Université Nice Sophia Antipolis - UFR Sciences
École Doctorale de Sciences Fondamentales et Appliquées

Thèse

pour obtenir le titre de

Docteur en Sciences

de l'Université Nice Sophia Antipolis

Discipline : Mathématiques

présentée et soutenue par

Julien CHEVALLIER

Modélisation de grands réseaux de neurones par processus de Hawkes

Modelling large neural networks via Hawkes processes

Thèse dirigée par
Patricia REYNAUD-BOURET
et
François DELARUE

soutenue le 9 septembre 2016.

Jury :

M.	François DELARUE	*Co-directeur de thèse*	Professeur
M.	Olivier FAUGERAS	*Examinateur*	DR INRIA
Mme.	Eva LÖCHERBACH	*Examinatrice*	Professeur
Mme.	Sylvie MÉLÉARD	*Rapporteuse*	Professeur
M.	Benoit PERTHAME	*Examinateur*	Professeur
Mme.	Patricia REYNAUD-BOURET	*Directrice de thèse*	DR CNRS
M.	Mathieu ROSENBAUM	*Rapporteur*	Professeur

Résumé. Comment fonctionne le cerveau ? Peut-on créer un cerveau artificiel ? Une étape essentielle en vue d'obtenir une réponse à ces questions est la modélisation mathématique des phénomènes à l'œuvre dans le cerveau. Ce manuscrit se focalise sur l'étude de modèles de réseaux de neurones inspirés de la réalité. Cette thèse se place à la rencontre entre trois grands domaines des mathématiques - l'étude des équations aux dérivées partielles (EDP), les probabilités et la statistique - et s'intéresse à leur application en neurobiologie.

Dans un premier temps, nous établissons les liens qui existent entre deux échelles de modélisation neurobiologique. À un niveau microscopique, l'activité électrique de chaque neurone est représentée par un processus ponctuel. À une plus grande échelle, un système d'EDP structuré en âge décrit la dynamique moyenne de ces activités. Il est alors montré que le modèle macroscopique (système d'EDP) peut se retrouver de deux manières distinctes : en étudiant la dynamique moyenne d'un neurone typique ou bien en étudiant la dynamique d'un réseau de n neurones en champ-moyen quand n tend vers $+\infty$. Dans le second cas, la convergence vers une dynamique limite est démontrée et les fluctuations de la dynamique microscopique autour de cette limite sont examinées.

Dans un second temps, nous construisons une procédure de test d'indépendance entre processus ponctuels, ces derniers étant destinés à modéliser l'activité de certains neurones. Ses performances sont contrôlées en termes de niveau asymptotique grâce à une approximation gaussienne valide sous l'hypothèse que les processus ponctuels sont des processus de Poisson homogènes. Cette validité théorique est vérifiée d'un point de vue pratique par une étude par simulations. Pour finir, notre procédure est appliquée sur de vraies données.

Abstract. How does the brain compute complex tasks? Is it possible to create en artificial brain? In order to answer these questions a key step is to build mathematical models for information processing in the brain. Hence this manuscript focuses on biological neural networks and their modelling. This thesis lies in between three domains of mathematics - the study of partial differential equations (PDE), probabilities and statistics - and deals with their application to neuroscience.

On the one hand, the bridges between two neural network models, involving two different scales, are highlighted. At a microscopic scale, the electrical activity of each neuron is described by a temporal point process. At a larger scale, an age structured system of PDE gives the global activity. There are two ways to derive the macroscopic model (PDE system) starting from the microscopic one: by studying the mean dynamics of one typical neuron or by investigating the dynamics of a mean-field network of n neurons when n goes to infinity. In the second case, we furthermore prove the convergence towards an explicit limit dynamics and inspect the fluctuations of the microscopic dynamics around its limit.

On the other hand, a method to detect synchronisations between two or more neurons is proposed. To do so, tests of independence between temporal point processes are constructed. The asymptotic level of the tests are controlled thanks to a Gaussian approximation which is valid as soon as the point processes are stationary Poisson processes. This theoretical control as well as the practical validity of the method are illustrated by a simulation study. Finally, the method is applied on real data.

Keywords: **Partial differential equations.** Age structured equations, measure solutions, mean-field system. **Probabilities.** Temporal point processes, Hawkes process, Poisson process, law of large numbers, mean-field interacting particle system, propagation of chaos, central limit theorem. **Statistics.** Independence test, multiple tests. **Neuroscience.** Neural network, multi-scale modelling, neuronal assemblies, Unitary Events method, coincidence pattern.

A celle qui partage ma vie,
mes envies et aussi mes soucis.

Remerciements

Avant tout, je tiens à remercier Patricia qui m'a offert un cadre de travail sérieux, pédagogique et familial : je suis sûr que tu trouveras d'autres schtroumpfs à guider vers le monde de la recherche. Avec François, vous avez grandement contribué à mon épanouissement mathématique et à la qualité de ce manuscrit grâce à vos lumières, conseils avisés et relectures attentives. Je ne saurais vous remercier assez pour tout le temps que vous m'avez accordé. J'ose espérer avoir hérité d'un peu de votre talent.

Je tiens à remercier mes rapporteurs, Sylvie Méléard et Mathieu Rosenbaum. Je suis profondément honoré par l'intérêt qu'ils ont porté à mon manuscrit. Les travaux de S. Méléard ont grandement contribué à la partie "champ-moyen" de cette thèse. Mes remerciements s'adressent également aux membres de mon jury : Eva Löcherbach, Olivier Faugeras et Benoit Perthame dont les discussions et/ou publications ont façonné ce manuscrit.

Trois personnes ont plus directement affecté ce manuscrit : María J. Cáceres, Marie Doumic et Thomas Laloë. Je souhaite que mes collaborations futures soient aussi enrichissantes personnellement et professionnellement.

Mes remerciements vont également aux membres du Laboratoire Jean-Alexandre Dieudonné et en particulier à mes collègues d'enseignement : Camillo, Cédric, Christine, Mathieu, Mélisande, Patricia, Pierre et Yannick. De plus, je tiens à noter l'aide que m'a apportée Hélène Politano, ma voisine de bureau : ainsi, je n'étais pas le seul à disputer, de manière infructueuse, mon ordinateur.

Il m'a été donné la chance de côtoyer des doctorants formidables au LJAD et ailleurs. Les soirées sur la plage, au cinéma ou plus modestement devant le TBI ont profondément marqué ces trois années. J'ai une pensée spéciale pour les derniers amateurs du RU, qu'ils soient du LJAD ou bien du LPMC.

Je tiens également à remercier la team PES : BZ, Joon, Pcorre et La Pétrides. Chaque année, une semaine passée hors du temps à peaufiner notre style (même si c'est de moins en moins vrai...).

Il y a ceux qui étaient, sont et seront toujours là. Mes parents m'ont soutenu tout au long de ce parcours en dents de scie, que ce soit sur le plan personnel ou professionnel. Tout ce qu'ils ont fait depuis toujours m'a façonné tel que je suis et je les en remercie. Mon frère a également toujours eu une oreille attentive (si ce n'est les deux). Lui et sa petite famille sont assurés de pouvoir compter sur moi à la hauteur de tout ce qu'ils ont fait pour moi. Les cheveux de blé et le sourire d'ange de Giulia sont une constante source d'inspiration.

Je n'oublie pas ma nouvelle famille d'adoption : Mary Alin, Antoine et Baptiste. Une moitié de moi les admire et l'autre les adore.

Je ne pouvais que finir ces remerciements par ma moitié. Nous nous connaissons depuis 3 ans (merci Patricia) et au fond de moi j'ai l'impression que c'est depuis toujours. Tu m'as permis de suivre tes pas de doctorante et d'amante. Que ce soit pour déjeuner, pour discuter, ou bien pour une pause café, je t'ai suivie, toi et ton sourire que tu me destines et qui m'enivre. Toi, qui m'as guidé sur le long périple de la thèse, je te suivrai à présent et à jamais.

Contents

Notations

n	number of neurons.	
$\mathbb{R}$, $\mathbb{Z}$, $\mathbb{N}$	the sets of real numbers, of integers and of natural numbers.	
$\mathcal{B}(\mathbb{R})$	the Borelian sigma-algebra on $\mathbb{R}$.	
$\mathbb{1}_A$	the indicator function of the set A.	
δ_x	the Dirac mass concentrated at x.	
$\mathcal{D}$, $\mathcal{C}$, $\mathcal{C}^k$	the spaces of càdlàg (right continuous with left limits) functions, of continuous functions and of functions that are k times continuously differentiable.	
$\mathcal{W}_0^{k,\alpha}$	weighted Sobolev spaces.	p.134
$\|\|\cdot\|\|_E$	the natural norm associated with a normed space E.	
$\mathcal{M}_{c,b}$, $\mathcal{C}_{c,b}^{\infty}$	particular test function spaces.	p.44
$\mathcal{M}(E)$, $\mathcal{P}(E)$	the spaces of Radon measures and of probability measures on E.	
$\mathcal{BC}(\mathbb{R}_+, \mathcal{M}(E))$	the space of bounded continuous curves on $\mathcal{M}(E)$.	p.92
$\langle w, e \rangle$	the dual action of w in the dual of E over $e \in E$.	
Card (A)	the cardinal of a set A.	
$A \setminus B$	the relative complement of B in A, i.e. $\{x \in A, x \notin B\}$.	
$A \Delta B$	the symmetric difference between A and B, i.e. $A \setminus B \cup B \setminus A$.	

$(\Omega, \mathcal{F}, \mathbb{P})$	a probability space.	
$\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$	a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$.	
$\mathbb{E}[Y]$, $\mathbf{V}\mathrm{ar}(Y)$	the expectation and the variance of a real random variable Y.	
$\mathcal{N}(m, \sigma^2)$	the standard Gaussian distribution with mean m and variance σ^2.	
$< M >_t$	angle bracket associated with a square integrable martingale $(M_t)_{t \geq 0}$.	
$\ll M \gg_t$	Doob-Meyer process associated with a square integrable Hilbert space valued martingale $(M_t)_{t \geq 0}$.	p.143

Point processes

N	a point process on $\mathbb{R}$ equipped with the Borelians $\mathcal{B}(\mathbb{R})$.	
$N(A)$, $N(dt)$	the number of points in $N \cap A$ and the point measure associated with N.	p.39
N_-, N_+	the trace of N on the non-positive times, that is $N \cap (-\infty, 0]$ and the trace of N on the positive times, that is $N \cap (0, +\infty)$.	
ζ_{N_-}	the distribution of the trace N_-.	
$(S_t)_{t \geq 0}$	the standard age process associated with a point process N.	p.40
$(S_{t-})_{t \geq 0}$	the predictable age process associated with a point process N.	p.40
$(A_t^k)_{t \geq 0}$	the successive age processes associated with a point process N.	p.41
$(\lambda_t)_{t \geq 0}$	the stochastic intensity of a point process N.	p.39
Π	a Poisson process with intensity equal to one on $\mathbb{R}_+ \times \mathbb{R}_+$.	p.43

Statistics

$\mathscr{L}$	a subset of $\{1, \ldots, n\}$ corresponding to the neurons whose independence is tested.	p.169
M	sample size, that is the number of trials.	p.171
$\mathcal{H}_0$, $\mathcal{H}_1$	the null and alternative hypotheses.	p.171
$\mathbf{F}_i$	the different simulation frameworks used.	p.175

Chapitre

1 Introduction

Contents

Cette thèse s'inscrit dans l'étude théorique des processus ponctuels temporels tant d'un point de vue probabiliste que statistique. Ceux-ci permettent la modélisation des temps d'occurrence de certains phénomènes physiques ou biologiques (naissance, mort, stimulus, etc). Les travaux de cette thèse portent sur leur utilisation en neurosciences, et plus particulièrement sur la modélisation des trains de décharge caractérisant la transmission d'un signal au sein d'un réseau de neurones.

La première partie de ce manuscrit, correspondant aux chapitres 2, 3 et 4, vise à établir un lien rigoureux entre ce point de vue probabiliste microscopique, dans lequel les neurones individuels forment l'échelle de base, et les modèles macroscopiques, de nature déterministe, utilisés en analyse et attachés à la description collective du réseau, tel le système d'équations aux dérivées partielles introduit par Pakdaman, Perthame et Salort dans [114]. D'une part, les résultats du chapitre 2 portent sur des réseaux de neurones indépendants modélisés par des processus ponctuels très généraux. D'autre part, les chapitres 3 et 4 proposent

d'étudier des réseaux de neurones modélisés par des processus de Hawkes en interaction de type champ-moyen : un phénomène de propagation du chaos (loi des grands nombres) est mis en évidence dans le chapitre 3 et l'étude des fluctuations (théorème central limite associé) est menée dans le chapitre 4.

Dans une seconde partie, correspondant au chapitre 5 de ce manuscrit, nous étudions la problématique de la détection de synchronisations dans les activités d'un petit nombre de neurones. Dans la lignée de la méthode introduite par Tuleau-Malot et ses co-auteurs pour deux neurones [158], nous y répondons en développant une méthode statistique capable de tester l'indépendance entre plusieurs processus de Poisson.

Les deux parties décrites ci-dessus sont indépendantes mais s'inscrivent toutes les deux dans le cadre de la modélisation de neurones à l'aide de processus ponctuels.

1.1 Processus ponctuels temporels : modélisation des temps d'occurrence de phénomènes aléatoires

Les processus ponctuels temporels sont utilisés dans des domaines d'application très variés pour modéliser les temps d'occurrence de phénomènes purement aléatoires ou dont la dynamique déterministe sous-jacente n'est pas connue/observable. Citons par exemple le cas historique des processus de naissance et mort [37, Chapter 1], la modélisation des temps d'arrivée à une file d'attente [16] ou encore le temps d'entrée d'une particule radioactive dans un compteur Geiger [157]. Cette liste ne se veut en aucun cas exhaustive : d'autres champs d'application des processus ponctuels, et en particulier des processus de Hawkes, sont exposés dans la section 1.1.c). Toutefois, il est important de rappeler que le cadre d'application privilégié de ce manuscrit est la modélisation des temps d'occurrence de potentiels d'action pour des neurones en interaction. Ce dernier sera amplement développé dans la section 1.2.

1.1.a) Définitions et notations

Commençons par quelques notations relatives aux processus ponctuels temporels. Tout au long de cette thèse, nous nous intéressons à des *processus ponctuels simples et localement finis*[1], généralement notés N, sur $\mathbb{R}$ muni de sa tribu borélienne $\mathcal{B}(\mathbb{R})$. Autrement dit, N est un ensemble aléatoire de points de $\mathbb{R}$ tel que, pour toute partie mesurable bornée $A \subset \mathbb{R}$, le cardinal (aléatoire) de l'ensemble $N \cap A$ est fini presque sûrement (p.s.). Ainsi, N peut être associé à une suite de temps croissants que nous notons $(T_n)_{n\in\mathbb{Z}}$ et nous supposerons que $T_0 \leq 0 < T_1$. Pour tout A dans $\mathcal{B}(\mathbb{R})$, nous notons $N(A)$ le cardinal de $N \cap A$. De plus, nous notons $N(dt)$ la mesure ponctuelle associée à N, i.e. la mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour toute fonction mesurable positive f, $\int_{\mathbb{R}} f(t)N(dt) = \sum_{i\in\mathbb{Z}} f(T_i)$.

Dans ce manuscrit, nous nous intéressons plus particulièrement au comportement de N pour les temps strictement positifs : nous supposerons que la trace de N sur les temps négatifs est caractérisée par sa loi notée ζ_{N_-}. Ainsi, nous nous intéressons de manière équivalente au processus ponctuel N ou au *processus de comptage* $(N_t)_{t\geq 0}$ associé, défini par $N_t := N(]0, t])$. Nous utilisons un espace probabilisé filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ tel que la *filtration canonique* associée à N, i.e. $(\mathcal{F}_t^N)_{t\geq 0}$ définie par $\mathcal{F}_t^N := \sigma(N \cap]-\infty, t])$, est contenue dans la filtration $\mathbb{F} := (\mathcal{F}_t)_{t\geq 0}$ au sens où pour tout $t \geq 0$, $\mathcal{F}_t^N \subset \mathcal{F}_t$. Introduisons également les processus d'âge (standard et prévisible) associés à un processus ponctuel N :

[1] le qualificatif *localement fini* peut, de manière équivalente, être remplacé par *non explosif*.

- le *processus d'âge* associé à N, noté $(S_t)_{t\geq 0}$, est défini par

$$S_t := t - \sup\{T \in N,\, T \leq t\} = t - T_{N_t}, \quad \text{pour tout } t \geq 0. \tag{1.1}$$

- le *processus d'âge prévisible* associé à N, noté $(S_{t-})_{t\geq 0}$, est défini par

$$S_{t-} := t - \sup\{T \in N,\, T < t\} = t - T_{N_{t-}}, \quad \text{pour tout } t > 0, \tag{1.2}$$

  et étendu par continuité en $t = 0$, où $N_{t-} = N(]0, t[)$. En particulier, $S_{0-} = S_0 = t - T_0$.

Remarquons que $(S_{t-})_{t\geq 0}$ est la version continue à gauche du processus d'âge standard.

Il est commun de simplifier l'étude d'un processus de comptage à celle de son intensité stochastique. Tout processus $\mathbb{F}$-prévisible $(\lambda_t)_{t\geq 0}$ tel que $(N_t - \int_0^t \lambda_s ds)_{t\geq 0}$ est une $\mathbb{F}$-martingale locale est appelé $\mathbb{F}$-*intensité*[2] *(prévisible)* de N (sur $]0, +\infty[$). De manière informelle, $\lambda_t dt$ représente la probabilité que le processus N admette un nouveau point dans l'intervalle $[t, t + dt]$ sachant $\mathcal{F}_{t-}$, et ce conditionnement permet en particulier de modéliser des interactions (attirance, répulsion) entre les différents points du processus. Sous certaines hypothèses qui seront vérifiées tout au long de cette thèse, un tel processus existe, est essentiellement unique et caractérise la loi du processus de comptage $(N_t)_{t\geq 0}$ (voir [16]). En particulier, sous ces hypothèses, la probabilité qu'un temps quelconque $t_0 > 0$ fixé appartienne au processus ponctuel N est nulle. Finalement, il est important de noter les faits suivants concernant les processus d'âge qui sont centraux dans cette thèse :

- les deux processus d'âge (adapté et prévisible) sont égaux pour tout $t \geq 0$ sauf si t est un point de N (ces derniers forment p.s. un ensemble de mesure nulle dans $\mathbb{R}_+$),
- pour tout $t \geq 0$, $S_{t-} = S_t$ p.s. (car N admet une intensité),
- la valeur $S_{0-} = S_0$ est déterminée par $N_- := N \cap \mathbb{R}_-$ et elle est finie s'il y a au moins un point dans ce dernier.

1.1.b) Exemples de processus ponctuels temporels

Nous donnons ici quelques familles de processus ponctuels en donnant la forme de leur intensité stochastique qui, nous le rappelons, caractérise la loi d'un processus ponctuel.

- *Processus de Poisson* : un processus de Poisson se caractérise par une intensité stochastique qui est une fonction déterministe de la variable temporelle t. Ce modèle possède des propriétés très intéressantes d'un point de vue mathématique, notamment l'indépendance entre les points du processus : le taux d'apparition d'un point n'est nullement influencé par les autres points. En particulier, pour peu que l'intensité soit non nulle, il est possible de trouver deux points arbitrairement proches.
- *Processus de renouvellement* : comme son nom l'indique, la dynamique est renouvelée après l'apparition d'un nouveau point du processus. Son intensité au temps t dépend du délai depuis l'apparition du dernier point (strictement avant le temps t), i.e. S_{t-}. Ce dernier étant aléatoire, l'intensité est bien stochastique ici. Notons que le

[2] L'intensité peut, de manière équivalente, se définir comme la projection prévisible de N. C'est à dire le processus $\mathbb{F}$-prévisible $(\lambda_t)_{t\geq 0}$ tel que pour tout processus positif et $\mathbb{F}$-prévisible $(C_t)_{t\geq 0}$, $\mathbb{E}[\int_0^{+\infty} C_t N(dt)] = \mathbb{E}[\int_0^{+\infty} C_t \lambda_t dt]$. Notons que cette dernière est la définition adoptée dans [16].

renouvellement du processus peut également se décrire par des délais inter-points (les variables $T_{i+1} - T_i$) qui forment une famille de variables aléatoires indépendantes et identiquement distribuées (i.i.d.). L'équivalence entre ces deux descriptions est abordée dans la section 2.3.b).

- *Processus de Wold* et généralisations : ces derniers peuvent se voir comme la généralisation des processus de renouvellement. Dans le cas du processus de Wold, en plus de dépendre de S_{t-}, l'intensité dépend du délai inter-points précédent, i.e. $T_{N_{t-}} - T_{N_{t-}-1}$. Plus généralement, on pourrait penser à un modèle dont l'intensité dépend de S_{t-} ainsi que des k délais inter-points précédents, où k est un paramètre entier du modèle.

- *Processus de Hawkes* : ce processus permet de modéliser des phénomènes d'auto-interaction à longue mémoire. Dans le cas général (non linéaire), son intensité est de la forme
$$\lambda_t = \Phi\left(\int_0^{t-} h(t-z)N(dz)\right), \tag{1.3}$$
où $\Phi : \mathbb{R} \to \mathbb{R}_+$ et $h : \mathbb{R}_+ \to \mathbb{R}$ sont respectivement appelées *fonction intensité* et *fonction d'auto-interaction*. Pour des raisons de stabilité, la fonction Φ est en général supposée être lipschitzienne. Notons que dans sa variante linéaire, c'est-à-dire quand Φ est une fonction affine et h est une fonction positive, il existe une représentation du processus de Hawkes par un processus de branchement : chaque point donne "naissance" à des points dans le futur selon un processus de Poisson d'intensité $h(t)$ (voir [71, 138] ou bien la section 2.8.c)).

Pour de plus amples informations sur les processus ponctuels (exemples, applications, résultats généraux, etc), nous renvoyons vers les deux tomes de Daley et Vere-Jones [37, 38]. Les exemples de processus ponctuels définis ci-dessus sont plus intensivement décrits dans le chapitre 2. Notons tout de même que les processus de Poisson et de Hawkes sont centraux dans ce manuscrit et se retrouvent tout au long des chapitres. De plus, la présentation de la classe des processus de Hawkes, ainsi que leurs généralisations, fait l'objet de la section suivante.

1.1.c) Processus de Hawkes

Le processus de Hawkes a été introduit par Hawkes [70] dans sa forme linéaire, c'est-à-dire (en comparaison avec (1.3))
$$\lambda_t = \mu + \int_0^{t-} h(t-z)N(dz), \tag{1.4}$$
pour modéliser le phénomène de réplique sismique. Le paramètre réel positif μ est appelé *taux d'apparition spontanée* et représente l'intensité du processus en l'absence de points. Le processus de Hawkes, et plus globalement toutes ses généralisations, possède un champ d'application très diversifié. De manière non exhaustive, ces processus sont utilisés pour modéliser :

- les temps d'occurrence de séismes en sismologie pour prendre en compte le phénomène de réplique, [85] ou [111] (modèle ETAS) ;

- les localisations d'un gêne sur un brin d'ADN en génomique [67, 139] ;

- les temps d'occurrence de vols en criminologie [104] ;
- les processus d'achat et vente en finance [7, 8] ;
- les temps d'occurrence de potentiel d'action en neurosciences [31, 68, 123, 136]. Nous rappelons que c'est ce dernier usage des processus de Hawkes qui est privilégié dans ce manuscrit.

On utilise très souvent la version multivariée des processus de Hawkes pour modéliser des phénomènes d'interactions. Un processus de Hawkes multivarié est un processus ponctuel multivarié, c'est-à-dire une famille finie de processus ponctuels possiblement dépendants, noté $(N^1, \dots, N^n)$ dont les intensités respectives pour $i = 1, \dots, n$ sont données par

$$\lambda_t^i = \Phi_i \left(\sum_{j=1}^{n} \int_0^{t-} h_{j \to i}(t - z) N^j(dz) \right), \tag{1.5}$$

où $\Phi_i : \mathbb{R} \to \mathbb{R}_+$ est la fonction intensité associée à la coordonnée i et $h_{j \to i}$ est la *fonction d'interaction* entre les coordonnées j et i décrivant l'influence de chaque point de N^j dans le passé sur le taux d'apparition d'un nouveau point de N^i, via la valeur de l'intensité λ^i. Dans la veine des généralisations des processus de Hawkes, signalons les systèmes de processus en interaction à mémoire variable étudiés dans [54] qui peuvent s'apparenter à la forme suivante d'intensité, en comparaison avec (1.5),

$$\lambda_t^i = \Phi_i \left(\sum_{j=1}^{n} \int_{t - S_{t-}^i}^{t-} h_{j \to i}(t - z) N^j(dz) \right),$$

où $(S_{t-}^i)_{t \geq 0}$ est le processus d'âge prévisible associé à N^i. Ici, la propriété de longue mémoire des processus de Hawkes est altérée : seule la dynamique depuis le dernier point du processus est intégrée.

Les propriétés théoriques de la classe des processus de Hawkes ont été largement étudiées ces dernières années :

- structure de branchement du processus de Hawkes linéaire (1.4) par Hawkes et Oakes [71] (voir également la section 2.8.c)) ;
- stationnarité et stabilité [17] ;
- asymptotique en temps long [80, 163] ;
- estimation des fonctions d'interaction [7, 68, 139] ;
- algorithmes de simulation [105, 106, 110] ;
- calcul des cumulants du processus de Hawkes linéaire [83].

Notons également le récent intérêt porté à l'approximation des processus de Hawkes multivariés dans un régime d'interaction de type champ-moyen [43, 44]. C'est dans cette lignée que s'inscrivent les chapitres 3 et 4.

1.1.d) Méthode de thinning

Bien que caractérisant la loi d'un processus ponctuel, son intensité stochastique ne donne pas directement d'informations sur la dynamique trajectorielle d'un processus ponctuel. La *méthode de thinning* permet en particulier de faire un lien entre la trajectoire d'un processus ponctuel et la trajectoire correspondante de son intensité. Le lien donné par cette approche est utilisé tout au long de cette thèse. Nous choisissons donc de donner ici quelques informations à propos de la méthode de thinning.

Les origines de cette méthode remontent à un article de Lewis et Shedler en 1978 [92] dans lequel est proposée une nouvelle technique de simulation de processus de Poisson inhomogène. Cette procédure de simulation a rapidement été généralisée par Ogata en 1981 [110] pour des processus ponctuels très généraux. Mais, en plus de fournir des algorithmes de simulation de processus ponctuels, l'approche par thinning permet, sur un plan plus théorique, de donner une représentation (trajectorielle) d'un processus ponctuel général en dimension 1 en fonction d'un processus de Poisson homogène en dimension 2. En effet, supposons que l'on s'intéresse au processus N d'intensité $(\lambda_t)_{t\geq 0}$. De manière informelle, il existe un processus de Poisson d'intensité 1 sur $\mathbb{R}_+ \times \mathbb{R}_+$ noté Π tel que presque sûrement, pour tout $t' \geq 0$, $N_{t'}$ est le nombre de points de Π dans la bande B sous le graphe de l'intensité $(\lambda_t)_{t\geq 0}$, c'est-à-dire,

$$N_{t'} = \mathrm{Card}\,(\Pi \cap B) \quad \text{avec } B = \{(t,x) \in [0,t'] \times \mathbb{R}_+,\ x \leq \lambda_t\}.$$

Cette représentation permet donc, dans une certaine mesure, de se ramener à l'étude beaucoup plus simple d'un processus de Poisson d'intensité 1.

La méthode de thinning est principalement utilisée dans ce manuscrit pour :

- décrire la dynamique du processus d'âge au moyen d'une équation aux dérivées partielles stochastique dirigée par un bruit poissonien (2.12)-(2.14),
- et pour fournir un couplage adéquat entre processus ponctuels, voir par exemple le couplage (3.38)-(3.39).

Pour finir, précisons tout de même que dans un but pédagogique, la description de la méthode du thinning ci-dessus utilise en fait le sens direct ainsi que sa réciproque. Dans un but de clarté, nous choisisons de ne pas les expliciter ici. L'énoncé du sens direct et sa preuve se trouvent dans la section 2.8.d). L'énoncé de la réciproque et sa preuve, accompagnés de compléments historiques, se trouvent dans l'annexe A.1. Nous fournissons dans ce manuscrit des preuves de ces deux résultats classiques compte tenu de la difficulté à en trouver des preuves détaillées dans la littérature.

1.1.e) Quid des processus ponctuels spatiaux ?

Finissons cette section en précisant que la famille des processus ponctuels est beaucoup plus large que les seuls processus ponctuels temporels. Les processus ponctuels spatiaux sont largement étudiés et peuvent modéliser la position d'arbres en sylviculture (en dimension 2), la position des agrégats et poches d'air dans un bloc de béton ou encore la position de neurones dans le cortex [131] (en dimension 3). Nous renvoyons vers les références [75] et [107] pour plus de détails sur les processus ponctuels spatiaux. L'exemple le plus simple est la généralisation du processus de Poisson aux dimensions supérieures[3]. L'absence

[3]Notons ici que nous utiliserons dans ce manuscrit un processus ponctuel spatial, à savoir le processus de Poisson d'intensité 1 en dimension 2, pour le thinning.

d'interaction (spatiale cette fois-ci) entre les points rend bien entendu son usage limité dans le cadre de l'analyse de configurations de points qui, la plupart du temps, exhibent des phénomènes d'attraction ou de répulsion entre les points. Parmi la pléthore de modèles de processus ponctuels spatiaux, notons la présence des processus de Gibbs [107] et processus déterminantaux [90] permettant de prendre en compte les dépendances entre points qui connaissent un essor récent.

1.2 Cadre d'application : modélisation de réseaux de neurones

La compréhension du fonctionnement du cerveau passe par l'analyse du réseau de neurones formé par celui-ci. Ce manuscrit se focalise sur sa modélisation par le biais de processus ponctuels temporels à l'échelle (microscopique) des neurones et de systèmes d'équations aux dérivées partielles à une échelle plus macroscopique.

1.2.a) Cadre neurobiologique

Au sein d'un réseau de neurones, l'information est transmise par l'intermédiaire de signaux électriques et chimiques émis et reçus par les neurones. Les potentiels d'actions constituent le substrat électrique de ce transport d'information. Un *potentiel d'action* correspond à une dépolarisation brève et stéréotypée (un même neurone génère les mêmes impulsions au cours du temps) du potentiel de la membrane plasmique d'un neurone en réponse à un stimulus, selon une loi du tout ou rien.

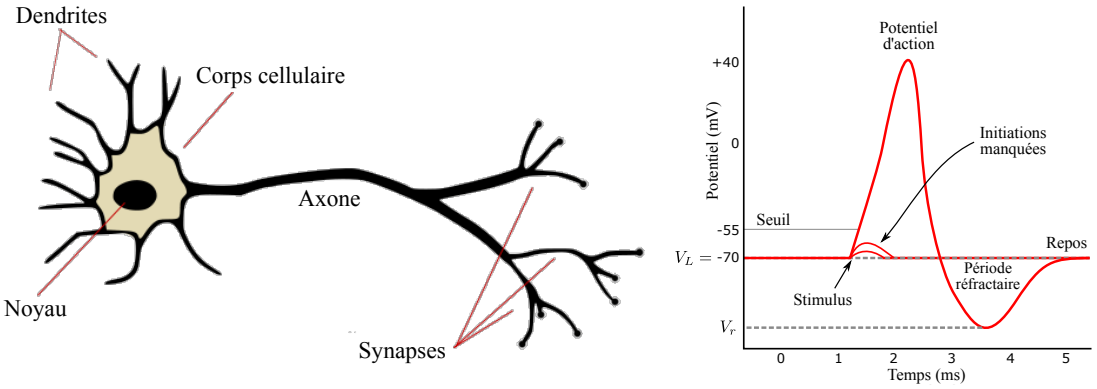


FIGURE 1.1 : Gauche : schéma classique d'un neurone. Droite : potentiel d'action.

La genèse d'un potentiel d'action a lieu au niveau du corps cellulaire du neurone. Il se manifeste ensuite sous la forme d'un signal électrique unidirectionnel transmis par l'axone, et on dit alors que le neurone *décharge*. Ce signal électrique se propage jusqu'aux synapses (voir la figure 1.1) où il provoque la libération de neurotransmetteurs. Ces neurotransmetteurs se fixent sur les dendrites des neurones voisins qualifiés de post-synaptiques. La fixation des neurotransmetteurs se traduit par la création d'un signal électrique au niveau du corps cellulaire du neurone post-synaptique concerné. En changeant de point de vue, un neurone (post-synaptique) reçoit, via ses dendrites et de manière cumulative, les informations émanant de plusieurs neurones qualifiés de pré-synaptiques. Si celui-ci est suffisamment sollicité par ce phénomène d'*intégration synaptique*, alors il émet à son tour un potentiel d'action. Ainsi se déroule le transport de l'information pour une relation

"père-fils" - au sens où il y a une connexion dirigée du père vers le fils. Cette opération se répétant de père en fils, l'information peut ainsi se transmettre. Notons qu'un même neurone peut être qualifié de pré-synaptique (père) ou bien de post-synaptique (fils) selon la connexion étudiée. L'information nécessaire à la compréhension du réseau peut être réduite à la connaissance des séquences des temps d'occurrence de ces potentiels d'action également appelés *trains de spikes* (un *spike* correspondant au temps d'occurrence d'un potentiel d'action). Le nombre moyen de spikes par unité de temps est appelé *taux de décharge*.

Outre les interactions décrites dans le paragraphe précédent, il faut également noter la présence, à l'échelle d'un neurone individuel, du phénomène de période réfractaire. La *période réfractaire* d'un neurone est l'intervalle de temps qui suit immédiatement sa décharge et au cours duquel il est peu enclin à, voire incapable, d'émettre un nouveau potentiel d'action (voir la figure 1.1).

Dans la suite, nous nous attacherons à proposer des modèles en accord avec les deux contraintes neuro-physiologiques que sont l'intégration synaptique et la période réfractaire.

1.2.b) Modélisation mathématique : plusieurs échelles possibles

Les structures neuronales présentent naturellement plusieurs échelles d'étude où dynamiques déterministes et stochastiques s'entremêlent. Par exemple, nous avons, de manière non exhaustive et selon une échelle croissante :

- les canaux ioniques, qui sont des portes perméables permettant le passage de particules chargées à travers la membrane d'un neurone. Leurs perméabilités aux différents ions (sodium, potassium, etc) varient au cours du temps, et cette dynamique est en fait responsable de la création d'un potentiel d'action [129]. Les canaux ioniques sont communément modélisés par des processus stochastiques [151].

- les neurones, qui communiquent grâce à l'émission de potentiels d'actions. En considérant qu'un neurone admet une infinité de canaux ioniques, la dynamique du potentiel de membrane peut être décrite par des systèmes dynamiques déterministes tels que le *système de Hodgkin-Huxley* [74] par exemple. Il est commun d'introduire du bruit stochastique dans ces systèmes dynamiques pour prendre en compte la dynamique (non observable) des canaux ioniques. En effet, le nombre de canaux ioniques étant fini, leur perméabilité moyenne fluctue (aléatoirement) autour de sa valeur limite (atteinte quand le nombre de canaux est supposé infini).

- les régions corticales décrites par le *local field potential* (potentiel de champ local), c'est-à-dire la somme des signaux électriques émanant de neurones localisés dans une petite zone de tissu nerveux. À cette échelle, la dynamique peut être décrite par des équations aux dérivées partielles où, une fois de plus, du bruit stochastique peut être introduit afin d'inclure l'aléa des neurones contribuant au *local field potential*. Là encore, le nombre de neurones étant fini, leur activité moyenne fluctue (aléatoirement) autour de sa valeur limite (atteinte quand le nombre de neurones est supposé infini).

À l'heure actuelle, la dynamique exacte du potentiel de membrane ne peut pas être mesurée de manière précise pour plus d'une poignée de neurones en simultané. Mais, comme nous l'avons déjà mentionné, la connaissance des trains de spikes (plusieurs dizaines de trains peuvent être enregistrés simultanément) est suffisante pour étudier un réseau de neurones et plus particulièrement le transport d'informations au sein de ce réseau. Dans

ce cadre là, les neurones individuels peuvent être modélisés par des processus ponctuels temporels qui décrivent exactement les trains de spikes correspondants et c'est ce que l'on fait dans cette thèse.

1.2.c) Deux échelles en particulier

Dans cette section, les deux échelles suivantes sont décrites : l'échelle microscopique des neurones modélisés par des processus ponctuels temporels (voir la section 1.1.a) pour les définitions et notations à ce sujet) et l'échelle macroscopique du réseau de neurones tout entier dont la dynamique est modélisée par un système d'équations aux dérivées partielles (EDP) structuré en âge.

Échelle microscopique. Nous faisons ici l'inventaire des exemples de processus ponctuels temporels présentés dans la section 1.1.b) mais du point de vue de la modélisation de neurones.

Le modèle le plus simple rencontré dans la littérature est le processus de Poisson homogène qui suppose un taux de décharge constant (voir le chapitre 5 ou [40, 66, 158]). Son intensité est donnée par $\lambda_t = \lambda$ constante. Cependant, de tels processus présentent l'inconvénient de ne pas reproduire le phénomène de période réfractaire. De plus, sur certains jeux de données réelles, l'hypothèse d'adéquation au modèle poissonien est rejetée [136].

En revanche, le processus de renouvellement, pour lequel le taux de décharge ne dépend que du dernier spike dans le passé, i.e. $\lambda_t = f(S_{t-})$ pour f une fonction mesurable donnée, permet, dans un cadre assez simple, d'imiter la période réfractaire. En effet, pour les processus de renouvellement, la loi du délai entre deux points consécutifs (ou *inter-spike interval* (ISI) dans notre cadre) est une donnée du modèle. Il suffit alors de prendre une loi qui ne charge pas les petits délais pour modéliser une période réfractaire. Dans le même esprit, le processus de Wold [89], un peu plus compliqué et pour lequel le taux de décharge ne dépend que des deux derniers spikes dans le passé, a montré de bons résultats d'un point de vue statistique [127].

Les trois modèles présentés ci-dessus ne permettent pas de prendre en compte les interactions entre neurones. Il existe plusieurs démarches visant à modéliser l'intégration synaptique dans le cadre de processus ponctuels. Parmi celles-ci, mentionnons la classe des processus de Hawkes abondamment utilisée dans le cadre de la modélisation de neurones [31, 88, 119, 136]. Rappelons ici la définition d'un processus de Hawkes multivarié. C'est un processus ponctuel multivarié $(N^1, \dots, N^n)$ dont les intensités respectives pour $i = 1, \dots, n$ sont données par

$$\lambda_t^i = \Phi_i \left(\sum_{j=1}^{n} \int_0^{t-} h_{j \to i}(t - z) N^j(dz) \right).$$

Ici, le taux de décharge dépend de tous les spikes de tous les neurones avant le temps t. La fonction $h_{j \to i}$ décrit l'influence de l'excitation du neurone j sur le potentiel de membrane, et donc sur le taux de décharge, du neurone i. La norme L^1 de $h_{j \to i}$, pour $j \neq i$, peut être rattachée au *poids synaptique* du neurone j sur le neurone i, c'est-à-dire la force de l'influence du neurone j sur le neurone i au travers de leur connexion synaptique.

Notons également l'utilisation de modèles linéaires généralisés dans le cadre de la modélisation de neurones [123]. Ceux-ci s'apparentent le plus souvent à des versions discrétisées d'un processus de Hawkes avec Φ égal à la fonction exponentielle.

Échelle macroscopique. À un niveau macroscopique, nous nous concentrons sur le point de vue proposé par Pakdaman, Perthame et Salort dans une série d'articles [114, 115, 116]. Ils proposent le système d'équations aux dérivées partielles (non-linéaires) structuré en âge suivant pour décrire la dynamique des spikes des neurones de tout le réseau :

$$\begin{cases} \dfrac{\partial n\,(t,s)}{\partial t} + \dfrac{\partial n\,(t,s)}{\partial s} + p\,(s, X\,(t))\,n\,(t,s) = 0, \\ m\,(t) := n\,(t,0) = \displaystyle\int_0^{+\infty} p\,(s, X\,(t))\,n\,(t,s)\,ds. \end{cases} \tag{PPS}$$

Ici, $n(t,s)$ représente la densité de probabilité de trouver un neurone (typique du réseau) d'âge s au temps t, l'âge d'un neurone étant le délai depuis sa dernière décharge. Bien entendu, la définition de l'âge d'un neurone correspond à la définition du processus d'âge associé à un processus ponctuel (voir l'équation (1.1)) dès l'instant où ce dernier modélise le train de spikes dudit neurone. La fonction p représente le taux de décharge d'un neurone qui dépend donc de son âge s. Notons que cette dépendance permet de modéliser le phénomène de période réfractaire en prenant, par exemple, $p(s,x) = \mathbb{1}_{s\geq\delta}$ pour un certain paramètre δ représentant la durée de la période réfractaire. De plus, la fonction p dépend également de la variable $X(t)$ représentant l'activité globale du réseau au temps t, sous la forme

$$X(t) := \int_0^t d(t-z)n(z,0)dz, \tag{1.6}$$

pour d une certaine fonction de délai. Cette intégrale de convolution modélise le temps de propagation de la décharge d'un neurone au reste du réseau.

Le système (PPS) décrit une dynamique de croissance/réinitialisation de la manière suivante : l'âge s d'un neurone au temps t peut

- augmenter linéairement avec vitesse 1 (si le neurone ne décharge pas au temps t),
- ou bien être réinitialisé en $s = 0$ (si le neurone décharge au temps t).

Bien entendu, cette dynamique conserve la masse dans le sens où pour tout $t \geq 0$, $\int_0^{+\infty} n(t,s)ds = \int_0^{+\infty} n(0,s)ds$.

La présence d'oscillations ainsi que la relaxation à l'équilibre ont été étudiées pour ce système [114, 115, 116]. Bien que présentant des caractéristiques qualitatives intéressantes du point de vue de la modélisation de réseaux de neurones, ce système a été justifié par une approche heuristique se basant sur l'étude du modèle *Intègre-et-décharge* effectuée dans [122]. Nous choisissons donc de donner ici quelques détails concernant le modèle intègre-et-décharge. Dans sa forme la plus simple, ce modèle décrit la dynamique du potentiel de membrane d'un neurone, noté $V(t)$, au moyen de l'équation différentielle ordinaire,

$$C_m \frac{dV}{dt} = -g_L(V - V_L) + I(t). \tag{1.7}$$

Ici, C_m est la capacité (de charge) de la membrane, g_L est la conductance de fuite (de l'intérieur du neurone vers la matrice extérieure) et V_L est le potentiel de repos (atteint pour l'état stationnaire avec $I(t) = 0$). Le *courant synaptique* $I(t)$ décrit l'influence des décharges des neurones pré-synaptiques sur le neurone dont le potentiel de membrane est modélisé par $V(t)$. Pour finir, la dynamique de décharge suivante est ajoutée : si $V(t)$ dépasse un certain seuil θ, le neurone décharge et $V(t)$ est réinitialisé en la valeur de

potentiel V_r. La différence de valeur entre le potentiel de repos V_L et le potentiel de réinitialisation V_r permet de modéliser le phénomène de période réfractaire : pour cela, on suppose que $V_r < V_L < \theta$ (voir la figure 1.1). Dans le cas particulier étudié dans [122], la solution explicite du modèle est donnée par

$$V(t) = V_L + (V_r - V_L)e^{-(t-T)/\tau_m} + \int_T^t d(t-z)N_{\text{input}}(dz). \tag{1.8}$$

Ici, T représente le dernier spike du neurone, $\tau_m = C_m/g_L$ est la constante de temps du système, la fonction d représente le potentiel excitateur post-synaptique (EPSP en anglais) et N_{input} est la somme des masses de Dirac localisées sur les spikes des neurones pré-synaptiques. Ainsi, notons que le terme intégral dans (1.8) modélise le phénomène d'intégration synaptique et en ce sens peut être rapproché de $X(t)$.

Signalons que la famille des modèles de type intègre-et-décharge est vaste et ne se résume pas au cadre décrit ci-dessus. En particulier, comme noté ci-dessous dans la section 1.3.d), des approches de type champ-moyen sont possibles pour des neurones de type intègre-et-décharge en interaction. Dans l'optique de l'utilisation de tels modèles à des données expérimentales, notons que ces dernières exhibent en général des constantes de temps qui varient au cours de l'expérience, et surtout un seuil gradué [79]. Or, un seuil (strict) de décharge est considéré dans le cadre de base du modèle intègre-et-décharge.

1.3 Première partie : lien entre deux échelles

La première partie de cette thèse a pour but de faire le lien entre, d'une part, certains types de processus ponctuels et, d'autre part, le système non-linéaire (PPS) en partant de la constatation suivante : la mesure $N(dt)$ représente l'activité (stochastique) d'un neurone selon la loi du tout ou rien et $n(t,0)$ représente l'activité moyenne d'un neurone typique. Ainsi, le processus de Hawkes (1.3) et le système (PPS) font tous les deux intervenir une intégrale de convolution entre une fonction de délai, h et d respectivement, et une représentation de l'activité instantanée des neurones, $N(dt)$ et $n(t,0)$ respectivement. Cette ressemblance est également visible avec le terme intégral apparaissant dans (1.8). Nous montrerons que cette ressemblance obéit en réalité à une correspondance mathématique, que nous expliciterons dès que possible dans cette section.

1.3.a) Lois probabilistes et équations aux dérivées partielles

Dans de nombreux domaines d'application, les équations aux dérivées partielles (EDP) décrivent la dynamique globale/moyenne (observable le plus souvent) associée à une dynamique plus individuelle souvent non observable compte tenu du nombre d'individus mis en jeu ou bien de leur taille microscopique par exemple. Le lien historique entre EDP et modélisation probabiliste (microscopique) remonte à l'intuition de Einstein (1905) qui consiste à voir le mouvement brownien (observé par Brown sur des grains de pollen) comme la dynamique microscopique sous-jacente à la diffusion modélisée par l'équation de la chaleur. Plus formellement, ce lien s'écrit de la manière suivante : soit $(B_t)_{t\geq 0}$ un mouvement brownien et $u^{\text{in}} : \mathbb{R} \to \mathbb{R}$ une fonction bornée. Alors, la fonction $u : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ définie par $u(t,x) := \mathbb{E}[u^{in}(x + B_t)]$ est solution de l'équation de la chaleur :

$$\frac{\partial}{\partial t}u(t,x) = \frac{1}{2}\Delta u(t,x),$$

avec pour condition initiale $u(0,x) = u^{\text{in}}(x)$. Ainsi, la "diffusion moyenne" d'une particule typique dans le milieu est décrite par l'équation de la chaleur. Ce résultat provient de la propriété de Markov du mouvement brownien. Plus généralement, il existe un lien étroit entre processus markoviens et équations aux dérivées partielles donné par l'équation de Fokker-Planck.

Entre les processus de diffusion (plus réguliers) et les processus de comptage (processus de saut pur) que nous étudions dans ce manuscrit, notons la présence des processus markoviens déterministes par morceaux (PDMP en anglais) introduits par Davis [39]. Dans ce cas, la dynamique d'une particule est dirigée par un flot déterministe jusqu'à l'apparition d'un "saut". Le taux d'apparition de ce "saut" ne dépend que de la position de la particule. Un "saut" peut correspondre à la fois à un saut de la position de la particule mais également à une transition de la dynamique déterministe de la particule, c'est-à-dire que le flot directeur de la dynamique peut être altéré par l'apparition d'un "saut". Dans le cadre de l'étude de cette thèse, précisons que des liens entre des modèles PDMP et des systèmes d'EDP ou d'équations différentielles ordinaires ont été étudiés récemment : communautés proies-prédateurs et le système de Lotka-Volterra [34], système de Hodgkin-Huxley pour la modélisation du potentiel de membrane d'un neurone [140].

À notre connaissance, la littérature traitant du lien entre processus ponctuels généraux et EDP est faible voire inexistante. La raison principale étant que les techniques usuelles utilisent l'hypothèse de Markov, et celle-ci n'est en général pas vérifiée par un processus ponctuel.

1.3.b) Approche par espérance mathématique

L'idée générale de la première approche développée dans le chapitre 2 réside sur le fait suivant :

> Une quantité déterministe qui décrit simplement le comportement moyen d'un variable aléatoire X est son espérance $\mathbb{E}[X]$.

Ainsi, cette première approche consiste à voir la quantité $n(t,s)$ qui apparait dans le système (PPS) comme l'espérance de son pendant aléatoire. Nous supposerons que la dynamique stochastique sous-jacente est caractérisée par un processus ponctuel N, d'intensité $(\lambda_t)_{t\geq 0}$, qui modélise le train de spikes d'un neurone typique du réseau.

Contributions de la thèse. Dans le chapitre 2 de ce manuscrit, nous étudions la loi de probabilité de l'âge S_t (ou de manière équivalente de S_{t-} puisque ces deux variables sont égales presque sûrement, comme nous l'avons vu dans la section 1.1.a)) associé à un processus ponctuel N admettant une intensité $(\lambda_t)_{t\geq 0}$ très générale. Plus précisément, nous cherchons à caractériser sa dynamique, en fonction du temps t, au moyen d'une EDP qui se rapproche de (PPS).

Nous construisons dans un premier temps, de manière ad hoc, une mesure (aléatoire) en temps t et en âge s, notée $U(dt,ds)$, qui se veut être l'analogue aléatoire de $n(t,s)$. Celle-ci est construite à partir de ses marginales en temps et en âge et vérifie, en particulier, pour tout $t>0$, $U(t,ds) = \delta_{S_{t-}}(ds)$ où $U(t,ds)$ est la marginale au temps t de la mesure $U(dt,ds)$. Cette construction est possible grâce à une propriété de Fubini vérifiée par les marginales en temps et en âge (voir la section 2.4.a)). La procédure de thinning, présentée dans la section 1.1.d) (avec quelques compléments en annexe A.1), permet d'écrire un

système d'équations aux dérivées partielles stochastiques vérifié, au sens faible[4], par la mesure U (voir la proposition 2.4.1).

De plus, sous certaines hypothèses d'intégrabilité sur l'intensité $(\lambda_t)_{t\geq 0}$, l'espérance de la mesure U est bien définie et nous la notons u. En particulier, pour tout $t > 0$, $u(t, ds) = \mathbb{E}[\delta_{S_{t-}}(ds)]$ c'est-à-dire que $u(t, ds)$ est la loi de probabilité de l'âge au temps t.

Le théorème 2.4.4 montre que cette mesure u satisfait, au sens faible, le système d'équations aux dérivées partielles (déterministe) suivant :

$$\begin{cases} \dfrac{\partial}{\partial t}u\left(dt, ds\right) + \dfrac{\partial}{\partial s}u\left(dt, ds\right) + \rho_{\lambda,\zeta_{N_-}}(t, s)u\left(dt, ds\right) = 0, \\ u\left(dt, 0\right) = \displaystyle\int_0^{+\infty} \rho_{\lambda,\zeta_{N_-}}(t, s)u\left(t, ds\right)\, dt, \end{cases} \tag{1.9}$$

où $\rho_{\lambda,\zeta_{N_-}}(t, s) := \mathbb{E}[\lambda_t | S_{t-} = s]$. Notons que $\rho_{\lambda,\zeta_{N_-}}$ dépend de l'intensité du processus ponctuel sous-jacent ainsi que de la condition initiale ζ_{N_-} (à savoir, la loi de $N \cap \mathbb{R}_-$). En comparant les systèmes (PPS) et (1.9), il apparaît que $n(t, s)$ est remplacé par $u(dt, ds)$ et que le terme $p(s, X(t))$ est remplacé par $\rho_{\lambda,\zeta_{N_-}}(t, s)$. Le premier remplacement découle de la construction de u. Le second remplacement par l'espérance conditionnelle sachant l'âge S_{t-} est expliqué par le fait que la mesure $u(t, \cdot)$ est la loi de S_{t-} : pour connaître le taux de décharge moyen d'un neurone d'âge s, on conditionne par l'évènement $S_{t-} = s$.

À notre connaissance, le coefficient $\rho_{\lambda,\zeta_{N_-}}(t, s)$ ne peut pas être explicité dans le cas général. Cependant, si l'intensité ne dépend que du temps t et du dernier point, i.e. $\lambda_t = f(t, S_{t-})$, alors, quelque soit ζ_{N_-}, $\rho_{\lambda,\zeta_{N_-}}(t, s) = f(t, s)$ et le système (1.9) est linéaire. Dans ce cas particulier, le processus d'âge prévisible est markovien et le résultat donné par le théorème 2.4.4 peut être obtenu plus simplement en utilisant le générateur infinitésimal de $(S_{t-})_{t\geq 0}$ et l'équation de Fokker-Planck (voir la section 2.5.a)). Ainsi, le théorème 2.4.4 peut être vu comme la généralisation de ce résultat à des dynamiques sous-jacentes non markoviennes, ce qui est le cas pour les processus de Hawkes.

Le coefficient $\rho_{\lambda,\zeta_{N_-}}(t, s)$ n'étant pas explicite dans le cas des processus de Hawkes (l'intensité dépend de tous les points du passé et pas simplement du dernier), une idée possible est d'approcher la dynamique associée à un processus de Hawkes par une dépendance par rapport aux k derniers spikes du neurone avec k suffisamment grand. Ceci amène à généraliser le système (1.9) qui porte sur la loi de l'âge, autrement dit du dernier spike du neurone, au système (2.27)-(2.29) qui donne la dynamique de la loi jointe des k derniers spikes.

Cependant, dans le cas particulier où le processus ponctuel sous-jacent est un processus de Hawkes linéaire (1.4), sa structure de branchement (voir la section 2.8.c)) permet de mieux appréhender l'espérance conditionnelle $\mathbb{E}[\lambda_t | S_{t-} = s]$. La conséquence pour le système (1.9) est la suivante (voir la proposition 2.5.4) :

> La fonction (de survie) v définie par $v(t, s) := \int_s^{+\infty} u(t, d\sigma)$ vérifie un système d'EDP fermé (ce qui n'est pas le cas pour le système (1.9)).

Notons en particulier que l'interprétation de $X(t)$, défini par l'équation (1.6), comme l'intégrale intervenant dans l'intensité d'un processus de Hawkes n'est pas vérifiée ici malgré la ressemblance mise en évidence à la fin de la section 1.2.c).

Finalement, il est possible de voir le système (1.9) comme le système vérifié (dans la limite d'une grande population de neurones) par la proportion de neurones d'âge s au temps

[4] La notion de solution au sens faible est précisée dans la proposition 2.4.1 pour ce cas particulier. Dans un cadre plus général, l'Appendice A.2 donne des précisions à ce sujet.

t dans le cas d'un réseau de n neurones modélisés par des processus ponctuels (très généraux) N^i $(i = 1, ..., n)$ indépendants et de même loi que N. Ce résultat est explicité dans le corollaire 2.4.5 et repose sur la loi des grands nombres pour les variables aléatoires réelles.

Cependant, cette hypothèse d'indépendance des neurones est trop forte (comme il apparaît clairement dans le chapitre 5). Nous nous sommes donc tournés vers une approche par limite de champ-moyen où des interactions entre les neurones sont considérées.

1.3.c) Généralités sur les limites de champ-moyen

Une limite de champ-moyen se propose d'étudier un système de n particules[5] en interactions symétriques, décrites par le n-uplet $(X_t^{n,1}, \dots, X_t^{n,n})$, lorsque n est très grand $(n \to +\infty)$. L'indice n dans $X_t^{n,1}$ par exemple, souligne le fait que la dynamique des particules dépend du nombre de particules dans le système : le fait de rajouter une particule perturbe la dynamique des autres particules car elles sont en interaction. En général, le système de particules est supposé être homogène (les particules sont échangeables[6]) et les interactions sont calibrées de telle sorte que la force d'interaction subie par une particule est asymptotiquement d'ordre 1. Par homogénéité, le terme d'interaction peut s'écrire comme une fonction de la mesure empirique des particules, disons $g(n^{-1}\sum_{i=1}^n \delta_{X_t^{n,i}})$. La force d'interaction entre deux particules données tend vers 0 à mesure que la taille du système n tend vers $+\infty$, ce qui suggère l'indépendance asymptotique entre deux particules données si celles-ci ont été initialisées de manière indépendante. Ce phénomène est communément appelé *propagation du chaos* et se trouve être intimement lié à la convergence de la mesure empirique des particules vers une loi limite (voir [101, Proposition 4.2] par exemple). De plus, la loi limite commune de toutes les particules est généralement caractérisée comme étant la solution d'une EDP non-linéaire. Cette non-linéarité, communément qualifiée de type *McKean-Vlasov*, provient du remplacement du terme d'interaction $g(n^{-1}\sum_{i=1}^n \delta_{X_t^{n,i}})$ par $g(P_t)$, où P_t est la loi limite d'une particule au temps t. Sous réserve d'hypothèse de continuité sur la fonction g, cela s'explique par l'heuristique suivante : l'indépendance asymptotique des particules, suggérée par la propagation du chaos, associée à la loi des grands nombres donne la convergence de la mesure empirique $n^{-1}\sum_{i=1}^n \delta_{X_t^{n,i}}$ vers P_t.

D'un point de vue technique, deux méthodes sont principalement utilisés : raisonnement par compacité ou bien argument de couplage (voir l'article fondateur de Sznitman [156]). Dans le premier cas, les preuves suivent le schéma suivant [101, 155] :

1. Montrer la tension (i.e. compacité faible) de la suite des mesures empiriques.

2. Identifier une équation vérifiée par tout point d'adhérence de cette suite.

3. Caractériser le point limite en prouvant l'unicité des solutions de cette équation.

4. La compacité combinée à l'unicité du point limite donne la convergence.

[5] Notons que la terminologie "particule" vient du fait que, à l'origine, les limites de champ-moyen concernaient principalement des systèmes de particules physiques. Ces particules en interaction hamiltonienne sont décrites à la limite par une équation cinétique (Boltzmann, Vlasov). La littérature sur ce sujet est très dense [84, 101, 152, 153, 156].

[6] La loi $(X_t^{n,1}, \dots, X_t^{n,n})$ est stable sous l'action d'une permutation des coordonnées.

Bien que fonctionnant dans des cadres très généraux, cette méthode a un caractère abstrait. En particulier, elle ne donne pas de vitesse de convergence. Par ailleurs, les preuves par couplage suivent le schéma suivant :

1. Trouver (souvent de manière heuristique dans un premier temps) un processus adéquat représentant la dynamique limite. Notons-le $(\overline{X}_t)_{t\geq 0}$ ici. La recherche de ce candidat repose sur l'heuristique de la loi des grands nombres décrite ci-dessus. Notons que la mesure limite P_t sera la loi de $\overline{X}_t$.

2. Montrer que ce processus est bien défini. Cette étape, commune avec l'approche par compacité, n'est pas triviale en général car la définition du processus limite $(\overline{X}_t)_{t\geq 0}$ fait intervenir un problème de point fixe de type McKean-Vlasov.

3. Coupler la "vraie" dynamique avec la dynamique limite de manière astucieuse. Pour cela, il faut construire des copies i.i.d. du processus limite $(\overline{X}_t)_{t\geq 0}$, notons-les $(\overline{X}^i_t)_{t\geq 0}$ pour tout entier $i \geq 1$, qui soient proches des processus dirigés par la "vraie" dynamique, i.e. les $(X^{n,i}_t)_{t\geq 0}$. Informellement, on cherche, en général, à diriger $(\overline{X}^i_t)_{t\geq 0}$ par le bruit de $(X^{n,i}_t)_{t\geq 0}$. Remarquons ici que les processus limites $(\overline{X}^i_t)_{t\geq 0}$ ne dépendent pas de n contrairement aux $(X^{n,i}_t)_{t\geq 0}$.

4. Finalement, si le processus limite trouvé à la première étape est bon et que le couplage a été bien fait, il est alors possible de montrer la convergence des $(X^{n,i}_t)_{t\geq 0}$ vers les copies i.i.d. $(\overline{X}^i_t)_{t\geq 0}$. En général, la distance entre les deux processus est quantifiable ce qui permet d'en déduire un majorant d'une distance de couplage (distance de Wasserstein par exemple) entre les deux distributions correspondantes.

Cette seconde méthode a l'avantage de fournir une vitesse de convergence, mais requiert en général des hypothèses plus fortes sur le modèle. Dans ce manuscrit, nous utilisons un argument de couplage pour montrer un résultat de type "loi des grands nombres" dans le chapitre 3 et la vitesse de convergence obtenue sert de pierre angulaire à l'étude du "théorème central limite" associé effectuée dans le chapitre 4.

1.3.d) Approche par limite de champ-moyen

Bien qu'étant particulièrement adaptées à l'étude de systèmes de particules physiques (gaz, plasma, ...), les approches de type champ-moyen (voir la section 1.3.c) pour plus de détails) sont courantes dans l'étude de grands réseaux de neurones en interaction[7]. Distinguons quatre types de modèles : les systèmes dynamiques qui génèrent des spikes de manière intrinsèque (FitzHugh–Nagumo [96, 97]), les modèles qui génèrent des spikes par seuillage (intègre-et-décharge [22, 41, 42, 76]), les oscillateurs couplés (Kuramoto [58, 59]) et les processus ponctuels ([52] ou [55, 73] qui proposent des modèles différents).

Les chapitres 3 et 4 de ce manuscrit proposent d'étudier la limite de champ-moyen d'un réseau de processus de Hawkes dépendants de l'âge (voir la définition 1.3.1 plus bas). La première partie de cette étude (i.e. le chapitre 3) est inspirée de l'étude d'un réseau de processus de Hawkes standards effectuée par Delattre, Fournier et Hoffmann

[7] L'hypothèse d'échangeabilité des neurones semble raisonnable dans certaines régions du système nerveux. De plus, le réseau neuronal est très connecté (voir [49] pour plus de détails sur les approches de type champ-moyen en neurosciences).

dans [43]. Ils montrent que des processus de Hawkes en interaction de type champ-moyen $(N^{n,1}, \dots, N^{n,n})$ dont les intensités sont données par

$$\lambda_t^{n,i} = \Phi\left(\frac{1}{n}\sum_{j=1}^{n}\int_0^{t-} h(t-z)N^{n,j}(dz)\right), \tag{1.10}$$

peuvent être approximés (quand $n \to +\infty$) par des processus de Poisson inhomogènes indépendants. Dans le modèle (1.10), en comparaison avec (1.5), les fonctions intensités Φ_i sont toutes égales ainsi que les fonctions d'interaction $h_{j\to i}$. En particulier, le graphe d'interaction entre les processus est le graphe complet et la valeur de l'intensité stochastique est la même pour tous les processus, d'où l'échangeabilité de $(N^{n,1}, \dots, N^{n,n})$. Originellement destinés à des applications en mathématiques financières, les résultats énoncés dans [43] ont été généralisés à un nombre fini de grandes populations de neurones par Ditlevsen et Löcherbach dans [44], chaque population pouvant être résumée, asymptotiquement, par son activité moyenne. Dans le cadre de plusieurs populations en interaction, un phénomène d'oscillations auto-entretenues peut apparaître. D'un point de vue biologique, ce phénomène peut permettre de comprendre comment le cerveau contrôle les rythmes biologiques (rythme cardiaque, pulmonaire, cycle de veille, etc).

Contributions de la thèse. Par souci de clarté, une version simplifiée[8] du modèle étudié dans les chapitres 3 et 4 est présentée ici. Ce modèle est basé sur les processus de Hawkes en interaction de type champ-moyen (1.10) qu'il généralise en ajoutant, entre autres, une dépendance par rapport à l'âge du processus lui-même. L'ajout de cette dépendance permet de contrecarrer l'une des principales carences du processus de Hawkes dans son utilisation en neurosciences : l'absence d'une description simple du phénomène de période réfractaire.

Définition 1.3.1. *Un réseau de processus de Hawkes dépendants de l'âge (ADHP) de paramètres (n, h, Ψ) est un processus ponctuel multivarié $(N^{n,i})_{i=1,..,n}$ dont les intensités sont respectivement*

$$\lambda_t^{n,i} = \Psi\left(S_{t-}^{n,i}, \frac{1}{n}\sum_{j=1}^{n}\int_0^{t-} h(t-z)N^{n,j}(dz)\right), \tag{1.11}$$

où $(S_{t-}^{n,i})_{t\geq 0}$ est le processus d'âge prévisible associé à $N^{n,i}$.

Le terme intégral dans (1.11) est homogène en i ce qui implique l'échangeabilité de $(N^{n,1}, \dots, N^{n,n})$. Par souci de simplicité, nous ne spécifions pas ici toutes les hypothèses techniques (qui sont décrites dans le chapitre 3), mais précisons tout de même l'hypothèse principale du modèle : la fonction Ψ est lipschitzienne en sa seconde variable (uniformément par rapport à la première).

Basé sur un argument de couplage, inspiré de [43] et reposant sur la méthode de thinning (introduite dans la section 1.1.d)), le théorème 3.4.1 montre que des ADHP peuvent être approchés par des processus ponctuels indépendants et identiquement distribués $\overline{N}^i$, $i \geq 1$. L'intensité de $\overline{N}^1$, notée $(\overline{\lambda}_t^1)_{t\geq 0}$, est une fonction du temps t et de l'âge prévisible associé $\overline{S}_{t-}^1$. En particulier, si les processus $N^{n,i}$ présentent des périodes réfractaires, alors il en va de même pour les processus limites $\overline{N}^i$.

[8] Précisons ici que, en comparaison avec (1.10), le graphe d'interaction associé au modèle étudié dans le chapitre 3 peut être de type Erdős-Rényi par exemple plutôt que complet.

Les processus d'âge (standards) associés aux ADHP, notés $(S_t^{n,i})_{t\geq 0}$, peuvent également être approximés et le corollaire 3.4.5 donne le résultat de type "loi des grands nombres" suivant : la mesure empirique des âges $\overline{\mu}^n_{S_t} := n^{-1}\sum_{i=1}^n \delta_{S_t^{n,i}}$ converge vers P_t qui est la loi de l'âge $\overline{S}^1_t$ associé à $\overline{N}^1$.

Le lien avec une EDP se fait via le processus limite. En effet, l'intensité $(\overline{\lambda}^1_t)_{t\geq 0}$ vérifie l'équation de point fixe de type McKean-Vlasov suivante :

$$\overline{\lambda}^1_t = \Psi\Big(\overline{S}^1_{t-}, \int_0^t h(t-z)\mathbb{E}[\overline{\lambda}^1_z]dz\Big).$$

L'intensité dépend de la propre loi du processus et cette dépendance ressemble fortement à la non-linéarité présente dans le système (PPS). En particulier, la proposition 3.3.8 (qui utilise des résultats du chapitre 2) montre que :

> Si la condition initiale $u(0,\cdot) = u^{\text{in}}$ est une densité, alors l'unique solution u du système suivant,
>
> $$\begin{cases} \dfrac{\partial u\,(t,s)}{\partial t} + \dfrac{\partial u\,(t,s)}{\partial s} + \Psi\,(s, X(t))\,u\,(t,s) = 0, \\ u\,(t,0) = \displaystyle\int_{s\in\mathbb{R}} \Psi\,(s, X(t))\,u\,(t,s)\,ds, \end{cases} \tag{1.12}$$
>
> où $X(t) = \int_0^t h(t-z)u(z,0)dz$, est telle que $u(t,\cdot)$ est la densité de P_t (la loi de l'âge $\overline{S}^1_t$).

Notons que le système ci-dessus est identique au système (PPS) à condition de remplacer $u(t,s)$ par $n(t,s)$, Ψ par p et h par d. Toutefois, nous avons choisi d'utiliser des notations différentes car les hypothèses supposées sur les fonctions Ψ et h dans ce manuscrit diffèrent de celles supposées sur p et d dans [114]. En particulier, les études de l'existence et de l'unicité des solutions de (PPS) et (1.12) diffèrent (légèrement). Nous démontrons dans ce manuscrit l'unicité des solutions de (1.12) pour des solutions à valeur mesure et l'existence d'une solution à valeur densité (voir le théorème 3.3.5). Finalement, le résultat de convergence de $\overline{\mu}^n_{S_t}$ vers la loi P_t combiné avec l'identification de P_t comme la solution de (1.12) permet de faire le lien entre les deux échelles de modélisation présentées en section 1.2.c) :

$$\overline{\mu}^n_{S_t} \xrightarrow[n\to+\infty]{} u(t,\cdot).$$

De plus, la ressemblance entre la variable $X(t)$ et l'intégrale intervenant dans l'intensité d'un processus de Hawkes, constatée à la fin de la section 1.2.c), est ici explicite.

Ayant obtenu un résultat de type "loi des grands nombres" pour la mesure empirique des âges, il est naturel de chercher le résultat de type "théorème central limite" associé. C'est l'analyse menée dans le chapitre 4. L'étude porte donc sur les "petites" fluctuations de $\overline{\mu}^n_{S_t}$ autour de sa limite P_t. Considérons le processus de fluctuation $(\eta^n_t)_{t\geq 0}$ défini par $\eta^n_t = \sqrt{n}(\overline{\mu}^n_{S_t} - P_t)$. Précisons ici que η^n_t est une mesure sur l'espace d'état des âges, à savoir $\mathbb{R}_+$. Une manière d'étudier cette mesure est donc de l'évaluer contre des fonctions test. En s'inspirant de travaux de Méléard et de ses co-auteurs [50, 82, 101], nous utilisons des fonctions tests φ dans un espace de Sobolev à poids. Plus précisément, nous considérons les espaces $\mathcal{W}_0^{k,\alpha}$, pour k entier (représentant l'ordre de régularité) et α dans $\mathbb{R}_+$ (représentant

le poids), définis comme la complétion de l'espace des fonctions $\mathcal{C}^\infty$ à support compact pour la norme suivante

$$||f||_{k,\alpha} := \left(\sum_{k'=0}^{k} \int_{\mathbb{R}} \frac{|f^{(k')}(x)|^2}{1+|x|^{2\alpha}} dx \right)^{1/2}.$$

Ainsi, la mesure η_t^n peut être vue comme une forme linéaire sur ces espaces $\mathcal{W}_0^{k,\alpha}$ (pour des valeurs particulières de k et α que nous choisissons de ne pas préciser ici). De manière équivalente, nous considérons η_t^n comme un élément du dual de $\mathcal{W}_0^{k,\alpha}$, noté $\mathcal{W}_0^{-k,\alpha}$. La suite repose sur un raisonnement par compacité. Nous montrons que le processus $(\eta_t^n)_{t\geq 0}$, qui prend ses valeurs dans $\mathcal{W}_0^{-k,\alpha}$, est tendu (théorème 4.4.14). De plus, le théorème 4.5.6 montre que tout point limite η vérifie un système de deux équations faisant intervenir un deuxième processus Γ (qui est à valeurs réelles et que nous choisissons de ne pas expliciter dans cette introduction). Finalement, l'unicité des solutions de ce système (proposition 4.5.11) implique l'unicité des points limites de $(\eta^n)_{n\geq 1}$ et donc la convergence de cette suite (théorème 4.5.12).

Pour finir, notons que ce résultat, bien que technique, ouvre la voie pour établir des procédures statistiques tels que des tests d'adéquation.

1.4 Seconde partie : détection de synchronisations entre neurones

Le chapitre 5 s'intéresse à la détection de motifs de dépendance dans l'activité électrique de neurones d'un point de vue statistique. Les résultats présentés reposent sur des travaux débutés lors du stage de Master 2 précédant cette thèse et se rapprochent des autres chapitres de ce manuscrit de part le fait qu'ils concernent des processus ponctuels dans le cadre de la modélisation de neurones.

