partial z}(x, z) = \left( \frac{\partial f}{\partial z_i}(x, z), i \in [1, \ldots, \ell] \right)
\]

and \((z, z')\) is the usual scalar product of \(z\) and \(z' \in \mathbb{R}_+^\ell\). The operator \(B_w^F\) is called the *infinitesimal generator* of the fast processes \((X^w(t), Z^w(t))\).

In Proposition 25 of Appendix C, we establish that, under the conditions of Sects. 2.2.2 and 2.2.3, the fast process \((X^w(t), Z^w(t))\) has a unique invariant distribution \(\Pi_w\).

### 3.2 Functionals of the Occupation Measure

We start with a rough, non-rigorous, picture of results that are usually established for slow-fast systems.

**Definition 3** (Occupation Measure) The occupation measure is the non-negative measure \(v_\varepsilon\) on \([0, T] \times \mathbb{R} \times \mathbb{R}_+^\ell\) such that

\[
v_\varepsilon(G) \overset{\text{def}}{=} \int_{[0,T] \times \mathbb{R} \times \mathbb{R}_+^\ell} G(s, x, z) v_\varepsilon(ds, dx, dz) \overset{\text{def}}{=} \int_{[0,T]} G(s, X_\varepsilon(s), Z_\varepsilon(s)) ds.
\]

for any non-negative Borelian function \(G\) on \([0, T] \times \mathbb{R} \times \mathbb{R}_+^\ell\).
The integration of Relation (15) gives the identity, for \( a = \{p, d\} \)
\[
\Omega_{\varepsilon, a}(t) = \omega_{0,a} - \alpha \int_0^t \Omega_{\varepsilon, a}(s) ds + \int_0^t n_{a,0}(Z_\varepsilon(s)) ds \\
+ \int_0^t n_{a,1}(Z_\varepsilon(s)) \varepsilon N_{\lambda, \varepsilon}(ds) + \int_0^t n_{a,2}(Z_\varepsilon(s)) \varepsilon N_{\beta, \varepsilon, X_\varepsilon}(ds).
\]

An averaging principle is said to hold when the convergence in distribution
\[
\lim_{\varepsilon \to 0} \left( \int_0^t G(X_\varepsilon(s), Z_\varepsilon(s)) ds \right) = \lim_{\varepsilon \to 0} \left( \int_{\mathbb{R} \times \mathbb{R}_+^L} G(x, z) \nu_\varepsilon(ds, dx, dz) \right) = \left( \int_0^t \int_{\mathbb{R} \times \mathbb{R}_+^L} G(x, z) \Pi_{w(s)}(dx, dz) ds \right),
\]
holds for a sufficiently rich class of Borelian functions \( G \). Usually, it is enough to prove the
weak convergence of the occupation measure for bounded Borelian functions \( G \).

In our case, there are important examples where \( G \) has a linear growth with respect to the
coordinates \( x \) or \( z = (z_j) \). The all-to-all models of pair-based rules for example, which are
widely used in computational neuroscience lead to unbounded functionals of the occupation
measure. See Sect. 3.3 of Robert and Vignoud [28]. Additionally, convergence results in
distribution of the jump processes such as,
\[
\lim_{\varepsilon \to 0} \left( \int_0^t G(X_\varepsilon(s), Z_\varepsilon(s)) \varepsilon N_{\lambda, \varepsilon}(ds) \right) = \left( \int_0^t \int_{\mathbb{R} \times \mathbb{R}_+^L} G(x, z) \Pi_{w(s)}(dx, dz) ds \right),
\]
and
\[
\lim_{\varepsilon \to 0} \left( \int_0^t G(X_\varepsilon(s), Z_\varepsilon(s)) \varepsilon N_{\beta, \varepsilon, X_\varepsilon}(ds) \right) = \left( \int_0^t \int_{\mathbb{R} \times \mathbb{R}_+^L} G(x, z) \beta(x) \Pi_{w(s)}(dx, dz) ds \right)
\]
are also required. They are not straightforward consequences of Relation (20) as it is usually
the case for bounded \( G \). See Kurtz [21] for example. These technical difficulties have to be
overcome to establish the tightness of the processes \( (\Omega_{\varepsilon}(t)) \), and consequently of \( (W_\varepsilon(t)) \).
As a result, additional limit results have to be established at this point, see Sect. 6. Furthermore,
as \( \varepsilon \) goes to 0, the process \( (\Omega_{\varepsilon, p}(t), \Omega_{\varepsilon, d}(t), W_\varepsilon(t)) \) should converge to a process
\( (\omega_p(t), \omega_d(t), w(t)) \) satisfying the relation,

\[
\begin{align*}
\omega_d(t) &= \omega_{d,0} - \alpha \int_0^t \omega(s) ds + \int_0^t \int_{\mathbb{R}_+^L} n_{a,0}(z) \Pi_{w(s)}(\mathbb{R}_+, dz) ds \\
+ \lambda \int_0^t \int_{\mathbb{R}_+^L} n_{a,1}(z) \Pi_{w(s)}(\mathbb{R}_+, dz) ds + \int_0^t \int_{\mathbb{R}_+^L} \beta(x) n_{a,2}(z) \Pi_{w(s)}(dx, dz) ds \\
\frac{dw}{dt}(t) &= M(\omega_p(t), \omega_d(t), w(t))
\end{align*}
\]

4 Averaging Principles Results

We fix \( T > 0 \), throughout the paper the convergence in distribution of processes is considered
on the bounded interval \( [0, T] \).
4.1 Main Result

We start by reviewing the assumptions detailed in Sect. 2.1 on the different parameters of the stochastic model.

Assumptions

(a) It is assumed that $\beta$ is a non-negative, continuous function on $\mathbb{R}$ and that $\beta(x)=0$ for $x \leq -c_\beta \leq 0$. Additionally, there exist a constant $C_\beta \geq 0$ such that

$$\beta(x) \leq C_\beta (1+|x|), \quad \forall x \in \mathbb{R};$$

(b) $g$ is continuous function on $\mathbb{R}$ and $0 \leq g(x) \leq \max(c_g, x)$ holds for all $x \in \mathbb{R}$, for some $c_g \geq 0$;

(c) All coordinates of the vector $\gamma$ are positive;

(d) There exists a constant $C_k \geq 0$ such that $0 \leq k_0 \leq C_k$ and functions $k_i, \ i = 1, 2$, in $C^1_b(\mathbb{R}_+^4, \mathbb{R}_+^3)$, are upper-bounded by $C_k \geq 0$;

(e) There exists a constant $C_n$ such that, for $j \in \{0, 1, 2\}$, $a \in \{p, d\}$, $n_{a,j}$ verifies,

$$n_{a,j}(z) \leq C_n (1+\|z\|),$$

where, for $z \in \mathbb{R}_+^\ell$, $\|z\|=z_1+\cdots+z_\ell$. Moreover, for any $w \in K_w$, the discontinuity points of

$$(x, z) \mapsto (n_{a,0}(z), n_{a,1}(z), \beta(x)n_{a,2}(z))$$

for $a \in \{p, d\}$, are negligible for the probability distribution $\Pi_w$ of Sect. C.

(f) $M$ can be decomposed as,

$$M(\omega_p, \omega_d, w)=M_p(\omega_p, w) - M_d(\omega_d, w) - \mu w,$$

where $M_a(\omega_a, w)$ is non-negative continuous function, non-decreasing on the first coordinate for a fixed $w \in K_w$, and,

$$M_a(\omega_a, w) \leq C_M (1+\omega_a),$$

for all $w \in K_w$, for $a \in \{p, d\}$.

The main result of the paper is the following theorem.

Theorem 4 (Asymptotic Time Evolution of Plasticity). Under the conditions of Sect. 2.1 and for initial conditions satisfying Relation (17), there exists $S_0 \in (0, +\infty]$, such that the family of processes $(\Omega_{\varepsilon,p}(t), \Omega_{\varepsilon,d}(t), W_\varepsilon(t), t<S_0)$, $\varepsilon \in (0, 1)$, of the system of Sect. 3, is tight for the convergence in distribution. As $\varepsilon$ goes to 0, any limiting point $(\omega_p(t), \omega_d(t), w(t), t<S_0)$, satisfies the ODEs, for $a \in \{p, d\}$,

$$\begin{cases}
\frac{d\omega_a(t)}{dt} = -\alpha \omega_a(t) + \int_{\mathbb{R} \times \mathbb{R}_+^\ell} (n_{a,0}(z)+\lambda n_{a,1}(z)+\beta(x)n_{a,2}(z)) \Pi_w(t)(dx, dz), \\
\frac{dw(t)}{dt} = M(\omega_p(t), \omega_d(t), w(t)),
\end{cases} \tag{21}$$

where, for $w \in K_w$, $\Pi_w$ is the unique invariant distribution $\Pi_w$ on $\mathbb{R} \times \mathbb{R}_+^\ell$ of the Markovian operator $B_w^F$. If $K_w$ is bounded, then $S_0=+\infty$ almost-surely.
Convergence in Distribution  As already mentioned, most of the efforts in this paper are devoted to the proof of the tightness property of \((\Omega_\varepsilon, p(t), \Omega_\varepsilon, d(t), W_\varepsilon(t))\). We note that our result identifies the limiting points, but it does not state any weak convergence results for the scaled processes. Regularity properties are actually required on \((\Pi_w)\) to have such results. For example, it would be sufficient to have that the mapping, 
\[ \Psi_a : w \mapsto \int_{\mathbb{R} \times \mathbb{R}_+^\ell} \left( n_{a,0}(z) + \lambda n_{a,1}(z) + \beta(x) n_{a,2}(z) \right) \Pi_w(dx, dz), \]
locally Lipschitz for \(w\), for \(a \in \{p, d\}\), so that Relation (21) has a unique solution.

Due to the generality of our model, we did not try to state a set of conditions that can ensure the desired regularity properties of the corresponding \(\Pi_w\). Uniqueness results are obtained in Sects. 5 and 6 of Robert and Vignoud [28] for several important cases. The same properties for the simple model are worked out in Sect. 8. However at this stage, a case by case analysis seems mandatory.

A Blow-up phenomenon  As it can be seen, when \(S_0 < +\infty\) the convergence is only proved on a bounded time interval. In the proof, the variable \(S_0\) results from the domain definition of the solution to a deterministic differential equation. This is not an artifact of our methods, see Proposition 22 in Sect. 8 for an example.

4.2 Steps of the Proof

The proof of the theorem is organized as follows. See also Fig. 1 of Appendix.

(a) Section 5. A stochastic upper-bound \(\bar{U}\) of the original process is introduced and a coupling argument is used to control \((W_\varepsilon(t))\). This is an important ingredient in the proof of tightness results for \((\Omega_\varepsilon, p(t), \Omega_\varepsilon, d(t), W_\varepsilon(t))\).

(b) Section 6. Under the temporary assumption that the process \((W(t))\) is bounded by \(K\), we establish tightness results for the truncated process \(U^K\), when \(\varepsilon\) goes to 0, of variables associated to fast processes \((X^K_\varepsilon(t), Z^K_\varepsilon(t))\) of the type
\[ \left( \int_0^T G \left( s, X^K_\varepsilon(s), Z^K_\varepsilon(s) \right) ds \right) \]
where \(G\) is a continuous Borelian function with a linear growth with respect to the coordinates \(x \in \mathbb{R}\) and \(z \in \mathbb{R}_+^\ell\). An averaging principle is shown for this truncated process.

(c) In Theorem 18 of Sect. 7, using monotonicity arguments, we are able to obtain a deterministic, analytical bound, uniform in \(K\), for the limiting points of the truncated process. From there, we prove an averaging principle for the dominating process \(\bar{U}\) (without truncation) in Proposition 19 where the explicit form of the ODE verified by the limiting points is known. As a direct consequence, we are able to prove that this limit is unique and that the scaled dominating process converges to the solution. Using the fact that the process \(W(t)\) is bounded by \(\bar{W}(t)\) and the previous convergence, we establish the desired results for the process \(U_\varepsilon(t)\) of Theorem 4.

4.3 Technical Results on Shot-Noise Processes

The processes \((X(t))\) and \((Z(t))\) are closely related to shot-noise processes and their generalizations. See for example Schottky [30], Rice [26] and Gilbert and Pollak [9] for an
introduction. We give a quick overview of their use in our proofs. In Appendix B, the results below and several technical lemmas for these processes are detailed and proved.

The following lemma gives an elementary representation result for general shot-noise process associated to a positive Radon measure. See Lemma 1 of Robert and Vignoud [28].

**Lemma 5** If \( \mu \) is a positive Radon measure on \( \mathbb{R}_+ \) and \( \gamma > 0 \), the unique càdlàg solution of the ODE

\[
dZ(t) = -\gamma Z(t) + \mu(dt),
\]
with initial point \( z_0 \in \mathbb{R}_+ \) is given by
\[
Z(t) = z_0 e^{-\gamma t} + \int_{(0,t]} e^{-\gamma(t-s)} \mu(ds).
\]
(22)

In view of SDEs (2) it is natural to introduce a scaled version of these processes.

**Definition 6** (Scaled Shot-Noise Process). For \( \varepsilon > 0 \), we define the shot-noise process \((S^\varepsilon_x(t))\), solution of the SDE
\[
dS^\varepsilon_x(t) = -S^\varepsilon_x(t)dt/\varepsilon + N^\varepsilon_x(\lambda/\varepsilon)(dt),
\]
where the initial point is \( x \geq 0 \).

**Proposition 7** For \( \xi \in \mathbb{R} \) and \( x \geq 0 \), the convergence in distribution of the processes
\[
\lim_{\varepsilon \downarrow 0} \left( \int_0^t e^{\xi S^\varepsilon_x(u)} du \right) = \left( E\left[e^{\xi S(\infty)}\right]\right)t
\]
holds, and
\[
\sup_{0 \leq \varepsilon \leq 1} E\left[e^{\xi S^\varepsilon_x(t)}\right] < +\infty.
\]
(23)

**Proof** See Sect. B.1 of Appendix.

We now introduce another shot-noise process \((R^\varepsilon_x(t))\) associated to the point process \( N^\varepsilon_xI/\varepsilon \) defined by Relation (6) where \( I(x) = x \), \( x \in \mathbb{R}_+ \). It is in fact a shot-noise process whose intensity function is \((S^\varepsilon_x(t))\),
\[
dR^\varepsilon_x(t) = -\gamma R^\varepsilon_x(t)dt/\varepsilon + N^\varepsilon_x(\lambda/\varepsilon)(dt),
\]
with the initial condition \( R^\varepsilon_x(0)=0 \).

It turns out that tightness properties of three families of linear functionals of such processes
\[
\left( \int_0^t R^\varepsilon_x(s) ds \right), \left( \int_0^t R^\varepsilon_x(s)eN^\varepsilon_x(\lambda/\varepsilon)(ds) \right), \left( \int_0^t R^\varepsilon_x(s)eN^\varepsilon_x(I/\varepsilon,S^\varepsilon_x)(ds) \right),
\]
are central to establish Theorem 4. The motivation comes from the three terms in the expression of \((\Omega_{\varepsilon,\alpha}(t))\), \( a \in \{p, d\} \) of Relation (15) and the fact that, with the condition of Relation (11), for \( j \in \{0, 1, 2\} \), \( n_j(z) \leq C_n^0 + C_n^1z \) for \( z \in \mathbb{R}_+ \).

The necessary results are stated in Proposition 8 which is used in Sect. 7.