1.4.a) Motivations biologiques

Historiquement, les neurones étaient considérés comme des entités agissant de manière indépendante [10]. Cette conception est principalement due au fait que pour des raisons technologiques, les neurobiologistes ne pouvaient enregistrer que l'activité d'un seul neurone à la fois[9]. Depuis l'utilisation de multi-électrodes, de nombreuses études ont porté sur les interactions entre les neurones et notamment leur synchronisation. Ces multi-électrodes (amas de plusieurs électrodes dans une petite région du cortex) enregistrent le potentiel électrique en des points très proches. Les potentiels d'action générés par les neurones proches du point d'implantation de l'électrode se retrouvent dans la dynamique du potentiel enregistré sous la forme de pics. Une triangulation spatiale, possible grâce à l'enregistrement redondant des multi-électrodes, permet de classifier les spikes selon le neurone qui les a générés (algorithme de *spike sorting*, voir [127]). Dans la suite, nous supposerons avoir accès à des données ayant subi cette classification comme pré-traitement. Ainsi, les données qui nous intéressent sont formées par les trains de spikes correspondant à un petit nombre $n \geq 2$ de

[9]Notons que deux neurones appartenant à une région corticale présentant de nombreux neurones en interaction de type champ-moyen peuvent présenter des activités électriques indépendantes (cf le phénomène de propagation du chaos présenté en section 1.3.d)). Réciproquement, l'hypothèse de champ-moyen n'est pas réaliste lorsque des neurones se synchronisent à certains moments au cours d'une tâche.

neurones enregistrés au cours d'une tâche sensori-motrice. De plus, nous supposerons avoir accès à M enregistrements indépendants et identiquement distribués correspondant à M répétitions de cette tâche.

Une *assemblée neuronale* est un groupe de neurones qui montrent une certaine tendance à exhiber des activités électriques synchronisées. Cette notion a été proposée dans [72] pour décrire le codage et le traitement de l'information par les neurones. La détection de synchronisations entre neurones a donc pour objectif d'identifier des assemblées neuronales intervenant à certains moments précis au cours de l'éxécution d'une tâche. Pour ce faire, de nombreuses méthodes sont possibles.

Une des méthodes les plus populaires à ce jour est la méthode des *Unitary Events* (UE) introduite dans la thèse de Grün [62]. Cette méthode est basée sur l'idée qu'une dépendance excitatrice du neurone A vers le neurone B doit être caractérisée par une surabondance du motif suivant : spike du neurone A suivi (dans un délai très court de l'ordre de la milliseconde) d'un spike du neurone B. Si un tel motif est sur-représenté, alors il est qualifié de Unitary Event. Utilisables sur des données discrétisées en temps, les différentes méthodes proposées par Grün et ses collaborateurs (par exemple [63, 64, 65, 66]) ont récemment été généralisées au cadre des processus ponctuels (données continues en temps) sous le nom de "Multiple Tests based on a Gaussian Approximation of the Unitary Events" (MTGAUE) dans un article de Tuleau-Malot et ses co-auteurs [158]. La méthode MTGAUE est construite pour tester l'indépendance entre deux processus ponctuels.

1.4.b) Contributions de la thèse

Nous généralisons (en partie) la méthode statistique introduite dans [158] au cas de $n \geq 2$ processus ponctuels représentant les trains de spikes de n neurones. Pour $N_1, \ldots, N_n$ des processus ponctuels sur $[a, b]$ et $\mathscr{L}$ un sous-ensemble de $\{1, ..., n\}$ donné, nous donnons une réponse statistique à la question suivante : est-ce que les processus N_l, $l \in \mathscr{L}$, sont indépendants ?

Nous répondons à cette question en construisant un test statistique de décision entre les deux hypothèses :

$$\begin{cases} (\mathcal{H}_0) \text{ "Les processus } N_l,\ l \in \mathscr{L} \text{ sont indépendants."} \\ (\mathcal{H}_1) \text{ "Les processus } N_l,\ l \in \mathscr{L} \text{ ne sont pas indépendants."} \end{cases}$$

La méthode MTGAUE de [158] est basée sur la notion de coïncidence avec délai entre deux neurones définie de la manière suivante : pour un paramètre $\delta > 0$, la variable

$$X_{\{1,2\}} = X_{\{1,2\}}(\delta) := \int_{[a,b]^2} \mathbf{1}_{|x-y| \leq \delta} N_1(dx) N_2(dy) = \sum_{X \in N_1} \sum_{Y \in N_2} \mathbf{1}_{|X-Y| \leq \delta}$$

désigne le nombre de coïncidences (de délai δ) entre les processus ponctuels N_1 et N_2. La généralisation à n processus que nous proposons est donnée par la définition qui suit.

Définition 1.4.1. *Soit $\mathscr{L}$ un sous-ensemble de $\{1, \ldots, n\}$. Notons $L \geq 2$ son cardinal et $\mathcal{L} = \{i_1 < \cdots < i_L\}$. Pour $\delta > 0$, le nombre $X_{\mathscr{L}}$ de coïncidences de délai δ entre les processus $N_{i_1}, \ldots, N_{i_L}$ est défini par :*

$$X_{\mathscr{L}} = X_{\mathscr{L}}(\delta) := \int_{[a,b]^L} \mathbf{1}_{\left| \max_{i \in \{1,\ldots,L\}} x_i - \min_{i \in \{1,\ldots,L\}} x_i \right| \leq \delta} N_{i_1}(dx_1) \ldots N_{i_L}(dx_L).$$

Cette notion de coïncidence entre plus de deux neurones est détaillée dans la section 5.2.b).

Sous $(\mathcal{H}_0)$, en supposant que les processus $N_1, \dots, N_n$ sont des processus de Poisson homogènes d'intensités respectives $\lambda_1, \dots, \lambda_n$ sur $[a, b]$, il est possible (voir le théorème 5.3.1 et proposition 5.3.2) de calculer explicitement l'espérance et la variance de $X_{\mathscr{L}}$ en fonction des paramètres, c'est-à-dire

$$\begin{cases} \mathbb{E}[X_{\mathscr{L}}] = m_0(n, L, \delta, a, b, \lambda_1, \dots, \lambda_n) = m_0 \\ \mathbf{V}\mathrm{ar}(X_{\mathscr{L}}) = v_0(n, L, \delta, a, b, \lambda_1, \dots, \lambda_n) = v_0. \end{cases}$$

De plus, si $(N_1^{(k)}, \dots, N_n^{(k)})_{1 \leq k \leq M}$ désigne M essais indépendants et de même loi que $(N_1, \dots, N_n)$, alors le théorème central limite donne la convergence en loi suivante[10] :

$$\sqrt{M}\,(\overline{m} - m_0) \xrightarrow[M \to \infty]{\text{loi}} \mathcal{N}(0, v_0)\,,$$

où $\overline{m} = \frac{1}{M} \sum_{k=1}^{M} X_{\mathscr{L}}^{(k)}$, et $X_{\mathscr{L}}^{(k)}$ est le nombre de coïncidences au k^{e} essai, c'est-à-dire entre les processus $N_{i_1}^{(k)}, \dots, N_{i_L}^{(k)}$. Cette convergence n'est pas directement utilisable d'un point de vue pratique car les intensités ne sont pas observables. Cependant, le lemme de Slutsky allié à la méthode delta garantit que le remplacement des vraies intensités λ_l par leurs estimateurs $\hat{\lambda}_l := \frac{1}{M(b-a)} \sum_{k=1}^{M} N_l^{(k)}([a, b])$ ne compromet pas le comportement gaussien de la limite, quitte à changer la variance. Plus précisément, nous montrons la convergence en loi suivante :

$$\sqrt{M} \frac{(\overline{m} - \hat{m}_0)}{\sqrt{\hat{\sigma}^2}} \xrightarrow[M \to \infty]{\text{loi}} \mathcal{N}(0, 1)\,,$$

où $\hat{m}_0 := m_0(n, L, \delta, a, b, \hat{\lambda}_1, \dots, \hat{\lambda}_n)$, $\hat{v}_0 := v_0(n, L, \delta, a, b, \hat{\lambda}_1, \dots, \hat{\lambda}_n)$ et

$$\hat{\sigma}^2 = \hat{v}_0 - f(n, L, \delta, a, b, \hat{\lambda}_1, \dots, \hat{\lambda}_n),$$

pour une certaine fonction positive f explicite. Cette modification de la variance provient du fait que les vraies intensités sont remplacées par leurs estimateurs. La valeur quantitative de cette modification est donnée par la méthode delta.

Cette étude asymptotique de la moyenne empirique du nombre de coïncidences permet de construire un test dont le niveau[11] asymptotique est contrôlé (voir la définition 5.3.4 et le corollaire 5.3.5). De plus, nous mettons en place une procédure de tests multiples, à partir de la procédure de test simple décrite ci-dessus, qui permet de détecter les sous-ensembles $\mathscr{L}$ de neurones montrant des évidences statistiques de dépendance. Cette procédure de tests multiples est basée sur la procédure de Benjamini-Hochberg [11].

L'utilisation de ces résultats théoriques sur des données simulées ($n = 4$ neurones dont les taux de décharges sont cohérents avec ceux mesurés expérimentalement) nous a permis de montrer que :

- l'approximation gaussienne est valide à partir d'un nombre d'essais M de l'ordre de 50 ;
- notre procédure reste raisonnable en pratique dans le cadre plus général des processus de Hawkes (modèle plus fidèle dans le cadre de la modélisation de neurones) ;

[10] La notation $\mathcal{N}(m, \sigma^2)$ désigne la loi normale de moyenne m et variance σ^2.

[11] Probabilité de rejeter l'hypothèse $(\mathcal{H}_0)$ alors que celle-ci est vraie.

- notre procédure de test simple semble aussi performante que la méthode UE décrite dans [64] en termes de puissance[12] (empirique) tout en garantissant le niveau (empirique), ce qui n'est pas le cas de la méthode UE ;
- la procédure de tests multiples sur tous les différents sous-ensembles $\mathscr{L} \subset \{1, \ldots, n\}$ montre des résultats satisfaisant (la multiplicité du test est $2^4 - 4 - 1 = 11$ ici).

Notons tout de même que le nombre de tests simultanés possibles avec la méthode de tests multiples est clairement limité (moins de trente). Cela provient du fait que l'approximation gaussienne n'est pas assez précise et que la fiabilité de la procédure n'est plus assurée.

Finalement, notre procédure est appliquée à des données issues d'une expérimentation effectuée sur un singe rhésus entraîné à répondre à des stimuli visuels. Le protocole expérimental est détaillé dans la section 5.6. Les résultats retournés par notre procédure suggèrent la présence de synchronisations lors d'un moment-clé de la tâche et que ces synchronisations ne sont pas présentes lors des moments de repos. Ces résultats sont en accord avec d'autres études effectuées sur les mêmes données [61, 158].

1.5 État des lieux

À l'heure de l'écriture de ce manuscrit :

- Le chapitre 2 correspond à des travaux, effectués en collaboration avec María José Cáceres[13], Marie Doumic[14,15] et Patricia Reynaud-Bouret[16]. Ils ont fait l'objet d'une publication à *Mathematical Models and Methods in Applied Sciences* [29].
- Le chapitre 3 correspond à une pré-publication disponible sur ArXiv [28] et soumise à un journal.
- Le chapitre 4 correspond à des travaux en cours et fera très prochainement l'objet d'une future soumission.
- Le chapitre 5 correspond à des travaux, effectués en collaboration avec Thomas Laloë[16]. Ils ont fait l'objet d'une publication à *Biometrical Journal* [30].

Nous avons laissé les chapitres 2, 3 et 5 sous leur forme d'article mise à part quelques modifications dans le but de préserver une cohérence entre les chapitres. En particulier, certaines considérations et définitions peuvent être redondantes avec cette introduction.

[12] Probabilité de rejeter $(\mathcal{H}_0)$ quand $(\mathcal{H}_0)$ est effectivement erronée.

[13] Departamento de Matemática Aplicada , Universidad de Granada, Campus de Fuentenueva E-18071 Grenade, Espagne.

[14] Inria Paris-Rocquencourt, Domaine de Voluceau BP 105, 78153 Le Chesnay, France.

[15] UPMC Université Paris 6, JLLL, 4 place Jussieu 75005 Paris, France.

[16] Laboratoire J. A. Dieudonné, UMR CNRS 6621, Université Nice Sophia Antipolis, Parc Valrose 06108 Nice Cedex 2, France.

Chapter

2 Microscopic approach of a time elapsed neural model

Abstract. The spike trains are the main components of the information processing in the brain. To model spike trains several point processes have been investigated in the literature. And more macroscopic approaches have also been studied, using partial differential equation models. The main aim of the present chapter is to build a bridge between several point processes models (Poisson, Wold, Hawkes) that have been proved to statistically fit real spike trains data and age-structured partial differential equations as introduced by Pakdaman, Perthame and Salort.

This chapter is the fruit of a collaboration with María José Cáceres[1], Marie Doumic[2,3] and Patricia Reynaud-Bouret[4]. The corresponding article [29] is published in *Mathematical Models and Methods in Applied Sciences*. However, the last section has not been published and can be considered as a supplementary material.

Contents

[1]Departamento de Matemática Aplicada , Universidad de Granada, Campus de Fuentenueva E-18071 Granada, Spain.
[2]Inria Paris-Rocquencourt, Domaine de Voluceau BP 105, 78153 Le Chesnay, France.
[3]UPMC University of Paris 6, JL Lions Lab., 4 place Jussieu 75005 Paris, France.
[4]Laboratoire J. A. Dieudonné, UMR CNRS 6621, Université Nice Sophia Antipolis, Parc Valrose 06108 Nice Cedex 2, France.

2.1 Introduction

In Neuroscience, the action potentials (spikes) are the main components of the real-time information processing in the brain. Indeed, thanks to the synaptic integration, the membrane voltage of a neuron depends on the action potentials emitted by some others, whereas if this membrane potential is sufficiently high, there is production of action potentials.

To access those phenomena, schematically, one can proceed in two ways: extracellularly record in vivo several neurons, at a same time, and have access to simultaneous spike trains (only the list of events corresponding to action potentials) or intracellularly record the whole membrane voltage of only one neuron at a time, being blind to the nearby neurons.

Many people focus on spike trains. Those data are fundamentally random and can be modelled easily by time point processes, i.e. random countable sets of points on $\mathbb{R}_+$. Several point processes models have been investigated in the literature, each of them reproducing different features of the neuronal reality. The easiest model is the homogeneous Poisson process, which can only reproduce a constant firing rate for the neuron, but which, in particular, fails to reproduce refractory periods[5]. It is commonly admitted that this model is too poor to be realistic. Indeed, in such a model, two points or spikes can be arbitrary close as soon as their overall frequency is respected in average. Another more realistic model is the renewal process [126], where the occurrence of a point or spike depends on the previous occurrence. More precisely, the distribution of delays between spikes (also called inter-spike intervals, ISI) is given and a distribution, which provides small weights to small delays, is able to mimic refractory periods. A deeper statistical analysis has shown that Wold processes is showing good results, with respect to goodness-of-fit test on real data sets [127]. Wold processes are point processes for which the next occurrence of a spike depends on the previous occurrence but also on the previous ISI. From another point of

[5]Biologically, a neuron cannot produce two spikes too closely in time.

view, the fact that spike trains are usually non stationary can be easily modelled by inhomogeneous Poisson processes [159]. All those models do not reflect one of the main features of spike trains, which is the synaptic integration and there has been various attempts to catch such phenomenon. One of the main model is the Hawkes model, which has been introduced in [31] and which has been recently shown to fit several stationary data [136]. Several studies have been done in similar directions (see for instance [18]). More recently a vast interest has been shown to generalized linear models [123], with which one can infer functional connectivity and which are just an exponential variant of Hawkes models.

There has also been several models of the full membrane voltage such as Hodgkin-Huxley models. It is possible to fit some of those probabilistic stochastic differential equations (SDE) on real voltage data [78] and to use them to estimate meaningful physiological parameters [45]. However, the lack of simultaneous data (voltages of different neurons at the same time) prevent these models to be used as statistical models that can be fitted on network data, to estimate network parameters. A simple SDE model taking synaptic integration into account is the well-known Integrate-and-Fire (IF) model. Several variations have been proposed to describe several features of real neural networks such as oscillations [21, 22]. In particular, there exists hybrid IF models including inhomogeneous voltage driven Poisson process [78] that are able to mimic real membrane potential data. However up to our knowledge and unlike point processes models, no statistical test have been applied to show that any of the previous variations of the IF model fit real network data.

Both, SDE and point processes, approaches are microscopic descriptions, where random noise explains the intrinsic variability. Many authors have argued that there must be some more macroscopic approach describing huge neural networks as a whole, using PDE formalism [33, 150]. Some authors have already been able to perform link between PDE approaches as the macroscopic system and SDE approach (in particular IF models) as the microscopic model [99, 113, 134]. Another macroscopic point of view on spike trains is proposed by Pakdaman, Perthame and Salort in a series of articles [114, 115, 116]. It uses a nonlinear age-structured equation to describe the spikes density. Adopting a population view, they aim at studying relaxation to equilibrium or spontaneous periodic oscillations. Their model is justified by a qualitative, heuristic approach. As many other models, their model shows several qualitative features such as oscillations that make it quite plausible for real networks, but once again there is no statistical proof of it, up to our knowledge.

In this context, the main purpose of the present chapter is to build a bridge between several point processes models that have been proved to statistically fit real spike trains data and age structured PDE of the type of Pakdaman, Perthame and Salort. The point processes are the microscopic models, the PDE being their meso-macroscopic counterpart. In this sense, it extends PDE approaches for IF models to models that statistically fit true spike trains data. In the first section, we introduce Pakdaman, Perthame and Salort PDE (PPS) via its heuristic informal and microscopic description, which is based on IF models. Then, in Section 2.3, we develop the different point process models, quite informally, to draw the main heuristic correspondences between both approaches. In particular, we introduce the conditional intensity of a point process and a fundamental construction, called Ogata's thinning [110], which allows a microscopic understanding of the dynamics of a point process. Thanks to Ogata's thinning, in Section 2.4, we have been able to rigorously derive a microscopic random weak version of (PPS) and to propose its expectation deterministic counterpart. An independent and identically distributed (i.i.d) population version

is also available. Several examples of applications are discussed in Section 2.5. To facilitate reading, technical results and proofs are included in two appendices. The present work is clearly just a first to link point processes and PDE: there are much more open questions than answered ones and this is discussed in the final conclusion. However, we think that this can be fundamental to acquire a deeper understanding of spike train models, their advantages as well as their limitations.

2.2 Synaptic integration and (PPS) equation

Based on the intuition that every neuron in the network should behave in the same way, Pakdaman, Perthame and Salort proposed in [114] a deterministic PDE denoted (PPS) in the sequel. The origin of this PDE is the classical (IF) model. In this section we describe the link between the (IF) microscopic model and the mesoscopic (PPS) model, the main aim being to show thereafter the relation between (PPS) model and other natural microscopic models for spike trains: point processes.

2.2.a) Integrate-and-fire

Integrate-and-fire models describe the time evolution of the membrane potential, $V(t)$, by means of ordinary differential equations as follows

$$C_m \frac{\mathrm{d}V}{\mathrm{d}t} = -g_L(V - V_L) + I(t), \tag{2.1}$$

where C_m is the capacitance of the membrane, g_L is the leak conductance and V_L is the leak reversal potential. If $V(t)$ exceeds a certain threshold $\theta > V_L$, the neuron fires / emits an action potential (spike) and $V(t)$ is reset to $V_r < V_L$. The *synaptic current* $I(t)$ takes into account the fact that other presynaptic neurons fire and excite the neuron of interest, whose potential is given by $V(t)$.

As stated in [114], the origin of (PPS) equation comes from [122], where the explicit solution of a classical IF model as (2.1) has been discussed. To be more precise the membrane voltage of one neuron at time t is described by:

$$V(t) = V_L + (V_r - V_L)e^{-(t-T)/\tau_m} + \int_T^t g(t-z)N_{input}(dz), \tag{2.2}$$

where T is the last spike emitted by the considered neuron, τ_m is the time constant of the system (normally $\tau_m = g_L/C_m$), h is the excitatory post synaptic potential (EPSP) and N_{input} is the sum of Dirac masses at each spike of the presynaptic neurons. Since after firing, $V(t)$ is reset to $V_r < V_L$, there is a refractory period when the neuron is less excitable than at rest. The constant time τ_m indicates whether the next spike can occur more or less rapidly. The other main quantity, $\int_T^t h(t-z)N_{input}(dz)$, is the *synaptic integration term*.

In [122], they consider a whole random network of such IF neurons and look at the behavior of this model, where the only randomness is in the network. In many other studies [21, 22, 23, 25, 99, 113, 150] IF models as (2.1) are considered to finally obtain other systems of partial differential equations (different to (PPS)) describing neural networks behavior. In these studies, each presynaptic neuron is assumed to fire as an independent Poisson process and via a diffusion approximation, the synaptic current is then approximated by a continuous in time stochastic process of Ornstein-Uhlenbeck.

2.2.b) The (PPS) equation

The deterministic PDE proposed by Pakdaman, Perthame and Salort, whose origin is also the microscopic IF model (2.2), is the following:

$$\begin{cases} \dfrac{\partial n\,(t,s)}{\partial t} + \dfrac{\partial n\,(t,s)}{\partial s} + p\,(s, X\,(t))\, n\,(t,s) = 0 \\ m\,(t) := n\,(t,0) = \displaystyle\int_0^{+\infty} p\,(s, X\,(t))\, n\,(t,s)\,\mathrm{d}s. \end{cases} \tag{PPS}$$

In this equation, $n(t,s)$ represents a probability density of neurons at time t having discharged at time $t-s$ for the last time. Therefore, s represents the time elapsed since the last discharge. The fact that the equation is an elapsed time structured equation is natural, because the IF model (2.2) clearly only depends on the time since the last spike. More informally, the variable s represents the "age" of the neuron.

The first equation of the system (PPS) represents a pure transport process and means that as time goes by, neurons of age s and past given by $X(t)$ are either aging linearly or reset to age 0 with rate $p\,(s, X\,(t))$.

The second equation of (PPS) describes the fact that when neurons spike, the age (the elapsed time) returns to 0. Therefore, $n(t,0)$ depicts the density of neurons undergoing a discharge at time t and it is denoted by $m(t)$. As a consequence of this boundary condition, for $n(\cdot,\cdot)$ at $s=0$, the following mass-conservation law is obtained:

$$\int_0^{+\infty} n\,(t,s)\,\mathrm{d}s = \int_0^{+\infty} n\,(0,s)\,\mathrm{d}s$$

This means that if $n\,(0,\cdot)$ is a probabilistic density then $n\,(t,\cdot)$ can be interpreted as a density at each time t. Denoting by $\mathrm{d}t$ the Lebesgue measure and since $m(t)$ is the density of firing neurons at time t in (PPS), $m(t)\mathrm{d}t$ can also be interpreted as the limit of $N_{input}(\mathrm{d}t)$ in (2.2) when the population of neurons becomes continuous.

The system (PPS) is nonlinear since the rate $p\,(s, X(t))$ depends on $n(t,0)$ by means of the quantity $X(t)$:

$$X(t) = \int_0^t h(t')m(t-t')\mathrm{d}t' = \int_0^t h(t')n(t-t',0)\mathrm{d}t'. \tag{2.3}$$

The quantity $X(t)$ represents the interactions between neurons. It "takes into account the averaged propagation time for the ionic pulse in this network" [114]. More precisely with respect to the IF models (2.2), this is the synaptic integration term, once the population becomes continuous. The only difference is that in (2.2) the memory is cancelled once the last spike has occurred and this is not the case here. However informally, both quantities have the same interpretation. Note nevertheless, that in [114], the function h can be much more general than the h of the IF models which clearly corresponds to EPSP. From now on and in the rest of the chapter, h is just a general non negative function without forcing the connection with EPSP.

The larger $p\,(s, X(t))$ the more likely neurons of age s and past $X(t)$ fire. Most of the time (but it is not a requisite), p is assumed to be less than 1 and is interpreted as the probability that neurons of age s fire. However, as shown in Section 2.4 and as interpreted in many population structured equation [32, 47, 121], $p(s, X(t))$ is closer to a hazard rate, i.e. a positive quantity such that $p\,(s, X(t))\,\mathrm{d}t$ is informally the probability to fire given that the neuron has not fired yet. In particular, it could be not bounded by 1 and does not

need to integrate to 1. A toy example is obtained if $p(s, X(t)) = \lambda > 0$, where a steady state solution is $n(t, s) = \lambda e^{-\lambda s}\mathbb{1}_{s\geq 0}$: this is the density of an exponential variable with parameter λ.

However, based on the interpretation of $p(s, X(t))$ as a probability bounded by 1, one of the main model that Pakdaman, Perthame and Salort consider is $p(s, X(t)) = \mathbb{1}_{s\geq\sigma(X(t))}$. This again can be easily interpreted by looking at (2.2). Indeed, since in the IF models the spike happens when the threshold θ is reached, one can consider that $p(s, X(t))$ should be equal to 1 whenever

$$V(t) = V_L + (V_r - V_L)e^{-(t-T)/\tau_m} + X(t) \geq \theta,$$

and 0 otherwise. Since $V_r - V_L < 0$, $p(s, X(t)) = 1$ is indeed equivalent to $s = t - T$ larger than some decreasing function of $X(t)$. This has the double advantage to give a formula for the refractory period ($\sigma(X(t))$) and to model excitatory systems: the refractory period decreases when the whole firing rate increases via $X(t)$ and this makes the neurons fire even more. This is for this particular case that Pakdaman, Perthame and Salort have shown existence of oscillatory behavior [115].

Another important parameter in the (PPS) model and introduced in [114] is J, which can be seen with our formalism as $\int h$ and which describes the network connectivity or the strength of the interaction. In [114] it has been proved that, for highly or weakly connected networks, (PPS) model exhibits relaxation to steady state and periodic solutions have also been numerically observed for moderately connected networks. The authors in [115] have quantified the regime where relaxation to a stationary solution occurs in terms of J and described periodic solution for intermediate values of J.

Recently, in [116], the (PPS) model has been extended including a fragmentation term, which describes the adaptation and fatigue of the neurons. In this sense, this new term incorporates the past activity of the neurons. For this new model, in the linear case there is exponential convergence to the steady states, while in the weakly nonlinear case a total desynchronization in the network is proved. Moreover, for greater nonlinearities, synchronization can again be numerically observed.

2.3 Point processes and conditional intensities as models for spike trains

We first start by quickly reviewing the main basic concepts and notations of point processes, in particular, conditional intensities and Ogata's thinning [110]. We refer the interested reader to [16] for exhaustiveness and to [19] for a much more condensed version, with the main useful notions.

2.3.a) Counting processes and conditional intensities

We focus on locally finite point processes on $\mathbb{R}$, equipped with the Borelians $\mathcal{B}(\mathbb{R})$.

Definition 2.3.1 (Locally finite point process)**.** *A locally finite point process N on $\mathbb{R}$ is a random set of points such that it has almost surely (a.s.) a finite number of points in finite intervals. Therefore, associated to N there is an ordered sequence of extended real valued random times $(T_z)_{z\in\mathbb{Z}}$: $\cdots \leq T_{-1} \leq T_0 \leq 0 < T_1 \leq \cdots$.*

For a measurable set A, N_A denotes the number of points of N in A. This is a random variable with values in $\mathbb{N} \cup \{+\infty\}$.

Definition 2.3.2 (Counting process associated to a point process)**.** *The process on $\mathbb{R}_+$ defined by $t \mapsto N_t := N_{(0,t]}$ is called the counting process associated to the point process N.*

The natural and the predictable filtrations are fundamental for the present work.

Definition 2.3.3 (Natural filtration of a point process)**.** *The natural filtration of N is the family $\left(\mathcal{F}_t^N\right)_{t\in\mathbb{R}}$ of σ-algebras defined by $\mathcal{F}_t^N = \sigma\left(N \cap (-\infty, t]\right)$.*

Definition 2.3.4 (Predictable filtration of a point process)**.** *The predictable filtration of N is the family of σ-algebra $\left(\mathcal{F}_{t-}^N\right)_{t\in\mathbb{R}}$ defined by $\mathcal{F}_{t-}^N = \sigma\left(N \cap (-\infty, t)\right)$.*

The intuition behind this concept is that $\mathcal{F}_t^N$ contains all the information given by the point process at time t. In particular, it contains the information whether t is a point of the process or not while $\mathcal{F}_{t-}^N$ only contains the information given by the point process strictly before t. Therefore, it does not contain (in general) the information whether t is a point or not. In this sense, $\mathcal{F}_{t-}^N$ represents (the information contained in) the past.

Under some rather classical conditions [16], which are always assumed to be satisfied here, one can associate to $(N_t)_{t\geq 0}$ a stochastic intensity λ_t[6], which is a non negative random quantity. The process $(N_t - \int_0^t \lambda_u \mathrm{d}u)_{t\geq 0}$ forms a local martingale [16]. Informally, $\lambda_t \mathrm{d}t$ represents the probability to have a new point in the interval $[t, t + \mathrm{d}t)$ given the past. Note that λ_t should not be understood as a function, in the same way as density is for random variables. It is a "recipe" explaining how the probability to find a new point at time t depends on the past configuration: since the past configuration depends on its own past, this is closer to a recursive formula. In this respect, the intensity should obviously depend on $N \cap (-\infty, t)$ and not on $N \cap (-\infty, t]$ to predict the occurrence at time t, since we cannot know whether t is already a point or not.

The distribution of the point process N on $\mathbb{R}$ is completely characterized by the knowledge of the intensity λ_t on $\mathbb{R}_+$ and the distribution of $N_- = N \cap \mathbb{R}_-$, which is denoted by ζ_{N_-} in the sequel. The information about ζ_{N_-} is necessary since each point of N may depend on the occurrence of all the previous points: if for all $t > 0$, one knows the "recipe" λ_t that gives the probability of a new point at time t given the past configuration, one still needs to know the distribution of N_- to obtain the whole process.
Two main assumptions are used depending on the type of results we seek:

$$\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,a.s.}\right) \;\middle|\; \text{for any } T \geq 0,\ \int_0^T \lambda_t \mathrm{d}t \text{ is finite a.s.}$$

$$\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,exp}\right) \;\middle|\; \text{for any } T \geq 0,\ \mathbb{E}\left[\int_0^T \lambda_t \mathrm{d}t\right] \text{ is finite.}$$

Clearly $\left(\mathcal{A}_{loc}^{\mathbb{L}^1,exp}\right)$ implies $\left(\mathcal{A}_{loc}^{\mathbb{L}^1,a.s.}\right)$. Note that $\left(\mathcal{A}_{loc}^{\mathbb{L}^1,a.s.}\right)$ implies non-explosion in finite time for the counting processes $(N_t)_{t\geq 0}$.

Definition 2.3.5 (Point measure associated to a point process)**.** *The point measure associated to N is denoted by $N(\mathrm{d}t)$ and defined by $N(\mathrm{d}t) = \sum_{i\in\mathbb{Z}} \delta_{T_i}(\mathrm{d}t)$, where δ_x is the Dirac mass in x.*

[6] In the article corresponding to this chapter, the intensity is denoted by $\lambda(t, \mathcal{F}_{t-}^N)$ instead of λ_t to emphasize its predictability.

By analogy with (PPS), and since points of point processes correspond to spikes (or times of discharge) for the considered neuron in spike train analysis, $N(\mathrm{d}t)$ is the microscopic equivalent of the distribution of discharging neurons $m(t)\mathrm{d}t$. Following this analogy, and since T_{N_t} is the last point less or equal to t for every $t \geq 0$, the age S_t at time t is defined by $S_t = t - T_{N_t}$. In particular, if t is a point of N, then $S_t = 0$. Note that S_t is $\mathcal{F}_t^N$ measurable for every $t \geq 0$ and therefore, $S_0 = -T_0$ is $\mathcal{F}_0^N$ measurable. To define an age at time $t = 0$, one assumes that

$$(\mathcal{A}_{T_0}) \,\big|\, \text{There exists a first point before 0 for the process } N_-, \text{ i.e. } -\infty < T_0.$$

As we have remarked before, the conditional intensity should depend on $N \cap (-\infty, t)$. Therefore, it cannot be function of S_t, since S_t informs us if t is a point or not. That is the main reason for considering the *predictable age process* defined by, for all $t \geq 0$,

$$S_{t-} = t - T_{N_{t-}}, \tag{2.4}$$

where $T_{N_{t-}}$ is the last point strictly before t (see Figure 2.1). Note also that knowing $(S_{t-})_{t\geq 0}$ or $(N_t)_{t\geq 0}$ is completely equivalent given $\mathcal{F}_0^N$.

The last and most crucial equivalence between (PPS) and the present point process set-up, consists in noting that the quantities $p(s, X(t))$ and λ_t have informally the same meaning: they both represent a firing rate, i.e. both give the rate of discharge as a function of the past. This dependence is made more explicit in $p(s, X(t))$ than in λ_t.

2.3.b) Examples

Let us review the basic point processes models of spike trains and see what kind of analogy is likely to exist between both models ((PPS) equation and point processes). These informal analogies are transformed, when possible, into exact mathematical results (see Section 2.5).

Homogeneous Poisson process This is the simplest case where $\lambda_t = \lambda$, with λ a fixed positive constant representing the firing rate. There is no dependence in time t (it is homogeneous) and no dependence with respect to the past. This case should be equivalent to $p(s, X(t)) = \lambda$ in (PPS). This can be made even more explicit. Indeed in the case where the Poisson process exists on the whole real line (stationary case), it is easy to see that

$$\mathbb{P}\left(S_{t-} > s\right) = \mathbb{P}\left(N_{[t-s,t)} = 0\right) = \exp(-\lambda s),$$

meaning that the age S_{t-} obeys an exponential distribution with parameter λ, i.e. the steady state of the toy example developed for (PPS) when $p(s, X(t)) = \lambda$.

Inhomogeneous Poisson process To model non stationarity, one can use $\lambda_t = \lambda(t)$, which only depends on time. This case should be equivalent to the replacement of $p(s, X(t))$ in (PPS) by $\lambda(t)$.

Renewal process This model is very useful to take refractory period into account. It corresponds to the case where the ISIs (delays between spikes) are independent and identically distributed (i.i.d.) with a certain given density ν on $\mathbb{R}_+$. The associated hazard rate is

$$f(s) = \frac{\nu(s)}{\int_s^{+\infty} \nu(x)\mathrm{d}x},$$

when $\int_s^{+\infty} \nu(x)\mathrm{d}x > 0$. Roughly speaking, $f(s)\mathrm{d}s$ is the probability that a neuron spikes with age s given that its age is larger than s. In this case, considering the set of spikes as the point process N, it is easy to show (see Section 2.8.a)) that its corresponding intensity is $\lambda_t = f(S_{t-})$ which only depends on the age[7]. One can also show quite easily that the process $(S_{t-})_{t>0}$, which is equal to $(S_t)_{t>0}$ almost everywhere (a.e.), is a Markovian process in time. This renewal setting should be equivalent in the (PPS) framework to $p(s, X(t)) = f(s)$.

Note that many people consider IF models (2.2) with Poissonian inputs with or without additive white noise. In both cases, the system erases all memory after each spike and therefore the ISIs are i.i.d. Therefore as long as we are only interested by the spike trains and their point process models, those IF models are just a particular case of renewal process [22, 24, 42, 122].

Wold process and more general structures Let A_t^1 be the delay (ISI) between the last point and the occurrence just before (see also Figure 2.1), i.e. $A_t^1 = T_{N_{t-}} - T_{N_{t-}-1}$. A Wold process [38, 89] is then characterized by $\lambda_t = f(S_{t-}, A_t^1)$. This model has been matched to several real data thanks to goodness-of-fit tests [127] and is therefore one of our main example with the next discussed Hawkes process case. One can show in this case that the successive ISI's form a Markov chain of order 1 and that the continuous time process (S_{t-}, A_t^1) is also Markovian.

This case should be equivalent to the replacement of $p(s, X(t))$ in (PPS) by $f(s, a^1)$, with a^1 denoting the delay between the two previous spikes. Naturally in this case, one should expect a PDE of higher dimension with third variable a^1.

More generally, one could define

$$A_t^k = T_{N_{t-}-(k-1)} - T_{N_{t-}-k}, \tag{2.5}$$

and point processes with intensity $\lambda_t = f(S_{t-}, A_t^1, ..., A_t^k)$. Those processes satisfy more generally that their ISI's form a Markov chain of order k and that the continuous time process $(S_{t-}, A_t^1, ..., A_t^k)$ is also Markovian (see Section 2.8.b)).

Remark 2.3.6. *The dynamics of the successive ages is pretty simple. On the one hand, the dynamics of the vector of the successive ages $(S_{t-}, A_t^1, ..., A_t^k)_{t>0}$ is deterministic between two jumping times. The first coordinate increases with rate 1. On the other hand, the dynamics at any jumping time T is given by the following shift:*

$$\begin{cases} \text{the age process goes to } 0, \text{ i.e. } S_T = 0, \\ \text{the first delay becomes the age, i.e. } A_{T+}^1 = S_{T-}, \\ \text{the other delays are shifted, i.e. } A_{T+}^i = A_T^{i-1} \text{ for all } i \leq k. \end{cases} \tag{2.6}$$

Hawkes processes The most classical setting is the linear (univariate) Hawkes process, which corresponds to

$$\lambda_t = \mu + \int_{-\infty}^{t-} h(t-z)N(dz),$$

where the positive parameter μ is called the spontaneous rate and the non negative function h, with support in $\mathbb{R}_+$, is called the interaction function, which is generally assumed to

[7] Remark that the intensity depends on the predictable age processes and not the standard ones since an intensity process must be predictable.

satisfy $\int_{\mathbb{R}_+} h < 1$ to guarantee the existence of a stationary version [38]. This model has also been matched to several real neuronal data thanks to goodness-of-fit tests [136]. Since it can mimic synaptic integration, as explained below, this represents the main example of the present work.

In the case where T_0 tends to $-\infty$, this is equivalent to say that there is no point on the negative half-line and in this case, one can rewrite

$$\lambda_t = \mu + \int_0^{t-} h(t-z)N(dz).$$

By analogy between $N(\mathrm{d}t)$ and $m(t)\mathrm{d}t$, one sees that $\int_0^{t-} h(t-z)N(dz)$ is indeed the analogous of $X(t)$ the synaptic integration in (2.3). So one could expect that the PDE analogue is given by $p(s, X(t)) = \mu + X(t)$. In Section 2.5, we show that this does not hold stricto sensu, whereas the other analogues work well.

Note that this model shares also some link with IF models. Indeed, the formula for the intensity is close to the formula for the voltage (2.2), with the same flavour for the synaptic integration term. The main difference comes from the fact that when the voltage reaches a certain threshold, it fires deterministically for the IF model, whereas the higher the intensity, the more likely is the spike for the Hawkes model, but without certainty. In this sense Hawkes models seem closer to (PPS) since as we discussed before, the term $p(s, X(t))$ is closer to a hazard rate and never imposes deterministically the presence of a spike.

To model inhibition (see [137] for instance), one can use functions h that may take negative values and in this case

$$\lambda_t = \left(\mu + \int_{-\infty}^{t-} h(t-z)N(dz)\right)_+,$$

which should correspond to $p(s, X(t)) = (\mu + X(t))_+$. Another possibility is

$$\lambda_t = \exp\left(\mu + \int_{-\infty}^{t-} h(t-z)N(dz)\right),$$

which is inspired by the generalized linear model as used by [123] and which should correspond to $p(s, X(t)) = \exp(\mu + X(t))$.

Note finally that Hawkes models in Neuroscience (and their variant) are usually multivariate meaning that they model interaction between spike trains thanks to interaction functions between point processes, each process representing a neuron. To keep the present analogy as simple as possible, we do not deal with those multivariate models in the present chapter. Some open questions in this direction are presented in conclusion.

2.3.c) Ogata's thinning algorithm

To turn the analogy between $p(s, X(t))$ and λ_t into a rigorous result on the PDE level, we need to understand the intrinsic dynamics of the point process. This dynamics is often not explicitly described in the literature (see e.g. the reference book by Brémaud [16]) because martingale theory provides a nice mathematical setting in which one can perform all the computations. However, when one wants to simulate point processes based on the

knowledge of their intensity, there is indeed a dynamics that is required to obtain a practical algorithm. This method has been described at first by Lewis in the Poisson setting [92] and generalized by Ogata in [110]. If there is a sketch of proof in [110], we have been unable to find any complete mathematical proof of this construction in the literature and we propose a full and mathematically complete version of this proof with minimal assumptions in Section 2.8.d). Let us just informally describe here, how this construction works.

The principle consists in assuming that is given an external homogeneous Poisson process Π of intensity 1 in $\mathbb{R}_+^2$ and with associated point measure given by

$$\Pi\left(\mathrm{d}t, \mathrm{d}x\right) = \sum_{(T,V)\in\Pi} \delta_{(T,V)}(\mathrm{d}t, \mathrm{d}x).$$

This means in particular that

$$\mathbb{E}\left[\Pi(\mathrm{d}t, \mathrm{d}x)\right] = \mathrm{d}t\ \mathrm{d}x. \tag{2.7}$$

Once a realisation of N_- fixed, which implies that $\mathcal{F}_0^N$ is known and which can be seen as an initial condition for the dynamics, the construction of the process N on $\mathbb{R}_+$ only depends on Π.

More precisely, if we know the intensity λ_t in the sense of the "recipe" that explicitly depends on t and $N\cap(-\infty, t)$, then once a realisation of Π and of N_- is fixed, the dynamics to build a point process N with intensity λ_t for $t \in \mathbb{R}_+$ is purely deterministic. It consists (see also Figure 2.1) in successively projecting on the abscissa axis the points that are below the graph of λ_t. Note that a point projection may change the shape of λ_t, just after the projection. Therefore the graph of λ_t evolves thanks to the realization of Π. For a more mathematical description, see Theorem 2.8.11 in Section 2.8.d). Note in particular that the construction ends on any finite interval $[0, T]$ a.s. if $\left(\mathcal{A}_{\lambda,loc}^{1,a.s}\right)$ holds.

Then the point process N, result of Ogata's thinning, is given by the union of N_- on $\mathbb{R}_-$ and the projected points on $\mathbb{R}_+$. It admits the desired intensity λ_t on $\mathbb{R}_+$. Moreover, the point measure can be represented by

$$\mathbb{1}_{t>0}\ N(\mathrm{d}t) = \sum_{\substack{(T,X)\in\Pi\ /\\ X\le\lambda_t}} \delta_T(\mathrm{d}t) = \left(\int_{x=0}^{\lambda_t} \Pi\left(\mathrm{d}t, \mathrm{d}x\right)\right). \tag{2.8}$$

NB: The last equality comes from the following convention. If $\delta_{(c,d)}$ is a Dirac mass in $(c, d) \in \mathbb{R}_+^2$, then $\int_{x=a}^{b} \delta_{(c,d)}(\mathrm{d}t, \mathrm{d}x)$, as a distribution in t, is $\delta_c(\mathrm{d}t)$ if $d \in [a, b]$ and 0 otherwise.

2.4 From point processes to PDE

Let us now present our main results. Informally, we want to describe the evolution of the distribution in s of the age S_t according to the time t. Note that at fixed time t, $S_{t-} = S_t$ a.s. and therefore it is the same as the distribution of S_{t-}. We prefer to study S_{t-} since its predictability, i.e. its dependence in $N \cap (-\infty, t)$, makes all definitions proper from a microscopic/random point of view. Microscopically, the interest lies in the evolution of $\delta_{S_{t-}}(\mathrm{d}s)$ as a random measure. But it should also be seen as a distribution in time, for equations like (PPS) to make sense. Therefore, we need to go from a distribution only in s

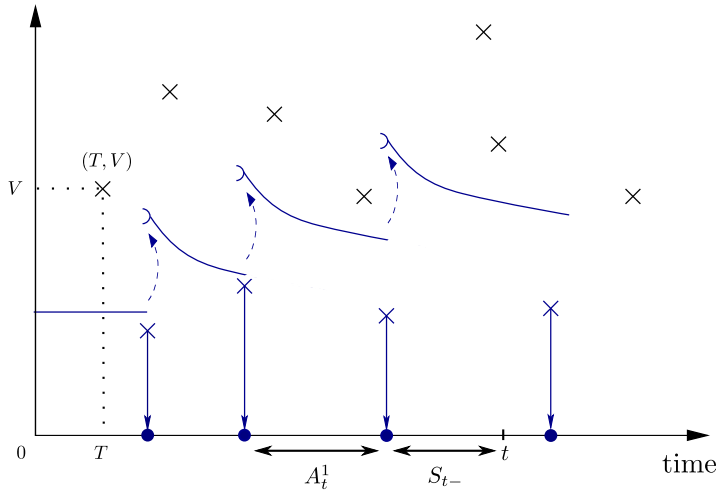


Figure 2.1: Example of Ogata's thinning algorithm on a linear Hawkes process with interaction function $h(u) = e^{-u}$ and no point before 0 (i.e. $N_- = \emptyset$). The crosses represent a realization of Π, Poisson process of intensity 1 on $\mathbb{R}_+^2$. The blue piecewise continuous line represents the intensity λ_t, which starts in 0 with value μ and then jumps each time a point of Π is present underneath it. The resulting Hawkes process (with intensity λ_t) is given by the blue circles. Age S_{t-} at time t and the quantity A_t^1 are also represented.

to a distribution in both s and t. Then one can either focus on the microscopic level, where the realisation of Π in Ogata's thinning construction is fixed or focus on the expectation of such a distribution.

2.4.a) A clean setting for bivariate distributions in age and time

In order to obtain, from a point process, (PPS) system we need to define bivariate distributions in s and t and marginals (at least in s), in such a way that weak solutions of (PPS) are correctly defined. Since we want to possibly consider more than two variables for generalized Wold processes, we consider the following definitions.

In the following, $< \varphi, \nu >$ denotes the integral of the integrable function φ with respect to the measure ν.

Let $k \in \mathbb{N}$. For every bounded measurable function φ of $(t, s, a_1, ..., a_k) \in \mathbb{R}_+^{k+2}$, one can define

$$\varphi_t^{(1)}(s, a_1, ..., a_k) = \varphi(t, s, a_1, ..., a_k) \quad \text{and} \quad \varphi_s^{(2)}(t, a_1, ..., a_k) = \varphi(t, s, a_1, ..., a_k).$$

Let us now define two sets of regularities for φ.

$\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$ | The function φ belongs to $\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$ if
- φ is a measurable bounded function,
- there exists $T > 0$ such that for all $t > T$, $\varphi_t^{(1)} = 0$.

$\mathcal{C}_{c,b}^{\infty}(\mathbb{R}_+^{k+2})$ | The function φ belongs to $\mathcal{C}_{c,b}^{\infty}(\mathbb{R}_+^{k+2})$ if
- φ is continuous, uniformly bounded,
- φ has uniformly bounded derivatives of every order,
- there exists $T > 0$ such that for all $t > T$, $\varphi_t^{(1)} = 0$.

Let $(\nu_1^t)_{t\geq 0}$ be a (measurable w.r.t. t) family of positive measures on $\mathbb{R}_+^{k+1}$, and $(\nu_2^s)_{s\geq 0}$ be a (measurable w.r.t. s) family of positive measures $\mathbb{R}_+^{k+1}$. Those families satisfy the Fubini property if

$$(\mathcal{P}_{Fubini}) \quad \text{for all } \varphi \in \mathcal{M}_{c,b}(\mathbb{R}_+^{k+2}), \qquad \int \langle \varphi_t^{(1)}, \nu_1^t \rangle \mathrm{d}t = \int \langle \varphi_s^{(2)}, \nu_2^s \rangle \mathrm{d}s.$$

In this case, one can define ν, measure on $\mathbb{R}_+^{k+2}$, by the unique measure on $\mathbb{R}_+^{k+2}$ such that for any test function φ in $\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$,

$$\langle \varphi, \nu \rangle = \int \langle \varphi_t^{(1)}, \nu_1^t \rangle \mathrm{d}t = \int \langle \varphi_s^{(2)}, \nu_2^s \rangle \mathrm{d}s.$$

To simplify notations, for any such measure $\nu(t, \mathrm{d}s, \mathrm{d}a_1, ..., \mathrm{d}a_k)$, we define

$$\begin{cases} \nu(t, \mathrm{d}s, \mathrm{d}a_1, ..., \mathrm{d}a_k) = \nu_1^t(\mathrm{d}s, \mathrm{d}a_1, ..., \mathrm{d}a_k), \\ \nu(\mathrm{d}t, s, \mathrm{d}a_1, ..., \mathrm{d}a_k) = \nu_2^s(\mathrm{d}t, \mathrm{d}a_1, ..., \mathrm{d}a_k). \end{cases}$$

In the sequel, we need in particular a measure on $\mathbb{R}_+^2$, η_x, defined for any real x by its marginals that satisfy $(\mathcal{P}_{Fubini})$ as follows

$$\forall\, t, s \geq 0, \qquad \eta_x(t, \mathrm{d}s) = \delta_{t-x}(\mathrm{d}s)\mathbb{1}_{t-x>0} \quad \text{and} \quad \eta_x(\mathrm{d}t, s) = \delta_{s+x}(\mathrm{d}t)\mathbb{1}_{s\geq 0}. \tag{2.9}$$

It represents a Dirac mass "travelling" on the positive diagonal originated in $(x, 0)$.

2.4.b) The microscopic construction of a random PDE

For a fixed realization of Π, we therefore want to define a random distribution $U(\mathrm{d}t, \mathrm{d}s)$ in terms of its marginals, thanks to $(\mathcal{P}_{Fubini})$, such that, $U(t, \mathrm{d}s)$ represents the distribution at time $t > 0$ of the age S_{t-}, i.e.

$$\forall\, t > 0, \quad U(t, \mathrm{d}s) = \delta_{S_{t-}}(\mathrm{d}s) \tag{2.10}$$

and satisfies similar equations as (PPS). This is done in the following proposition.

Proposition 2.4.1. *Let Π, $\mathcal{F}_0^N$ and an intensity $(\lambda_t)_{t>0}$ be given as in Section 2.3.c) and satisfying $(\mathcal{A}_{T_0})$ and $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,a.s.}\right)$. On the event Ω of probability 1, where Ogata's thinning is well defined, let N be the point process on $\mathbb{R}$ that is constructed thanks to Ogata's thinning with associated predictable age process $(S_{t-})_{t>0}$ and whose points are denoted $(T_i)_{i\in\mathbb{Z}}$.*

Let the (random) measure U and its corresponding marginals be defined by

$$U(\mathrm{d}t, \mathrm{d}s) = \sum_{i=0}^{+\infty} \eta_{T_i}(\mathrm{d}t, \mathrm{d}s)\ \mathbb{1}_{0\leq t\leq T_{i+1}}. \tag{2.11}$$

Then, on Ω, U satisfies $(\mathcal{P}_{Fubini})$ and $U(t, \mathrm{d}s) = \delta_{S_{t-}}(\mathrm{d}s)$. Moreover, on Ω, U is a solution in the weak sense of the following system

$$\frac{\partial}{\partial t}U(\mathrm{d}t, \mathrm{d}s) + \frac{\partial}{\partial s}U(\mathrm{d}t, \mathrm{d}s) + \left(\int_{x=0}^{\lambda_t} \Pi(\mathrm{d}t, \mathrm{d}x)\right) U(t, \mathrm{d}s) = 0, \tag{2.12}$$

$$U(\mathrm{d}t, 0) = \int_{s\in\mathbb{R}_+} \left(\int_{x=0}^{\lambda_t} \Pi(\mathrm{d}t, \mathrm{d}x)\right) U(t, \mathrm{d}s) + \delta_0(\mathrm{d}t)\mathbb{1}_{T_0=0}, \tag{2.13}$$

$$U(0, \mathrm{d}s) = \delta_{-T_0}(\mathrm{d}s)\mathbb{1}_{T_0<0} = U^{in}(\mathrm{d}s)\mathbb{1}_{s>0}, \tag{2.14}$$

where $U^{in}(ds) = \delta_{-T_0}(ds)$. *The weak sense means that for any* $\varphi \in \mathcal{C}^{\infty}_{c,b}(\mathbb{R}^2_+)$,

$$\int_{\mathbb{R}_+\times\mathbb{R}_+} \left(\frac{\partial}{\partial t}\varphi(t,s) + \frac{\partial}{\partial s}\varphi(t,s)\right) U(\mathrm{d}t,\mathrm{d}s) + \int_{\mathbb{R}_+\times\mathbb{R}_+} [\varphi(t,0) - \varphi(t,s)] \left(\int_{x=0}^{\lambda_t} \Pi(\mathrm{d}t,\mathrm{d}x)\right) U(t,\mathrm{d}s) + \varphi(0,-T_0) = 0. \quad (2.15)$$

The proof of Proposition 2.4.1 is included in Section 2.7.a). Note also that thanks to the Fubini property, the boundary condition (2.13) is satisfied also in a strong sense.

System (2.12)–(2.14) is a random microscopic version of (PPS) if $T_0 < 0$, where $n(s,t)$ the density of the age at time t is replaced by $U(t,\cdot) = \delta_{S_t-}$, the Dirac mass in the age at time t. The assumption $T_0 < 0$ is satisfied *a.s.* if T_0 has a density, but this may not be the case for instance if the experimental device gives an impulse at time zero (e.g. [127] studied Peristimulus time histograms (PSTH), where the spike trains are locked on a stimulus given at time 0).

This result may seem rather poor from a PDE point of view. However, since this equation is satisfied at a microscopic level, we are able to define correctly all the important quantities at a macroscopic level. Indeed, the analogy between $p(s,X(t))$ and λ_t is actually on the random microscopic scale a replacement of $p(s,X(t))$ by $\int_{x=0}^{\lambda_t}\Pi(\mathrm{d}t,\mathrm{d}x)$, whose expectancy given the past is, heuristically speaking, equal to λ_t because the mean behaviour of Π is given by the Lebesgue measure (see (2.7)). Thus, the main question at this stage is : can we make this argument valid by taking the expectation of U? This is addressed in the next section.

The property $(\mathcal{P}_{Fubini})$ and the quantities η_{T_i} mainly allows to define $U(\mathrm{d}t,0)$ as well as $U(t,\mathrm{d}s)$. As expected, with this definition, (2.10) holds as well as

$$U(\mathrm{d}t,0) = \mathbb{1}_{t\geq 0}\ N(\mathrm{d}t), \quad (2.16)$$

i.e. the spiking measure (the measure in time with age 0) is the point measure.

Note also that the initial condition is given by $\mathcal{F}_0^N$, since $\mathcal{F}_0^N$ fixes in particular the value of T_0 and $(\mathcal{A}_{T_0})$ is required to give sense to the age at time 0. To understand the initial condition, remark that if $T_0 = 0$, then $U(0,\cdot) = 0 \neq \lim_{t\to 0^+} U(t,\cdot) = \delta_0$ by definitions of η_{T_i} but that if $T_0 < 0$, $U(0,\cdot) = \lim_{t\to 0^+} U(t,\cdot) = \delta_{-T_0}$.

The mass-conservativeness (i.e. for all $t \geq 0$, $\int_0^{+\infty} U(t,\mathrm{d}s) = 1$) is obtained by using (a sequence of test functions converging to) $\varphi = \mathbb{1}_{t\leq T}$.

Proposition 2.4.1 shows that the (random) measure U, defined by (2.11), in terms of a given point process N, is a weak solution of System (2.12)-(2.14). The study of the well-posedness of this system could be addressed following, for instance, the ideas given in [26]. In this case U should be the unique solution of system (2.12)–(2.14).

As last comment about Proposition 2.4.1, we analyse the particular case of the linear Hawkes process, in the following remark.

Remark 2.4.2. *In the particular case of the linear Hawkes process,*

$$\lambda_t = \mu + \int_{-\infty}^{t-} h(t-z)N(\mathrm{d}z).$$

Thanks to (2.16) *one decomposes the intensity into a term given by the initial condition plus a term given by the measure* U*,*

$$\lambda_t = \mu + F_0(t) + \int_0^{t-} h(t-z)U(\mathrm{d}z,0),$$

where $F_0(t) = \int_{-\infty}^0 h(t-z)N_-(\mathrm{d}z)$ *is* $(\mathcal{F}_0^N)$*-measurable and considered as an initial condition. Hence,* (2.12)–(2.14) *becomes a closed system in the sense that* λ_t *is now an explicit function of the solution of the system. This is not true in general.*

2.4.c) The PDE satisfied in expectation

In this section, we want to find the system satisfied by the expectation of the random measure U. First, we need to give a proper definition of such an object. The construction is based on the construction of U and is summarized in the following proposition. (The proofs of all the results of this subsection are in Section 2.7.a)).

Proposition 2.4.3. *Let* Π, $\mathcal{F}_0^N$ *and an intensity* $(\lambda_t)_{t>0}$ *be given as in Section 2.3.c) and satisfying* $(\mathcal{A}_{T_0})$ *and* $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,exp}\right)$*. Let* N *be the process resulting of Ogata's thinning and let* U *be the random measure defined by* (2.11)*. Let* $\mathbb{E}$ *denote the expectation with respect to* Π *and* $\mathcal{F}_0^N$*.*

Then for any test function φ *in* $\mathcal{M}_{c,b}(\mathbb{R}_+^2)$*, both expectations* $\mathbb{E}\left[\int \varphi(t,s)U(t,\mathrm{d}s)\right]$ *and* $\mathbb{E}\left[\int \varphi(t,s)U(\mathrm{d}t,s)\right]$ *are finite and one can define* $u(t,\mathrm{d}s)$ *and* $u(\mathrm{d}t,s)$ *by*

$$\begin{cases} \forall\, t \geq 0, & \displaystyle\int \varphi(t,s)u(t,\mathrm{d}s) = \mathbb{E}\left[\int \varphi(t,s)U(t,\mathrm{d}s)\right], \\ \forall\, s \geq 0, & \displaystyle\int \varphi(t,s)u(\mathrm{d}t,s) = \mathbb{E}\left[\int \varphi(t,s)U(\mathrm{d}t,s)\right]. \end{cases}$$

Moreover, $u(t,\mathrm{d}s)$ *and* $u(\mathrm{d}t,s)$ *satisfy* $(\mathcal{P}_{Fubini})$ *and one can define* $u(\mathrm{d}t,\mathrm{d}s) = u(t,\mathrm{d}s)\mathrm{d}t = u(\mathrm{d}t,s)\mathrm{d}s$ *on* $\mathbb{R}_+^2$*, such that for any test function* φ *in* $\mathcal{M}_{c,b}(\mathbb{R}_+^2)$*,*

$$\int \varphi(t,s)u(\mathrm{d}t,\mathrm{d}s) = \mathbb{E}\left[\int \varphi(t,s)U(\mathrm{d}t,\mathrm{d}s)\right],$$

quantity which is finite.

In particular, since $\int \varphi(t,s)u(t,\mathrm{d}s) = \mathbb{E}\left[\int \varphi(t,s)U(t,\mathrm{d}s)\right] = \mathbb{E}\left[\varphi(t,S_{t-})\right]$, $u(t,\cdot)$ is therefore the distribution of S_{t-}, the (predictable version of the) age at time t. Now let us show that as expected, u satisfies a system similar to (PPS).

Theorem 2.4.4. *Let* Π, $\mathcal{F}_0^N$ *and an intensity* $(\lambda_t)_{t>0}$ *be given as in Section 2.3.c) and satisfying* $(\mathcal{A}_{T_0})$ *and* $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,exp}\right)$*. If* N *is the process resulting of Ogata's thinning,* $(S_{t-})_{t>0}$ *its associated predictable age process,* U *its associated random measure, defined by* (2.11)*, and* u *its associated mean measure, defined in Proposition 2.4.3, then, there exists a bivariate measurable function* $\rho_{\lambda,\zeta_{N_-}}$ *satisfying*

$$\begin{cases} \forall\, T \geq 0, \displaystyle\int_0^T \int_s \rho_{\lambda,\zeta_{N_-}}(t,s)u(\mathrm{d}t,\mathrm{d}s) < +\infty, \\ \rho_{\lambda,\zeta_{N_-}}(t,s) = \mathbb{E}\left[\lambda\left(t,\mathcal{F}_{t-}^N\right)\middle| S_{t-} = s\right] \quad u(\mathrm{d}t,\mathrm{d}s)\text{-a.e} \end{cases} \tag{2.17}$$

and such that u *is solution in the weak sense of the following system*

$$\frac{\partial}{\partial t}u\left(\mathrm{d}t,\mathrm{d}s\right) + \frac{\partial}{\partial s}u\left(\mathrm{d}t,\mathrm{d}s\right) + \rho_{\lambda,\zeta_{N_-}}(t,s)u\left(\mathrm{d}t,\mathrm{d}s\right) = 0, \tag{2.18}$$

$$u\left(\mathrm{d}t,0\right) = \int_{s\in\mathbb{R}_+} \rho_{\lambda,\zeta_{N_-}}(t,s)u\left(t,\mathrm{d}s\right)\,\mathrm{d}t + \delta_0(\mathrm{d}t)u^{in}(\{0\}), \tag{2.19}$$

$$u\left(0,\mathrm{d}s\right) = u^{in}(\mathrm{d}s)\mathbb{1}_{s>0}, \tag{2.20}$$

where u^{in} is the law of $-T_0$. The weak sense means here that for any φ in $\mathcal{C}^\infty_{c,b}(\mathbb{R}^2_+)$,

$$\int_{\mathbb{R}_+\times\mathbb{R}_+}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial s}\right)\varphi(t,s)\,u(\mathrm{d}t,\mathrm{d}s)+\int_{\mathbb{R}_+\times\mathbb{R}_+}[\varphi(t,0)-\varphi(t,s)]\rho_{\lambda,\zeta_{N_-}}(t,s)u(\mathrm{d}t,\mathrm{d}s)+\int_{\mathbb{R}_+}\varphi(0,s)u^{in}(\mathrm{d}s)=0, \quad (2.21)$$

Comparing this system to (PPS), one first sees that $n(t,\cdot)$, the density of the age at time t, is replaced by the mean measure $u(t,\cdot)$. If $u^{in}\in L^1(\mathbb{R}_+)$ we have $u^{in}(\{0\})=0$ so we get an equation which is exactly of renewal type, as (PPS). In the general case where u^{in} is only a probability measure, the difference with (PPS) lies in the term $\delta_0(\mathrm{d}t)u^{in}(\{0\})$ in the boundary condition for $s=0$ and in the term $\mathbb{1}_{s>0}$ in the initial condition for $t=0$. Both these extra terms are linked to the possibility for the initial measure u^{in} to charge zero. This possibility is not considered in [114] - else, a similar extra term would be needed in the setting of [114] as well. As said above in the comment of Proposition 2.4.1, we want to keep this term here since it models the case where there is a specific stimulus at time zero [127].

In general and without more assumptions on λ, it is not clear that u is not only a measure satisfying ($\mathcal{P}_{Fubini}$) but also absolutely continuous w.r.t. to $\mathrm{d}t\ \mathrm{d}s$ and that the equations can be satisfied in a strong sense.

Concerning $p(s,X(t))$, which has always been thought of as the equivalent of λ_t, it is not replaced by λ_t, which would have no meaning in general since this is a random quantity, nor by $\mathbb{E}[\lambda_t]$ which would have been a first possible guess; it is replaced by $\mathbb{E}[\lambda_t|S_{t-}=s]$. Indeed intuitively, since

$$\mathbb{E}\left[\int_{x=0}^{\lambda_t}\Pi(\mathrm{d}t,\mathrm{d}x)\middle|\mathcal{F}^N_{t-}\right]=\lambda_t\mathrm{d}t,$$

the corresponding weak term can be interpreted as, for any test function φ,

$$\begin{aligned}\mathbb{E}\left[\int\varphi(t,s)\left(\int_{x=0}^{\lambda_t}\Pi(\mathrm{d}t,\mathrm{d}x)\right)U(t,\mathrm{d}s)\right]&\\ =\mathbb{E}\left[\int\varphi(t,s)\,\lambda_t\delta_{S_{t-}}(\mathrm{d}s)\mathrm{d}t\right]&=\int_t\mathbb{E}[\varphi(t,S_{t-})\,\lambda_t]\,\mathrm{d}t\\ &=\int_t\mathbb{E}[\varphi(t,S_{t-})\,\mathbb{E}[\lambda_t|S_{t-}]]\,\mathrm{d}t,\end{aligned}$$

which is exactly $\int\varphi(t,s)\rho_{\lambda,\zeta_{N_-}}(t,s)u(\mathrm{d}t,\mathrm{d}s)$.

This conditional expectation makes dependencies particularly complex, but this also enables to derive equations even in non-Markovian setting (as Hawkes processes for instance, see Section 2.5). More explicitly, $\rho_{\lambda,\zeta_{N_-}}(t,s)$ is a function of the time t, of the age s, but it also depends on λ, the shape of the intensity of the underlying process and on the distribution of the initial condition N_-, that is ζ_{N_-}. As explained in Section 2.3, it is both the knowledge of ζ_{N_-} and λ that characterizes the distribution of the process and in general the conditional expectation cannot be reduced to something depending on less than that. In Section 2.5, we discuss several examples of point processes where one can (or cannot) reduce the dependence.

Note that here again, we can prove that the equation is mass-conservative by taking (a sequence of functions converging to) $\varphi=\mathbb{1}_{t\le T}$ as a test function.

A direct corollary of Theorem 2.4.4 can be deduced thanks to the law of large numbers. This can be seen as the interpretation of (PPS) equation at a macroscopic level, when the population of neurons is i.i.d..

Corollary 2.4.5. *Let $(N^i)_{i=1}^{\infty}$ be some i.i.d. point processes with respective intensities given by λ_t^i on $(0,+\infty)$ satisfying $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,exp}\right)$ and associated predictable age processes $(S_{t-}^i)_{t>0}$. Suppose furthermore that the distribution of N^1 on $(-\infty,0]$ is given by ζ_{N_-} which is such that $\zeta_{N_-}(N_-^1=\emptyset)=0$.*

Then there exists a measure u satisfying $(\mathcal{P}_{Fubini})$, weak solution of Equations (2.18) and (2.19), with $\rho_{\lambda,\zeta_{N_-}}$ defined by

$$\rho_{\lambda,\zeta_{N_-}}(t,s)=\mathbb{E}\left[\lambda_t^1|S_{t-}^1=s\right],\quad u(\mathrm{d}t,\mathrm{d}s)\text{-a.e.}$$

and with u^{in} distribution of the age at time 0, such that, for any $\varphi\in\mathcal{C}_{c,b}^{\infty}(\mathbb{R}_+^2)$,

$$\forall\, t>0,\quad \int\varphi(t,s)\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{S_t^i}(\mathrm{d}s)\right)\xrightarrow[n\to+\infty]{a.s.}\int\varphi(t,s)u(t,\mathrm{d}s),\tag{2.22}$$

In particular, informally, the fraction of neurons at time t with age in $[s,s+\mathrm{d}s)$ in this i.i.d. population of neurons indeed tends to $u(t,\mathrm{d}s)$.

2.5 Application to the various examples

Let us now apply these results to the examples presented in Section 2.3.b).

2.5.a) When the intensity only depends on time and age

If $\lambda_t=f(t,S_{t-})$ (homogeneous and inhomogeneous Poisson processes and renewal processes are particular examples) then the intuition giving that $p(s,X(t))$ is analogous to λ_t works. Let us assume that $f(t,s)\in L^{\infty}(\mathbb{R}_+^2)$. We have $\mathbb{E}\left[\lambda_t|S_{t-}=s\right]=f(t,s)$. Under this assumption, we may apply Theorem 2.4.4, so that we know that the mean measure u associated to the random process is a solution of System (2.18)–(2.20). Therefore the mean measure u satisfies a completely explicit PDE of the type (PPS) with $\rho_{\lambda,\zeta_{N_-}}(t,s)=f(t,s)$ replacing $p(s,X(t))$. In particular, in this case $\rho_{\lambda,\zeta_{N_-}}(t,s)$ depends on the initial distribution ζ_{N_-} only through the distribution of $-T_0$, that is the initial condition u^{in}. Since $f(\cdot,\cdot)\in L^{\infty}\big([0,T]\times\mathbb{R}_+\big)$, assuming also $u^{in}\in L^1(\mathbb{R}_+)$, it is well-known that there exists a unique solution u such that $(t\mapsto u(t,\cdot))\in\mathcal{C}\big([0,T],L^1(\mathbb{R}_+)\big)$, see for instance [121] Section 3.3. p.60.

Note that, following [26], uniqueness for measure solutions may also be established, hence the mean measure u associated to the random process is the unique solution of System (2.18)–(2.20), and it is in $\mathcal{C}\big([0,T],L^1(\mathbb{R}_+)\big)$: the PDE formulation, together with existence and uniqueness, has provided a regularity result on u which is obtained under weaker assumptions than through Fokker-Planck / Kolmogorov equations. This is another possible application field of our results: using the PDE formulation to gain regularity. Let us now develop the Fokker-Planck / Kolmogorov approach for renewal processes.

Renewal processes The renewal process, i.e. when $\lambda_t = f(S_{t-})$, with f a continuous function on $\mathbb{R}_+$, has particular properties. As noted in Section 2.3.b), the renewal age process $(S_{t-})_{t>0}$ is an homogeneous Markovian process. It is known for a long time that it is easy to derive PDE on the corresponding density through Fokker-Planck / Kolmogorov equations, once the variable of interest (here the age) is Markovian (see for instance [14]). Here we briefly follow this line to see what kind of PDE can be derived through the Markovian properties and to compare the equation with the (PPS) type system derived in Theorem 2.4.4.

Since f is continuous, the infinitesimal generator[8] of $(S_t)_{t>0}$ is given by

$$(\mathcal{G}\phi)(x) = \phi'(x) + f(x)\left(\phi(0) - \phi(x)\right), \tag{2.23}$$

for all $\phi \in \mathcal{C}^1(\mathbb{R}_+)$ (see [15]). Note that, since for every $t > 0$ $S_{t-} = S_t$ a.s., the process $(S_{t-})_{t>0}$ is also Markovian with the same infinitesimal generator.

Let us now define for all $t > 0$ and all $\phi \in \mathcal{C}^1(\mathbb{R}_+)$,

$$P_t\phi(x) = \mathbb{E}\left[\phi(S_{t-})|S_0 = x\right] = \int \phi(s)u_x(t, \mathrm{d}s),$$

where $x \in \mathbb{R}_+$ and $u_x(t, \cdot)$ is the distribution of S_{t-} given that $S_0 = x$. Note that $u_x(t, \mathrm{d}s)$ corresponds to the marginal in the sense of $(\mathcal{P}_{Fubini})$ of the mean measure u_x given by Theorem 2.4.4 with $\rho_{\lambda,\zeta_{N_-}}(t, s) = f(s)$ and initial condition δ_x, i.e. $T_0 = -x$ a.s.

In this homogeneous Markovian case, the forward Kolmogorov equation gives

$$\frac{\partial}{\partial t}P_t = P_t\mathcal{G}.$$

Let $\varphi \in \mathcal{C}^\infty_{c,b}(\mathbb{R}^2_+)$ and let $t > 0$. This implies that

$$\begin{aligned}\frac{\partial}{\partial t}\left(P_t\varphi(t, s)\right) &= P_t\mathcal{G}\varphi(t, s) + P_t\frac{\partial}{\partial t}\varphi(t, s)\\ &= P_t\left[\frac{\partial}{\partial s}\varphi(t, s) + f(s)\left(\varphi(t, 0) - \varphi(t, s)\right) + \frac{\partial}{\partial t}\varphi(t, s)\right].\end{aligned}$$

Since φ is compactly supported in time, an integration with respect to t yields

$$-P_0\varphi(0, s) = \int P_t\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\varphi(t, s)\mathrm{d}t + \int P_t f(s)\left(\varphi(t, 0) - \varphi(t, s)\right)\mathrm{d}t,$$

or equivalently

$$-\varphi(0, x) = \int\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\varphi(t, s)\, u_x(t, \mathrm{d}s)\,\mathrm{d}t - \int\left(\varphi(t, s) - \varphi(t, 0)\right)f(s)u_x(t, \mathrm{d}s)\mathrm{d}t, \tag{2.24}$$

in terms of u_x. This is exactly Equation (2.21) with $u^{in} = \delta_x$.

The result of Theorem 2.4.4 is stronger than the application of the forward Kolmogorov equation on homogeneous Markovian systems since the result of Theorem 2.4.4 never used

[8]The infinitesimal generator of an homogeneous Markov process $(Z_t)_{t\geq 0}$ is the operator $\mathcal{G}$ which is defined to act on every function $\phi : \mathbb{R}^n \to \mathbb{R}$ in a suitable space $\mathcal{D}$ by

$$\mathcal{G}\phi(x) = \lim_{t\to 0^+}\frac{\mathbb{E}\left[\phi(Z_t)|Z_0 = x\right] - \phi(x)}{t}.$$

the Markov assumption and can be applied to non Markovian processes (see Section 2.5.c)). So the present work is a general set-up where one can deduce PDE even from non Markovian microscopic random dynamics. Note also that only boundedness assumptions and not continuity ones are necessary to directly obtain (2.24) via Theorem 2.4.4: to obtain the classical Kolmogorov theorem, one would have assumed $f \in \mathcal{C}^0(\mathbb{R}_+^2)$ rather than $f \in L^\infty(\mathbb{R}_+^2)$.

2.5.b) Generalized Wold process

In the case where $\lambda_t = f(S_{t-}, A_t^1, ..., A_t^k)$, with f being a non-negative function, one can define in a similar way $u_k(t, s, a_1, \dots, a_k)$ which is informally the distribution at time t of the processes with age s and past given by $a_1, ...a_k$ for the last k ISI's. We want to investigate this case not for its Markovian properties, which are nevertheless presented in Proposition 2.8.2 in the Proof Section for sake of completeness, but because this is the first basic example where the initial condition is indeed impacting $\rho_{\lambda,\zeta_{N_-}}$ in Theorem 2.4.4.

To do so, the whole machinery applied on $u(\mathrm{d}t, \mathrm{d}s)$ is first extended in the next result to $u_k\left(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a^1, \dots, \mathrm{d}a^k\right)$ which represents the dynamics of the age and the last k ISI's. This could have been done in a very general way by an easy generalisation of Theorem 2.4.4. However to avoid too cumbersome equations, we express it only for generalized Wold processes to provide a clean setting to illustrate the impact of the initial conditions on $\rho_{\lambda,\zeta_{N_-}}$. Hence, we similarly define a random distribution $U_k(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, \dots, \mathrm{d}a_k)$ such that its evaluation at any given time t exists and is

$$U_k(t, \mathrm{d}s, \mathrm{d}a_1, \dots, \mathrm{d}a_k) = \delta_{(S_{t-}, A_t^1, ..., A_t^k)}(\mathrm{d}s, \mathrm{d}a_1, \dots, \mathrm{d}a_k). \tag{2.25}$$

The following result states the PDE satisfied by $u_k = \mathbb{E}[U_k]$.

Proposition 2.5.1. *Let k be a positive integer and f be some non negative function on $\mathbb{R}_+^{k+1}$. Let N be a generalized Wold process with predictable age process $(S_{t-})_{t>0}$, associated points $(T_i)_{i\in\mathbb{Z}}$ and intensity $\lambda_t = f(S_{t-}, A_t^1, ..., A_t^k)$ satisfying $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,exp}\right)$, where $A_t^1, \dots, A_t^k$ are the successive ages defined by (2.5). Suppose that ζ_{N_-} is such that $\zeta_{N_-}(T_{-k} > -\infty) = 1$. Let U_k be defined by*

$$U_k(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, \dots, \mathrm{d}a_k) = \sum_{i=0}^{+\infty} \eta_{T_i}(\mathrm{d}t, \mathrm{d}s) \prod_{j=1}^{k} \delta_{A_{T_i}^j}(\mathrm{d}a_j)\ \mathbb{1}_{0\le t\le T_{i+1}}, \tag{2.26}$$

If N is the result of Ogata's thinning on the Poisson process Π, then U_k satisfies (2.25) and $(\mathcal{P}_{Fubini})$ a.s. in Π and $\mathcal{F}_0^N$. Assume that the initial condition u_k^{in}, defined as the distribution of $(-T_0, A_0^1, \dots, A_0^k)$ which is a random vector in $\mathbb{R}^{k+1}$, is such that $u_k^{in}(\{0\} \times \mathbb{R}_+^k) = 0$. Then U_k admits a mean measure u_k which also satisfies $(\mathcal{P}_{Fubini})$ and the following system in the weak sense: on $\mathbb{R}_+ \times \mathbb{R}_+^{k+1}$,

$$\Big\{\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big\} u_k(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, ..., \mathrm{d}a_k) + f(s, a_1, ..., a_k) u_k(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, ..., \mathrm{d}a_k) = 0, \tag{2.27}$$

$$u_k(\mathrm{d}t, 0, \mathrm{d}s, \mathrm{d}a_1, ..., \mathrm{d}a_{k-1}) = \int_{a_k=0}^{+\infty} f(s, a_1, ..., a_k)\, u_k(t, \mathrm{d}s, \mathrm{d}a_1, ..., \mathrm{d}a_k)\ \mathrm{d}t, \tag{2.28}$$

$$u_k(0, \mathrm{d}s, \mathrm{d}a_1, \dots, \mathrm{d}a_k) = u_k^{in}(\mathrm{d}s, \mathrm{d}a_1, \dots, \mathrm{d}a_k). \tag{2.29}$$

We have assumed $u_k^{in}(\{0\}\times\mathbb{R}_+^k) = 0$ (i.e. $T_0 \neq 0$ a.s.) for the sake of simplicity, but this assumption may of course be relaxed and Dirac masses at 0 should then be added in a similar way as in Theorem 2.4.4.

If $f \in L^\infty(\mathbb{R}_+^{k+1})$, we may apply Proposition 2.5.1, so that the mean measure u_k satisfy System (2.27)–(2.29). Assuming an initial condition $u_k^{in} \in L^1(\mathbb{R}_+^{k+1})$, we can prove exactly as for the renewal equation (with a Banach fixed point argument for instance) that there exists a unique solution u_k such that $(t \mapsto u_k(t,\cdot)) \in \mathcal{C}\big(\mathbb{R}_+, L^1(\mathbb{R}_+^{k+1})\big)$ [121] to the generalized Wold case, the boundary assumption on the kth penultimate point before time 0 being necessary to give sense to the successive ages at time 0. By uniqueness, this proves that the mean measure u_k is this solution, so that it belongs to $\mathcal{C}\big(\mathbb{R}_+, L^1(\mathbb{R}_+^{k+1})\big)$: Proposition 2.5.1 leads to a regularity result on the mean measure.

Now that we have clarified the dynamics of the successive ages, one can look at this system from the point of view of Theorem 2.4.4, that is when only two variables s and t are considered. In this respect, let us note that U defined by (2.11) is such that

$$U(\mathrm{d}t, \mathrm{d}s) = \int_{a_1,\dots,a_k} U_k(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, \dots, \mathrm{d}a_k).$$

Since the integrals and the expectations are exchangeable in the weak sense, the mean measure u defined in Proposition 2.4.3 is such that

$$u(\mathrm{d}t, \mathrm{d}s) = \int_{a_1,\dots,a_k} u_k(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, \dots, \mathrm{d}a_k).$$

But (2.27) in the weak sense means, for all $\varphi \in \mathcal{C}_{c,b}^\infty(\mathbb{R}^{k+2})$,

$$\begin{aligned}\int \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \varphi(t,s,a_1,...,a_k) u_k\left(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, \dots, \mathrm{d}a_k\right) \\ + \int \left[\varphi\left(t,0,a_1,\dots,a_k\right) - \varphi(t,s,a_1,\dots,a_k)\right] f\left(s,a_1,\dots,a_k\right) u_k\left(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, \dots, \mathrm{d}a_k\right) \\ + \int \varphi\left(0,s,a_1,\dots,a_k\right) u_k^{in}\left(\mathrm{d}s, \mathrm{d}a_1, \dots, \mathrm{d}a_k\right) = 0. \quad (2.30)\end{aligned}$$

Letting $\psi \in \mathcal{C}_{c,b}^\infty(\mathbb{R}^2)$ and $\varphi \in \mathcal{C}_{c,b}^\infty(\mathbb{R}^{k+2})$ being such that

$$\forall\, t, s, a_1, \dots, a_k, \quad \varphi(t,s,a_1,\dots,a_k) = \psi(t,s),$$

we end up proving that the function $\rho_{\lambda,\zeta_{N_-}}$ defined in Theorem 2.4.4 satisfies

$$\rho_{\lambda,\zeta_{N_-}}(t,s) u\left(\mathrm{d}t, \mathrm{d}s\right) = \int_{a_1,\dots,a_k} f\left(s,a_1,\dots,a_k\right) u_k\left(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, \dots, \mathrm{d}a_k\right), \quad (2.31)$$

$u(\mathrm{d}t, \mathrm{d}s)$-a.e. Equation (2.31) means exactly from a probabilistic point of view that

$$\rho_{\lambda,\zeta_{N_-}}(t,s) = \mathbb{E}\left[f(S_{t-}, A_t^1, ..., A_t^k) | S_{t-} = s\right], \quad u(\mathrm{d}t, \mathrm{d}s)\text{-}a.e.$$

Therefore, in the particular case of generalized Wold process, the quantity $\rho_{\lambda,\zeta_{N_-}}$ depends on the shape of the intensity (here the function f) and also on u_k. But, by Proposition 2.5.1, u_k depends on its initial condition given by the distribution of $(-T_0, A_0^1, \dots, A_0^k)$, and not only $-T_0$ as in the initial condition for u. That is, as announced in the remarks

following Theorem 2.4.4, $\rho_{\lambda,\zeta_{N_-}}$ depends in particular on the whole distribution of the underlying process before time 0, namely ζ_{N_-} and not only on the initial condition for u. Here, for generalized Wold processes, it only depends on the last k points before time 0.

For more general non Markovian settings, the integration cannot be simply described by a measure u_k in dimension $(k+2)$ being integrated with respect to $\mathrm{d}a^1...\mathrm{d}a^k$. In general, the integration has to be done on all the "randomness" hidden behind the dependence of λ_t with respect to the past once S_{t-} is fixed and in this sense it depends on the whole distribution ζ_{N_-} of N_-. This is made even clearer on the following non Markovian example: the Hawkes process.

2.5.c) Hawkes process

As we have seen in Section 2.3.b), there are many different examples of Hawkes processes that can all be expressed as

$$\lambda_t = \phi\left(\int_{-\infty}^{t-} h\left(t-x\right) N(\mathrm{d}x)\right),$$

where the main case is $\phi(\theta) = \mu + \theta$, for μ some positive constant, which is the linear case.

When there is no point before 0, $\lambda_t = \phi\left(\int_0^{t-} h\left(t-x\right) N(\mathrm{d}x)\right)$. In this case, the interpretation is so close to (PPS) that the first guess, which is wrong, would be that the analogous in (PPS) is

$$p(s, X(t)) = \phi(X(t)), \tag{2.32}$$

where $X(t) = \mathbb{E}\left[\int_0^{t-} h\left(t-x\right) N(\mathrm{d}x)\right] = \int_0^t h\left(t-x\right) u(\mathrm{d}x, 0)$. This is wrong, even in the linear case since λ_t depends on all the previous points. Therefore $\rho_{\lambda,\zeta_{N_-}}$ defined by (2.17) corresponds to a conditioning given only the last point.

By looking at this problem through the generalized Wold approach, one can hope that for h decreasing fast enough:

$$\lambda_t \simeq \phi\left(h(S_{t-}) + h(S_{t-} + A_t^1) + ... + h(S_{t-} + A_t^1 + ... + A_t^k)\right).$$

In this sense and with respect to generalized Wold processes described in the previous section, we are informally integrating on "all the previous points" except the last one and not integrating over all the previous points. This is informally why (2.32) is wrong even in the linear case.

Actually, $\rho_{\lambda,\zeta_{N_-}}$ is computable for linear Hawkes processes: we show in the next section that

$$\rho_{\lambda,\zeta_{N_-}}(t,s) \neq \phi\Big(\int_{-\infty}^{t} h(t-x)u(\mathrm{d}x,0)\Big) = \mu + \int_0^{\infty} h(t-x)u(\mathrm{d}x,0)$$

and that $\rho_{\lambda,\zeta_{N_-}}$ explicitly depends on ζ_{N_-}.

Linear Hawkes process

We are interested in Hawkes processes with a past before time 0 given by $\mathcal{F}_0^N$, which is not necessarily the past given by a stationary Hawkes process. To illustrate the fact that the past is impacting the value of $\rho_{\lambda,\zeta_{N_-}}$, we focus on two particular cases:

$$\left(\mathcal{A}_{N_-}^1\right) \;\Big|\; N_- = \{T_0\} \text{ a.s. and } T_0 \text{ admits a bounded density } f_0 \text{ on } \mathbb{R}_-$$

$$\left(\mathcal{A}^2_{N_-}\right) \mid N_- \text{ is an homogeneous Poisson process with intensity } \alpha \text{ on } \mathbb{R}_-$$

Before stating the main result, we need some technical definitions. Indeed the proof is based on the underlying branching structure of the linear Hawkes process described in Section 2.8.c) and the following functions (L_s, G_s) are naturally linked to this branching decomposition (see Lemma 2.8.7).

Lemma 2.5.2. *Let $h \in L^1(\mathbb{R}_+)$ such that $\|h\|_{L^1} < 1$. For all $s \geq 0$, there exist a unique solution $(L_s, G_s) \in L^1(\mathbb{R}_+) \times L^\infty(\mathbb{R}_+)$ of the following system*

$$\log(G_s(x)) = \int_0^{(x-s)\vee 0} G_s(x-w)h(w)\mathrm{d}w - \int_0^x h(w)\mathrm{d}w, \tag{2.33}$$

$$L_s(x) = \int_{s\wedge x}^x \left(h\left(w\right) + L_s(w)\right) G_s(w)h(x-w)\,\mathrm{d}w, \tag{2.34}$$

where $a\vee b$ (resp. $a\wedge b$) denotes the maximum (resp. minimum) between a and b. Moreover, $L_s(x \leq s) \equiv 0$, $G_s : \mathbb{R}_+ \to [0,1]$, and L_s is uniformly bounded in L^1.

This result allows to define two other important quantities, K_s and q, by, for all $s, t \geq 0, z \in \mathbb{R}$,

$$K_s(t,z) := \int_0^{(t-s)\vee 0} \left[h(t-x) + L_s(t-x)\right] G_s(t-x)h(x-z)\mathrm{d}x,$$

$$\log(q(t,s,z)) := -\int_{(t-s)\vee 0}^t h(x-z)\mathrm{d}x - \int_0^{(t-s)\vee 0} \left[1 - G_s(t-x)\right] h(x-z)\mathrm{d}x. \tag{2.35}$$

Finally, the following result is just an obvious remark that helps to understand the resulting system.

Remark 2.5.3. *For a non negative $\Phi \in L^\infty(\mathbb{R}_+)$ and $v^{in} \in L^\infty(\mathbb{R}_+)$, there exists a unique solution $v \in L^\infty(\mathbb{R}_+^2)$ in the weak sense to the following system,*

$$\frac{\partial}{\partial t}v(t,s) + \frac{\partial}{\partial s}v(t,s) + \Phi(t,s)v(t,s) = 0, \tag{2.36}$$

$$v(t,0) = 1 \qquad v(t=0,s) = v^{in}(s) \tag{2.37}$$

Moreover $t \mapsto v(t,.)$ is in $\mathcal{C}(R_+, L^1_{loc}(\mathbb{R}_+))$.

If v^{in} is a survival function (i.e. non increasing from 0 to 1), then $v(t,.)$ is a survival function and $-\partial_s v$ is a probability measure for all $t > 0$.

Proposition 2.5.4. *Using the notations of Theorem 2.4.4, let N be a Hawkes process with past before 0 given by N_- satisfying either $\left(\mathcal{A}^1_{N_-}\right)$ or $\left(\mathcal{A}^2_{N_-}\right)$ and with intensity on $\mathbb{R}_+$ given by*

$$\lambda_t = \mu + \int_{-\infty}^{t-} h(t-x)N(\mathrm{d}x),$$

where μ is a positive real number and $h \in L^\infty(\mathbb{R}_+)$ is a non-negative function with support in $\mathbb{R}_+$ such that $\int h < 1$.

Then, the mean measure u defined in Proposition 2.4.3 satisfies Theorem 2.4.4 and moreover its integral $v(t,s) := \int_s^{+\infty} u(t, d\sigma)$ is the unique solution of the system (2.36)–(2.37) where v^{in} is the survival function of $-T_0$, and where $\Phi = \Phi^{\mu,h}_{\zeta_{N_-}} \in L^\infty(\mathbb{R}_+)$ is defined by

$$\Phi^{\mu,h}_{\zeta_{N_-}} = \Phi^{\mu,h}_{+} + \Phi^{h}_{-,\zeta_{N_-}}, \tag{2.38}$$

where, for all non negative s, t,

$$\Phi_{+}^{\mu,h}(t,s) = \mu\left(1 + \int_{s\wedge t}^{t} (h(x) + L_s(x))G_s(x)\mathrm{d}x\right), \tag{2.39}$$

and where, under Assumption $\left(\mathcal{A}_{N_-}^1\right)$,

$$\Phi_{-,\zeta_{N_-}}^{h}(t,s) = \frac{\int_{-\infty}^{0\wedge(t-s)} \left(h(t-t_0) + K_s(t,t_0)\right) q(t,s,t_0) f_0(t_0)\mathrm{d}t_0}{\int_{-\infty}^{0\wedge(t-s)} q(t,s,t_0) f_0(t_0)\mathrm{d}t_0}, \tag{2.40}$$

or, under Assumption $\left(\mathcal{A}_{N_-}^2\right)$,

$$\Phi_{-,\zeta_{N_-}}^{h}(t,s) = \alpha \int_{-\infty}^{0\wedge(t-s)} \left(h(t-z) + K_s(t,z)\right) q(t,s,z)\mathrm{d}z. \tag{2.41}$$

In these formulae, L_s, G_s, K_s and q are given by Lemma 2.5.2 and (2.35). Moreover

$$\forall\, s \geq 0, \quad \int_s^{+\infty} \rho_{\lambda,\zeta_{N_-}}(t,x)u(t,\mathrm{d}x) = \Phi_{\zeta_{N_-}}^{\mu,h}(t,s) \int_s^{+\infty} u(t,\mathrm{d}x). \tag{2.42}$$

The proof is included in Section 2.8.c). Proposition 2.5.4 gives a purely analytical definition for v, and thus for u, in two specific cases, namely $\left(\mathcal{A}_{N_-}^1\right)$ or $\left(\mathcal{A}_{N_-}^2\right)$. In the general case, treated in Proposition 2.8.5, there remains a dependence with respect to the initial condition ζ_{N_-}, *via* the function $\Phi_{-,\zeta_{N_-}}^{h}$.

Remark 2.5.5. *Contrarily to the general result in Theorem 2.4.4, Proposition 2.5.4 focuses on the equation satisfied by $v(\mathrm{d}t, s) = \int_s^{+\infty} u(\mathrm{d}t, \mathrm{d}x)$ because in Equation (2.36) the function parameter $\Phi = \Phi_{\zeta_{N_-}}^{\mu,h}$ may be defined independently of the definitions of v or u, which is* not *the case for the rate $\rho_{\lambda,\zeta_{N_-}}$ appearing in Equation (2.18). Thus, it is possible to depart from the system of equations defining v, study it, prove existence, uniqueness and regularity for v under some assumptions on the initial distribution u^{in} as well as on the birth function h, and then deduce regularity or asymptotic properties for u without any previous knowledge on the underlying process.*

In Sections 2.5.a) and 2.5.b), we were able to use the PDE formulation to prove that the distribution of the ages u has a density. Here, since we only obtain a closed formula for v and not for u, we would need to derive Equation (2.36) in s to obtain a similar result, so that we need to prove more regularity on $\Phi_{\zeta_{N_-}}^{\mu,h}$. Such regularity for $\Phi_{\zeta_{N_-}}^{\mu,h}$ is not obvious since it depends strongly on the assumptions on N_-. This paves the way for future research, where the PDE formulation would provide regularity on the distribution of the ages, as done above for renewal and Wold processes.

Remark 2.5.6. *These two cases $\left(\mathcal{A}_{N_-}^1\right)$ and $\left(\mathcal{A}_{N_-}^2\right)$ highlight the dependence with respect to all the past before time 0 (i.e. ζ_{N_-}) and not only the initial condition (i.e. the age at time 0). In fact, they can give the same initial condition u^{in}.*

For instance, $\left(\mathcal{A}_{N_-}^1\right)$ with $-T_0$ exponentially distributed with parameter $\alpha > 0$ gives the same law for $-T_0$ as $\left(\mathcal{A}_{N_-}^2\right)$ with parameter α. However, if we fix some non-negative real number s, one can show that $\Phi_{-,\zeta_{N_-}}^{h}(0,s)$ is different in those two cases. It is clear from the definitions that for every real number z, $q(0,s,z) = 1$ and $K_s(0,z) = 0$. Thus, in the first case,

$$\Phi_{-,\zeta_{N_-}}^{h}(0,s) = \frac{\int_{-\infty}^{-s} h(-t_0)\alpha e^{\alpha t_0}\mathrm{d}t_0}{\int_{-\infty}^{-s} \alpha e^{\alpha t_0}\mathrm{d}t_0} = \frac{\int_s^{+\infty} h(z)\alpha e^{-\alpha z}\mathrm{d}z}{\int_s^{+\infty} \alpha e^{-\alpha z}\mathrm{d}z},$$

while in the second case,

$$\Phi^h_{-,\zeta_{N_-}}(0,s) = \alpha \int_{-\infty}^{-s} h(-z)\mathrm{d}z = \alpha \int_s^{+\infty} h(w)\mathrm{d}w.$$

Therefore $\Phi^h_{-,\zeta_{N_-}}$ *clearly depends on* ζ_{N_-} *and not just on the distribution of the last point before 0, and so is* $\rho_{\lambda,\zeta_{N_-}}$.

Remark 2.5.7. *If we follow our first guest,* $\rho_{\lambda,\zeta_{N_-}}$ *would be either*

$$\mu + \int_0^t h(t-x)u(\mathrm{d}x,0) \quad \textit{or} \quad \mu + \int_{-\infty}^t h(t-x)u(\mathrm{d}x,0).$$

In particular, it would not depend on the age s. *Therefore by* (2.42), *so would* $\Phi^{\mu,h}_{\zeta_{N_-}}$. *But for instance at time* $t = 0$, *when* N_- *is an homogeneous Poisson process of parameter* α, $\Phi^{\mu,h}_{\zeta_{N_-}}(0,s) = \mu + \alpha\int_s^{+\infty} h(w)\mathrm{d}w$, *which obviously depends on* s. *Therefore the intuition linking Hawkes processes and (PPS) does not apply.*

Remark 2.5.8. *Since the mean measure* u *defined in Proposition 2.4.3 is, in the distribution sense, the opposite of the* $s-$*derivative of the unique solution of* (2.36)-(2.37), $v \in L^\infty(\mathbb{R}^2_+)$, *the regularity of* v *implies regularity of* u. *In Hawkes processes with* $\Phi^{\mu,h}_{-,\zeta_{N_-}} = 0$ *and spontaneous rate constant,* $\mu > 0$, *as in Proposition 2.8.5,* u *presents a discontinuity for* $t = s$. *This happens because the probability to have a spike just before 0 or just after 0 may be different. However, if the spontaneous rate depends on time, i.e.* μ *is replaced by* $\mu(t)$, *and* $\mu(0) = 0$ *this discontinuity disappears because* $\Phi^{\mu,h}_+(t,s) = \mu(t) + \int_s^t (h(z) + L_s(z))\, G_s(z)\mu(t-z)\mathrm{d}z$, *when* $t > s$ *and* $\Phi^{\mu,h}_+(t,s) = \mu(t)$ *when* $s < t$. *In the cases described in Proposition 2.8.10, where* $\Phi^{\mu,h}_{-,\zeta_{N_-}} \neq 0$, *some conditions have to be guaranteed for* u *to be continous. For instance, if* N_- *is reduced to one point* T_0, *its density* f_0 *should have support in* $(-\infty, a]$ *where* $a < 0$ *and the support of the interaction function* h *should be in* $(0, -a)$.

Linear Hawkes process with no past before time 0

A classical framework in point processes theory is the case in $\left(\mathcal{A}^1_{N_-}\right)$ where $T_0 \to -\infty$, or equivalently, when N has intensity $\lambda_t = \mu + \int_0^{t-} h(t-x)N(\mathrm{d}x)$. The problem in this case is that the age at time 0 is not finite. The age is only finite for times greater than the first spiking time T_1.

Here again, the quantity $v(t,s)$ reveals more informative and easier to use: having the distribution of T_0 going to $-\infty$ means that $\mathrm{Supp}(u^{in})$ goes to $+\infty$, so that the initial condition for v tends to value uniformly 1 for any $0 \leq s < +\infty$. If we can prove that the contribution of $\Phi^h_{-,\zeta_{N_-}}$ vanishes, the following system is a good candidate to be the limit system:

$$\frac{\partial}{\partial t}v^\infty(t,s) + \frac{\partial}{\partial s}v^\infty(t,s) + \Phi^{\mu,h}_+(t,s)\, v^\infty(t,s) = 0, \tag{2.43}$$

$$v^\infty(t,0) = 1, \qquad v^\infty(0,s) = 1, \tag{2.44}$$

where $\Phi^{\mu,h}_+$ is defined in Proposition 2.5.4. This leads us to the following proposition.

Proposition 2.5.9. *Under the assumptions and notations of Proposition 2.5.4, consider, for all $M \geq 0$, v_M the unique solution of system (2.36)-(2.37) with Φ given by Proposition 2.5.4, case $\left(\mathcal{A}^1_{N_-}\right)$, with T_0 uniformly distributed in $[-M-1,-M]$. Then, as M goes to infinity, v_M converges uniformly on any set of the type $(0,T)\times(0,S)$ towards the unique solution v^∞ of System (2.43)-(2.44).*

2.6 Conclusion

We present in this chapter a bridge between univariate point processes, that can model the behaviour of one neuron through its spike train, and a deterministic age structured PDE introduced by Pakdaman, Perthame and Salort, named (PPS). More precisely Theorem 2.4.4 present a PDE that is satisfied by the distribution u of the age s at time t, where the age represents the delay between time t and the last spike before t. This is done in a very weak sense and some technical structure, namely $(\mathcal{P}_{Fubini})$, is required.

The main point is that the "firing rate" which is a deterministic quantity written as $p(s, X(t))$ in (PPS) becomes the conditional expectation of the intensity given the age at time t in Theorem 2.4.4. This first makes clear that $p(s, X(t))$ should be interpreted as a hazard rate, which gives the probability that a neuron fires given that it has not fired yet. Next, it makes clearly rigorous several "easy guess" bridges between both set-ups when the intensity only depends on the age. But it also explained why when the intensity has a more complex shape (Wold, Hawkes), this term can keep in particular the memory of all that has happened before time 0.

One of the main point of the present study is the Hawkes process, for which what was clearly expected was a legitimation of the term $X(t)$ in the firing rate $p(s, X(t))$ of (PPS), which models the synaptic integration. This is not the case, and the interlinked equations that have been found for the cumulative distribution function $v(t,\cdot)$ do not have a simple nor direct deterministic interpretation. However one should keep in mind that the present bridge, in particular in the population wide approach, has been done for independent neurons. This has been done to keep the complexity of the present work reasonable as a first step. But it is also quite obvious that interacting neurons cannot be independent. So one of the main question is: can we recover (PPS) as a limit with precisely a term of the form $X(t)$ if we consider multivariate Hawkes processes that really model interacting neurons ?

2.7 Proofs linked with the PDE

2.7.a) Proof of Proposition 2.4.1

First, let us verify that U satisfies Equation (2.10). For any $t > 0$,

$$U(t, \mathrm{d}s) = \sum_{i \geq 0} \eta_{T_i}(t, \mathrm{d}s) \mathbb{1}_{0 \leq t \leq T_{i+1}},$$

by definition of U. Yet, $\eta_{T_i}(t, \mathrm{d}s) = \delta_{t-T_i}(\mathrm{d}s)\mathbb{1}_{t>T_i}$, and the only integer i such that $T_i < t \leq T_{i+1}$ is $i = N_{t-}$. So, for all $t > 0$, $U(t, \mathrm{d}s) = \delta_{t-T_{N_{t-}}}(\mathrm{d}s) = \delta_{S_{t-}}(\mathrm{d}s)$.

Secondly, let us check $(\mathcal{P}_{Fubini})$. Let $\varphi \in \mathcal{M}_{c,b}(\mathbb{R}_+^2)$, and let T be such that for all $t > T$, $\varphi_t^{(1)} = 0$. Then, since $U(t, \mathrm{d}s) = \sum_{i=0}^{+\infty} \eta_{T_i}(t, \mathrm{d}s)\mathbb{1}_{0 \leq t \leq T_{i+1}}$,

$$\begin{aligned}
\left| \int_{\mathbb{R}_+} \left(\int_{\mathbb{R}_+} \varphi(t,s) U(t, \mathrm{d}s) \right) \mathrm{d}t \right| &\leq \int_{\mathbb{R}_+} \left(\int_{\mathbb{R}_+} |\varphi(t,s)| \sum_{i \geq 0} \eta_{T_i}(t, \mathrm{d}s) \mathbb{1}_{0 \leq t \leq T_{i+1}} \right) \mathrm{d}t \\
&= \sum_{i \geq 0} \int_{\mathbb{R}_+} |\varphi(t, t - T_i)| \mathbb{1}_{t > T_i} \mathbb{1}_{0 \leq t \leq T_{i+1}} \mathrm{d}t = \sum_{i \geq 0} \int_{\max(0,T_i)}^{T_{i+1}} |\varphi(t, t - T_i)| \mathrm{d}t \\
&= \int_0^{T_1} |\varphi(t, t - T_0)| + \sum_{i/0 < T_i < T} \int_{T_i}^{T_{i+1}} |\varphi(t, t - T_i)| \mathrm{d}t.
\end{aligned}$$

Since there is a finite number of points of N between 0 and T, on Ω, this quantity is finite and one can exchange $\sum_{i \geq 0}$ and $\int_{t=0}^{+\infty} \int_{s=0}^{+\infty}$. Therefore, since all the η_{T_i} satisfy $(\mathcal{P}_{Fubini})$ and $\varphi(t,s)\mathbb{1}_{0 \leq t \leq T_{i+1}}$ is in $\mathcal{M}_{c,b}(\mathbb{R}_+^2)$, so does U.

For the dynamics of U, similar computations lead for every $\varphi \in \mathcal{C}_{c,b}^{\infty}({\mathbb{R}_+}^2)$ to

$$\int \varphi(t,s)\, U(\mathrm{d}t, \mathrm{d}s) = \sum_{i \geq 0} \int_{\max(0,-T_i)}^{T_{i+1}-T_i} \varphi(s + T_i, s)\, \mathrm{d}s.$$

We also have

$$\begin{aligned}
\int \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s} \right) \varphi(t,s)\, U(\mathrm{d}t, \mathrm{d}s) &= \sum_{i \geq 0} \int_{\max(0,-T_i)}^{T_{i+1}-T_i} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s} \right) \varphi(s + T_i, s)\, \mathrm{d}s \\
&= \sum_{i \geq 1} \left[\varphi(T_{i+1}, T_{i+1} - T_i) - \varphi(T_i, 0) \right] + \varphi(T_1, T_1 - T_0) - \varphi(0, -T_0). \quad (2.45)
\end{aligned}$$

It remains to express the term with $\int_{x=0}^{\lambda_t} \Pi(\mathrm{d}t, \mathrm{d}x) = \sum_{i \geq 0} \delta_{T_{i+1}}(\mathrm{d}t)$, that is

$$\begin{aligned}
\int \varphi(t,s)\, U(t, \mathrm{d}s) \sum_{i \geq 0} \delta_{T_{i+1}}(\mathrm{d}t) &= \int \left(\int \varphi(t,s)\, U(t, \mathrm{d}s) \right) \sum_{i \geq 0} \delta_{T_{i+1}}(\mathrm{d}t) \\
&= \int \varphi(t, S_{t-}) \sum_{i \geq 0} \delta_{T_{i+1}}(\mathrm{d}t) = \sum_{i \geq 0} \varphi(T_{i+1}, T_{i+1} - T_i), \quad (2.46)
\end{aligned}$$

and, since $\int U(t, \mathrm{d}s) = 1$ for all $t > 0$,

$$\int \int \varphi(t, 0)\, U(t, \mathrm{d}s) \sum_{i \geq 0} \delta_{T_{i+1}}(\mathrm{d}t) = \sum_{i \geq 0} \varphi(T_{i+1}.0), \quad (2.47)$$

Identifying all the terms in the right-hand side of Equation (2.45), this lead to Equation (2.15), which is the weak formulation of System (2.12)–(2.14).

2.7.b) Proof of Proposition 2.4.3

Let $\varphi \in \mathcal{M}_{c,b}(\mathbb{R}_+^2)$, and let T be such that for all $t > T$, $\varphi_t^{(1)} = 0$. Then,

$$\int |\varphi(t,s)| U(t,\mathrm{d}s) \leq ||\varphi||_{L^\infty} \mathbb{1}_{0\leq t\leq T}, \tag{2.48}$$

since at any fixed time $t > 0$, $\int U(t,\mathrm{d}s) = 1$. Hence, the expectation $\mathbb{E}\left[\int \varphi(t,s)U(t,\mathrm{d}s)\right]$ is well-defined and finite and so $u(t,.)$ is well-defined.

On the other hand, at any fixed age s,

$$\begin{aligned}\int |\varphi(t,s)| U(\mathrm{d}t,s) &= \sum_{i=0}^{+\infty} |\varphi(s+T_i,s)| \mathbb{1}_{0\leq s\leq T_{i+1}-T_i} \\ &= \sum_{i\geq 0} |\varphi(s+T_i,s)| \mathbb{1}_{0\leq s+T_i\leq T} \mathbb{1}_{0\leq s\leq T_{i+1}-T_i},\end{aligned}$$

because for all $t > T$, $\varphi_t^{(1)} = 0$. Then, one can deduce the following bound

$$\begin{aligned}&\int |\varphi(t,s)| U(\mathrm{d}t,s) \\ &\quad\leq |\varphi(s+T_0,s)| \mathbb{1}_{-T_0\leq s\leq T-T_0} \mathbb{1}_{0\leq s\leq T_1-T_0} + \sum_{i\geq 1} |\varphi(s+T_i,s)| \mathbb{1}_{0\leq s\leq T} \mathbb{1}_{T_i\leq T} \\ &\quad\leq ||\varphi||_{L^\infty} \left(\mathbb{1}_{-T_0\leq s\leq T-T_0} + N_T \mathbb{1}_{0\leq s\leq T}\right).\end{aligned}$$

Since the intensity is L^1_{loc} in expectation, $\mathbb{E}[N_T] = \mathbb{E}\left[\int_0^T \lambda_t \mathrm{d}t\right] < +\infty$. So, the following expectation

$$\mathbb{E}\left[\int |\varphi(t,s)| U(\mathrm{d}t,s)\right] \leq ||\varphi||_{L^\infty} \left(\mathbb{E}\left[\mathbb{1}_{-T_0\leq s\leq T-T_0}\right] + \mathbb{E}[N_T]\, \mathbb{1}_{0\leq s\leq T}\right), \tag{2.49}$$

is well-defined and finite and so $u(\cdot,s)$ is well-defined.

Now, let us show $(\mathcal{P}_{Fubini})$. First, Equation (2.48) implies

$$\int \mathbb{E}\left[\int |\varphi(t,s)| U(t,\mathrm{d}s)\right] \mathrm{d}t \leq T||\varphi||_{L^\infty},$$

and Fubini's theorem implies that the following integrals are well-defined and that the following equality holds,

$$\int \mathbb{E}\left[\int \varphi(t,s) U(t,\mathrm{d}s)\right] \mathrm{d}t = \mathbb{E}\left[\int\int \varphi(t,s) U(t,\mathrm{d}s)\mathrm{d}t\right]. \tag{2.50}$$

Secondly, Equation (2.49) implies

$$\int \mathbb{E}\left[\int |\varphi(t,s)| U(\mathrm{d}t,s)\right] \mathrm{d}s \leq ||\varphi||_{L^\infty} \left(T + T\mathbb{E}[N_T]\right),$$

by exchanging the integral with the expectation and Fubini's theorem implies that the following integrals are well-defined and that the following equality holds,

$$\int \mathbb{E}\left[\int \varphi(t,s) U(\mathrm{d}t,s)\right] \mathrm{d}s = \mathbb{E}\left[\int\int \varphi(t,s) U(\mathrm{d}t,s)\mathrm{d}s\right]. \tag{2.51}$$

Now, it only remains to use $(\mathcal{P}_{Fubini})$ for U to deduce that the right members of Equations (2.50) and (2.51) are equal. Moreover, $(\mathcal{P}_{Fubini})$ for U tells that these two quantities are equal to $\mathbb{E}\left[\int\int \varphi(t,s) U(\mathrm{d}t,\mathrm{d}s)\right]$. This concludes the proof.

2.7.c) Proof of Theorem 2.4.4

Let $\rho_{\lambda,\zeta_{N_-}}$ be defined as follows, for every $t>0$ and $s\geq 0$,

$$\rho_{\lambda,\zeta_{N_-}}(t,s) := \liminf_{\varepsilon\downarrow 0} \frac{\mathbb{E}\left[\lambda_t \mathbb{1}_{|S_{t-}-s|\leq\varepsilon}\right]}{\mathbb{P}\left(|S_{t-}-s|\leq\varepsilon\right)}.$$

Since $(\lambda_t)_{t>0}$ and $(S_{t-})_{t>0}$ are predictable processes, and a fortiori progressive processes (see page 9 in [16]), $\rho_{\lambda,\zeta_{N_-}}$ is a measurable function of (t,s).

For every $t>0$, let μ_t be the measure defined by $\mu_t(A)=\mathbb{E}\left[\lambda_t \mathbb{1}_A(S_{t-})\right]$ for all measurable set A. Since Assumption $\left(\mathcal{A}^{\mathbb{L}^1,exp}_{\lambda,loc}\right)$ implies that $\mathrm{d}t$-a.e. $\mathbb{E}\left[\lambda_t\right]<+\infty$ and since $u(t,\mathrm{d}s)$ is the distribution of S_{t-}, μ_t is absolutely continuous with respect to $u(t,\mathrm{d}s)$ for $\mathrm{d}t$-almost every t.

Let f_t denote the Radon Nikodym derivative of μ_t with respect to $u(t,\mathrm{d}s)$. For $u(t,\mathrm{d}s)$-a.e. s, $f_t(s)=\mathbb{E}\left[\lambda_t|\, S_{t-}=s\right]$ by definition of the conditional expectation. Moreover, a Theorem of Besicovitch [100, Corollary 2.14] claims that for $u(t,\mathrm{d}s)$-a.e. s, $f_t(s)=\rho_{\lambda,\zeta_{N_-}}(t,s)$.

Hence, the equality $\rho_{\lambda,\zeta_{N_-}}(t,s)=\mathbb{E}\left[\lambda_t|\, S_{t-}=s\right]$ holds $u(t,\mathrm{d}s)\mathrm{d}t=u(\mathrm{d}t,\mathrm{d}s)$-a.e. Next, in order to use $(\mathcal{P}_{Fubini})$, let us note that for any $T,K>0$,

$$\rho^{K,T}_{\lambda,\zeta_{N_-}} : (t,s)\mapsto \left(\rho_{\lambda,\zeta_{N_-}}(t,s)\wedge K\right)\mathbb{1}_{0\leq t\leq T}\in\mathcal{M}_{c,b}(\mathbb{R}^2_+). \tag{2.52}$$

Hence, $\int\int \rho^{K,T}_{\lambda,\zeta_{N_-}}(t,s)u(\mathrm{d}t,\mathrm{d}s)=\int\left(\int \rho^{K,T}_{\lambda,\zeta_{N_-}}(t,s)u(t,\mathrm{d}s)\right)\mathrm{d}t$ which is always upper bounded by

$$\int_0^T\left(\int \rho_{\lambda,\zeta_{N_-}}(t,s)u(t,\mathrm{d}s)\right)\mathrm{d}t=\int_0^T\mu_t(\mathbb{R}_+)\mathrm{d}t=\int_0^T\mathbb{E}\left[\lambda_t\right]\mathrm{d}t<+\infty.$$

Letting $K\to+\infty$, one has that $\int_0^T\int \rho_{\lambda,\zeta_{N_-}}(t,s)u(\mathrm{d}t,\mathrm{d}s)$ is finite for all $T>0$. Once $\rho_{\lambda,\zeta_{N_-}}$ correctly defined, the proof of Theorem 2.4.4 is a direct consequence of Proposition 2.4.1.

More precisely, let us show that (2.15) implies (2.21). Taking the expectation of (2.15) gives that for all $\varphi\in\mathcal{C}^\infty_{c,b}(\mathbb{R}^2_+)$,

$$\mathbb{E}\left[\int\left[\varphi\left(t,s\right)-\varphi(t,0)\right]\left(\int_{x=0}^{\lambda_t}\Pi\left(\mathrm{d}t,\mathrm{d}x\right)\right)U\left(t,\mathrm{d}s\right)\right]-\int\varphi\left(0,s\right)u^{in}\left(\mathrm{d}s\right)$$
$$-\int\left(\partial_t+\partial_s\right)\varphi\left(t,s\right)u\left(\mathrm{d}t,\mathrm{d}s\right)=0. \tag{2.53}$$

Let us denote $\psi(t,s):=\varphi(t,s)-\varphi(t,0)$. Due to Ogata's thinning construction,

$$\int_{x=0}^{\lambda_t}\Pi\left(\mathrm{d}t,\mathrm{d}x\right)=N(\mathrm{d}t)\mathbb{1}_{t>0}$$

where N is the point process constructed by thinning, and so,

$$\mathbb{E}\left[\int\psi\left(t,s\right)\left(\int_{x=0}^{\lambda_t}\Pi\left(\mathrm{d}t,\mathrm{d}x\right)\right)U\left(t,\mathrm{d}s\right)\right]=\mathbb{E}\left[\int_{t>0}\psi\left(t,S_{t-}\right)N(\mathrm{d}t)\right]. \tag{2.54}$$

But $\psi(t, S_{t-})$ is a $(\mathcal{F}_t^N)$-predictable process and

$$\mathbb{E}\left[\int_{t>0} |\psi(t, S_{t-})|\lambda_t \mathrm{d}t\right] \leq \|\psi\|_{L^\infty}\, \mathbb{E}\left[\int_0^T \lambda_t \mathrm{d}t\right] < +\infty,$$

hence, using the martingale property of the predictable intensity,

$$\mathbb{E}\left[\int_{t>0} \psi\left(t, S_{t-}\right) N(\mathrm{d}t)\right] = \mathbb{E}\left[\int_{t>0} \psi\left(t, S_{t-}\right) \lambda_t \mathrm{d}t\right]. \tag{2.55}$$

Moreover, thanks to Fubini's Theorem, the right-hand term is finite and equal to $\int \mathbb{E}[\psi\left(t, S_{t-}\right) \lambda_t]\mathrm{d}t$, which can also be seen as

$$\int \mathbb{E}\left[\psi\left(t, S_{t-}\right) \rho_{\lambda,\zeta_{N_-}}(t, S_{t-})\right] \mathrm{d}t = \int \psi(t,s)\rho_{\lambda,\zeta_{N_-}}(t,s)u(t,\mathrm{d}s)\mathrm{d}t. \tag{2.56}$$

For all $K > 0$, $\Big((t,s) \mapsto \psi(t,s)\left(\rho_{\lambda,\zeta_{N_-}}(t,s) \wedge K\right)\Big) \in \mathcal{M}_{c,b}(\mathbb{R}_+^2)$ and, from $(\mathcal{P}_{Fubini})$, it is clear that

$$\int \psi(t,s)\left(\rho_{\lambda,\zeta_{N_-}}(t,s) \wedge K\right) u(t,\mathrm{d}s)\mathrm{d}t = \int \psi(t,s)\left(\rho_{\lambda,\zeta_{N_-}}(t,s) \wedge K\right) u(\mathrm{d}t,\mathrm{d}s).$$

For any $K > 0$, the right-hand side is upper-bounded in absolute value by $\|\psi\|_{L^\infty} \int_0^T \int_s \rho_{\lambda,\zeta_{N_-}}(t,s)u(\mathrm{d}t,\mathrm{d}s)$ which is finite. Letting $K \to +\infty$ one can show that

$$\int \psi(t,s)\rho_{\lambda,\zeta_{N_-}}(t,s)u(t,\mathrm{d}s)\mathrm{d}t = \int \psi(t,s)\rho_{\lambda,\zeta_{N_-}}(t,s)u(\mathrm{d}t,\mathrm{d}s). \tag{2.57}$$

Gathering (2.54)-(2.57) with (2.53) gives (2.21).

2.7.d) Proof of Corollary 2.4.5

For all $i \in \mathbb{N}^*$, let us denote $N_+^i = N^i \cap (0, +\infty)$ and $N_-^i = N^i \cap \mathbb{R}_-$. Thanks to Theorem A.1.1, the processes N_+^i can be seen as constructed via thinning of independent Poisson processes on $\mathbb{R}_+^2$. Let $(\Pi^i)_{i\in\mathbb{N}}$ be the sequence of point measures associated to independent Poisson processes of intensity 1 on $\mathbb{R}_+^2$ given by Theorem A.1.1. Let T_0^i denote the closest point to 0 in N_-^i. In particular, $(T_0^i)_{i\in\mathbb{N}^*}$ is a sequence of i.i.d. random variables.

For each i, let U^i denote the solution of the microscopic equation corresponding to Π^i and T_0^i as defined in Proposition 2.4.1 by (2.11). Using (2.10), it is clear that $\sum_{i=1}^n \delta_{S_{t-}^i}(\mathrm{d}s) = \sum_{i=1}^n U^i(t,\mathrm{d}s)$ for all $t > 0$.

Then, for every $\varphi \in \mathcal{C}_{c,b}^\infty(\mathbb{R}_+^2)$,

$$\int \varphi(t,s)\left(\frac{1}{n}\sum_{i=1}^n \delta_{S_t^i}(\mathrm{d}s)\right) = \frac{1}{n}\sum_{i=1}^n \int \varphi(t,s)U^i(t,\mathrm{d}s).$$

The right-hand side is a sum n i.i.d. random variables with mean $\int \varphi(t,s)u(t,\mathrm{d}s)$, so (2.22) clearly follows from the standard law of large numbers.

2.8 Proofs linked with the various examples

2.8.a) Renewal process

Proposition 2.8.1. *With the notations of Section 2.3, let N be a point process on $\mathbb{R}$, with predictable age process $(S_{t-})_{t>0}$, such that $T_0 = 0$ a.s. The following statements are equivalent:*

(i) $N_+ = N \cap (0, +\infty)$ is a renewal process with ISI's distribution given by some density $\nu : \mathbb{R}_+ \to \mathbb{R}_+$.

(ii) N admits $\lambda_t = f(S_{t-})$ as an intensity on $(0, +\infty)$ and $(\lambda_t)_{t>0}$ satisfies $\left(\mathcal{A}^{\mathbb{L}^1,a.s.}_{\lambda,loc}\right)$, for some $f : \mathbb{R}_+ \to \mathbb{R}_+$.

In such a case, for all $x \geq 0$, f and ν satisfy

$$\bullet\ \nu(x) = f(x)\exp(-\int_0^x f(y)dy) \text{ with the convention } \exp(-\infty) = 0, \tag{2.58}$$

$$\bullet\ f(x) = \frac{\nu(x)}{\int_x^{+\infty}\nu(y)dy} \quad \text{if } \int_x^{+\infty}\nu(y)dy \neq 0, \text{ else } f(x) = 0. \tag{2.59}$$

Proof. For (ii) $\Rightarrow$ (i). Since $T_0 = 0$ a.s., Point (2) of Proposition 2.8.2 given later on for the general Wold case implies that the ISI's of N forms a Markov chain of order 0 i.e. they are i.i.d. with density given by (2.58).

For (i) $\Rightarrow$ (ii). Let $x_0 = \inf\{x \geq 0, \int_x^{+\infty}\nu(y)dy = 0\}$. It may be infinite. Let us define f by (2.59) for every $0 \leq x < x_0$ and let $\tilde{N}$ be a point process on $\mathbb{R}$ such that $\tilde{N}_- = N_-$ and $\tilde{N}$ admits $\tilde{\lambda}_t = f(S^{\tilde{N}}_{t-})$ as an intensity on $(0, +\infty)$ where $(S^{\tilde{N}}_{t-})_{t>0}$ is the predictable age process associated to $\tilde{N}$. Applying (ii) $\Rightarrow$ (i) to $\tilde{N}$ gives that the ISI's of $\tilde{N}$ are i.i.d. with density given by

$$\tilde{\nu}(x) = \frac{\nu(x)}{\int_x^{+\infty}\nu(y)dy}\exp\left(-\int_0^x \frac{\nu(y)}{\int_y^{+\infty}\nu(z)\mathrm{d}z}dy\right),$$

for every $0 \leq x < x_0$ and $\tilde{\nu}(x) = 0$ for $x \geq x_0$. It is clear that $\nu = \tilde{\nu}$ since the function

$$x \mapsto \frac{1}{\int_x^{+\infty}\nu(y)dy}\exp\left(-\int_0^x \frac{\nu(y)}{\int_y^{+\infty}\nu(z)\mathrm{d}z}dy\right)$$

is differentiable with derivative equal to 0. Since N and $\tilde{N}$ are renewal processes with same density ν and same first point $T_0 = 0$, they have the same distribution. Since the intensity characterizes a point process, N also admits $\lambda_t = f(S^N_{t-})$ as an intensity on $(0, +\infty)$.

Moreover, since N is a renewal process, it is non-explosive in finite time and so $(\lambda_t)_{t>0}$ satisfies $\left(\mathcal{A}^{\mathbb{L}^1,a.s.}_{\lambda,loc}\right)$. □

2.8.b) Generalized Wold processes

In this Section, we suppose that there exists $k \geq 0$ such that the underlying point process N has intensity

$$\lambda_t = f(S_{t-}, A^1_t, ..., A^k_t), \tag{2.60}$$

where f is a function and the A^i's are defined by Equation (2.5).

Markovian property and the resulting PDE

Let N be a point process of intensity given by (2.60). If $T_{-k} > -\infty$, its associated age process $(S_t)_t$ can be defined for all t in $(T_{-k}, +\infty)$. Then let, for any integer $i \geq -k$,

$$\mathbb{A}_i = T_{i+1} - T_i = S_{T_{i+1}-} \tag{2.61}$$

and denote $(\mathcal{F}_i^{\mathbb{A}})_{i\geq -k}$ the natural filtration associated to $(\mathbb{A}_i)_{i\geq -k}$.

For any $t \geq 0$, and point process Π on $\mathbb{R}_+^2$, let us denote $\Pi_{\geq t}$ (resp. $\Pi_{>t}$) the restriction to $\mathbb{R}_+^2$ (resp. $(0,+\infty)\times\mathbb{R}_+$) of the point process Π shifted t time units to the left on the first coordinate. That is, $\Pi_{\geq t}(C\times D) = \Pi((t+C)\times D)$ for all $C \in \mathcal{B}(\mathbb{R}_+), D \in \mathcal{B}(\mathbb{R}_+)$ (resp. $C \in \mathcal{B}((0,+\infty))$).

Proposition 2.8.2. *Let consider k a non-negative integer, f some non negative function on $\mathbb{R}_+^{k+1}$ and N a generalized Wold process of intensity given by (2.60). Suppose that ζ_{N_-} is such that $\zeta_{N_-}(T_{-k} > -\infty) = 1$ and that $(\lambda_t)_{t>0}$ satisfies $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,a.s.}\right)$. Then,*

(i) If $(X_t)_{t\geq 0} = \left((S_{t-}, A_t^1, ..., A_t^k)\right)_{t\geq 0}$, then for any finite non-negative stopping time τ, $(X_t^\tau)_{t\geq 0} = (X_{t+\tau})_{t\geq 0}$ is independent of $\mathcal{F}_{\tau-}^N$ given X_τ.

(ii) the process $(\mathbb{A}_i)_{i\geq 1}$ given by (2.61) forms a Markov chain of order k with transition measure given by

$$\nu(\mathrm{d}x, y_1, ..., y_k) = f(x, y_1, ..., y_k)\exp\left(-\int_0^x f(z, y_1, ..., y_k)\mathrm{d}z\right)\mathrm{d}x. \tag{2.62}$$

If $T_0 = 0$ a.s., this holds for $(\mathbb{A}_i)_{i\geq 0}$.

Furthermore, if f is continuous then $\mathcal{G}$, the infinitesimal generator of $(X_t)_{t\geq 0}$, is given by, for all ϕ in $\mathcal{C}^1(\mathbb{R}_+^{k+1})$,

$$(\mathcal{G}\phi)(s, a_1, ..., a_k) = \frac{\partial}{\partial s}\phi(s, a_1, ..., a_k) + f(s, a_1, ..., a_k)\left(\phi(0, s, a_1, ..., a_{k-1}) - \phi(s, a_1, ..., a_k)\right). \tag{2.63}$$

Proof. First, let us show the first point of the Proposition. Let Π be such that N is the process resulting of Ogata's thinning with Poisson measure Π. The existence of such a measure is assured by Theorem A.1.1. We show that for any finite stopping time τ, the process $(X_t^\tau)_{t\geq 0}$ can be expressed as a function of X_τ and $\Pi_{\geq\tau}$ which is the restriction to $\mathbb{R}_+^2$ of the Poisson process Π shifted τ time units to the left on the first coordinate. Let $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^{k+1}$.

For all $t \geq 0$, let $Y_t = X_\tau + te_1$ and define

$$R_0 = \inf\left\{t \geq 0, \int_{[0,t]}\int_{x=0}^{f(Y_w)} \Pi_{\geq\tau}(\mathrm{d}w, \mathrm{d}x) = 1\right\}.$$

Note that R_0 may be null, in particular when τ is a jumping time of the underlying point process N. It is easy to check that R_0 can be expressed as a measurable function of X_τ and $\Pi_{\geq\tau}$. Moreover, it is clear that $X_{t\wedge R_0}^\tau = Y_{t\wedge R_0}$ for all $t \geq 0$. So, R_0 can be seen as the delay until the first point of the underlying process N after time τ. Suppose that R_p, the delay until the $(p+1)$th point, is constructed for some $p \geq 0$ and let us show how R_{p+1} can be

constructed. For $t \geq R_p$, let $Z_t = \theta(X^\tau_{R_p}) + te_1$, where $\theta : (x_1, \dots, x_{k+1}) \mapsto (0, x_1, \dots, x_k)$ is a right shift operator modelling the dynamics described by (2.6). Let us define

$$R_{p+1} = \inf \left\{ t > R_p, \int_{(R_p, R_p+t]} \int_{x=0}^{f(Z_w)} \Pi_{\geq\tau}(\mathrm{d}w, \mathrm{d}x) = 1 \right\}. \tag{2.64}$$

Note that for any $p \geq 0$, R_{p+1} cannot be null. It is coherent with the fact that the counting process $(N_t)_{t>0}$ only admits jumps with height 1. It is easy to check that R_{p+1} can be expressed as a measurable function of $\theta(X^\tau_{R_p})$ and $\Pi_{>\tau+R_p}$. It is also clear that $X^\tau_{t\wedge R_{p+1}} = Z_{t\wedge R_{p+1}}$ for all $t \geq R_p$. So, R_{p+1} can be seen as the delay until the $(p+2)$th point of the process N after time τ. By induction, $X^\tau_{R_p}$ can be expressed as a function of X_τ and $\Pi_{\geq\tau}$, and this holds for R_{p+1} and $X^\tau_{R_{p+1}}$ too.

To conclude, remark that the process $(X^\tau_t)_{t\geq 0}$ is a measurable function of X_τ and all the R_p's for $p \geq 0$. Thanks to the independence of the Poisson measure Π, $\mathcal{F}^N_{\tau-}$ is independent of $\Pi_{\geq\tau}$. Then, since $(X^\tau_t)_{t\geq 0}$ is a function of X_τ and $\Pi_{\geq\tau}$, $(X^\tau_t)_{t\geq 0}$ is independent of $\mathcal{F}^N_{\tau-}$ given X_τ which concludes the first point.

For Point (ii), fix $i \geq 1$ and apply Point (i) with $\tau = T_i$. It appears that in this case, $R_0 = 0$ and $R_1 = \mathbb{A}_i$. Moreover, $R_1 = \mathbb{A}_i$ can be expressed as a function of $\theta(X_\tau)$ and $\Pi_{>\tau}$. However, $\theta(X_\tau) = (0, \mathbb{A}_{i-1}, \dots, \mathbb{A}_{i-k})$ and $\mathcal{F}^{\mathbb{A}}_{i-1} \subset \mathcal{F}^N_{T_i}$. Since $\tau = T_i$, $\Pi_{>\tau}$ is independent of $\mathcal{F}^N_{T_i}$ and so $\mathbb{A}_i$ is independent of $\mathcal{F}^{\mathbb{A}}_{i-1}$ given $(\mathbb{A}_{i-1}, \dots, \mathbb{A}_{i-k})$. That is, $(\mathbb{A}_i)_{i\geq 1}$ forms a Markov chain of order k.

Note that if $T_0 = 0$ a.s. (in particular it is non-negative), then one can use the previous line of argument with $\tau = 0$ and conclude that the Markov chain starts one time step earlier, i.e. $(\mathbb{A}_i)_{i\geq 0}$ forms a Markov chain of order k.

To show (2.62), note that $R_1 = \mathbb{A}_i$, defined by (2.64), has the same distribution as the first point of a Poisson process with intensity $\lambda(t) = f(t, \mathbb{A}_{i-1}, \dots, \mathbb{A}_{i-k})$ thanks to the thinning Theorem. Hence, the transition measure of $(\mathbb{A}_i)_{i\geq 1}$ is given by (2.62).

Now that $(X_t)_{t\geq 0}$ is Markovian, one can compute its infinitesimal generator. Suppose that f is continuous and let $\phi \in \mathcal{C}^1_b(\mathbb{R}^{k+1}_+)$, The generator of $(X_t)_{t\geq 0}$ is defined by $\mathcal{G}\phi(s, a_1, \dots, a_k) = \lim_{h\to 0^+} \frac{P_h - \mathrm{Id}}{h}\phi$, where

$$\begin{aligned} P_h\phi(s, a_1, \dots, a_k) &= \mathbb{E}\left[\phi(X_h) | X_0 = (s, a_1, \dots, a_k)\right] \\ &= \mathbb{E}\left[\phi(X_h)\mathbb{1}_{\{N([0,h])=0\}} \middle| X_0 = (s, a_1, \dots, a_k)\right] \\ &\quad + \mathbb{E}\left[\phi(X_h)\mathbb{1}_{\{N([0,h])>0\}} \middle| X_0 = (s, a_1, \dots, a_k)\right] \\ &= E_0 + E_{>0}. \end{aligned}$$

The case with no jump is easy to compute,

$$E_0 = \phi(s + h, a_1, \dots, a_k)(1 - f(s, a_1, \dots, a_k)h) + o(h), \tag{2.65}$$

thanks to the continuity of f. When h is small, the probability to have more than two jumps in $[0, h]$ is a $o(h)$, so the second case can be reduced to the case with exactly one random jump (namely T),

$$\begin{aligned} E_{>0} &= \mathbb{E}\left[\phi(X_h)\mathbb{1}_{\{N([0,h])=1\}} \middle| X_0 = (s, \mathbf{a}_k)\right] + o(h) \\ &= \mathbb{E}\left[\phi(\theta(X_0 + T) + (h - T)e_1)\mathbb{1}_{\{N\cap[0,h]=\{T\}\}} \middle| X_0 = (s, \mathbf{a}_k)\right] + o(h) \\ &= \mathbb{E}\left[(\phi(0, s, \mathbf{a}_{k-1}) + o(1))\mathbb{1}_{\{N\cap[0,h]=\{T\}\}} \middle| X_0 = (s, \mathbf{a}_k)\right] + o(h) \\ &= \phi(0, s, \mathbf{a}_{k-1})(f(s, \mathbf{a}_k)h) + o(h), \end{aligned} \tag{2.66}$$

with $\mathbf{a}_k := (a_1, \dots, a_k)$, thanks to the continuity of ϕ and f. Gathering (2.65) and (2.66) with the definition of the generator gives (2.63). □

Sketch of proof of Proposition 2.5.1

Let N be the point process construct by Ogata's thinning of the Poisson process Π and U_k be as defined in Proposition 2.5.1. By an easy generalisation of Proposition 2.4.1, one can prove that on the event Ω of probability 1, where Ogata's thinning is well defined, and where $T_0 < 0$, U_k satisfies $(\mathcal{P}_{Fubini})$, (2.25) and on $\mathbb{R}_+ \times \mathbb{R}_+^{k+1}$, the following system in the weak sense

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) U_k\left(\mathrm{d}t, \mathrm{d}s, \mathrm{d}\mathbf{a}\right) + \left(\int_{x=0}^{f(s,a_1,\ldots,a_k)} \Pi\left(\mathrm{d}t, \mathrm{d}x\right)\right) U_k\left(t, \mathrm{d}s, \mathrm{d}\mathbf{a}\right) = 0,$$

$$U_k\left(\mathrm{d}t, 0, \mathrm{d}s, \mathrm{d}a_1, ..., \mathrm{d}a_{k-1}\right) = \int_{a_k \in \mathbb{R}} \left(\int_{x=0}^{f(s,a_1,\ldots,a_k)} \Pi\left(\mathrm{d}t, \mathrm{d}x\right)\right) U_k\left(t, \mathrm{d}s, \mathrm{d}\mathbf{a}\right),$$

with $\mathrm{d}\mathbf{a} = \mathrm{d}a_1 \times ... \times \mathrm{d}a_k$ and initial condition $U^{in} = \delta_{(-T_0, A_0^1, \ldots, A_0^k)}$.

Similarly to Proposition 2.4.3, one can also prove that for any test function φ in $\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$, $\mathbb{E}\left[\int \varphi(t,s,\mathbf{a}) U_k(t, \mathrm{d}s, \mathrm{d}\mathbf{a})\right]$ and $\mathbb{E}\left[\int \varphi(t,s,\mathbf{a}) U_k(\mathrm{d}t, s, \mathrm{d}\mathbf{a})\right]$ are finite and one can define $u_k(t, \mathrm{d}s, \mathrm{d}\mathbf{a})$ and $u_k(\mathrm{d}t, s, \mathrm{d}\mathbf{a})$ by, for all φ in $\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$,

$$\int \varphi(t,s,\mathbf{a}) u_k(t, \mathrm{d}s, \mathrm{d}\mathbf{a}) = \mathbb{E}\left[\int \varphi(t,s,\mathbf{a}) U_k(t, \mathrm{d}s, \mathrm{d}\mathbf{a})\right],$$

for all $t \geq 0$, and

$$\int \varphi(t,s,\mathbf{a}) u_k(\mathrm{d}t, s, \mathrm{d}\mathbf{a}) = \mathbb{E}\left[\int \varphi(t,s,\mathbf{a}) U_k(\mathrm{d}t, s, \mathrm{d}\mathbf{a})\right],$$

for all $s \geq 0$. Moreover, $u_k(t, \mathrm{d}s, \mathrm{d}\mathbf{a})$ and $u_k(\mathrm{d}t, s, \mathrm{d}\mathbf{a})$ satisfy $(\mathcal{P}_{Fubini})$ and one can define $u_k(\mathrm{d}t, \mathrm{d}s, \mathrm{d}\mathbf{a}) = u_k(t, \mathrm{d}s, \mathrm{d}\mathbf{a})\mathrm{d}t = u_k(\mathrm{d}t, s, \mathrm{d}\mathbf{a})\mathrm{d}s$ on $\mathbb{R}_+^2$, such that for any test function φ in $\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$,

$$\int \varphi(t,s,\mathbf{a}) u_k(\mathrm{d}t, \mathrm{d}s, \mathrm{d}\mathbf{a}) = \mathbb{E}\left[\int \varphi(t,s,\mathbf{a}) U_k(\mathrm{d}t, \mathrm{d}s, \mathrm{d}\mathbf{a})\right],$$

quantity which is finite. The end of the proof is completely analogous to the one of Theorem 2.4.4.

2.8.c) Linear Hawkes processes

Cluster decomposition

Proposition 2.8.3. *Let g be a non negative $L^1_{loc}(\mathbb{R}_+)$ function and h a non negative $L^1(\mathbb{R}_+)$ function such that $\|h\|_1 < 1$. Then the branching point process N is defined as $\cup_{k=0}^{\infty} N_k$ the set of all the points in all generations constructed as follows:*

- *Ancestral points are N_{anc} distributed as a Poisson process of intensity g; $N_0 := N_{anc}$ can be seen as the points of generation 0.*
- *Conditionally to N_{anc}, each ancestor $a \in N_{anc}$ gives birth, independently of anything else, to children points $N_{1,a}$ according to a Poisson process of intensity $h(. - a)$; $N_1 = \cup_{a \in N_{anc}} N_{1,a}$ forms the first generation points.*

Then the construction is recursive in k, the number of generations:

- *Denoting N_k the set of points in generation k, then conditionally to N_k, each point $x \in N_k$ gives birth, independently of anything else, to children points $N_{k+1,x}$ according to a Poisson process of intensity $h(. - x)$; $N_{k+1} = \cup_{x \in N_k} N_{k+1,x}$ forms the points of the $(k+1)$th generation.*

This construction ends almost surely in every finite interval. Moreover the intensity of N exists and is given by

$$\lambda_t = g(t) + \int_0^{t-} h(t-x)N(\mathrm{d}x).$$

This is the cluster representation of the Hawkes process. When $g \equiv \nu$, this has been proved in [71]. However up to our knowledge this has not been written for a general function g.

Proof. First, let us fix some $A > 0$. The process ends up almost surely in $[0, A]$ because there is a.s. a finite number of ancestors in $[0, A]$: if we consider the family of points attached to one particular ancestor, the number of points in each generation form a subcritical Galton Watson process with reproduction distribution, a Poisson variable with mean $\int h < 1$ and whose extinction is consequently almost sure.

Next, to prove that N has intensity

$$H(t) = g(t) + \int_0^{t-} h(t-x)N(\mathrm{d}x),$$

we exhibit a particular thinning construction, where on the one hand, N is indeed a branching process by construction as defined by the proposition and, which, on the other hand, guarantees that Ogata's thinning project the points below $H(t)$. We can always assume that $h(0) = 0$, since changing the intensity of Poisson process in the branching structure at one particular point has no impact. Hence $H(t) = g(t) + \int_0^t h(t-x)N(\mathrm{d}x)$.

The construction is recursive in the same way. Fix some realisation Π of a Poisson process on $\mathbb{R}_+^2$.

For N_{anc}, project the points below the curve $t \to g(t)$ on $[0, A]$. By construction, N_{anc} is a Poisson process of intensity $g(t)$ on $[0, A]$. Note that for the identification (see Theorem 2.8.11) we just need to do it on finite intervals and that the ancestors that may be born after time A do not have any descendants in $[0, A]$, so we can discard them, since they do not appear in $H(t)$, for $t \leq A$.

Enumerate the points in $N_{anc} \cap [0, A]$ from T_1 to $T_{N_{0,\infty}}$.

- The children of T_1, N_{1,T_1}, are given by the projection of the points of Π whose ordinates are in the strip $t \mapsto (g(t), g(t) + h(t - T_1)]$. As before, by the property of spatial independence of Π, this is a Poisson process of intensity $h(. - T_1)$ conditionally to N_{anc}.

- Repeat until $T_{N_{0,\infty}}$, where $N_{1,T_{N_{0,\infty}}}$ are given by the projection of the points of Π whose ordinates are in the strip $t \mapsto (g(t) + \sum_{i=1}^{N_{0,\infty}-1} h(t - T_i), g(t) + \sum_{i=1}^{N_{0,\infty}} h(t - T_i)]$. As before, by the property of independence of Π, this is a Poisson process of intensity $h(. - T_{N_{0,\infty}})$ conditionally to N_{anc} and because the consecutive strips do not overlap, this process is completely independent of the previous processes (N_{1,T_i})'s that have been constructed.

Note that at the end of this first generation, $N_1 = \cup_{T\in N_{anc}} N_{1,T}$ consists of the projection of points of Π in the strip $t \mapsto (g(t), g(t) + \sum_{i=1}^{N_{0,\infty}} h(t-T_i)]$. They therefore form a Poisson process of intensity $\sum_{i=1}^{N_{0,\infty}} h(t-T_i) = \int h(t-z)N_{anc}(dz)$, conditionally to N_{anc}.

For generation $k+1$ replace in the previous construction N_{anc} by N_k and $g(t)$ by $g(t) + \sum_{j=0}^{k-1} \int h(t-u)dN_j(u)$. Once again we end up for each point x in N_k with a process of children $N_{k+1,x}$ which is a Poisson process of intensity $h(t-x)$ conditionally to N_k and which is totally independent of the other $N_{k+1,y}$'s. Note also that as before, $N_{k+1} = \cup_{x\in N_k} N_{k+1,x}$ is a Poisson process of intensity $\int h(t-z)N_k(dz)$, conditionally to $N_0, ..., N_k$.

Hence we are indeed constructing a branching process as defined by the proposition. Because the underlying Galton Watson process ends almost surely, as shown before, it means that there exists a.s. one generation N_{k^*} which will be completely empty and our recursive construction ends up too.

The main point is to realize that at the end the points in $N = \cup_{k=0}^{\infty} N_k$ are exactly the projection of the points in Π that are below

$$t \mapsto g(t) + \sum_{k=0}^{+\infty} \int h(t-z)N_k(dz) = g(t) + \sum_{k=0}^{+\infty} \int_0^t h(t-z)N_k(dz)$$

hence below

$$t \mapsto g(t) + \int_0^t h(t-z)N(dz) = H(t).$$

Moreover $H(t)$ is $\mathcal{F}_t^N$ predictable. Therefore by Theorem 2.8.11, N has intensity $H(t)$, which concludes the proof. □

A cluster process N_c, is a branching process, as defined before, which admits intensity $\lambda_t^c = h(t) + \int_0^{t-} h(t-z)N_c(\mathrm{d}z)$. Its distribution only depends on the function h. It corresponds to the family generated by one ancestor at time 0 in Proposition 2.8.3. Therefore, by Proposition2.8.3, a Hawkes process N with empty past ($N_- = \emptyset$) of intensity $\lambda_t = g(t) + \int_0^{t-} h(t-z)N(\mathrm{d}z)$ can always be seen as the union of N_{anc} and of all the $a + N_c^a$ for $a \in N_{anc}$ where the N_c^a are i.i.d. cluster processes.

For a Hawkes process N with non empty past, N_-, this is more technical. Let N_{anc} be a Poisson process of intensity g on $\mathbb{R}_+$ and $\left(N_c^V\right)_{V\in N_{anc}}$ be a sequence of i.i.d. cluster processes associated to h. Let also

$$N_{>0} = N_{anc} \cup \left(\bigcup_{V\in N_{anc}} V + N_c^V \right). \tag{2.67}$$

As we prove below, this represents the points in N that do not depend on N_-. The points that are depending on N_- are constructed as follows independently of $N_{>0}$. Given N_-, let $\left(N_1^T\right)_{T\in N_-}$ denote a sequence of independent Poisson processes with respective intensities $\lambda_T(v) = h(v-T)\mathbb{1}_{(0,\infty)}(v)$. Then, given N_- and $\left(N_1^T\right)_{T\in N_-}$, let $\left(N_c^{T,V}\right)_{V\in N_1^T, T\in N_-}$ be a sequence of i.i.d. cluster processes associated to h. The points depending on the past N_- are given by the following formula as proved in the next Proposition:

$$N_{\leq 0} = N_- \cup \left(\bigcup_{T\in N_-} N_1^T \cup \left(\bigcup_{V\in N_1^T} V + N_c^{T,V} \right) \right). \tag{2.68}$$

Proposition 2.8.4. *Let $N = N_{\leq 0} \cup N_{>0}$, where $N_{>0}$ and $N_{\leq 0}$ are given by (2.67) and (2.68). Then N is a linear Hawkes process with past given by N_- and intensity on $(0, +\infty)$ given by $\lambda_t = g(t) + \int_{-\infty}^{t-} h(t-x)N(\mathrm{d}x)$, where g and h are as in Proposition 2.8.3.*

Proof. Proposition2.8.3 yields that $N_{>0}$ has $\mathcal{F}_{t-}^{N_{>0}}$-intensity

$$\lambda_t^{N_{>0}} = g(t) + \int_0^{t-} h(t-x)N_{>0}(\mathrm{d}x), \tag{2.69}$$

and that, given N_-, for any $T \in N_-$, $N_H^T = N_1^T \cup \left(\bigcup_{V \in N_1^T} V + N_c^{T,V}\right)$ has $\mathcal{F}_{t-}^{N_H^T}$-intensity

$$\lambda_t^{N_H^T} = h(t-T) + \int_0^{t-} h(t-x)N_H^T(\mathrm{d}x), \tag{2.70}$$

Moreover, all these processes are independent given N_-. For any $t \geq 0$, one can note that

$$\mathcal{F}_t^{N_{\leq 0}} \subset \mathcal{G}_t := \mathcal{F}_0^{N_-} \vee \left(\bigvee_{T \in N_-} \mathcal{F}_t^{N_H^T} \right),$$

and so $N_{\leq 0}$ has $\mathcal{G}_{t-}$-intensity

$$\lambda_t^{N_{\leq 0}} = \sum_{T \in N_-} \lambda_t^{N_H^T} = \int_{-\infty}^{t-} h(t-x)N_{\leq 0}(\mathrm{d}x) \tag{2.71}$$

on $(0, +\infty)$. Since this last expression is $\mathcal{F}_t^{N_{\leq 0}}$-predictable, by [16, page 27], this is also the $\mathcal{F}_{t-}^{N_{\leq 0}}$-intensity of $N_{\leq 0}$. Moreover, $N_{\leq 0}$ and $N_{>0}$ are independent by construction and, for any $t \geq 0$, $\mathcal{F}_t^N \subset \mathcal{F}_t^{N_{\leq 0}} \vee \mathcal{F}_t^{N_{>0}}$. Hence, as before, N has $\mathcal{F}_{t-}^N$-intensity on $(0, +\infty)$ given by

$$\lambda_t = \lambda_t^{N_{\leq 0}} + \lambda_t^{N_{>0}} = g(t) + \int_{-\infty}^{t-} h(t-x)N(\mathrm{d}x).$$

□

A general result for linear Hawkes processes

The following proposition is a consequence of Theorem 2.4.4 applied to Hawkes processes with general past N_-.