**Proposition 8** For \( H \in \{S^\varepsilon_x, R^\varepsilon_x\} \), the families of processes
\[
\left( \int_0^t H^\varepsilon_x(u) du \right), \left( \int_0^t H^\varepsilon_x(u)^2 du \right) \text{ and } \left( \int_0^t R^\varepsilon_x(u)S^\varepsilon_x(u) du \right), \varepsilon \in (0, 1),
\]
are tight for the convergence in distribution.

**Proof** We first prove the tightness of the second family of processes. With Cauchy-Schwartz’ Inequality, we have
\[
\int_s^t H^\varepsilon_x(u)^2 du \leq \sqrt{t-s} \sqrt{\int_s^t H^\varepsilon_x(u)^4 du}.
\]
Moreover, Relation (23) gives an estimate of $S_\varepsilon$ moments and Proposition 24 of Appendix states that

$$\sup_{\varepsilon \in (0,1), t \geq 0} E[R_\varepsilon(t)^4] < +\infty$$

Gathering up previous estimates, we show that there exists a constant $C$ independent of $\varepsilon$ and $s, t$ such that

$$E\left[\left(\int_s^t H_\varepsilon(u)^2 \, du\right)^2\right] \leq (t-s)^2.$$

Kolmogorov-Čentsov’s Criterion, see Theorem 2.8 and Problem 4.11 of Karatzas and Shreve [17], implies that the family of variables

$$\left(\int_0^t H_\varepsilon(u)^2 \, du\right)$$

is tight.

By using repeatedly Cauchy-Schwartz’ Inequality, for $0 \leq s \leq t$, we have

$$E\left[\left(\int_s^t H_\varepsilon(u) \, du\right)^4\right] \leq (t-s)^2 E\left[\left(\int_s^t H_\varepsilon(u)^2 \, du\right)^2\right] \leq C(t-s)^4$$

and

$$E\left[\left(\int_s^t R_\varepsilon(u)S_\varepsilon(u) \, du\right)^2\right] \leq \sqrt{E\left[\left(\int_s^t R_\varepsilon(u)^2 \, du\right)^2\right]} \sqrt{E\left[\left(\int_s^t S_\varepsilon(u)^2 \, du\right)^2\right]} \leq C(t-s)^2$$

Kolmogorov-Čentsov’s Criterion can then also be applied to the two other families of processes of the proposition. The proposition is proved.

\[\square\]

5 A Coupling Property

In this section a process

$$(\bar{U}(t))=(\bar{X}(t), \bar{Z}(t), \bar{\Omega}(t), \bar{W}(t))$$

in $\mathcal{D}([0, T], \mathbb{R}_+^4)$ is introduced. It has similarities with the process $(U(t))$ of Definition 1 but fewer coordinates and simpler parameters. More importantly, all its coordinates are non-negative. We first prove, via a coupling, that the sample paths of the processes $(U(t))$ and $(\bar{U}(t))$ can be compared in a sense to be made precise. Secondly, we derive several technical estimates for $(\bar{U}(t))$ which are important to prove the tightness of the scaled processes $((\bar{\Omega}_\varepsilon(t), \bar{W}_\varepsilon(t)))$ defined in Sect. 3.
The process \(\bar{U}(t)\) is the solution of the SDEs
\[
\begin{align*}
\frac{d\bar{X}(t)}{dt} &= -\bar{X}(t)dt + \bar{W}(t)N_\lambda(dt), \\
\frac{d\bar{Z}(t)}{dt} &= -\gamma\bar{Z}(t) + C_k \delta + C_kN_\lambda(dt) + C_kN_{\bar{\beta},\bar{X}}(dt), \\
\frac{d\bar{\Omega}(t)}{dt} &= -\alpha \bar{\Omega}(t)dt + C_n \left(1 + \ell\bar{Z}(t)\right) dt + C_n \left(1 + \ell\bar{Z}(t)\right) \lambda(dt) \\
+ C_n \left(1 + \ell\bar{Z}(t)\right) N_{\bar{\beta},\bar{X}}(ds) \\
\frac{d\bar{W}(t)}{dt} &= C_M \left(1 + \bar{\Omega}(t)\right) dt,
\end{align*}
\tag{25}
\]
with \(\bar{\beta}(x) = \bar{\beta}_1 + x\) and with initial condition \(\bar{U}(0)\) given by
\[
(\bar{x}_0, \bar{z}_0, \bar{\omega}_0, \bar{W}_0) = (\max(x_0, 0), \max_{i \in \{1, \ldots, \ell\}} \z_0, \max_{a \in [a, p]} \{\omega_{0,a}, |w_0|\}).
\]

\(C_\beta, C_n, C_k\) and \(C_M\) are non-negative constants associated to the conditions of Sect. 2, and \(\gamma = \min(\gamma_i : i = 1, \ldots, \ell)\).

Throughout this section, for \(t \geq 0\), if \((U(t))\) is a solution of Relations (7), the inequality \(U(t) \leq \bar{U}(t)\) will stand for the four relations, \(X(t) \leq \bar{X}(t)\),
\[
\max_{i \in \{1, \ldots, \ell\}} \{Z_i(t)\} \leq \bar{Z}(t), \max_{a \in [p, d]} (\Omega_a(t)) \leq \bar{\Omega}(t), \text{ and } |W(t)| \leq \bar{W}(t).
\]

### 5.1 A Coupling Property

We start by proving a monotonicity property of the behavior of both systems “between” jumps.

Define \(u(t) = (x(t), z(t), \omega_p(t), \omega_d(t), w(t))\) that follows,
\[
\begin{align*}
\frac{dx(t)}{dt} &= -x(t)dt, \\
\frac{dz_i(t)}{dt} &= (-\gamma i z_i(t) + k_0) dt, \quad i \in \{1, \ldots, \ell\}, \\
\frac{d\omega_a(t)}{dt} &= -\alpha \omega_a(t)dr + n_{0,a}(z(t))dr, \quad a \in [p, d], \\
\frac{dw(t)}{dt} &= M \left(\omega_p(t), \omega_d(t), w(t)\right) dt.
\end{align*}
\]

and \(\underline{u}(t) = (\underline{x}(t), \underline{z}(t), \underline{\omega}(t), \underline{w}(t))\) with,
\[
\begin{align*}
\frac{d\underline{x}(t)}{dt} &= -\underline{x}(t)dt, \\
\frac{d\underline{z}(t)}{dt} &= \left(-\gamma \underline{z}(t) + C_k\right) dt, \\
\frac{d\underline{\omega}(t)}{dt} &= -\alpha \underline{\omega}(t)dt + C_n \left(1 + \ell\underline{z}(t)\right) dt, \\
\frac{d\underline{w}(t)}{dt} &= C_M \left(1 + \underline{\omega}(t)\right) dt.
\end{align*}
\]

**Lemma 9** Under the conditions of Sects. 2.2.3, 2.2.4, and 2.2.5 and if the initial conditions are such that \(u(0) \leq \underline{u}(0)\), then for all \(t \geq 0\), \(u(t) \leq \underline{u}(t)\).

**Proof** The result is clear for the function \((x(t))\) and also for the functions \((z_i(t))\). For \((\omega_a(t))\), with \(a \in [p, d]\), we have
\[
\frac{d(\omega_a(t) - \omega_{0,a}(t))}{dt} = -\alpha (\omega(t) - \omega_{0,a}(t))dt + \left(C_n \left(1 + \ell\underline{z}(t)\right) - n_{0,a}(z(t))\right) dt,
\]
by Condition (11), we obtain
\[
C_n \left(1 + \ell\underline{z}(t)\right) - n_{0,a}(z(t)) \geq C_n \left(1 + \|z(t)\|\right) - n_{0,a}(z(t)) \geq 0.
\]
Lemma 5 gives the relation
\[ e^{at} (\omega(t) - \omega_d(t)) = (\omega(0) - \omega_d(0)) + \int_0^t e^{-as} C_n \left( 1 + \ell \overline{\alpha}(s) - n_{0,a}(z(s)) \right) ds \geq 0. \]

Finally, again with Lemma 5, Condition 2.2.5 and the last inequality, we have, for \( t \geq 0 \),
\[ d(\omega(t) - w(t)) = -\mu(\omega(s) - w(s)) dt + \left( C_M(1 + \overline{\omega}(s)) - M_p(\omega_p(s), w(s)) \right) dt \]

This leads to,
\[ w(t) \leq \omega(t). \]

In the same way, we can prove that,
\[ w(t) \geq -\omega(t). \]

The lemma is proved. \( \square \)

**Proposition 10 (Coupling).** Under the conditions of Sect. 2.1, there exists a coupling of \( (\overline{U}(t)) \) and \( (U(t)) \) such that, almost surely, for all \( t > 0 \), \( (U(t)) \leq (\overline{U}(t)) \), in particular
\[ |W(t)| \leq \overline{W}(t), \quad \forall t \geq 0. \]

**Proof** All we have to prove is that if \( U(0) \leq \overline{U}(0) \) and if \( \tau \) is the first jump of either \( (U(t)) \) or \( (\overline{U}(t)) \), then \( U(t) \leq \overline{U}(t) \) for \( t \leq \tau \). Our statement is then easily proved by induction on the sequence of jumps of both processes.

Since \( (\overline{U}(t)) \) and \( (U(t)) \) are governed by the deterministic ODEs of Lemma 9, the relation \( U(t) \leq \overline{U}(t) \) holds for \( 0 \leq t < \tau \). The processes \( (W(t)) \) and \( (\overline{W}(t)) \) being continuous, \( |W(t)| = |W(\tau - )| \leq \overline{W}(\tau - ) = \overline{W}(\tau) \).

The instant of jump \( \tau \) is the minimum of \( \tau_1, \tau_2 \) and \( \overline{\tau}_2 \), with
\[
\begin{align*}
\tau_1 &= \inf \{ t > 0 : N_{\alpha}(0,t) \neq 0 \}, \\
\tau_2 &= \inf \left\{ t > 0 : N_{\beta,X}(0,t) = \int_{(0,t]} \mathcal{P}_2 ((0, \beta(X(s-)), ds) \neq 0 \right\}, \\
\overline{\tau}_2 &= \inf \left\{ t > 0 : N_{\overline{\beta},X}(0,t) = \int_{(0,t]} \mathcal{P}_2 ((0, \overline{\beta}(X(s-)), ds) \neq 0 \right\}.
\end{align*}
\]

Since, for \( x \leq \overline{x} \), \( \beta(x) \leq \overline{\beta}(x) \leq \overline{\beta}(\overline{x}) \) is a non-decreasing function and that \( X \leq \overline{X} \) holds until the first jump, the inequality \( \overline{\tau}_2 \leq \tau_2 \) holds almost surely.

If \( \tau_1 < \overline{\tau}_2 \), then
\[ X(\tau) = X(\tau-) + W(\tau-) \leq \overline{X}(\tau-) + \overline{W}(\tau-) = \overline{X}(\tau). \]

For \( i \in \{1, \ldots, \ell \} \),
\[ Z_i(\tau) = Z_i(\tau-) + k_i(Z(\tau-)) \leq \overline{Z}(\tau-) + C_k = \overline{Z}(\tau) \]

and
\[ \Omega(\tau) = \Omega(\tau-) + n_{a,1}(Z(\tau-)) \leq \Omega(\tau-) + C_n (1 + \|Z(\tau)\|) \leq \overline{\Omega}(\tau-) + C_n (1 + \ell \overline{Z}(\tau)) = \overline{\Omega}(\tau). \]

Thus we have \( U(\tau) \leq \overline{U}(\tau) \). The same arguments work in a similar way when \( \tau_2 = \overline{\tau}_2 < \tau_1 \).
In this case, we have
\[ X(\tau) = X(\tau -) - g(X(\tau -)) \leq \overline{X}(\tau -) = \overline{X}(\tau). \]

The last case \( \tau_2 < \min(\tau_2, \tau_1) \) is not more difficult, since the components of \( (U(t)) \) do not experience jumps and those of \( (\overline{U}(t)) \) have positive jumps due to \( \mathcal{N}_\beta, \overline{X} \). The proposition is proved. \[ \square \]

The process \( (\overline{U}_\varepsilon(t)) \) is defined by the SDEs,
\[
\begin{align*}
\mathrm{d}X_\varepsilon(t) &= -X_\varepsilon(t)\mathrm{d}t/\varepsilon + \overline{W}_\varepsilon(t)\mathcal{N}_{\lambda/\varepsilon}(\mathrm{d}t), \\
\mathrm{d}Z_\varepsilon(t) &= \left(-\gamma Z_\varepsilon(t) + C_k\right)\mathrm{d}t/\varepsilon + C_k\mathcal{N}_{\lambda/\varepsilon}(\mathrm{d}t) + C_k\mathcal{N}_{\beta/\varepsilon, \overline{X}_\varepsilon}(\mathrm{d}t), \\
\mathrm{d}\Omega_\varepsilon(t) &= -\alpha\Omega_\varepsilon(t)\mathrm{d}t + C_n \left(1 + \ell Z_\varepsilon(t)\right)\mathrm{d}t \\
&\quad + C_n \left(1 + \ell Z_\varepsilon(t-)ight) \left(\varepsilon\mathcal{N}_{\lambda/\varepsilon}(\mathrm{d}t) + \varepsilon\mathcal{N}_{\beta/\varepsilon, \overline{X}_\varepsilon}(\mathrm{d}t)\right), \\
\mathrm{d}W_\varepsilon(t) &= C_M \left(1 + \overline{\Omega}_\varepsilon(t)\right)\mathrm{d}t,
\end{align*}
\]
and with \( \overline{U}_\varepsilon(0) = \overline{U}(0) = (\overline{x}_0, \overline{z}_0, \overline{w}_0, \overline{w}_0) \).

The infinitesimal generator \( \overline{B}_w^F \) of Relation (18) is given by, for \( v = (x, z) \) and \( f \in \mathcal{C}_1^\beta(\mathbb{R} \times \mathbb{R}_+) \),
\[
\overline{B}_w^F(f)(v) \overset{\text{def}}{=} -x \frac{\partial f}{\partial x}(v) + (-\gamma z + C_k) \frac{\partial f}{\partial z}(v) \\
+ \lambda \left(f(v + w e_1 + C_k e_2) - f(v)\right) + C_\beta (x + 1) \left(f(v + C_k e_2) - f(v)\right),
\]
where \( e_1 = (1, 0) \) and \( e_2 = (0, 1) \).

6 Asymptotic Results for the Truncated Process

In this section we study the scaling properties of fast processes of \( (\overline{U}(t)) \) defined by Relation (26). In this section, we fix \( K > 0 \) and consider an analogue process for which the impact of the connection intensity is truncated at \( K \). In Sect. 6.3 an averaging principle will be established for this process as a first step in the proof of the main result of the paper.