Proposition 2.8.5. *Using the notations of Theorem 2.4.4, let N be a Hawkes process with past before 0 given by N_- of distribution ζ_{N_-} and with intensity on $\mathbb{R}_+$ given by*

$$\lambda_t = \mu + \int_{-\infty}^{t-} h(t-x)N(dx),$$

where μ is a positive real number and h is a non-negative function with support in $\mathbb{R}_+$ such that $\int h < 1$. Suppose that ζ_{N_-} is such that

$$\sup_{t \geq 0} \mathbb{E}\left[\int_{-\infty}^{0} h(t-x)N_-(dx) \right] < +\infty. \tag{2.72}$$

Then, the mean measure u defined in Proposition 2.4.3 satisfies Theorem 2.4.4 and moreover its integral $v(t,s) := \int_s^{+\infty} u(t, d\sigma)$ is a solution of the system (2.36)–(2.37) where v^{in} is the survival function of $-T_0$, and where $\Phi = \Phi^{\mu,h}_{\zeta_{N_-}}$ is given by $\Phi^{\mu,h}_{\zeta_{N_-}} = \Phi^{\mu,h}_{+} + \Phi^{\mu,h}_{-,\zeta_{N_-}}$, with $\Phi^{\mu,h}_{+}$ given by (2.39) and $\Phi^{\mu,h}_{-,\zeta_{N_-}}$ given by,

$$\forall\, s,t \geq 0, \quad \Phi^{\mu,h}_{-,\zeta_{N_-}}(t,s) = \mathbb{E}\left[\int_{-\infty}^{t-} h(t-z)N_{\leq 0}(dz)\,\middle|\, N_{\leq 0}\left([t-s,t)\right) = 0\right]. \tag{2.73}$$

Moreover, (2.42) holds.

Proof of the general result of Proposition 2.8.5

Before proving Proposition 2.8.5, we need some technical preliminaries.

Events of the type $\{S_{t-} \geq s\}$ are equivalent to the fact that the underlying process has no point between $t-s$ and t. Therefore, for any point process N and any real numbers $t, s \geq 0$, let

$$\mathcal{E}_{t,s}(N) = \{N \cap [t-s,t) = \emptyset\}. \tag{2.74}$$

Various sets $\mathcal{E}_{t,s}(N)$ are used in the sequel and the following lemma is applied several times to those sets.

Lemma 2.8.6. *Let Y be some random variable and $I(Y)$ some countable set of indices depending on Y. Suppose that $(X_i)_{i\in I(Y)}$ is a sequence of random variables which are independent conditionally on Y. Let $A(Y)$ be some event depending on Y and $\forall\, j \in I(Y)$, $B_j = B_j(Y, X_j)$ be some event depending on Y and X_j. Then, for any $i \in I(Y)$, and for all sequence of measurable functions $(f_i)_{i\in I(Y)}$ such that the following quantities exist,*

$$\mathbb{E}\left[\sum_{i\in I(Y)} f_i(Y,X_i)\,\middle|\, A\#B\right] = \mathbb{E}\left[\sum_{i\in I(Y)} \mathbb{E}\left[f_i(Y,X_i)\middle|\, Y, B_i\right]\,\middle|\, A\#B\right],$$

where $\mathbb{E}\left[f_i(Y,X_i)\middle|\, Y, B_i\right] = \frac{\mathbb{E}\left[f_i(Y,X_i)\mathbb{1}_{B_i}\middle|\, Y\right]}{\mathbb{P}(B_i|\, Y)}$ and $A\#B = A(Y) \cap \left(\bigcap_{j\in I(Y)} B_j\right)$.

Proof. By definition,

$$\begin{aligned}
\mathbb{E}\left[\sum_{i\in I(Y)} f_i(Y,X_i)\,\middle|\, A\#B\right] &= \frac{\mathbb{E}\left[\sum_{i\in I(Y)} f_i(Y,X_i)\mathbb{1}_{A(Y)}\prod_{j\in I(Y)}\mathbb{1}_{B_j}\right]}{\mathbb{P}\left(A\#B\right)} \\
&= \frac{\mathbb{E}\left[\sum_{i\in I(Y)} \mathbb{E}\left[f_i(Y,X_i)\prod_{j\in I(Y)}\mathbb{1}_{B_j}\middle| Y\right]\mathbb{1}_{A(Y)}\right]}{\mathbb{P}\left(A\#B\right)} \\
&= \frac{\mathbb{E}\left[\sum_{i\in I(Y)} \mathbb{E}\left[f_i(Y,X_i)\mathbb{1}_{B_i}|Y\right]\prod_{j\neq i}\mathbb{P}(B_j|Y)\mathbb{1}_{A(Y)}\right]}{\mathbb{P}\left(A\#B\right)},
\end{aligned}$$

thanks to the conditional independence of the X_j's (and so the B_j's) given Y. By definition, $\mathbb{E}\left[f_i(Y,X_i)\mathbb{1}_{B_i}|Y\right] = \mathbb{E}\left[f_i(Y,X_i)|Y,B_i\right]\mathbb{P}\left(B_i|Y\right)$, and thanks to the definition of conditional

probability, we end up with

$$\begin{aligned}\mathbb{E}\left[\sum_{i\in I(Y)} f_i(Y,X_i)\middle| A\#B\right] &= \frac{\mathbb{E}\left[\sum_{i\in I(Y)}\mathbb{E}\left[f_i(Y,X_i)|Y,B_i\right]\prod_{j\in I(Y)}\mathbb{P}(B_j|Y)\mathbb{1}_{A(Y)}\right]}{\mathbb{P}\left(A\#B\right)}\\ &= \frac{\mathbb{E}\left[\sum_{i\in I(Y)}\mathbb{E}\left[f_i(Y,X_i)|Y,B_i\right]\prod_{j\in I(Y)}\mathbb{1}_{B_j}\mathbb{1}_{A(Y)}\right]}{\mathbb{P}\left(A\#B\right)}\\ &= \mathbb{E}\left[\sum_{i\in I(Y)}\mathbb{E}\left[f_i(Y,X_i)|Y,B_i\right]\middle| A\#B\right].\end{aligned}$$

□

The following lemma is linked to Lemma 2.5.2.

Lemma 2.8.7. *Let N be a linear Hawkes process with no past before time 0 (i.e. $N_- = \emptyset$) and intensity on $(0,+\infty)$ given by $\lambda_t = g(t) + \int_0^{t-} h(t-x)N(\mathrm{d}x)$, where g and h are as in Proposition 2.8.3 and let, for any $x, s \geq 0$,*

$$\begin{cases} L_s^{g,h}(x) = \mathbb{E}\left[\int_0^x h(x-z)N(\mathrm{d}z)\middle|\mathcal{E}_{x,s}(N)\right] \\ G_s^{g,h}(x) = \mathbb{P}\left(\mathcal{E}_{x,s}(N)\right), \end{cases}$$

Then, for any $x, s \geq 0$,

$$L_s^{g,h}(x) = \int_{s\wedge x}^x \left(h\left(z\right) + L_s^{h,h}(z)\right) G_s^{h,h}(z) g(x-z)\,\mathrm{d}z, \tag{2.75}$$

and

$$\log(G_s^{g,h}(x)) = \int_0^{(x-s)\vee 0} G_s^{h,h}(x-z)g(z)\mathrm{d}z - \int_0^x g(z)\mathrm{d}z. \tag{2.76}$$

In particular, $(L_s^{h,h}, G_s^{h,h})$ is in $L^1 \times L^\infty$ and is a solution of (2.33)-(2.34).

Proof. The statement only depends on the distribution of N. Hence, thanks to Proposition 2.8.4, it is sufficient to consider $N = N_{anc} \cup \left(\cup_{V\in N_{anc}} V + N_c^V\right)$.

Let us show (2.75). First, let us write $L_s^{g,h}(x) = \mathbb{E}\left[\sum_{X\in N} h(x-X)\middle|\mathcal{E}_{x,s}(N)\right]$, and note that $L_s^{g,h}(x) = 0$ if $x \leq s$. The following decomposition holds

$$L_s^{g,h}(x) = \mathbb{E}\left[\sum_{V\in N_{anc}}\left(h(x-V) + \sum_{W\in N_c^V} h(x-V-W)\right)\middle|\mathcal{E}_{x,s}(N)\right].$$

According to Lemma 2.8.6 and the following decomposition,

$$\mathcal{E}_{x,s}(N) = \mathcal{E}_{x,s}(N_{anc}) \cap \left(\bigcap_{V\in N_{anc}} \mathcal{E}_{x-V,s}(N_c^V)\right), \tag{2.77}$$

let us denote $Y = N_{anc}$, $X_V = N_c^V$ and $B_V = \mathcal{E}_{x-V,s}(N_c^V)$ for all $V \in N_{anc}$. Let us fix $V \in N_{anc}$ and compute the conditional expectation of the inner sum with respect to the filtration of N_{anc} which is

$$\begin{aligned}\mathbb{E}\left[\sum_{W\in N_c^V} h(x-V-W)\middle| Y, B_V\right] &= \mathbb{E}\left[\sum_{W\in N_c} h((x-V)-W)\middle|\mathcal{E}_{x-V,s}(N_c)\right]\\ &= L_s^{h,h}(x-V),\end{aligned} \tag{2.78}$$

since, conditionally on N_{anc}, N_c^V has the same distribution as N_c which is a linear Hawkes process with conditional intensity $\lambda_t^c = h(t) + \int_0^{t-} h(t-z)N_c(\mathrm{d}z)$. Using the conditional independence of the cluster processes with respect to N_{anc}, one can apply Lemma 2.8.6 and deduce that

$$L_s^{g,h}(x) = \mathbb{E}\left[\sum_{V\in N_{anc}} \big(h(x-V) + L_s^{h,h}(x-V)\big)\middle|\mathcal{E}_{x,s}(N)\right]$$

The following argument is inspired by Moller [105]. For every $V \in N_{anc}$, we say that V has mark 0 if V has no descendant or himself in $[x-s,x)$ and mark 1 otherwise. Let us denote N_{anc}^0 the set of points with mark 0 and $N_{anc}^1 = N_{anc} \setminus N_{anc}^0$. For any $V \in N_{anc}$, we have $\mathbb{P}\left(V \in N_{anc}^0 | N_{anc}\right) = G_s^{h,h}(x-V)\mathbb{1}_{[x-s,x)^c}(V)$, and all the marks are chosen independently given N_{anc}. Hence, N_{anc}^0 and N_{anc}^1 are independent Poisson processes and the intensity of N_{anc}^0 is given by $\lambda(v) = g(v)G_s^{h,h}(x-v)\mathbb{1}_{[x-s,x)^c}(v)$. Moreover, the event $\{N_{anc}^1 = \emptyset\}$ can be identified to $\mathcal{E}_{x,s}(N)$ and

$$\begin{aligned} L_s^{g,h}(x) &= \mathbb{E}\left[\sum_{V\in N_{anc}^0} \big(h(x-V) + L_s^{h,h}(x-V)\big)\middle| N_{anc}^1 = \emptyset\right] \\ &= \int_{-\infty}^{x-} \left(h\left(x-w\right) + L_s^{h,h}(x-w)\right) g(w)G_s^{h,h}(x-w)\mathbb{1}_{[x-s,x)^c}(w)\mathrm{d}w \\ &= \int_0^{(x-s)\vee 0} \left(h\left(x-w\right) + L_s^{h,h}(x-w)\right) G_s^{h,h}(x-w)g(w)\,\mathrm{d}w, \end{aligned}$$

where we used the independence between the two Poisson processes. It suffices to substitute w by $z = x - w$ in the integral to get the desired formula. Since $G_s^{h,h}$ is bounded, it is obvious that $L_s^{h,h}$ is L^1.

Then, let us show (2.76). First note that if $x < 0$, $G_s^{g,h}(x) = 1$. Next, following (2.77) one has $G_s^{g,h}(x) = \mathbb{E}\left[\mathbb{1}_{\mathcal{E}_{x,s}(N_{anc})}\prod_{X\in N_{anc}}\mathbb{1}_{\mathcal{E}_{x-X,s}(N_c^X)}\right]$. This is also

$$\begin{aligned} G_s^{g,h}(x) &= \mathbb{E}\left[\mathbb{1}_{N_{anc}\cap[x-s,x)=\emptyset}\prod_{V\in N_{anc}\cap[x-s,x)^c}\mathbb{1}_{\mathcal{E}_{x-V,s}(N_c^V)}\right], \\ &= \mathbb{E}\left[\mathbb{1}_{N_{anc}\cap[x-s,x)=\emptyset}\prod_{V\in N_{anc}\cap[x-s,x)^c}G_s^{h,h}(x-V)\right], \end{aligned}$$

by conditioning with respect to N_{anc}. Since $N_{anc}\cap[x-s,x)$ is independent of $N_{anc}\cap[x-s,x)^c$, this gives

$$G_s^{g,h}(x) = \exp(-\int_{x-s}^x g(z)\mathrm{d}z)\mathbb{E}\left[\exp\left(\int_{[x-s,x)^c}\log(G_s^{h,h}(x-z))N_{anc}(\mathrm{d}z)\right)\right].$$

This leads to $\log(G_s^{g,h}(x)) = -\int_{x-s}^x g(z)\mathrm{d}z + \int_{[x-s,x)^c}(G_s^{h,h}(x-z)-1)g(z)\mathrm{d}z$, thanks to Campbell's Theorem [86]. Then, (2.76) clearly follows from the facts that if $z > x > 0$ then $G_s^{h,h}(x-z) = 1$ and $g(z) = 0$ as soon as $z < 0$. □

Proof of Lemma 2.5.2 In turn, we use a Banach fixed point argument to prove that for all $s \geq 0$ there exists a unique couple $(L_s, G_s) \in L^1(\mathbb{R}_+) \times L^\infty(\mathbb{R}_+)$ solution to these

equations. To do so, let us first study Equation (2.33) and define $T_{G,s} : L^\infty(\mathbb{R}_+) \to L^\infty(\mathbb{R}_+)$ by

$$T_{G,s}(f)(x) := \exp\left(\int_0^{(x-s)\vee 0} f(x-z)h(z)\mathrm{d}z - \int_0^x h(z)\mathrm{d}z\right).$$

The right-hand side is well-defined since $h \in L^1$ and $f \in L^\infty$. Moreover we have

$$T_{G,s}(f)(x) \leq e^{\|f\|_{L^\infty}\left(\int_0^{(x-s)\vee 0} h(z)\mathrm{d}z - \int_0^x h(z)\mathrm{d}z\right)} \leq e^{(\|f\|_{L^\infty}-1)\int_0^{(x-s)\vee 0} h(z)\mathrm{d}z}.$$

This shows that $T_{G,s}$ maps the ball of radius 1 of L^∞ into itself, and more precisely into the intersection of the positive cone and the ball. We distinguish two cases:

– If $x < s$, then $T_{G,s}(f)(x) = \exp(-\int_0^x h(z)\mathrm{d}z)$ for any f, thus, the unique fixed point is given by $G_s : x \mapsto \exp(-\int_0^x h(z)\mathrm{d}z)$, which does not depend on $s > x$.

– And if $x > s$, the functional $T_{G,s}$ is a $k-$contraction in $\{f \in L^\infty(\mathbb{R}_+), \|f\|_{L^\infty} \leq 1\}$, with $k \leq \int_0^{+\infty} h(z)\mathrm{d}z < 1$, by convexity of the exponential. More precisely, using that for all x, y, $|e^x - e^y| \leq e^{\max(x,y)}|x-y|$ we end up with, for $\|f\|, \|g\|_{L^\infty} \leq 1$,

$$\begin{aligned}\left|T_{G,s}(f)(x) - T_{G,s}(g)(x)\right| &\leq e^{-\int_0^x h(z)\mathrm{d}z} e^{\int_0^{x-s} h(z)\mathrm{d}z} \|f-g\|_{L^\infty} \int_0^{(x-s)} h(z)\mathrm{d}z \\ &\leq \|f-g\|_{L^\infty} \int_{\mathbb{R}_+} h(z)\mathrm{d}z.\end{aligned}$$

Hence there exists only one fixed point G_s that we can identify with $G_s^{h,h}$ given in Proposition 2.8.7 and $G_s^{h,h}$ being a probability, G_s takes values in $[0,1]$.

Analogously, we define the functional $T_{L,s} : L^1(\mathbb{R}_+) \to L^1(\mathbb{R}_+)$ by

$$T_{L,s}(f)(x) := \int_{s\wedge x}^x \left(h\left(z\right) + f(z)\right) G_s(z)h(x-z)\,\mathrm{d}z,$$

and it is easy to check that $T_{L,s}$ is well-defined as well. We similarly distinguish the two cases:

– If $x < s$, then the unique fixed point is given by $L_s(x) = 0$.

– And if $x > s$, thus $T_{L,s}$ is a $k-$contraction with $k \leq \int_0^{+\infty} h(y)dy < 1$ in $L^1((s,+\infty))$ since $\|G_s\|_{L^\infty} \leq 1$:

$$\begin{aligned}\|T_{L,s}(f) - T_{L,s}(g)\|_{L^1} &= \int_s^{+\infty} \Big|\int_s^x \big(f(z)-g(z)\big)G_s(z)h(x-z)\mathrm{d}z\Big|\mathrm{d}x \\ &\leq \|G_s\|_{L^\infty} \int_s^{+\infty}\int_{-\infty}^z \big|f(z)-g(z)\big|h(x-z)\mathrm{d}z\mathrm{d}x \\ &= \|G_s\|_{L^\infty}\|f-g\|_{L^1((s,+\infty))} \int_0^{+\infty} h(y)dy.\end{aligned}$$

In the same way, there exists only one fixed point $L_s = L_s^{h,h}$ given by Proposition 2.8.7. In particular $L_s(x \leq s) \equiv 0$.

Finally, as a consequence of Equation (2.34) we find that if L_s is the unique fixed point of $T_{L,s}$, then

$$\|L_s\|_{L^1(\mathbb{R}_+)} \leq \frac{\left(\int_0^{+\infty} h(y)\ dy\right)^2}{1 - \int_0^{+\infty} h(y)\ dy}$$

and therefore L_s is uniformly bounded in L^1 with respect to s.

Lemma 2.8.8. *Let N be a linear Hawkes process with past before time 0 given by N_- and intensity on $(0,+\infty)$ given by $\lambda_t = \mu + \int_{-\infty}^{t-} h(t-x)N(\mathrm{d}x)$, where μ is a positive real number and h is a non-negative function with support in $\mathbb{R}_+$, such that $||h||_{L^1} < 1$. If the distribution of N_- satisfies (2.72) then $(\mathcal{A}^{\mathbb{L}^1,exp}_{\lambda,loc})$ is satisfied.*

Proof. For all $t > 0$, let $\overline{\lambda}(t) = \mathbb{E}[\lambda_t]$. By Proposition 2.8.4,

$$\overline{\lambda}(t) = \mathbb{E}\left[\mu + \int_0^{t-} h(t-x)N_{>0}(\mathrm{d}x)\right] + \mathbb{E}\left[\int_{-\infty}^{t-} h(t-x)N_{\leq 0}(\mathrm{d}x)\right]$$

which is possibly infinite.

Let us apply Proposition 2.8.7 with $g \equiv \mu$ and $s = 0$, the choice $s = 0$ implying that $\mathcal{E}_{t,0}(N_{>0})$ is of probability 1. Therefore

$$\mathbb{E}\left[\mu + \int_0^{t-} h(t-x)N_{>0}(\mathrm{d}x)\right] = \mu\left(1 + \int_0^t (h(x) + L_0(x))\mathrm{d}x\right),$$

where $(L_0, G_0 = 1)$ is the solution of Lemma 2.5.2 for $s = 0$, by identification of Proposition 2.8.7. Hence $\mathbb{E}\left[\mu + \int_0^{t-} h(t-x)N_{>0}(\mathrm{d}x)\right] \leq \mu(1 + ||h||_{L^1} + ||L_0||_{L^1})$.

On the other hand, thanks to Lemma 2.8.9, we have

$$\mathbb{E}\left[\int_{-\infty}^{t-} h(t-x)N_{\leq 0}(\mathrm{d}x)\right] = \mathbb{E}\left[\sum_{T\in N_-}\left(h(t-T) + \int_0^t [h(t-x) + L_0(t-x)]\,h(x-T)\mathrm{d}x\right)\right].$$

Since all the quantities are non negative, one can exchange all the integrals and deduce that

$$\mathbb{E}\left[\int_{-\infty}^{t-} h(t-x)N_{\leq 0}(\mathrm{d}x)\right] \leq M(1 + ||h||_{L^1} + ||L_0||_{L^1}),$$

with $M = \sup_{t\geq 0}\mathbb{E}\left[\int_{-\infty}^0 h(t-x)N_-(\mathrm{d}x)\right]$ which is finite by assumption. Hence, $\overline{\lambda}(t) \leq (\mu + M)(1 + ||h||_{L^1} + ||L_0||_{L^1})$, and therefore $(\mathcal{A}^{\mathbb{L}^1,exp}_{\lambda,loc})$ is satisfied.

□

Proof of Proposition 2.8.5 First, by Proposition 2.8.4

$$\begin{aligned}\mathbb{E}[\lambda_t |\, S_{t-} \geq s] = &\\ &\mu + \mathbb{E}\left[\int_0^{t-} h(t-z)N_{>0}(\mathrm{d}z)\middle|\mathcal{E}_{t,s}(N)\right] + \mathbb{E}\left[\int_{-\infty}^{t-} h(t-z)N_{\leq 0}(\mathrm{d}z)\middle|\mathcal{E}_{t,s}(N)\right]\\ &= \mu + \mathbb{E}\left[\int_0^{t-} h(t-z)N_{>0}(\mathrm{d}z)\middle|\mathcal{E}_{t,s}(N_{>0})\right] + \mathbb{E}\left[\int_{-\infty}^{t-} h(t-z)N_{\leq 0}(\mathrm{d}z)\middle|\mathcal{E}_{t,s}(N_{\leq 0})\right]\end{aligned}$$

By Lemma 2.8.7, we obtain $\mathbb{E}[\lambda_t |\, S_{t-} \geq s] = \mu + L^{\mu,h}_s(t) + \Phi^h_{-,\zeta_{N_-}}(t,s)$. Identifying by Lemma 2.5.2, $L_s = L^{h,h}_s$ and $G_s = G^{h,h}_s$, we obtain

$$\mathbb{E}[\lambda_t |\, S_{t-} \geq s] = \Phi^{\mu,h}_+(t,s) + \Phi^h_{-,\zeta_{N_-}}(t,s).$$

Hence $\Phi^{\mu,h}_{\zeta_{N_-}}(t,s) = \mathbb{E}\left[\lambda_t \middle| S_{t-} \geq s\right]$.

Lemma 2.8.8 ensures that the assumptions of Theorem 2.4.4 are fulfilled. Let u and $\rho^{\mu,h}_{\zeta_{N_-}} = \rho_{\lambda,\zeta_{N_-}}$ be defined accordingly as in Theorem 2.4.4. With respect to the PDE system, there are two possibilities to express $\mathbb{E}\left[\lambda_t \mathbb{1}_{\{S_{t-}\geq s\}}\right]$. The first one involves $\rho_{\lambda,\zeta_{N_-}}$ and is $\mathbb{E}\left[\rho^{\mu,h}_{\zeta_{N_-}}(t,S_{t-})\mathbb{1}_{S_{t-}\geq s}\right]$, whereas the second one involves $\Phi^{\mu,h}_{\zeta_{N_-}}$ and is $\Phi^{\mu,h}_{\zeta_{N_-}}(t,s)\mathbb{P}\left(S_{t-}\geq s\right)$.

This leads to $\int_s^{+\infty} \rho^{\mu,h}_{\zeta_{N_-}}(t,x)u(t,\mathrm{d}x) = \Phi^{\mu,h}_{\zeta_{N_-}}(t,s)\int_s^{+\infty} u(t,\mathrm{d}x)$, since $u(t,\mathrm{d}s)$ is the distribution of S_{t-}. Let us denote $v(t,s) = \int_s^{+\infty} u(t,\mathrm{d}x)$: this relation, together with Equation (2.18) for u, immediately gives us that v satisfies Equation (2.36) with $\Phi = \Phi^{\mu,h}_{\zeta_{N_-}}$. Moreover, $\int_0^{+\infty} u(t,\mathrm{d}x) = 1$, which gives us the boundary condition in (2.37).

Study of the general case for $\Phi^h_{-,\zeta_{N_-}}$ in Proposition 2.8.5

Lemma 2.8.9. *Let consider h a non-negative function with support in $\mathbb{R}_+$ such that $\int h < 1$, N_- a point process on $\mathbb{R}_-$ with distribution ζ_{N_-} and $N_{\leq 0}$ defined by (2.68). If $\Phi^h_{-,\zeta_{N_-}}(t,s) := \mathbb{E}\left[\int_{-\infty}^{t-} h(t-z)N_{\leq 0}(\mathrm{d}z)\middle| \mathcal{E}_{t,s}(N_{\leq 0})\right]$, for all $s,t\geq 0$, then,*

$$\Phi^h_{-,\zeta_{N_-}}(t,s) = \mathbb{E}\left[\sum_{T\in N_-}\left(h(t-T)+K_s(t,T)\right)\middle|\mathcal{E}_{t,s}(N_{\leq 0})\right], \tag{2.79}$$

where $K_s(t,u)$ is given by (2.35).

Proof. Following the decomposition given in Proposition 2.8.4, one has

$$\Phi^h_{-,\zeta_{N_-}}(t,s) = \mathbb{E}\left[\sum_{T\in N_-}\left(h(t-T) + \sum_{V\in N_1^T}\left(h(t-V)+\sum_{W\in N_c^{T,V}} h(t-V-W)\right)\right)\middle|\mathcal{E}_{t,s}(N_{\leq 0})\right],$$

where $\mathcal{E}_{t,s}(N_{\leq 0}) = \mathcal{E}_{t,s}(N_-)\bigcap_{T'\in N_-}\left(\mathcal{E}_{t,s}(N_1^T)\bigcap_{V'\in N_1^T}\mathcal{E}_{t-V',s}(N_c^{V'})\right)$. Let us fix $T\in N_-$, $V\in N_1^T$ and compute the conditional expectation of the inner sum with respect to N_- and N_1^T. In the same way as for (2.78) we end up with

$$\mathbb{E}\left[\sum_{W\in N_c^{T,V}} h(t-V-W)\middle| N_-, N_1^T, \mathcal{E}_{t-V,s}(N_c^{T,V})\right] = L^{h,h}_s(t-V),$$

since, conditionally on N_- and N_1^T, $N_c^{T,V}$ has the same distribution as N_c. Using the conditional independence of the cluster processes $(N_c^{T,V})_{V\in N_1^T}$ with respect to $\left(N_-,(N_1^T)_{T\in N_-}\right)$, one can apply Lemma 2.8.6 with $Y = \left(N_-,(N_1^T)_{T\in N_-}\right)$ and $X_{(T,V)} = N_c^{T,V}$ and deduce that

$$\Phi^h_{-,\zeta_{N_-}}(t,s) = \mathbb{E}\left[\sum_{T\in N_-}\left(h(t-T)+\sum_{V\in N_1^T}\left(h(t-V)+L^{h,h}_s(t-V)\right)\right)\middle|\mathcal{E}_{t,s}(N_{\leq 0})\right].$$

Let us fix $T \in N_-$ and compute the conditional expectation of the inner sum with respect to N_- which is

$$\Gamma := \mathbb{E}\left[\sum_{V\in N_1^T}\big(h(t-V)+L_s^{h,h}(t-V)\big)\middle| N_-, A_{t,s}^T\right], \tag{2.80}$$

where $A_{t,s}^T = \mathcal{E}_{t,s}(N_1^T)\cap\left(\bigcap_{V'\in N_1^T}\mathcal{E}_{t-V',s}(N_c^{T,V'})\right)$. For every $V\in N_1^T$, we say that V has mark 0 if V has no descendant or himself in $[t-s,t)$ and mark 1 otherwise. Let us denote $N_1^{T,0}$ the set of points with mark 0 and $N_1^{T,1} = N_1^T \setminus N_1^{T,0}$.

For any $V\in N_1^T$, $\mathbb{P}\left(V\in N_1^{T,0}\middle| N_1^T\right) = G_s^{h,h}(t-V)\mathbb{1}_{[t-s,t)^c}(V)$ and all the marks are chosen independently given N_1^T. Hence, $N_1^{T,0}$ and $N_1^{T,1}$ are independent Poisson processes and the intensity of $N_1^{T,0}$ is given by $\lambda(v) = h(v-T)\mathbb{1}_{[0,+\infty)}(v)G_s^{h,h}(t-v)\mathbb{1}_{[t-s,t)^c}(v)$. Moreover, $A_{t,s}^T$ is the event $\left\{N_1^{T,1}=\emptyset\right\}$, so

$$\begin{aligned}
\Gamma &= \mathbb{E}\left[\sum_{V\in N_1^{T,0}}\big(h(t-V)+L_s^{h,h}(t-V)\big)\middle| N_-, \left\{N_1^{T,1}=\emptyset\right\}\right]\\
&= \int_{-\infty}^{t-}\left[h(t-v)+L_s^{h,h}(t-v)\right]h(v-T)\mathbb{1}_{[0,+\infty)}(v)G_s^{h,h}(t-v)\mathbb{1}_{[t-s,t)^c}(v)dv\\
&= K_s(t,T).
\end{aligned}$$

Using the independence of the cluster processes, one can apply Lemma 2.8.6 with $Y = N_-$ and $X_T = \left(N_1^T, (N_c^{T,V})_{V\in N_1^T}\right)$ and (2.79) clearly follows. □

Lemma 2.8.10. *Under the assumptions and notations of Proposition 2.8.5 and Lemma 2.5.2, the function $\Phi^h_{-,\zeta_{N_-}}$ of Proposition 2.8.5 can be identified with (2.40) under $(\mathcal{A}^1_{N_-})$ and with (2.41) under $(\mathcal{A}^2_{N_-})$ and (2.72) is satisfied in those two cases.*

Proof. First, remind the expression of $\Phi^h_{-,\zeta_{N_-}}$ obtained in Lemma 2.8.9, that is

$$\Phi^h_{-,\zeta_{N_-}}(t,s) = \mathbb{E}\left[\sum_{T\in N_-}\big(h(t-T)+K_s(t,T)\big)\middle|\mathcal{E}_{t,s}(N_{\le 0})\right].$$

– **Under $\left(\mathcal{A}^1_{N_-}\right)$.** On the one hand, for every $t\ge 0$,

$$\begin{aligned}
\mathbb{E}\left[\int_{-\infty}^0 h(t-x)N_-(\mathrm{d}x)\right] &= \mathbb{E}\left[h(t-T_0)\right]\\
&= \int_{-\infty}^0 h(t-t_0)f_0(t_0)\mathrm{d}t_0 \le ||f_0||_{L^\infty}\int_0^{+\infty}h(y)dy,
\end{aligned}$$

hence ζ_{N_-} satisfies (2.72). On the other hand, since N_- is reduced to one point T_0,

$$\Phi^h_{-,\zeta_{N_-}}(t,s) = \frac{1}{\mathbb{P}\left(\mathcal{E}_{t,s}(N_{\le 0})\right)}\mathbb{E}\left[\left(h(t-T_0)+K_s(t,T_0)\right)\mathbb{1}_{\mathcal{E}_{t,s}(N_{\le 0})}\right],$$

using the definition of the conditional expectation. To compute $\mathbb{P}(\mathcal{E}_{t,s}(N_{\leq 0})|T_0)$, we use the decomposition

$$\mathcal{E}_{t,s}(N_{\leq 0}) = \{T_0 < t-s\} \cap \mathcal{E}_{t,s}(N_1^{T_0}) \cap \left(\bigcap_{V \in N_1^{T_0}} \mathcal{E}_{t-V,s}(N_c^{T_0,V}) \right)$$

and the fact that, conditionally on $N_1^{T_0}$, for all $V \in N_1^{T_0}$, $N_c^{T_0,V}$ has the same distribution as N_c to deduce that

$$\mathbb{E}\left[\mathbb{1}_{\mathcal{E}_{t,s}(N_{\leq 0})} \middle| T_0\right] = \mathbb{1}_{T_0<t-s}\mathbb{E}\left[\mathbb{1}_{\mathcal{E}_{t,s}(N_1^{T_0})}\middle|T_0\right]\mathbb{E}\left[\prod_{V\in N_1^{T_0}\cap[t-s,t)^c} G_s(t-V)\middle|T_0\right],$$

because the event $\mathcal{E}_{t,s}(N_1^{T_0})$ involves $N_1^{T_0}\cap[t-s,t)$ whereas the product involves $N_1^{T_0}\cap[t-s,t)^c$, both of those processes being two independent Poisson processes. Their respective intensities are $\lambda(x) = h(x-T_0)\mathbb{1}_{[(t-s)\vee 0,t)}(x)$ and $\lambda(x) = h(x-T_0)\mathbb{1}_{[0,(t-s)\vee 0)}(x)$, so we end up with

$$\begin{cases} \mathbb{E}\left[\mathbb{1}_{\mathcal{E}_{t,s}(N_1^{T_0})}\middle|T_0\right] = \exp\left(-\int_{t-s}^t h(x-T_0)\mathbb{1}_{[0,+\infty)}(x)\mathrm{d}x\right) \\ \mathbb{E}\left[\prod_{V\in N_1^{T_0}\cap[t-s,t)^c} G_s(t-V)\middle|T_0\right] = \exp\left(-\int_0^{(t-s)\vee 0}\left[1-G_s(t-x)\right]h(x-T_0)\mathrm{d}x\right). \end{cases}$$

The product of these two last quantities is exactly $q(t,s,T_0)$ given by (2.35). Note that $q(t,s,T_0)$ is exactly the probability that T_0 has no descendant in $[t-s,t)$ given T_0. Hence, $\mathbb{P}\left(\mathcal{E}_{t,s}(N_{\leq 0})\right) = \int_{-\infty}^{0\wedge(t-s)} q(t,s,t_0)f_0(t_0)\mathrm{d}t_0$ and (2.40) clearly follows.

– **Under $\left(\mathcal{A}^2_{N_-}\right)$.** On the one hand, for any $t \geq 0$,

$$\mathbb{E}\left[\int_{-\infty}^0 h(t-x)N_-(\mathrm{d}x)\right] = \mathbb{E}\left[\int_{-\infty}^0 h(t-x)\alpha\mathrm{d}x\right] \leq \alpha\int_0^{+\infty} h(y)dy,$$

hence ζ_{N_-} satisfies (2.72). On the other hand, since we are dealing with a Poisson process, we can use the same argumentation of marked Poisson processes as in the proof of Lemma 2.8.7. For every $T \in N_-$, we say that T has mark 0 if T has no descendant or himself in $[t-s,t)$ and mark 1 otherwise. Let us denote N_-^0 the set of points with mark 0 and $N_-^1 = N_- \setminus N_-^0$. For any $T \in N_-$, we have

$$\mathbb{P}\left(T \in N_-^0 \middle| N_-\right) = q(t,s,T)\mathbb{1}_{[t-s,t)^c}(T),$$

and all the marks are chosen independently given N_-. Hence, N_-^0 and N_-^1 are independent Poisson processes and the intensity of N_-^0 is given by

$$\lambda(z) = \alpha\mathbb{1}_{z\leq 0}\, q(t,s,z)\mathbb{1}_{[t-s,t)^c}(z)$$

Moreover, $\mathcal{E}_{t,s}(N_{\leq 0}) = \left\{N_-^1 = \emptyset\right\}$. Hence,

$$\Phi^h_{-,\zeta_{N_-}}(t,s) = \mathbb{E}\left[\sum_{T\in N_-^0}\left(h(t-T)+K_s(t,T)\right)\middle| N_-^1 = \emptyset\right]$$

which gives (2.41) thanks to the independence of N_-^0 and N_-^1.

□

Proof of Propositions 2.5.4 and 2.5.9

Since we already proved Proposition 2.8.5 and Lemma 2.8.10, to obtain Proposition 2.5.4 it only remains to prove that $\Phi^{\mu,h}_{\zeta_{N_-}} \in L^\infty(\mathbb{R}^2_+)$, to ensure uniqueness of the solution by Remark 2.5.3. To do so, it is easy to see that the assumption $h \in L^\infty(\mathbb{R}_+)$ combined with Lemma 2.5.2 giving that $G_s \in [0,1]$ and $L_s \in L^1(\mathbb{R}_+)$ ensures that $\Phi^{\mu,h}_+, q$ and K_s are in $L^\infty(\mathbb{R}_+)$. In turn, this implies that $\Phi^h_{-,\zeta_{N_-}}$ in both (2.40) and (2.41) is in $L^\infty(\mathbb{R}_+)$, which concludes the proof of Proposition 2.5.4.

Proof of Proposition 2.5.9 The method of characteristics leads us to rewrite the solution v of (2.36)–(2.37) by defining $f^{in} \equiv v^{in}$ on $\mathbb{R}_+$, $f^{in} \equiv 1$ on $\mathbb{R}_-$ such that

$$v(t,s) = \begin{cases} f^{in}(s-t)e^{-\int_{(t-s)\vee 0}^{t} \Phi(y,s-t+y)\,\mathrm{d}y}, & \text{when } s \geq t \\ f^{in}(s-t)e^{-\int_{(s-t)\vee 0}^{s} \Phi(y+t-s,y)\,\mathrm{d}y}, & \text{when } t \geq s. \end{cases} \tag{2.81}$$

Let $\zeta^M_{N_-}$ be the distribution of the past given by $\left(\mathcal{A}^1_{N_-}\right)$ and $T_0 \sim \mathcal{U}([-M-1,-M])$. By Proposition 2.5.4, let v_M be the solution of System (2.36)–(2.37) with $\Phi = \Phi^{\mu,h}_{\zeta^M_{N_-}}$ and $v^{in} = v^{in}_M$, (i.e. the survival function of a uniform variable on $[-M-1,-M]$). Let also v^∞_M be the solution of System (2.36)–(2.37) with $\Phi = \Phi^{\mu,h}_{\zeta^M_{N_-}}$ and $v^{in} \equiv 1$, and v_∞ the solution of (2.43)-(2.44). Then,

$$\|v_M - v^\infty\|_{L^\infty((0,T)\times(0,S))} \leq \|v_M - v^\infty_M\|_{L^\infty((0,T)\times(0,S))} + \|v^\infty_M - v^\infty\|_{L^\infty((0,T)\times(0,S))}.$$

By definition of v^{in}_M, it is clear that $v^{in}_M(s) = 1$ for $s \leq M$, so that Formula (2.81) implies that $v_M(t,s) = v^\infty_M(t,s)$ as soon as $s - t \leq M$ and so $\|v_M - v^\infty_M\|_{L^\infty((0,T)\times(0,S))} = 0$ as soon as $M \geq S$.

To evaluate the distance $\|v^\infty_M - v^\infty\|_{L^\infty((0,T)\times(0,S))}$, it remains to prove that

$$\exp\left(-\int_0^t \Phi^h_{-,\zeta^M_{N_-}}(y, s-t+y)\,\mathrm{d}y\right) \to 1$$

uniformly on $(0,T)\times(0,S)$ for any $T > 0$, $S > 0$. For this, it suffices to prove that $\Phi^h_{-,\zeta^M_{N_-}}(t,s) \to 0$ uniformly on $(0,T)\times(0,S)$. Since q given by (2.35) takes values in $[\exp(-2||h||_{L^1}), 1]$, (2.40) implies

$$\Phi^h_{-,\zeta^M_{N_-}}(t,s) \leq \frac{\int_{-\infty}^{0\wedge(t-s)} (h(t-t_0) + K_s(t,t_0))\,\mathbb{1}_{[-M-1,-M]}(t_0)\mathrm{d}t_0}{\int_{-\infty}^{0\wedge(t-s)} \exp(-2||h||_{L^1})\mathbb{1}_{[-M-1,-M]}(t_0)\mathrm{d}t_0}.$$

Since $||G_s||_{L^\infty} \leq 1$, L_s and h are non-negative, it is clear that

$$K_s(t,t_0) \leq \int_0^{+\infty} [h(t-x) + L_s(t-x)]\,h(x-t_0)\mathrm{d}x,$$

and so

$$\begin{aligned} \int_{-M-1}^{-M} K_s(t,t_0)\mathrm{d}t_0 &\leq \int_0^{+\infty} [h(t-x) + L_s(t-x)] \left(\int_{-M-1}^{-M} h(x-t_0)\mathrm{d}t_0\right)\mathrm{d}x \\ &\leq \int_M^{+\infty} h(y)dy \int_0^{+\infty} [h(t-x) + L_s(t-x)]\,\mathrm{d}x \\ &\leq \int_M^{+\infty} h(y)dy\,[||h||_{L^1} + ||L_s||_{L^1}]\,. \end{aligned}$$

Hence, for M large enough

$$\Phi^h_{-,\zeta^M_{N_-}}(t,s) \leq \frac{\int_M^{+\infty} h(y)dy\,[||h||_{L^1} + ||L_s||_{L^1}]}{\exp(-2||h||_{L^1})} \to 0,$$

uniformly in (t,s) since L_s is uniformly bounded in L^1, which concludes the proof.

2.8.d) Thinning

The demonstration of Ogata's thinning algorithm uses a generalization of point processes, namely the marked point processes. However, only the basic properties of simple and marked point processes are needed (see [16] for a good overview of point processes theory). Here $(\mathcal{F}_t)_{t>0}$ denotes a general filtration such that $\mathcal{F}^N_t \subset \mathcal{F}_t$ for all $t>0$, and not necessarily the natural one, i.e. $(\mathcal{F}^N_t)_{t>0}$.

Theorem 2.8.11. *Let Π be a $(\mathcal{F}_t)$-Poisson process with intensity 1 on $\mathbb{R}^2_+$. Let $(\lambda_t)_{t\geq 0}$ be a non-negative $(\mathcal{F}_t)$-predictable process which is L^1_{loc} a.s. and define the point process N by*

$$N(C) = \int_{C\times\mathbb{R}_+} \mathbf{1}_{[0,\lambda_t]}(z)\,\Pi(\mathrm{d}t\times\mathrm{d}z),$$

for all $C\in\mathcal{B}(\mathbb{R}_+)$. Then N admits λ_t as a $(\mathcal{F}_t)$-predictable intensity.

Moreover, if $(\lambda_t)_{t\geq 0}$ is in fact $\left(\mathcal{F}^N_t\right)$-predictable, then N admits λ_t as a $\left(\mathcal{F}^N_t\right)$-predictable intensity.

Proof. The goal is to apply the martingale characterization of the intensity (Chapter II, Theorem 9 in [16]). We cannot consider Π as a point process on $\mathbb{R}_+$ marked in $\mathbb{R}_+$ (in particular, the point with the smallest abscissa cannot be defined). However, for every $k\in\mathbb{N}$, we can define $\Pi^{(k)}$, the restriction of Π to the points with ordinate smaller than k, by $\Pi^{(k)}(C) = \int_C \Pi(\mathrm{d}t\times\mathrm{d}z)$ for all $C\in\mathcal{B}(\mathbb{R}_+\times[0,k])$. Then $\Pi^{(k)}$ can be seen as a point process on $\mathbb{R}_+$ marked in $E_k := [0,k]$ with intensity kernel $1.\mathrm{d}z$ with respect to $(\mathcal{F}_t)$. In the same way, we define $N^{(k)}$ by

$$N^{(k)}(C) = \int_{C\times\mathbb{R}_+} \mathbf{1}_{z\in[0,\lambda_t]}\,\Pi^{(k)}(\mathrm{d}t\times\mathrm{d}z) \quad \text{for all} \quad C\in\mathcal{B}(\mathbb{R}_+).$$

Let $\mathcal{P}(\mathcal{F}_t)$ be the predictable σ-algebra (see [16, page 8]). Let us denote $\mathcal{E}_k = \mathcal{B}([0,k])$ and $\tilde{\mathcal{P}}_k(\mathcal{F}_t) = \mathcal{P}(\mathcal{F}_t)\otimes\mathcal{E}_k$ the associated marked predictable σ-algebra.

For any fixed z in E, $\{(t',\omega)\in\mathbb{R}_+\times\Omega \text{ such that } \lambda_{t'}(\omega)\geq z\}$ belongs to the predictable σ-algebra $\mathcal{P}(\mathcal{F}_t)$ since λ is predictable. Let us denote

$$\Gamma_k := \{(t',\omega,z)\in\mathbb{R}_+\times\Omega\times E_k,\ \lambda_{t'}(\omega)\geq z\}$$

and remark that for all k,

$$\Gamma_k = \bigcap_{n\in\mathbb{N}^*}\bigcup_{q\in\mathbb{Q}_+} \{(t',\omega)\in\mathbb{R}_+\times\Omega,\ \lambda_{t'}(\omega)\geq q\}\times\left(\left[0,q+\frac{1}{n}\right]\cap E_k\right).$$

So, $\Gamma_k\in\tilde{\mathcal{P}}_k(\mathcal{F}_t)$ and $\mathbf{1}_{z\in[0,\lambda_{t'}]\cap E_k}$ is $\tilde{\mathcal{P}}_k(\mathcal{F}_t)$-measurable. Hence, one can apply the Integration Theorem (Chapter VIII, Corollary 4 in [16]). So,

$$(X_t)_{t\geq 0} := \left(\int_0^t\int_{E_k} \mathbf{1}_{z\in[0,\lambda_{t'}]}\,\bar{M}^{(k)}(dt'\times dz)\right)_{t\geq 0} \text{ is a } (\mathcal{F}_t)\text{-local martingale}$$

where $\bar{M}^{(k)}(dt' \times dz) = \Pi^{(k)}(dt' \times dz) - dzdt'$. In fact,

$$X_t = N_t^{(k)} - \int_0^t \min(\lambda_{t'}, k)\, dt'.$$

Yet, $N_t^{(k)}$ (respectively $\int_0^t \min(\lambda_{t'}, k)\, dt'$) is non-decreasingly converging towards N_t (resp. $\int_0^t \lambda_{t'} dt'$). Both of the limits are finite a.s. thanks to the local integrability of the intensity (see page 27 of [16]). Thanks to monotone convergence we deduce that $\left(N_t - \int_0^t \lambda_{t'} dt'\right)_{t\geq 0}$ is a $(\mathcal{F}_t)$-local martingale. Then, thanks to the martingale characterization of the intensity, N_t admits λ_t as an $(\mathcal{F}_t)$-intensity. The last point of the Theorem is a reduction of the filtration. Since $(\lambda_t)_{t\geq 0}$ is $\left(\mathcal{F}_t^N\right)$-predictable, it is a fortiori $\left(\mathcal{F}_t^N\right)$-progressive and the desired result follows (see [16, page 27]). □

2.9 Unpublished material

This section has not been published in *Mathematical Models and Methods in Applied Sciences* as the previous sections. Here is given some insights about: the link between strong and weak formulations and the route from the equation satisfied by u to the equation satisfied by $v := \int u$.

2.9.a) From the strong to the weak formulation

In this section, we give heuristics on the transition from System (2.12)-(2.14) to Equation (2.15). Integrating Equation (2.12) with respect to a test function φ would give

$$\int_{\mathbb{R}_+\times\mathbb{R}_+} \varphi(t,s)\Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big)U(\mathrm{d}t, \mathrm{d}s) + \int_{\mathbb{R}_+\times\mathbb{R}_+} \varphi(t,s)\left(\int_{x=0}^{\lambda_t} \Pi(\mathrm{d}t, \mathrm{d}x)\right) U(t, \mathrm{d}s) = 0,$$

yet $(\frac{\partial}{\partial t} + \frac{\partial}{\partial s})U$ does not make sense since U is not differentiable in general (U is a measure). However, the heuristics given by the integration by parts formula yields

$$\begin{aligned}\int_{\mathbb{R}_+\times\mathbb{R}_+} \varphi(t,s)\Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big)U(\mathrm{d}t, \mathrm{d}s) = &- \int_{\mathbb{R}_+\times\mathbb{R}_+} \Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big)\varphi(t,s)U(\mathrm{d}t, \mathrm{d}s) \\ &+ \int_{\mathbb{R}_+} [\varphi(t,s)U(t, \mathrm{d}s)]_{t=0}^{+\infty} + \int_{\mathbb{R}_+} [\varphi(t,s)U(\mathrm{d}t, s)]_{s=0}^{+\infty}\end{aligned}$$

yet, since φ is compactly supported,

$$[\varphi(t,s)U(t, \mathrm{d}s)]_{t=0}^{+\infty} = -\varphi(0,s)U(0, \mathrm{d}s) \quad \text{and} \quad [\varphi(t,s)U(\mathrm{d}t, s)]_{s=0}^{+\infty} = -\varphi(t,0)U(\mathrm{d}t, 0).$$

Using the boundary condition (2.13) and the initial condition (2.14) gives

$$\begin{aligned}\int_{\mathbb{R}_+} \varphi(0,s)U(0, \mathrm{d}s) + \int_{\mathbb{R}_+} \varphi(t,0)U(\mathrm{d}t, 0) = \varphi(0, -T_0) + \\ \int_{\mathbb{R}_+\times\mathbb{R}_+} \varphi(t,0)\left(\int_{x=0}^{\lambda_t} \Pi(\mathrm{d}t, \mathrm{d}x)\right) U(t, \mathrm{d}s).\end{aligned}$$

Hence (2.15) clearly follows.

2.9.b) From Equation (2.21) to System (2.36)-(2.37)

In terms of strong sense formulation, Relation (2.42) together with System (2.18)-(2.20) satisfied by u gives that v satisfies (2.36)-(2.37). Yet this is not so clear when the equations are to be understood in the weak sense which is the case here since the solutions we are looking at are possibly measures. However, the following trick connects the two systems of equations in the weak sense.

First, remind that $v(t,s) = \int_s^{+\infty} u(t, \mathrm{d}\sigma)$ so that, by integration by parts, for any test function φ in $\mathcal{C}^\infty_{c,b}(\mathbb{R}^2)$,

$$\begin{aligned}\int_{\mathbb{R}_+\times\mathbb{R}_+} \varphi(t,s)u(\mathrm{d}t,\mathrm{d}s) &= \int_{\mathbb{R}_+\times\mathbb{R}_+} \frac{\partial}{\partial s}\varphi(t,s)v(t,s)\mathrm{d}t\mathrm{d}s - \int_{t\in\mathbb{R}_+} [\varphi(t,s)v(t,s)]_{s=0}^{+\infty}\mathrm{d}t \\ &= \int_{\mathbb{R}_+\times\mathbb{R}_+} \frac{\partial}{\partial s}\varphi(t,s)v(t,s)\mathrm{d}t\mathrm{d}s + \int_{t\in\mathbb{R}_+} \varphi(t,0)v(t,0)\mathrm{d}t. \quad (2.82)\end{aligned}$$

Then, by integration by parts, Equation (2.42) gives

$$\begin{aligned}\int_{\mathbb{R}_+\times\mathbb{R}_+} \varphi(t,s)\rho_{\lambda,\zeta_{N_-}}(t,s)u(\mathrm{d}t,\mathrm{d}s) = \int_{\mathbb{R}_+\times\mathbb{R}_+} \Big(\frac{\partial}{\partial s}\varphi(t,s)\Big)\Phi^{\mu,h}_{\zeta_{N_-}}(t,s)v(t,s)\mathrm{d}t\mathrm{d}s \\ + \int_{t\in\mathbb{R}_+} \varphi(t,0)\Phi^{\mu,h}_{\zeta_{N_-}}(t,0)v(t,0)\mathrm{d}t, \quad (2.83)\end{aligned}$$

where we used that φ is compactly supported. Finally, integration by parts once again gives, with $v^{in} := v(0,\cdot)$,

$$\int_{\mathbb{R}_+} \varphi(0,s)u^{in}(\mathrm{d}s) = \int_{\mathbb{R}_+} \frac{\partial}{\partial s}\varphi(0,s)v^{in}(s)\mathrm{d}s + \varphi(0,0)v^{in}(0). \quad (2.84)$$

Hence, starting from the weak formulation (2.21), using (2.83), (2.84) and (2.82) applied to $(\frac{\partial}{\partial t} + \frac{\partial}{\partial s})\varphi$ gives, with $\tilde{\varphi} := \frac{\partial}{\partial s}\varphi$,

$$\begin{aligned}\int_{\mathbb{R}_+\times\mathbb{R}_+} \Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big)\tilde{\varphi}(t,s)v(t,s)\mathrm{d}t\mathrm{d}s + \int_{\mathbb{R}_+} \tilde{\varphi}(t,0)v(t,0)\mathrm{d}t + \int_{\mathbb{R}_+} \frac{\partial}{\partial t}\varphi(t,0)v(t,0)\mathrm{d}t \\ - \int_{\mathbb{R}_+\times\mathbb{R}_+} \tilde{\varphi}(t,s)\Phi^{\mu,h}_{\zeta_{N_-}}(t,s)v(t,s)\mathrm{d}t\mathrm{d}s + \int_{\mathbb{R}_+} \tilde{\varphi}(0,s)v^{in}(s)\mathrm{d}s + \varphi(0,0)v^{in}(0) = 0\end{aligned}$$

which is exactly the weak formulation of System (2.36)-(2.37) since $\int \frac{\partial}{\partial t}\varphi(t,0)v(t,0)\mathrm{d}t = -\varphi(0)v^{in}(0)$ by integration by parts.

Chapter

3 Mean-field limit of generalized Hawkes processes

Abstract. We generalize multivariate Hawkes processes mainly by including a dependence with respect to the age of the process, i.e. the delay since the last point.

Within this class, we investigate the limit behaviour, when n goes to infinity, of a system of n mean-field interacting age-dependent Hawkes processes. We prove that such a system can be approximated by independent and identically distributed age dependent point processes interacting with their own mean intensity. This result generalizes the study performed in [43].

In continuity with the previous one, the second of goal this chapter is to give a proper link between these generalized Hawkes processes as microscopic models of individual neurons and the age-structured system of partial differential equations introduced by Pakdaman, Perthame and Salort in [114] as macroscopic model of neurons.

This chapter has been submitted and can be found on ArXiv [28].

Contents

3.1 Introduction

In the recent years, the self-exciting point process known as the Hawkes process [70] has been used in very diverse areas. First introduced to model earthquake replicas [85] or [111] (ETAS model), it has been used in criminology to model burglary [104], in genomic data analysis to model occurrences of genes [67, 139], in social networks analysis to model viewing or popularity [9, 35], as well as in finance [7, 8]. We refer to [93] or [164] for more extensive reviews on applications of Hawkes processes. A univariate (nonlinear) Hawkes process is a point process N admitting a stochastic intensity of the form

$$\lambda_t = \Phi\left(\int_0^{t-} h(t-z)N(dz)\right), \tag{3.1}$$

where $\Phi : \mathbb{R} \to \mathbb{R}_+$ is called the intensity function, $h : \mathbb{R}_+ \to \mathbb{R}$ is called the self-interaction function and $N(dz)$ denotes the point measure associated with N.

Such a form of the intensity is motivated by practical cases where all the previous points of the process may impact the rate of appearance of a new point. The influence of the past points is formulated in terms of the delay between those past occurrences and the present time, through the weight function h. In the natural framework where h is non-negative and Φ increasing, this choice of interaction models an excitatory phenomenon: each time the process has a jump, it excites itself in the sense that it increases its intensity and thus the probability of finding a new point. A classical case is the *linear* Hawkes process for which h is non-negative and $\Phi(x) = \mu + x$ where μ is a positive constant called the spontaneous rate. Note however that Hawkes processes can also describe inhibitory phenomena. For example, the function h may take negative values, Φ being the positive part modulo the spontaneous rate μ, i.e. $\Phi(x) = \max(0, \mu + x)$.

Hawkes processes are also really suitable to model interacting particles such as bids and asks orders in limit order books [109]. Multivariate Hawkes processes consist of multivariate point processes $(N^1, \dots, N^n)$ whose intensities are respectively given for $i = 1, \dots, n$ by

$$\lambda_t^i = \Phi_i\left(\sum_{j=1}^n \int_0^{t-} h_{j\to i}(t-z)N^j(dz)\right), \tag{3.2}$$

where $\Phi_i : \mathbb{R} \to \mathbb{R}_+$ is the intensity function associated with the particle i and $h_{j\to i}$ is the *interaction function* describing the influence of each point of N^j on the appearance of a new point onto N^i, via its intensity λ^i.

When the number of interacting particles is huge (as, for instance, financial or social networks agents), one may be willing to let the number of particles goes to infinity. This

is especially true for multivariate Hawkes processes subject to mean-field interactions. In such a case, we may indeed expect propagation of chaos, namely the particles are expected to become asymptotically independent, provided that they start from independent and identically distributed (i.i.d.) initial conditions and submitted to i.i.d. sources of noise. Mean-field type interactions involve some homogeneity and some symmetry through coefficients that depend upon the empirical measure of the processes: In the limit regime, the coefficients depend upon the common asymptotic distribution of the particles, which satisfies nonlinear dynamics, sometimes called of McKean-Vlasov type.

The study of mean-field situations for Hawkes processes was initiated by Delattre et al. [43] by considering the following form of intensity

$$\lambda_t^i = \Phi\left(\frac{1}{n}\sum_{j=1}^{n}\int_0^{t-} h(t-z)N^j(dz)\right), \tag{3.3}$$

where, in comparison with (3.2), all the Φ_i's and the $h_{j\to i}$'s are the same. In particular, it is shown in [43] that mean-field interacting Hawkes processes are well approximated, when the size of the network n goes to infinity, by i.i.d. Poisson processes of the McKean-Vlasov type in the sense that their intensity is given by the following implicit formula $\overline{\lambda}(t) = \Phi(\int_0^t h(t-z)\overline{\lambda}(z)dz)$.

In the present chapter, a generalized version of Hawkes processes with mean-field interactions, namely Age Dependent Random Hawkes Processes (ADRHP for short), is studied. For any point process N, we call *predictable age process* associated with N the predictable process $(S_{t-})_{t\geq 0}$ given by

$$S_{t-} = t - \sup\{T \in N,\, T < t\}, \qquad \text{for all } t > 0,$$

and extended by continuity in $t = 0$. In particular, its value in $t = 0$ is entirely determined by $N \cap \mathbb{R}_-$ and is well-defined as soon as there is a point therein. In comparison with the standard mean-field type Hawkes processes studied in [43] we assume here that the intensity function Φ in (3.3) (which is denoted by Ψ to avoid confusion) may also depend on the predictable age process $(S^i_{t-})_{t\geq 0}$ associated with the point process N^i, like for instance

$$\lambda_t^i = \Psi\left(S^i_{t-}, \frac{1}{n}\sum_{j=1}^{n}\int_0^{t-} h(t-z)N^j(dz)\right). \tag{3.4}$$

This more general choice for the intensity makes the main difference with [43], where the intensity is assumed to be of the simpler form (3.3) only. We then show that, instead of Poisson processes of the McKean-Vlasov type, the limit processes associated with mean-field interacting age-dependent Hawkes processes are point processes of the McKean-Vlasov type whose stochastic intensity not only depends on the time but also on the age. More precisely, for the toy example (3.4), the intensity of the limit process $\overline{N}$ would be given by the following implicit formula $\overline{\lambda}_t = \Psi(\overline{S}_{t-}, \int_0^t h(t-z)\mathbb{E}\left[\overline{\lambda}_z\right] dz)$ where $(\overline{S}_{t-})_{t\geq 0}$ is the predictable age process associated with $\overline{N}$.

Part of our analysis finds its motivation in the use of Hawkes processes for the modelling in neuroscience. First of all, at a microscopic scale, Hawkes processes are commonly used in theoretical studies [31, 68, 123, 136] to describe the time occurrences of the action potentials of different neurons. These action potentials are associated with brutal changes of the membrane potential, called *spikes* in the rest of the chapter. The motivation for using

Hawkes process is well-understood and linked with the *synaptic integration* phenomenon: the interaction functions $h_{j\to i}$ describe the fact that, whenever a neuron spikes, the membrane potential of the other neurons in the network (and thus their firing rate as well) may change. In that sense, the L^1 norm of the interaction function $h_{j\to i}$, for $j \neq i$, is the analogue of the synaptic weight of neuron j over neuron i, that is the strength of the influence of neuron j over neuron i through their synaptic connection. For example, if one considers $h_{j\to i} = \alpha_{j\to i} h$ for a fixed function h then $\alpha_{j\to i}$ represents the (relative) synaptic weight of neuron j over neuron i. Notice that in the present work we allow the functions $h_{j\to i}$ to be random and thus the synaptic weights to be random as well (as in [49] for instance).

To model a transition in the behaviour of the network at the shifting time $t = 0$, the distribution of $N \cap \mathbb{R}_-$ is considered as an initial condition of the dynamics of the point process and may be different from the distribution of a Hawkes process. Therefore, to specify the dependence of the dynamics (on $\mathbb{R}_+$) upon the initial condition, the following form of intensity can be considered:

$$\lambda_t = \Phi\left(\int_{-\infty}^{t-} h(t-z)N(dz)\right) = \Phi\left(\int_0^{t-} h(t-z)N(dz) + F(t)\right), \tag{3.5}$$

where $F(t) := \int_{-\infty}^{0} h(t-z)N(dz)$ models, in a Hawkes manner, the influence of the initial condition. This choice of F is taken from Chapter 2. However, other choices are conceivable. For example, more general functions F may describe a stimulus at a given time $t_0 < 0$ which is more convenient for peristimulus analyses like [127].

However, standard Hawkes processes fail to model, in a convenient way, the neurophysiological constraint known as *refractory period*, that is the fact that a neuron cannot spike twice in a too short delay. This is the main reason why we allow the intensity of the Hawkes process to depend upon the age in the present study. In comparison with (3.1), one may represent strict refractory period by considering, for instance, the following form of intensity:

$$\lambda_t = \Phi\left(\int_0^{t-} h(t-z)N(dz)\right) \mathbb{1}_{S_{t-}\geq\delta}, \tag{3.6}$$

where $(S_{t-})_{t\geq 0}$ is the predictable age process associated with N and δ is a parameter corresponding to the time length of the strict refractory period of a neuron. This sounds as an alternative to the strategy used in Chapter 5. Therein, refractory periods are described by choosing, in the standard formulation of Hawkes processes, strongly negative self-interaction functions at a very short range. The strategy used in the present chapter is more flexible: synaptic integration and refractory period involve different aspects of the physiology of a neuron and so we prefer to describe each of them by different elements in the modelling.

Mean-field approaches have been already used to pass from a microscopic to a macroscopic description of neural networks. Taking for granted that the network is symmetric enough, the mean-field modelling sounds quite fair. Indeed neural networks admit a large number of vertices and are highly connected (see [49] for a review). One may distinguish three types of models: intrinsically spike generating models (like the FitzHugh–Nagumo model [96]), threshold spike generating models (like the integrate-and-fire model [22, 42, 41]) and point processes models ([52] or [55, 73]).

As usual with McKean-Vlasov dynamics, the asymptotic evolution (when n goes to infinity) of the distribution of the population at hand can be described as the solution

of a nonlinear partial differential equation (PDE). In the present chapter, the candidate to describe the dynamics at a macroscopic level is the following age structured system of nonlinear PDEs studied by Pakdaman, Perthame and Salort in a series of articles [114, 115, 116].

$$\begin{cases} \dfrac{\partial n\left(s,t\right)}{\partial t} + \dfrac{\partial n\left(t,s\right)}{\partial s} + p\left(s, X\left(t\right)\right) n\left(t,s\right) = 0, \\ m\left(t\right) := n\left(t,0\right) = \displaystyle\int_0^{+\infty} p\left(s, X\left(t\right)\right) n\left(t,s\right) ds. \end{cases} \tag{PPS}$$

Here, $n(t,s)$ represents the probability density of the age s of a neuron at time t where the age of a neuron is the delay since its last spike. Of course, the definition of the age of a neuron fits with the definition of the age associated with a point process as soon as the spike train is modelled by a point process. The function p represents the firing rate which may depend on the age s. As already explained, this dependence describes for instance the phenomenon of refractory period (e.g. $p(s,x) = \mathbb{1}_{s\geq\delta}$ for some $\delta > 0$). The function p may also depend on the global activity of the network which is denoted by $X(t) := \int_0^t d(z)n(t-z,0)dz$ where d is some delay function. This global (deterministic) variable $X(t)$ corresponds to the mean of the integral that appears in (3.1). This correspondence forms the basis of Chapter 2 where a bridge is made between a modified version of (PPS) and the distribution of the age of a single neuron (modelled by a point process). From a neural network point of view, this distribution can of course be recovered as the limit of the empirical distribution associated with a network of i.i.d. neurons.

The study of the link between the (PPS) system and a mean-field interacting neural network (modelled by point processes) was left as an open question in the previous chapter. The heuristic of this mean-field interpretation comes from the specific structure of the variable $X(t)$ which brings out a non-linearity of the McKean-Vlasov type. One of the main purpose of the present chapter is to answer that left open question. To be precise, this kind of study is performed in a preliminary work [130] for a firing rate p that is continuous and non-decreasing in both variables and under Markovian assumptions. Transposed to the Hawkes framework, this last point corresponds to interaction functions of the form $h_{j\to i}(t) = e^{-\beta(t-\tau_j)}\mathbb{1}_{[\tau_j,+\infty)}(t)$ where β is a constant and the τ_j's are i.i.d. random variables describing the propagation time of the signal from the neuron to the network. The convergence of the empirical measure is discussed in [130] when p is continuous only but without any rate of convergence. In the present study, rates of convergence are given for non Markovian Hawkes processes (that is non necessary exponential interaction functions) as well as for firing rates that are discontinuous with respect to the age, like (3.6) for instance. However, we make the crucial assumption that the firing rate p is Lipschitz continuous with respect to the second variable.

To sum up, we call Age Dependent Random Hawkes Process (ADRHP) a multivariate age dependent Hawkes process (like (3.6) for instance) with some general dependence with respect to the initial condition (3.5) and with some randomness regarding the interaction functions $h_{j\to i}$. This chapter has two main purposes: extend the mean-field approximation obtained in [43] to this generalization of Hawkes processes and establish a proper link between the microscopic modelling of individual neurons given by a n-particle system of mean-field interacting age-dependent Hawkes processes (like (3.4) for instance) and the macroscopic modelling given by the (PPS) system.

The chapter is organized as follows. In section 3.2, we introduce ADRHPs and we show how to represent them as solutions of an SDE driven by a Poisson noise. As a by-product

of this representation, we get, on the one hand, the existence of such processes, and on the other hand, an efficient way to get a coupling between our n-particle system and n i.i.d. limit processes. As a first step towards the mean-field approximation, the limit dynamics is studied in Section 3.3. Existence and uniqueness of a solution of the (PPS) system, which is our candidate to drive the limit dynamics, are proved in Theorem 3.3.5. As a consequence, we get the existence of point processes of the McKean-Vlasov type whose intensity depends on both the time and the age. In Section 3.4, these processes are proved to be the mean-field approximation of age dependent random Hawkes processes (Theorem 3.4.1 and Corollary 3.4.5) using coupling arguments under either of the two following main assumptions: the intensity is bounded or the intensity does not depend on the age. Notice that even when the intensity does not depend on the age, the results presented here extend the ones given in [43] since random interaction functions $h_{j\to i}$ as well as dependences with respect to the dynamics before time 0 cannot be taken into account in [43]. Finally, the link between age dependent random Hawkes processes and the (PPS) system is given by Corollary 3.4.5. For sake of readability, most of the computations and technical lemmas are given in two appendices.

General notations

- The space of continuous function from E to $\mathbb{R}$ is denoted by $\mathcal{C}(E)$.
- The space of Radon (resp. probability) measures on E is denoted by $\mathcal{M}(E)$ (resp. $\mathcal{P}(E)$).
- For ν in $\mathcal{P}(E)$, $X \sim \nu$ means that X is a random variable distributed according to ν.
- For $f : \mathbb{R} \to \mathbb{R}$, $||f||_1$, $||f||_2$ and $||f||_\infty$ respectively denote the L^1, L^2 and L^∞ norms of f.

3.2 Age dependent random Hawkes processes

In all the sequel, we focus on locally finite point processes, N, on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that are random countable sets of points of $\mathbb{R}$ such that for any bounded measurable set $A \subset \mathbb{R}$, the number of points in $N \cap A$ is finite almost surely (a.s.). The associated points define an ordered sequence of points $(T_n)_{n\in\mathbb{Z}}$. For a measurable set A, $N(A)$ denotes the number of points of N in A. We are interested in the behaviour of N on $(0, +\infty)$ and we denote $t \in \mathbb{R}_+ \mapsto N_t := N((0, t])$ the associated counting process. Furthermore, the point measure associated with N is denoted by $N(dt)$. In particular, for any non-negative measurable function f, $\int_{\mathbb{R}} f(t)N(dt) = \sum_{i\in\mathbb{Z}} f(T_i)$. For any point process N, we call *age process* associated with N the process $(S_t)_{t\geq 0}$ given by

$$S_t = t - \sup\{T \in N,\, T \leq t\}, \qquad \text{for all } t \geq 0. \tag{3.7}$$

In comparison with the age process, we call *predictable age process* associated with N the predictable process $(S_{t-})_{t\geq 0}$ given by

$$S_{t-} = t - \sup\{T \in N,\, T < t\}, \qquad \text{for all } t > 0, \tag{3.8}$$

and extended by continuity in $t = 0$.

We work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and suppose that the canonical filtration associated with N, namely $(\mathcal{F}_t^N)_{t\geq 0}$ defined by $\mathcal{F}_t^N := \sigma(N \cap (-\infty, t])$, is such that for all $t \geq 0$, $\mathcal{F}_t^N \subset \mathcal{F}_t$. Let us denote $\mathbb{F} := (\mathcal{F}_t)_{t\geq 0}$. We call $\mathbb{F}$-(predictable) intensity of N any non-negative $\mathbb{F}$-predictable process $(\lambda_t)_{t\geq 0}$ such that $(N_t - \int_0^t \lambda_s ds)_{t\geq 0}$ is an $\mathbb{F}$-local martingale. Informally, $\lambda_t dt$ represents the probability that the process N has a new point in $[t, t+dt]$ given $\mathcal{F}_{t-}$. Under some assumptions that are supposed here, this intensity process exists, is essentially unique and characterizes the point process (see [16] for more insights). In particular, since N admits an intensity, for any $t \geq 0$, the probability that t belongs to N is null. Moreover, notice the following properties satisfied by the age processes:

- the two age processes are equal for all $t \geq 0$ except the positive times T in N (almost surely a set of null measure in $\mathbb{R}_+$),
- for any $t \geq 0$, $S_{t-} = S_t$ almost surely (since N admits an intensity),
- and the value $S_{0-} = S_0$ is entirely determined by $N \cap \mathbb{R}_-$ and is well-defined as soon as there is a point therein.