6.1 Definition of the Truncated Process

We define \( (\overline{U}_\varepsilon^K(t)) \) as the solution of the SDEs,
\[
\begin{align*}
\mathrm{d}X_\varepsilon^K(t) &= -X_\varepsilon^K(t)\mathrm{d}t/\varepsilon + K \wedge \overline{W}_\varepsilon^K(t)\mathcal{N}_{\lambda/\varepsilon}(\mathrm{d}t), \\
\mathrm{d}Z_\varepsilon^K(t) &= \left(-\gamma Z_\varepsilon^K(t) + C_k\right)\mathrm{d}t/\varepsilon + C_k\mathcal{N}_{\lambda/\varepsilon}(\mathrm{d}t) + C_k\mathcal{N}_{\beta/\varepsilon, \overline{X}_\varepsilon^K}(\mathrm{d}t), \\
\mathrm{d}\Omega_\varepsilon^K(t) &= -\alpha\Omega_\varepsilon^K(t)\mathrm{d}t + C_n \left(1 + \ell Z_\varepsilon^K(t)\right)\mathrm{d}t \\
&\quad + C_n \left(1 + \ell Z_\varepsilon^K(t-)ight) \left(\varepsilon\mathcal{N}_{\lambda/\varepsilon}(\mathrm{d}t) + \varepsilon\mathcal{N}_{\beta/\varepsilon, \overline{X}_\varepsilon^K}(\mathrm{d}t)\right) \\
\mathrm{d}W_\varepsilon^K(t) &= C_M \left(1 + \overline{\Omega}_\varepsilon^K(t)\right)\mathrm{d}t,
\end{align*}
\]
and with \( \overline{U}_\varepsilon^K(0) = \overline{U}(0) = (\overline{x}_0, \overline{z}_0, \overline{w}_0, \overline{w}_0) \).

We begin with a lemma giving a stochastic upper bound of \( (\overline{X}_\varepsilon^K(t)) \) in terms of a standard shot-noise process.
Lemma 11 There exists a constant $C_X > 0$ independent of $\varepsilon$ such that the relation
\begin{equation}
X^K_\varepsilon (t) \leq C_X + KS_\varepsilon (t)
\end{equation}
holds for $t \geq 0$, where $(S_\varepsilon (t))$ is the shot-noise process of Definition 6. For any $\eta > 0$, there exists a compact subset $K$ of $\mathbb{R}^2_+$ such that,
\[
\sup_{0 < \varepsilon < 1} \mathbb{P} \left( \left( X^K_\varepsilon (t), Z^K_\varepsilon (t) \right) \notin K \right) \leq \eta.
\]

Proof With Relation (22), we have, for $t \geq 0$,
\[
X^K_\varepsilon (t) = x_0 e^{-t/\varepsilon} + \int_0^t e^{-(t-u)/\varepsilon} K \land \mathcal{W}_\varepsilon (u-) \mathcal{N}_{\lambda/\varepsilon} (du).
\]
Which gives, for $s \leq t \leq T$,
\[
X^K_\varepsilon (t) \leq x_0 e^{-t/\varepsilon} + K \int_0^t e^{-(t-u)/\varepsilon} \mathcal{N}_{\lambda/\varepsilon} (du) \leq x_0 + KS_\varepsilon (t),
\]
and therefore
\[
\mathbb{E} \left[ X^K_\varepsilon (t) \right] \leq C_X + K \mathbb{E} [S_\varepsilon (t)] \leq C_X + \lambda KT.
\]
Relation (22) gives the inequality
\[
\mathbb{E} \left[ Z^K_\varepsilon (t) \right] \leq z_0 + C_k \int_0^t \exp(-\gamma (t - s)) \left( 1 + \lambda + C_\beta (1 + \mathbb{E} \left[ X^K_\varepsilon (s) \right]) \right) ds
\]
which leads to,
\[
\sup_{0 < \varepsilon < 1} \mathbb{E} \left[ Z^K_\varepsilon (t) \right] < +\infty.
\]
We conclude by using Markov’s inequality.

6.2 Tightness of the Truncated Process

The next important lemma is used to prove tightness properties of the processes $(\Omega_\varepsilon (t))$.

Lemma 12 (Tightness of linear functionals of the fast processes) The family of processes
\[
\left( \int_0^t X^K_\varepsilon (u) du \right), \left( \int_0^t Z^K_\varepsilon (u) du \right), \left( \int_0^t X^K_\varepsilon (u) Z^K_\varepsilon (u) du \right), \varepsilon \in (0, 1),
\]
are tight for the convergence in distribution. The processes
\[
\left( \overline{M}^{K, 1}_\varepsilon (t) \right) \overset{\text{def.}}{=} \left( \int_0^t Z^K_\varepsilon (u-) \left[ \varepsilon \mathcal{N}_{\lambda/\varepsilon} (du) - \lambda du \right] \right)
\]
\[
\left( \overline{M}^{K, 2}_\varepsilon (t) \right) \overset{\text{def.}}{=} \left( \int_0^t Z^K_\varepsilon (u-) \left[ \varepsilon \mathcal{N}_{\beta/\varepsilon, X^K_\varepsilon} (du) - \beta \left( X^K_\varepsilon (u) \right) du \right] \right)
\]
converge in distribution to 0 as $\varepsilon$ goes to 0.
Proof Relation (29) gives for $0 \leq s \leq t$,
\[ \int_s^t \overline{X}_e^K(u)\,du \leq C_X(t-s) + K \int_s^t S_e(u)\,du, \]
The tightness of the three processes results from this relation and Proposition 8.
Indeed, Relation (22) shows that, for $t \geq 0$,
\begin{align*}
\overline{Z}_e^K(t) - z_0 &= C_k \int_0^t e^{-\gamma(t-s)/\varepsilon} \, ds + C_k \int_0^t e^{-\gamma(t-s)/\varepsilon} \, N_\lambda/\varepsilon(\,ds) \\
&\quad + C_k \int_0^t e^{-\gamma(t-s)/\varepsilon} \, p_2 \left( \left[ 0, \frac{1+\overline{X}_e^K(s)}{\varepsilon} \right] \right) \, ds \\
&\leq \frac{C_k}{\gamma} + C_k \int_0^t e^{-\gamma(t-s)/\varepsilon} \, N_\lambda/\varepsilon(\,ds) \\
&\quad + C_k \int_0^t e^{-\gamma(t-s)/\varepsilon} \, p_2 \left( \left[ 0, \frac{1+C_X}{\varepsilon} \right] \right) \, ds \\
&\quad + C_k \int_0^t e^{-\gamma(t-s)/\varepsilon} \, p_2 \left( \left[ \frac{1+C_X}{\varepsilon}, \frac{1+C_X}{\varepsilon} + C_k \frac{S_e(s)}{\varepsilon} \right] \right) \, ds.
\end{align*}
The first two terms of the right-hand side of last relation are, up to the constants $\gamma$ instead of 1 and $C_\beta(1+C_X)$ instead of $\lambda$, equal to $S_e(t)$. Similarly, up to the constant $C_\beta K$ of $S_e(t)$ instead of 1, the last term is equal to $R_e(t)$.

The two processes $(\overline{M}_{e,i}^K(t))$, $i=1, 2$ are martingales with previsible increasing processes
\[ \left( \varepsilon \lambda \int_0^t \overline{Z}_e^K(u)^2 \, du \right) \quad \text{and} \quad \left( \varepsilon \int_0^t \overline{Z}_e^K(u)^2 C_\beta \left( 1+\overline{X}_e^K(u) \right) \, du \right). \]
For $t \geq 0$, we have
\[ \mathbb{E} \left[ \overline{M}_{e,1}^K(t)^2 \right] \leq \int_0^t \mathbb{E} \left[ \overline{Z}_e^K(u)^2 \right] \, du, \]
and,
\begin{align*}
\mathbb{E} \left[ \overline{M}_{e,2}^K(t)^2 \right] &= \varepsilon C_\beta \left( \mathbb{E} \left[ \overline{M}_{e,1}^K(t)^2 \right] + \int_0^t \mathbb{E} \left[ \overline{Z}_e^K(u)^2 \overline{X}_e^K(u) \right] \, du \right) \\
&\leq \varepsilon C_\beta \int_0^t \left( \mathbb{E} \left[ \overline{Z}_e^K(u)^2 \right] + \sqrt{\mathbb{E} \left[ \overline{Z}_e^K(u)^4 \right]} \sqrt{\mathbb{E} \left[ \overline{X}_e^K(u)^2 \right]} \right) \, du,
\end{align*}
with Cauchy-Schwartz’ inequality.

Using the upper-bounds for $(\overline{X}_e^K(t))$ and $(\overline{Z}_e^K(t))$ and Relation (23) for $S_e$ and Proposition 24 of the Appendix for $R_e$, we obtain that the quantity $\mathbb{E} \left[ \overline{M}_{e,2}^K(t)^2 \right]$ converges to 0 as $\varepsilon$ goes to 0. The last statement of the lemma follows from Doob’s inequality.

We now define the associated occupation measure $\overline{\nu}_e^K$ in the same way as in Sect. 3. Let $G$ be a non-negative Borelian function on $[0, T] \times \mathbb{R}_+^2$, define $\overline{\nu}_e^K$ the non-negative measure on $[0, T] \times \mathbb{R}_+^2$ by
\[ \int_{[0,T] \times \mathbb{R}_+^2} G(s, x, z) \overline{\nu}_e^K(\,ds, \, dx, \, dz) \overset{\text{def.}}{=} \int_{[0,T]} G \left( s, \overline{X}_e^K(s), \overline{Z}_e^K(s) \right) \, ds. \]
Lemma 13 The family of random Radon measures $\overline{\nu}_\epsilon^K$, $\epsilon \in (0, 1)$, is tight for the convergence in distribution and for any bounded Borelian function on $[0, T] \times \mathbb{R}_+^2$, the set of processes

$$(I_G(t)) \overset{def}{=} \left( \int_0^t G \left( u, X^K_\epsilon(u), Z^K_\epsilon(u) \right) \, du, \, 0 \leq t \leq T \right)$$

is tight for the convergence in distribution.

Proof For $a > 0$, $\epsilon \in (0, 1)$, and $K$ a Borelian subset of $\mathbb{R}_+^2$, we have

$$\mathbb{P} \left( \overline{\nu}_\epsilon^K \left( [0, T] \times K^c \right) > a \right) \leq \frac{1}{a} \mathbb{E} \left[ \nu^K_\epsilon \left( [0, T] \times K^c \right) \right] \leq \frac{T}{a} \sup_{0 \leq t \leq T} \mathbb{P} \left( \overline{X}_\epsilon^K(t), \overline{Z}_\epsilon^K(t) \notin K \right)$$

Lemma 11 shows the existence of a compact set $K \subset \mathbb{R}_+^2$ such that the last term of the right-hand side of this inequality can be made arbitrarily small. Lemma 1.3 of Kurtz [21] gives that the family of random measures $\overline{\nu}_\epsilon^K$, $0 < \epsilon < 1$ is tight.

for the last part of the proposition, we use the criterion of modulus of continuity, see Billingsley [3]. This is a simple consequence of the inequality, for $0 \leq s, t \leq T$, $|I_G(t) - I_G(s)| \leq \|G\|_\infty |t - s|$. The lemma is proved. $\square$

Proposition 14 The family of random variables $(\overline{\Omega}_\epsilon^K(t), \overline{W}_\epsilon^K(t), \overline{\nu}_\epsilon^K)$, $\epsilon \in (0, 1)$, is tight.

Proof Tightness properties of $(\overline{\nu}_\epsilon^K)$ have been proved in Lemma 13. Relation (22) gives the relation, for $t \geq 0$,

$$\overline{\Omega}_\epsilon^K(t) - \overline{\omega}_0 e^{-\alpha t} = \int_0^t e^{-\alpha(t-s)} C_n \left( 1 + \epsilon \overline{Z}_\epsilon^K(s) \right) \left( 1 + \lambda + C_\beta (1 + \overline{X}_\epsilon^K(s)) \right) ds + \int_0^t e^{-\alpha(t-s)} C_n \left( 1 + \epsilon \overline{Z}_\epsilon^K(s) \right) \left[ \epsilon \mathbb{N}_{\lambda/\epsilon}(ds) - \lambda ds + \epsilon \mathbb{N}_{\beta/\epsilon, \overline{X}_\epsilon^K}(ds) - \beta \left( \overline{X}_\epsilon^K(s) \right) ds \right].$$

(30)

Lemma 12 shows that the family of processes associated to the first term of the right-hand side of this identity is tight, and that the process of the second term is vanishing in distribution as $\epsilon$ goes to 0. The family of processes $(\overline{\Omega}_\epsilon^K(t))$ is therefore tight and the tightness of $(\overline{W}_\epsilon^K(t))$ follows from its representation with $(\overline{\Omega}_\epsilon^K(t))$. The proposition is proved. $\square$

6.3 Averaging Principle for the Truncated Process $(\overline{U}_\epsilon^K(t))$

The goal of this section is to prove the following averaging principles for the truncated process. We start by stating the two following lemmas that are proved in Appendix A and that are essential to the proof of averaging principle.

We fix a sequence $(\epsilon_n)$ such that $(\overline{\nu}_\epsilon^K)$ is converging in distribution to $\overline{\nu}^K$. The first result focus on identifying the limiting linear functional of $\overline{\nu}^K$.

Lemma 15 For any continuous bounded Borelian function $G$ and $a, b, c \in \mathbb{R}_+$, the sequence of processes

$$\left( \int_0^t \left( a \overline{X}_\epsilon^K(s) + b \overline{Z}_\epsilon^K(s) + c \overline{X}_\epsilon^K(s) \overline{Z}_\epsilon^K(s) \right) G \left( \overline{X}_\epsilon^K(s), \overline{Z}_\epsilon^K(s) \right) ds \right)$$

is tight for the convergence in distribution.
converges in distribution to
\[
\left( \int_0^t \int_{\mathbb{R}^2} (ax + bz + cxz)G(x, z)\Pi^K(ds, dx, dz) \right).
\]

The second lemma shows that \( \Pi^K \) can be expressed as the product of the invariant measure \( \Pi_w \) and the Lebesgue measure. This is Lemma (1.4) of Kurtz [21].

**Lemma 16** For any non-negative Borelian function \( F \) on \( \mathbb{R}_+ \times \mathbb{R}_+^2 \), almost surely,
\[
\int_0^T F(s, x, z)\Pi^K(ds, dx, dz) = \int_0^T F(s, x, z)\Pi_{K \land \Pi^K(s)}(dx, dz)ds,
\]
where, for \( w \in \mathbb{R}_+ \), \( \Pi_w \) is the unique invariant distribution of the Markov process associated to the infinitesimal generator \( B^F_w \) defined by Relation (27).

With these two lemmas, Proposition 17 can be established.

**Proposition 17** Any limiting point \((\bar{\omega}^K(t), \bar{w}^K(t))\) of the family of processes \((\bar{\Omega}^K_\varepsilon(t), \bar{W}^K_\varepsilon(t))\), when \( \varepsilon \) goes to 0, verifies, almost surely for all \( t \geq 0 \), the ODE
\[
\begin{cases}
\bar{\omega}^K(t) = \bar{\omega}_0 - \alpha \int_0^t \bar{\omega}^K(s)ds \\
\quad + \int_0^t \int_{\mathbb{R}^2} C_n(1+\ell z)(1+\lambda + C\beta(1+x))\Pi_{\varepsilon}(dx, dz) ds \\
\bar{w}^K(t) = \bar{w}_0 + \int_0^t CM \left( 1 + \bar{\omega}^K(s) \right) ds,
\end{cases}
\]
holds, where \( \Pi_w \) is the unique invariant distribution of the Markov process associated to the infinitesimal generator \( B^F_w \) defined by Relation (27).