In analogy with the study of the dynamics of a variable over time, we use a dichotomy between the behaviour of the point process before time 0 (which is treated as an initial condition) and its behaviour after time 0 (which is supposed to admit a "Hawkes type" intensity). For every point process N, we denote $N_- = N \cap \mathbb{R}_-$ and $N_+ = N \cap (0, +\infty)$. In the rest of the chapter, a point process on $\mathbb{R}$ is characterized by:

1. the distribution of N_-, namely ζ_{N_-}, which gives the dynamics of N on $\mathbb{R}_-$;
2. the $\mathbb{F}$-predictable intensity λ_t, which gives the dynamics of N on $(0, \infty)$.

In particular, ζ_{N_-} characterizes the distribution of T_0 that is the last point (spike) before time 0. Notice that the σ-algebra $\mathcal{F}_0$ is such that N_- is $\mathcal{F}_0$-measurable.

3.2.a) Parameters of the model

The definition of an *age dependent random Hawkes process* (ADRHP) is given bellow, but let us first introduce the parameters of the model:

- a positive integer n which is the number of particles (e.g. neurons) in the network (for $i = 1, \dots, n$, N^i represents the occurrences of the events (e.g. spikes) associated with the particle i);
- a distribution ζ_{N_-} determining the initial conditions $(N_-^i)_{i=1,..,n}$ which are i.i.d. point processes on $\mathbb{R}_-$ distributed according to ζ_{N_-};
- a distribution μ_H determining the matrix of interaction functions $\mathbf{H} = (H_{ij})_{1\leq i,j\leq n}$ where $H_{ij} : \mathbb{R}_+ \to \mathbb{R}$ are $\mathcal{F}_0$-measurable random functions distributed according to μ_H such that

$$\begin{cases} \text{for any fixed } i = 1, \dots, n, \text{ the variables } H_{i1}, \dots, H_{in} \text{ are independent,} \\ \text{the vectors } (H_{i1}, \dots, H_{in}) \text{ are exchangeable (with respect to } i), \\ \text{the matrix } \mathbf{H} \text{ is independent from the initial conditions } (N_-^i)_{i=1,..,n}; \end{cases} \tag{3.9}$$

- a distribution ν_F determining the matrix of functions $\mathbf{F} = (F_{ij})_{1\leq i,j\leq n}$ where $F_{ij} : \mathbb{R}_+ \to \mathbb{R}$ are $\mathcal{F}_0$-measurable random functions distributed according to ν_F such that

$$\begin{cases} \text{for any fixed } i = 1, \dots, n, \text{ the variables } F_{i1}, \dots, F_{in} \text{ are independent,} \\ \text{the vectors } (F_{i1}, \dots, F_{in}) \text{ are exchangeable (with respect to } i); \end{cases} \tag{3.10}$$

- an intensity function $\Psi : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$.

Note that the functions H_{ij}'s can in particular be equal to a given deterministic function h which corresponds to more standard Hawkes processes.

3.2.b) Definition via the intensity

The definition of an age dependent random Hawkes process is given by providing the form of its intensity.

Definition 3.2.1. *An age dependent random Hawkes process (ADRHP) with parameters $(n, \mu_H, \nu_F, \Psi, \zeta_{N_-})$ is a family $(N^i)_{i=1,..,n}$ of point processes on $\mathbb{R}$ such that $(N^i_-)_{i=1,..,n}$ is a family of i.i.d. point processes on $\mathbb{R}_-$ distributed according to ζ_{N_-} and $(N^i_+)_{i=1,..,n}$ is a family of point processes on $\mathbb{R}_+$ with $\mathbb{F}$-intensity given for all $i = 1, \dots, n$ by*

$$\lambda^i_t = \Psi\left(S^i_{t-}, \frac{1}{n}\sum_{j=1}^{n}\left(\int_0^{t-} H_{ij}(t-z)N^j_+(dz) + F_{ij}(t)\right)\right), \tag{3.11}$$

where $(S^i_{t-})_{t\geq 0}$ is the predictable age process associated with N^i defined in (3.8) and $(H_{ij})_{1\leq i,j\leq n}$ (respectively $(F_{ij})_{1\leq i,j\leq n}$) is a random matrix with entries distributed according to μ_H (resp. ν_F) and satisfying (3.9) (resp. (3.10)).

Remark that the intensities depend on the predictable age processes and not the standard ones since an intensity process must be predictable. An age dependent random Hawkes process admits two different behaviours:

1. before time 0, the processes $(N^i_-)_{i=1,\dots,n}$ are independent and identically distributed;
2. after time 0, the processes $(N^i_+)_{i=1,\dots,n}$ are dependent (in general) and driven by their respective intensities which can be different from one process to another.

Remark 3.2.2.

1. *The dichotomy of behaviours can model a change of regime at time $t = 0$. It should be interesting to see whether the results could be extended to initial conditions given by a mean-field dynamics and not necessarily i.i.d. ones. However, it is not in the scope of this chapter.*
2. *Assumptions (3.9) and (3.10) mean that a single particle receives i.i.d. interactions from its neighbours and that the particles are exchangeable: one can permute the particles without modifying their joint distribution.*
3. *Given $\mathcal{F}_0$, the randomness of λ^i_t in Equation (3.11) only lies in the point measures $N^j_+(dz)$ and the predictable age process $(S^i_{t-})_{t\geq 0}$. These intensities, and so the point processes, are not exchangeable given $\mathcal{F}_0$. However, they are exchangeable when they are considered with respect to all the randomness (including the N^i_-'s, H_{ij}'s and F_{ij}'s).*

4. *The link between synaptic weights (that is the strength with which one given neuron influences an other one) and interaction functions can be well emphasized by the following choice of interaction functions. Consider a fixed function* $h : \mathbb{R}_+ \to \mathbb{R}$ *and, independently of everything else, a sequence* $(\alpha_j)_{j=1,\dots,n}$ *of i.i.d. random variables with values in* $[0,1]$*. Then,* $(H_{ij})_{1\leq i,j\leq n}$ *defined by* $H_{ij} = \alpha_j h$ *satisfies (3.9). The* α_j*'s represent the (relative) synaptic weight of neuron* j *over all the other ones.*

   The interaction functions, even if they are random, are fixed at time 0*. The dynamics of synaptic weights is not taken into account here.*

5. *As presented in the introduction (see Equation (3.6)), a particular case we have in mind in this study is when there exists a function* $\Phi : \mathbb{R} \to \mathbb{R}_+$ *and a non-negative real number* δ *such that*

$$\Psi(s,x) = \Phi(x)\mathbb{1}_{s\geq\delta}. \tag{3.12}$$

   This particular choice of Ψ *provides an interesting modelling of the strict refractory period of a neuron. Furthermore, when* $\delta = 0$*, there is no refractory period and one recovers more standard Hawkes processes. In particular, if* μ_H *is the Dirac mass located at some fixed function* h *and* ν_F *is the Dirac mass located at the null function, then one recovers the Hawkes processes studied in [43]. Remark that the exchangeability of the Hawkes processes studied in [43] is obvious since they have the same intensity at each time* t*.*

6. *Since the auto-interaction given by* H_{ii} *is scaled by* $1/n$*, it vanishes when* n *goes to infinity and so the asymptotic behaviour proved in this chapter (Corollary 3.4.5) remains the same if one assumes that* $H_{ii} = 0$*.*

7. *Unlike the matrix* $\mathbf{H}$*, the matrix* $\mathbf{F}$ *can depend on the initial conditions as it can be seen in the following particular case which is derived from (3.5) for instance. For the same matrix* $\mathbf{H} = (H_{ij})_{1\leq i,j\leq n}$ *as in (3.9), we may choose for all* $1 \leq i,j \leq n$*, the function* $F_{ij} : \mathbb{R}_+ \to \mathbb{R}$ *defined for all* $t \geq 0$ *by*

$$F_{ij}(t) = \int_{-\infty}^{0} H_{ij}(t-z)N_-^j(dz). \tag{3.13}$$

   These random functions are $\mathcal{F}_0$*-measurable and they satisfy the first two lines of (3.10) thanks to the independence of the* H_{ij}*'s and the* N_-^j*'s. Hence, one can consider the intensity given by (3.11) with such a choice of* F *to represent the contribution of the processes* $(N_-^i)_{i=1,\dots,n}$ *to the dynamics after time* 0*. In this example, the* F_{ij}*'s are obviously dependent from the* N_-^j*'s.*

8. *In the case of neurons modelling, one can model external inputs via the functions* F_{ij}*. For example, one could take* $F_{ij} = H_{ij}(t-\tau)$ *where* τ *is some non-positive real number that may be random (independent of anything else) modelling that all the neurons have spiked at the same time* $\tau < 0$ *thanks to a common stimulus.*

3.2.c) List of assumptions

In the present chapter, several assumptions on the parameters of the model are used depending on the context. For sake of simplicity, all these assumptions are gathered here.

Main assumptions

$\left(\mathcal{A}_{u^{\text{in}}}^{\zeta_{N_-}}\right)$:	If N_- is distributed according to ζ_{N_-} ($N_- \sim \zeta_{N_-}$) and T_0 denotes the closest point of N_- to 0, then $-T_0$ admits a density with respect to the Lebesgue measure denoted by u^{in} ("in" stands for "initial"). Furthermore, u^{in} is uniformly bounded.
$(\mathcal{A}_{\infty}^{\mu_H})$:	If $H \sim \mu_H$, then there exists a deterministic function $G : \mathbb{R}_+ \to \mathbb{R}$ such that a.s., for all $t \geq 0$, $\lvert H(t)\rvert \leq G(t)$. The smallest possible deterministic function G, denoted by M_{μ_H}, is moreover supposed to be locally integrable. In particular, $\mathbb{E}[H(t)]$ is well-defined and we let $m_{\mu_H}(t) := \mathbb{E}[H(t)]$.
$(\mathcal{A}_{1}^{\nu_F})$:	If $F \sim \nu_F$, then $t \in \mathbb{R}_+ \mapsto \mathbb{E}[\lvert F(t)\rvert]$ is locally bounded. In particular, for all $t \geq 0$, $\mathbb{E}[F(t)]$ is well-defined and we let $m_{\nu_F}(t) := \mathbb{E}[F(t)]$.
$(\mathcal{A}_{\text{Lip}}^{\Psi})$:	The function $\Psi : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$ is uniformly Lipschitz continuous with respect to the second coordinate: there exists a constant $C > 0$ such that, for all $s \geq 0$, the function $x \mapsto \Psi(s, x)$ is Lipschitz with constant C. The smallest constant C is denoted by $\text{Lip}(\Psi)$. Furthermore, $s \in \mathbb{R}_+ \mapsto \Psi(s, 0)$ is uniformly bounded.
$(\mathcal{A}_{\infty}^{\Psi})$:	The function Ψ is uniformly bounded, that is $\lvert\lvert\Psi\rvert\rvert_\infty < +\infty$.
$(\mathcal{A}_{\Psi=\Psi_0})$:	There exists a function $\Psi_0 : \mathbb{R} \to \mathbb{R}_+$ such that, for all $s \geq 0$, $\Psi(s, \cdot) = \Psi_0(\cdot)$. In this case, if $(\mathcal{A}_{\text{Lip}}^{\Psi})$ is satisfied then $\text{Lip}(\Psi)$ is rather denoted by $\text{Lip}(\Psi_0)$.

Additional assumptions

$\left(\mathcal{A}_{\infty}^{\zeta_{N_-}}\right)$:	If $N_- \sim \zeta_{N_-}$ and T_0 denotes the closest point of N_- to 0, then $-T_0$ is upper bounded a.s. that is there exists a constant $C > 0$ such that $-T_0 \leq C$ a.s. The smallest possible constant C is denoted by M_{T_0}.
$(\mathcal{A}_{\infty,2}^{\mu_H})$:	$(\mathcal{A}_{\infty}^{\mu_H})$ is satisfied and M_{μ_H} is furthermore locally square integrable.
$(\mathcal{A}_{2}^{\nu_F})$:	$(\mathcal{A}_{1}^{\nu_F})$ is satisfied and if $F \sim \nu_F$, then for all $t \geq 0$, $F(t)$ admits a variance denoted by $V_{\nu_F}(t)$ satisfying that for all $t \geq 0$, $\int_0^t V_{\nu_F}(t')^{1/2} dt' < +\infty$. Furthermore, m_{ν_F} is a continuous function.

Remark that:

- Assumptions $(\mathcal{A}_{\infty}^{\mu_H})$, $(\mathcal{A}_{1}^{\nu_F})$ and $(\mathcal{A}_{\text{Lip}}^{\Psi})$ are used to prove the existence of the n-particle system.
- Assumptions $(\mathcal{A}_{\infty,2}^{\mu_H})$, $(\mathcal{A}_{2}^{\nu_F})$ and $(\mathcal{A}_{\text{Lip}}^{\Psi})$ are used to prove the mean-field approximation under the addition of either $(\mathcal{A}_{\infty}^{\Psi})$ and $(\mathcal{A}_{u^{\text{in}}}^{\zeta_{N_-}})$ or $(\mathcal{A}_{\Psi=\Psi_0})$.

3.2.d) Representation via a stochastic differential equation

Definition 3.2.1 describes ADRHPs as weak solutions. It characterizes their distribution but not their path-wise dynamics. As it is well emphasized in [98], point processes can be either represented as weak solutions thanks to their stochastic intensity or represented as strong solutions of a stochastic differential equation (SDE) driven by Poisson noise. The idea to represent point processes as *strong* solutions of SDEs driven by Poisson measures was first

introduced by Lewis and Shedler [92], for inhomogeneous Poisson process, and extended by Ogata [110] (thinning procedure) under some weak assumptions on the intensity. It says that, if N admits $(\lambda_t)_{t\geq 0}$ as a $\mathbb{F}$-predictable intensity, then the point measure associated with N can be represented by $N(dt) = \Pi(dt \times [0, \lambda_t])$ where Π is a Poisson measure with intensity 1 on $\mathbb{R}_+^2$. This has been used to show existence or stability results for some classes of point processes by Brémaud and Massoulié in [17] or [98] and more recently to exhibit some suitable coupling between interacting Hawkes processes and their mean field approximation in [43]. We introduce here the representation of ADRHPs based on such a thinning procedure.

Representation 3.2.3. *Let $(N_-^i)_{i\geq 1}$ be some i.i.d. point processes on $\mathbb{R}_-$ distributed according to ζ_{N_-}. Let $(H_{ij})_{1\leq i,j\leq n}$ (respectively $(F_{ij})_{1\leq i,j\leq n}$) be a random matrix with entries distributed according to μ_H (resp. ν_F) and satisfying (3.9) (resp. (3.10)). Let $(\Pi^i(dt,dx))_{i\geq 1}$ be some i.i.d. $\mathbb{F}$-Poisson measures with intensity 1 on $\mathbb{R}_+^2$.*

Let $(N_t^i)_{t\geq 0}^{i=1,..,n}$ be a family of counting processes such that, for $i = 1,..,n$ and all $t \geq 0$,

$$N_t^i = \int_0^t \int_0^\infty \mathbb{1}_{\left\{x \leq \Psi\left(S_{t'-}^i, \frac{1}{n}\sum_{j=1}^n \left(\int_0^{t'-} H_{ij}(t'-z)N_+^j(dz) + F_{ij}(t')\right)\right)\right\}} \Pi^i(dt', dx), \tag{3.14}$$

where $(S_{t-}^i)_{t\geq 0}$ is the predictable age process associated with $N^i = N_-^i \cup N_+^i$ and N_+^i is the point process associated with the counting process $(N_t^i)_{t\geq 0}$. Then, $(N^i)_{i=1,..,n}$ is an age dependent random Hawkes process with parameters $(n, \mu_H, \nu_F, \Psi, \zeta_{N_-})$.

This representation is mainly used in this chapter in order to provide a suitable coupling between ADRHPs and i.i.d. point processes describing the mean-field dynamics.

Going back and forth between the weak solution of Definition 3.2.1 and the strong solution of Representation 3.2.3 is classic: the thinning Theorem (see [17, Lemma 2] or Theorem 2.8.11 for a complete proof) states that a strong solution is also a weak solution; and the Poisson inversion [17, Lemma 4] states that, from a weak solution $(N^i)_{i=1,...,n}$, one can construct Poisson measures on an enlarged probability space such that (3.14) is satisfied.

At this stage, one has two equivalent concepts of ADRHPs but no result on the existence of such processes. Indeed, if there is too much self-excitation, then there may be an infinite number of points in finite time. In the present chapter, point processes that do not explode in finite time are considered and, thanks to Representation 3.2.3, one can prove existence of these non-explosive processes.

Proposition 3.2.4. *Under $(\mathcal{A}_\infty^{\mu_H})$, $(\mathcal{A}_1^{\nu_F})$ and $(\mathcal{A}_{\text{Lip}}^{\Psi})$, there exists an ADRHP $(N^i)_{i=1,..,n}$ with parameters $(n, \mu_H, \nu_F, \Psi, \zeta_{N_-})$ such that $t \mapsto \mathbb{E}[N_t^1]$ is locally bounded.*

This result can be challenging in an infinite dimensional framework like in [43]. However, in our finite dimensional framework, it is quite clear since the ADRHP can be stochastically dominated by some multivariate *linear* Hawkes process (thanks to the Lipschitz assumption $(\mathcal{A}_{\text{Lip}}^{\Psi})$). Yet the well-posedness of linear Hawkes processes is standard thanks to their branching structure [71]. Nevertheless, a proof of Proposition 3.2.4 is given in Appendix 3.6.a).

3.3 Study of the limit dynamics

The interactions between the point processes involved in the definition of an ADRHP are of mean-field type. Therefore, the limit version of Equation (3.14) is proposed below in Equation (3.15) (informally, the empirical means involved in (3.14) are replaced by their expected values). The *limit equation* with parameters $(h, f_0, \Psi, \zeta_{N_-})$ is given by

$$\forall t > 0,\ \overline{N}_t = \int_0^t \int_0^\infty \mathbb{1}_{\left\{x \leq \Psi\left(\overline{S}_{t'-}, \int_0^{t'-} h(t'-z)\mathbb{E}\left[\overline{N}_+(dz)\right] + f_0(t')\right)\right\}} \Pi(dt', dx), \tag{3.15}$$

where h and f_0 are some functions from $\mathbb{R}_+$ to $\mathbb{R}$, $\Pi(dt', dx)$ is an $\mathbb{F}$-Poisson measure on $\mathbb{R}_+^2$ with intensity 1 and $(\overline{S}_{t-})_{t\geq 0}$ is the predictable age process associated with $\overline{N} = \overline{N}_- \cup \overline{N}_+$ where $\overline{N}_-$ is a point process distributed according to ζ_{N_-} and $\overline{N}_+$ is the point process associated with the counting process $(\overline{N}_t)_{t\geq 0}$.

Looking simultaneously at Equations (3.14) and (3.15) shows that the empirical mean of the random interaction functions H_{ij} (*respectively the random functions* F_{ij}) are replaced by h (*resp.* f_0) which should be the mean interaction function m_{μ_H} (*resp.* m_{ν_F}). Moreover, the empirical mean of the point measures $N_+^j(dz)$ in (3.14) is replaced by the expectation of the point measure $\overline{N}_+(dz)$.

Finally, let us note that the dependence with respect to the predictable age process is still present in the limit equation. This matches with experimental data in neuroscience where refractory periods are highlighted. By comparison, there is no such dependence in the limit process given in [43] which is an inhomogeneous Poisson process.

This limit equation is used in the next section to provide suitable couplings to prove the mean-field approximation. Hence, the main point of this section is to prove the well-posedness of the limit equation (3.15). However, to study the probabilistic formulation of the mean-field dynamics described in Equation (3.15), one first needs to find a representation of the distribution of a possible solution of (3.15). As a first step, we prove existence/uniqueness results for a linearisation of the (PPS) system (Proposition 3.3.1) as well as we give a representation of the solution given by the method of characteristics (Proposition 3.3.2). The second step is to deduce existence/uniqueness results for the (PPS) system (Theorem 3.3.5) from the linearised system via a fixed point argument. Then, the well-posedness of the limit equation (3.15) is proved thanks to the results obtained for the (PPS) system. Finally, the link between the (PPS) system and the processes defined by the limit equation is fully investigated.

Note that the analytical study of the fixed point equation satisfied by the expectation of the solution of (3.15) (as it is done in [43]) can be extended to the case when the intensity does depend on the age. However, the results for the (PPS) system are valid in a more general framework so they are favoured here.

3.3.a) Study of the linear system

In comparison with the (PPS) system, the linear system studied below corresponds to the case where the firing rate p in (PPS) is a function of the time t and the age s only. More

precisely, we consider the system

$$\begin{cases} \dfrac{\partial u\,(t,s)}{\partial t} + \dfrac{\partial u\,(t,s)}{\partial s} + f(t,s)u\,(t,s) = 0, \\ u\,(t,0) = \displaystyle\int_{s\in\mathbb{R}} f(t,s)u\,(t,s)\,ds, \end{cases} \tag{3.16}$$

where f is a bounded function.

We state uniqueness of the solution of this system in a measure space as a consequence of the uniqueness result stated in [26]. More precisely, the result is stated in $\mathcal{BC}(\mathbb{R}_+, \mathcal{M}(\mathbb{R}))$ that is the space of bounded continuous curves on $\mathcal{M}(\mathbb{R})$ (the space of Radon measures on $\mathbb{R}$) endowed with the *bounded Lipschitz norm* as considered in [26]. As we are interested in probability measures, let us remark that the bounded Lipschitz norm on $\mathcal{P}(\mathbb{R}_+)$ is equivalent, thanks to the duality of Kantorovich-Rubinstein, to the modified 1-Wasserstein distance defined by

$$\tilde{W}_1(\mu,\nu) := \inf \mathbb{E}\left[\min(|X-Y|,1)\right], \tag{3.17}$$

where the infimum is taken over all joint distributions of the random variables X and Y with marginals μ and ν respectively.

Since measure solutions are considered, a weak form of the system is given. The following set of test functions is used[1]:

$\mathcal{C}^\infty_{c,b}(\mathbb{R}^2_+)$

The function φ belongs to $\mathcal{C}^\infty_{c,b}(\mathbb{R}^2_+)$ if
- φ is continuous, uniformly bounded,
- φ has uniformly bounded derivatives of every order,
- there exists $T > 0$ such that $\varphi(t,s) = 0$ for all $t > T$ and $s \geq 0$.

The result stated below is a consequence of [26, Theorem 2.4.] in the same essence than the one presented in [26, Section 3.3.]. Its proof is given in Appendix 3.6.b).

Proposition 3.3.1. *Assume that $f : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ is bounded and continuous (uniformly in the second variable) with respect to the first variable. Assume that u^{in} belongs to $\mathcal{M}(\mathbb{R}_+)$.*

Then, there exists a unique solution in the weak sense u such that $t \mapsto u(t,\cdot)$ belongs to $\mathcal{BC}(\mathbb{R}_+, \mathcal{M}(\mathbb{R}_+))$ of the system (3.16) with initial condition $u(0,\cdot) = u^{\text{in}}$. The weak sense means here that for every φ in $\mathcal{C}^\infty_{c,b}(\mathbb{R}^2_+)$,

$$\int_{\mathbb{R}^2_+} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \varphi\,(t,s)\,u\,(t,ds)\,dt + \int_{\mathbb{R}_+} \varphi(0,s)u^{\text{in}}(ds) + \int_{\mathbb{R}^2_+} [\varphi(t,0) - \varphi(t,s)] f(t,s)u(t,ds)dt = 0. \tag{3.18}$$

Remark that the system is mass-conservative (e.g. take a sequence of functions converging to $t \mapsto \mathbb{1}_{[0,T]}(t)$ as test functions in the weak equation (3.18)). As we are interested in probability measures as solutions, let us remark that the mass-conservation alone cannot ensure that the solution is a probability even if the initial condition is a probability. However, when the initial condition is a probability which admits a density, the method of characteristics shows that the solution of (3.16) is a probability density function for all time $t \geq 0$.

[1] It is the same as in Chapter 2

Proposition 3.3.2. *Under the assumptions of Proposition 3.3.1, assume that u^{in} is a probability which admits a density (denoted by u^{in} as well) with respect to the Lebesgue measure. Then, there exists a unique locally bounded function $u_0 : \mathbb{R}_+ \to \mathbb{R}$ (which is furthermore non-negative) such that u defined by*

$$\begin{cases} u(t,s) = u^{\text{in}}(s-t)\exp\left(-\displaystyle\int_0^t f(t', s-t+t')dt'\right), & \text{for } s \geq t \qquad (3.19) \\ u(t,s) = u_0(t-s)\exp\left(-\displaystyle\int_0^s f(t-s+s', s')ds'\right), & \text{for } t \geq s \qquad (3.20) \end{cases}$$

is the unique solution of (3.16). In particular,

- *u satisfies the second equation of (3.16) in a strong sense,*
- *since u_0 is non-negative and the system is mass-conservative, the function $u(t,\cdot)$ is a density for all time $t \geq 0$.*

A detailed proof of this result is given in Appendix 3.6.c). Here are listed some properties of the solution u of the linear system.

Proposition 3.3.3. *Under the assumptions of Proposition 3.3.2, assume furthermore that there exists $M > 0$ such that for all $s \geq 0$, $0 \leq u^{\text{in}}(s) \leq M$.*

Then, the solution u of (3.16) is such that the function $t \mapsto u(t,\cdot)$ belongs to $\mathcal{C}(\mathbb{R}_+, L^1(\mathbb{R}_+))$, the function $t \mapsto u(t,0)$ is continuous and

$$0 \leq u(t,s) \leq \max(M, ||f||_\infty), \quad \text{for all } t, s \geq 0. \qquad (3.21)$$

Proof. The first continuity property is rather classic thanks to a fixed point argument in the space $\mathcal{C}([0,T], L^1(\mathbb{R}_+))$ for a good choice of $T > 0$ (see [121, Section 3.3.] for instance). The second one is given by the second equation of (3.16) (which is satisfied in a strong sense by the solution given by the characteristics). Indeed,

$$\begin{aligned} |u(t+t',0) - u(t,0)| &\leq \int_0^{+\infty} f(t+t',s)|u(t+t',s) - u(t,s)|ds \\ &\quad + \int_0^{+\infty} |f(t+t',s) - f(t,s)|u(t,s)ds \\ &\leq ||f||_\infty ||u(t+t',\cdot) - u(t,\cdot)||_1 + \sup_{s\geq 0} |f(t+t',s) - f(t,s)|. \end{aligned}$$

Yet, the continuity properties of both functions f and $t \mapsto u(t,\cdot)$ and give that $|u(t+t',0) - u(t,0)|$ goes to 0 as t' goes to 0, hence the continuity of $t \mapsto u(t,0)$.

Finally, one can prove that u satisfies (3.21) thanks to the representation given by the characteristics. On the one hand, the function u_0 given in Proposition 3.3.2 is non-negative and so is u. On the other hand, it follows from (3.19) that for $s \geq t$, $u(t,s) \leq M$ and it follows from the second equation of (3.16) that for all $t \geq 0$, $u(t,0) \leq ||f||_\infty$ and so (3.20) implies that for $t \geq s$, $u(t,s) \leq ||f||_\infty$. □

3.3.b) Study of the (PPS) system

Here, a global existence result for the nonlinear system (PPS) is stated under suitable assumptions. Since this result is one of the cornerstone of this work, its proof is given even if its sketch is pretty similar to the proof of [114, Theorem 5.1].

In comparison with the uniqueness for the linear system which takes place in $\mathcal{BC}(\mathbb{R}_+, \mathcal{M}(\mathbb{R}_+))$, the uniqueness result stated in this section takes place in $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$. However, this last result is sufficient for our purpose since it is applied to measures that are probabilities a priori.

First of all, a technical lemma is needed to fully understand the non linearity involved in the system (PPS).

Lemma 3.3.4. *Under ($\mathcal{A}^{\Psi}_{\mathrm{Lip}}$) and ($\mathcal{A}^{\Psi}_{\infty}$), assume that $h : \mathbb{R}_+ \to \mathbb{R}$ is locally integrable and that $f_0 : \mathbb{R}_+ \to \mathbb{R}$ is continuous. Then, for all u in $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$ there exists a unique function $X_u : \mathbb{R}_+ \mapsto \mathbb{R}$ such that*

$$X_u(t) = \int_{z=0}^{t} \int_{s=0}^{+\infty} h(t-z)\Psi\left(s, X_u(z) + f_0(z)\right) u\left(z, ds\right) dz. \tag{3.22}$$

Furthermore, the function X_u is continuous.

Proof. The proof is divided in three steps:

-1. Establish a priori estimates on X_u to show that it is locally bounded. Indeed, using the boundedness of Ψ and the fact that u belongs to $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$, one deduces that

$$X_u(t) \leq ||\Psi||_\infty \int_{z=0}^{t} \int_{s=0}^{+\infty} |h(t-z)| u\left(z, ds\right) dz \leq ||\Psi||_\infty \int_{z=0}^{t} |h(t-z)| dz.$$

Hence, the local integrability of h implies the local boundedness of X_u.

-2. Show that X_u exists and is unique as a fixed point. For any $T > 0$, consider the following application

$$\begin{array}{rccc} G_T : & L^\infty([0,T]) & \longrightarrow & L^\infty([0,T]) \\ & X & \longmapsto & \left(t \mapsto \displaystyle\int_{z=0}^{t} \int_{s=0}^{+\infty} h(t-z)\Psi\left(s, X(z) + f_0(z)\right) u\left(dz, ds\right)\right). \end{array}$$

The Lipschitz continuity of Ψ leads, for any X_1, X_2 in $L^1([0,T])$ and t in $[0,T]$, to

$$\begin{aligned} |G_T(X_1)(t) - G_T(X_2)(t)| &\leq \mathrm{Lip}(\Psi) \int_0^t |h(t-z)| \, |X_1(z) - X_2(z)| \int_{s=0}^{+\infty} u(z, ds) dz \\ &\leq \mathrm{Lip}(\Psi) \int_0^t |h(t-z)| \, |X_1(z) - X_2(z)| \, dz \\ &\leq \mathrm{Lip}(\Psi) ||X_1 - X_2||_{L^\infty([0,T])} \int_0^T |h(z)| dz. \end{aligned}$$

Fix $T > 0$ such that $\mathrm{Lip}(\Psi) \int_0^T h(z) dz \leq 1/2$ so that G_T is a contraction and admits a unique fixed point. Iterating this fixed point gives the existence and uniqueness of X_u in the space of locally bounded functions.

For instance, we give the idea for the first iteration. Denoting W the fixed point of G_T, one can consider the following application

$$\begin{array}{rccc} G^W_{2T} : & L^\infty([T,2T]) & \longrightarrow & L^\infty([T,2T]) \\ & X & \longmapsto & \left(t \mapsto \displaystyle\int_{z=0}^{t} \int_{s=0}^{+\infty} h(t-z)\Psi\left(s, \tilde{X}(z) + f_0(z)\right) u\left(z, ds\right) dz\right), \end{array}$$

where $\tilde{X}(t) = W(t)$ if $0 \leq t < T$, $\tilde{X}(t) = X(t)$ if $T \leq t \leq 2T$, and $\tilde{X}(t) = 0$ otherwise. Applying the same argument as for the fixed point of G_T leads to existence and uniqueness of the trace of X_u on $[0, 2T]$.

-3. Finally, let us show that X_u is continuous thanks to a generalized Gronwall lemma. Using the Lipschitz continuity and the boundedness of Ψ, one deduces from (3.22) that

$$|X_u(t+t') - X_u(t)| \leq \mathrm{Lip}(\Psi) \int_0^t |h(y)||(X_u + f_0)(t-y) - (X_u + f_0)(t+t'-y)|dy + ||\Psi||_\infty \int_t^{t+t'} |h(y)|dy$$

and so

$$|X_u(t+t') - X_u(t)| \leq \mathrm{Lip}(\Psi)\Big[\int_0^t |h(t-z)||X_u(z+t') - X_u(z)|dz + \int_0^t |h(t-z)||f_0(z+t') - f_0(z)|dz\Big] + ||\Psi||_\infty \int_t^{t+t'} |h(y)|dy.$$

This means that the function $Y_u^{(t')} := |X_u(\cdot + t') - X_u(\cdot)|$ satisfies

$$Y_u^{(t')} \leq g^{(t')}(t) + \mathrm{Lip}(\Psi) \int_0^t |h(t-z)| Y_u^{(t')}(z)dz$$

where $g^{(t')}(t) := \mathrm{Lip}(\Psi) \int_0^t |h(t-z)||f_0(z+t') - f_0(z)|dz + ||\Psi||_\infty \int_t^{t+t'} |h(y)|dy$. Applying Lemma 3.7.4-*(i)* gives, for any $T > 0$, $\sup_{t\in[0,T]} Y_u^{(t')}(t) \leq C_T \sup_{t\in[0,T]} g^{(t')}(t)$. Yet the continuity (hence uniform continuity on compact time intervals) of f_0 and the local integrability of h gives that $\sup_{t\in[0,T]} g^{(t')}(t) \to 0$ as $t' \to 0$. □

Now, we have all the ingredients to state the existence/uniqueness result for the (PPS) system in a measure space of possible solutions. Notice that the existence/uniqueness result used for the linear system would not directly apply to the non-linear system. In that sense, we extend the result stated in [26].

Theorem 3.3.5. *Under ($\mathcal{A}^\Psi_{\mathrm{Lip}}$) and ($\mathcal{A}^\Psi_\infty$), assume that $h : \mathbb{R}_+ \to \mathbb{R}$ is locally integrable and that $f_0 : \mathbb{R}_+ \to \mathbb{R}$ is continuous. Assume that u^{in} is a non-negative function such that both $\int_0^{+\infty} u^{\mathrm{in}}(s)ds = 1$ and there exists $M > 0$ such that for all $s \geq 0$, $0 \leq u^{\mathrm{in}}(s) \leq M$.*

Then, there exists a unique solution in the weak sense u such that $t \mapsto u(t, \cdot)$ belongs to $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$ of the following (PPS) system

$$\begin{cases} \dfrac{\partial u(t,s)}{\partial t} + \dfrac{\partial u(t,s)}{\partial s} + \Psi(s, X(t) + f_0(t))\, u(t,s) = 0, \\ u(t,0) = \displaystyle\int_{s\in\mathbb{R}} \Psi(s, X(t) + f_0(t))\, u(t,s)\, ds, \end{cases} \tag{3.23}$$

with initial condition that $u(0, \cdot) = u^{\mathrm{in}}$, where for all $t \geq 0$, $X(t) = \int_0^t h(t-z)u(z, 0)dz$. The weak sense means here that for every φ in $\mathcal{C}^\infty_{c,b}(\mathbb{R}^2_+)$,

$$\int_{\mathbb{R}^2_+} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \varphi(t,s)\, u(t, ds)\, dt + \int_{\mathbb{R}_+} \varphi(0,s) u^{\mathrm{in}}(s) ds + \int_{\mathbb{R}^2_+} [\varphi(t,0) - \varphi(t,s)] \Psi(s, X(t) + f_0(t))\, u(t, ds)dt = 0 \tag{3.24}$$

where X is the continuous function given by Lemma 3.3.4 and satisfying

$$X(t) = \int_{z=0}^{t} \int_{s=0}^{+\infty} h(t-z)\Psi\left(s, X(z) + f_0(z)\right) u\left(z, ds\right) dz. \tag{3.25}$$

Moreover, the solution u is such that, for all $t \geq 0$, the measure $u(t,\cdot)$ is a probability and admits a density which is identified to the solution itself. Furthermore, the function $t \mapsto u(t,\cdot)$ belongs to $\mathcal{C}(\mathbb{R}_+, L^1(\mathbb{R}_+))$, the function $t \mapsto u(t,0)$ is continuous and

$$0 \leq u(t,s) \leq \max(M, ||\Psi||_\infty), \quad \textit{for all } t, s \geq 0. \tag{3.26}$$

As for the linear case, the system is mass-conservative (e.g. take a sequence of functions converging to $t \mapsto \mathbb{1}_{[0,T]}(t)$ as test functions in the weak equation (3.24)).

Proof. The proof is divided in two steps. First, we apply the results of Section 3.3.a) to a linearised version of the non-linear system (3.23) and then we find the auxiliary function X corresponding to the solution u as a fixed point in a space of continuous functions in order to deal with the non linearity of the system (3.23).

-1. The linearised version of the system takes the form

$$\begin{cases} \dfrac{\partial u\left(t,s\right)}{\partial t} + \dfrac{\partial u\left(t,s\right)}{\partial s} + \Psi\left(s, Y(t) + f_0(t)\right) u\left(t,s\right) = 0, \\ u\left(t,0\right) = \displaystyle\int_{s\in\mathbb{R}} \Psi\left(s, Y(t) + f_0(t)\right) u\left(t,s\right) ds, \end{cases} \tag{3.27}$$

for a fixed continuous function Y. Note that the funciton $f : (t,s) \mapsto \Psi\left(s, Y(t) + f_0(t)\right)$ is bounded and continuous (uniformly in s) with respect to t. So the assumptions of Propositions 3.3.1, 3.3.2 and 3.3.3 are satisfied. In particular, for any continuous function Y, there exists a unique solution u_Y (with initial condition u^{in}) in $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+)) \subset \mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{M}_+))$ of the system (3.27) which furthermore satisfies the properties listed in Proposition 3.3.3.

-2. Let us notice that for all $T > 0$, if Y belongs to the Banach space $\mathcal{C}([0,T])$ equipped with the sup norm then $t \mapsto \int_0^t h(t-z)u_Y(z,0)dz$ belongs to $\mathcal{C}([0,T])$ too. Indeed, remind that $t \mapsto u_Y(t,0)$ is continuous thanks to Proposition 3.3.3 and that $\int_0^t h(t-z)u_Y(z,0)dz = \int_0^t h(z)u_Y(t-z,0)dz$. Hence one can consider the application

$$\begin{array}{rccc} F_T : & \mathcal{C}([0,T]) & \longrightarrow & \mathcal{C}([0,T]) \\ & Y & \longmapsto & \left(t \mapsto \displaystyle\int_0^t h(t-z)u_Y(z,0)dz\right), \end{array} \tag{3.28}$$

and show that it admits a fixed point for a good choice of T. Computations given in Appendix 3.6.d) provide the following statement

$$\exists T > 0, \forall Y_1, Y_2 \in L^\infty([0,T]), \quad ||F_T(Y_1) - F_T(Y_2)||_{L^\infty([0,T])} \leq \frac{1}{2}||Y_1 - Y_2||_{L^\infty([0,T])}, \tag{3.29}$$

where T depends neither on the initial condition nor on f_0.

Until the end of the proof, let fix such a T. Then, there exists a unique W in $\mathcal{C}([0,T])$ such that $F_T(W) = W$. In particular, u_W is a solution of (3.23) on $[0,T]$ so we have existence on $[0,T]$.

For the uniqueness, let us consider the trace of a solution $u \in \mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$ to (3.23) on $[0,T]$. Then, the auxiliary function X_u associated with u defined in Lemma 3.3.4 is

continuous. Since u is a solution of (3.23), the trace of X_u on $[0,T]$ is a fixed point of F_T and so $X_u = W$ and $u = u_W$. This gives the uniqueness of the solution of (3.23) on $[0,T]$ in $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$.

Taking $u_W(T,\cdot)$ instead of u^{in} as initial condition, the function $t \mapsto f_0(t+T) + \int_0^T h(t-z)u_W(z,0)dz$ instead of f_0 and applying the same find of fixed point argument gives the trace of the solution on $[T,2T]$. Iterating this fixed point argument, one deduces that there exists a unique solution u of (3.23) on $\mathbb{R}_+$ (remind that T depends neither on the initial condition nor on f_0). In particular, the iteration is possible since the boundedness of the initial condition is carried on by the equation (see Equation (3.21)).

The regularity and boundedness of the solution u, i.e. the continuity properties and Equation (3.26) listed at the end of the statement, come from the regularity and boundedness of the solutions u_Y since u is one of the u_Y's.

□

3.3.c) Limit process

Let us remind the limit equation (3.15), that is

$$\forall t > 0,\ \overline{N}_t = \int_0^t \int_0^\infty \mathbb{1}_{\left\{x \le \Psi\left(\overline{S}_{t'-}, \int_0^{t'-} h(t'-z)\mathbb{E}\left[\overline{N}_+(dz)\right] + f_0(t')\right)\right\}} \Pi(dt', dx).$$

This limit equation describes an age dependent point process interacting with its own mean intensity in an Hawkes manner. More precisely, a solution $(\overline{N}_t)_{t\ge 0}$ of (3.15), if it exists, admits an intensity $\overline{\lambda}_t$ which depends on the time t and the age $\overline{S}_{t-}$ in a McKean-Vlasov manner in the sense that it satisfies the following implicit equation $\overline{\lambda}_t = \Psi(\overline{S}_{t-}, \int_0^{t'-} h(t'-z)\mathbb{E}\left[\overline{\lambda}_z\right] dz + f_0(t'))$. Equation (3.15) is in particular a non trivial fixed point problem of the McKean-Vlasov type. Notice that the dependence of $\overline{N}$ with respect to the past $\overline{N}_-$ is reduced to the age at time 0, i.e. $\overline{S}_{0-} = -T_0$. Throughout this chapter, a solution of the limit equation is called a *point process of the McKean-Vlasov type whose intensity depends on time and on the age*.

The fixed point problem of the limit equation (3.15) is proved to be well-posed in two cases as given in the next two statements. In either case, the idea of the proof is first to compute the mean intensity (denoted by $\overline{\lambda}(t)$) of a possible solution of (3.15). The next two propositions state the same result under different sets of assumptions and can be summarized as follows:

- in the first case, the intensity is bounded, i.e. Ψ satisfies $(\mathcal{A}^{\Psi}_{\infty})$, and the mean intensity is given by the system (3.23). More precisely, $\overline{\lambda}(t) = u(t,0)$ with u given by Theorem 3.3.5.

- in the second case, the intensity does not depend on the predictable age process, i.e. Ψ satisfies $(\mathcal{A}_{\Psi=\Psi_0})$, and the mean intensity is given by a generalization of [43, Lemma 24].

Proposition 3.3.6. *Under $(\mathcal{A}^{\zeta_{N_-}}_{u^{\text{in}}})$, $(\mathcal{A}^{\Psi}_{\text{Lip}})$ and $(\mathcal{A}^{\Psi}_{\infty})$, assume that $h : \mathbb{R}_+ \to \mathbb{R}$ is locally integrable and that $f_0 : \mathbb{R}_+ \to \mathbb{R}$ is continuous. Denote by u the unique solution of (3.23) with initial condition u^{in} as given by Theorem 3.3.5 and let, for all $t \ge 0$, $\overline{\lambda}(t) := u(t,0)$. Then, the following statements hold*

(i) if $(\overline{N}_t)_{t\geq 0}$ *is a solution of* (3.15) *then* $\mathbb{E}\left[\overline{N}_+(dt)\right] = \overline{\lambda}(t)dt$,

(ii) there exists a unique (once Π *and* $\overline{N}_-$ *are fixed) solution* $(\overline{N}_t)_{t\geq 0}$ *of the following system*

$$\begin{cases} \overline{N}_t = \displaystyle\int_0^t \int_0^\infty \mathbb{1}_{\left\{x\leq \Psi\left(\overline{S}_{t'-}, \int_0^{t'} h(t'-z)\overline{\lambda}(z)dz + f_0(t')\right)\right\}} \Pi(dt', dx), \\ \mathbb{E}\left[\overline{N}_t\right] = \displaystyle\int_0^t \overline{\lambda}(t')dt', \end{cases} \tag{3.30}$$

where $(\overline{S}_{t-})_{t\geq 0}$ *is the predictable age process associated with* $\overline{N} = \overline{N}_- \cup \overline{N}_+$ *where* $\overline{N}_-$ *is a point process distributed according to* ζ_{N_-} *and* $\overline{N}_+$ *is the point process associated with the counting process* $(\overline{N}_t)_{t\geq 0}$

In particular, $\overline{\lambda}$ *is a continuous function satisfying*

$$\overline{\lambda}(t) = \mathbb{E}\left[\Psi\left(\overline{S}_{t-}, \int_0^{t-} h(t-z)\overline{\lambda}(z)dz + f_0(t)\right)\right]$$

and the solution of (3.30) *is the unique (once* Π *and* $\overline{N}_-$ *are fixed) solution of* (3.15).

Proof. -(i) Suppose that $(\overline{N}_t)_{t\geq 0}$ is a solution of (3.15). The thinning procedure implies that $(\overline{N}_t)_{t\geq 0}$ admits an intensity which only depends on the time t and the age $\overline{S}_{t-}$. This allows us to denote by f the bivariate function such that the intensity of $\overline{N}$ at time t is given by $\overline{\lambda}_t = f(t, \overline{S}_{t-})$. It satisfies for all $t, s \geq 0$,

$$f(t,s) = \Psi\left(s, \int_0^{t-} h(t-z)\mathbb{E}\left[f(z, \overline{S}_{z-})\right] dz + f_0(t)\right). \tag{3.31}$$

In particular, the intensity is bounded since Ψ is bounded. So, if we denote $w_t = w(t, \cdot)$ the distribution of the age $\overline{S}_{t-}$, Lemma 3.7.2 gives that w belongs to $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$ and Section 2.5.a) implies that w satisfies the system

$$\begin{cases} \dfrac{\partial w(t,s)}{\partial t} + \dfrac{\partial w(t,s)}{\partial s} + f(t,s)w(t,s) = 0, \\ w(t,0) = \displaystyle\int_0^{+\infty} f(t,s)w(t,s)\, ds. \end{cases}$$

Yet, by definition of w, $\mathbb{E}\left[f(z, \overline{S}_{z-})\right] = \int_{s=0}^{+\infty} f(z,s)w(z, ds)$, so (3.31) rewrites as

$$f(t,s) = \Psi\left(s, X(t) + f_0(t)\right), \tag{3.32}$$

where X satisfies

$$X(t) = \int_0^t h(t-z) \int_0^{+\infty} \Psi(s, X(z) + f_0(z))w(z, ds)dz.$$

Hence, w is a solution in $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$ of the system (3.23). Yet, the solution of (3.23) is unique (Theorem 3.3.5) so we have $w = u$ (defined in Proposition 3.3.6) and in particular $\mathbb{E}\left[\overline{N}_+(dt)\right] = \mathbb{E}\left[f(t, \overline{S}_{t-})\right] dt = u(t,0)dt = \overline{\lambda}(t)dt$.

-(ii) The first equation of (3.30) is a standard thinning equation with $\overline{\lambda}$ given by the first step so its solution $(\overline{N}_t)_{t\geq 0}$ is a measurable function of Π and $\overline{N}_-$ hence it is unique (once Π and $\overline{N}_-$ are fixed).

To conclude this step, it suffices to check that $(\overline{N}_t)_{t\geq 0}$ satisfies the second equation of (3.30). Identifying $\overline{\lambda}(t)$ with $\mathbb{E}\left[f(t,\overline{S}_{t-})\right]$ in (3.31), the intensity of $\overline{N}$ is given by

$$f(t,\overline{S}_{t-}) = \Psi\left(\overline{S}_{t-}, \int_0^{t-} h(t-z)\overline{\lambda}(z)dz + f_0(t)\right) \tag{3.33}$$

which is bounded. Hence, Section 2.5.a) implies that the distribution of the age $\overline{S}_{t-}$ denoted by $v(t,\cdot)$ is the unique solution of

$$\begin{cases} \dfrac{\partial v\,(t,s)}{\partial t} + \dfrac{\partial v\,(t,s)}{\partial s} + \Psi\left(s, \displaystyle\int_0^t h(t-z)\overline{\lambda}(z)dz + f_0(t)\right) v\,(t,s) = 0, \\ v\,(t,0) = \displaystyle\int_0^\infty \Psi\left(s, \int_0^t h(t-z)\overline{\lambda}(z)dz + f_0(t)\right) v\,(t,s)\,ds. \end{cases} \tag{3.34}$$

Since $\overline{\lambda}(t) = u(t,0)$ and u is a solution of (3.23), it is clear that u satisfies this system, so $u(t,\cdot)$ is the density of $\overline{S}_{t-}$. Finally, using Fubini's Theorem we have

$$\begin{aligned} \mathbb{E}\left[\overline{N}_t\right] = \int_0^t \mathbb{E}\Big[f(t',\overline{S}_{t'-})\Big]dt' &= \int_0^t \mathbb{E}\left[\Psi\left(\overline{S}_{t'-}, \int_0^{t'-} h(t'-z)\overline{\lambda}(z)dz + f_0(t')\right)\right]dt' \\ &= \int_0^t \int_0^\infty \Psi\left(s, \int_0^{t'} h(t'-z)\overline{\lambda}(z)dz + f_0(t')\right) u(t',s)dsdt' \\ &= \int_0^{t'} \overline{\lambda}(t')dt', \end{aligned}$$

since u satisfies the second equation of (3.34).

Finally, the three remaining points are rather simple. Firstly, the continuity of $\overline{\lambda}$ comes from Theorem 3.3.5. Secondly, using (3.33) and (i) one has

$$\overline{\lambda}(t) = \mathbb{E}\left[f(t,\overline{S}_{t-})\right] = \mathbb{E}\left[\Psi\left(\overline{S}_{t-}, \int_0^{t-} h(t-z)\overline{\lambda}(z)dz + f_0(t)\right)\right].$$

Lastly, the solution of (3.30) is clearly a solution of (3.15) and (i) tells that a solution of (3.15) is necessarily a solution of (3.30) which gives uniqueness. □

Proposition 3.3.7. *Under $(\mathcal{A}^{\Psi}_{\text{Lip}})$ and $(\mathcal{A}_{\Psi=\Psi_0})$ assume that $h : \mathbb{R}_+ \to \mathbb{R}$ is locally integrable and that $f_0 : \mathbb{R}_+ \to \mathbb{R}$ is continuous.*

Then, there exists a unique function $\overline{\lambda}$ (which is furthermore continuous on $\mathbb{R}_+$) depending only on Ψ_0, h and f_0 such that the following statements hold

(i) if $(\overline{N}_t)_{t\geq 0}$ is a solution of (3.15) then $\mathbb{E}\left[\overline{N}_+(dt)\right] = \overline{\lambda}(t)dt$,

(ii) there exists a unique (once Π is fixed) solution $(\overline{N}_t)_{t\geq 0}$ to the following system

$$\begin{cases} \overline{N}_t = \displaystyle\int_0^t \int_0^\infty \mathbb{1}_{\left\{x\leq \Psi_0\left(\int_0^{t'} h(t'-z)\overline{\lambda}(z)dz + f_0(t')\right)\right\}} \Pi(dt',dx), \\ \mathbb{E}\left[\overline{N}_t\right] = \displaystyle\int_0^t \overline{\lambda}(t')dt'. \end{cases} \tag{3.35}$$

In particular, $\overline{\lambda}(t) = \Psi_0\left(\int_0^{t-} h(t-z)\overline{\lambda}(z)dz + f_0(t)\right)$ and the solution of (3.35) is the unique (once Π is fixed) solution of (3.15).

The proof follows [43, Theorem 8-*(i)*] and is given in Appendix 3.6.e) for sake of exhaustiveness.

3.3.d) Link via the age process

The link between the limit equation (3.15) and system (3.23) is even deeper than what is stated in Proposition 3.3.6. Indeed, the distribution of the age process (either the predictable or the standard one) associated with a solution of the limit equation is a solution of (3.23) as described in the next statement.

Proposition 3.3.8. *Under the assumptions of Proposition 3.3.6, the unique solution u to the system (3.23) with initial condition that $u(0,\cdot) = u^{\text{in}}$ is such that $u(t,\cdot)$ is the density of the age $\overline{S}_{t-}$ (or $\overline{S}_t$ since they are equal a.s.) associated with the solution of the limit equation (3.15) given in Proposition 3.3.6.*

This result is in fact given in the proof of Proposition 3.3.6 below Equation (3.34).

3.4 Mean-field dynamics

The convergence to the limit dynamics is proved by using a path-wise coupling (like in [43, 101]) between the processes given by Representation 3.2.3 on the one hand and by the limit equation on the other hand. Then, this coupling is studied in two different cases: when the intensity is bounded, i.e. Ψ satisfies $(\mathcal{A}^{\Psi}_{\infty})$, or when the intensity does not depend on the predictable age process, i.e. Ψ satisfies $(\mathcal{A}_{\Psi=\Psi_0})$.

The precise statement regarding the convergence of the n-particle system towards point processes of the McKean-Vlasov type whose intensity depends on time and on the age is given in Corollary 3.4.5.

3.4.a) Coupling

Once the limit equation is well-posed, following the ideas of Sznitman in [156], it is easy to construct a suitable coupling between ADRHPs and i.i.d. solutions of the limit equation (3.15). More precisely, consider

- a sequence $(N^i_-)_{i\geq 1}$ of i.i.d. point processes distributed according to ζ_{N_-};
- an infinite matrix $(H_{ij})_{i,j\geq 1}$ (independent of $(N^i_-)_{i\geq 1}$) with entries distributed according to μ_H such that

$$\begin{cases} \text{for any fixed } i \geq 1, \text{ the variables } H_{i1}, \dots, H_{in}, \dots \text{ are independent,} \\ \text{the sequences } H_{i1}, \dots, H_{in}, \dots \text{ are exchangeable (with respect to } i), \\ \text{the matrix } (H_{ij})_{i,j\geq 1} \text{ is independent from the initial conditions } (N^i_-)_{i=1,..,n}; \end{cases} \tag{3.36}$$

- an infinite matrix $(F_{ij})_{i,j\geq 1}$ with entries distributed according to ν_F such that

$$\begin{cases} \text{for any fixed } i \geq 1, \text{ the variables } F_{i1}, \dots, F_{in}, \dots \text{ are independent,} \\ \text{the sequences } F_{i1}, \dots, F_{in}, \dots \text{ are exchangeable (with respect to } i), \end{cases} \tag{3.37}$$

- a sequence $(\Pi^i(dt', dx))_{i\geq 1}$ of i.i.d. $\mathbb{F}$-Poisson measures with intensity 1 on $\mathbb{R}^2_+$.

Notice that (3.36) (resp. (3.37)) is the equivalent of (3.9) (resp. (3.10)) for infinite matrices. Under $(\mathcal{A}^{\mu_H}_{\infty,2})$, $(\mathcal{A}^{\nu_F}_2)$ and $(\mathcal{A}^{\Psi}_{\text{Lip}})$, the assumptions of Proposition 3.2.4 are satisfied. Furthermore, if we assume either:

- H_1: "$(\mathcal{A}_\infty^\Psi)$ and $(\mathcal{A}_{u^{\text{in}}}^{\zeta_{N_-}})$ are satisfied",
- H_2: "$(\mathcal{A}_{\Psi=\Psi_0})$ is satisfied",

then m_{μ_H} and m_{ν_F} (defined in $(\mathcal{A}_\infty^{\mu_H})$ and $(\mathcal{A}_1^{\nu_F})$) satisfy the assumptions of either Proposition 3.3.6 (under H_1) or 3.3.7 (under H_2) so one can build simultaneously:
- a sequence $((N^{n,i})_{i=1,\dots,n})_{n\geq 1}$ such that for all n, $(N^{n,i})_{i=1,\dots,n}$ is an ADRHP with parameters $(n, \mu_H, \nu_F, \Psi, \zeta_{N_-})$ according to Representation 3.2.3 that is

$$N_t^{n,i} = \int_0^t \int_0^\infty \mathbb{1}\left\{ x \leq \Psi\left(S_{t'-}^{n,i}, \frac{1}{n}\sum_{j=1}^n \left[\int_0^{t'-} H_{ij}(t'-z) N_+^{n,j}(dz) + F_{ij}(t') \right] \right) \right\} \Pi^i(dt', dx) \tag{3.38}$$

with predictable age processes $(S_{t-}^{n,i})_{t\geq 0}^{i=1,\dots,n}$ and past given by N_-^i,
- and a sequence $(\overline{N}_t^i)_{t\geq 0}^{i\geq 1}$ of iindependent and identically distributed solutions of the limit equation with parameters $(m_{\mu_H}, m_{\nu_F}, \Psi, \zeta_{N_-})$ that is

$$\overline{N}_t^i = \int_0^t \int_0^\infty \mathbb{1}\left\{ x \leq \Psi\left(\overline{S}_{t'-}^i, \int_0^{t'} m_{\mu_H}(t'-z)\overline{\lambda}(z)dz + m_{\nu_F}(t') \right) \right\} \Pi^i(dt', dx), \tag{3.39}$$

where $\overline{\lambda}$ is defined either in Proposition 3.3.6 (under H_1) or 3.3.7 (under H_2) and $(\overline{S}_{t-}^i)_{t\geq 0}$ is the predictable age process associated with $\overline{N}^i = N_-^i \cup \overline{N}_+^i$.

Notice that this coupling is based on the sharing of a common past $(N_-^i)_{i\geq 1}$ and a common underlying randomness, that are the $\mathbb{F}$-Poisson measures $(\Pi^i(dt', dx))_{i\geq 1}$. Note that the sequence of ADRHPs is indexed by the size of the network n whereas the solutions of the limit equation that represent the behaviour under the mean field approximation are not.

The following result states the control of the mean-field approximation.

Theorem 3.4.1. *Under $(\mathcal{A}_{\infty,2}^{\mu_H})$, $(\mathcal{A}_2^{\nu_F})$ and $(\mathcal{A}_{\text{Lip}}^\Psi)$, assume either H_1 or H_2. Then, the sequence of ADRHP $(N^{n,i})_{i=1,..,n}$ (with $\mathbb{F}$-intensities on $\mathbb{R}_+$ denoted by $(\lambda^{n,i})_{i=1,..,n}$) defined by (3.38) and the i.i.d. copies $(\overline{N}_t^i)_{t\geq 0}^{i\geq 1}$ of the solution of the limit equation (with $\mathbb{F}$-intensities on $\mathbb{R}_+$ denoted by $(\overline{\lambda}^i)_{i=1,..,n}$) defined by (3.39) are such that for all $i = 1, \dots, n$ and $\theta > 0$,*

$$\mathbb{E}\left[\sup_{t\in[0,\theta]} |N_t^{n,i} - \overline{N}_t^i| \right] \leq \int_0^\theta \mathbb{E}\left[|\lambda_t^{n,i} - \overline{\lambda}_t^i| \right] dt \leq C(\theta, \Psi, \mu_H, \nu_F)\, n^{-1/2}, \tag{3.40}$$

where the constant $C(\theta, \Psi, \mu_H, \nu_F)$ does not depend on n.

Proof. First, let us note that in each cases, the coupling is well-defined thanks to either Proposition 3.3.6 or 3.3.7. Let us fix some $n \geq 1$. Let us denote $\Delta_n^i(t) = \int_0^t |N_+^{n,i}(dt') - \overline{N}_+^i(dt')|$ and $\delta_n^i(t) = \mathbb{E}[\Delta_n^i(t)]$ its expectation. Denoting $A \triangle B$ the symmetric difference of the sets A and B and $\text{Card}\,(A)$ the cardinal of the set A we have

$$\Delta_n^i(t) = \int_0^t |N_+^{n,i}(dt') - \overline{N}_+^i(dt')| = \text{Card}\left((N^{n,i} \triangle \overline{N}^i) \cap [0,t] \right), \tag{3.41}$$

that is the number of points that are not common to $N^{n,i}$ and $\overline{N}^i$ between 0 and t. Then, it is clear that, for all $i = 1, \dots, n$ and $\theta > 0$, $\sup_{t\in[0,\theta]} |N_t^i - \overline{N}_t^i| \leq \Delta_n^i(\theta)$ and so

$$\mathbb{E}\left[\sup_{t\in[0,\theta]} |N_t^{n,i} - \overline{N}_t^i|\right] \leq \mathbb{E}\Big[\Delta_n^i(\theta)\Big] = \delta_n^i(\theta). \tag{3.42}$$

On the one hand, the $\overline{N}^i$'s are i.i.d. hence exchangeable. On the other hand, thanks to the form of the intensity and the assumptions on the matrices $(H_{ij})_{i,j\geq 1}$ and $(F_{ij})_{i,j\geq 1}$ - (3.36) and (3.37) - the family $(N^{n,i})_{i=1,\dots,n}$ is exchangeable too. Hence δ_n^i does not depend on i and is simply denoted by δ_n in the sequel. Let us focus on the case $i = 1$. First, let us remind that $\overline{\lambda}_t^1$ is the intensity of $\overline{N}^1$, so

$$\Delta_n^1(\theta) = \int_0^\theta \int_0^\infty \left| \mathbb{1}_{\left\{x\leq \lambda_t^{n,1}\right\}} - \mathbb{1}_{\left\{x\leq \overline{\lambda}_t^1\right\}} \right| \Pi^1(dt, dx).$$

Taking expectation we find

$$\begin{aligned}
\delta_n(\theta) &= \mathbb{E}\left[\int_0^\theta \mathbb{E}\left[\int_0^\infty \left| \mathbb{1}_{\left\{x\leq \lambda_t^{n,1}\right\}} - \mathbb{1}_{\left\{x\leq \overline{\lambda}_t^1\right\}} \right| \Pi^1(dt, dx) \middle| \mathcal{F}_{t-}\right]\right] \\
&= \mathbb{E}\left[\int_0^\theta \int_0^\infty \left| \mathbb{1}_{\left\{x\leq \lambda_t^{n,1}\right\}} - \mathbb{1}_{\left\{x\leq \overline{\lambda}_t^1\right\}} \right| dx dt\right] \\
&= \mathbb{E}\left[\int_0^\theta |\lambda_t^{n,1} - \overline{\lambda}_t^1| dt\right] = \int_0^\theta \mathbb{E}\left[\left|\lambda_t^{n,1} - \overline{\lambda}_t^1\right|\right] dt,
\end{aligned} \tag{3.43}$$

where the last equality comes from Fubini's Theorem. It remains to show that the rate of convergence is $n^{-1/2}$.

Computations given in Section 3.6.f) show that in each cases there exists some locally bounded function g depending on Ψ, μ_H and ν_F such that δ_n satisfies,

$$\begin{cases}
\delta_n(\theta) \leq n^{-1/2} g(\theta) + \displaystyle\int_0^\theta \left[||\Psi||_\infty + \mathrm{Lip}(\Psi) M_{\mu_H}(\theta - z)\right] \delta_n(z) dz & \text{(under } H_1\text{)}, \\
\delta_n(\theta) \leq n^{-1/2} g(\theta) + \displaystyle\int_0^\theta \mathrm{Lip}(\Psi) M_{\mu_H}(\theta - z) \delta_n(z) dz & \text{(under } H_2\text{)}.
\end{cases} \tag{3.44}$$

Remark that the only dependence with respect to n lies in δ_n. Since g is locally bounded and M_{μ_H} is locally integrable, using Lemma 3.7.4-*(i)*, we end up with

$$\delta_n(\theta) \leq C(\theta, \Psi, \mu_H, \nu_F)\, n^{-1/2}$$

where C does not depend on n. □

There are mainly two reasons for the dichotomy of the assumptions H_1 and H_2. Firstly, up to our knowledge, existence/uniqueness results on the macroscopic system (3.23) are only valid if the function Ψ is bounded. Secondly, as it appears in (3.44), when the intensity of the n-particle system depends on the age, the control of $\delta_n(\theta)$ involves the L^∞ norm of the function Ψ. This boundedness condition is used in order to control the coupling as soon as the ages of the n-particle system on the one hand and the i.i.d. copies of the solution of the limit equation on the other hand are different. Notice that even under H_2, this coupling result extends [43, Theorem 8-*(ii)*] since there are two novelties in the present

article: random interaction functions H_{ij} as well as dependences with respect to the past F_{ij}.

Under more restrictive assumptions (corresponding informally to uniform controls instead of local ones), the rate (with respect to θ) of $C(\theta, \Psi, \mu_H, \nu_F)$ given in Theorem 3.4.1 is linear in comparison with a rate which is at least exponential in general. The main assumption corresponds to the stability criterion of Hawkes processes [17].

Proposition 3.4.2. *Under $(\mathcal{A}^{\mu_H}_{\infty,2})$, $(\mathcal{A}^{\nu_F}_{2})$ and $(\mathcal{A}^{\Psi}_{\mathrm{Lip}})$, assume either H_1 or H_2. Furthermore, assume that both:*

- *the functions M_{μ_H} and Ψ are such that $\alpha := \mathrm{Lip}(\Psi)||M_{\mu_H}||_1 < 1$ and $||M_{\mu_H}||_2 < \infty$;*
- *the functions m_{ν_F} and V_{ν_F} are uniformly bounded.*

Then, the constant $C(\theta, \Psi, \mu_H, \nu_F)$ given in Theorem 3.4.1 can be bounded by $\beta(\Psi, \mu_H, \nu_F)\,\theta$ where the constant $\beta(\Psi, \mu_H, \nu_F)$ depends neither on θ nor on n. This bound holds for all $\theta \geq 0$ under H_2 whereas it holds for $\theta < (1-\alpha)/||\Psi||_\infty$ under H_1.

A proof is given in Appendix 3.6.g) where an explicit expression of β can be found in Equation (3.82) or (3.84) depending on the context.

As said in the introduction, Hawkes processes seem to be the microscopic point processes underpinning the (PPS) system introduced in [114]. There is a striking similarity, modulo a change of time, between :

- on the one hand, the mean intensity, denoted by $m(t)$, of a Hawkes process which is a function of $\int_0^t h(t-z)m(z)dz$,
- and on the other hand, the firing rate p in (PPS) which is a function of $\int_0^t d(z)n(t-z,0)dz$.

A first step in this direction has been made in the previous chapter in the framework of a network of i.i.d. Hawkes processes. In that case, there is no direct bridge between Hawkes processes and the (PPS) system as it is shown in the previous chapter. Indeed, when the size of the network goes to infinity, one recovers conditional expectation of the intensity with respect to the age (instead of the mean intensity). By comparison, the limit system of the mean-field interacting age-dependent Hawkes processes system considered in the present article involves the mean intensity as it can be seen in Equation (3.15). Furthermore, this mean intensity is, in some cases, given by the macroscopic system.

System (3.38)-(3.39) provides an efficient coupling between the spikes attached with the n-particle system and the spikes associated with the limiting process. In order to go one step further, a natural question is to wonder about a possible coupling between the ages associated with the two dynamics. This question is not addressed in [43] in which the propagation of chaos is discussed at the level of the counting processes only. In comparison, we are here willing to investigate this question carefully. The underlying motivation is not of a mathematical essence only: exhibiting a suitable coupling between the ages of the ADRHP and the ages of the point processes of the McKean-Vlasov type whose intensity depends on time and on the age is the right and proper way to make the connection between the microscopic description of the neural dynamics and the macroscopic equation (PPS).

In the sequel, Assumption $(\mathcal{A}^{\zeta_{N^-}}_{\infty})$ is used. It appears that the dependence of the age at time 0 with respect to the initial condition generates additional difficulties for investigating

the mean-field approximation. To limit the complexity of the analysis, it is quite convenient to assume that the age at time 0 is bounded, which is precisely what Assumption $(\mathcal{A}_\infty^{\zeta_{N-}})$ says.

Corollary 3.4.3. *With the notations and assumptions of Theorem 3.4.1, assume that $(\mathcal{A}_\infty^{\zeta_{N-}})$ is satisfied.*

Then, the predictable age processes $(S_{t-}^{n,i})_{t\geq 0}^{i=1,\dots,n}$ associated with the sequence of ADRHP $(N^{n,i})_{i=1,..,n}$ and the predictable age processes $(\overline{S}_{t-}^{i})_{t\geq 0}^{i\geq 1}$ associated with the i.i.d. solutions $(\overline{N}_t^i)_{t\geq 0}^{i\geq 1}$ of the limit equation satisfy for all $i = 1,\dots,n$ and $\theta > 0$,

$$\mathbb{E}\left[\sup_{t\in[0,\theta]} |S_{t-}^{n,i} - \overline{S}_{t-}^{i}|\right] \leq (M_{T_0} + \theta)\, C(\theta, \Psi, \mu_H, \nu_F)\, n^{-1/2}, \tag{3.45}$$

where $C(\theta, \Psi, \mu_H, \nu_F)$ is given in Theorem 3.4.1 and M_{T_0} is defined in $(\mathcal{A}_\infty^{\zeta_{N-}})$.

Remark 3.4.4. *Reminding the strong connection between the predictable age process and the standard one, stated below Equation (3.8), it is clear that the bound (3.45) is also valid when replacing the predictable age processes by their standard counterparts.*

Proof. Let us note that, for all $n \geq 1$ and $i = 1,\dots,n$, $N^{n,i}$ and $\overline{N}^i$ coincide on the non-positive part, i.e. $N_-^{n,i} = \overline{N}_-^i$. Therefore, $S_0^{n,i} = \overline{S}_0^i$ and $\sup_{t\in[0,\theta]} |S_{t-}^{n,i} - \overline{S}_{t-}^i|$ is a.s. upper bounded by $M_{T_0} + \theta$ when the trajectories $(S_{t-}^{n,i})_{t\in[0,\theta]}$ and $(\overline{S}_{t-}^i)_{t\in[0,\theta]}$ are different and is equal to 0 otherwise. Therefore, we have the following bound

$$\mathbb{E}\left[\sup_{t\in[0,\theta]} |S_{t-}^{n,i} - \overline{S}_{t-}^{i}|\right] \leq (M_{T_0} + \theta)\mathbb{P}\left(\left(S_{t-}^{n,i}\right)_{t\in[0,\theta]} \neq \left(\overline{S}_{t-}^{i}\right)_{t\in[0,\theta]}\right).$$

Yet if the trajectories are different, there is at least one point between 0 and θ which is not common to both $N_+^{n,i}$ and $\overline{N}_+^i$, that is $\sup_{t\in[0,\theta]} |N_t^{n,i} - \overline{N}_t^i| \neq 0$, hence

$$\mathbb{P}\left(\left(S_{t-}^{n,i}\right)_{t\in[0,\theta]} \neq \left(\overline{S}_{t-}^{i}\right)_{t\in[0,\theta]}\right) = \mathbb{P}\left(\sup_{t\in[0,\theta]} |N_t^{n,i} - \overline{N}_t^i| \neq 0\right). \tag{3.46}$$

Moreover, since counting processes are piecewise constant with jumps of height 1 a.s., it is clear that

$$\mathbb{P}\left(\sup_{t\in[0,\theta]} |N_t^{n,1} - \overline{N}_t^1| \neq 0\right) = \mathbb{P}\left(\sup_{t\in[0,\theta]} |N_t^{n,1} - \overline{N}_t^1| \geq 1\right) \leq \mathbb{E}\left[\sup_{t\in[0,\theta]} |N_t^{n,1} - \overline{N}_t^1|\right], \tag{3.47}$$

where we used Markov's inequality. Finally, inequality (3.45) clearly follows from Theorem 3.4.1. □

3.4.b) Mean-field approximations

Inspired by the seminal work of Sznitman [156], we now obtain, from the results of the previous section, the convergence of the n-particle system towards the limit equation:

- the empirical distribution of the point processes associated with the n-particle system converges to the distribution of the point process solution of the limit equation,

- the empirical distribution of the age processes associated with the n-particle system converges to the distribution of the age process associated with the solution of the limit equation,

This result, together with the ones from the previous paragraphs, are typical of what is known as the *propagation of chaos theory of interacting particle system*. In particular, it says that k fixed neurons behave independently and identically when the size of the network goes to infinity. Their spiking dynamics being described by the limit equation (3.15).