**Proof** Relation (30) gives the identity, for \( t \geq 0 \),
\[
\bar{\Omega}^K_\varepsilon(t) = \bar{\omega}_0 e^{-\alpha t} + e^{-\alpha t}\bar{M}^K_\varepsilon(t) + e^{-\alpha t} \int_0^t e^{\alpha s} C_n(1+\ell \bar{Z}_\varepsilon(s))(1+\lambda + C\beta(1+\bar{X}_\varepsilon^K(s))) ds,
\]
with
\[
\bar{M}^K_\varepsilon(t) \stackrel{\text{def.}}{=} \int_0^t e^{\alpha s} C_n \left( 1+\ell \bar{Z}_\varepsilon^K(s) \right) \left[ \varepsilon N_{\lambda/\varepsilon}(ds) + \varepsilon N_{\beta/\varepsilon, \bar{X}_\varepsilon^K}(ds) - (\lambda + \beta(\bar{X}_\varepsilon^K(s)))ds \right].
\]
Lemma 12 shows that \( \bar{M}^K_\varepsilon(t) \) is converging in distribution to 0 when \( \varepsilon \) goes to 0. We now use Lemmas 15 and 16 and we get that \((\bar{\omega}^K(t), \bar{w}^K(t))\) satisfies the desired relation.

\[\square\]

7 **Proof of an Averaging Principle**

Finally, this section gathers all the results from the previous sections to prove Theorem 4. The proof is done in two steps:

(a) Using an analytical result, an averaging principle for \((\bar{U}(t))\) is proved.
(b) The coupling of Sect. 5 is then used to show that a stochastic averaging result also holds in the general case.
7.1 Averaging Principle for the Coupled Process (\(\bar{U}(t)\))

We now turn to an analytical result by considering the dynamical system of Proposition 17 when \(K = +\infty\) and by showing an existence and uniqueness results which will be crucial in the proof of the general theorem.

For \(w \geq 0\), \(\Pi_w\) is the invariant distribution of the Markov process \((\overline{X}^w(t), \overline{Z}^w(t))\) satisfying the SDE

\[
\begin{align*}
\frac{d\overline{X}^w}{dt} &= -\overline{X}^w(t) dt + w \overline{\mathcal{N}}_\lambda(t), \\
\frac{d\overline{Z}^w}{dt} &= \left( -\gamma \overline{Z}^w(t) + C_k \right) dt + C_k \overline{\mathcal{N}}_\lambda(t) + C_k \overline{\mathcal{N}}_\beta, \overline{X}^w(t).
\end{align*}
\]

Its existence is a consequence of Proposition 25.

**Theorem 18** Under conditions of Sect. 2.2.5, there exists \(S_0 \in (0, +\infty]\) and a unique continuous function \((\overline{\omega}(t), \overline{w}(t))\) on \([0, S_0]\), solution of the ODE, for \(0 \leq t < S_0\),

\[
\begin{align*}
\frac{d\overline{\omega}}{dt} &= \overline{\omega}_0 - \alpha \int_0^t \overline{\omega}(s) ds + \int_0^t \int_{\mathbb{R}_+^2} C_n (1+\ell z) (1+\lambda + C_\beta (1+x)) \Pi_{\pi(s)} (dx, dz) ds, \\
\frac{d\overline{w}}{dt} &= \overline{w}_0 + \int_0^t C_M (1+\overline{\omega}(s)) ds,
\end{align*}
\]

with \((\overline{\omega}_0, \overline{w}_0) \in \mathbb{R}_+^2\).

Any limiting point \((\overline{\omega}^K(t), \overline{w}^K(t))\) of the family of processes \((\overline{\Omega}_\varepsilon^K(t), \overline{W}_\varepsilon^K(t))\), when \(\varepsilon\) goes to 0 is such that, for all \(0 \leq t < S_0\),

\[
\overline{\omega}^K(t) \leq \overline{\omega}(t) \text{ and } \overline{w}^K(t) \leq \overline{w}(t).
\]

**Proof** The existence of limiting points of the processes \((\overline{\Omega}_\varepsilon^K(t), \overline{W}_\varepsilon^K(t))\) is due to Proposition 17. If \((\overline{X}^w, \overline{Z}^w)\) is a random variable with distribution \(\Pi_w\), we have

\[
\overline{X}^w \overset{\text{dist.}}{=} w \int_0^{+\infty} e^{-s} \overline{\mathcal{N}}_\lambda (ds),
\]

and, with standard calculations, we obtain the relations

\[
\mathbb{E} \left[ \overline{X}^w \right] = \lambda w, \quad \mathbb{E} \left[ \left( \overline{X}^w \right)^2 \right] = \left( \lambda^2 + \frac{\lambda}{2} \right) w^2,
\]

and, consequently,

\[
\gamma \mathbb{E} \left[ \overline{Z}^w \right] = C_k \left( 1+\lambda + C_\beta \left( 1+\mathbb{E} \left[ \overline{X}^w \right] \right) \right) = C_k \left( 1+\lambda + C_\beta (1+\lambda w) \right).
\]

The SDEs for \((\overline{X}^w(t))\) and \((\overline{Z}^w(t))\) give

\[
\begin{align*}
\frac{d\overline{X}^w}{dt} &= \left( -(\gamma+1) \overline{X}^w(t) \overline{Z}^w(t) + C_k \overline{X}^w(t) \right) dt \\
&\quad + \left( w \overline{Z}^w(t) - C_k \overline{X}^w(t) + C_k w \right) \overline{\mathcal{N}}_\lambda (dt) + C_k \overline{\mathcal{N}}_\beta, \overline{X}^w (dt),
\end{align*}
\]

and thus, at equilibrium, we obtain the relation

\[
\mathbb{E} \left[ \overline{X}^w \overline{Z}^w \right] = C_k \frac{1}{\gamma+1} \left( \lambda w \left( 1+\mathbb{E} \left[ \overline{Z}^w \right] \right) + (1+\lambda + C_\beta) \mathbb{E} \left[ \overline{X}^w \right] + C_\beta \mathbb{E} \left[ \left( \overline{X}^w \right)^2 \right] \right).
\]
We have therefore that the function
\[ w \mapsto C_n \int_{\mathbb{R}^2_+} (1+\ell z)(1+\lambda + C_\beta (1+x)) \Pi_w(dx, dz) \]
is a non-decreasing and locally Lipschitz function. The existence and uniqueness follows from standard results for ODEs. There exists some \( S_0 > 0 \), such that, on the time interval \([0, S_0]\), the solution \((\vec{\omega}(t), \vec{w}(t))\) of the ODE is the limit of a Picard’s scheme \((\vec{\omega}_n(t), \vec{w}_n(t))\) associated to Relation (31) with
\[ (\vec{\omega}_0(t), \vec{w}_0(t)) = (\vec{\omega}^K(t), \vec{w}^K(t)) , \]
for all \( K \in \mathbb{R}_+ \). See Sect. 3 of Chapter 8 of Hirsch and Smale [15] for example. We now prove by induction that \((\vec{\omega}^K(t), \vec{w}^K(t)) \leq (\vec{\omega}_n(t), \vec{w}_n(t))\) holds on \([0, S_0]\) for all \( n \geq 1 \). If this is true for \( n \), then
\[
\begin{align*}
\vec{\omega}_{n+1}(t) &= \vec{\omega}_0 e^{-\alpha t} + \int_0^t e^{-\alpha (t-s)} \int_{\mathbb{R}^2} C_n (1+\ell z)(1+\lambda + C_\beta (1+x)) \Pi_{\vec{w}_n(s)}(dx, dz) \, ds \\
&\geq \vec{\omega}_0 e^{-\alpha t} + \int_0^t e^{-\alpha (t-s)} \int_{\mathbb{R}^2} C_n (1+\ell z)(1+\lambda + C_\beta (1+x)) \Pi_{\vec{w}^K(s) \wedge K}(dx, dz) \, ds \\
&= \vec{\omega}^K(t),
\end{align*}
\]
and the relation \( \vec{\omega}_{n+1}(t) \geq \vec{w}^K(t) \) follows directly. The proof by induction is completed. We just have to let \( n \) go to infinity to obtain the last statement of our proposition. \( \square \)

**Proposition 19** Under conditions of Sect. 2.2.5, for the convergence in distribution,
\[
\lim_{\varepsilon \to 0} ((\vec{\Omega}_\varepsilon(t), \vec{W}_\varepsilon(t)), t < S_0) = ((\vec{\omega}(t), \vec{w}(t)), t < S_0),
\]
where \((\vec{\Omega}_\varepsilon(t), \vec{W}_\varepsilon(t))\) is the process defined by SDEs (26) and \((\vec{\omega}(t), \vec{w}(t)), t < S_0\) by ODE (31).

**Proof** From Proposition 17, let \((\vec{\omega}^K(t), \vec{w}^K(t))\) be a limiting point, there exists a sequence \((\varepsilon_n)\) such that the sequence of processes \((\vec{\Omega}_{\varepsilon_n}(t), \vec{W}_{\varepsilon_n}(t))\) is converging to a continuous process \((\vec{\omega}^K(t), \vec{w}^K(t))\).

With the same notations as in Proposition 18, for any \( T < S_0 \), by continuity of \((\vec{\omega}(t), \vec{w}(t))\) on \([0, T]\), the quantity
\[
K_0 \overset{\text{def}}{=} 1 + \sup_{t \leq T} \vec{w}(t)
\]
is finite. Since \( \vec{w}^K(t) \leq \vec{w}(t) \) holds for all \( t \geq 0 \), the uniqueness result of Proposition 18 gives the identity
\[
((\vec{\omega}^K(t), \vec{w}^K(t)), t \leq T) = ((\vec{\omega}(t), \vec{w}(t)), t \leq T)
\]
for all \( K \geq K_0 \). Consequently, for any \( \eta > 0 \), there exists \( n_0 > 0 \) such that for \( n \geq n_0 \),
\[
P \left( \sup_{s \leq T} \vec{W}_{\varepsilon_n}^{K_0}(s) \geq K_0 \right) \leq P \left( \sup_{s \leq T} \vec{W}_{\varepsilon_n}^{K_0}(s) \geq 1 + \sup_{t \leq T} \vec{w}^{K_0}(t) \right) \leq \eta,
\]
since the process \((\vec{\omega}^{K_0}(t), \vec{w}^{K_0}(t))\) is upper-bounded, coordinate by coordinate on the time interval \([0, S_0]\), by \((\omega(t), w(t))\), defined by Relation (31). Note that \( S_0 \) is independent of the
sequence \((\varepsilon_n)\). Hence, for \(n \geq n_0\), Relation (28) gives

\[
P \left( \begin{align*}
\overline{X}_{\varepsilon_n}(s) &= \overline{X}_{\varepsilon_n}(s), \\
\overline{Z}_{\varepsilon_n}(s) &= \overline{Z}_{\varepsilon_n}(s)
\end{align*} \right)
\]

\[\geq P \left( \sup_{s \leq T} \overline{W}_{\varepsilon_n}(s) \leq K_0 \right) \geq 1 - \eta. \tag{33}\]

This shows that the sequence of processes \(((\overline{X}_{\varepsilon_n}(t), \overline{W}_{\varepsilon_n}(t)), t \leq T)\) is converging in distribution to \(((\overline{\omega}(t), \overline{w}(t)), t \leq T)\). The proposition is proved. \(\square\)

### 7.2 Averaging Principle for the Process \((U_\varepsilon(t))\)

We now conclude this section with the proof of Theorem 4. We fix \(T < S_0\).

The coupling property of Proposition 10 and Relation (33) give the existence of \(K_0\) and \(n_0\) such that for \(n \geq n_0\), the inequality

\[
P \left( \sup_{t \leq T} |W_{\varepsilon_n}(t)| \leq K_0 \right) \geq 1 - \eta
\]

holds.

We get that the results of Lemma 12 hold with \(X_{\varepsilon_n}^{K}\) and \(Z_{\varepsilon_n}^{K}\) replaced by \(X_{\varepsilon_n}\) and \(\|Z_{\varepsilon_n}\|\), and \(\overline{\beta}\) by \(\beta\). With the same arguments as in the proof of Lemma 13, the family of random measures \((\nu_{\varepsilon_n})\) defined by Relation (19) is tight and Lemma 15 holds for \((X_{\varepsilon_n}(t))\) and \((\|Z_{\varepsilon_n}\|(t))\).

With Relations (15) and (16), we get, for \(a \in \{p, d\}\) and \(t \geq 0\),

\[
\Omega_{\varepsilon, a}(t) = \omega_0.a \cdot \int_0^t e^{-\alpha(t-s)}n_{a,0}(Z_\varepsilon(s))ds + \int_0^t e^{-\alpha(t-s)}n_{a,1}(Z_\varepsilon(s-))\varepsilon N_{\ell,0}(ds) + \int_0^t e^{-\alpha(t-s)}n_{a,2}(Z_\varepsilon(s-))\varepsilon N_{\beta/\varepsilon, X_\varepsilon}(ds),
\]

and

\[
W_\varepsilon(t) = W_\varepsilon(0) + \int_0^t M \left( \Omega_{\varepsilon, p}(s), \Omega_{\varepsilon, d}(s), W(t) \right) ds.
\]

Relation (11) of the assumptions of Sect. 2.2.4 gives that \(n_{a,j}(z) \leq C_n(1 + \|z\|)\) for \(z \in /\ell\), \(a \in \{p, d\}\) and \(j \in \{0, 1, 2\}\). Lemma 12 applied to \((X_{\varepsilon_n}(t))\) and \((\|Z_{\varepsilon_n}\|(t))\) shows that the family of processes \((\Omega_{\varepsilon, a}(t))\) is tight, consequently that the same property hold for \((W_\varepsilon(t))\) due to the relation satisfied by \(M\) of Sect. 2.2.5. We have thus obtained the tightness of the sequence of processes

\[
(\Omega_{\varepsilon_n, p}(t), \Omega_{\varepsilon_n, d}(t), W_{\varepsilon_n}(t), \nu_{\varepsilon_n}).
\]

By taking a subsequence, we can assume it is converging in distribution to some process \((\alpha_p(t), \alpha_d(t), w(t), v)\).

We now identify the measure \(\nu\). Lemma (1.4) of Kurtz [21] shows that, for any bounded Borelian function \(G\) on \(\mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}^\ell\),

\[
\int_0^T G(s, x, z)\nu(ds, dx, dz) = \int_0^T \int_{\mathbb{R} \times \mathbb{R}^\ell} G(s, x, z)\gamma(s)(dx, dz)ds,
\]

\(\triangleleft\) Springer
where \((\gamma(s))\) is a previsible measure-valued process.

The continuity of the different functions \(g(\cdot)\) and \(k_i(\cdot)\) give that, for \(f\in C^1_b(\mathbb{R}\times\mathbb{R}_+^\ell)\), \((x, z, w)\mapsto B^F_w(f)(x, z)\) is continuous, where \(B^F_w\) is the operator defined by Relation (18).

Moreover, using

- Condition of Sect. 2.2.2 for the growth of the function \(\beta\);
- The boundedness of \(k_i, i \in \{1, 2\}\) of Condition of Sect. 2.2.3,

we have that, for \(f\in C^1_b(\mathbb{R}\times\mathbb{R}_+^\ell)\), there exists a constant \(C_0\) such that, for \((x, z)\in\mathbb{R}\times\mathbb{R}_+^\ell\),

\[
|B^F_w(f)(x, z)| \leq C_0 (1 + |x| + \|z\|) \left( \|f\|_\infty + \|\nabla f\|_\infty \right).
\]

We apply the equivalent of Lemma 15 to \((X_{\varepsilon_n}(t))\) and \((Z_{\varepsilon_n}(t))\) to get the relation, for \(T > 0\),

\[
\lim_{n \to +\infty} \int_0^T B^F_{w_{\varepsilon_n}(s)}(f)(X_{\varepsilon_n}(s), Z_{\varepsilon_n}(s)) \, ds = \int_0^T \int_{\mathbb{R}\times\mathbb{R}_+^\ell} B^F_w(f)(x, z) \nu(ds, dx, dz).
\]

From the SDEs (14) and (14) we have

\[
f \left( X_{\varepsilon_n}(s), Z_{\varepsilon_n}(s) \right) = f(x_0, z_0) + M^f_{\varepsilon_n}(t)
\]

\[
+ \frac{1}{\varepsilon} \int_0^t B^F_{w_{\varepsilon_n}(s)}(f) \left( X_{\varepsilon_n}(s), Z_{\varepsilon_n}(s) \right) \, ds,
\]

(34)

where \((M^f_{\varepsilon_n}(t))\) is the associated martingale.

In the same way as in the proof of Lemma 16, we show that \((\varepsilon_n M^f_{\varepsilon_n}(t))\) is converging in distribution to 0 which leads to the identity, almost surely for any \(t \leq T\),

\[
\int_0^t \int_{\mathbb{R}\times\mathbb{R}_+^\ell} B^F_{w(s)}(f)(x, z) \nu(ds, dx, dz) = 0,
\]

for \(f\) in a countable dense subset \(S\) of the functions of \(C^1_b(\mathbb{R}\times\mathbb{R}_+^\ell)\) with compact support. Consequently,

\[
\int_0^t \int_{\mathbb{R}\times\mathbb{R}_+^\ell} B^F_{w(s)}(f)(x, z) \gamma(s)(dx, dz) \, ds = 0,
\]

which gives that for almost all \(t\in[0, T]\)

\[
\int_{\mathbb{R}\times\mathbb{R}_+^\ell} B^F_{w(t)}(f)(x, z) \gamma(t)(dx, dz) = 0, \quad \forall f\in S.
\]

Proposition 25 of the Appendix gives therefore that, \(\gamma(t) = \Pi_{w(t)}\), for almost all \(t\in[0, T]\) (for Lebesgue’s measure), so that

\[
\int_0^T G(s, x, z) \nu(ds, dx, dz) = \int_0^T \int_{\mathbb{R}\times\mathbb{R}_+^\ell} G(s, x, z) \Pi_{w(s)}(dx, dz) \, ds,
\]

holds almost surely for all bounded continuous functions on \(\mathbb{R}_+\times\mathbb{R}\times\mathbb{R}_+^\ell\).

To establish the first identity of Relation (31), we need the convergence in distribution

\[
\lim_{n \to +\infty} \left( \int_0^t e^{-\alpha(t-u)} \begin{pmatrix} n_0(Z_{\varepsilon_n}(u)) \\ n_1(Z_{\varepsilon_n}(u)) \\ \beta(X_{\varepsilon_n}(u)) n_2(Z_{\varepsilon_n}(u)) \end{pmatrix} \, du \right).
\]
the random variable
\((x, z) \mapsto (n_0(z), n_1(z), \beta(x)n_2(z))\),

is \(\Pi_w\) almost everywhere continuous.

The theorem is therefore proved.

### 8 The Simple Model

In this section we consider the simple model defined in Sect. 1.1. Recall that the associated SDEs are

\[
\begin{align*}
\mathrm{d}X(t) &= -X(t)\mathrm{d}t + W(t-)\mathcal{N}_\beta(\mathrm{d}t), \\
\mathrm{d}Z(t) &= -\gamma Z(t)\mathrm{d}t + B_1\mathcal{N}_\beta(\mathrm{d}t) + B_2\mathcal{N}_{\beta,X}(\mathrm{d}t), \\
\mathrm{d}W(t) &= Z(t-)\mathcal{N}_{\beta,X}(\mathrm{d}t),
\end{align*}
\]

with \(\gamma > 0, B_1, B_2 \in \mathbb{R}_+,\) and \(\beta\) is assumed to be a Lipschitz function on \(\mathbb{R}_+\).

This is not, strictly speaking, a special case of the processes defined by Relations (7), but the tightness results of Sect. A of Appendix concerning occupation times of fast processes can obviously be used.

Let, for \(w \geq 0, (X^w(t), Z^w(t))\) be the fast processes associated to the model of Definition 2. Proposition 25 shows that \((X^w(t), Z^w(t))\) has a unique invariant distribution \(\Pi_w\). We denote by \((X^w_\infty, Z^w_\infty)\) a random variable with distribution \(\Pi_w\).

**Proposition 20** The function

\[
w \mapsto \mathbb{E}\left[Z^w_\infty \beta(X^w_\infty) \right] = \int_{\mathbb{R}_+^2} z\beta(x)\Pi_w(\mathrm{d}x, \mathrm{d}z)
\]

is locally Lipschitz on \(\mathbb{R}_+\).

**Proof** Assuming \(X^w(0) = Z^w(0) = 0\), Lemma 5 and Definition 6 give the relations

\[
\begin{align*}
X^w(t) &= w \int_0^t e^{-(t-s)}\mathcal{N}_\beta(\mathrm{d}s) = wX^1(t) \\
Z^w(t) &= B_1 \int_0^t e^{-\gamma(t-s)}\mathcal{N}_\beta(\mathrm{d}s) + B_2 \int_0^t e^{-\gamma(t-s)}\mathcal{P}_2((0, \beta(wX^1(s-)), \mathrm{d}s).
\end{align*}
\]

the random variable \((X^w(t), Z^w(t))\) is converging in distribution to \((X^w_\infty, Z^w_\infty)\) as well as any of its moments. Define

\[
\Psi_t(u) \overset{\text{def}}{=} \mathbb{E}\left[\beta(wX^1(t)) \int_0^t e^{-\gamma(t-s)}\mathcal{P}_2((0, \beta(wX^1(s-)), \mathrm{d}s) \right],
\]

for \(x, y \geq 0,\)

\[
|\Psi_t(x) - \Psi_t(y)| \\
\leq \mathbb{E}\left[|\beta(xX^1(t)) - \beta(yX^1(t))| \int_0^t e^{-\gamma(t-s)}\mathcal{P}_2((0, \beta(wX^1(s-)), \mathrm{d}s) \right].
\]
We note that \((X^1(t))\) is a functional of \(N_x\) and is therefore independent of the Poisson process \(\mathcal{P}_2\). We now take care of the two terms of the right-hand side of the last expression.

For the first term, if \(L_\beta\) is the Lipschitz constant of the function \(\beta\), we obtain

\[
\mathbb{E}
\left|
\beta (x X^1(t)) - \beta (y X^1(t)) \right|
\int_0^t e^{-\gamma(t-s)} \mathcal{P}_2((\beta(x X^1(s)), \beta(y X^1(s))), ds)
\right| \leq L_\beta |x-y| \mathbb{E}
\left[
X^1(t) \int_0^t e^{-\gamma(t-s)} \beta(x X^1(s)) ds
\right],
\]  

and, for the second term, if \(|y-x| \leq 1\),

\[
\mathbb{E}
\left|
\beta (y X^1(t)) \int_0^t e^{-\gamma(t-s)} \mathcal{P}_2((\beta(x X^1(s)), \beta(y X^1(s))), ds)
\right| \leq \mathbb{E}
\left[
\beta (y X^1(t)) \int_0^t e^{-\gamma(t-s)} \beta(x X^1(s)) - \beta(y X^1(s)) ds
\right] \leq L_\beta |x-y| \mathbb{E}
\left[
(\beta(0) + L_\beta (1+x) X^1(t)) \int_0^t e^{-\gamma(t-s)} X^1(s) ds
\right].
\]  

Fubini’s Theorem gives the relation,

\[
\mathbb{E}
\left[
X^1(t) \int_0^t e^{-\gamma(t-s)} \beta(x X^1(s)) ds
\right] = \int_0^t e^{-\gamma s} \mathbb{E}
\left[
X^1(t) \beta(x X^1(t-s))
\right] ds,
\]

and the convergence in distribution of the Markov process \((X^1(t))\) implies the convergence of \((\mathbb{E}
\left[
X^1(t) \beta(x X^1(t-s))
\right])\) to a finite limit when \(t\) goes to infinity. With Relations (35) and (36) and the expressions of \((X^w(t))\) and \((Z^w(t))\), we deduce that for \(x \geq 0\), there exists a constant \(F_x\) independent of \(t\) such that

\[
\left| \mathbb{E}
\left[
Z^x(t) \beta (X^x(t))
\right] - \mathbb{E}
\left[
Z^y(t) \beta (X^y(t))
\right] \right| \leq F_x |x-y|
\]

holds for all \(t \geq 0\) and \(y\) such that \(|y-x| \leq 1\). We conclude the proof of the proposition by letting \(t\) go to infinity. \(\square\)

The averaging principle for the simple model, announced in Sect. 1.2 of the introduction can now be stated.

**Theorem 21** If the function \(\beta\) is Lipschitz, there exists some \(S_0(w_0) > 0\), such that the family of processes \((W_\varepsilon(t), t < S_0(w_0))\) defined by Relation (2) converges in distribution to \((w(t), t < S_0(w_0)))\), the unique solution of the ODE

\[
\frac{dw}{dt}(t) = \int_{\mathbb{R}_+^2} z \beta(x) \Pi_w(t)(dx, dz)
\]

with \(w(0) = w_0\).

**Proof** For \(t \geq 0\),

\[
W_\varepsilon(t) = w_0 + \int_0^t Z_\varepsilon(s-) \varepsilon \mathcal{N}_{\beta/\varepsilon, X_\varepsilon}(ds),
\]

we then proceed as in Sect. 7 by using in particular the analogue of Lemma 12 and 15. \(\square\)
An explicit representation of the limiting connection intensity process can be obtained when linear activation functions.

**Proposition 22** If the activation function $\beta$ is such that $\beta(x) = v + \beta_0 x$ and

$$\Lambda_2 = \lambda \beta_0^2 B_2 \left( \frac{\lambda}{\gamma} + \frac{1}{2(\gamma+1)} \right), \quad \Lambda_1 = \lambda \beta_0 \left( \frac{B_1}{\gamma+1} + \frac{\lambda B_1 + 2v B_2}{\gamma} \right), \quad \Lambda_0 = \frac{v}{\gamma} (\lambda B_1 + v B_2),$$

then if $\Lambda > 0$, the asymptotic weight process $(w(t), 0 \leq t < \delta_0(0))$ of Theorem 21 with initial point $w_0 \geq 0$ can be expressed as:

(a) If $\Delta \overset{\text{def}}{=} \Lambda_1^2 - 4 \Lambda_2 \Lambda_0 > 0$, then

$$w(t) = \frac{s_2(w_0 + s_1)e^{\sqrt{\Delta}t} - s_1(w_0 + s_2)}{(w_0 + s_2) - (w_0 + s_1)e^{\sqrt{\Delta}t}}, \quad S_0(w_0) = \frac{1}{\sqrt{\Delta}} \ln \left( \frac{w_0 + s_2}{w_0 + s_1} \right),$$

with

$$s_1 \overset{\text{def}}{=} \frac{\Lambda_1 - \sqrt{\Delta}}{2 \Lambda_2} \quad \text{and} \quad s_2 \overset{\text{def}}{=} \frac{\Lambda_1 + \sqrt{\Delta}}{2 \Lambda_2}.$$

(b) If $\Delta = 0$, then

$$w(t) = \frac{2w_0 \Lambda_2 + \Lambda_1}{\Lambda_2 (2 - (2 \Lambda_2 w_0 + \Lambda_1) t)} - \frac{\Lambda_1}{2 \Lambda_2}, \quad S_0(w_0) = \frac{2}{2w_0 \Lambda_2 + \Lambda_1}.$$

(c) If $\Delta < 0$, then

$$w(t) = \frac{\sqrt{-\Delta}}{2 \Lambda_2} \left( \tan \left( \frac{1}{2} \sqrt{-\Delta} \cdot t + \arctan(z_0) \right) + \left\lfloor \frac{z_0 + 1}{2} \right\rfloor \pi \right) - \frac{\Lambda_1}{2 \Lambda_2},$$

with

$$S_0(w_0) = \frac{2}{\sqrt{-\Delta}} \left( \frac{\pi}{2} - \arctan(z_0) \right) \quad \text{and} \quad z_0 \overset{\text{def}}{=} \frac{2w_0 \Lambda_2 + \Lambda_1}{\sqrt{-\Delta}}.$$

It should be noted that under the conditions of this proposition, this model always exhibits a blow-up phenomenon.

**Proof** The SDEs give the relation

$$dX^w Z^w(t) = -(\gamma + 1) X^w Z^w(t) dt + (w Z^w(t-)+B_1 w + B_1 X^w(t-)) d\tilde{N}_t^w + 2 X^w(t-) Z^w(t-) d\tilde{N}_t^\beta,w.$$  

If the initial point has the same distribution as $(X^w_0, Z^w_0)$, by integrating and by taking the expected valued of this SDE, we obtain the identity

$$(\gamma + 1) \mathbb{E} \left[ X^w N^w_\infty \right] = \lambda w B_1 + (\lambda B_1 + v B_2) \mathbb{E} \left[ X^w_\infty \right] + \beta_0 B_2 \mathbb{E} \left[ (X^w_\infty)^2 \right] + \lambda w \mathbb{E} [Z^w_\infty].$$

With Relations (32), we have

$$\mathbb{E} \left[ X^w_\infty \right] = \lambda w, \quad \mathbb{E} \left[ (X^w_\infty)^2 \right] = \left( \lambda^2 + \frac{\lambda}{2} \right) w^2,$$

and, similarly,

$$\gamma \mathbb{E} [Z^w_\infty] = \lambda B_1 + B_2 (v + \beta_0 \mathbb{E}[X^w_\infty]) = \lambda B_1 + B_2 (v + \beta_0 \lambda w).$$
By using Theorem 21, with these identities, we obtain that \( (w(t)) \) satisfies the ODE
\[
\frac{dw}{dt}(t) = E\left[Z^w_\infty (v + \beta_0 X^w_\infty)\right] = \lambda_2 w^2 + \lambda_1 w + \lambda_0,
\]
on its domain of definition. We conclude the proof with trite calculations. \(\square\)

Appendix A: Proofs of Technical Results for Occupation Times

A.1 Proof of Lemma 15

Denote, for \( t \geq 0, a, b, c \in \mathbb{R}_+, \) and \( \varepsilon > 0, \)
\[
L_\varepsilon(t) \overset{\text{def.}}{=} aX^K_\varepsilon(t) + bZ^K_\varepsilon(t) + cX^K_\varepsilon(t)Z^K_\varepsilon(t).
\]
Let \( G \) be a continuous bounded Borelian function. From Proposition 12, we can extract a sub-sequence \( (\varepsilon_n) \) such that, for the convergence in distribution
\[
\lim_{n \to +\infty} \left( \int_0^t L_{\varepsilon_n}(u)G\left(X^K_{\varepsilon_n}(u), Z^K_{\varepsilon_n}(u)\right) \, du \right) = (L(t)),
\]
where \( (L(t)) \) is a continuous càdlàg process.