Corollary 3.4.5. *Let $\mathcal{L}(X)$ denote the distribution of some random variable X and $\mathcal{D}(\mathbb{R}_+)$ denote the space of càdlàg functions from $\mathbb{R}_+$ to $\mathbb{R}$ endowed with the Skorokhod topology.*

With notations and assumptions of Theorem 3.4.1, we have the following mean-field approximations:

- *the weak convergence in $\mathcal{P}(\mathcal{D}(\mathbb{R}_+))$ of the empirical measure of counting processes,*

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{(N_t^{n,i})_{t\geq 0}} \xrightarrow[n\to\infty]{} \mathcal{L}\big((\overline{N}_t^1)_{t\geq 0}\big); \tag{3.48}$$

- *if furthermore ($\mathcal{A}_\infty^{\zeta_{N-}}$) holds, the weak convergence in $\mathcal{P}(\mathcal{D}(\mathbb{R}_+))$ of the empirical measure of the standard age processes,*

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{(S_t^{n,i})_{t\geq 0}} \xrightarrow[n\to\infty]{} \mathcal{L}\big((\overline{S}_t^1)_{t\geq 0}\big). \tag{3.49}$$

Both convergences also hold in probability since the limits are constant in $\mathcal{P}(\mathcal{D}(\mathbb{R}_+))$.

Finally, under the assumptions of Corollary 3.4.3, if furthermore ($\mathcal{A}_\infty^{\Psi}$) and ($\mathcal{A}_{u^{\mathrm{in}}}^{\zeta_{N-}}$) hold, then the unique solution u of the system (3.23) with initial condition that $u(0,\cdot)=u^{\mathrm{in}}$ is such that $u_t := u(t,\cdot)$ is the density of the age $\overline{S}_{t-}^1$ and for all $\theta>0$,

$$\sup_{t\in[0,\theta]}\mathbb{E}\left[W_1\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{S_{t-}^{n,i}},u_t\right)\right]\leq D(\theta,\Psi,\mu_H,\nu_F,M_{T_0})\,n^{-1/2}, \tag{3.50}$$

where the constant $D(\theta,\Psi,\mu_H,\nu_F,M_{T_0})$ does not depend on n and W_1 denotes the standard 1-Wasserstein distance.

Remark 3.4.6. *Of course, the convergence (3.49) is also valid when replacing the standard age processes by their predictable counterparts. However, let us mention that the predictable age processes belong to $\mathcal{G}(\mathbb{R}_+)$, the space of càglàd functions (continuous to the left with right limits). Hence the convergence of the empirical measure of the predictable age processes holds in $\mathcal{P}(\mathcal{G}(\mathbb{R}_+))$, where we endow $\mathcal{G}(\mathbb{R}_+)$ with an analoguous of the Skorokhod topology*

Proof. The space of càdlàg functions $\mathcal{D}(\mathbb{R}_+)$ endowed with the Skorokhod topology is a Polish space. So, according to [156, Proposition 2.2] or [101, Proposition 4.2], to show the first limit (3.48), it suffices to check that $((N_t^{n,1})_{t\geq 0},(N_t^{n,2})_{t\geq 0})$ converges in distribution, as $n\to+\infty$, to two independent copies of $(\overline{N}_t^1)_{t\geq 0}$. Since the convergence with respect to Lipschitz continuous test functions is sufficient in order to prove the convergence in distribution (Portemanteau Theorem [87]), the first limit clearly follows from both (3.40) and the fact that the uniform convergence topology on compact time intervals is finer than

the Skorokhod topology. The proof of the second limit is similar with the difference that it follows from (3.45) (remind Remark 3.4.4) instead of (3.40).

The link between the solution of (3.23) and the age processes associated with the solution of the limit equation (3.15) is given by Proposition 3.3.8.

The rate of convergence for the 1-Wasserstein distance stated in (3.50) is a consequence of the rate of convergence for i.i.d. real valued random variables. Fix $\theta > 0$ and let t be in $[0, \theta]$. First, using the exchangeability of the particles, it follows from Corollary 3.4.3 that there exists a constant $C(\theta, \Psi, \mu_H, \nu_F, M_{T_0})$ such that

$$\mathbb{E}\left[W_1\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{S^{n,i}_{t-}}, \frac{1}{n}\sum_{i=1}^{n}\delta_{\overline{S}^i_{t-}}\right)\right] \leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\sup_{t\in[0,\theta]}|S^{n,i}_{t-} - \overline{S}^i_{t-}|\right] \leq C(\theta, \Psi, \mu_H, \nu_F, M_{T_0})\, n^{-1/2}.$$

Then, applying [51, Theorem 1] to the i.i.d. random variables $\overline{S}^i_{t-}$, that are bounded by $M_{T_0} + \theta$, we deduce that there exists a constant $\tilde{C}(\theta, M_{T_0})$ such that

$$\mathbb{E}\big[W_1\big(\frac{1}{n}\sum_{i=1}^{n}\delta_{\overline{S}^i_{t-}}, u_t\big)\big] \leq \tilde{C}(M_{T_0}, \theta)n^{-1/2}.$$

Finally, the triangular inequality for the 1-Wasserstein distance gives (3.50). □

The first mean-field approximation (3.48) is a generalization of the one given in [43, Theorem 8-*(iii)*] where mean-field interacting Hawkes processes are approximated by Poisson processes of the McKean-Vlasov type. Here, the limit processes are point processes of the McKean-Vlasov type whose intensity depends on time (like Poisson processes) and on the age.

Moreover, Equation (3.50) extends the result on the rate of convergence for the age processes given in [130, Section 5].

3.5 Conclusion

We present a generalization of mean-field interacting Hawkes processes, namely age dependent random Hawkes processes (ADRHPs), which are well-adapted to neuroscience modelling. From a biological point of view, they encompass some interesting features such as refractory period, synaptic integration or random synaptic weights. These processes are studied in a mean-field situation and we show in Theorem 3.4.1 and Corollary 3.4.5 that, as the number of particles goes to infinity, they can be well approximated by point processes of the McKean-Vlasov type whose intensity depends on time and on the age. These limit point processes are closely related to the age structured PDE system introduced by Pakdaman, Perthame and Salort, namely (PPS), as shown in Proposition 3.3.8.

Hence, using the theory of mean-field approximations, the present article makes a bridge between the microscopic modelling given by Hawkes processes, or more generally age dependent random Hawkes processes, and the macroscopic modelling given by the (PPS) system. This bridge is presented under the main assumption that the intensity of the microscopic point processes is bounded. In this sense, the present article offers an answer to the question left open in the previous chapter. This legitimises the convolution term $X(t)$ in the (PPS) system as well as opens the way to the study of new assumptions on the

spiking rate p appearing in the (PPS) system from a more analytical point of view. Up to our knowledge, this has not been done yet.

The present article gives somehow the law of large numbers for a generalization of Hawkes processes. It could be interesting to investigate how these processes fluctuates around their mean limit or in other words find some kind of functional central limit theorem for Hawkes processes in a mean-field framework.

As noted, random synaptic weights can be considered in this study. However, they are supposed to be, in some sense, independent and identically distributed which can be considered as an unrealistic assumption. Inspired by [48], it could be interesting to see how correlated synaptic weights could be handled in the Hawkes processes framework.

On a different path, it could be interesting to see how locally stationary Hawkes processes, as introduced in [143], behave in a mean-field situation. Indeed, these processes may take into account the dynamics of the synaptic weights occurring in the neural network.

3.6 Proofs

3.6.a) Proof of Proposition 3.2.4

Let us denote $G_i : t \mapsto \sup_{s\geq 0} \Psi(s,0) + \mathrm{Lip}(\Psi) n^{-1} \sum_{j=1}^n |F_{ij}(t)|$. Thanks to $(\mathcal{A}_1^{\nu_F})$ and Fubini's Theorem we have, for all $T > 0$, $\mathbb{E}[\int_0^T |F_{11}(t)|dt] = \int_0^T \mathbb{E}[|F_{11}(t)|]dt < +\infty$. In particular, for all $1 \leq i,j \leq n$, $t \mapsto |F_{ij}(t)|$ is locally integrable almost surely. Hence, there exists a subset Ω of probability 1 such that, on Ω, G_i is locally integrable for all i. Fixing the G_i's, one can apply Lemma 3.7.1 (with $a_i = \mathrm{Lip}(\Psi)$ and $g_i = G_i$) to deduce that the processes $(N_t^i)_{t\geq 0}^{i=1,..,n}$ are dominated by the processes $(\tilde{N}_t^i)_{t\geq 0}^{i=1,..,n}$ (defined by (3.86)) and so are well-defined.

It remains to show that the function $t \mapsto \mathbb{E}[N_t^1]$ is locally bounded. First, let us study the dominating processes. We have

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \tilde{N}_t^i\right] &\leq \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \int_0^t G_i(t')dt'\right] \\
&\quad + \mathrm{Lip}(\Psi)\frac{1}{n}\sum_{i=1}^n \int_0^t \int_0^{t'} M_{\mu_H}(t'-z)\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^n \tilde{N}^j(dz)\right]dt' \\
&\leq \frac{1}{n}\sum_{i=1}^n \int_0^t \mathbb{E}\left[G_i(t')\right]dt' + \mathrm{Lip}(\Psi)\int_0^t M_{\mu_H}(t-t')\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^n \tilde{N}_{t'}^j\right]dt',
\end{aligned}$$

where we used Lemma 3.7.3. Next, $t \mapsto \mathbb{E}[G_i(t)] = \sup_{s\geq 0}\Psi(s,0) + \mathrm{Lip}(\Psi)\mathbb{E}[|F_{11}(t)|]$ is locally bounded and M_{μ_H} is locally integrable so Lemma 3.7.4-*(i)* gives that $t \mapsto \mathbb{E}\left[n^{-1}\sum_{i=1}^n \tilde{N}_t^i\right] = \mathbb{E}\left[\tilde{N}_t^1\right]$ is locally bounded. Finally, the stochastic domination (in particular, $\mathbb{E}[N_t^1] \leq \mathbb{E}[\tilde{N}_t^1]$) gives the result.

3.6.b) Proof of Proposition 3.3.1

First, in order to be consistent with the formalism used in [26] we must rewrite the system (3.16) in a single equation in the following way

$$\frac{\partial u(t,s)}{\partial t} + \frac{\partial u(t,s)}{\partial s} = N_1(t,s,u) + \delta_{s=0} n_2(t,u), \tag{3.51}$$

with initial condition u^{in} where

$$\begin{cases} N_1(t,s,u) := -f(t,s)u(t,s) \\ n_2(t,u) := \int_0^{+\infty} f(t,s')u(t,s')ds'. \end{cases}$$

The use of the Dirac mass localized in age equal to 0 represents the boundary condition that is the second equation of (3.16).

Note that the general result [26, Theorem 2.4.] gives existence and uniqueness of solution in $BC(\mathbb{R}_+, \mathcal{M}(\mathbb{R}))$ and not $BC(\mathbb{R}_+, \mathcal{M}(\mathbb{R}_+))$ even if the initial condition has support contained in $\mathbb{R}_+$. However, as explained in [26, Section 3.3.], it suffices to extend the equation for s in $\mathbb{R}$, to apply [26, Theorem 2.4.] and then to check that the support of the solution is conserved in the sense that: if u^{in} has support on $\mathbb{R}_+$ then the unique solution given by the Theorem has also support contained in $\mathbb{R}_+$ for all time $t \geq 0$.

Hence, consider Equation (3.51) but with s being in $\mathbb{R}$ by mirror symmetry for definiteness (that is $N_1(t,s,u) = N_1(t,-s,u)$ and so $f(t,s) = f(t,-s)$). Let us check that the assumptions of [26, Theorem 2.4.] are satisfied.

-(H1) and (H2) are clearly satisfied.

-(H3). We need to verify that N_1 and n_2 are continuous in t with respect to the usual topology and in u with respect to the topology induced by the bounded Lipschitz norm (denoted by $||.||_{BL}$). On the one hand, using the boundedness of f, we have, with u_t denoting the measure $u(t,\cdot)$,

$$||N_1(t+t',s,u+\tilde{u}) - N_1(t,s,u)||_{BL} \leq \sup_{s\in\mathbb{R}} |f(t+t',s) - f(t,s)|||u_t||_{BL} + ||f||_\infty||\tilde{u}_{t+t'}||_{BL} + ||f||_\infty||u_{t+t'} - u_t||_{BL}.$$

As t' and $\tilde{u}$ converge to 0, the first term converges to 0 since f is (uniformly in s) continuous with respect to t, the second one clearly converges to 0 and the third one converges to 0 since u belongs to $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$.

On the other hand, using once again the boundedness of f, we have,

$$|n_2(t+t',u+\tilde{u}) - n_2(t,u)| \leq \sup_{s\in\mathbb{R}} |f(t+t',s) - f(t,s)|||u_t||_{BL} + ||f||_\infty||u_{t+t'} - u_t||_{BL},$$

which converges to 0 as t' tends to 0 thanks to the continuity property of f.

-(H4). It suffices to show that N_1 and n_2 are Lipschitz continuous with respect to the variable u. On the one hand, we have

$$||N_1(t,s,u) - N_1(t,s,v)||_{BL} \leq ||f||_\infty||u_t - v_t||_{BL}.$$

On the other hand, we have

$$|n_2(t,u) - n_2(t,v)| \leq ||f||_\infty||u_t - v_t||_{BL}.$$

-(H5). Here, it suffices to check that $N_1(t,s,u) + \delta_{s=0}n_2(t,u)$ carries bounded sets in total variation norm to bounded sets in total variation norm. Denoting $||.||_{TV}$ the total variation norm, we have $||N_1(t,s,u) + \delta_{s=0}n_2(t,u)||_{TV} \leq 2||f||_\infty||u||_{TV}$.

Finally, the argument to prove conservation of the support for solutions being the same as the one elaborated in [26, Section 3.3.], it is not reproduced here.

3.6.c) Proof of Proposition 3.3.2

The method of characteristics applied to the first equation of (3.16) suggests to consider:

- for all $z \geq 0$, $u^z : t \mapsto u(t, t+z)$ satisfying

$$\frac{d}{dt}u^z(t) = -f(t, z+t)u^z(t)$$

hence for all $t \geq 0$,

$$u^z(t) = u^z(0)\exp\left(-\int_0^t f(t', z+t')dt'\right)$$

and so, using that $u^z(0) = u^{\text{in}}(z)$ and letting $s = z+t$, one has (3.19).

- for all $z \geq 0$, $u^z : s \mapsto u(s+z,s)$ satisfying

$$\frac{d}{ds}u^z(s) = -f(s+z,s)u^z(s)$$

hence for all $s \geq 0$,

$$u^z(s) = u^z(0)\exp\left(-\int_0^s f(s'+z,s')ds'\right)$$

and so, using that $u^z(0) = u(z,0)$ and letting $t = z+s$, one has

$$u(t,s) = u(t-s,0)\exp\left(-\int_0^s f(t-s+s',s')ds'\right), \quad \text{for } t \geq s,$$

which is not exactly (3.20). Here, $u(t-s,0)$ is just a parameter which is not constrained by the first equation of (3.16). However, it is characterized by the second equation of (3.16) as explained below.

For any $T > 0$, consider the application

$$\begin{array}{rccc} G_T: & L^\infty([0,T]) & \longrightarrow & L^\infty([0,T]) \\ & u_0 & \longmapsto & (t \mapsto G(u_0)(t)), \end{array}$$

where

$$G(u_0)(t) := \int_0^t f(t,s)u_0(t-s)\exp\left(-\int_0^s f(t-s+s',s')ds'\right)ds + \int_t^{+\infty} f(t,s)u^{\text{in}}(s-t)\exp\left(-\int_0^t f(t',s-t+t')dt'\right)ds.$$

Note that the characteristics and the second equation of (3.16) suggest that, denoting u the solution of (3.16), $u(\cdot,0)$ is bounded and that its trace on $[0,T]$ is a fixed point of G_T. Using the boundedness of f and the fact that the argument in the exponential is non-positive, we have for any u_0, v_0 in $L^\infty([0,T])$,

$$|G_T(u_0)(t) - G_T(v_0)(t)| \leq ||f||_\infty \int_0^t |u_0(t-s) - v_0(t-s)|ds \leq T||f||_\infty ||u_0 - v_0||_{L^\infty([0,T])}.$$

Now, fix $T > 0$ such that $T||f||_\infty \leq 1/2$ so that G_T is a contraction and admits a unique fixed point. Note that G_T maps non-negative functions to non-negative functions, so that the fixed point of G_T is a non-negative function. Iterating this fixed point gives the existence and uniqueness of a locally bounded function u_0 (which is non-negative) such that for all $t \geq 0$, $u_0(t) = G(u_0)(t)$ (see the end of the proof of Lemma 3.3.4 for the same kind of argument in a more detailed form). Until the end of the proof, u_0 will denote this fixed point.

It only remains to check that u is a solution of (3.16) in the weak sense. Let φ be in $\mathcal{C}^\infty_{c,b}(\mathbb{R}^2_+)$, let us compute $\int_{\mathbb{R}^2_+}\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\varphi(t,s)\,u(t,s)\,dtds$. Sticking with the decomposition of the representation given by (3.19)-(3.20) and using integration by parts to go backward in the heuristic given by the method of characteristics, one has

$$\int_{s\geq t}\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\varphi(t,s)\,u(t,s)\,dtds = \int_{s\geq t}\varphi(t,s)f(t,s)u(t,s)dtds - \int_{s=0}^{+\infty}\varphi(0,s)u^{\text{in}}(s)ds, \tag{3.52}$$

and

$$\int_{t\geq s}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial s}\right)\varphi\left(t,s\right)u\left(t,s\right)dtds=\int_{t\geq s}\varphi(t,s)f(t,s)u(t,s)dtds-\int_{t=0}^{+\infty}\varphi(t,0)u_0(t)dt. \tag{3.53}$$

Remarking that the definition of u_0 as the fixed point of G gives that u defined by (3.19)-(3.20) satisfies the second equation of (3.16) in a strong sense, one deduces that (3.18) is satisfied by gathering (3.52) and (3.53).

3.6.d) Proof of (3.29)

Let Y_1 and Y_2 be two fixed locally bounded functions and denote $u_1 := u_{Y_1}$ and $u_2 := u_{Y_2}$ the two solutions associated with Y_1 and Y_2 (with same initial condition u^{in}) and $\tilde{u}(t) := u_1(t,0)-u_2(t,0)$ for all $t\geq 0$. Using the characteristics, i.e. (3.19) and (3.20), one deduces from the second equation in (3.27) that for $i=1$ or 2, $u_i(t,0)=v_i^{0,t}+v_i^{t,\infty}$ where

$$\begin{cases} v_i^{0,t} := \displaystyle\int_0^t \Psi(s,Y_i(t)+f_0(t))u_i(t-s,0)\exp\left(-\int_0^s \Psi(s',Y_i(t-s+s')+f_0(t-s+s'))ds'\right)ds \\ v_i^{t,\infty} := \displaystyle\int_t^{+\infty} \Psi(s,Y_i(t)+f_0(t))u^{\text{in}}(s-t)\exp\left(-\int_0^t \Psi(s-t+t',Y_i(t')+f_0(t'))dt'\right)ds. \end{cases}$$

and so $\tilde{u}(t)=A(t)+B(t)$ where $A(t):=v_1^{0,t}-v_2^{0,t}$ and $B(t):=v_1^{t,\infty}-v_2^{t,\infty}$. Before studying the functions A and B, let us remark that in order to prove (3.29), it suffices to prove that there exists a non-decreasing function C (which depends neither on the initial condition nor on f_0) such that for all $T\geq 0$,

$$||\tilde{u}||_{L^\infty([0,T])}\leq C(T)||Y_1-Y_2||_{L^\infty([0,T])}. \tag{3.54}$$

Indeed, using the definition of F_T given by (3.28), one then deduces that for all $T\geq 0$,

$$||F_T(Y_1)-F_T(Y_2)||_{L^\infty([0,T])}\leq C(T)||Y_1-Y_2||_{L^\infty([0,T])}\int_0^T|h(z)|dz$$

and since h is locally integrable and C is non-decreasing, there exists $T>0$ small enough such that $C(T)\int_0^T|h(z)|dz\leq 1/2$ which ends the proof. To prove (3.54), let t be a positive real number.

Study of A. We have $A=A_1+A_2+A_3$ with

$$\begin{cases} A_1(t) := \displaystyle\int_0^t [\Psi(s,Y_1(t)+f_0(t))-\Psi(s,Y_2(t)+f_0(t))]u_1(t-s,0)e^{-\int_0^s\Psi(s',Y_1+f_0)ds'}ds \\ A_2(t) := \displaystyle\int_0^t \Psi(s,Y_2(t)+f_0(t))[u_1(t-s,0)-u_2(t-s,0)]e^{-\int_0^s\Psi(s',Y_1+f_0)ds'}ds \\ A_3(t) := \displaystyle\int_0^t \Psi(s,Y_2(t)+f_0(t))u_2(t-s,0)[e^{-\int_0^s\Psi(s',Y_1+f_0)ds'}-e^{-\int_0^s\Psi(s',Y_2+f_0)ds'}]ds. \end{cases}$$

where the arguments "$t-s+s'$" in the exponentials are not written for simplicity.

- Study of A_1. Using the Lipschitz continuity of Ψ, the fact that the argument in the exponential is non-positive and the a priori bound on u_1, we have

$$\begin{aligned}|A_1(t)|\leq \text{Lip}(\Psi)\int_0^t u_1(t-s,0)|Y_1(t)-Y_2(t)|ds &\leq \text{Lip}(\Psi)||u_1||_{L^\infty(\mathbb{R}_+^2)}\,t\,|Y_1(t)-Y_2(t)| \\ &\leq \text{Lip}(\Psi)\max(M,||\Psi||_\infty)\,t\,||Y_1-Y_2||_{L^\infty([0,t])}. \end{aligned}\tag{3.55}$$

- Study of A_2. Using the boundedness of Ψ and the fact that the argument in the exponential is non-positive, we have

$$|A_2(t)| \leq ||\Psi||_\infty \int_0^t |u_1(z,0) - u_2(z,0)|dz = ||\Psi||_\infty \int_0^t |\tilde{u}(z,0)|dz. \tag{3.56}$$

- Study of A_3. The arguments of the exponentials are non-positive and the exponential function is Lipschitz with constant 1 on $\mathbb{R}_-$ so, using the a priori bound on u_2, we have

$$\begin{aligned} |A_3(t)| &\leq ||\Psi||_\infty ||u_2||_{L^\infty(\mathbb{R}_+^2)} \int_0^t \left| \int_0^s \Psi(s', Y_1 + f_0)ds' - \int_0^s \Psi(s', Y_2 + f_0)ds' \right| ds \\ &\leq ||\Psi||_\infty \max(M, ||\Psi||_\infty) \mathrm{Lip}(\Psi) \int_0^t \int_0^s |Y_1(t-s+s') - Y_2(t-s+s')| \, ds'ds \\ &\leq ||\Psi||_\infty \max(M, ||\Psi||_\infty) \mathrm{Lip}(\Psi) \, t^2 \, ||Y_1 - Y_2||_{L^\infty([0,t])}. \end{aligned} \tag{3.57}$$

where the arguments "$t - s + s'$" are not written in the first equation for simplicity.

Study of B. We have $B = B_1 + B_2$ with

$$\begin{cases} B_1(t) := \displaystyle\int_t^{+\infty} [\Psi(s, Y_1(t) + f_0(t)) - \Psi(s, Y_2(t) + f_0(t))] \\ \qquad\qquad u^{\mathrm{in}}(s-t) \exp\left(- \displaystyle\int_0^t \Psi(s - t + t', Y_1(t') + f_0(t'))dt' \right) ds \\ B_2(t) := \displaystyle\int_t^{+\infty} \Psi(s, Y_2(t) + f_0(t)) u^{\mathrm{in}}(s-t) \\ \qquad\qquad [e^{-\int_0^t \Psi(s-t+t', Y_1(t') + f_0(t'))dt'} - e^{-\int_0^t \Psi(s-t+t', Y_2(t') + f_0(t'))dt'}]ds. \end{cases}$$

- Study of B_1. Using the Lipschitz continuity of Ψ and the fact that the argument in the exponential is non-positive, we have

$$|B_1(t)| \leq \mathrm{Lip}(\Psi) \int_t^{+\infty} |Y_1(t) - Y_2(t)| u^{\mathrm{in}}(s-t)ds \leq \mathrm{Lip}(\Psi) ||Y_1 - Y_2||_{L^\infty([0,t])}, \tag{3.58}$$

where we used that $\int_0^{+\infty} u^{\mathrm{in}}(s)ds = 1$.

- Study of B_2. As for A_3, we have

$$\begin{aligned} |B_2(t)| &\leq ||\Psi||_\infty \int_t^{+\infty} u^{\mathrm{in}}(s-t) \\ &\qquad \left| \int_0^t \Psi(s - t + t', Y_1(t') + f_0(t')) - \Psi(s - t + t', Y_2(t') + f_0(t'))dt' \right| ds \\ &\leq ||\Psi||_\infty \mathrm{Lip}(\Psi) \int_t^{+\infty} u^{\mathrm{in}}(s-t) \int_0^t |Y_1(t') - Y_2(t')| \, dt'ds \\ &\leq ||\Psi||_\infty \mathrm{Lip}(\Psi) \, t \, ||Y_1 - Y_2||_{L^\infty([0,t])}, \end{aligned} \tag{3.59}$$

where we used once again that $\int_0^{+\infty} u^{\mathrm{in}}(s)ds = 1$.

Gathering (3.55), (3.56), (3.57), (3.58) and (3.59), we get that there exists a non-decreasing function g such that for all $t > 0$,

$$|\tilde{u}(t)| \leq g(t) ||Y_1 - Y_2||_{L^\infty([0,t])} + ||\Psi||_\infty \int_0^t |\tilde{u}(z)|dz.$$

Then, Lemma 3.7.4-*(i)* gives that for all $T \geq 0$,

$$||\tilde{u}||_{L^\infty([0,T])} \leq C_T \sup_{t\in[0,T]} \left\{ g(t)||Y_1 - Y_2||_{L^\infty([0,t])} \right\},$$

with C_T being a non-decreasing function of T. Then, (3.54) follows since g is non-decreasing.

3.6.e) Proof of Proposition 3.3.7

-(i) Suppose that $(\overline{N}_t)_{t\geq 0}$ is a solution of (3.15). Then, the mean cumulative intensity $\overline{\Lambda}(t) = \mathbb{E}[\overline{N}_t]$ is a non-decreasing locally bounded function satisfying

$$\overline{\Lambda}(t) = \int_0^t \Psi_0 \left(\int_0^{t'} h(t'-z) d\overline{\Lambda}_z + f_0(t') \right) dt' \text{ for every } t \geq 0. \tag{3.60}$$

By Lemma 3.7.5, we know that (3.60) admits a unique solution which is furthermore of class $\mathcal{C}^1$ and we denote $\overline{\lambda}$ its derivative. Thus, we have $\mathbb{E}\left[\overline{N}_+(dt)\right] = \overline{\Lambda}'(t)dt = \overline{\lambda}(t)dt$.

-(ii) The first equation of (3.35) is a classical thinning equation so its solution $(\overline{N}_t)_{t\geq 0}$ is a measurable function of Π hence it is unique (once Π is fixed).

To conclude this step, it suffices to check that $(\overline{N}_t)_{t\geq 0}$ satisfies the second equation of (3.35) where we remind that $\overline{\lambda}$ is the derivative of $\overline{\Lambda}$ which is the unique solution of (3.60). But $\mathbb{E}[\overline{N}_t] = \int_0^t \Psi_0(\int_0^{t'} h(t'-z)\overline{\lambda}(z)dz + f_0(t'))dt'$, which is equal to $\overline{\Lambda}(t) = \int_0^t \overline{\lambda}(t')dt'$ since $\overline{\Lambda}$ is a solution of (3.60).

Finally, the two remaining points are rather simple. First, taking the derivative of (3.60) gives that $\overline{\lambda}(t) = \Psi_0\left(\int_0^{t-} h(t-z)\overline{\lambda}(z)dz + f_0(t)\right)$. Secondly, the solution of (3.35) is clearly a solution of (3.15) and (i) tells that a solution of (3.15) is necessarily a solution of (3.35) which gives uniqueness.

3.6.f) Proof of (3.44)

For simplicity of notations in (3.38) and (3.39), let us denote for all $t \geq 0$,

$$\begin{cases} \gamma_t^{n,i} := \dfrac{1}{n}\displaystyle\sum_{j=1}^n \left(\int_0^{t-} H_{ij}(t-z) N_+^{n,j}(dz) + F_{ij}(t) \right), \\ \overline{\gamma}(t) := \displaystyle\int_0^t m_{\mu_H}(t-z)\overline{\lambda}(z)dz + m_{\nu_F}(t), \end{cases} \tag{3.61}$$

where $\overline{\lambda}$ is defined either in Proposition 3.3.6 (under H_1) or 3.3.7 (under H_2). We have $\lambda_t^{n,i} = \Psi(S_{t-}^{n,i}, \gamma_t^{n,i})$ and $\overline{\lambda}_t^i = \Psi(\overline{S}_{t-}^i, \overline{\gamma}(t))$. Notice that $\overline{\gamma}$ is a deterministic function (which does not depend on i) whereas $\gamma^{n,i}$ is random.

First point of (3.44). Assume that H_1 is satisfied. Then one can use the decomposition,

$$1 = \mathbb{1}_{\{S_{t-}^{n,1} = \overline{S}_{t-}^1\}} + \mathbb{1}_{\{S_{t-}^{n,1} \neq \overline{S}_{t-}^1\}} \tag{3.62}$$

and the fact that $|\lambda_t^{n,1} - \overline{\lambda}_t^1| \leq ||\Psi||_\infty$ to deduce from (3.43) that

$$\delta_n(\theta) \leq \int_0^\theta \mathbb{E}\left[\left|\lambda_t^{n,1} - \overline{\lambda}_t^1\right| \mathbb{1}_{\{S_{t-}^{n,1} = \overline{S}_{t-}^1\}}\right] dt + ||\Psi||_\infty \int_0^\theta \mathbb{P}\left(S_{t-}^{n,1} \neq \overline{S}_{t-}^1\right) dt. \tag{3.63}$$

Let us denote

$$\begin{cases} A^n(\theta) := \displaystyle\int_0^\theta \mathbb{E}\left[\left|\lambda_t^{n,1} - \overline{\lambda}_t^1\right| \mathbb{1}_{\{S_{t-}^{n,1} = \overline{S}_{t-}^1\}}\right] dt, \\ D^n(\theta) := \displaystyle\int_0^\theta \mathbb{P}\left(S_{t-}^{n,1} \neq \overline{S}_{t-}^1\right) dt. \end{cases}$$

Study of $A^n(\theta)$. Using the Lipschitz continuity of Ψ, it is clear that for all t in $[0,\theta]$, if $S_{t-}^{n,1} = \overline{S}_{t-}^1$, then $|\lambda_t^{n,1} - \overline{\lambda}_t^1| \leq \mathrm{Lip}(\Psi)|\gamma_t^{n,1} - \overline{\gamma}(t)|$. So, one deduces that

$$A^n(\theta) \leq \mathrm{Lip}(\Psi)(B_1^n(\theta) + B_2^n(\theta) + B_3^n(\theta) + C^n(\theta)), \tag{3.64}$$

where

$$\begin{cases} B_1^n(\theta) := \displaystyle\int_0^\theta \mathbb{E}\Big[\Big| \int_0^t \frac{1}{n}\sum_{j=1}^n H_{1j}(t-z)[N_+^{n,j}(dz) - \overline{N}_+^j(dz)]\Big|\Big] dt, \\ B_2^n(\theta) := \displaystyle\int_0^\theta \mathbb{E}\Big[\Big| \int_0^t \frac{1}{n}\sum_{j=1}^n H_{1j}(t-z)[\overline{N}_+^j(dz) - \overline{\lambda}_z^j dz]\Big|\Big] dt, \\ B_3^n(\theta) := \displaystyle\int_0^\theta \mathbb{E}\Big[\Big| \int_0^t \frac{1}{n}\sum_{j=1}^n \Big(H_{1j}(t-z)\overline{\lambda}_z^j - m_{\mu_H}(t-z)\overline{\lambda}(z)\Big) dz\Big|\Big] dt, \\ C^n(\theta) := \displaystyle\int_0^\theta \mathbb{E}\Big[\Big|\frac{1}{n}\sum_{j=1}^n F_{1j}(t) - m_{\nu_F}(t)\Big|\Big] dt. \end{cases} \tag{3.65}$$

- Study of B_1^n. Firstly, using $(\mathcal{A}_\infty^{\mu_H})$ and then Lemma 3.7.3, we have

$$\begin{aligned} B_1^n(\theta) &\leq \int_0^\theta \mathbb{E}\Big[\int_0^t \frac{1}{n}\sum_{j=1}^n \Big|H_{1j}(t-z)\Big|\Big|N_+^{n,j}(dz) - \overline{N}_+^j(dz)\Big|\Big] dt \\ &\leq \int_0^\theta \mathbb{E}\Big[\int_0^t \frac{1}{n}\sum_{j=1}^n M_{\mu_H}(t-z)\Big|N_+^{n,j}(dz) - \overline{N}_+^j(dz)\Big|\Big] dt \\ &= \int_0^\theta \frac{1}{n}\sum_{j=1}^n \mathbb{E}\Big[\int_0^t M_{\mu_H}(t-z) d\Delta_n^j(z)\Big] dt = \int_0^\theta M_{\mu_H}(\theta - z)\delta_n(z) dz, \end{aligned} \tag{3.66}$$

where the Δ_n^j's are given in (3.41).

- Study of B_2^n. Secondly, using Cauchy-Schwartz inequality, we have

$$\begin{aligned} B_2^n(\theta) &\leq \frac{1}{n}\int_0^\theta \mathbb{E}\Big[\Big|\sum_{j=1}^n \int_0^t H_{1j}(t-z)[\overline{N}_+^j(dz) - \overline{\lambda}_z^j dz]\Big|^2\Big]^{1/2} dt \\ &= \frac{1}{n}\int_0^\theta \mathbb{E}\Big[\sum_{j=1}^n \int_0^t H_{1j}(t-z)^2 \overline{\lambda}_z^j dz\Big]^{1/2} dt \\ &\leq \frac{1}{n}\int_0^\theta \Big(n\int_0^t M_{\mu_H}(t-z)^2 ||\Psi||_\infty dz\Big)^{1/2} dt \\ &= \frac{1}{\sqrt{n}}||\Psi||_\infty^{1/2} \int_0^\theta \Big(\int_0^t M_{\mu_H}(t-z)^2 dz\Big)^{1/2} dt := \frac{1}{\sqrt{n}}\tilde{B}_2(\theta), \end{aligned} \tag{3.67}$$

by computing the bracket of a sum over a compensated point process (see [60, Proposition II.4.1.]), using $(\mathcal{A}_\infty^{\mu_H})$ and the fact that $\overline{\lambda}^j$ is bounded by $||\Psi||_\infty$.

- Study of B_3^n. Let us fix some t in $[0,\theta]$ and z in $[0,t]$ and denote $Y_j := H_{1j}(t-z)\overline{\lambda}_z^j$. Since $\overline{\lambda}_z^j$ is the intensity of a solution of the limit equation, it is independent of H_{1j}. Moreover the H_{1j}'s are independent (see (3.9)) and the $\overline{\lambda}^j$'s are independent so the Y_j's are independent. Hence,

$$\mathbb{E}\left[Y_j\right] = \mathbb{E}\left[H_{1j}(t-z)\right]\mathbb{E}\left[\overline{\lambda}_z^j\right] = m_{\mu_H}(t-z)\overline{\lambda}(z),$$

and

$$\mathbb{V}\mathrm{ar}\left(Y_j\right) = \mathbb{V}\mathrm{ar}\left(H_{1j}(t-z)\right)\mathbb{V}\mathrm{ar}\left(\overline{\lambda}_z^j\right) + \mathbb{V}\mathrm{ar}\left(H_{1j}(t-z)\right)\overline{\lambda}(z)^2 + m_{\mu_H}(t-z)^2\mathbb{V}\mathrm{ar}\left(\overline{\lambda}_z^j\right). \quad (3.68)$$

On the one hand, it clearly follows from $(\mathcal{A}_\infty^{\mu_H})$ that $\mathbb{V}\mathrm{ar}(H_{1j}(s)) \le M_{\mu_H}(s)^2$ for all $s \ge 0$. On the other hand, it clearly follows from $(\mathcal{A}_\infty^{\Psi})$ that $\mathbb{V}\mathrm{ar}(\overline{\lambda}_z^j) \le ||\Psi||_\infty^2$ for all $z \ge 0$. Finally (3.68) leads to

$$\mathbb{V}\mathrm{ar}\left(Y_j\right) \le M_{\mu_H}(t-z)^2||\Psi||_\infty^2 + M_{\mu_H}(t-z)^2||\Psi||_\infty^2 + m_{\mu_H}(t-z)^2||\Psi||_\infty^2 \le 3M_{\mu_H}(t-z)^2||\Psi||_\infty^2$$

since m_{μ_H} is dominated by M_{μ_H} and $\overline{\lambda}$ is bounded by $||\Psi||_\infty$. Using the fact that the Y_j's are independent, one has

$$\mathbb{V}\mathrm{ar}\left(\frac{1}{n}\sum_{j=1}^n Y_j\right) = \mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n H_{1j}(t-z)\overline{\lambda}_z^j - m_{\mu_H}(t-z)\overline{\lambda}(z)\right|^2\right] \le \frac{3}{n}M_{\mu_H}(t-z)^2||\Psi||_\infty^2.$$

So, thanks to Cauchy-Schwartz inequality,

$$\begin{aligned} B_3^n(\theta) &\le \int_0^\theta\int_0^t \left(\frac{3}{n}M_{\mu_H}(t-z)^2||\Psi||_\infty^2\right)^{1/2} dzdt \\ &\le \frac{\sqrt{3}}{\sqrt{n}}||\Psi||_\infty \int_0^\theta\int_0^t M_{\mu_H}(t-z)dzdt := \frac{1}{\sqrt{n}}\tilde{B}_3(\theta) \end{aligned} \quad (3.69)$$

- Study of C^n. Since for all $t \ge 0$, $F_{11}(t),\dots,F_{1n}(t)$ are i.i.d. random variables (see (3.10)) with expectation $m_{\nu_F}(t)$ and variance $V_{\nu_F}(t)$, one deduces from Cauchy-Schwartz inequality that

$$C^n(\theta) \le \frac{1}{\sqrt{n}}\int_0^\theta V_{\nu_F}(t)^{1/2}dt := \frac{1}{\sqrt{n}}\tilde{C}(\theta). \quad (3.70)$$

Finally, one deduces from (3.64), (3.66), (3.67), (3.69) and (3.70) that

$$A^n(\theta) \le \frac{\mathrm{Lip}(\Psi)}{\sqrt{n}}\left(\tilde{B}_2(\theta)+\tilde{B}_3(\theta)+\tilde{C}(\theta)\right) + \mathrm{Lip}(\Psi)\int_0^\theta M_{\mu_H}(\theta-z)\delta_n(z)dz. \quad (3.71)$$

Study of $D^n(\theta)$. Since the initial conditions of $N^{n,1}$ and $\overline{N}^1$ are the same (equal to N^1_-), it holds that $S^{n,1}_0 = \overline{S}^1_0$ a.s. If, at a fixed time $t \geq 0$, $S^{n,1}_{t-} \neq \overline{S}^1_{t-}$, then there is at least one point between 0 and t which is not common to both $N^{n,1}_+$ and $\overline{N}^1_+$, that is $\sup_{t' \in [0,t]} |N^{n,i}_{t'} - \overline{N}^i_{t'}| \neq 0$, hence $\mathbb{P}(S^{n,1}_{t-} \neq \overline{S}^1_{t-}) \leq \mathbb{P}(\sup_{t' \in [0,t]} |N^{n,i}_{t'} - \overline{N}^i_{t'}| \neq 0)$. Moreover, since counting processes are piecewise constant with jumps of height 1 a.s., it is clear that

$$\mathbb{P}\left(\sup_{t' \in [0,t]} |N^{n,1}_{t'} - \overline{N}^1_{t'}| \neq 0\right) = \mathbb{P}\left(\sup_{t' \in [0,t]} |N^{n,1}_{t'} - \overline{N}^1_{t'}| \geq 1\right) \leq \mathbb{E}\left[\sup_{t' \in [0,t]} |N^{n,1}_{t'} - \overline{N}^1_{t'}|\right],$$

where we used Markov's inequality. Using (3.42), one has $\mathbb{P}\left(S^{n,1}_{t-} \neq \overline{S}^1_{t-}\right) \leq \delta_n(t)$ and so

$$D^n(\theta) \leq \int_0^\theta \delta_n(t)dt. \tag{3.72}$$

Rewriting (3.63) under the form $\delta_n(\theta) \leq A^n(\theta) + ||\Psi||_\infty D^n(\theta)$, one deduces from (3.71) and (3.72) that

$$\delta_n(\theta) \leq \frac{\mathrm{Lip}(\Psi)}{\sqrt{n}}\left(\tilde{B}_2(\theta) + \tilde{B}_3(\theta) + \tilde{C}(\theta)\right) + \int_0^\theta \left[||\Psi||_\infty + \mathrm{Lip}(\Psi) M_{\mu_H}(\theta - z)\right] \delta_n(z)dz \tag{3.73}$$

where $\tilde{B}_2$, $\tilde{B}_3$, $\tilde{C}$ are respectively defined in (3.67), (3.69) and (3.70). Since M_{μ_H} is locally square integrable, $\theta \mapsto \tilde{B}_2(\theta)$ is locally bounded; since M_{μ_H} is locally integrable, $\theta \mapsto \tilde{B}_3(\theta)$ is locally bounded; since V_{ν_F} is locally square root integrable, $\theta \mapsto \tilde{C}(\theta)$ is locally bounded. Hence we proved the first point of (3.44).

Second point of (3.44). Assume that H_2 is satisfied. Then the decomposition (3.62) is not helpful anymore. Whatever the predictable age processes are, one always has $\lambda^{n,i}_t = \Psi_0(\gamma^{n,i}_t)$ and $\overline{\lambda}^i_t = \Psi_0(\overline{\gamma}(t))$ for all $i = 1, \ldots, n$ and $t > 0$. Remark that in this case, the intensities $\overline{\lambda}^i$ of the limit processes defined by (3.39) are deterministic and equal to $\overline{\lambda}$ defined in Proposition 3.3.7. Instead of (3.63) one should start from $\delta_n(\theta) = \int_0^\theta \mathbb{E}[|\lambda^{n,1}_t - \overline{\lambda}^1_t|]dt$. One can prove in the same way as above that

$$\delta_n(\theta) \leq \mathrm{Lip}(\Psi_0)(B^n_1(\theta) + B^n_2(\theta) + B^n_3(\theta) + C^n(\theta)), \tag{3.74}$$

where B^n_1, B^n_2, B^n_3 and C^n are defined by (3.65). Since the uniform boundedness of Ψ was not used in the study of B^n_1 and C^n then (3.66) and (3.70) still hold. It remains to control B^n_2 and B^n_3 under 1ssumption ($\mathcal{A}_{\Psi=\Psi_0}$).

- Study of B^n_2. Firstly, remind that for all j, $\overline{\lambda}^j_t = \overline{\lambda}(t)$ so, using Cauchy-Schwartz inequality, we have

$$\begin{aligned}
B^n_2(\theta) &\leq \frac{1}{n}\int_0^\theta \mathbb{E}\Big[\Big|\sum_{j=1}^n \int_0^t H_{1j}(t-z)[\overline{N}^j_+(dz) - \overline{\lambda}(z)dz]\Big|^2\Big]^{1/2} dt \\
&= \frac{1}{n}\int_0^\theta \mathbb{E}\Big[\sum_{j=1}^n \int_0^t H_{1j}(t-z)^2 \overline{\lambda}(z)dz\Big]^{1/2} dt \\
&\leq \frac{1}{\sqrt{n}}\int_0^\theta \left(\int_0^t M_{\mu_H}(t-z)^2\overline{\lambda}(z)dz\right)^{1/2} dt := \frac{1}{\sqrt{n}}\overline{B}_2(\theta),
\end{aligned} \tag{3.75}$$

by computing the bracket of a sum over a compensated point process (see [60, Proposition II.4.1.]) and then using $(\mathcal{A}_\infty^{\mu_H})$.

- Study of B_3^n. Secondly, since $\overline{\lambda}_t^j = \overline{\lambda}(t)$ for all j, then B_3^n rewrites as

$$B_3^n(\theta) = \int_0^\theta \mathbb{E}\Big[\Big| \int_0^t \frac{1}{n}\sum_{j=1}^n (H_{1j}(t-z) - m_{\mu_H}(t-z))\,\overline{\lambda}(z)dz\Big|\Big]dt$$
$$\leq \int_0^\theta \int_0^t \mathbb{E}\Big[\Big|\frac{1}{n}\sum_{j=1}^n (H_{1j}(t-z) - m_{\mu_H}(t-z))\,\Big|\Big]\overline{\lambda}(z)dzdt.$$

Using Cauchy-Schwartz inequality, the fact that the H_{1j}'s are i.i.d. with mean function m_{μ_H} and that for all $s \geq 0$, $\mathbb{V}\mathrm{ar}(H_{1j}(s)) \leq M_{\mu_H}(s)^2$ (which follows from $(\mathcal{A}_\infty^{\mu_H})$), we have

$$B_3^n(\theta) \leq \int_0^\theta \int_0^t \frac{1}{\sqrt{n}}\mathbb{V}\mathrm{ar}\,(H_{1j}(t-z))^{1/2}\,\overline{\lambda}(z)dzdt$$
$$\leq \frac{1}{\sqrt{n}}\int_0^\theta \int_0^t M_{\mu_H}(t-z)\overline{\lambda}(z)dzdt := \frac{1}{\sqrt{n}}\overline{B}_3(\theta). \quad (3.76)$$

Finally, one deduces from (3.74), (3.66), (3.75), (3.76) and (3.70) that

$$\delta_n(\theta) \leq \frac{\mathrm{Lip}(\Psi_0)}{\sqrt{n}}\left(\overline{B}_2(\theta) + \overline{B}_3(\theta) + \tilde{C}(\theta)\right) + \int_0^\theta \mathrm{Lip}(\Psi_0)M_{\mu_H}(\theta - z)\delta_n(z)dz, \quad (3.77)$$

where $\overline{B}_2$, $\overline{B}_3$, $\tilde{C}$ are respectively defined in (3.75), (3.76) and (3.70).

Since M_{μ_H} is locally square integrable and $\overline{\lambda}$ is continuous, $\theta \mapsto \overline{B}_2(\theta)$ is locally bounded; since M_{μ_H} is locally integrable and $\overline{\lambda}$ is continuous, $\theta \mapsto \overline{B}_3(\theta)$ is locally bounded; since V_{ν_F} is locally square root integrable, $\theta \mapsto \tilde{C}(\theta)$ is locally bounded. Hence we proved the second point of (3.44).

3.6.g) Proof of Proposition 3.4.2

-1. Assume that H_1 is satisfied. The mean intensity $\overline{\lambda}$ defined in Proposition 3.3.6 is clearly uniformly bounded by $||\Psi||_\infty$.

Looking at (3.73), one wants to find some uniform bounds on $\tilde{B}_2$, $\tilde{B}_3$, $\tilde{C}$ respectively defined in (3.67), (3.69) and (3.70). Firstly, since M_{μ_H} is square integrable and $\overline{\lambda}$ is uniformly bounded, it is clear from (3.67) that

$$\tilde{B}_2(\theta) \leq ||M_{\mu_H}||_2||\Psi||_\infty^{1/2}\theta. \quad (3.78)$$

Secondly, using the integrability of M_{μ_H} and the boundedness of $\overline{\lambda}$, one deduces from (3.69) that

$$\tilde{B}_3(\theta) \leq \sqrt{3}||M_{\mu_H}||_1||\Psi||_\infty\theta. \quad (3.79)$$

Finally, since V_{ν_F} is uniformly bounded, one deduces from (3.70) that

$$\tilde{C}(\theta) \leq ||V_{\nu_F}||_\infty^{1/2}\theta. \quad (3.80)$$

Moreover, using the fact that δ_n is a non-decreasing function, we find, with the notation $\alpha = \mathrm{Lip}(\Psi)||M_{\mu_H}||_1$,

$$\begin{cases} \int_0^\theta \mathrm{Lip}(\Psi)M_{\mu_H}(\theta - z)\delta_n(z)dz \leq \alpha\delta_n(\theta) \\ \int_0^\theta ||\Psi||_\infty\delta_n(z)dz \leq ||\Psi||_\infty\theta\delta_n(\theta). \end{cases} \quad (3.81)$$

Gathering (3.78), (3.79), (3.80) and (3.81) one deduces from (3.73) that

$$\delta_n(\theta) \leq \frac{\mathrm{Lip}(\Psi)}{\sqrt{n}} \left(||M_{\mu_H}||_2 ||\Psi||_\infty^{1/2} + \sqrt{3}||M_{\mu_H}||_1 ||\Psi||_\infty + ||V_{\nu_F}||_\infty^{1/2} \right) \theta + (\alpha + ||\Psi||_\infty \theta)\delta_n(\theta),$$

which leads to $\delta_n(\theta) \leq \beta(\Psi, \mu_H, \nu_F)\theta n^{-1/2}$ with

$$\beta(\Psi, \mu_H, \nu_F) := \frac{\mathrm{Lip}(\Psi)}{1 - (\alpha + ||\Psi||_\infty \theta)} \left(||M_{\mu_H}||_2 ||\Psi||_\infty^{1/2} + \sqrt{3}||M_{\mu_H}||_1 ||\Psi||_\infty + ||V_{\nu_F}||_\infty^{1/2} \right), \tag{3.82}$$

as soon as $\alpha + ||\Psi||_\infty \theta < 1$, i.e. $\theta < (1-\alpha)/||\Psi||_\infty$.

-2. Assume H_2 is satisfied. Under the assumptions of Proposition 3.4.2, the mean intensity $\overline{\lambda}$ defined in Proposition 3.3.7 is uniformly bounded thanks to Lemma 3.7.6.

Looking at (3.77), one wants to find some uniform bounds on $\overline{B}_2$, $\overline{B}_3$, $\tilde{C}$ respectively defined in (3.75), (3.76) and (3.70). In the same way as above, one can deduce from (3.75), (3.76) and (3.70) that

$$\begin{cases} \overline{B}_2(\theta) \leq ||M_{\mu_H}||_2 ||\overline{\lambda}||_\infty^{1/2} \theta \\ \overline{B}_3(\theta) \leq ||M_{\mu_H}||_1 ||\overline{\lambda}||_\infty \theta \\ \tilde{C}(\theta) \leq ||V_{\nu_F}||_\infty^{1/2} \theta. \end{cases} \tag{3.83}$$

Using (3.83) and the first equation of (3.81) which is still valid is this case, one deduces from (3.77) that

$$\delta_n(\theta) \leq \frac{\mathrm{Lip}(\Psi)}{\sqrt{n}} \left(||M_{\mu_H}||_2 ||\overline{\lambda}||_\infty^{1/2} + ||M_{\mu_H}||_1 ||\overline{\lambda}||_\infty + ||V_{\nu_F}||_\infty^{1/2} \right) \theta + \alpha\delta_n(\theta),$$

which leads to $\delta_n(\theta) \leq \beta(\Psi, \mu_H, \nu_F)\theta n^{-1/2}$ with

$$\beta(\Psi, \mu_H, \nu_F) := \frac{\mathrm{Lip}(\Psi)}{1-\alpha} \left(||M_{\mu_H}||_2 ||\overline{\lambda}||_\infty^{1/2} + ||M_{\mu_H}||_1 ||\overline{\lambda}||_\infty + ||V_{\nu_F}||_\infty^{1/2} \right), \tag{3.84}$$

for every $\theta \geq 0$. Notice that an explicit expression of $||\overline{\lambda}||_\infty$ with respect to M_{μ_H}, m_{ν_F} and Ψ can be obtained thanks to Lemma 3.7.6.

3.7 Lemmas

3.7.a) Point processes

Here we collect some technical lemmas about point processes.

The following lemma is used to show the well-posedness of the studied point processes.

Lemma 3.7.1. *Let $n \geq 1$ be an integer, let $(g_i)_{i=1,\dots,n}$ be a family of locally integrable functions, $(a_i)_{i=1,\dots,n}$ be a family of non-negative real numbers and $h : \mathbb{R}_+ \to \mathbb{R}$ be a locally integrable function. Let $(\Pi^i(dt, dx))_{i\geq 1}$ be some i.i.d. $\mathbb{F}$-Poisson measures with intensity 1 on $\mathbb{R}_+^2$.*

Let $(N_t^i)_{t\geq 0}^{i=1,..,n}$ be a family of counting processes such that for $i = 1, .., n$ and all $t \geq 0$,

$$N_t^i = \int_0^t \int_0^\infty \mathbb{1}_{\{x \leq \lambda_{t'}^i\}} \Pi^i(dt', dx), \tag{3.85}$$

where the λ^i's are $\mathbb{F}$-predictable processes such that $\lambda_t^i \leq g_i(t)+a_i\frac{1}{n}\sum_{j=1}^n \int_0^{t-} |h(t-u)|N^j(du)$. Then, the linear multivariate Hawkes process $(\tilde{N}_t^i)_{t\geq 0}^{i=1,..,n}$ defined by

$$\tilde{N}_t^i = \int_0^t \int_0^\infty \mathbb{1}_{\left\{x \leq g_i(t') + a_i \frac{1}{n}\sum_{j=1}^n \int_0^{t'-} |h(t'-z)|\tilde{N}^j(dz)\right\}} \Pi^i(dt', dx), \tag{3.86}$$

is such that for all $i = 1, \ldots, n$, $\tilde{N}^i$ stochastically dominates N^i in the sense that $N^i \subset \tilde{N}^i$ where N^i (resp. $\tilde{N}^i$) denotes the point process associated with the counting process $(N_t^i)_{t\geq 0}$ (resp. $(\tilde{N}_t^i)_{t\geq 0}$). In particular, the processes $(N_t^i)_{t\geq 0}^{i=1,..,n}$ are well-defined.

Proof. First, let us note that the processes $(\tilde{N}_t^i)_{t\geq 0}^{i=1,..,n}$ are well-defined by the Galton-Watson representation of the linear Hawkes process introduced in [71] when the g_i's are constant in time (see Proposition 2.8.3 when the g_i's are more generally locally integrable functions).

We are going to show by induction that

$$\forall t \geq 0,\ \lambda_t^i \leq \tilde{\lambda}_t^i := g_i(t) + a_i \frac{1}{n}\sum_{j=1}^n \int_0^{t-} |h(t-u)|\tilde{N}^j(du).$$

Indeed, for all time t less than the first point of either $N := \cup_{i=1}^n N^i$ or $\tilde{N} := \cup_{i=1}^n \tilde{N}^i$, the respective intensities are such that $\lambda_t^i \leq g_i(t) = \tilde{\lambda}_t^i$. Hence, the first point of $N \cup \tilde{N}$ (denoted by T_1) is a point of $\tilde{N}$ (and possibly a point of N). Let us denote $(T_k)_{k\geq 1}$ the ordered sequence of the points of $N \cup \tilde{N}$.

Let us fix some $k_0 \geq 1$. Suppose that for every $i = 1, \ldots, n$, $\lambda_t^i \leq \tilde{\lambda}_t^i$, for all $t \leq T_{k_0}$. Then, it is clear that for every $k = 1, \ldots, k_0$, $T_k \in \tilde{N}$, hence for every $i = 1, \ldots, n$, $\tilde{N}^i$ stochastically dominates N^i up to time T_{k_0+1-}. Moreover, it implies that for every $i = 1, \ldots, n$ and for all t in $(T_{k_0}, T_{k_0+1}]$,

$$\begin{aligned}
\lambda_t^i &\leq g_i(t) + a_i \frac{1}{n}\sum_{j=1}^n \int_0^{t-} |h(t-z)|N^j(dz) \\
&\leq g_i(t) + a_i \frac{1}{n}\sum_{j=1}^n \int_0^{t-} |h(t-z)|\tilde{N}^j(dz) = \tilde{\lambda}_t^i,
\end{aligned}$$

since $|h|$ is a non negative function. Therefore, by induction on k, the desired stochastic domination holds true for all time. In particular, the dominated processes are well-defined. □

Lemma 3.7.2. *If N admits the bounded $\mathbb{F}$-intensity λ_t and $(S_{t-})_{t\geq 0}$ denote its associated predictable age process, then the distribution of S_{t-} denoted by w_t is such that $t \mapsto w_t$ belongs to $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$.*

Proof. This continuity result comes from the fact that the probability that N has a point in an interval goes to 0 as the size of the interval goes to 0. Indeed, let t, t' be positive real numbers, $\mathbb{P}(N([t, t+t')) \neq 0) \leq \mathbb{E}[N([t, t+t'))] = \mathbb{E}[\int_t^{t+t'} \lambda_z dz]$ goes to 0 as t' goes to 0. Moreover, $S_{(t+t')-} = S_{t-} + t'$ as soon as there is no point of N in the interval $[t, t+t')$ and so one has

$$\tilde{W}_1(w_{t+t'}, w_t) \leq \mathbb{E}\left[\min\left(\left|S_{(t+t')-} - S_{t-}\right|, 1\right)\right] \leq t' + \mathbb{P}(N([t, t+t')) \neq 0) \xrightarrow[t'\to 0]{} 0,$$

reminding that $\tilde{W}_1$ denotes the modified Wasserstein distance defined in (3.17). The same argument for $t' < 0$ gives continuity. □

3.7.b) Analytic lemmas

Here, we collect some analytic lemmas regarding the convolution equations used throughout the present chapter. First, here are two lemmas introduced in [43].

The first one is an easy application of Fubini's Theorem [43, Lemma 22].

Lemma 3.7.3. *Let $\Phi : \mathbb{R}_+ \to \mathbb{R}$ be locally integrable and let $\alpha : \mathbb{R}_+ \to \mathbb{R}$ have bounded variations on compact intervals, satisfying $\alpha(0) = 0$. Then, for all $t \geq 0$,*

$$\int_0^t \int_0^s \Phi(s-u)\, d\alpha(u)\, ds = \int_0^t \Phi(t-s)\, \alpha(s)\, ds,$$

where the integral has to be understood in the Stieltjes' sense.

The second one is a rather classical generalization of Grönwall Lemma [43, Lemma 23].

Lemma 3.7.4. *Let $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be locally integrable and $g : \mathbb{R}_+ \to \mathbb{R}_+$ be locally bounded.*

(i) Consider a locally bounded non-negative function u such that for all $t \geq 0$, $u_t \leq g_t + \int_0^t \Phi(t-s)\, u_s ds$. Then, $\sup_{t\in[0,T]} u_t \leq C_T \sup_{t\in[0,T]} g_t$, for some constant C_T depending only on $T > 0$ and Φ. Moreover, C_T can be taken as a non-decreasing function of T.

(ii) Consider a sequence of locally bounded non-negative functions u^n such that for all $t \geq 0$, all $n \geq 0$, $u_t^{n+1} \leq \int_0^t \Phi(t-s)\, u_s^n ds$. Then, $\sup_{t\in[0,T]} \sum_{n\geq 0} u_t^n \leq C_T$, for some constant C_T depending only on $T > 0$, u^0 and Φ.

(iii) Consider a sequence of locally bounded non-negative functions u^n such that for all $t \geq 0$, all $n \geq 0$, $u_t^{n+1} \leq g_t + \int_0^t \Phi(t-s)\, u_s^n ds$. Then, $\sup_{t\in[0,T]} \sup_{n\geq 0} u_t^n \leq C_T$, for some constant C_T depending only on $T > 0$, u^0, g and Φ.

Note that we added to the first statement that C_T can be taken as a non-decreasing function of T. It is not given in [43, Lemma 23] but it is a direct consequence of the proof.

Then, here is a well-posedness result which is a generalization of [43, Lemma 24].

Lemma 3.7.5. *Let $\Phi : \mathbb{R} \to \mathbb{R}_+$ be Lipschitz-continuous, $h : \mathbb{R}_+ \to \mathbb{R}$ be locally integrable and $f_0 : \mathbb{R}_+ \to \mathbb{R}$ be continuous. The equation*

$$m_t = \int_0^t \Phi\left(\int_0^{t'} h(t'-z) dm_z + f_0(t')\right) dt' \tag{3.87}$$

has a unique locally bounded solution. Furthermore, m is of class $\mathcal{C}^1$ on $\mathbb{R}_+$.

Proof. The proof is similar to [43, Lemma 24]. We refrain from reproducing it here; instead, we only indicate the minor changes that are required to make it fit the current framework, i.e. the addition of the function f_0. The "uniqueness" part is exactly the same. The "existence" part requires f_0 to be locally integrable in order to have locally boundedness in the Picard iteration. Finally, we need f_0 to be continuous to show by induction that at each step of the Picard iteration the function is $\mathcal{C}^1$ on $\mathbb{R}_+$ and so it is for the limit, that is the solution of (3.87). □

Here is given an analytic result which is used to give a uniform bound on the mean intensity of a Hawkes process under stationary conditions.

Lemma 3.7.6. *Let $\Phi : \mathbb{R} \to \mathbb{R}_+$ be Lipschitz-continuous and $h : \mathbb{R}_+ \to \mathbb{R}$ be integrable such that $\mathrm{Lip}(\Phi)||h||_1 < 1$. Moreover, let $f_0 : \mathbb{R}_+ \to \mathbb{R}$ be uniformly bounded.*

If $g : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous function satisfying

$$g(t) \leq \Phi\left(\int_0^t h(t-u)g(u)du + f_0(t)\right) \tag{3.88}$$

for every $t > 0$, then g is uniformly upper bounded by

$$M := \frac{\Phi(0) + \mathrm{Lip}(\Phi)||f_0||_\infty}{1 - \mathrm{Lip}(\Phi)||h||_1}.$$

Proof. For any $t > 0$,

$$g(t) \leq \Phi(0) + \mathrm{Lip}(\Phi)\left(\int_0^t |h(t-u)|g(u)du + ||f_0||_\infty\right)$$

hence, thanks to the continuity of g, for every $T \geq 0$,

$$\sup_{t\in[0,T]} g(t) \leq \Phi(0) + \mathrm{Lip}(\Phi)\left(||h||_1 \sup_{t\in[0,T]} g(t) + ||f_0||_\infty\right)$$

and

$$\sup_{t\in[0,T]} g(t) \leq \frac{1}{1 - \mathrm{Lip}(\Phi)||h||_1}\left[\Phi(0) + \mathrm{Lip}(\Phi)||f_0||_\infty\right] = M.$$

□

Chapter

4 Fluctuations for generalized Hawkes processes

Abstract. A particular case of the mean field interacting age-dependent Hawkes processes is studied in the present chapter. The propagation of chaos and associated law of large numbers (when the number of processes n goes to $+\infty$) being granted by the study performed in the previous chapter (Corollary 3.4.5 and Equation (3.50) in particular), the aim of the present chapter is to prove the resulting functional central limit theorem. It involves the study of a measure-valued process describing the fluctuations (at scale $n^{-1/2}$) of the empirical measure of the ages around its limit value. This fluctuation process is proved to converge towards a limit process characterized by a limit system of stochastic differential equations driven by a Gaussian noise instead of Poisson.

Contents

4.1 Introduction

We refer to Section 3.1 for an introduction to the notations, some application fields of Hawkes processes in general and the suitability of age-dependent random Hawkes processes (ADRHP) for modelling in neuroscience.

In Chapter 3, age-dependent Hawkes processes are introduced and the asymptotic behaviour of n mean-field interacting age-dependent Hawkes processes $(N^{n,i})_{i=1,\dots,n}$ when n

goes to infinity is investigated. Furthermore, the associated age processes $((S_t^{n,i})_{t\geq 0})_{i=1,\dots,n}$ are also studied within this mean-field framework: law of large numbers (3.49) is proved, the rate of convergence (for fixed time t) being at most $n^{-1/2}$ (3.50). Below we denote by $\overline{\mu}^n$ the empirical measure of the ages and by P its limit. In light of the bound obtained on the rate of convergence, the fluctuation process defined by $\eta^n = \sqrt{n}(\overline{\mu}^n - P)$ is expected to describe, on the right scale, the second order term appearing in the expansion of the mean-field approximation, the first order term being given by the law of large numbers.

Following the approach developed in [50], we prove in the present chapter that the fluctuations satisfy a functional central limit theorem (CLT) in a suitable distributional space: the limit of the normalized fluctuations is described by means of a stochastic differential equation in infinite dimension driven by a Gaussian noise in comparison with the Poisson noise appearing in Chapter 3. To do so, we regard the fluctuation process η^n as taking values in a Hilbert space, namely the dual of some Sobolev space of test functions[1]. The index of regularity of the dual space, in one-to-one correspondence with the regularity of the test functions in the Sobolev space, is prescribed by the tightness property we are able to provide to the sequence $(\eta^n)_{n\geq 1}$ and by the form of the generator of the limiting McKean-Vlasov dynamics identified in Chapter 3. Let us precise that this generator is the one associated with the renewal dynamics of the (PPS) system as highlighted by Proposition 3.3.8.

Although the choice of this index of regularity is rather constrained, the choice of the domain supporting the Sobolev space is somewhat larger. Indeed, two options are available, depending on the way we consider the process η^n, either over a finite time horizon, namely $(\eta_t^n)_{0\leq t\leq \theta}$ for some $\theta \geq 0$, or in infinite horizon, namely $(\eta_t^n)_{t\geq 0}$.

In the first case, we may use the fact that there exists a compact K_θ (which is growing with θ) such that η_t^n is supported in K_θ for all t in $[0,\theta]$. Hence, one could regard, for all $\theta \geq 0$, the fluctuation process $(\eta_t^n)_{0\leq t\leq \theta}$ as a process with values in the dual of a standard Sobolev space of functions with support in K_θ. The main drawback of such an approach is that the space of trajectories within which the CLT takes place depends on the time horizon θ. To bypass this issue, one may be willing to work directly on the entire positive time line $\mathbb{R}_+$, but then, it is not possible anymore to find a compact subset K supporting the measures η_t^n, for all $t \geq 0$, since $\cup_{\theta\geq 0} K_\theta = \mathbb{R}_+$. A convenient strategy to sidestep this fact is to use a Sobolev space supported by the entire $\mathbb{R}_+$. Yet, standard Sobolev spaces supported by $\mathbb{R}_+$ fail to accommodate with our purpose, since, as made clear by the proof below, constant functions are required to belong to the space of test functions. Therefore, instead of a standard Sobolev space, we may use a weighted Sobolev space, provided that the weight satisfies suitable integrability properties.

In order to state our CLT on the whole time interval, the second approach is preferred. Furthermore, we also feel convenient to choose Sobolev spaces with polynomial weights, the definition of which is recalled in Section 4.4.a) below. This choice is quite comfortable because Sobolev spaces with polynomial weights are well-documented in the literature. In particular, results on the connection between spaces weighted by different powers, Sobolev embedding theorems and Maurin's theorem, are well-known. It is worth noting that, provided that constant functions can be chosen as test functions, the precise value of the power in the polynomial weight of the Sobolev space do not really matter in our analysis: more generally, a different choice of family of weights would have been possible and, somehow,

[1]One could regard η^n as taking values in the Sobolev space directly thanks to Riesz representation theorem. However, a major issue comes from this representation. This issue is tackled in Appendix A.3

it would have led to a result equivalent to ours. In this regard, we stress, at the end of the chapter, the fact that our result in infinite horizon is in fact equivalent to what we would have obtained by implementing the first of the two approaches mentioned above instead of the second one: roughly speaking, one can recover our result by sticking together the CLTs obtained on each finite interval of the form $[0, \theta]$, for $\theta \geq 0$; conversely, one can prove, from our statement, that, on any finite interval $[0, \theta]$, the CLT holds true in the dual space of a standard Sobolev space supported by K_θ.

The Hilbertian approach used in this chapter has been already implemented in the diffusion processes framework [50, 82, 97, 101]. Let us mention here what are the main differences between these earlier results and ours:

- Under general non-degeneracy conditions, the marginal laws of a diffusion process are not compactly supported. The unboundedness of the support imposes the choice of weighted Sobolev spaces even in finite time horizon. In this framework, Sobolev spaces with polynomial weights are especially adapted to carry solutions with moments that are finite up to some order only. In that case, the choice of the power in the weight is explicitly prescribed by the maximal order up to which the solution has a finite moment. As already mentioned, this differs from our case: in the present chapter, particles (namely, the ages of the neurons) are compactly supported over any finite time interval and thus, have finite moments of any order. Once again, this is the reason why the choice of the power, and more generally of the weight, in the Sobolev space is much larger.

- Unlike point processes, diffusion processes are time continuous. Also, their generator is both local and of second order, whereas the generator for the point process identified in the mean-field limit in Chapter 3 is both of the first order and nonlocal. As a first consequence, the indices of regularity of the various Sobolev spaces used in this chapter differ from those used in the diffusive framework. Also, the space of trajectories cannot be the same: although the limit process in our CLT has continuous trajectories, we must work with a space of càdlàg functions in order to accommodate with the jumps of the fluctuation process. Surprisingly, jumps do not just affect the choice of the functional space used to state the CLT (namely space of càdlàg versus space of continuous functions) but it also dictates the metric used to estimate the error in the Sznitman coupling between the age-dependent Hawkes process investigated in Chapter 3 and its mean-field limit. Indeed, counting processes take integer values so the standard trick used for diffusion processes that consists in getting stronger estimates for the Sznitman coupling by considering L^p-norms, for $p > 2$, is completely useless here. Therefore, we develop a specific approach by providing higher order estimates of the error in the Sznitman coupling in the total variation sense. Up to our knowledge, this argument is completely new.

Let us mention that the fluctuations of jumps processes have been the object of previous publications [140, 161]. However, the CLTs are established in the fluid limit, namely small jumps at high frequency so that the jumps vanish at the limit. The techniques developed in those articles are useless here since the framework of the present chapter does not fall into the fluid limit framework: in our case, the limit processes are also jump processes.

The present chapter is organized as follows. The model is given in Section 4.2: it is a particular case of the model introduced in Chapter 3. Then, the main estimates required

in this work are given in Section 4.3. These can be seen as the extension, to higher orders, of the estimates used in Chapter 3 to get the bound $n^{-1/2}$ on the rate of convergence. These key estimates are used to prove tightness for the distribution η^n in a Hilbert space that is the dual of some weighted Sobolev space. Under regularity assumptions on the intensity function Ψ and the interaction function h, we finally prove in Section 4.5.b) the convergence of the fluctuation process. Furthermore, its limit is characterized by a system of stochastic differential equations, driven by a Gaussian process with explicit covariance, and involving an auxiliary process with values in $\mathbb{R}$ (Theorem 4.5.12).

General notations

- Statistical distributions are referred to as laws of random variables to avoid confusion with distributions in the analytical sense that are linear forms acting on some test function space.
- The space of bounded functions of class $\mathcal{C}^k$, with bounded derivatives of each order less than k is denoted by $\mathcal{C}_b^k$.
- The space of càdlàg (right continuous with left limits) functions is denoted by $\mathcal{D}$.
- For μ a measure on E and φ a function on E, we denote $\langle \mu, \varphi \rangle := \int_E \varphi(x)\mu(dx)$ when it makes sense.
- If a quantity Q depends on the time variable t, then we most often use the notation Q_t when it is a random process in comparison with $Q(t)$ when it is a deterministic function.
- We say that the quantity $Q_n(\sigma)$, which depends on an integer n and a parameter $\sigma \in \mathbb{R}^d$, is bounded up to a locally bounded function (which does not depend on n) by $f(n)$, denoted by $Q_n(\sigma) \lesssim_\sigma f(n)$, if there exists a locally bounded function $C : \mathbb{R}^d \to \mathbb{R}_+$ such that, for all n, $|Q_n(\sigma)| \leq C(\sigma) f(n)$.
- Throughout this chapter, C denotes a constant that may change from line to line.

4.2 Definitions and propagation of chaos

The mathematical framework used in this chapter being the same as the one used in Chapter 3, we refer to Section 3.2 for the main notations and definitions. The model considered here is a particular case of the age dependent random Hawkes process (ADRHP) introduced in Chapter 3. The main differences are:

- the interaction functions H_{ij} are now supposed to be equal to a deterministic function $h : \mathbb{R}_+ \to \mathbb{R}$,
- the functions F_{ij} are now supposed to be equal to the null function.

Hence, the parameters of the model are: the number of particles n, the interaction function h, the intensity function Ψ and the law of the initial conditions[2] ζ_{N_-}.