We will now prove that the process \( (L(t)) \) is such that,
\[
(L(t)) = \left( \int_0^t \int_{\mathbb{R}^2} (ax + bz + cxz)G(x, z)\nu^K(ds, dx, dz) \right).
\]

For \( A > 0, \) the convergence of \( (\nu^K_{\varepsilon_n}) \) to \( \nu^K \) gives the convergence in distribution,
\[
\lim_{n \to +\infty} \left( \int_0^t A \wedge L_{\varepsilon_n}(s)G\left(X^K_{\varepsilon_n}(s), Z^K_{\varepsilon_n}(s)\right) \, ds \right)
\]
\[
= \left( \int_0^t \int_{\mathbb{R}^2} A \wedge (ax + bz + cxz)G(x, z)\nu^K(ds, dx, dz) \right). \tag{38}
\]

Using again the upper-bound, Relation (29), for \( (X^K_{\varepsilon}(t)) \) with Relation (23), and Proposition 24 for \( R_\varepsilon, \) we obtain that
\[
C_L \overset{\text{def.}}{=} \sup_{0 < \varepsilon < 1, 0 \leq t \leq T} E\left[ L_\varepsilon(s)^2 \right] < + \infty,
\]
hence, for \( \eta > 0, \)
\[
\mathbb{P}\left( \int_0^T (L_\varepsilon(s) - A)^+ ds \geq \eta \right) \leq \frac{1}{\eta} \int_0^T E\left[ (L_\varepsilon(s) - A)^+ \right] ds
\]
\[
\leq \frac{1}{\eta A} \int_0^T E\left[ L_\varepsilon(s)^2 \right] ds \leq \frac{C_L T}{\eta A}.
\]
Since \( G \) is bounded, with the elementary relation \( x = x \wedge A + (x - A)^+, x \geq 0, \) then, for \( n \geq 1, \)
\[
\mathbb{P}\left( \sup_{0 \leq s \leq T} \left| \int_0^t L_{\varepsilon_n}(u)G\left(X^K_{\varepsilon_n}(u), Z^K_{\varepsilon_n}(u)\right) \, du \right| \right).
\]
For any $A > 0$ and $n \geq 1$, Cauchy-Schwartz’s inequality gives the relation

$$
\mathbb{E} \left[ \int_0^T A \wedge L_{\varepsilon_n}(u) G \left( \frac{X_{\varepsilon_n}^K(u)}{\eta}, \frac{Z_{\varepsilon_n}^K(u)}{\eta} \right) \, du \right] \leq C L T \| G \|_\infty.
$$

(39)

With Relation (38) and the integrability properties proven just above, we obtain that, for $n \to \infty$, we get the inequality

$$
\mathbb{E} \left[ \int_0^T A \wedge \left( ax + bz + cz \right) G(x, z) \nu^K \left( ds, dx, dz \right) \right] \leq \sqrt{C L T} \| G \|_\infty.
$$

By letting $A$ go to infinity, the monotone convergence theorem shows that

$$
\lim_{A \to +\infty} \mathbb{E} \left[ \int_0^T \int_{\mathbb{R}_+^2} A \wedge \left( ax + bz + cz \right) G(x, z) \nu^K \left( ds, dx, dz \right) \right] = \mathbb{E} \left[ \int_0^T \int_{\mathbb{R}_+^2} (ax + bz + cz) G(x, z) \nu^K \left( ds, dx, dz \right) \right] \leq \sqrt{C L T} \| G \|_\infty. + \infty.
$$

(40)

With Relation (39) and the integrability properties proven just above, we obtain that, for $\varepsilon > 0$, there exists $n_0$ such that if $n \geq n_0$, then the relation

$$
\mathbb{P} \left( \sup_{0 \leq t \leq T} \left| \int_0^t L_{\varepsilon_n}(u) G \left( \frac{X_{\varepsilon_n}^K(u)}{\eta}, \frac{Z_{\varepsilon_n}^K(u)}{\eta} \right) \, du \right| + \int_0^T \int_{\mathbb{R}_+^2} (ax + bz + cz) G(x, z) \nu^K \left( ds, dx, dz \right) \geq \eta \right) \leq \varepsilon
$$

holds. Lemma 15 is proved.

### A.2 Proof of Lemma 16

Following Papanicolaou et al. [24] and Kurtz [21], we first show that there exists an optional process $(\Gamma_{\varepsilon_n}^K)$, with values in the set of probability distributions on $\mathbb{R}_+^2$, such that, almost surely, for any bounded Borelian function $G$ on $\mathbb{R}_+ \times \mathbb{R}_+^2$, the relation

$$
\int_{\mathbb{R}_+ \times \mathbb{R}_+^2} G(s, x, z) \nu^K \left( ds, dx, dz \right) = \int_{\mathbb{R}_+ \times \mathbb{R}_+^2} G(s, x, z) \Gamma_{\varepsilon_n}^K \left( dx, dz \right) \, ds.
$$

(40)

Recall that the optional $\sigma$-algebra is the smallest $\sigma$-algebra containing adapted càdlàg processes. See Sect. VI.4 of Rogers and Williams [29] for example. This is a simple consequence of Lemma 1.4 of Kurtz [21] and the fact that, due to Relation (19), the measure $\nu^K \left( ds, \mathbb{R}_+^2 \right)$ is the Lebesgue measure on $[0, T]$.

Let $f \in C^1_{\varepsilon_n}(\mathbb{R}_+^2)$ be a bounded $C^1$-function on $\mathbb{R}_+^2$ with bounded partial derivatives, we have the relation

$$
\varepsilon f \left( \frac{X_{\varepsilon_n}^K(t)}{\eta}, \frac{Z_{\varepsilon_n}^K(t)}{\eta} \right) = \varepsilon f \left( \bar{x}_0, \bar{z}_0 \right) + \varepsilon M^f_{\varepsilon_n}(t)
$$

$$
+ \int_0^t B^{f}_{K \wedge \nu^K} \left( f \left( \frac{X_{\varepsilon_n}^K(s)}{\eta}, \frac{Z_{\varepsilon_n}^K(s)}{\eta} \right) \right) ds,
$$

(41)
where, for $t \geq 0$, if $(\overline{V}_e^K(s))^{\text{def}} = (\overline{X}_e^K(s), \overline{Z}_e^K(s))$,

$$
\overline{M}_e^f(t) \stackrel{\text{def}}{=} \int_0^t \left( f \left( \overline{V}_e^K(s-) + (K \wedge \overline{W}_e^K(s), 1) \right) - f \left( \overline{V}_e^K(s-) \right) \right) \left[ N_{\lambda/\varepsilon}(ds) - \frac{\lambda}{\varepsilon} ds \right] 
+ \int_0^t \left( f \left( \overline{V}_e^K(s-) + (0, 1) \right) - f \left( \overline{V}_e^K(s-) \right) \right) \left[ N_{\beta/\varepsilon,X_e^K}(ds) - \frac{\beta}{\varepsilon} (X_e^K(s)) ds \right].
$$

Proposition 12 shows that the martingale $(\varepsilon \overline{M}_e^f(t))$ is converging in distribution to 0 as $\varepsilon$ goes to 0.

Relation (41) gives therefore the convergence in distribution

$$
\lim_{\varepsilon \to 0} \left( \int_0^T B_{K \wedge \overline{W}_e^K(s)}^F (f) \left( \overline{X}_e^K(s), \overline{Z}_e^K(s) \right) ds \right) = 0.
$$

The convergence in distribution of $(\overline{S}_e^K(t), \overline{W}_e^K(t), \overline{V}_e^K)$, Proposition 15 and Relation (40) give that, for any $f \in C^1_b(\mathbb{R}^2_+)$, the relation

$$
\left( \int_0^T \int_0^T B_{K \wedge \overline{W}_e^K(s)}^F (f)(x, z) \Gamma^K_{s,t} (dx, dz) ds, 0 \leq t \leq T \right) = (0, 0 \leq t \leq T).
$$
holds with probability 1.

Let $(f_n)$ be a dense countable sequence in $C^1_b(\mathbb{R}^2_+)$ and $\mathcal{E}_1$ be the event, where Relation (42) holds for all $f = f_n$, $n \geq 1$. Note that that $P(\mathcal{E}_1) = 1$. On $\mathcal{E}_1$, there exists a (random) subset $S_1$ of $[0, T]$ with Lebesgue measure $T$ such that

$$
\int_{\mathbb{R}^2_+} B_{K \wedge \overline{W}_e^K(s)}^F (f_n)(x, z) \overline{\Pi}^K_s (dx, dz) = 0, \forall s \in S_1 \text{ and } \forall n \geq 1,
$$
and, consequently,

$$
\int_{\mathbb{R}^2_+} B_{K \wedge \overline{W}_e^K(s)}^F (f)(x, z) \overline{\Pi}^K_s (dx, dz) = 0, \forall s \in S_1 \text{ and } \forall f \in C^1_b(\mathbb{R}^2_+).
$$

By Proposition 25, for $s \in S_1$, the probability distribution $\overline{\Pi}^K_s$ is the invariant distribution $\Pi_{K \wedge \overline{W}_e^K(s)}$. Lemma 16 is proved.

**Appendix B: Shot-Noise Processes**

This section presents several technical results on shot-noise processes which are crucial for the proof of Theorem 4. See Schottky [30], Rice [26] and Gilbert and Pollak [9] for an introduction.

**B.1 A Scaled Shot-Noise Process**

Recall that $(S_x^0(t))$, with initial point $x \geq 0$, has been introduced by Definition 6. We will have the following conventions,

$$
(S_x(t))^{\text{def}} = (S_x^0(t)), (S_x^1(t))^{\text{def}} = (S^0_1(t)) \text{ and } (S(t))^{\text{def}} = (S^0_1(t)).
$$
The process \((S(t))\) is in fact the standard shot-noise process of Lemma 5 associated to the Poisson process \(N_\lambda\), for \(t \geq 0\),

\[
S(t) = \int_0^t e^{-(t-s)}N_\lambda(ds) \overset{\text{dist.}}{=} \int_0^t e^{-s}N_\lambda(ds).
\]

(43)

In particular \((S(t))\) is a stochastically non-decreasing process, i.e. for \(y \geq 0\) and \(s \leq t\),

\[
\mathbb{P}(S(s) \geq y) \leq \mathbb{P}(S(t) \geq y).
\]

(44)

A classical formula for Poisson processes, see Proposition 1.5 of Robert [27] for example, gives the relation, for \(\xi \in \mathbb{R}\),

\[
\mathbb{E}[e^{\xi S(t)}] = \exp\left(-\lambda\int_0^t (1 - \exp(\xi e^{-s}))\,ds\right),
\]

(45)

in particular \(\mathbb{E}[S^x(t)] = x \exp(-t) + \lambda(1 - \exp(-t))\). It also shows that \((S^x(t))\) is converging in distribution to \(S(\infty)\) such that,

\[
\mathbb{E}[e^{\xi S(\infty)}] = \exp\left(-\lambda\int_0^{+\infty} (1 - \exp(\xi e^{-s}))\,ds\right) < +\infty.
\]

It is easily seen that \((S^x_\varepsilon(t)) \overset{\text{dist.}}{=} (S^x(t/\varepsilon))\) and thus with Relation (22), \((S^x_\varepsilon(t))\) can be represented as, for \(t \geq 0\),

\[
S^x_\varepsilon(t) \overset{\text{def.}}{=} xe^{-t/\varepsilon} + \int_0^t e^{-(t-s)/\varepsilon}N_{\lambda/\varepsilon}(ds) = xe^{-t/\varepsilon} + S_x(t).
\]

(46)

We remind here the results of Proposition 7, that will be proved in the following paragraph. For \(\xi \in \mathbb{R}\) and \(x \geq 0\), the convergence in distribution of the processes

\[
\lim_{\varepsilon \to 0} \left(\int_0^t e^{\xi S^x_\varepsilon(u)}\,du\right) = \left(\mathbb{E}[e^{\xi S(\infty)}]\right)t
\]

holds, and

\[
\sup_{0<\varepsilon \leq 1} \mathbb{E}[e^{\xi S_x(t)}] < +\infty.
\]

Proof of Proposition 7 Let \(T_1\) and \(T_2\) be two stopping times bounded by \(N, \theta > 0\), and verifying \(0 \leq T_2 - T_1 \leq \theta\). Using Relation (46) and the strong Markov property of Poisson processes, we have that

\[
\mathbb{E}\left[\int_{T_1}^{T_2} e^{\xi S^x_\varepsilon(u)}\,du\right] = \varepsilon \mathbb{E}\left[\int_{T_1/\varepsilon}^{T_2/\varepsilon} e^{\xi S^x(u)}\,du\right]
\]

\[
= \varepsilon \mathbb{E}\left[\int_0^{(T_2 - T_1)/\varepsilon} e^{\xi S^x(T_1/\varepsilon)(u)}\,du\right] \leq \varepsilon \mathbb{E}\left[e^{\xi S(T_1/\varepsilon)}\mathbb{E}\left[\int_0^{\theta/\varepsilon} e^{\xi S(u)}\,du\right]\right]
\]

\[
\leq \theta e^{\xi \mathbb{E}\left[e^{\xi S(\theta/\varepsilon)}\right]} \mathbb{E}\left[e^{\xi S(\infty)}\right] \leq \theta e^{\xi \mathbb{E}\left[e^{\xi S(\infty)}\right]^2},
\]

holds, by stochastic monotonicity of \((S(t))\) of Relation (44).

Aldous’ Criterion, see Theorem VI.4.5 of Jacod and Shiryaev [16] gives that the family of processes

\[
\left(\int_0^t e^{\xi S_x(u)}\,du\right),
\]

\(\xi \in \mathbb{R}\) an
is tight when $\varepsilon$ goes to 0. For $p \geq 1$ and a fixed vector $(t_i) \in \mathbb{R}_+^P$, the ergodic theorem for the Markov process $(S(t))$ gives the almost-sure convergence of

$$
\lim_{\varepsilon \to 0} \left( \int_0^{t_i} e^{\xi S_\varepsilon(u)} \, du, \; i = 1, \ldots, p \right) = \lim_{\varepsilon \to 0} \left( \varepsilon \int_0^{t_i/\varepsilon} e^{\xi S(u)} \, du, \; i = 1, \ldots, p \right)
$$

$$
= \left( \mathbb{E} \left[ e^{\xi S(\infty)} \right] t_i, \; i = 1, \ldots, p \right).
$$

Hence, due to the tightness property, the convergence also holds in distribution for the processes

$$
\lim_{\varepsilon \to 0} \left( \int_0^t e^{\xi S_\varepsilon(u)} \, du \right) = \left( \mathbb{E} \left[ e^{\xi S(\infty)} \right] t \right).
$$

The last part is a direct consequence of the identity $(S^\varepsilon(t)) \overset{\text{dist.}}{=} (S^\varepsilon(t/\varepsilon))$, and of Relation (45), which gives

$$
\mathbb{E} \left[ e^{\xi S_\varepsilon(t)} \right] = \exp \left( -\lambda \int_0^{t/\varepsilon} \left( 1 - \exp (\xi e^{-s}) \right) ds \right)
$$

$$
\leq \exp \left( -\lambda \int_0^{+\infty} \left( 1 - \exp (\xi e^{-s}) \right) ds \right).
$$

The proposition is proved. \hfill \Box

### B.2 Interacting Shot-Noise Processes

Recall that $(R(t))$ defined by Relation (24) is a shot-noise process with intensity equal to the shot-noise process $(S_\varepsilon(t))$.

We start with a simple result on moments of functionals of Poisson processes.

**Lemma 23** If $Q$ is a Poisson point process on $\mathbb{R}_+$ with a positive Radon intensity measure $\mu$ and $f$ is a Borelian function such that

$$
I_k(f) \overset{\text{def.}}{=} \int f(u)^k \mu(du) < +\infty, \; 1 \leq k \leq 4,
$$

then

$$
\mathbb{E} \left[ \left( \int f(u) Q(du) \right)^4 \right] = (I_4 + 6I_1^2 I_2 + 4I_1 I_3 + 3I_2^2 + I_1^4) (f).
$$

**Proof** It is enough to prove the inequality for non-negative bounded Borelian functions $f$ with compact support on $\mathbb{R}_+$.

The formula for the Laplace transform of Poisson point processes, see Proposition 1.5 of Robert [27], gives for $\xi \geq 0$,

$$
\mathbb{E} \left[ \exp \left( \xi \int_0^{+\infty} f(u) Q(du) \right) \right] = \exp \left( \int_0^{+\infty} \left( e^{\xi f(u)} - 1 \right) \mu(du) \right).
$$

The proof is done in a straightforward way by differentiating the last identity with respect to $\xi$ four times and then set $\xi = 0$. \hfill \Box

 Springer
**Proposition 24** The inequality

\[
\sup_{\varepsilon \in (0,1)} \mathbb{E} \left[ R_\varepsilon(t)^4 \right] < +\infty
\]

holds.

**Proof** Denote, for \( t \geq 0 \),

\[
J_{k,\varepsilon}(t) \overset{\text{def}}{=} \int_0^t e^{-\gamma k(t-u)/\varepsilon} \frac{S_\varepsilon(u)}{\varepsilon} du,
\]

the identity \((S_\varepsilon(t))^{\text{dist}} = (S(t/\varepsilon))\) and Relation (43) coupled with Fubini’s Theorem give the relations

\[
J_{k,\varepsilon}(t) = \int_0^{t/\varepsilon} e^{-\gamma k(t/\varepsilon-u)/\varepsilon} S(u) du = \frac{1}{\gamma k-1} \int_0^{t/\varepsilon} \left( e^{-(t/\varepsilon-v)} - e^{-\gamma k(t/\varepsilon-v)} \right) N_\lambda(dv)
\]

\[
\overset{\text{dist}}{=} \frac{1}{\gamma k-1} \int_0^{t/\varepsilon} \left( e^{-v} - e^{-\gamma k v} \right) N_\lambda(dv) \leq \frac{1}{|\gamma k-1|} J_k
\]

with

\[
\overline{J}_k \overset{\text{def}}{=} \int_0^{+\infty} \left( e^{-k\gamma v} + e^{-v} \right) N_\lambda(dv).
\]

Relation (22) applied to \( R_\varepsilon(t) \) gives

\[
R_\varepsilon(t) = \int_0^t e^{-\gamma(t-u)/\varepsilon} \mathcal{P}_2 \left( \left( 0, \frac{S_\varepsilon(u)}{\varepsilon} \right) \right), du.
\]

The quantity \( S_\varepsilon(u) \) is a functional of the point process \( \mathcal{P}_1 \) and is therefore independent of the Poisson point process \( \mathcal{P}_2 \). Lemma 23 gives therefore that

\[
\mathbb{E} \left[ R_\varepsilon(t)^4 \mid \mathcal{P}_1 \right] = J_{4,\varepsilon}(t) + 6 J_{1,\varepsilon}(t)^2 J_{2,\varepsilon}(t) + 4 J_{1,\varepsilon}(t) J_{3,\varepsilon}(t) + 3 J_{2,\varepsilon}(t)^2 + J_{1,\varepsilon}(t)^4,
\]

hence,

\[
\mathbb{E} \left[ R_\varepsilon(t)^4 \right] \leq \mathbb{E} \left[ \frac{\overline{J}_4}{|\gamma-1|^4} + \frac{6 \overline{J}_1^2 \overline{J}_2}{|\gamma-1|^2 |2\gamma-1|^2} + \frac{4 \overline{J}_1 \overline{J}_3}{|\gamma-1| |3\gamma-1|} + \frac{3 \overline{J}_2^2}{|2\gamma-1|^2} + \overline{J}_4 \right].
\]

Again with Proposition 7 we obtain that, for \( k \geq 1 \), the variable \( \overline{J}_k \) has finite moments of all orders, therefore by Cauchy-Shwartz’ Inequality, the right-hand side of the last inequality is finite. The proposition is proved.

**Appendix C: Equilibrium of Fast Processes**

For \( w \in K_w \), recall that the Markov process \((X^w(t), Z^w(t))\) of Definition 2 is such that

\[
\begin{align*}
\mathrm{d}X^w(t) &= -X^w(t) \mathrm{d}t + w \mathcal{N}_\lambda(\mathrm{d}t) - g \left( X^w(t) - \right) \mathcal{N}_{\tilde{\beta}, X^w}(\mathrm{d}t) \quad (47) \\
\mathrm{d}Z^w(t) &= (-\gamma \odot Z^w(t) + k_0) \mathrm{d}t + k_1 (Z^w(t) -) \mathcal{N}_\lambda(\mathrm{d}t) + k_2 (Z^w(t) -) \mathcal{N}_{\tilde{\beta}, X^w}(\mathrm{d}t). \quad (48)
\end{align*}
\]
Proposition 25 Under the conditions of Sections 2.2.2 and 2.2.3, the Markov process \((X^w(t), Z^w(t))\) solution of the SDEs (47) and (48) has a unique invariant distribution \(\Pi_w\), i.e. the unique probability distribution \(\mu\) on \(\mathbb{R} \times \mathbb{R}^+\) such that

\[
\left(\mu, B^F_w(f)\right) = \int_{\mathbb{R} \times \mathbb{R}^+} B^F_w(f)(x, z) \mu(dx, dz) = 0,
\]

for any \(f \in C^1_b(\mathbb{R} \times \mathbb{R}^+)\), where \(B^F_w\) is the operator defined by Relation (18).

Proof We denote by \((X^w_n, Z^w_n)\) the embedded Markov chain of the Markov process \((X^w(t), Z^w(t))\), i.e. the sequence of states visited by \((X^w(t), Z^w(t))\) after each jump, of either \(N_\lambda\) or \(N_{\beta,x}\).

The proof of the proposition is done in three steps. We first show that the return time of \((X^w(t), Z^w(t))\) to a compact set of \(\mathbb{R} \times \mathbb{R}^+\) is integrable. Then we prove that the Markov chain \((X^w_n, Z^w_n)\) is Harris ergodic, and consequently that it has a unique invariant measure. For a general introduction on Harris Markov chains, see Nummelin [23] and Meyn and Tweedie [22]. Finally, the proof of the proposition uses the classical framework of stationary point processes.

C.1 Integrability of Return Times to a Compact Subset

Suppose that \(w \geq 0\). The conditions of Sect. 2.2.2 on the functions \(\beta\) and \(g\), and Relation (47) show that \(X^w(t) \geq -c_0\), for all \(t \geq 0\), if \(X^w(0) \geq -c_0\), with \(c_0 = C_\beta + c_g\). The state space of the Markov process \((X^w(t), Z^w(t))\) can be taken as \(S^d = [-c_0, +\infty) \times \mathbb{R}^+\).

Define, for \((x, z) \in S\) and \(0 < a \leq 1\),

\[
H(x, z) = x + a||z||, \quad \text{with} \quad ||z|| = \sum_{i=1}^\ell z_i,
\]

we get that

\[
B^F_w(H)(x, z) = -x + \left(-a \sum_{i=1}^\ell \gamma_i z_i + k_{0,i}\right) + \lambda \left(w + a \sum_{i=1}^\ell k_{1,i}(z)\right) + \beta(x) \left(-g(x) + a \sum_{i=1}^\ell k_{2,i}(z)\right)
\]

hence, with the assumptions of Sect. 2.2.3 and 2.2.2 on the function \(k\) and \(\beta\), and \(a \leq 1\),

\[
B^F_w(H)(x, z) \leq -x - a\gamma ||z|| + \ell C_k + \lambda (w + \ell a C_k) + C_{\beta}(1 + x) \ell a C_k
\]

\[
\leq (\ell a C_{\beta} C_k - 1) x - a\gamma ||z|| + (\ell C_k + \ell a C_k) + C_{\beta}(1 + x) \ell a C_k
\]

\[
\leq (\ell a C_{\beta} C_k - 1) x - a\gamma ||z|| + C,
\]

where \(\gamma > 0\) is the minimum of the coordinates of \(\gamma\) and \(C\) is a constant independent of \(x, z\) and \(a\). We fix \(0 < a \leq 1\) sufficiently small so that \(\ell a C_{\beta} C_k < 1\) and \(K > c_0\) such that

\[
C < \gamma K /2 - 1 \quad \text{and} \quad C < (1 - \ell a C_{\beta} C_k) K /2 - 1.
\]

If \(H(x, z) > K\) then \(\max(x, a||z||) > K/2\) and therefore \(B^F_w(H)(x, z) \leq -1\), \(H\) is therefore a Lyapunov function for \(B^F_w\). One deduces that the same result holds for the return time of Markov chain, \((X^w_n, Z^w_n)\) in the set \(I_K = \{(x, z) : H(x, z) \leq K\}\).
C.2 Harris Ergodicity of \((X^w_n, Z^w_n)\)

Proposition 5.10 of Nummelin [23] is used to show that \(I_K\) is a recurrent set. A regeneration property would be sufficient to conclude. In particular, we can prove that \(I_K\) is a small set, that is, there exists some positive, non-trivial, Radon measure \(\nu\) on \(S\) such that,

\[
\mathbb{P}_{(x_0, z_0)} \left((X^w_1, Z^w_1) \in S\right) \geq \nu(S),
\]

(50)

for any Borelian subset \(S\) of \(S\) and all \((x_0, z_0) \in I_K\).

We denote by \(s_1\), resp. \(t_1\), the first instant of \(N_x\), resp. of \(N_{123}\), then, for \((X^w_0, Z^w_0) = (x_0, z_0) \in I_K\), by using the deterministic differential equations between jumps, we get

\[
\mathbb{P}_{(x_0, z_0)} (s_1 < t_1) = \mathbb{E}\left[\exp\left(-\int_0^{s_1} \beta(x_0 \exp(-s)) \, ds\right)\right] \geq \mathbb{E}\left[\exp(-c_1^1 \beta s_1)\right] = p_0 = \frac{\lambda}{\lambda + c_1^1},
\]

since \(\beta\) is bounded by some constant \(c_1^1\) on the interval \([-c_0, K]\).

In the following argument, we restrict \(X\) to be non-negative, the extension to \([-c_0, +\infty\) is straightforward. For \(A = [0, A] \in \mathcal{B}(\mathbb{R}_+)\) and \(B = [0, B] \in \mathcal{B}(\mathbb{R}^\ell_+),\) from Equations (47) and (48), we obtain the relation

\[
\mathbb{P}_{(x_0, z_0)} (X^w_1, Z^w_1) \in A \times B \geq p_0 \mathbb{P}\left((X^w_1, Z^w_1) \in A \times B \mid s_1 < t_1\right)
\]

\[
= p_0 \mathbb{P}\left(x_0 e^{-s_1} + w, A, (z_0 - k_0) \cap e^{-\gamma s_1} + k_1 \cap (z_0 - k_0) \cap e^{-\gamma s_1} + k_0 \in B \mid s_1 < t_1\right)
\]

\[
\geq \mathbb{P}\left(H((z_0 - k_0) \cap e^{-\gamma s_1} + k_1), (z_0 - k_0) \cap e^{-\gamma s_1} + k_0) \in B \mid s_1 < t_1\right)
\]

\[
= p_0 \mathbb{P}\left(H((z_0 - k_0) \cap e^{-\gamma s_1} + k_1), (z_0 - k_0) \cap e^{-\gamma s_1} + k_0) \in B\right),
\]

where \(H(z) = z + k_1(z), s_1^{\text{dist}} = (s_1 \mid s_1 \leq t_1)\). By using the fact that \(k_1\) is in \(c_1^1(\mathbb{R}^\ell_+, \mathbb{R}_+)\) by the conditions of Sect. 2.2.3 and in the same way as Example of Sect. 4.3.3 page 98 of Meyn and Tweedie [22], we can prove that the random variable

\[
\left(x_0 e^{-\tilde{s}_1} + w, H((z_0 - k_0) \cap e^{-\gamma \tilde{s}_1} + k_0)\right)
\]

has a density, uniformly bounded below by a positive function \(h\) on \(\mathbb{R}_+ \times \mathbb{R}^\ell_+\), so that

\[
\mathbb{P}_{(x_0, z_0)} (X^w, Z^w) \in A \times B \geq \int_{A \times B} h(x, z) \, dx \, dz, \forall A \in \mathcal{B}(\mathbb{R}_+), B \in \mathcal{B}(\mathbb{R}^\ell_+),
\]

for all \((x_0, z_0) \in I_K\). This relation is then extended to all Borelian subsets \(S\) of \(S\), so that Relation (50) holds. Proposition 5.10 of Nummelin [23] gives therefore that \((X^w_n, Z^w_n)\) is Harris ergodic.

If \(w < 0\), the last two steps can be done in a similar way. In this case, the process \((-X^w(t))\) satisfies an analogous equation with the difference that the process \(N_{123} X^w\) does not jump when \((-X^w(t)) > c_1^1\) since \(\beta(x) = 0\) for \(x \leq -c_1^1\).

C.3 Characterization of \(\Pi_w\)

Let \(\widehat{\Pi}_w\) be the invariant probability distribution of \((X^w_n, Z^w_n)\). With the above notations,

\[
\mathbb{E}_{\widehat{\Pi}_w} \left[\min(s_1, t_1)\right] \leq \mathbb{E}_{\widehat{\Pi}_w} \left[s_1\right] = \frac{1}{\lambda} < +\infty,
\]

the probability defined by the classical cycle formula,
for any bounded Borelian function on \( \mathbb{R} \times \mathbb{R}_+^\ell \) is an invariant distribution for the process 
\((X^w(t), Z^w(t))\).

Proposition 9.2 of Ethier and Kurtz [7] shows that any distribution is invariant for 
\((X^w(t), Z^w(t))\) if and only if it satisfies Relation (49). It remains to prove the uniqueness 
of the invariant distribution, using the fact that the embedded Markov chain has a unique 
invariant distribution.

Although this is a natural result, we have not been able to find a reference in the literature. 
Most results are stated for discrete time, the continuous time is usually treated by looking at 
the process on a “discrete skeleton”, i.e. at instants multiple of some positive constant. See 
Proposition 3.8 of Asmussen [1] for example. As this technique is not adapted to our system, 
we derive a different proof using the Palm measure of the associated stationary point process.

If \( \mu \) is some invariant distribution of the Markov process \((X^w(t), Z^w(t))\), we build a 
stationary version \(((X^w(t), Z^w(t)), t \in \mathbb{R})\) of it on the whole real line. In particular, we have 
that \((X^w(t), Z^w(t)) \overset{\text{dist}}{=} \mu\), for all \( t \in \mathbb{R} \).