[2]In fact, since the functions F_{ij} (which may depend on the initial conditions of an ADRHP) are null, the model used in the present chapter depends on the weaker information given by the law of the age at time 0, i.e. $S_0 = S_{0-}$, instead of depending on the law of the whole past ζ_{N_-}. Nevertheless, we keep ζ_{N_-} to preserve coherence among the chapters.

For sake of simplicity, all the assumptions made on the parameters are gathered here:

$\left(\mathcal{A}^{\zeta_{N_-}}_{u^{\text{in}},\infty}\right)$:	If N_- is distributed according to ζ_{N_-} and T_0 denotes the closest point of N_- to 0, then $-T_0$ admits a density with respect to the Lebesgue measure, denoted by u^{in}, which is uniformly bounded with compact support so that there exists a constant $C > 0$ such that $-T_0 \leq C$ almost surely (a.s.). The smallest possible constant C is denoted by M_{T_0}.
$(\mathcal{A}^h_\infty)$:	The interaction function h is locally bounded. Denote by, for all $t \geq 0$, $h_\infty(t) := \max_{s\in[0,t]} h(s) < +\infty$.
$(\mathcal{A}^h_{\text{Höl}})$:	There exist two positive constants denoted by $\text{Höl}(h)$ and $\beta(h)$ such that for all $t, s \geq 0$, $\lvert h(t) - h(s)\rvert \leq \text{Höl}(h)\lvert t - s\rvert^{\beta(h)}$.
$\left(\mathcal{A}^\Psi_{y,\mathcal{C}^2}\right)$:	For all $s \geq 0$, the function $\Psi_s : y \mapsto \Psi(s,y)$ is of class $\mathcal{C}^2$. Furthermore, $\lvert\lvert\frac{\partial\Psi}{\partial y}\rvert\rvert_\infty := \sup_{s,y} \lvert\frac{\partial\Psi}{\partial y}(s,y)\rvert < +\infty$ and $\lvert\lvert\frac{\partial^2\Psi}{\partial y^2}\rvert\rvert_\infty < +\infty$. The constant $\lvert\lvert\frac{\partial\Psi}{\partial y}\rvert\rvert_\infty$ is denoted by $\text{Lip}(\Psi)$ (this in order to be coherent with the notation in Chapter 3).
$(\mathcal{A}^\Psi_\infty)$:	The function Ψ is uniformly bounded, that is $\lvert\lvert\Psi\rvert\rvert_\infty < +\infty$.
$\left(\mathcal{A}^\Psi_{s,\mathcal{C}^2_b}\right)$:	For all y in $\mathbb{R}$, the functions $s \mapsto \Psi(s,y)$ and $s \mapsto \frac{\partial\Psi}{\partial y}(s,y)$ respectively belong to $\mathcal{C}^2_b$ and $\mathcal{C}^1_b$. Furthermore, the functions $y \mapsto \lvert\lvert\Psi(\cdot,y)\rvert\rvert_{\mathcal{C}^2_b}$ and $y \mapsto \lvert\lvert\frac{\partial\Psi}{\partial y}(\cdot,y)\rvert\rvert_{\mathcal{C}^1_b}$ are locally bounded[3].
$\left(\mathcal{A}^\Psi_{s,\mathcal{C}^4_b}\right)$:	For all y in $\mathbb{R}$, the function $s \mapsto \Psi(s,y)$ belongs to $\mathcal{C}^4_b$ and $y \mapsto \lvert\lvert\Psi(\cdot,y)\rvert\rvert_{\mathcal{C}^4_b}$ is locally bounded.

Remark 4.2.1. *Note that:*

- *Assumption $(\mathcal{A}^h_{\text{Höl}})$ implies Assumption $(\mathcal{A}^h_\infty)$,*
- *the assumptions regarding the intensity function Ψ are rather technical, nevertheless Assumptions $(\mathcal{A}^\Psi_{y,\mathcal{C}^2})$, $(\mathcal{A}^\Psi_\infty)$ and $(\mathcal{A}^\Psi_{s,\mathcal{C}^2_b})$ are satisfied as soon as Ψ belongs to $\mathcal{C}^2_b$. Furthermore, Assumption $(\mathcal{A}^\Psi_{s,\mathcal{C}^4_b})$ is satisfied if Ψ is in $\mathcal{C}^4_b$.*

Let $(\mathcal{A}_{\text{LLN}})$ be satisfied if $(\mathcal{A}^{\zeta_{N_-}}_{u^{\text{in}},\infty})$, $(\mathcal{A}^h_\infty)$, $(\mathcal{A}^\Psi_{y,\mathcal{C}^2})$ and $(\mathcal{A}^\Psi_\infty)$ are satisfied. These four assumptions also appear in Chapter 3, where they are used to prove propagation of chaos. Furthermore, let $(\mathcal{A}_{\text{TGN}})$ be satisfied if $(\mathcal{A}_{\text{LLN}})$ and $(\mathcal{A}^\Psi_{s,\mathcal{C}^2_b})$ are satisfied. It is used in the present chapter to prove tightness of the fluctuations. Finally, let $(\mathcal{A}_{\text{CLT}})$ be satisfied if $(\mathcal{A}_{\text{TGN}})$, $(\mathcal{A}^h_{\text{Höl}})$ and $(\mathcal{A}^\Psi_{s,\mathcal{C}^4_b})$ are satisfied. It is used in the present chapter to prove convergence of the fluctuations.

Notice that Assumption $(\mathcal{A}^{\zeta_{N_-}}_{u^{\text{in}},\infty})$ implies that if N_- is distributed according to ζ_{N_-}, then the age processes associated with N are such that, almost surely,

$$\text{for all } t \geq 0,\ S_t \leq M_{T_0} + t \text{ and } S_{t-} \leq M_{T_0} + t. \tag{4.1}$$

As a particular case of the ADRHP defined in Chapter 3, we consider in this chapter age dependent Hawkes processes as described below. We drop the adjective "random" since the interaction function h is deterministic. In comparison with Representation 3.2.3, the latter are described by the following representation induced by the thinning procedure.

[3]The definitions of the norms $\lvert\lvert\cdot\rvert\rvert_{\mathcal{C}^k_b}$, for all $k \geq 0$, can be found in Section 4.4.a)

Representation 4.2.2. *Let $(\Pi^i(dt,dx))_{i\geq 1}$ be some i.i.d. $\mathbb{F}$-Poisson measures with intensity 1 on $\mathbb{R}_+^2$. Let $(N_-^i)_{i\geq 1}$ be some i.i.d. point processes on $\mathbb{R}_-$ distributed according to ζ_{N_-}.*

Let $(N_t^i)_{t\geq 0}^{i=1,..,n}$ be a family of counting processes such that, for $i=1,..,n$, and all $t\geq 0$,

$$N_t^i = \int_0^t \int_0^{+\infty} \mathbb{1}_{\left\{x \leq \Psi\left(S_{t'-}^i, \frac{1}{n}\sum_{j=1}^n \left(\int_0^{t'-} h(t'-z)N_+^j(dz)\right)\right)\right\}} \Pi^i(dt',dx), \tag{4.2}$$

where $(S_{t-}^i)_{t\geq 0}$ is the predictable age process associated with $N^i = N_-^i \cup N_+^i$ and N_+^i is the point process associated with the counting process $(N_t^i)_{t\geq 0}$. Then, $(N^i)_{i=1,..,n}$ is an age dependent Hawkes process (ADHP) with parameters (n,h,Ψ,ζ_{N_-}).

Remark 4.2.3. *Note that an ADHP is in fact a (deterministic) measurable function of the Poisson measures $(\Pi^i(dt,dx))_{i\geq 1}$. More classically, as in Definition 3.2.1, an ADHP can be characterized by its stochastic intensity as follows. If $(N^i)_{i=1,..,n}$ is an ADHP with parameters (n,h,Ψ,ζ_{N_-}), then, for all $i=1,\dots,n$, the point process N^i admits an $\mathbb{F}$-intensity λ_t^i defined, for all $t\geq 0$, by*

$$\lambda_t^i = \Psi\left(S_{t-}^i, \frac{1}{n}\sum_{j=1}^n \int_0^{t-} h(t-z)N_+^j(dz)\right). \tag{4.3}$$

Going back and forth between the definition via the intensities (4.3) and Representation 4.2.2 is standard (see Section 3.2.d) for more insights). Furthermore, Proposition 3.2.4 gives that, under Assumption ($\mathcal{A}_{LLN}$), there exists an ADHP $(N^i)_{i=1,..,n}$ with parameters (n,h,Ψ,ζ_{N_-}) such that $t\mapsto \mathbb{E}[N_t^1]$ is locally bounded.

Notice that, since the initial conditions $(N_-^i)_{i=1,..,n}$ are i.i.d. and the Poisson measures $(\Pi^i(dt,dx))_{i\geq 1}$ are i.i.d., the processes N^i, $i=1,\dots,n$, defined by (4.2) are exchangeable. Being interested in the mean-field limit of ADHPs, let us note that, according to the analysis performed in Chapter 3, the *limit equation* (in comparison with (4.2)) with parameters (h,Ψ,ζ_{N_-}) is given by

$$\forall t>0,\ \overline{N}_t = \int_0^t \int_0^{+\infty} \mathbb{1}_{\left\{x \leq \Psi\left(\overline{S}_{t'-}, \int_0^{t'-} h(t'-z)\mathbb{E}\left[\overline{N}_+(dz)\right]\right)\right\}} \Pi(dt',dx), \tag{4.4}$$

where $\Pi(dt',dx)$ is an $\mathbb{F}$-Poisson measure on $\mathbb{R}_+^2$ with intensity 1 and $(\overline{S}_{t-})_{t\geq 0}$ is the predictable age process associated with $\overline{N} = \overline{N}_- \cup \overline{N}_+$ where $\overline{N}_-$ is a point process distributed according to ζ_{N_-} and $\overline{N}_+$ is the point process associated with the counting process $(\overline{N}_t)_{t\geq 0}$.

Under Assumption ($\mathcal{A}_{\text{LLN}}$), Proposition 3.3.6 states existence and uniqueness of the limit process $\overline{N}$. In particular, there exists a continuous function $\overline{\lambda}:\mathbb{R}_+\to\mathbb{R}$ (which depends on the parameters h, Ψ and ζ_{N_-}) such that if $(\overline{N}_t)_{t\geq 0}$ is a solution of (4.4) then $\mathbb{E}[\overline{N}_+(dt)] = \overline{\lambda}(t)dt$. Let us define the deterministic function $\overline{\gamma}$ by, for all $t\geq 0$,

$$\overline{\gamma}(t) := \int_0^t h(t-z)\overline{\lambda}(z)dz. \tag{4.5}$$

Notice that $\overline{\gamma}(t')$ is the integral term $\int_0^{t'-} h(t'-z)\mathbb{E}\left[\overline{N}_+(dz)\right]$ appearing in (4.4).

Once the limit equation is well-posed, following the ideas of Sznitman in [156], it is easy to construct a suitable coupling between ADHPs and i.i.d. solutions of the limit equation (4.4). More precisely, consider

- a sequence $(N_-^i)_{i\geq 1}$ of i.i.d. point processes distributed according to ζ_{N_-};
- a sequence $(\Pi^i(dt', dx))_{i\geq 1}$ of i.i.d. $\mathbb{F}$-Poisson measures with intensity 1 on $\mathbb{R}_+^2$.

Under Assumption $(\mathcal{A}_{\text{LLN}})$, we have existence of both ADHPs and the limit process $\overline{N}$. Hence, one can build simultaneously:
- a sequence (indexed by $n \geq 1$) $(N^{n,i})_{i=1,\dots,n}$ of ADHPs with parameters $(n, h, \Psi, \zeta_{N_-})$ according to Representation 4.2.2 namely

$$N_t^{n,i} = \int_0^t \int_0^{+\infty} \mathbb{1}_{\left\{x \leq \Psi\left(S_{t'-}^{n,i}, \gamma_{t'}^n\right)\right\}} \Pi^i(dt', dx) \tag{4.6}$$

and past given by N_-^i, where $\gamma_{t'}^n := n^{-1} \sum_{j=1}^n \int_0^{t'-} h(t'-z) N_+^{n,j}(dz)$,
- and a sequence $(\overline{N}_t^i)_{t\geq 0}^{i\geq 1}$ of i.i.d. solutions of the limit equation namely

$$\overline{N}_t^i = \int_0^t \int_0^{+\infty} \mathbb{1}_{\left\{x \leq \Psi\left(\overline{S}_{t'-}^i, \overline{\gamma}(t')\right)\right\}} \Pi^i(dt', dx), \tag{4.7}$$

and past given by N_-^i, where $\overline{\gamma}$ is defined by (4.5). Moreover, denote by $\lambda_t^{n,i} := \Psi(S_{t-}^{n,i}, \gamma_t^n)$ and $\overline{\lambda}_t^i := \Psi(\overline{S}_{t-}^i, \overline{\gamma}(t))$ the respective intensities of $N^{n,i}$ and $\overline{N}^i$.

Notice that the coupling above is based on the sharing of a common past $(N_-^i)_{i\geq 1}$ and a common underlying randomness, that are the $\mathbb{F}$-Poisson measures $(\Pi^i(dt', dx))_{i\geq 1}$, thanks to the thinning procedure. Note that the sequence of ADHPs is indexed by the size of the network n whereas the solutions of the limit equation which represent the behaviour under the mean field approximation are not.

The proof of the convergence (as $n \to +\infty$) of the empirical measure $\overline{\mu}_{S_t}^n := \frac{1}{n} \sum_{i=1}^n \delta_{S_t^{n,i}}$ towards the law of $\overline{S}_t^1$, for all $t \geq 0$, given in Chapter 3 relies on the following estimates (Corollaries 3.4.3 and 3.4.5): for all $i = 1, \dots, n$ and $\theta > 0$,

$$\mathbb{E}\left[\sup_{t\in[0,\theta]} |S_{t-}^{n,i} - \overline{S}_{t-}^i|\right] \lesssim_\theta \mathbb{P}\left(\left(S_{t-}^{n,i}\right)_{t\in[0,\theta]} \neq \left(\overline{S}_{t-}^i\right)_{t\in[0,\theta]}\right) \lesssim_\theta n^{-1/2}. \tag{4.8}$$

As a straight follow-up to the convergence of the empirical measure $\overline{\mu}_{S_t}^n$, we are interested in the dynamics of the fluctuations of this empirical measure around its limit. For any $t \geq 0$, let P_t denote the law of both $\overline{S}_t^1$ and $\overline{S}_{t-}^1$ (they have the same law since they are equal almost surely). Notice that P_t admits the density $u(t, \cdot)$ with respect to the Lebesgue measure, where u is the unique solution of (3.23) according to Proposition 3.3.8, thus

$$\langle P_t, \varphi \rangle = \int_0^{+\infty} \varphi(s) u(t, s) ds.$$

The analysis of the coupling (Equation (4.8)) gives a rate of convergence in, at least, $n^{-1/2}$ so we want to find the limit law of the fluctuation process defined, for all $t \geq 0$, by

$$\eta_t^n := \sqrt{n}\left(\overline{\mu}_{S_t}^n - P_t\right). \tag{4.9}$$

Notice that η_t^n is a distribution in the functional analysis sense on the state space of the ages, i.e. $\mathbb{R}_+$, and is devoted to be considered as a linear form acting on test functions φ by means of $\langle \eta_t^n, \varphi \rangle$.

4.3 Estimates in total variation norm

The bound $(n^{-1/2})$ on the rate of convergence, given by (4.8), is not sufficient in order to prove convergence or even tightness of the fluctuation process η^n. Some refined estimates are necessary. For instance, when dealing with diffusions, one looks for higher order moment estimates on the difference between the particles driven by the real dynamics and the limit particles (see [50, 82, 97, 101] for instance). Here, we deal with pure jump processes and, up to our knowledge, there is no reason why one could obtain better rates for higher order moments. A simple way to catch this fact is by looking at the coupling between the counting processes. Indeed, the difference between two counting processes, say $\delta_t^{n,i} = |N_t^{n,i} - \overline{N}_t^i|$, takes value in $\mathbb{N}$ so that for all $p \geq 1$, $(\delta_t^{n,i})^p \geq \delta_t^{n,i}$, and the moment of order p is greater than the moment of order one.

In order to accommodate this fact, the key idea is to estimate the coupling (4.6)-(4.7) in the total variation distance. Hence, the estimates needed in the next section (and proved in the present section) are the analogous of higher order moments but with respect to the total variation norm, i.e. the probabilities

$$\begin{aligned} \chi_n^{(k)}(\theta) &:= \mathbb{P}\left((S_{t-}^{n,k'})_{t\in[0,\theta]} \neq (\overline{S}_{t-}^{k'})_{t\in[0,\theta]} \text{ for every } k' = 1, ..., k\right) \\ &= \mathbb{P}\left((S_t^{n,k'})_{t\in[0,\theta]} \neq (\overline{S}_t^{k'})_{t\in[0,\theta]} \text{ for every } k' = 1, ..., k\right), \end{aligned} \tag{4.10}$$

for all positive integer k and real number $\theta \geq 0$.

The heuristics underlying the result stated below, in Proposition 4.3.1, relies on the asymptotic independence between the k age processes $(S_{t-}^{n,k'})_{t\in[0,\theta]}$, $k' = 1, ..., k$. Indeed, if they were independent then we would have (remind (4.8)),

$$\chi_n^{(k)}(\theta) = \prod_{k'=1}^{k} \mathbb{P}\big((S_{t-}^{n,k'})_{t\in[0,\theta]} \neq (\overline{S}_{t-}^{k'})_{t\in[0,\theta]}\big) = (\chi_n^{(1)}(\theta))^k \lesssim_\theta n^{-k/2},$$

which is exactly the rate of convergence we find below.

Proposition 4.3.1. *Under Assumption ($\mathcal{A}_{LLN}$),*

$$\chi_n^{(k)}(\theta) \lesssim_{(\theta,k)} n^{-k/2} \quad \text{and} \quad \xi_n^{(k)}(t) := \mathbb{E}\left[|\gamma_t^n - \overline{\gamma}(t)|^k\right] \lesssim_{(t,k)} n^{-k/2}.$$

Remark 4.3.2. *In addition to the explanation given in the beginning of this section, let us mention that the analogous to the higher moment estimates obtained for diffusions is obtained here for the difference between γ_t^n and $\overline{\gamma}(t)$. Indeed, as k grows, the convergence of $\xi_n^{(k)}(t)$ quickens. However, this gain in the rate of convergence does not apply when looking at the difference between the ages $S_t^{n,1}$ and $\overline{S}_t^1$ or the difference between the intensities $\lambda_t^{n,1}$ and $\overline{\lambda}_t^1$ (except if Ψ does not depend on s).*

Proof. Denote by $A \triangle B$ the symmetric difference of the sets A and B. Then, for any $i \leq n$, let us define $\Delta^{n,i} := N^{n,i} \Delta \overline{N}^i$ that is the set of points that are not common to $N^{n,i}$ and $\overline{N}^i$. From (4.6)-(4.7), one has

$$\Delta_t^{n,i} = \int_0^t \int_0^{+\infty} \mathbb{1}_{\left\{x \in [[\lambda_{t'}^{n,i}, \overline{\lambda}_{t'}^i]]\right\}} \Pi^i(dt', dx),$$

where $[[\lambda^{n,i}_{t'}, \overline{\lambda}^i_{t'}]]$ is the non empty interval which is either $[\lambda^{n,i}_{t'}, \overline{\lambda}^i_{t'}]$ or $[\overline{\lambda}^i_{t'}, \lambda^{n,i}_{t'}]$. Then, the intensity of the point process $\Delta^{n,i}$ is given by $\lambda^{\Delta,n,i}_t := |\lambda^{n,i}_t - \overline{\lambda}^i_t|$.

Note that, for all $n \geq 1$ and $i = 1, \ldots, n$, $N^{n,i}$ and $\overline{N}^i$ coincide on the non-positive part, i.e. $N^{n,i}_- = \overline{N}^i_-$. Therefore, $S^{n,i}_{0-} = \overline{S}^i_{0-}$ so that the equality between the processes $(S^{n,1}_{t-})_{t\in[0,\theta]}$ and $(\overline{S}^1_{t-})_{t\in[0,\theta]}$ is equivalent to $\Delta^{n,1}_{\theta-} = 0$. In particular, one has

$$\chi^{(k)}_n(\theta) \leq \mathbb{E}\left[\prod_{i=1}^k \Delta^{n,i}_{\theta-}\right], \tag{4.11}$$

since counting processes take value in $\mathbb{N}$. For any positive integers k and p, let us denote, for all $n \geq k$,

$$\varepsilon^{(k,p)}_n(\theta) := \mathbb{E}\left[\prod_{i=1}^k \left(\Delta^{n,i}_{\theta-}\right)^p\right].$$

Let us show, by induction on k, that

$$\varepsilon^{(k,p)}_n(\theta) \lesssim_{(\theta,k,p)} n^{-k/2} \tag{4.12}$$

which will end the proof thanks to (4.11). First, note that the case $k = 1$ and $p = 1$ is already treated in Chapter 3. Indeed, Theorem 3.4.1 gives

$$\varepsilon^{(1,1)}_n(\theta) = \int_0^\theta \mathbb{E}\left[|\lambda^{n,1}_t - \overline{\lambda}^1_t|\right] dt \lesssim_\theta n^{-1/2}. \tag{4.13}$$

Then, note that for any two positive integers p and q,

$$\varepsilon^{(k,p)}_n(\theta) \leq \varepsilon^{(k,q)}_n(\theta) \text{ as soon as } p \leq q. \tag{4.14}$$

This is due to the fact that counting processes take value in $\mathbb{N}$. The rest of the proof is divided in two steps: initialization and inductive step.

Step one. For $k = 1$ and p a positive integer, it holds that

$$(\Delta^{n,1}_{\theta-})^p = \sum_{p'=0}^{p-1} \binom{p}{p'} \int_0^{\theta-} (\Delta^{n,1}_{t-})^{p'} \Delta^{n,1}(dt). \tag{4.15}$$

Indeed, each time the process $(\Delta^{n,1}_t)_{t\geq 0}$ jumps (from $\Delta^{n,1}_{t-}$ to $\Delta^{n,1}_{t-} + 1$) then $(\Delta^{n,1}_{t-})^p$ jumps from $(\Delta^{n,1}_{t-})^p$ to $(\Delta^{n,1}_{t-} + 1)^p$ so the infinitesimal variation is

$$(\Delta^{n,1}_{t-} + 1)^p - (\Delta^{n,1}_{t-})^p = \sum_{p'=0}^{p-1} \binom{p}{p'} (\Delta^{n,1}_{t-})^{p'}.$$

The right-hand side of (4.15) involves integrals of predictable processes, that are the $(\Delta^{n,1}_{t-})^{p'}$, with respect to a point measure under which it is convenient to take expectation.

More precisely, since $(\Delta^{n,1}_{t-})^{p'} \leq (\Delta^{n,1}_{t-})^p$ as soon as $0 < p' \leq p - 1$, it holds that

$$\begin{aligned}
\varepsilon^{(1,p)}_n(\theta) = \mathbb{E}\left[(\Delta^{n,1}_{\theta-})^p\right] &\leq \mathbb{E}\left[\int_0^\theta \Delta^{n,1}(dt)\right] + 2^p \mathbb{E}\left[\int_0^\theta (\Delta^{n,1}_{t-})^p \Delta^{n,1}(dt)\right]. \\
&\leq \varepsilon^{(1,1)}_n(\theta) + 2^p \int_0^\theta \mathbb{E}\left[(\Delta^{n,1}_{t-})^p \lambda^{\Delta,n,1}_t\right] dt.
\end{aligned} \tag{4.16}$$

Yet the intensity $\lambda_t^{\Delta,n,1}$ is bounded by $||\Psi||_\infty$ and $\varepsilon_n^{(1,1)}(\theta) \lesssim_\theta n^{-1/2}$, see (4.13), so

$$\varepsilon_n^{(1,p)}(\theta) \lesssim_{(\theta,p)} n^{-1/2} + \int_0^\theta \varepsilon_n^{(1,p)}(t)dt,$$

and Lemma 4.7.1 gives $\varepsilon_n^{(1,p)}(\theta) \lesssim_{(\theta,p)} n^{-1/2}$.

Step two. For all integers $k \geq 2$ and $p \geq 1$, one can generalize the argument given to prove (4.15). Indeed, for any j in $\{1, ..., k\}$, each time the process $(\Delta_t^{n,j})_{t\geq 0}$ jumps (from $\Delta_{t-}^{n,j}$ to $\Delta_{t-}^{n,j}+1$) then $\prod_{i=1}^k (\Delta_{t-}^{n,i})^p$ jumps, almost surely, from $(\Delta_{t-}^{n,j})^p \prod_{i\neq j, i=1}^k (\Delta_{t-}^{n,i})^p$ to $(\Delta_{t-}^{n,1}+1)^p \prod_{i\neq j, i=1}^k (\Delta_{t-}^{n,i})^p$ since there are almost surely no common jumps between any two of the point processes $\Delta^{n,i}$. So the infinitesimal variation of $(\prod_{i=1}^k (\Delta_{t-}^{n,i})^p)_{t\geq 0}$ is a.s.

$$\sum_{p'=0}^{p-1} \binom{p}{p'} \prod_{i\neq j, i=1}^k (\Delta_{t-}^{n,i})^p (\Delta_{t-}^{n,j})^{p'}$$

as soon as t is a jumping time of $(\Delta_t^{n,j})_{t\geq 0}$. It then follows that

$$\prod_{i=1}^k (\Delta_{\theta-}^{n,i})^p = \sum_{j=1}^k \sum_{p'=0}^{p-1} \binom{p}{p'} \int_0^{\theta-} \prod_{i\neq j, i=1}^k (\Delta_{t-}^{n,i})^p (\Delta_{t-}^{n,j})^{p'} \Delta^{n,j}(dt), \quad \text{almost surely.}$$

Hence, thanks to the exchangeability of the processes $(\Delta^{n,i})_{i=1,...,n}$ and the predictability of the integrated processes, we have

$$\begin{aligned}
\varepsilon_n^{(k,p)}(\theta) &= \sum_{j=1}^k \sum_{p'=0}^{p-1} \binom{p}{p'} \mathbb{E}\left[\int_0^\theta \prod_{i\neq j, i=1}^k (\Delta_{t-}^{n,i})^p (\Delta_{t-}^{n,j})^{p'} \Delta^{n,j}(dt)\right] \\
&= k \sum_{p'=0}^{p-1} \binom{p}{p'} \int_0^\theta \mathbb{E}\left[(\Delta_{t-}^{n,1})^{p'} \prod_{i=2}^k (\Delta_{t-}^{n,i})^p \lambda_t^{\Delta,n,1}\right] dt \\
&\leq k \int_0^\theta \mathbb{E}\left[\prod_{i=2}^k (\Delta_{t-}^{n,i})^p \lambda_t^{\Delta,n,1}\right] + 2^p \mathbb{E}\left[(\Delta_{t-}^{n,1})^p \prod_{i=2}^k (\Delta_{t-}^{n,i})^p \lambda_t^{\Delta,n,1}\right] dt, \quad (4.17)
\end{aligned}$$

where we used that $(\Delta_{t-}^{n,1})^{p'} \leq (\Delta_{t-}^{n,1})^p$ as soon as $0 < p' \leq p-1$.

On the one hand, using that $\lambda_t^{\Delta,n,1} \leq ||\Psi||_\infty$, the second expectation in (4.17) is bounded by $||\Psi||_\infty \varepsilon_n^{(k,p)}(t)$. On the other hand, we use $(\mathcal{A}_{y,\mathcal{C}^2}^\Psi)$ which gives the following bound on the intensity,

$$\lambda_t^{\Delta,n,1} \leq \mathrm{Lip}(\Psi)|\gamma_t^n - \overline{\gamma}(t)| + ||\Psi||_\infty \mathbb{1}_{S_{t-}^{n,1} \neq \overline{S}_{t-}^1} \leq \mathrm{Lip}(\Psi)|\gamma_t^n - \overline{\gamma}(t)| + ||\Psi||_\infty (\Delta_{t-}^{n,1})^p.$$

Hence the first expectation in (4.17) is bounded by

$$\mathrm{Lip}(\Psi) D(t) + ||\Psi||_\infty \varepsilon_n^{(k,p)}(t), \quad (4.18)$$

with $D(t) := \mathbb{E}[\prod_{i=2}^k (\Delta_{t-}^{n,i})^p |\gamma_t^n - \overline{\gamma}(t)|]$. The second term of (4.18) is convenient to use a Gronwall-type lemma. To deal with the first term, we use a trick involving the exchangeability of the particles. Indeed, using the exchangeability we can replace each of the $k-1$ terms $(\Delta_{t-}^{n,i})^p$ in the expression of $D(t)$ by the following sum

$$\frac{1}{\lfloor \frac{n}{k} \rfloor} \sum_{j_i=(i-1)\lfloor \frac{n}{k} \rfloor + 1}^{i\lfloor \frac{n}{k} \rfloor} (\Delta_{t-}^{n,j_i})^p$$

without modifying the value of the expectation since the sums are taken on disjoined indices. Hence, using for the second line a generalization of Hölder's inequality with k exponents equal to $1/k$, we have

$$\begin{aligned} D(t) &\leq \mathbb{E}\left[\prod_{i=2}^{k}\left(\frac{1}{\lfloor\frac{n}{k}\rfloor}\sum_{j_i=(i-1)\lfloor\frac{n}{k}\rfloor+1}^{i\lfloor\frac{n}{k}\rfloor}(\Delta_{t-}^{n,j_i})^p\right)|\gamma_t^n-\overline{\gamma}(t)|\right] \\ &\leq \left(\prod_{i=2}^{k}\mathbb{E}\left[\left(\frac{1}{\lfloor\frac{n}{k}\rfloor}\sum_{j=1}^{\lfloor\frac{n}{k}\rfloor}(\Delta_{t-}^{n,j})^p\right)^k\right]^{1/k}\right)\xi_n^{(k)}(t)^{1/k} \leq E_{n,k,p}(t)^{\frac{k-1}{k}}\xi_n^{(k)}(t)^{1/k}, \quad (4.19) \end{aligned}$$

with $E_{n,k,p}(t) := \mathbb{E}[((1/\lfloor\frac{n}{k}\rfloor)\sum_{j=1}^{\lfloor\frac{n}{k}\rfloor}(\Delta_{t-}^{n,j})^p)^k]$. Yet, computations given in Sections 4.6.a) and 4.6.b) give the two following statements: there exists a constant $C(k)$ which does not depend on n or p such that

$$E_{n,k,p}(t) \leq C(k)\left(\sum_{k'=1}^{k-1} n^{k'-k}\varepsilon_n^{(k',pk)}(t) + \varepsilon_n^{(k,p)}(t)\right), \quad (4.20)$$

and $\xi_n^{(k)}(t)$ satisfy the following bound,

$$\xi_n^{(k)}(t) \lesssim_{(t,k)} n^{-k/2} + \sum_{k'=1}^{k-1} n^{k'-k}\varepsilon_n^{(k',k)}(t) + \varepsilon_n^{(k,1)}(t). \quad (4.21)$$

Then, using the induction hypothesis, that is for all $1 \leq k' \leq k-1$ and for all positive integer p, $\varepsilon_n^{(k',p)}(t) \lesssim_{(t,k,p)} n^{-k'/2}$, one has,

$$\begin{cases} E_{n,k,p}(t) \lesssim_{(t,k,p)} \sum_{k'=1}^{k-1} n^{k'-k}n^{-k'/2} + \varepsilon_n^{(k,p)}(t) \lesssim_{(t,k,p)} n^{-(k+1)/2} + \varepsilon_n^{(k,p)}(t) \\ \xi_n^{(k)}(t) \lesssim_{(t,k,p)} n^{-k/2} + \sum_{k'=1}^{k-1} n^{k'-k}n^{-k'/2} + \varepsilon_n^{(k,1)}(t) \lesssim_{(t,k,p)} n^{-k/2} + \varepsilon_n^{(k,1)}(t). \end{cases} \quad (4.22)$$

Gathering (4.17), (4.18), (4.19) and (4.22) gives (remind that $\varepsilon_n^{(k,1)}(t) \leq \varepsilon_n^{(k,p)}(t)$)

$$\varepsilon_n^{(k,p)}(\theta) \lesssim_{(\theta,k,p)} n^{-k/2} + \int_0^\theta \varepsilon_n^{(k,p)}(t)dt,$$

and so the Grönwall-type Lemma 4.7.1 gives $\varepsilon_n^{(k,p)}(\theta) \lesssim_{(\theta,k,p)} n^{-k/2}$ which ends the proof thanks to (4.11).

□

4.4 Tightness

The aim of this section is to prove tightness of the sequence of the laws of $(\eta^n)_{n\geq 1}$ regarded as stochastic processes (in time) with values in a suitable space of distributions. Thus, we consider $(\eta_t^n)_{t\geq 0}$ as a random process with values in the dual space of some well-chosen space of test functions. In Section 4.4.a), we give the definition of these spaces of test functions. Following the Hilbertian approach developed in [50], we work with weighted Sobolev Hilbert spaces. Finally, the tightness result is stated in Theorem 4.4.14.

The following study takes benefit of the Hilbert structure of the Sobolev spaces considered. This prompts us to recall the following Aldous tightness criterion for Hilbert space valued stochastic processes (cf. [81, p. 34-35]). Let H be a separable Hilbert space. A sequence of processes[4] $(X^n)_{n\geq 1}$ in $\mathcal{D}(\mathbb{R}_+, H)$ defined on the respective filtered[5] probability spaces $(\Omega^n, \mathcal{F}^n, (\mathcal{F}_t^n)_{t\geq 0}, \mathbb{P}^n)$ is tight if both conditions below hold true:

$(\mathtt{A}_1)$: for every $t \geq 0$ and $\varepsilon > 0$, there exists a compact set $K \subset H$ such that

$$\sup_{n\geq 1} \mathbb{P}^n \left(X_t^n \notin K \right) \leq \varepsilon,$$

$(\mathtt{A}_2)$: for every $\varepsilon_1, \varepsilon_2 > 0$ and $\theta \geq 0$, there exists $\delta_0 > 0$ and an integer n_0 such that for all $(\mathcal{F}_t^n)_{t\geq 0}$-stopping time $\tau_n \leq \theta$,

$$\sup_{n\geq n_0} \sup_{\delta\leq \delta_0} \mathbb{P}^n \left(||X^n_{\tau_n+\delta} - X^n_{\tau_n}||_H \geq \varepsilon_1 \right) \leq \varepsilon_2.$$

Note that $(\mathtt{A}_1)$ is implied by the condition $(\mathtt{A}_{1'})$ stated below which is much easier to ensue.

$(\mathtt{A}_{1'})$: There exists a Hilbert space H_0 such that $H_0 \hookrightarrow_K H$ and, for all $t \geq 0$,

$$\sup_{n\geq 1} \mathbb{E}^n[||X_t^n||^2_{H_0}] < +\infty,$$

where the notation $\hookrightarrow_K$ means that the embedding is compact and $\mathbb{E}^n$ denotes the expectation associated with the probability $\mathbb{P}^n$.

The fact that $(\mathtt{A}_{1'})$ implies $(\mathtt{A}_1)$ is easily checked: by compactness of the embedding, closed balls in H_0 are compact in H so, Markov's inequality gives $(\mathtt{A}_1)$.

4.4.a) Preliminaries on weighted Sobolev spaces

Here are listed some definitions and technical results about the weighted Sobolev spaces used in the present chapter. To avoid confusion, let us stress the fact that the test functions we use are supported in the state space of the ages, namely $\mathbb{R}_+$. For any integer k and any real α in $\mathbb{R}_+$, we denote by $\mathcal{W}_0^{k,\alpha} := \mathcal{W}_0^{k,\alpha}(\mathbb{R}_+)$ the completion of the set of compactly supported (in $\mathbb{R}_+$) functions of class $\mathcal{C}^\infty$ for the following norm

$$||f||_{k,\alpha} := \left(\sum_{k'=0}^{k} \int_{\mathbb{R}_+} \frac{|f^{(k')}(x)|^2}{1+|x|^{2\alpha}} dx \right)^{1/2},$$

where $f^{(j)}$ denotes the j^{th} derivative of f. Then, $\mathcal{W}_0^{k,\alpha}$ equipped with the norm $||\cdot||_{k,\alpha}$ is a separable Hilbert space and we denote $(\mathcal{W}_0^{-k,\alpha}, ||\cdot||_{-k,\alpha})$ its dual space. Notice that

$$\begin{cases} \text{if } k' \geq k, \text{ then } ||.||_{k,\alpha} \leq ||.||_{k',\alpha} \text{ and } ||.||_{-k',\alpha} \leq ||.||_{-k,\alpha}, \\ \text{if } \alpha' \geq \alpha, \text{ then } \mathcal{W}_0^{k,\alpha} \hookrightarrow \mathcal{W}_0^{k,\alpha'} \text{ and } \mathcal{W}_0^{-k,\alpha'} \hookrightarrow \mathcal{W}_0^{-k,\alpha}, \end{cases} \tag{4.23}$$

where the notation $\hookrightarrow$ means that the embedding is continuous.

[4] Complements about Hilbert space valued stochastic processes can be found in Appendix A.4.

[5] We assume here that $(\mathcal{F}_t^n)_{t\geq 0}$ is such that X^n is $(\mathcal{F}_t^n)_{t\geq 0}$-adapted.

Let $\mathcal{C}^{k,\alpha}$ be the space of functions f on $\mathbb{R}_+$ with continuous derivatives up to order k such that, for all $k' \leq k$, $\sup_{x\in\mathbb{R}_+} |f^{(k')}(x)|/(1+|x|^\alpha) < +\infty$. We equip this space with the norm

$$||f||_{\mathcal{C}^{k,\alpha}} := \sum_{k'=0}^{k} \sup_{x\in\mathbb{R}_+} \frac{|f^{(k')}(x)|}{1+|x|^\alpha}.$$

Recall that $\mathcal{C}_b^k$ is the space of bounded functions of class $\mathcal{C}^k$ with bounded derivatives of every order less than k. Notice that $\mathcal{C}_b^k = \mathcal{C}^{k,0}$ as normed spaces. Denote by $\mathcal{C}_b^{-k}$ its dual space. For any $\alpha > 1/2$ and any integer k (so that $\int_{\mathbb{R}_+} 1/(1+|x|^{2\alpha})dx < +\infty$), $\mathcal{C}_b^k \hookrightarrow \mathcal{W}_0^{k,\alpha}$, i.e. there exists a constant C such that

$$||\cdot||_{k,\alpha} \leq C||\cdot||_{\mathcal{C}_b^k}. \tag{4.24}$$

We recall the following Sobolev embeddings (see [50, Section 2.1.]):

(i) Sobolev embedding theorem: $\mathcal{W}_0^{m+k,\alpha} \hookrightarrow \mathcal{C}^{k,\alpha}$ for $m \geq 1$, $k \geq 0$ and α in $\mathbb{R}_+$, i.e. there exists a constant C such that

$$||f||_{\mathcal{C}^{k,\alpha}} \leq C||f||_{m+k,\alpha}. \tag{4.25}$$

(ii) Maurin's theorem: $\mathcal{W}_0^{m+k,\alpha} \hookrightarrow_{H.S.} \mathcal{W}_0^{k,\alpha+\beta}$ for $m \geq 1$, $k \geq 0$, α in $\mathbb{R}_+$ and $\beta > 1/2$, where $H.S.$ means that the embedding is of Hilbert-Schmidt type[6]. In particular, the embedding is compact and there exists a constant C such that

$$||f||_{k,\alpha+\beta} \leq C||f||_{k+m,\alpha}. \tag{4.26}$$

Hence, the following dual embeddings hold true:

$$\begin{cases} \mathcal{W}_0^{-k,\alpha} \hookrightarrow \mathcal{C}_b^{-k}, \text{ for } k \geq 0 \text{ and } \alpha > 1/2, \text{ (dual embedding of (4.24))} \\ \mathcal{W}_0^{-k,\alpha+\beta} \hookrightarrow_{H.S.} \mathcal{W}_0^{-(m+k),\alpha}, \text{ for } m \geq 1,\ k \geq 0,\ \alpha \text{ in } \mathbb{R}_+ \text{ and } \beta > 1/2. \end{cases} \tag{4.27}$$

In some of the proofs given in the next section, we consider an orthonormal basis $(\varphi_j)_{j\geq 1}$ of $\mathcal{W}_0^{k,\alpha}$ composed of $\mathcal{C}^\infty$ functions with compact support. The existence of such a basis follows from the fact that the functions of class $\mathcal{C}^\infty$ with compact support are dense in $\mathcal{W}_0^{k,\alpha}$. Furthermore, if $(\varphi_j)_{j\geq 1}$ is an orthonormal basis of $\mathcal{W}_0^{k,\alpha}$ and w belongs to $\mathcal{W}_0^{-k,\alpha}$, then $||w||_{-k,\alpha}^2 = \sum_{j\geq 1} \langle w, \varphi_j\rangle^2$ thanks to Parseval's identity. Let us precise that we stick with the notation $(\varphi_j)_{j\geq 1}$ even if the space $\mathcal{W}_0^{k,\alpha}$ (in particular the regularity k) may differ from page to page.

The two lemmas below are useful throughout the analysis.

Lemma 4.4.1. *For every test function φ in $\mathcal{W}_0^{2,\alpha}$, $||\varphi'||_{1,\alpha} \leq ||\varphi||_{2,\alpha}$. If f belongs to $\mathcal{C}_b^k$ for some $k \geq 1$ then, for any fixed α in $\mathbb{R}_+$, there exists a constant C such that for every test function φ in $\mathcal{W}_0^{k,\alpha}$, $||f\varphi||_{k,\alpha} \leq C||f||_{\mathcal{C}_b^k}||\varphi||_{k,\alpha}$.*

Proof. The first assertion follows from the definition of $||\cdot||_{2,\alpha}$, and the second one follows from Leibniz's rule and the definition of $||\cdot||_{k,\alpha}$. □

[6]Here, it means that $\sum_{j\geq 1} ||\varphi_j||_{k,\alpha+\beta}^2 < +\infty$ if $(\varphi_j)_{j\geq 1}$ is an orthonormal basis of $\mathcal{W}_0^{m+k,\alpha}$.

Lemma 4.4.2. *For any fixed α in $\mathbb{R}_+$ and x, y in $\mathbb{R}$, the mappings δ_x and $D_{x,y} : \mathcal{W}_0^{1,\alpha} \to \mathbb{R}$, defined by $\delta_x(\varphi) := \varphi(x)$ and $D_{x,y}(\varphi) := \varphi(x) - \varphi(y)$ are linear continuous. In particular, for all α in $\mathbb{R}_+$, there exist some positive constants C_1 and C_2 such that, if x and y are bounded by some constant M, i.e. $|x| \leq M$ and $|y| \leq M$, then*

$$\begin{cases} ||\delta_x||_{-2,\alpha} \leq ||\delta_x||_{-1,\alpha} \leq C_1(1 + M^\alpha), \\ ||D_{x,y}||_{-2,\alpha} \leq ||D_{x,y}||_{-1,\alpha} \leq C_2(1 + M^\alpha). \end{cases} \tag{4.28}$$

Proof. Remark that $|D_{x,y}(\varphi)| \leq |\varphi(x)| + |\varphi(y)| = |\delta_x(\varphi)| + |\delta_y(\varphi)|$. Hence, it suffices to show that there exists some positive constant C such that $||\delta_x||_{-1,\alpha} \leq C(1 + |x|^\alpha)$. Yet, $|\delta_x(\varphi)| = |\varphi(x)| \leq ||\varphi||_{\mathcal{C}^{0,\alpha}}(1 + |x|^\alpha) \leq C||\varphi||_{1,\alpha}(1 + |x|^\alpha)$ by (4.25). □

Remark 4.4.3. *At this point, let us mention two reasons why weighted Sobolev spaces are more appropriate than standard (non-weighted) Sobolev spaces of functions on $\mathbb{R}_+$:*

- *we want to be able to consider functions of $\mathcal{C}_b^k$ as test functions: indeed, Ψ must be considered as a test function in Equation (4.55) below, yet we do not want Ψ to be compactly supported with respect to the age s or even to rapidly decrease when s goes to infinity. The natural space to which Ψ belongs is some $\mathcal{C}_b^k$ space,*
- *in order to ensue criterion ($\mathcal{A}_{1'}$), a compact embedding is required but Maurin's theorem does not apply for standard Sobolev spaces on $\mathbb{R}_+$ (see [2, Theorem 6.37]).*

To satisfy the first point in the remark above, the weight α is assumed to be (strictly) greater than $1/2$ in all the next sections so that (4.24) holds true.

4.4.b) Decomposition of the fluctuations

Here, we give a semi-martingale representation of η^n used to simplify the study of tightness. Let us denote R (for *reset*) the linear mapping defined by $R\varphi := \varphi(0) - \varphi(\cdot)$ where φ is some test function. This mapping naturally appears in our problem since the age process jumps to the value 0 at each point of the underlying point process. The announced decomposition of the dynamics of $(\eta_t^n)_{t\geq 0}$ is described in the Proposition below.

Proposition 4.4.4. *Under Assumption ($\mathcal{A}_{LLN}$), for every test function φ in $\mathcal{C}_b^1$ and $t \geq 0$,*

$$\langle \eta_t^n, \varphi\rangle - \langle \eta_0^n, \varphi\rangle = \int_0^t \big(\langle \eta_z^n, L_z\varphi\rangle + A_z^n(\varphi)\big)dz + M_t^n(\varphi), \tag{4.29}$$

with $L_z\varphi(s) = \varphi'(s) + \Psi(s, \overline{\gamma}(z))R\varphi(s)$ for all $z \geq 0$ and s in $\mathbb{R}$, where $\overline{\gamma}$ is defined by (4.5), and

$$\begin{cases} M_t^n(\varphi) := n^{-1/2} \displaystyle\sum_{i=1}^n \int_0^t R\varphi(S_{z-}^{n,i}) \left(N^{n,i}(dz) - \lambda_z^{n,i}dz\right), \\ A_z^n(\varphi) := n^{-1/2} \displaystyle\sum_{i=1}^n R\varphi(S_{z-}^{n,i}) \left(\lambda_z^{n,i} - \Psi(S_{z-}^{n,i}, \overline{\gamma}(z))\right). \end{cases} \tag{4.30}$$

Furthermore, for any φ in $\mathcal{C}_b^1$, $(M_t^n(\varphi))_{t\geq 0}$ is a real valued $\mathbb{F}$-martingale with angle bracket given by

$$< M^n(\varphi) >_t = \frac{1}{n} \sum_{i=1}^n \int_0^t R\varphi \left(S_{z-}^{n,i}\right)^2 \lambda_z^{n,i} dz. \tag{4.31}$$

Remark 4.4.5. *To avoid confusion, let us mention that (4.30) defines M_t^n and A_z^n as distributions acting on test functions. More precisely, we show below that they can be seen as distributions in $\mathcal{W}_0^{-2,\alpha}$ (Propositions 4.4.7 and 4.4.10). However, we do not use the notation for the dual action $\langle \cdot, \cdot \rangle$ to avoid tricky notation involving several angle brackets in (4.31) for instance.*

Proof. By definition of η^n (Equation (4.9)),

$$\langle \eta_t^n, \varphi \rangle - \langle \eta_0^n, \varphi \rangle = \sqrt{n} \left[\frac{1}{n} \sum_{i=1}^{n} \left(\left\langle \delta_{S_t^{n,i}}, \varphi \right\rangle - \left\langle \delta_{S_0^{n,i}}, \varphi \right\rangle \right) - \left(\langle P_t, \varphi \rangle - \langle P_0, \varphi \rangle \right) \right].$$

Since, for all $i = 1, \dots, n$, the age process $(S_t^{n,i})_{t \geq 0}$ is piece-wise continuous, increasing with rate 1 and jumps from $S_{t-}^{n,i}$ to 0 when $N_t^{n,i} - N_{t-}^{n,i} = 1$, we have

$$\left\langle \delta_{S_t^{n,i}}, \varphi \right\rangle - \left\langle \delta_{S_0^{n,i}}, \varphi \right\rangle = \int_0^t \varphi' \left(S_z^{n,i} \right) dz + \int_0^t R\varphi \left(S_{z-}^{n,i} \right) N^{n,i}(dz)$$

and so

$$\left\langle \overline{\mu}_{S_t}^n, \varphi \right\rangle - \left\langle \overline{\mu}_{S_0}^n, \varphi \right\rangle = \int_0^t \left\langle \overline{\mu}_{S_z}^n, \varphi' \right\rangle dz + \frac{1}{n} \sum_{i=1}^{n} \int_0^t R\varphi \left(S_{z-}^{n,i} \right) N^{n,i}(dz). \tag{4.32}$$

Now, we have in the same way

$$\left\langle \delta_{\overline{S}_t^1}, \varphi \right\rangle - \left\langle \delta_{\overline{S}_0^1}, \varphi \right\rangle = \int_0^t \varphi' \left(\overline{S}_z^1 \right) dz + \int_0^t R\varphi \left(\overline{S}_{z-}^1 \right) \overline{N}^1(dz)$$

and, by definition of $(P_t)_{t \geq 0}$,

$$\langle P_t, \varphi \rangle - \langle P_0, \varphi \rangle = \mathbb{E} \left[\int_0^t \varphi' \left(\overline{S}_z^1 \right) dz \right] + \mathbb{E} \left[\int_0^t R\varphi \left(\overline{S}_{z-}^1 \right) \overline{N}^1(dz) \right]. \tag{4.33}$$

Yet, since φ' is bounded, Fubini's theorem gives that $\mathbb{E}[\int_0^t \varphi'(\overline{S}_z^1) dz] = \int_0^t \langle P_z, \varphi' \rangle \, dz$. Moreover, remind that the intensity of $\overline{N}^1$ is $\overline{\lambda}_t^1 = \Psi(\overline{S}_{t-}^1, \overline{\gamma}(t))$. Yet, since φ and Ψ are bounded,

$$\mathbb{E} \left[\int_0^t \left| R\varphi \left(\overline{S}_{z-}^1 \right) \right| \overline{\lambda}_z^1 dz \right] < +\infty,$$

and so $\mathbb{E}[\int_0^t R\varphi(\overline{S}_{z-}^1) \overline{N}^1(dz)] = \mathbb{E}[\int_0^t R\varphi(\overline{S}_{z-}^1) \overline{\lambda}_z^1 dz]$ since $(\overline{S}_{t-}^1)_{t \geq 0}$ is a predictable process (see [16, II. T8]). Using once again Fubini's theorem, we end up with

$$\mathbb{E} \left[\int_0^t R\varphi \left(\overline{S}_{z-}^1 \right) \overline{N}^1(dz) \right] = \int_0^t \langle P_z, \Psi(\cdot, \overline{\gamma}(z)) R\varphi \rangle \, dz$$

and so (4.33) becomes

$$\langle P_t, \varphi \rangle - \langle P_0, \varphi \rangle = \int_0^t \langle P_z, \varphi' \rangle \, dz + \int_0^t \langle P_z, \Psi(\cdot, \overline{\gamma}(z)) R\varphi \rangle \, dz. \tag{4.34}$$

Gathering (4.32) and (4.34) gives

$$\begin{aligned} \langle \eta_t^n, \varphi \rangle - \langle \eta_0^n, \varphi \rangle = & \int_0^t \langle \eta_z^n, \varphi' \rangle \, dz \\ & + \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} \int_0^t R\varphi \left(S_{z-}^{n,i} \right) N^{n,i}(dz) - \int_0^t \langle P_z, \Psi(\cdot, \overline{\gamma}(z)) R\varphi \rangle \, dz \right) \end{aligned}$$

and so

$$\langle \eta_t^n, \varphi\rangle - \langle \eta_0^n, \varphi\rangle = \int_0^t \langle \eta_z^n, L_z\varphi\rangle\, dz + \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \int_0^t R\varphi\left(S_{z-}^{n,i}\right) N^{n,i}(dz) - \int_0^t \left\langle \overline{\mu}^n_{S_{z-}}, \Psi(\cdot, \overline{\gamma}(z))R\varphi\right\rangle dz\right), \quad (4.35)$$

where we used that, almost surely, $\overline{\mu}^n_{S_{z-}} = \overline{\mu}^n_{S_z}$ for almost every z in $\mathbb{R}_+$. Then, the second term in the right-hand side of (4.35) rewrites as $M_t^n(\varphi) + \int_0^t A_z^n(\varphi)dz$.

It remains to show that $(M_t^n(\varphi))_{t\geq 0}$ is an $\mathbb{F}$-martingale. Yet, for all $i = 1, \ldots, n$,

$$\mathbb{E}\left[\int_0^t \left|R\varphi\left(S_{z-}^{n,i}\right)\right| \lambda_z^{n,i} dz\right] \leq 2||\varphi||_\infty \mathbb{E}\left[\int_0^t \lambda_z^{n,i} dz\right] = 2||\varphi||_\infty \mathbb{E}\left[N_t^{n,i}\right] < +\infty,$$

and the $\mathbb{F}$-predictability of the age processes $(S_{t-}^{n,i})_{t\geq 0}$ gives the result (see [16, II. T8]). Finally, the expression of the angle bracket (4.31) follows from standard computations for point processes (see [60, Proposition II.4.1.]). □

4.4.c) Estimates in dual spaces

Below are stated estimates of the terms η^n, A^n and M^n appearing in (4.29) and regarded as distributions. More precisely, the estimates given in this section are stated in terms of the norm on either $\mathcal{W}_0^{-1,\alpha}$ or $\mathcal{W}_0^{-2,\alpha}$ for any $\alpha > 1/2$ (in comparison with $\mathcal{W}_0^{-2,2}$ and $\mathcal{W}_0^{-4,1}$ in [82] for instance). Usually, like in [50, 82, 97, 101], the weight is linked to the maximal order of the moment estimates obtained on the positions of the particles. However, the age processes are bounded in finite time (remind (4.1)) so the weight α of the Sobolev space can be taken as large as wanted. The weighted Sobolev spaces are nevertheless interesting here since, in particular, the distribution η_t^n belongs to $\mathcal{W}_0^{-1,\alpha}$ for all $t \geq 0$ (see Proposition 4.4.6 below). We refer to the introductory discussion in Section 4.1 for complements on the usefulness of the weights.

We first estimate η^n in the smaller space $\mathcal{W}_0^{-1,\alpha}$. This is later used in order to satisfy condition ($\mathtt{A}_{1'}$) of the Aldous type criterion stated on page 134.

Proposition 4.4.6. *Under Assumption ($\mathcal{A}_{LLN}$), for any $\alpha > 1/2$ and $\theta \geq 0$,*

$$\sup_{n\geq 1}\sup_{t\in[0,\theta]} \mathbb{E}\left[||\eta_t^n||^2_{-1,\alpha}\right] < +\infty. \quad (4.36)$$

Proof. Let $(\varphi_k)_{k\geq 1}$ be an orthonormal basis of $\mathcal{W}_0^{1,\alpha}$ so that, in particular $||\eta_t^n||^2_{-1,\alpha} = \sum_{k\geq 1}\langle \eta_t^n, \varphi_k\rangle^2$. Using the coupling (4.6)-(4.7), we have for every k and $t \leq \theta$,

$$\langle \eta_t^n, \varphi_k\rangle = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \varphi_k(S_t^{n,i}) - \mathbb{E}\left[\varphi_k(\overline{S}_t^i)\right]\right) = S_t^n(\varphi_k) + T_t^n(\varphi_k),$$

where

$$\begin{cases} S_t^n(\varphi_k) := n^{-1/2}\sum_{i=1}^n \varphi_k(S_t^{n,i}) - \varphi_k(\overline{S}_t^i) \\ T_t^n(\varphi_k) := n^{-1/2}\sum_{i=1}^n \varphi_k(\overline{S}_t^i) - \mathbb{E}[\varphi_k(\overline{S}_t^i)]. \end{cases}$$

Step one. Using the independence of the age processes $(\overline{S}_t^i)_{t\geq 0}$, we have

$$\begin{aligned}\mathbb{E}\left[\sum_{k\geq 1} T_t^n(\varphi_k)^2\right] &= \sum_{k\geq 1}\mathbb{E}\left[\frac{1}{n}\left(\sum_{i=1}^n \varphi_k(\overline{S}_t^i) - \mathbb{E}\left[\varphi_k(\overline{S}_t^i)\right]\right)^2\right] \\ &= \sum_{k\geq 1}\mathbb{E}\left[\left(\varphi_k(\overline{S}_t^1) - \mathbb{E}\left[\varphi_k(\overline{S}_t^1)\right]\right)^2\right] \leq \sum_{k\geq 1}\mathbb{E}\left[\left(\varphi_k(\overline{S}_t^1)\right)^2\right] \\ &\leq \mathbb{E}\left[\sum_{k\geq 1}(\delta_{\overline{S}_t^1}(\varphi_k))^2\right] = \mathbb{E}\left[||\delta_{\overline{S}_t^1}||_{-1,\alpha}^2\right].\end{aligned}$$

Then, using Lemma 4.4.2 and the fact that the age $\overline{S}_t^1$ is upper bounded by $M_{T_0}+t \leq M_{T_0}+\theta$ (thanks to $(\mathcal{A}_{u^{\text{in}},\infty}^{\zeta_{N-}})$, remind (4.1)), it follows that

$$\mathbb{E}\left[\sum_{k\geq 1} T_t^n(\varphi_k)^2\right] \leq (C_1)^2(1+(M_{T_0}+\theta)^\alpha)^2$$

and so $\sup_{n\geq 1}\sup_{t\in[0,\theta]}\mathbb{E}[\sum_{k\geq 1} T_t^n(\varphi_k)^2] < +\infty$.

Step two. Expanding the square and using exchangeability of the age processes $(S_t^{n,i})_{t\geq 0}$, one has

$$\begin{aligned}\mathbb{E}\left[\sum_{k\geq 1} S_t^n(\varphi_k)^2\right] &= n^{-1}\mathbb{E}\left[\sum_{k\geq 1}\left(\sum_{i=1}^n \varphi_k(S_t^{n,i}) - \varphi_k(\overline{S}_t^i)\right)^2\right] \\ &= (n-1)\mathbb{E}\left[\sum_{k\geq 1}(\varphi_k(S_t^{n,1}) - \varphi_k(\overline{S}_t^1))(\varphi_k(S_t^{n,2}) - \varphi_k(\overline{S}_t^2))\right] \\ &\quad + \mathbb{E}\left[\sum_{k\geq 1}(\varphi_k(S_t^{n,1}) - \varphi_k(\overline{S}_t^1))^2\right]. \qquad (4.37)\end{aligned}$$

Since the ages $S_t^{n,1}$, $\overline{S}_t^1$, $S_t^{n,2}$ and $\overline{S}_t^2$ are upper bounded by $M_{T_0}+\theta$ and $(\varphi_k(x_1)-\varphi_k(x_2))(\varphi_k(y_1)-\varphi_k(y_2)) = 0$ as soon as $x_1 = x_2$ or $y_1 = y_2$, we have

$$\mathbb{E}\left[\sum_{k\geq 1}(\varphi_k(S_t^{n,1}) - \varphi_k(\overline{S}_t^1))(\varphi_k(S_t^{n,2}) - \varphi_k(\overline{S}_t^2))\right] \leq \chi_n^{(2)}(\theta)\sup_{x,y\leq M_{T_0}+\theta}\sum_{k\geq 1}|\varphi_k(x)-\varphi_k(y)|^2, \qquad (4.38)$$

where $\chi_n^{(2)}(\theta)$ is defined by (4.10). Yet, since $(\varphi_k)_{k\geq 1}$ is an orthonormal basis of $\mathcal{W}_0^{1,\alpha}$, we have

$$\begin{cases}\sum_{k\geq 1}(\varphi_k(S_t^{n,1}) - \varphi_k(\overline{S}_t^1))^2 = \sum_{k\geq 1}\left\langle D_{S_t^{n,1},\overline{S}_t^1}, \varphi_k\right\rangle^2 = ||D_{S_t^{n,1},\overline{S}_t^1}||_{-1,\alpha}^2 \\ \sup_{x,y\leq M_{T_0}+\theta}\sum_{k\geq 1}|\varphi_k(x)-\varphi_k(y)|^2 = \sup_{x,y\leq M_{T_0}+\theta}||D_{x,y}||_{-1,\alpha}^2.\end{cases} \qquad (4.39)$$

Hence, using Lemma 4.4.2 and once again the fact that the ages $S_t^{n,1}$ and $\overline{S}_t^1$ are upper bounded by $M_{T_0}+\theta$, we have, by gathering (4.37)-(4.39),

$$\mathbb{E}\left[\sum_{k\geq 1} S_t^n(\varphi_k)^2\right] \leq (n-1)\chi_n^{(2)}(\theta)(C_2)^2(1+(M_{T_0}+\theta)^\alpha)^2 + (C_2)^2(1+(M_{T_0}+\theta)^\alpha)^2,$$

and it follows from Proposition 4.3.1 that $\sup_{n\geq 1}\sup_{t\in[0,\theta]}\mathbb{E}[\sum_{k\geq 1}S_t^n(\varphi_k)^2]<+\infty$.

Finally, by convexity of the square function, $||\eta_t^n||_{-1,\alpha}^2\leq 2\sum_{k\geq 1}S_t^n(\varphi_k)^2+T_t^n(\varphi_k)^2$ so that (4.36) follows from the two steps above. □

Proposition 4.4.7. *Under Assumption ($\mathcal{A}_{LLN}$), for any integer n and any real $\alpha>1/2$, the process $(M_t^n)_{t\geq 0}$, defined by (4.30), is an $\mathbb{F}$-martingale which belongs to $\mathcal{D}(\mathbb{R}_+,\mathcal{W}_0^{-1,\alpha})$ almost surely. Furthermore, for any $\theta\geq 0$,*

$$\sup_{n\geq 1}\mathbb{E}\left[\sup_{t\in[0,\theta]}||M_t^n||_{-1,\alpha}^2\right]<+\infty. \tag{4.40}$$

Remark 4.4.8. *In comparison with Proposition 4.4.7, we do not prove that $t\mapsto\eta_t^n$ belongs to $\mathcal{D}(\mathbb{R}_+,\mathcal{W}_0^{-1,\alpha})$ almost surely in Proposition 4.4.6. For our purpose, it is sufficient to prove that $t\mapsto\eta^n$ belongs to the bigger space $\mathcal{D}(\mathbb{R}_+,\mathcal{W}_0^{-2,\alpha})$. The corresponding tightness property is proved below in Proposition 4.4.13.*

Proof. We first show (4.40) and then use it in order to prove that $(M_t^n)_{t\geq 0}$ is càdlàg. Let $(\varphi_k)_{k\geq 1}$ be an orthonormal basis of $\mathcal{W}_0^{1,\alpha}$ composed of $\mathcal{C}^\infty$ functions with compact support. For all $k\geq 1$, the test function φ_k belongs to $\mathcal{C}_b^1$ so that $(M_t^n(\varphi_k))_{t\geq 0}$ is an $\mathbb{F}$-martingale (Proposition 4.4.4). Using Doob's inequality for real-valued martingales [77, Theorem 1.43.] and Equation (4.31), one has

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,\theta]}||M_t^n||_{-1,\alpha}^2\right]&\leq\sum_{k\geq 1}\mathbb{E}\left[\sup_{t\in[0,\theta]}M_t^n(\varphi_k)^2\right]\\&\leq C\sum_{k\geq 1}\mathbb{E}\left[M_\theta^n(\varphi_k)^2\right]\leq C||\Psi||_\infty\mathbb{E}\left[\int_0^\theta\sum_{k\geq 1}R\varphi_k(S_{z-}^{n,1})^2dz\right],\end{aligned}$$

where the last inequality comes from exchangeability and boundedness of the intensity. Noticing that $R\varphi_k(S_{z-}^{n,1})=D_{0,S_{z-}^{n,1}}(\varphi_k)$ and then using Lemma 4.4.2 as we have done in the proof of Proposition 4.4.6, it follows that

$$\mathbb{E}\left[\int_0^\theta\sum_{k\geq 1}R\varphi_k(S_{z-}^{n,1})^2dz\right]\leq(C_2)^2\int_0^\theta(1+(M_{T_0}+\theta)^\alpha)^2dz,$$

which does not depend on n and gives (4.40). Moreover, gathering the integrability property given by (4.40) and the fact that, for all $k\geq 1$, the process $(M_t^n(\varphi_k))_{t\geq 0}$ is an $\mathbb{F}$-martingale, we have that M^n is a $\mathcal{W}_0^{-1,\alpha}$-valued $\mathbb{F}$-martingale (see Appendix A.4 for more insights on Hilbert space valued martingales).

It remains to show that $(M_t^n)_{t\geq 0}$ is càdlàg. First remark that for any k, the $\mathbb{F}$-martingale $(M_t^n(\varphi_k))_{t\geq 0}$ is càdlàg. Let $\varepsilon>0$ and $t_0>0$. For any $n\geq 1$,

$$\mathbb{E}\left[\sum_{k\geq 1}\sup_{t\in[0,t_0+1]}M_t^n(\varphi_k)^2\right]<+\infty,$$

so there exists a set Ω^n such that $\mathbb{P}(\Omega^n)=1$ and for all ω in Ω^n,

$$\sum_{k\geq 1}\sup_{t\in[0,t_0+1]}\langle M_t^n(\omega),\varphi_k\rangle^2<+\infty.$$

Once ω is fixed in Ω^n, there exists an integer k_0 (which depends on ω) such that $\sum_{k>k_0} \sup_{t\in[0,t_0+1]} \langle M_t^n(\omega), \varphi_k\rangle^2 < \varepsilon$. Let t be such that $t_0 < t \leq t_0 + 1$, using the right continuity of $t \mapsto \langle M_t^n(\omega), \varphi_k\rangle$, we have, dropping ω for simplicity of notations,

$$\begin{aligned} ||M_t^n - M_{t_0}^n||_{-1,\alpha}^2 &= \sum_{k\geq 1} (M_t^n(\varphi_k) - M_{t_0}^n(\varphi_k))^2 \\ &\leq \sum_{k=1}^{k_0} (M_t^n(\varphi_k) - M_{t_0}^n(\varphi_k))^2 + 2\sum_{k>k_0} [M_t^n(\varphi_k)^2 + M_{t_0}^n(\varphi_k)^2] \\ &\leq \sum_{k=1}^{k_0} \varepsilon + 4\varepsilon = (k_0+4)\varepsilon, \end{aligned}$$

as soon as $|t - t_0|$ is small enough. Hence, $t \mapsto M_t^n(\omega)$ is right continuous with values in $\mathcal{W}_0^{-1,\alpha}$. In the same way, let $(t_m)_{m\geq 1}$ be a sequence such that $t_m < t_0$ and $t_m \to t_0$. For any integers m and ℓ, we have, dropping ω for simplicity of notations,

$$\begin{aligned} ||M_{t_m}^n - M_{t_\ell}^n||_{-1,\alpha}^2 &= \sum_{k\geq 1} (M_{t_m}^n(\varphi_k) - M_{t_\ell}^n(\varphi_k))^2 \\ &\leq \sum_{k=1}^{k_0} (M_{t_m}^n(\varphi_k) - M_{t_\ell}^n(\varphi_k))^2 + 4\varepsilon. \end{aligned}$$

Yet, for all $k = 1, \ldots, k_0$, the sequence $(M_{t_m}^n(\varphi_k))_{m\geq 1}$ is convergent hence Cauchy. It follows that $(M_{t_m}^n(\omega))_{m\geq 1}$ is a Cauchy sequence and so converges in $\mathcal{W}_0^{-1,\alpha}$. Hence, $t \mapsto M_t^n(\omega)$ admits left limits in $\mathcal{W}_0^{-1,\alpha}$. Finally, $t \mapsto M_t^n$ belongs to $\mathcal{D}(\mathbb{R}_+, \mathcal{W}_0^{-1,\alpha})$ almost surely. □

A bound for the linear operator L_z appearing in Equation (4.29) is given in the following proposition by means of Lemma 4.4.1. Hence, let us note here that, under Assumption $(\mathcal{A}^{\Psi}_{s,\mathcal{C}_b^2})$, the functions

$$t \mapsto ||\Psi(\cdot, \overline{\gamma}(t))||_{\mathcal{C}_b^2} \text{ and } t \mapsto \left\| \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(t)) \right\|_{\mathcal{C}_b^1} \text{ are locally bounded,} \tag{4.41}$$

since $t \mapsto \overline{\gamma}(t)$ is locally bounded. In the same way, under Assumption $(\mathcal{A}^{\Psi}_{s,\mathcal{C}_b^4})$, the function

$$t \mapsto ||\Psi(\cdot, \overline{\gamma}(t))||_{\mathcal{C}_b^4} \text{ is locally bounded.} \tag{4.42}$$

Proposition 4.4.9. *Under $(\mathcal{A}^{\Psi}_{s,\mathcal{C}_b^2})$, for any z in $\mathbb{R}_+$, the application L_z defined in Proposition 4.4.4 is a linear continuous mapping from $\mathcal{W}_0^{2,\alpha}$ to $\mathcal{W}_0^{1,\alpha}$ for any $\alpha > 1/2$. Moreover, for all φ in $\mathcal{W}_0^{2,\alpha}$ and $\theta \geq 0$,*

$$\sup_{z\in[0,\theta]} \frac{||L_z\varphi||_{1,\alpha}^2}{||\varphi||_{2,\alpha}^2} < +\infty. \tag{4.43}$$

Proof. By definition of L_z and the triangular inequality,

$$||L_z\varphi||_{1,\alpha}^2 \leq 2(||\varphi'||_{1,\alpha}^2 + ||\Psi(\cdot, \overline{\gamma}(z))R\varphi||_{1,\alpha}^2).$$

Firstly, $||\varphi'||_{1,\alpha}^2 \leq ||\varphi||_{2,\alpha}^2$. Secondly, by Lemma 4.4.1, for all $z \leq \theta$,

$$||\Psi(\cdot, \overline{\gamma}(z))R\varphi||_{1,\alpha}^2 \leq C \sup_{z\in[0,\theta]} ||\Psi(\cdot, \overline{\gamma}(z))||_{\mathcal{C}_b^1}^2 ||R\varphi||_{1,\alpha}^2.$$

Using the convexity of the square function, it appears that

$$||R\varphi||^2_{1,\alpha} \leq 2\int_{\mathbb{R}_+} \frac{1}{1+|x|^{2\alpha}}dx|\varphi(0)|^2 + 2||\varphi||^2_{1,\alpha}.$$

Yet, by (4.25), $|\varphi(0)| \leq ||\varphi||_{\mathcal{C}^{0,\alpha}} \leq C||\varphi||_{1,\alpha}$ so that, for any fixed $\alpha > 1/2$, $||R\varphi||^2_{1,\alpha} \leq C||\varphi||^2_{1,\alpha}$ and so, by using (4.23), we have $||R\varphi||^2_{1,\alpha} \leq C||\varphi||^2_{2,\alpha}$. Finally, (4.41) gives (4.43). □

Before giving estimates for A^n, let us precise the expansion we use in the proof. Namely, using that $\lambda^{n,i}_t = \Psi(S^{n,i}_{t-}, \gamma^n_t)$ and $(\mathcal{A}^{\Psi}_{y,\mathcal{C}^2})$, it follows from Taylor's inequality that for φ in $\mathcal{W}^{2,\alpha}_0$,

$$A^n_t(\varphi) = \frac{1}{n}\sum_{i=1}^n R\varphi(S^{n,i}_{t-})\frac{\partial\Psi}{\partial y}(S^{n,i}_{t-}, \overline{\gamma}(t))\left(\sqrt{n}(\gamma^n_t - \overline{\gamma}(t)) + \sqrt{n}r^{n,i}_t\right), \tag{4.44}$$

with the rests satisfying $|r^{n,i}_t| \leq \sup_{s,y}|\frac{\partial^2\Psi}{\partial y^2}(s,y)||\gamma^n_t - \overline{\gamma}(t)|^2/2$. This upper-bound does not depend on φ. Let us denote $\Gamma^n_{t-} := \sqrt{n}(\gamma^n_t - \overline{\gamma}(t))$ and

$$R^{n,(1)}_t(\varphi) := \frac{1}{n}\sum_{i=1}^n \left(R\varphi(S^{n,i}_{t-})\frac{\partial\Psi}{\partial y}(S^{n,i}_{t-}, \overline{\gamma}(t))\sqrt{n}r^{n,i}_t\right),$$

so that (4.44) rewrites as

$$A^n_t(\varphi) = \left\langle \overline{\mu}^n_{S_t}, \frac{\partial\Psi}{\partial y}(\cdot, \overline{\gamma}(t))R\varphi \right\rangle \Gamma^n_{t-} + R^{n,(1)}_t(\varphi). \tag{4.45}$$

Proposition 4.4.10. *Under Assumption $(\mathcal{A}_{LLN})$, for any integer n and any real numbers t in $\mathbb{R}_+$ and $\alpha > 1/2$, the application A^n_t, defined by (4.30), is a linear continuous mapping from $\mathcal{W}^{2,\alpha}_0$ to $\mathbb{R}$ which satisfies, for any $\theta \geq 0$,*

$$\sup_{n\geq 1}\sup_{t\in[0,\theta]} \mathbb{E}\left[||A^n_t||^2_{-2,\alpha}\right] < +\infty. \tag{4.46}$$

Proof. Starting from (4.45), we have, by convexity of the square function,

$$A^n_t(\varphi)^2 \leq 2\Big(\left\langle \overline{\mu}^n_{S_t}, \frac{\partial\Psi}{\partial y}(\cdot, \overline{\gamma}(t))R\varphi \right\rangle^2 (\Gamma^n_{t-})^2 + R^{n,(1)}_t(\varphi)^2\Big)$$

Let $(\varphi_k)_{k\geq 1}$ be an orthonormal basis of $\mathcal{W}^{2,\alpha}_0$ so that $||A^n_t||^2_{-2,\alpha} = \sum_{k\geq 1} A^n_t(\varphi_k)^2$. Noticing that $R\varphi_k(S^{n,i}_{t-}) = D_{0,S^{n,i}_{t-}}(\varphi_k)$ and then using Lemma 4.4.2 as we have done in the proof of Proposition 4.4.6, it follows that

$$\begin{aligned}\sum_{k\geq 1}\left\langle \overline{\mu}^n_{S_t}, \frac{\partial\Psi}{\partial y}(\cdot, \overline{\gamma}(t))R\varphi_k \right\rangle^2 &\leq \mathrm{Lip}(\Psi)^2\frac{1}{n}\sum_{i=1}^n\left(\sum_{k\geq 1} R\varphi_k(S^{n,i}_{t-})^2\right)\\ &\leq \mathrm{Lip}(\Psi)^2(C_2)^2(1+(M_{T_0}+\theta)^\alpha)^2,\end{aligned}$$

and in the same way,

$$\sum_{k\geq 1} R^{n,(1)}_t(\varphi_k)^2 \leq \mathrm{Lip}(\Psi)^2(C_2)^2(1+(M_{T_0}+\theta)^\alpha)^2\frac{1}{n}\sum_{i=1}^n(\sqrt{n}r^{n,i}_t)^2.$$

Hence,

$$\sum_{k\geq 1} A_t^n(\varphi_k)^2 \leq 2\mathrm{Lip}(\Psi)^2(C_2)^2(1+(M_{T_0}+\theta)^\alpha)^2\left((\Gamma_{t-}^n)^2+\frac{1}{n}\sum_{i=1}^n(\sqrt{n}r_t^{n,i})^2\right).$$

Yet, as a consequence of Proposition 4.3.1, $\xi_n^{(2)}(t)=\mathbb{E}[|\Gamma_{t-}^n|^2]/n \lesssim_\theta n^{-1}$ and $\xi_n^{(4)}(t)=\mathbb{E}[|\Gamma_t^n|^4]/n^2 \lesssim_\theta n^{-2}$. In particular, uniformly in $t\leq\theta$, the L^1 norm of $(\Gamma_{t-}^n)^2$ is of order 1 while the L^1 norm of the rest term satisfies

$$\frac{1}{n}\sum_{i=1}^n(\sqrt{n}r_t^{n,i})^2 \leq n\Big(\sup_{s,y}|\frac{\partial^2\Psi}{\partial y^2}(s,y)|\Big)^2|\gamma_t^n-\overline{\gamma}(t)|^4/4$$

and so vanishes to 0 as n goes to infinity. Hence,

$$\sup_{n\geq 1}\sup_{t\in[0,\theta]}\mathbb{E}\left[||A_t^n||_{-2,\alpha}^2\right]=\sup_{n\geq 1}\sup_{t\in[0,\theta]}\mathbb{E}\left[\sum_{k\geq 1}A_t^n(\varphi_k)^2\right]<+\infty.$$

□

To prove tightness of $(\eta^n)_{n\geq 1}$ in $\mathcal{D}(\mathbb{R}_+,\mathcal{W}_0^{-2,\alpha})$, we use (as a consequence of (4.29)) the following decomposition in $\mathcal{W}_0^{-2,\alpha}$,

$$\eta_t^n-\eta_0^n=\int_0^t L_z^*\eta_z^n dz+\int_0^t A_z^n dz+M_t^n, \tag{4.47}$$

where L_z^* is the adjoint operator of L_z.