We denote by \((S_n, n \in \mathbb{Z})\) the non-decreasing sequence of the jumps (due to \(N^\beta\) and 
\(\hat{N}_{\beta, X^w}\)), with the convention \(S_0 \leq 0 < S_1\) The sequence \(((X^w(S_n), Z^w(S_n)), n \geq 0)\) has the 
same distribution as the process \(((X^w_n, Z^w_n), n \geq 0)\), the Markov chain with initial state 
\((X^w(S_0), Z^w(S_0))\). Since, for any \( t \in \mathbb{R} \),
\[
\left((X^w(s+t), Z^w(s+t)), s \in \mathbb{R}\right) \overset{\text{dist}}{=} \left((X^w(s), Z^w(s)), s \in \mathbb{R}\right),
\]
the marked point process \(T \overset{\text{def}}{=} (S_n, (X^w(S_n), Z^w(S_n)), n \in \mathbb{Z})\) is a stationary point process, 
i.e.
\[
((S_n, X^w(S_n), Z^w(S_n), n \in \mathbb{Z}) \overset{\text{dist}}{=} ((S_n-t, X^w(S_n), Z^w(S_n), n \in \mathbb{Z}), \forall t \in \mathbb{R}.
\]
The Palm measure of \(T\) is a probability distribution \(\hat{Q}\) such that the sequence 
\(((S_n-S_{n-1}, X^w(S_n), Z^w(S_n), n \in \mathbb{Z})\) is stationary. See Chapter 11 of Robert [27] for a quick 
presentation of stationary point processes and Palm measures.

Under \(\hat{Q}\), the Markov chain \(((X^w_n, Z^w_n), n \geq 0)\) is at equilibrium. Using Harris 
ergodicity, we have proved in the previous section that the Markov chain \(((X^w_n, Z^w_n), n \geq 0)\) has a unique invariant measure. Considering that both sequences \(((X^w_n, Z^w_n), n \geq 0)\) and 
\(((X^w(S_n), Z^w(S_n)), n \geq 0)\) have the same distribution, we have that \(\hat{Q}(\mathbb{R}_+^\ell, \cdot)\) is uniquely 
determined.

Moreover, remembering that,
\[
\hat{Q} (S_n-S_{n-1}>t) = \mathbb{E}_{\hat{Q}} \left[ \exp \left( - \int_0^t \beta \left( X^w_n(s) e^{-s} \right) \, ds \right) \right]
\]
We have that \(\hat{Q}\) is entirely determined by the ergodic distribution of the embedded Markov 
chain and consequently that the Palm measure \(\hat{Q}\) is unique. By Proposition 11.5 of 
Robert [27], the distribution of \(T\) is expressed with \(\hat{Q}\).

We have, for every bounded function \(f\),
\[
\mathbb{E}_\mu \left[ f(X^w(0), Z^w(0)) \right] = \mathbb{E}_T \left[ f(X^w(S_0)e^{S_0}, Z^w(S_0) \odot e^{\gamma S_0}) \right],
\]
which uniquely determines the invariant distribution \(\mu\).

The proposition is proved.  \(\square\)
Appendix D: Averaging Principles for Discrete Models of Plasticity

In this section, we present a general discrete model of plasticity, state the associated averaging principle theorem and give a sketch of its proof. We will only point out the differences with the proof of the main result of this paper, Theorem 4.

For this model of plasticity, the membrane potential $X$, the plasticity processes $Z$ and the synaptic weight $W$ are integer-valued variables. This system is illustrated in Sect. 7 of Robert and Vignoud [28] for calcium-based models. It amounts to represent these three quantities $X$, $Z$ and $W$ as multiple of a “quantum” instead of a continuous variable. The leaking mechanism in particular, the term corresponding to $-\gamma Y(t)\, dt$ in the continuous model, $Y\in\{X, Z, W\}$ and $\gamma>0$, in the SDEs, is represented by the fact that each quantum leaves the system at a fixed rate $\gamma$.

The main advantage of this model is that simple analytical expressions of the invariant distribution are available.

**Definition 26** The SDEs for the discrete model are

\[
\begin{align*}
\text{d}X(t) &= -\mathcal{N}_{I,X}(dt) + W(t-)\mathcal{N}_\lambda(dt) - \mathcal{N}_{I,\beta X}(dt), \\
\text{d}Z(t) &= -\mathcal{N}_{I,Y Z}(dt) + B_1\mathcal{N}_{\lambda_1}(dt) + B_2\mathcal{N}_{I,\beta X}(dt), \\
\text{d}\Omega_a(t) &= -\alpha\sum_{a,i} \Omega_{a,i}(t)dt + n_{a,i}(Z(t))dt \\
&\quad + n_{a,1}(Z(t-))\mathcal{N}_{\lambda_1}(dt) + n_{a,2}(Z(t-))\mathcal{N}_{I,\beta X}(dt), \quad a\in\{p, d\}, \\
\text{d}W(t) &= -\mathcal{N}_{I,\mu W}(dt) + A_{p}\mathcal{N}_{I,\Omega_p}(dt) - A_d\mathbb{1}_{\{w(t-)=A_d\}}\mathcal{N}_{I,\Omega_d}(dt),
\end{align*}
\]  

(51)

where $\beta, \gamma, \mu$ are non-negative real numbers, $B_1, B_2\in\mathbb{N}^\ell$ and, for $a\in\{p, d\}$, $A_a\in\mathbb{N}$. The functions $n_{a,i}$ are assumed to be bounded by $C_n$.

For $a\in\{p, d\}$, the function $I$ of $\mathcal{N}_{I,G}$ for $G\in\{X, \beta X, \gamma Z, \mu W, \Omega_p, \Omega_d\}$, defined by relation (6), is the identity function $I(x)=x$, $x\in\mathbb{R}$ and $\mathcal{N}_\lambda$ is a Poisson process on $\mathbb{R}_+$ with rate $\lambda$. All associated Poisson processes are assumed to be independent.

**Definition 27** For a fixed $w$, the process of the fast variables $(X^w(t), Z^w(t))$ on $\mathbb{N}\times\mathbb{N}^\ell$ of the SDEs is the Markov process whose transition rates are given by, for $(x, z)\in\mathbb{N}\times\mathbb{N}^\ell$,

\[
(x, z) \mapsto \begin{cases} 
(x+w, z+B_1) & \lambda, \\
(x-1, z) & x, \\
(x-1, z+B_2) & \beta x.
\end{cases}
\]

**Theorem 28** (Averaging Principle for a Discrete Model) If the assumptions of Definition 26 are verified, the family of scaled processes $(W_\varepsilon(t))$ associated to Relations (51) is converging in distribution, as $\varepsilon$ goes to 0, to the càdlàg integer-valued process $(w(t))$ satisfying the ODE

\[
\text{d}w(t) = -\mathcal{N}_{I,Y w}(dt) + A_p\mathcal{N}_{I,\omega_p}(dt) - A_d\mathbb{1}_{\{w(t-)=A_d\}}\mathcal{N}_{I,\omega_d}(dt),
\]

(52)

and, for $a\in\{p, d\}$,

\[
\frac{d\omega_a(t)}{dt} = -\alpha\omega_a(t) + \int_{\mathbb{N}\times\mathbb{N}^\ell} (n_{a,0}(z) + \lambda n_{a,1}(z) + \beta(x)n_{a,2}(z))\Pi_w(dx, dz),
\]

where $\Pi_w$ is the invariant distribution of the Markov process of Definition 27.
Proof. Again, we have to show that, on a fixed finite interval, the process \((W(t))\) is bounded with high probability. A coupled process that stochastically bounds from above the discrete process is also defined.

Definition 29 The process \((\overline{X}(t), \overline{Z}(t), \overline{\Omega}(t), \overline{W}(t))\) satisfies the following SDEs

\[
\begin{align*}
\frac{d\overline{X}(t)}{dt} &= -N_{1,\overline{X}}(dt) + W(t-)N_{\lambda}(dt), \\
\frac{d\overline{Z}(t)}{dt} &= -N_{1,\overline{Z}}(dt) + B_1N_{\lambda}(dt) + B_2N_{1,\overline{X}}(dt), \\
\frac{d\overline{\Omega}(t)}{dt} &= -\alpha\overline{\Omega}(t)dt + C_n\overline{\Omega}(dt) + C_nN_{1,\overline{X}}(dt), \\
\frac{d\overline{W}(t)}{dt} &= A_pN_{1,\overline{W}}(dt),
\end{align*}
\]

where \(B_1, B_2 \in \mathbb{N}^\ell\) and, for \(a \in \{p, d\}, A_p \in \mathbb{N}\).

It is not difficult to prove that this process is indeed a coupling that verifies the relation \(W(t) \leq \overline{W}(t)\), for all \(t \geq 0\) and that the process \((\overline{W}(t))\) is non-decreasing.

From the SDEs governing the scaled version of the coupled system, we obtain

\[
\begin{align*}
\mathbb{E}[\overline{W}_\varepsilon(t) - w_0] &\leq A_p\mathbb{E}\left[\int_0^t N_{1,\overline{\Omega}}(s)\,ds\right] \\
&\leq A_p t \mathbb{E}\left[\omega_0 + C_n \sup_{s \leq t} \int_0^s e^{-\alpha(s-u)} \left(\varepsilon N_{1,\overline{W}}(u) + \varepsilon N_{1,\overline{W}}(u)\right)\,du\right] \\
&\leq A_p t \left(\omega_0 + C_n \frac{(1+\lambda)}{\alpha} + \mathbb{E}\left[\int_0^t \beta \overline{X}(u)\,du\right]\right) \\
&\leq A_p t \left(\omega_0 + C_n \frac{(1+\lambda)}{\alpha} + \frac{\beta}{\alpha} \mathbb{E}\left[\int_0^t \overline{W}_\varepsilon(u)\,du\right]\right) \\
&\leq D + D \int_0^t \mathbb{E}[\overline{W}_\varepsilon(u)]\,du,
\end{align*}
\]

for all \(t \leq T\), for some constant \(D \geq 0\). Gronwall’s Lemma gives a uniform bound, with respect to \(\varepsilon\),

\[
\mathbb{E}\left[\sup_{t \leq T} \overline{W}_\varepsilon(t)\right] = \mathbb{E}[\overline{W}_\varepsilon(T)] \leq (D + w_0)e^{DT}.
\]

Using Markov inequality, we have then that, for any \(\eta > 0\), the existence of \(K_0\) and \(n_0\) such that \(n \geq n_0\), the inequality

\[
\mathbb{P}\left(\sup_{t \leq T} \overline{W}_\varepsilon(t) \leq K_0\right) \geq 1 - \eta
\]

holds. We can then finish the proof in the same way as in Sect. 7.2. The tightness property of the family of càdlàg processes \((\overline{W}_\varepsilon(t)), \varepsilon \in (0, 1)\) are proved with Aldous’ criterion, see Theorem VI.4.5 of Jacod and Shiryaev [16].

We have to prove the uniqueness of the solution of Relation (52) and the convergence in distribution of the scaled process to the process \(w(t)\). For this, we need to have some Lipschitz property on the limiting system, and finite first moments for the invariant distribution of \((X^w(t), Z^w(t))\). This is proved in Sect. 7 of Robert and Vignoud [28] for the case where \(Z\) is a one-dimensional process, the extension to multi-dimensional \(Z\) is straightforward. \(\square\)

References

1. Asmussen, S.: Applied Probability and Queues: Stochastic Modelling and Applied Probability, 2nd edn. Springer, New York (2003)
2. Beznea, L., Cîmpean, I., Röckner, M.: A new approach to the existence of invariant measures for Markovian semigroups. Ann. Inst. Henri Poincaré Probab. Stat. 55(2), 977–1000 (2019)
3. Billingsley, P.: Convergence of Probability Measures. Wiley, New York (1999)
4. Cáceres, M.J., Carrillo, J.A., Perthame, B.: Analysis of nonlinear noisy integrate & fire neuron models: blow-up and steady states. J. Math. Neurosci. 1(1), 7 (2011)
5. Crane, R., Sornette, D.: Robust dynamic classes revealed by measuring the response function of a social system. Proc. Natl. Acad. Sci. 105(41), 15649–15653 (2008)
6. Embrechts, P., Liniger, T., Lin, L.: Multivariate Hawkes processes: an application to financial data. J. Appl. Probab. 48(A), 367–378 (2011)
7. Ethier, S.N., Kurtz, T.G.: Markov Processes: Characterization and Convergence. Wiley, New York (2009)
8. Freidlin, M.I., Wentzell, A.D.: Random Perturbations of Dynamical Systems, 2nd edn. Springer, New York (1998)
9. Gilbert, E.N., Pollak, H.O.: Amplitude distribution of shot noise. Bell Syst. Tech. J. 39(2), 333–350 (1960)
10. Gust, G., Schbath, S.: FADO: a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes’ model. In: Statistical Applications in Genetics and Molecular Biology 4 (2005)
11. Has’minski˘ı, R.Z.: Stochastic stability of differential equations. Alphen aan den Rijn: Sijthoff & Noordhoff, pp. xvi+344 (1980)
12. Hawkes, A.G.: Spectra of some self-exciting and mutually exciting point processes. Biometrika 58(1), 83–90 (1971)
13. Hawkes, A.G., Oakes, D.: A cluster process representation of a self-exciting process. J. Appl. Probab. 11(3), 493–503 (1974)
14. Helson, P.: A new stochastic STDP rule in a neural network model (2018). arXiv:1706.00364 [math]
15. Hirsch, M.W., Smale, S.: Differential equations, dynamical systems, and linear algebra. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York, pp. xi+358 (1974)
16. Jacod, J., Shiryaev, A.N.: Limit Theorems for Stochastic Processes. Grundlehren der Mathematischen Wissenschaften, vol. 288, p. xvi+601. Springer, Berlin (1987)
17. Karatzas, I., Shreve, S.: Brownian Motion and Stochastic Calculus. Graduate Texts in Mathematics, vol. 113, 2nd edn. Springer, New York (1998)
18. Kempter, R., Gerstner, W., van Hemmen, J.L.: Hebbian learning and spiking neurons. Phys. Rev. E 59(4), 4498–4514 (1999)
19. Kerstan, J.: Teilprozesse Poissonscher Prozesse. In: Trans. Third Prague Conf. Information Theory, Statist. Decision Functions, Random Processes (Liblice, 1962). Publ. House Czech. Acad. Sci., Prague, pp. 377–403 (1964)
20. Kingman, J.F.C.: Poisson Processes. Clarendon Press, Oxford (1992)
21. Kurtz, T.G.: Averaging for martingale problems and stochastic approximation. In: Karatzas, I., Ocone, D. (eds.) Applied Stochastic Analysis, vol. 177, pp. 186–209. Springer, Berlin (1992)
22. Meyn, S., Tweedie, R.: Markov Chains and Stochastic Stability. Communications and Control Engineering Series. Springer, New York (1993)
23. Nummelin, E.: General Irreducible Markov Chains and Non-Negative Operators. Cambridge University Press, Cambridge (2004)
24. Papanicolaou, G., Stroock, D.W., Varadhan, S.R.S.: Martingale approach to some limit theorems. In: Proc. 1976. Duke Conf. On Turbulence. III. Duke University Math (1977)
25. Reynaud-Bouret, P., Schbath, S.: Adaptive estimation for Hawkes processes; application to genome analysis. Ann. Stat. 38(5), 2781–2822 (2010)
26. Rice, S.O.: Mathematical analysis of random noise. Bell Syst. Tech. J. 23(3), 282–332 (1944)
27. Robert, P.: Stochastic Networks and Queues. Stochastic Modelling and Applied Probability. Springer, Berlin (2003)
28. Robert, P., Vignoud, G.: Stochastic models of synaptic plasticity in neural networks (2020). arXiv:2010.08195
29. Rogers, L.C.G., Williams, D.: Diffusions, Markov Processes and Martingales: Itô Calculus, vol. 2. Cambridge University Press, Cambridge (2000)
30. Schottky, W.: Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern. Annalen der physik 362(23), 541–567 (1918)

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