Remark 4.4.11. *As a corollary of Proposition 4.4.9, one has, for all $\alpha>1/2$, all w in $\mathcal{W}_0^{-1,\alpha}$ and all $\theta\geq 0$,*

$$\sup_{z\in[0,\theta]}\frac{||L_z^*w||_{-2,\alpha}^2}{||w||_{-1,\alpha}^2}<+\infty. \tag{4.48}$$

Indeed, both $||L_z^*w||_{-2,\alpha}^2\leq\sup_{||\varphi||_{2,\alpha}=1}||L_z\varphi||_{1,\alpha}^2||w||_{-1,\alpha}^2$ and Equation (4.43) give the result.

Furthermore, the Doob-Meyer process[7] $(\ll M^n\gg_t)_{t\geq 0}$ associated with the square integrable $\mathbb{F}$-martingale $(M_t^n)_{t\geq 0}$ satisfies the following: for any $t\geq 0$, $\ll M^n\gg_t$ is the linear continuous mapping from $\mathcal{W}_0^{2,\alpha}$ to $\mathcal{W}_0^{-2,\alpha}$ given, for all φ_1, φ_2 in $\mathcal{W}_0^{2,\alpha}$, by

$$\langle\ll M^n\gg_t(\varphi_1),\varphi_2\rangle=\frac{1}{n}\sum_{i=1}^n\int_0^t R\varphi_1(S_{z-}^{n,i})R\varphi_2(S_{z-}^{n,i})\lambda_z^{n,i}dz.$$

This last equation can be retrieved thanks to the polarization identity from (4.31). Moreover, to give sense to Equation (4.47), we need the lemma stated below.

Lemma 4.4.12. *Under ($\mathcal{A}_{TGN}$), the integrals $\int_0^t L_z^*\eta_z^n dz$ and $\int_0^t A_z^n dz$ are almost surely well defined as Bochner integrals[8] in $\mathcal{W}_0^{-2,\alpha}$ for any $\alpha>1/2$. In particular, the functions $t\mapsto\int_0^t L_z^*\eta_z^n dz$ and $t\mapsto\int_0^t A_z^n dz$ are almost surely strongly continuous in $\mathcal{W}_0^{-2,\alpha}$.*

[7] It is the generalization of the angle bracket for real valued martingales to Hilbert space valued martingales (see Appendix A.4).

[8] It is the generalization of Lebesgue integral to functions that take values in a Banach space, as the limit of integrals of simple functions. In particular, as for the Lebesgue integral, one has $||\int f||\leq\int||f||$.

Proof. Since $\mathcal{W}_0^{-2,\alpha}$ is separable, it suffices to verify that (see Yosida [162, p. 133]):

(i) for every φ in $\mathcal{W}_0^{2,\alpha}$, the functions $z \mapsto \langle L_z^* \eta_z^n, \varphi \rangle = \langle \eta_z^n, L_z \varphi \rangle$ and $z \mapsto A_z^n(\varphi)$ are measurable,

(ii) the integrals $\int_0^t ||L_z^* \eta_z^n||_{-2,\alpha} dz$ and $\int_0^t ||A_z^n||_{-2,\alpha} dz$ are finite almost surely.

The first condition is immediate. The second one follows from the controls we have shown.

Indeed, on the one hand, it follows from Equation (4.48) that $\int_0^t ||L_z^* \eta_z^n||_{-2,\alpha} dz \lesssim_t \int_0^t ||\eta_z^n||_{-1,\alpha} dz$ and Proposition 4.4.6 implies $\mathbb{E}[\int_0^t ||\eta_z^n||_{-1,\alpha+1} dz] < +\infty$ so that $\int_0^t ||L_z^* \eta_z^n||_{-2,\alpha} dz$ is finite a.s.

On the other hand, Proposition 4.4.10 gives that $\mathbb{E}[\int_0^t ||A_z^n||_{-2,\alpha} dz]$ is finite and so $\int_0^t ||A_z^n||_{-2,\alpha} dz$ is finite a.s. □

Now, using the decomposition (4.47) we are able to somehow exchange the expectation with the supremum in the control of η, i.e. Equation (4.36). Indeed, in comparison with Proposition 4.4.6, we are able to prove the following statement.

Proposition 4.4.13. *Under ($\mathcal{A}_{TGN}$), for every $\alpha > 1/2$ and $\theta \geq 0$,*

$$\sup_{n \geq 1} \mathbb{E}\left[\sup_{t \in [0,\theta]} ||\eta_t^n||_{-2,\alpha}^2 \right] < +\infty, \tag{4.49}$$

and $t \mapsto \eta_t^n$ belongs to $\mathcal{D}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$ almost surely.

Proof. Starting from (4.47), we have by convexity of the square function

$$\sup_{t \in [0,\theta]} ||\eta_t^n||_{-2,\alpha}^2 \leq 4\Big[||\eta_0^n||_{-2,\alpha}^2 + \theta \int_0^\theta (||L_z^* \eta_z^n||_{-2,\alpha}^2 + ||A_z^n||_{-2,\alpha}^2) dz + \sup_{t \in [0,\theta]} ||M_t^n||_{-2,\alpha}^2 \Big].$$

We deduce from Equation (4.43) that $\int_0^\theta \mathbb{E}[||L_z^* \eta_z^n||_{-2,\alpha}^2] dz \lesssim_\theta \sup_{z \in [0,\theta]} \mathbb{E}[||\eta_z^n||_{-1,\alpha}^2]$. Hence, taking the expectation in both sides of the inequality above and applying Propositions 4.4.6, 4.4.10 and 4.4.7 (remind (4.27)), we get (4.49). Starting from (4.47) and using that the integrals are continuous from Lemma 4.4.12 and M^n is càdlàg from Proposition 4.4.7, it follows that η^n is càdlàg.

□

4.4.d) Tightness result

Using the estimates proved in Section 4.4.c), the tightness criterion stated on page 134 can be checked.

Theorem 4.4.14. *Under ($\mathcal{A}_{TGN}$), for any $\alpha > 1/2$, the sequences of the laws of $(M^n)_{n \geq 1}$ and of $(\eta^n)_{n \geq 1}$ are tight in the space $\mathcal{D}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$.*

Proof. Condition ($\mathtt{A}_{1'}$) with $H_0 = \mathcal{W}_0^{-1,\alpha+1}$ and $H = \mathcal{W}_0^{-2,\alpha}$ is satisfied for both processes as a consequence of embedding (4.27) (remind that Hilbert-Schmidt operators are compact) and Propositions 4.4.6 and 4.4.7.

On the one hand, condition ($\mathtt{A}_2$) holds for $(M^n)_{n \geq 1}$ as soon as it holds for the trace of the processes $(\ll M^n \gg)_{n \geq 1}$ given below (4.47) [81, Rebolledo's theorem, p. 40]. Let

$(\varphi_k)_{k\geq 1}$ be an orthonormal basis of $\mathcal{W}_0^{2,\alpha}$. Let $\theta \geq 0$, $\delta_0 > 0$ and $\delta \leq \delta_0$. Furthermore, let τ_n be an $\mathbb{F}$-stopping time smaller than θ.

$$\begin{aligned}
&|\mathrm{Tr} \ll M^n \gg_{\tau_n+\delta} - \mathrm{Tr} \ll M^n \gg_{\tau_n}| \\
&\qquad = \left| \sum_{k\geq 1} \langle \ll M^n \gg_{\tau_n+\delta} (\varphi_k), \varphi_k \rangle - \langle \ll M^n \gg_{\tau_n} (\varphi_k), \varphi_k \rangle \right| \\
&\leq \sum_{k\geq 1} \frac{1}{n} \sum_{i=1}^{n} \int_{\tau_n}^{\tau_n+\delta} [R\varphi_k \left(S_{z-}^{n,i}\right)]^2 \lambda_z^{n,i} dz \leq ||\Psi||_\infty \frac{1}{n} \sum_{i=1}^{n} \int_{\tau_n}^{\tau_n+\delta} \sum_{k\geq 1} R\varphi_k \left(S_{z-}^{n,i}\right)^2 dz.
\end{aligned}$$

Noticing that $R\varphi_k(S_{z-}^{n,i}) = D_{0,S_{z-}^{n,i}}(\varphi_k)$ and then using Lemma 4.4.2 as we have done in the proof of Proposition 4.4.6, it follows that

$$\mathbb{E}\left[|\mathrm{Tr} \ll M^n \gg_{\tau_n+\delta} - \mathrm{Tr} \ll M^n \gg_{\tau_n}|\right] \leq \delta_0 ||\Psi||_\infty (C_2)^2 \left(1 + (M_{T_0} + \theta + \delta_0)^\alpha\right)^2.$$

This last bound is arbitrarily small for δ_0 small enough which gives condition $(\mathtt{A}_2)$ thanks to Markov's inequality.

On the other hand, using decomposition (4.47) and the fact that $(M^n)_{n\geq 1}$ is tight, it suffices to show the tightness of the remaining terms $(R_t^n = \eta_0^n + \int_0^t L_z^* \eta_z^n dz + \int_0^t A_z^n dz)_{n\geq 1}$ in order to show tightness of $(\eta^n)_{n\geq 1}$. Yet, using Equation (4.48), we have

$$\begin{aligned}
||R_{\tau_n+\delta}^n - R_{\tau_n}^n||_{-2,\alpha}^2 &= \left\| \int_{\tau_n}^{\tau_n+\delta} L_z^* \eta_z^n + A_z^n dz \right\|_{-2,\alpha}^2 \\
&\leq 2\delta \int_{\tau_n}^{\tau_n+\delta} (||L_z^* \eta_z^n||_{-2,\alpha}^2 + ||A_z^n||_{-2,\alpha}^2) dz \leq 2\delta_0 \int_0^{\theta+\delta_0} (C||\eta_z^n||_{-1,\alpha+1}^2 + ||A_z^n||_{-2,\alpha}^2) dz,
\end{aligned}$$

where C depends on θ and δ_0. Then, Propositions 4.4.6 and 4.4.10 imply that $\sup_{n\geq 1} \mathbb{E}[||R_{\tau_n+\delta}^n - R_{\tau_n}^n||_{-2,\alpha}^2] \leq C\delta_0$ for δ_0 small enough. Finally, Markov's inequality gives condition $(\mathtt{A}_2)$ for $(R^n)_{n\geq 1}$ and so the tightness of $(\eta^n)_{n\geq 1}$.

□

Remark 4.4.15. *For any $\alpha > 1/2$, every limit (with respect to the convergence in law) M (respectively η) in $\mathcal{D}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$ of the sequence $(M^n)_{n\geq 1}$ (resp. $(\eta^n)_{n\geq 1}$) satisfies*

$$\mathbb{E}\left[\sup_{t\in[0,\theta]} ||M_t||_{-2,\alpha}^2\right] < +\infty \quad \left(\textit{resp. } \mathbb{E}\left[\sup_{t\in[0,\theta]} ||\eta_t||_{-2,\alpha}^2\right] < +\infty\right). \tag{4.50}$$

Moreover, the limit laws are supported in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$.

Proof. Let us first show that the limit points are continuous. According to [12, Theorem 13.4.], it suffices to prove that for all $\theta \geq 0$, the maximal jump size of M^n and η^n on $[0,\theta]$ converge to 0 almost surely in order to prove the last point. Yet, for all φ in $\mathcal{W}_0^{2,\alpha}$,

$$\Delta M_t^n(\varphi) := |M_t^n(\varphi) - M_{t-}^n(\varphi)| = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} D_{0,S_{t-}^{n,i}}(\varphi) \mathbb{1}_{t\in N^{n,i}},$$

where we use the definition of $M_t^n(\varphi)$ given by (4.30) for φ in $\mathcal{C}_b^1$ and a density argument to extend it to φ in $\mathcal{W}_0^{2,\alpha}$, and

$$\langle \Delta\eta_t^n, \varphi \rangle := |\langle \eta_t^n, \varphi \rangle - \langle \eta_{t-}^n, \varphi \rangle| = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} D_{0,S_{t-}^{n,i}}(\varphi) \mathbb{1}_{t\in N^{n,i}}$$

where we used the fact that $(P_t)_{t\geq 0}$ is continuous in $\mathcal{W}_0^{-2,\alpha}$ (see Lemma 4.7.2). Since almost surely there is no common point to any two of the point processes $(N^{n,i})_{i=1,\dots,n}$, there is, almost surely, for all $t \geq 0$, at most one of the $\mathbb{1}_{t\in N^{n,i}}$ which is non null. Then, Lemma 4.4.2 implies

$$\begin{cases} \sup_{t\in[0,\theta]} ||\Delta M_t^n||_{-2,\alpha} \leq \frac{1}{\sqrt{n}} C_2(1+(M_{T_0}+\theta)^\alpha), \\ \sup_{t\in[0,\theta]} ||\Delta \eta_t^n||_{-2,\alpha} \leq \frac{1}{\sqrt{n}} C_2(1+(M_{T_0}+\theta)^\alpha), \end{cases}$$

which gives the desired convergence to 0.

Finally, (4.50) are consequences of Propositions 4.4.7 and 4.4.13 (remind (4.27)) where we use the previous step and the fact that the mapping $g \mapsto \sup_{t\in[0,\theta]} ||g_t||^2_{-2,\alpha}$ from $\mathcal{D}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$ to $\mathbb{R}$ is continuous at every point g^0 in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$. □

4.5 Characterization of the limit

The aim of this section is to prove convergence of the sequence $(\eta^n)_{n\geq 1}$ by identifying the limit fluctuation process η as the unique solution of a (SDE) in infinite dimension. We first prove, in Section 4.5.a), that every limit η satisfies a certain SDE (Theorem 4.5.6). Then, we show, in Section 4.5.b), that this SDE uniquely characterizes the limit law. This completes the proof of the convergence in law of $(\eta^n)_{n\geq 1}$ to η.

4.5.a) Candidate for the limit equation

In this section, the limit version of Equation (4.47) is stated. Apart from η^n, there are two random processes in (4.47) that are A^n and M^n. The following notation encompasses the source of the stochasticity of both A^n and M^n and is mainly used in order to track the correlations between those two quantities. For all $n \geq 1$, let W^n be the $\mathcal{W}_0^{-1,\alpha}$-valued martingale defined, for all $t \geq 0$ and φ in $\mathcal{W}_0^{1,\alpha}$, by

$$W_t^n(\varphi) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \varphi(S_{z-}^{n,i})(N^{n,i}(dz) - \lambda_z^{n,i} dz).$$

Notice that $M_t^n(\varphi) = W_t^n(R\varphi)$. Furthermore, as for M^n, the Doob-Meyer process $(\ll W^n \gg_t)_{t\geq 0}$ associated with $(W_t^n)_{t\geq 0}$ satisfies the following: for any $t \geq 0$, $\ll W^n \gg_t$ is the linear continuous mapping from $\mathcal{W}_0^{2,\alpha}$ to $\mathcal{W}_0^{-2,\alpha}$ given, for all φ_1 and φ_2 in $\mathcal{W}_0^{2,\alpha}$, by

$$\langle \ll W^n \gg_t (\varphi_1), \varphi_2 \rangle = \frac{1}{n} \sum_{i=1}^n \int_0^t \varphi_1(S_{z-}^{n,i}) \varphi_2(S_{z-}^{n,i}) \lambda_z^{n,i} dz. \tag{4.51}$$

All the results given for M^n in the previous section can be extended to W^n. In particular, the sequence $(W^n)_{n\geq 1}$ is tight in $\mathcal{D}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$. Next, we prove that it converges towards the Gaussian process W defined below.

Definition 4.5.1. *For any $\alpha > 1/2$, let W be a continuous centred Gaussian process with values in $\mathcal{W}_0^{-2,\alpha}$ with covariance given, for all φ_1 and φ_2 in $\mathcal{W}_0^{2,\alpha}$, for all t and $t' \geq 0$, by*

$$\begin{aligned} \mathbb{E}\left[W_t(\varphi_1) W_{t'}(\varphi_2)\right] &= \int_0^{t\wedge t'} \langle P_z, \varphi_1 \varphi_2 \Psi(\cdot, \overline{\gamma}(z)) \rangle \, dz \\ &= \int_0^{t\wedge t'} \int_0^{+\infty} \varphi_1(s)\varphi_2(s)\Psi(s, \overline{\gamma}(z)) u(z,s) ds dz, \end{aligned} \tag{4.52}$$

where u is the unique solution of (3.23).

Remark 4.5.2. *We refer to Appendix A.4 for the existence and uniqueness in law of such a process W. Furthermore, the process W defined above does not depend on the weight α in the sense that the definition is consistent with respect to the weights. Indeed, say W^α and W^β are two processes is the sense of Definition 4.5.1 with values in $\mathcal{W}_0^{-2,\alpha}$ and $\mathcal{W}_0^{-2,\beta}$ respectively. Assume for instance that $\beta > \alpha$. Then, W^β can be seen as a process with values in $\mathcal{W}_0^{-2,\alpha}$ via the canonical embedding $\mathcal{W}_0^{-2,\beta} \hookrightarrow \mathcal{W}_0^{-2,\alpha}$. Yet, the covariance structure (4.52) does not depend on the weights α and β so W^β is also a Gaussian process with values in $\mathcal{W}_0^{-2,\alpha}$ with the prescribed covariance and the uniqueness in law guaranties the equality of the laws of W^α and W^β as $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$-valued random variables.*

Proposition 4.5.3. *Under ($\mathcal{A}_{TGN}$), for any $\alpha > 1/2$, the sequence $(W^n)_{n\geq 1}$ of processes in $\mathcal{D}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$ converges in law to W.*

Proof. As already stated, the sequence $(W^n)_{n\geq 1}$ is tight. Then, let us consider the following decomposition, for any φ_1 and φ_2 in $\mathcal{W}_0^{2,\alpha}$,

$$\langle \ll W^n \gg_t (\varphi_1), \varphi_2 \rangle - \int_0^t \langle P_z, \varphi_1\varphi_2\Psi(\cdot, \overline{\gamma}(z)\rangle \, dz = B_t^n + C_t^n,$$

with

$$\begin{cases} B_t^n := \dfrac{1}{n}\displaystyle\sum_{i=1}^n \int_0^t \varphi_1(S_{z-}^{n,i})\varphi_2(S_{z-}^{n,i}) \left(\lambda_z^{n,i} - \Psi(S_{z-}^{n,i}, \overline{\gamma}(z))\right) dz, \\ C_t^n := \displaystyle\int_0^t \left\langle \overline{\mu}_{S_z}^n - P_z, \varphi_1\varphi_2\Psi(\cdot, \overline{\gamma}(z))\right\rangle dz, \end{cases}$$

where we used the fact that, almost surely, $\overline{\mu}_{S_{z-}}^n = \overline{\mu}_{S_z}^n$ for almost every z in $\mathbb{R}_+$. The first term B^n converges in L^1 to 0 by using the Lipschitz continuity of Ψ and the convergence of γ^n to $\overline{\gamma}$ given by Proposition 4.3.1. From the convergence

$$\frac{1}{n}\sum_{i=1}^n \delta_{(S_t^{n,i})_{t\geq 0}} \xrightarrow[n\to\infty]{} \mathcal{L}\big((\overline{S}_t^1)_{t\geq 0}\big),$$

which is a consequence of the propagation of chaos (Corollary 3.4.5), one can deduce that for almost every z, $\frac{1}{n}\sum_{i=1}^n \delta_{S_z^{n,i}} \xrightarrow[n\to\infty]{} P_z$ (see for instance [77, Proposition VI.3.14 and Lemma VI.3.12]). Then, dominated convergence implies that the second term C^n converges in expectation to 0. Hence, the bracket of W^n (4.51) converges to the covariance (4.52) for $t' = t$.

Furthermore, as for M^n (see the proof of Remark 4.4.15), the maximum jump size of W^n converges to 0. Hence, Rebolledo's central limit theorem for local martingales [132] gives, for every $\varphi_1, \dots, \varphi_k$ in $\mathcal{W}_0^{2,\alpha}$ and $t_1, \dots, t_k \geq 0$, the convergence of $(W_{t_1}^n(\varphi_1), \dots, W_{t_k}^n(\varphi_k))$ to a Gaussian vector with the prescribed covariance (4.52). The limit law of $(W^n)_{n\geq 1}$ is then characterized as the law of a continuous Gaussian process with covariance (4.52). □

Denote by $\mathbf{1} : \mathbb{R}_+ \to \mathbb{R}$ the constant function equal to 1 (which belongs to $\mathcal{W}_0^{2,\alpha}$ since we assume $\alpha > 1/2$) and note that $W_t^n(\mathbf{1})$ is the rescaled canonical martingale associated with the system of age-dependent Hawkes processes, namely

$$W_t^n(\mathbf{1}) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n N_t^{n,i} - \int_0^t \lambda_z^{n,i} dz\right).$$

Now, let us expand the decomposition (4.47) in order to get a closed equation. Let us recall the expansion of A^n given by (4.45), that is

$$A_t^n(\varphi) = \left\langle \overline{\mu}_{S_t}^n, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(t)) R\varphi \right\rangle \Gamma_{t-}^n + R_t^{n,(1)}(\varphi),$$

with $\Gamma_{t-}^n = \sqrt{n}(\gamma_t^n - \overline{\gamma}(t))$ and the rest term:

$$R_t^{n,(1)}(\varphi) := \frac{1}{n} \sum_{i=1}^n \left(R\varphi(S_{t-}^{n,i}) \frac{\partial \Psi}{\partial y}(S_{t-}^{n,i}, \overline{\gamma}(t)) \sqrt{n} r_t^{n,i} \right).$$

Below, we use the fact that this rest term converges to 0 in L^1 norm: indeed, recall that

$$|r_t^{n,i}| \lesssim |\gamma_t^n - \overline{\gamma}(t)|^2 \tag{4.53}$$

and, thanks to Proposition 4.3.1,

$$\mathbb{E}\left[|\gamma_t^n - \overline{\gamma}(t)|^2 \right] \lesssim_t n^{-1}.$$

Since Γ_{t-}^n (as part of $A_t^n(\varphi)$) only appears in (4.47) as an integrand and is only discontinuous on a set of Lebesgue measure equal to zero, we can replace it by its càdlàg version denoted by Γ_t^n. Let us consider the decomposition $\Gamma_t^n = \Upsilon_t^1 + \Upsilon_t^2 + \Upsilon_t^3$, with

$$\begin{cases} \Upsilon_t^1 := \sqrt{n} \displaystyle\int_0^t h(t-z) \left(\frac{1}{n} \sum_{i=1}^n N^{n,i}(dz) - \lambda_z^{n,i} dz \right) = \int_0^t h(t-z) dW_z^n(\mathbf{1}), \\ \Upsilon_t^2 := \sqrt{n} \displaystyle\int_0^t h(t-z) \frac{1}{n} \sum_{i=1}^n (\lambda_z^{n,i} - \Psi(S_{z-}^{n,i}, \overline{\gamma}(z))) dz, \\ \Upsilon_t^3 := \sqrt{n} \displaystyle\int_0^t h(t-z) \frac{1}{n} \sum_{i=1}^n (\Psi(S_{z-}^{n,i}, \overline{\gamma}(z)) - \overline{\lambda}(z)) dz = \int_0^t h(t-z) \left\langle \eta_z^n, \Psi(\cdot, \overline{\gamma}(z) \right\rangle dz, \end{cases}$$

where we used, in the last line, the fact that $\overline{\mu}_{S_{z-}}^n = \overline{\mu}_{S_z}^n$ for almost every z in $\mathbb{R}_+$, and $\overline{\lambda}(z) = \langle P_z, \Psi(\cdot, \overline{\gamma}(z)) \rangle$.

Based on Assumption $(\mathcal{A}_{y,\mathcal{C}^2}^{\Psi})$, as for Equation (4.44), one can give the Taylor expansion of the term

$$\Upsilon_t^2 = \sqrt{n} \int_0^t h(t-z) \frac{1}{n} \sum_{i=1}^n (\Psi(S_{z-}^{n,i}, \gamma_z^n) - \Psi(S_{z-}^{n,i}, \overline{\gamma}(z))) dz.$$

On the one hand, gathering the decomposition (4.29) with (4.45) and on the other hand gathering $\Gamma_t^n = \Upsilon_t^1 + \Upsilon_t^2 + \Upsilon_t^3$ with the Taylor expansion of Υ_t^2 give that (η^n, Γ^n) satisfies the following closed system for all φ in $\mathcal{W}_0^{2,\alpha}$,

$$\begin{aligned} \langle \eta_t^n, \varphi \rangle - \langle \eta_0^n, \varphi \rangle - \int_0^t \langle \eta_z^n, L_z \varphi \rangle dz - \int_0^t \left\langle \overline{\mu}_{S_z}^n, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) R\varphi \right\rangle \Gamma_z^n dz \\ - \int_0^t R_z^{n,(1)}(\varphi) dz = W_t^n(R\varphi), \end{aligned} \tag{4.54}$$

$$\begin{aligned} \Gamma_t^n - \int_0^t h(t-z) \left\langle \overline{\mu}_{S_z}^n, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) \right\rangle \Gamma_z^n dz - \int_0^t h(t-z) R_z^{n,(2)} dz \\ - \int_0^t h(t-z) \left\langle \eta_z^n, \Psi(\cdot, \overline{\gamma}(z) \right\rangle dz = \int_0^t h(t-z) dW_z^n(\mathbf{1}), \end{aligned} \tag{4.55}$$

where the rest term $R_z^{n,(2)}$ is defined by

$$R_z^{n,(2)} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \Psi}{\partial y}(S_{t-}^{n,i}, \overline{\gamma}(t)) r_t^{n,i}.$$

Note that, once again, Γ_{z-}^n, which naturally appears in the first integral term of (4.55), is replaced by its càdlàg version Γ_z^n since they are equal except on a null measure set.

The next step is to prove the tightness of $(\Gamma^n)_{n\geq 1}$. First note that, as a consequence of Proposition 4.3.1, for all $k \geq 0$ and $\theta \geq 0$,

$$\sup_{t\in[0,\theta]} \mathbb{E}\left[|\Gamma_t^n|^k\right] < +\infty, \tag{4.56}$$

since $\sup_{t\in[0,\theta]} \mathbb{E}\left[|\Gamma_t^n|^k\right] = \sup_{t\in[0,\theta]} \mathbb{E}\left[|\Gamma_{t-}^n|^k\right]$ because the underlying point processes admit intensities so that there is almost surely no jump at time θ.

Proposition 4.5.4. *Under ($\mathcal{A}_{TGN}$) and ($\mathcal{A}^h_{\text{Höl}}$), the sequence of the laws of $(\Gamma^n)_{n\geq 1}$ is tight in $\mathcal{D}(\mathbb{R}_+, \mathbb{R})$.*

Proof. The idea is to use (4.55). The first step is to simplify (4.55) by using the following convergences

$$\begin{cases} \mathbb{E}\left[\left|\sup_{t\in[0,\theta]} \int_0^t h(t-z) R_z^{n,(2)} dz\right|\right] \to 0, \\ \mathbb{E}\left[\left|\sup_{t\in[0,\theta]} \int_0^t h(t-z) \left\langle \overline{\mu}_{S_z}^n - P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) \right\rangle \Gamma_z^n dz\right|\right] \to 0. \end{cases} \tag{4.57}$$

These two convergences follow from the two following claims: by (4.56),

$$\sup_{z\in[0,\theta]} \mathbb{E}\left[\left|R_z^{n,(2)}\right|\right] \leq \mathrm{Lip}(\Psi) C n^{-1/2} \sup_{z\in[0,\theta]} \mathbb{E}\left[|\Gamma_{z-}^n|^2\right] \to 0,$$

and, by Cauchy-Schwarz inequality,

$$\sup_{z\in[0,\theta]} \mathbb{E}\left[\left|\left\langle \overline{\mu}_{S_z}^n - P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) \right\rangle \Gamma_z^n\right|\right] \to 0. \tag{4.58}$$

Indeed,

$$\begin{aligned} \mathbb{E}\left[\left|\left\langle \overline{\mu}_{S_z}^n - P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) \right\rangle \Gamma_z^n\right|\right] &\leq \mathbb{E}\left[\left|\frac{1}{\sqrt{n}} \left\langle \eta_z^n, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) \right\rangle\right|^2\right]^{1/2} \mathbb{E}\left[|\Gamma_z^n|^2\right]^{1/2} \\ &\leq \frac{1}{\sqrt{n}} \mathbb{E}\left[||\eta_z^n||_{-1,\alpha}^2\right]^{1/2} \left\|\frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z))\right\|_{1,\alpha} \mathbb{E}\left[|\Gamma_z^n|^2\right]^{1/2}, \end{aligned}$$

for any $\alpha > 1/2$. Then, (4.58) follows from Proposition 4.4.6 and Equations (4.24), (4.41) and (4.56).

Return to (4.55). The right-hand side is tight since it is convergent (Corollary 4.5.5) and the last term in the left hand side is tight since $(\eta^n)_{n\geq 1}$ is tight (with continuous limit

points) and $\eta \mapsto \int_0^t h(t-z)\,\langle \eta_z, \Psi(\cdot, \overline{\gamma}(z))\rangle\, dz$ is continuous at every point η in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$ thanks to Lemma 4.7.3 (remind (4.24) and (4.41)). Moreover, the term in the middle may be simplified by means of (4.57). Hence it remains to prove the tightness of the sequence of continuous processes $(I^n)_{n\geq 1}$ defined, for all $t \geq 0$, by

$$I_t^n := \int_0^t h(t-z) \left\langle P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) \right\rangle \Gamma_z^n dz.$$

We use Aldous criterion [12, Theorem 16.10.], that is the simplified version of the one stated on page 134 but for real valued processes. First, for all $\theta \geq 0$,

$$\mathbb{E}\left[\sup_{t\in[0,\theta]} |I_t^n|\right] \leq h_\infty(\theta)\mathrm{Lip}(\Psi) \int_0^\theta \mathbb{E}\left[|\Gamma_z^n|\right] dz,$$

is bounded uniformly with respect to n thanks to Equation (4.56). And Markov's inequality implies that, for every $\theta \geq 0$ and $\varepsilon > 0$, there exists $a > 0$ such that

$$\sup_{n\geq 1} \mathbb{P}\left(\sup_{t\in[0,\theta]} |I_t^n| \geq a\right) \leq \varepsilon,$$

which is the standard compactness condition.

Then, for the Aldous criterion, let us consider $\delta_0 > 0$, $\delta \leq \delta_0$ and for all $n \geq 1$, an $\mathbb{F}$-stopping time smaller than θ denoted by τ_n. Assume for a while that $h(0) = 0$ and extend the function h to the whole real line by setting 0 on the negative real numbers. As for Equation (4.59), we have

$$|I_{\tau_n+\delta}^n - I_{\tau_n}^n| \leq \mathrm{Höl}(h)\delta_0^{\beta(h)}\mathrm{Lip}(\Psi) \int_0^{\theta+\delta_0} |\Gamma_z^n| dz.$$

Hence, as before, (4.56) implies that $\sup_{n\geq 1} \mathbb{E}[|I_{\tau_n+\delta}^n - I_{\tau_n}^n|] \leq C(\theta+\delta_0)\delta_0^{\beta(h)}$ which is arbitrary small for δ_0 small enough and Markov's inequality gives that, for any $\varepsilon_1, \varepsilon_2 > 0$, there exists δ_0 such that $\sup_{n\geq 1} \sup_{\delta\leq\delta_0} \mathbb{P}\left(|I_{\tau_n+\delta}^n - I_{\tau_n}^n| \geq \varepsilon_1\right) \leq \varepsilon_2$, that is Aldous criterion. Hence, $(I^n)_{n\geq 1}$ is tight in $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$.

Now, if $h(0) \neq 0$, one can use the following decomposition,

$$I^n(t) = \int_0^t (h(t-z) - h(0)) \left\langle P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) \right\rangle \Gamma_z^n dz + h(0) \int_0^t \left\langle P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) \right\rangle \Gamma_z^n dz.$$

The first term is tight thanks to what we have done in the case $h(0) = 0$ whereas the tightness of the second one is simpler and left to the reader (use Equation (4.56)). □

Let us denote $V_t^n := \int_0^t h(t-z) dW_z^n(\mathbf{1})$ and $V_t := \int_0^t h(t-z) dW_z(\mathbf{1})$. The convergence of the sources of stochasticity in the system (4.54)-(4.55) is stated in the following corollary of Proposition 4.5.3.

Corollary 4.5.5. *Under ($\mathcal{A}_{TGN}$) and ($\mathcal{A}_{\mathrm{Höl}}^h$), the following convergence in law holds true in $\mathcal{D}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha} \times \mathbb{R})$,*

$$\left(R^* W_t^n, V_t^n\right)_{t\geq 0} \Rightarrow \left(R^* W_t, V_t\right)_{t\geq 0},$$

where R^* denotes the adjoint of R.

Proof. First, the tightness (and convergence) of $(R^*W^n)_{n\geq 1}$ comes from the continuity of R^* as a mapping from $\mathcal{W}_0^{-2,\alpha}$ to $\mathcal{W}_0^{-2,\alpha}$ which comes from the continuity of R as a mapping from $\mathcal{W}_0^{2,\alpha}$ to $\mathcal{W}_0^{2,\alpha}$ (this last point is proved similarly to the continuity from $\mathcal{W}_0^{1,\alpha}$ to $\mathcal{W}_0^{1,\alpha}$ proved at the end of the proof of Proposition 4.4.9). Then, let us show that $(V^n)_{n\geq 1}$ is tight in $\mathcal{D}(\mathbb{R}_+,\mathbb{R})$.

Assume that $h(0)=0$ and extend the function h to the whole real line by the value 0 on the negative real numbers.

-*(i)* For all $n\geq 1$, $V_0^n=0$ a.s. so $(V_0^n)_{n\geq 1}$ is clearly tight.

For any $t>r\geq 0$, since $h(r-z)=0$ as soon as $z\geq r$, one has

$$V_t^n - V_r^n = \int_0^t \left[h(t-z)-h(r-z)\right] dW_z^n(\mathbf{1}).$$

Let us denote, for all $x\geq 0$, $\mathtt{V}_{r,t}^n(x)=\int_0^x \left[h\left(t-z\right)-h\left(r-z\right)\right] dW_z^n(\mathbf{1})$. It is a martingale with respect to x. Burkholder-Davis-Gundy inequality [148, p. 894] gives the existence of a universal constant C_p such that

$$\mathbb{E}\left[\sup_{x\leq t}\left|\mathtt{V}_{r,t}^n(x)\right|^{2p}\right] \leq C_p \mathbb{E}\left[\left[\mathtt{V}_{r,t}^n\right]_t^p\right].$$

Yet, the quadratic variation of $\mathtt{V}_{r,t}^n$ is given by

$$\left[\mathtt{V}_{r,t}^n\right]_x = \frac{1}{n}\sum_{j=1}^n \int_0^x \left[h\left(t-z\right)-h\left(r-z\right)\right]^2 N^{n,j}(dz) \leq \mathrm{Höl}(h)^2\left|t-r\right|^{2\beta(h)} \frac{1}{n}\sum_{j=1}^n N_x^{n,j}. \tag{4.59}$$

So, using the exchangeability, we have for all $p\geq 0$,

$$\mathbb{E}\left[|V_t^n - V_r^n|^{2p}\right] \leq C_p \mathrm{Höl}(h)^{2p}\left|t-r\right|^{2\beta(h)p}\mathbb{E}\left[|N_t^{n,1}|^p\right].$$

Yet, the intensity of $N^{n,1}$ is bounded so that $N^{n,1}$ is stochastically dominated by a Poisson process with intensity $||\Psi||_\infty$. Hence, $\mathbb{E}[|N_t^{n,1}|^p]\leq \mathbb{E}[\mathrm{Poiss}(t||\Psi||_\infty)^p]$ where $\mathrm{Poiss}(t||\Psi||_\infty)$ is a Poisson variable with parameter $t||\Psi||_\infty$. This implies that $\mathbb{E}[|N_t^{n,1}|^p]$ is bounded uniformly in n by a locally bounded function of the time t, say $\tilde{C}_p(t)$ (which can be assumed to be increasing continuous without any loss of generality). Then, taking $p=1$, $t=\delta$ and $r=0$ and using Markov's inequality gives

-*(ii)* for all $\varepsilon>0$, $\lim_{\delta\to 0}\limsup_n \mathbb{P}(|V_\delta^n - V_0^n|>\varepsilon)=0$.

Finally, taking $p=1/\beta(h)$ and using Markov's inequality gives

-*(iii)* for all $\nu>0$, $\mathbb{P}(|V_t^n-V_r^n|>\nu)\leq \nu^{-2/\beta(h)}|F(t)-F(s)|^2$,

where $F(t):=(C_{\frac{1}{\beta(h)}}\tilde{C}_{\frac{1}{\beta(h)}}(t)\mathrm{Höl}(h)^{2/\beta(h)})^{1/2}\,t$ defines an increasing continuous function.

Hence, (i), (ii) and (iii) allow to apply Billingsley's criterion for tightness [77, Theorem VI.4.1] to deduce that $(V^n)_{n\geq 1}$ is tight.

Now, if $h(0)\neq 0$, one can use the following decomposition,

$$V_t^n = \int_0^t \left[h\left(t-z\right)-h\left(0\right)\right] dW_z^n(\mathbf{1}) + h\left(0\right) W_t^n(\mathbf{1}).$$

The first term is tight thanks to what we have done in the case $h\left(0\right)=0$ whereas the second term is converging since $(W^n)_n$ is converging, whence $(V^n)_{n\geq 1}$ is tight.

Now, since the limit trajectories of R^*W^n are continuous, the couple $(R^*W^n, V^n)_{n\geq 1}$ is tight in $\mathcal{D}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha} \times \mathbb{R})$. It now suffices to characterize the limiting finite dimensional distributions. Recall that $V_t = \int_0^t h(t-z)dW_z(\mathbf{1})$ and denote by $(t_1, \dots, t_k)$ a k-tuple of positive times.

First, suppose that h is piecewise constant. In that case, the convergence of W^n towards W easily implies the convergence of $((R^*W^n_{t_1}, V^n_{t_1}), \dots, (R^*W^n_{t_k}, V^n_{t_k}))$ to $((R^*W_{t_1}, V_{t_1}), \dots, (R^*W_{t_k}, V_{t_k}))$ (use the fact that h is a piecewise function to write $V^n_{t_i}$ as a sum of increments of $W^n(\mathbf{1})$).

Then, since h is continuous, one can find, for each $\varepsilon > 0$, a piecewise constant function h^ε such that $||h - h^\varepsilon||_\infty \leq \varepsilon$. Denote $V^{n,\varepsilon}_t := \int_0^t h^\varepsilon(t-z)dW^n_z(\mathbf{1})$ and notice that $\mathbb{E}\left[|V^n_t - V^{n,\varepsilon}_t|^2\right] \leq 2\varepsilon^2\mathbb{E}\left[< W^n(\mathbf{1}) >_t\right] \leq 2\varepsilon^2||\Psi||_\infty t \to 0$ as $\varepsilon \to 0$. In the same way, denote $V^\varepsilon_t := \int_0^t h^\varepsilon(t-z)dW_z(\mathbf{1})$ and remark that $\mathbb{E}\left[|V_t - V^\varepsilon_t|^2\right] \leq 2\varepsilon^2||\Psi||_\infty t \to 0$ as $\varepsilon \to 0$. Yet, the previous point gives the convergence, in terms of finite dimensional distributions, of $V^{n,\varepsilon}$ to V^ε for all $\varepsilon > 0$ so the convergence, in terms of finite dimensional distributions, of V^n to V follows which ends the proof. □

Both sequences $(\eta^n)_{n\geq 1}$ and $(\Gamma^n)_{n\geq 1}$ are tight with continuous limit trajectories. Tightness of the couple $(\eta^n, \Gamma^n)_{n\geq 1}$ hence follows and we are now in position to give the system satisfied by any limit (η, Γ).

Theorem 4.5.6. *Under ($\mathcal{A}_{TGN}$) and ($\mathcal{A}^h_{\text{Höl}}$), for all $\alpha > 1/2$, any limit (η, Γ) of the sequence $(\eta^n, \Gamma^n)_{n\geq 1}$ is a solution in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha} \times \mathbb{R})$ of the following system (formulated in $\mathcal{W}_0^{-3,\alpha} \times \mathbb{R}$),*

$$\forall \varphi \in \mathcal{W}_0^{3,\alpha}, \quad \langle \eta_t, \varphi\rangle - \langle \eta_0, \varphi\rangle - \int_0^t \langle \eta_z, L_z\varphi\rangle\, dz - \int_0^t \left\langle P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z))R\varphi \right\rangle \Gamma_z dz = W_t(R\varphi), \quad (4.60)$$

$$\Gamma_t - \int_0^t h(t-z)\,\langle \eta_z, \Psi(\cdot, \overline{\gamma}(z)\rangle\, dz - \int_0^t h(t-z)\left\langle P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z))\right\rangle \Gamma_z dz = \int_0^t h(t-z)dW_z(\mathbf{1}). \quad (4.61)$$

Remark 4.5.7. *The linear operator L_z appearing in (4.54) and (4.60) reduces the regularity of the test functions by 1. Hence, if we consider Equation (4.54) for test functions φ in $\mathcal{W}_0^{2,\alpha}$ then we must consider η^n as taking values in $\mathcal{W}_0^{-1,\alpha}$ when dealing with the integral term $\int_0^t \langle \eta^n_z, L_z\varphi\rangle\, dz$. Yet $(\eta^n)_{n\geq 1}$ is not tight in this space. Thus we consider (4.54) for test functions in $\mathcal{W}_0^{3,\alpha}$ so that every term is tight. That is why the limit equation (4.60) is formulated in $\mathcal{W}_0^{-3,\alpha}$. However, the limit process η takes values in the smaller space $\mathcal{W}_0^{-2,\alpha}$.*

In the proof below, we use the two following statements whose proofs are similar to those of Proposition 4.4.9 and Remark 4.4.11: for any $\alpha > 1/2$ and $\theta \geq 0$,

$$\text{for all } \varphi \text{ in } \mathcal{W}_0^{3,\alpha}, \quad \sup_{z\in[0,\theta]} \frac{||L_z\varphi||^2_{2,\alpha}}{||\varphi||^2_{3,\alpha}} < +\infty, \quad (4.62)$$

$$\text{and, for all } w \text{ in } \mathcal{W}_0^{-2,\alpha}, \quad \sup_{z\in[0,\theta]} \frac{||L^*_z w||^2_{-3,\alpha}}{||w||^2_{-2,\alpha}} < +\infty. \quad (4.63)$$

Remark 4.5.8. *The initial condition η_0 of the system (4.60)-(4.61) is determined by the distribution ζ_{N_-} of the initial conditions for the underlying point processes. First, η_0 is well defined as the limit in $\mathcal{W}_0^{-3,\alpha}$ of η_0^n. Indeed, the sequence $(\eta_0^n)_{n\geq 1}$ is tight in $\mathcal{W}_0^{-3,\alpha}$ (it is tight in $\mathcal{W}_0^{-2,\alpha}$ and there is a continuous embedding of $\mathcal{W}_0^{-2,\alpha}$ into $\mathcal{W}_0^{-3,\alpha}$) and for any φ in $\mathcal{W}_0^{3,\alpha}$, we have the convergence of the real-valued random variables $\langle \eta_0^n, \varphi\rangle = \sqrt{n}\left\langle \overline{\mu}_{S_0}^n - P_0, \varphi\right\rangle$ by applying the standard central limit theorem since the initial conditions are i.i.d.*

Proof. As a consequence of tightness and continuity of the limit trajectories, we have tightness of the process $(\eta^n, \Gamma^n, W^n, V^n)_{n\geq 1}$ in $\mathcal{D}(\mathbb{R}_+, \mathcal{W}^{-2,\alpha}\times\mathbb{R}\times\mathcal{W}_0^{-2,\alpha}\times\mathbb{R})$. Hence, let us assume without loss of generality that the sequence converges to (η,Γ,W,V) in $\mathcal{D}(\mathbb{R}_+, \mathcal{W}^{-2,\alpha}\times\mathbb{R}\times\mathcal{W}_0^{-2,\alpha}\times\mathbb{R})$.

Then, let $(\varphi_k)_{k\geq 1}$ be an orthonormal basis of $\mathcal{W}_0^{3,\alpha}$ and define the following applications: for all $k\geq 1$, $F_k : \mathcal{D}(\mathbb{R}_+, \mathcal{W}^{-2,\alpha}\times\mathbb{R}\times\mathcal{W}_0^{-2,\alpha}) \to \mathcal{D}(\mathbb{R}_+,\mathbb{R})$ satisfy for all $t\geq 0$,

$$F_k(f^1,f^2,f^3)(t) := \left\langle f_t^1, \varphi_k\right\rangle - \left\langle f_0^1,\varphi_k\right\rangle - \int_0^t \left\langle f_z^1, L_z\varphi_k\right\rangle dz \\ - \int_0^t \left\langle P_z, R\varphi_k \frac{\partial\Psi}{\partial y}(\cdot,\overline{\gamma}(z))\right\rangle f_z^2 dz - f_t^3(R\varphi_k),$$

and $G : \mathcal{D}(\mathbb{R}_+, \mathcal{W}^{-2,\alpha}\times\mathbb{R}\times\mathbb{R}) \to \mathcal{D}(\mathbb{R}_+,\mathbb{R})$ satisfy for all $t\geq 0$,

$$G(g^1,g^2,g^3)(t) := g_t^2 - \int_0^t h(t-z)\left\langle g_z^1, \Psi(\cdot,\overline{\gamma}(z)\right\rangle dz \\ - \int_0^t h(t-z)\left\langle P_z, \frac{\partial\Psi}{\partial y}(\cdot,\overline{\gamma}(z))\right\rangle g_z^2 dz - g_t^3.$$

Notice that the system (4.60)-(4.61) is equivalent to

$$\begin{cases} \forall k\geq 1, \quad F_k(\eta,\Gamma,W) = 0 \\ G(\eta,\Gamma,V) = 0. \end{cases} \tag{4.64}$$

Step one. Let us show that the first line of (4.64) is satisfied. First, we prove that for all $k\geq 1$, F_k is continuous at every point $(\tilde{f}^1,\tilde{f}^2,\tilde{f}^3)$ in $\mathcal{C}(\mathbb{R}_+,\mathcal{W}^{-2,\alpha}\times\mathbb{R}\times\mathcal{W}_0^{-2,\alpha})$. To state continuity of F_k at a continuous trajectory, it suffices to show continuity with respect to each coordinate f^1, f^2 and f^3.

- Equation (4.62) implies that $z\mapsto ||L_z\varphi_k||_{2,\alpha}$ is locally bounded and Lemma 4.7.3 gives the following:

$$f^1 \mapsto \left(t\mapsto \left\langle f^1(t),\varphi_k\right\rangle - \left\langle f^1(0),\varphi_k\right\rangle - \int_0^t \left\langle f^1(z), L_z\varphi_k\right\rangle dz\right)$$

is a mapping from $\mathcal{D}(\mathbb{R}_+,\mathcal{W}_0^{-2,\alpha})$ into $\mathcal{D}(\mathbb{R}_+,\mathbb{R})$ which is continuous at every point $\tilde{f}^1$ in $\mathcal{C}(\mathbb{R}_+,\mathcal{W}_0^{-2,\alpha})$.

- Then, notice that $\left|\left\langle P_z, R\varphi_k \frac{\partial\Psi}{\partial y}(\cdot,\overline{\gamma}(z))\right\rangle\right| \leq \mathrm{Lip}(\Psi)\mathbb{E}[|D_{0,\overline{S}_z^1}(\varphi_k)|]$ and Lemma 4.4.2 gives that

$$z\mapsto \left\langle P_z, R\varphi_k\frac{\partial\Psi}{\partial y}(\cdot,\overline{\gamma}(z))\right\rangle \text{ is locally bounded,}$$

so that applying Lemma 4.7.4 gives the continuity of F_k with respect to f^2.

- Finally, F_k is clearly continuous with respect to f^3.

Notice that (4.54) gives for any $k \geq 1$,

$$\forall t \geq 0,\ F_k(\eta^n, \Gamma^n, W^n)(t) - \int_0^t R_z^{n,(1)}(\varphi_k)dz - \int_0^t \left\langle \overline{\mu}_{S_z}^n - P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z))R\varphi_k \right\rangle \Gamma_z^n dz = 0. \tag{4.65}$$

Yet, we have, on the one hand, for all $\theta \geq 0$,

$$\begin{aligned} \sup_{z\in[0,\theta]} \mathbb{E}\left[\left|R_z^{n,(1)}(\varphi_k)\right|\right] &\leq \sqrt{n}\mathrm{Lip}(\Psi) \sup_{z\in[0,\theta]} \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n |D_{0,S_z^{n,i}}(\varphi_k)||r_z^{n,i}|\right] \\ &\leq n^{-1/2}C\mathrm{Lip}(\Psi)(1+(M_{T_0}+\theta)^\alpha)||\varphi_k||_{2,\alpha} \sup_{z\in[0,\theta]} \mathbb{E}\left[|\Gamma_{z-}^n|^2\right] \to 0, \end{aligned}$$

where we used Lemma 4.4.2, Equations (4.53) and (4.56), and on the other hand,

$$\sup_{z\in[0,\theta]} \mathbb{E}\left[\left|\left\langle \overline{\mu}_{S_z}^n - P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z))R\varphi_k \right\rangle \Gamma_z^n\right|\right] \to 0,$$

which follows from Cauchy-Schwarz inequality as we have done for (4.58).

These two convergences above imply

$$\begin{cases} \mathbb{E}\left[\left|\sup_{t\in[0,\theta]} \int_0^t R_z^{n,(1)}(\varphi_k)dz\right|\right] \to 0, \\ \mathbb{E}\left[\left|\sup_{t\in[0,\theta]} \int_0^t \left\langle \overline{\mu}_{S_z}^n - P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z))R\varphi_k \right\rangle \Gamma_z^n dz\right|\right] \to 0. \end{cases} \tag{4.66}$$

On the one hand, gathering (4.65) and (4.66) gives the convergence of $F_k(\eta^n, \Gamma^n, W^n)$ to 0 in probability and, on the other hand, applying the continuous mapping theorem gives the convergence in law of $F_k(\eta^n, \Gamma^n, W^n)$ to $F_k(\eta, \Gamma, W)$. Identifying the limits gives $F_k(\eta, \Gamma, W) = 0$ which ends this step.

Step two. Let us show that the second line of (4.64) is satisfied. First, we prove that G is continuous at every point $(\tilde{g}^1, \tilde{g}^2, \tilde{g}^3)$ in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}^{-2,\alpha} \times \mathbb{R} \times \mathbb{R})$. To state continuity of G at a continuous trajectory, it suffices to show continuity with respect to each coordinate g^1, g^2 and g^3.

- Equations (4.24), (4.41) and Lemma 4.7.3 give the following:

$$g^1 \mapsto \left(t \mapsto \int_0^t h(t-z)\left\langle g_z^1, \Psi(\cdot, \overline{\gamma}(z)\right\rangle dz\right)$$

is a mapping from $\mathcal{D}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$ into $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ which is continuous at every point $\tilde{g}^1$ in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$.

- Then, notice that $|\langle P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z))\rangle| \leq \mathrm{Lip}(\Psi)$ so that

$$z \mapsto \left\langle P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) \right\rangle \text{ is locally bounded,}$$

so that applying Lemma 4.7.4 gives the continuity of G with respect to g^2.

- Finally, G is clearly continuous with respect to g^3.

Notice that (4.55) gives, for all $t \geq 0$,

$$G(\eta^n, \Gamma^n, V^n)(t) - \int_0^t h(t-z) R_z^{n,(2)} dz - \int_0^t h(t-z) \left\langle \overline{\mu}_{S_z}^n - P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) \right\rangle \Gamma_z^n dz = 0.$$

Finally, the argument used to end the previous step also applies here.

To conclude, the two steps above give (4.64) which gives that the process (η, Γ) is a solution of (4.60)-(4.61). Finally, its trajectories are supported in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha} \times \mathbb{R})$ since η is supported in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$ and Γ is supported in $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ as a solution of (4.61) (remind that h is Hölder continuous). □

4.5.b) Uniqueness of the limit law

The next step in order to prove convergence of the sequence $(\eta^n, \Gamma^n)_{n \geq 1}$ is to prove uniqueness of the solutions of the limit system (4.60)-(4.61). Since the system is linear, the standard argument is to consider the system satisfied by the difference between two solutions and show that its unique solution is trivial. Let (η, Γ) and $(\hat{\eta}, \hat{\Gamma})$ be two solutions associated with the same "noise" W and the same initial condition η_0. Denote by $\tilde{\eta} := \eta - \hat{\eta}$ and $\tilde{\Gamma} := \Gamma - \hat{\Gamma}$ the differences. Then, $(\tilde{\eta}, \tilde{\Gamma})$ is a solution of the following system

$$\forall \varphi \in \mathcal{W}_0^{3,\alpha}, \quad \langle \tilde{\eta}_t, \varphi \rangle - \int_0^t \langle \tilde{\eta}_z, L_z \varphi \rangle \, dz - \int_0^t \left\langle P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) R\varphi \right\rangle \tilde{\Gamma}_z dz = 0, \tag{4.67}$$

$$\tilde{\Gamma}_t - \int_0^t h(t-z) \langle \tilde{\eta}_z, \Psi(\cdot, \overline{\gamma}(z)) \rangle \, dz - \int_0^t h(t-z) \left\langle P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) \right\rangle \tilde{\Gamma}_z dz = 0. \tag{4.68}$$

The standard follow-up is to use Gronwall's lemma. Let us show here why it does not work in our case. For instance, assume we want to prove that $||\tilde{\eta}||_{-3,\alpha} = 0$: heuristically, when applied to (4.68), Gronwall's argument gives that $|\tilde{\Gamma}_t|$ is bounded by some locally bounded function of t times the integral $\int_0^t ||\tilde{\eta}_z||_{-3,\alpha} dz$. However, even if we use this bound for $\tilde{\Gamma}$ in (4.67), Gronwall's argument cannot be applied since the term $\int_0^t \langle \tilde{\eta}_z, L_z \varphi \rangle \, dz$ involves $||\tilde{\eta}_z||_{-2,\alpha}$ which is greater than the desired norm $||\tilde{\eta}_z||_{-3,\alpha}$. This problem cannot be bypassed by upgrading the regularity as we have done before to deal with the fact that the operator L_z reduces the regularity of the test functions.

Since the main limitation comes from the differential part of the operator L_z, let us consider L_z as the sum of the first order differential operator plus a perturbation. More precisely, let $\mathcal{L} : \varphi \mapsto \varphi'$ and $G_t : \varphi \mapsto \Psi(\cdot, \overline{\gamma}(t)) R\varphi$ so that $L_t = \mathcal{L} + G_t$. Let us present here the heuristics behind the argument we use to bypass the issue induced by the differential operator $\mathcal{L}$: instead of studying the time derivative $\frac{d}{dt} \langle \tilde{\eta}_t, \varphi \rangle$ in (4.67), the idea is to find some family of test functions $(\varphi_t)_{t \geq 0}$ such that $\left\langle \tilde{\eta}_t, \frac{d}{dt} \varphi_t \right\rangle = - \langle \tilde{\eta}_t, \mathcal{L} \varphi_t \rangle$; thus the differential operator $\mathcal{L}$ vanishes in $\frac{d}{dt} \langle \tilde{\eta}_t, \varphi_t \rangle$ and Gronwall's argument can be applied.

More precisely, let us introduce the shift operators $\tau_t : \varphi \mapsto \varphi(\cdot + t)$ for all $t \geq 0$. Notice that these shift operators are linked with the method of characteristics applied to a transport equation with constant speed equal to 1 which is exactly the dynamics described by the differential operator $\mathcal{L}$. Below are given some bounds for the operators $\mathcal{L}$, G_t and τ_t when acting on the space $\mathcal{C}_b^4$.

Lemma 4.5.9. *Let φ be in $\mathcal{C}_b^4$. Assume that $t \mapsto ||\Psi(\cdot, \overline{\gamma}(t))||_{\mathcal{C}_b^4}$ is locally bounded. Then, $||\mathcal{L}\varphi||_{\mathcal{C}_b^3} \leq ||\varphi||_{\mathcal{C}_b^4}$, for all $t \geq 0$, $||\tau_t\varphi||_{\mathcal{C}_b^4} = ||\varphi||_{\mathcal{C}_b^4}$ and*

$$t \mapsto \frac{||G_t\varphi||_{\mathcal{C}_b^4}}{||\varphi||_{\mathcal{C}_b^4}} \quad \text{is locally bounded.}$$

Proof. The first two assertions follow from the definition of the norms $||\cdot||_{\mathcal{C}_b^k}$. The third and last one follows from Leibniz rule. □

Remark 4.5.10. *From now on, the test functions are considered in $\mathcal{C}_b^4$. Thus, we prove that η is characterized by the limit equation as a process with values in the dual space $\mathcal{C}_b^{-4}$. Nevertheless, since $\mathcal{C}_b^4$ is dense in $\mathcal{W}_0^{3,\alpha}$, it is also characterized by the limit equation as a process with values in $\mathcal{W}_0^{-3,\alpha}$ for instance.*

Let $t \geq t'$ and s in $\mathbb{R}$. Then,

$$\int_{t'}^t \tau_{t-z}\varphi'(s)dz = \int_{t'}^t \varphi'(s+t-z)dz = \varphi(s+t-t') - \varphi(s) = \tau_{t-t'}\varphi(s) - \varphi(s).$$

Moreover, since τ_t and $\mathcal{L}$ commute, one has

$$\tau_{t-t'}\varphi(s) - \varphi(s) = \int_{t'}^t \mathcal{L}(\tau_{t-z}\varphi)(s)dz.$$

Yet, Lemma 4.5.9 gives that $||\mathcal{L}(\tau_{t-z}\varphi)||_{\mathcal{C}_b^3} \leq ||\varphi||_{\mathcal{C}_b^4}$ thus $\int_{t'}^t \mathcal{L}\tau_{t-z}\varphi dz$ makes sense as a Bochner integral in $\mathcal{C}_b^3$ as soon as φ is in $\mathcal{C}_b^4$. Hence, in the proof below we use the following statement: for all φ in $\mathcal{C}_b^4$,

$$\tau_{t-t'}\varphi - \varphi = \int_{t'}^t \mathcal{L}(\tau_{t-z}\varphi)dz, \quad \text{as points in } \mathcal{C}_b^3. \tag{4.69}$$

Proposition 4.5.11. *Under ($\mathcal{A}_{CLT}$), the system (4.60)-(4.61) has no more than one solution in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha} \times \mathbb{R})$ once the initial condition η_0 and the "noise" W are fixed.*

Proof. Let (η, Γ) and $(\hat{\eta}, \hat{\Gamma})$ be two solutions of (4.60)-(4.61) in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha} \times \mathbb{R})$ associated with the the same "noise" W and the same initial condition η_0. Denote by $\tilde{\eta} := \eta - \hat{\eta}$ and $\tilde{\Gamma} := \Gamma - \hat{\Gamma}$ the differences. Since $\alpha > 1/2$, we have $\mathcal{W}_0^{-2,\alpha} \subset \mathcal{C}_b^{-4}$ (remind (4.27)) so $\tilde{\eta}$ belongs to $\mathcal{C}_b^{-4}$ and we will prove that $||\tilde{\eta}||_{\mathcal{C}_b^{-4}} = 0$.

Starting from (4.68), one has

$$|\tilde{\Gamma}_t| \leq h_\infty(t)||\Psi(\cdot, \overline{\gamma}(t))||_{\mathcal{C}_b^4} \int_0^t ||\tilde{\eta}_z||_{\mathcal{C}_b^{-4}} dz + h_\infty(t)\mathrm{Lip}(\Psi) \int_0^t |\tilde{\Gamma}|_z dz,$$

and Lemma 4.7.1 gives $|\tilde{\Gamma}_t| \lesssim_t \int_0^t ||\tilde{\eta}_z||_{\mathcal{C}_b^{-4}} dz$. Now, let φ be in $\mathcal{C}_b^4$ and use (4.69) and the fact that $\tilde{\eta}$ is in $\mathcal{W}_0^{-2,\alpha} \subset \mathcal{C}_b^{-3}$ to get $\langle \tilde{\eta}_t, \varphi \rangle = D_1 - D_2$ where

$$\begin{cases} D_1 := \displaystyle\int_0^t \langle \tilde{\eta}_{t'}, (\mathcal{L} + G_{t'})(\tau_{t-t'}\varphi) \rangle \, dt' + \int_0^t \left\langle P_{t'}, R\tau_{t-t'}\varphi \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(t')) \right\rangle \tilde{\Gamma}_{t'} dt'. \\ D_2 := \displaystyle\int_0^t \left\langle \tilde{\eta}_{t'}, (\mathcal{L} + G_{t'})(\int_{t'}^t \mathcal{L}(\tau_{t-z}\varphi)dz) \right\rangle dt' \\ \qquad\qquad + \displaystyle\int_0^t \left\langle P_{t'}, R\int_{t'}^t \mathcal{L}(\tau_{t-z}\varphi)dz \, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(t')) \right\rangle \tilde{\Gamma}_{t'} dt'. \end{cases} \tag{4.70}$$

The linearity of the operators allows to write $D_2 = D_{2,A} + D_{2,B}$ with

$$\begin{cases} D_{2,A} := \displaystyle\int_0^t \int_{t'}^t \langle \tilde{\eta}_{t'}, (\mathcal{L} + G_{t'})(\mathcal{L}(\tau_{t-z}\varphi))\rangle \, dz dt' \\ D_{2,B} := \displaystyle\int_0^t \int_{t'}^t \left\langle P_{t'}, R\mathcal{L}\tau_{t-z}\varphi \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(t'))\right\rangle \tilde{\Gamma}_{t'} dz dt'. \end{cases}$$

Then, the idea is to use Fubini's theorem to exchange the two integrals $\int_0^t$ and $\int_{t'}^t$.

On the one hand,

$$\int_0^t \int_{t'}^t |\, \langle \tilde{\eta}_{t'}, (\mathcal{L} + G_{t'})(\mathcal{L}(\tau_{t-z}\varphi))\rangle \,| dz dt' \leq \int_0^t \int_{t'}^t ||\tilde{\eta}_{t'}||_{\mathcal{C}_b^{-2}} ||(\mathcal{L} + G_{t'})(\mathcal{L}(\tau_{t-z}\varphi))||_{\mathcal{C}_b^2} dz dt'.$$

Notice that $\sup_{t'\in[0,t]} ||\tilde{\eta}_{t'}||_{\mathcal{C}_b^{-2}} \leq C(\sup_{t'\in[0,t]} ||\eta_{t'}||_{-2,\alpha} + \sup_{t'\in[0,t]} ||\hat{\eta}_{t'}||_{-2,\alpha}) < +\infty$ since η and $\hat{\eta}$ takes values in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$ and that, thanks to Lemma 4.5.9, for all $t' \leq t$, $||(\mathcal{L} + G_{t'})(\mathcal{L}(\tau_{t-z}\varphi))||_{\mathcal{C}_b^2} \lesssim_t ||\varphi||_{\mathcal{C}_b^4} < +\infty$. Hence, Fubini's theorem gives

$$D_{2,A} = \int_0^t \int_0^z \langle \tilde{\eta}_{t'}, (\mathcal{L} + G_{t'})(\mathcal{L}(\tau_{t-z}\varphi))\rangle \, dt' dz. \tag{4.71}$$

On the other hand,

$$\int_0^t \int_{t'}^t \left| \left\langle P_{t'}, R\mathcal{L}\tau_{t-z}\varphi \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(t'))\right\rangle \right| |\tilde{\Gamma}_{t'}| dz dt' \leq 2\mathrm{Lip}(\Psi) \int_0^t \int_{t'}^t ||\mathcal{L}(\tau_{t-z}\varphi)||_\infty |\tilde{\Gamma}_{t'}| dz dt'.$$

Remark that $||\mathcal{L}(\tau_{t-z}\varphi)||_\infty \leq ||\varphi||_{\mathcal{C}_b^4}$ and $\sup_{t'\in[0,t]} |\tilde{\Gamma}_t| \leq \sup_{t'\in[0,t]} |\Gamma_t| + \sup_{t'\in[0,t]} |\hat{\Gamma}_t| < +\infty$ since Γ and $\hat{\Gamma}$ takes values in $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$. Hence, Fubini's theorem gives

$$D_{2,B} = \int_0^t \int_0^z \left\langle P_{t'}, R\mathcal{L}\tau_{t-z}\varphi \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(t'))\right\rangle \tilde{\Gamma}_{t'} dt' dz. \tag{4.72}$$

Now, for any z in $[0,t]$, Equation (4.67) with $\varphi = \mathcal{L}(\tau_{t-z}\varphi)$ (it is a valid test function since it belongs to $\mathcal{C}_b^3 \subset \mathcal{W}_0^{3,\alpha}$) gives

$$\begin{aligned} \langle \tilde{\eta}_z, \mathcal{L}(\tau_{t-z}\varphi)\rangle = \int_0^z \langle \tilde{\eta}_{t'}, (\mathcal{L} + G_{t'})(\mathcal{L}(\tau_{t-z}\varphi))\rangle \, dt' \\ + \int_0^z \left\langle P_{t'}, R\mathcal{L}\tau_{t-z}\varphi \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(t'))\right\rangle \tilde{\Gamma}_{t'} dt'. \end{aligned}$$

Gathering the equation above with (4.71) and (4.72) gives

$$D_2 = \int_0^t \langle \tilde{\eta}_z, \mathcal{L}(\tau_{t-z}\varphi)\rangle \, dz,$$

which is exactly the term driven by $\mathcal{L}$ in the definition of D_1 (4.70) so that, coming back to $D_1 - D_2$, we have

$$D_1 - D_2 = \langle \tilde{\eta}_t, \varphi\rangle = \int_0^t \langle \tilde{\eta}_{t'}, G_{t'}(\tau_{t-t'}\varphi)\rangle \, dt' + \int_0^t \left\langle P_{t'}, R\tau_{t-t'}\varphi \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(t'))\right\rangle \tilde{\Gamma}_{t'} dt'.$$

Hence, using the bound we proved on $\tilde{\Gamma}$, we have for all φ in $\mathcal{C}_b^4$,

$$|\, \langle \tilde{\eta}_t, \varphi\rangle \,| \lesssim_t \int_0^t ||\tilde{\eta}_{t'}||_{\mathcal{C}_b^{-4}} ||G_{t'}(\tau_{t-t'}\varphi)||_{\mathcal{C}_b^4} dt' + 2||\varphi||_{\mathcal{C}_b^4} \mathrm{Lip}(\Psi) \int_0^t \int_0^{t'} ||\tilde{\eta}_z||_{\mathcal{C}_b^{-4}} dz,$$

and so $||\tilde{\eta}_t||_{\mathcal{C}_b^{-4}} \lesssim_t \int_0^t ||\tilde{\eta}_{t'}||_{\mathcal{C}_b^{-4}} dt'$ and Lemma 4.7.1 gives that for all $t \geq 0$, $||\tilde{\eta}_t||_{\mathcal{C}_b^{-4}} = 0$ thus $|\tilde{\Gamma}_t| = 0$ thanks to the bound we proved on $\tilde{\Gamma}$. Finally, since $\mathcal{C}_b^4$ is dense in $\mathcal{W}_0^{2,\alpha}$, we have $||\tilde{\eta}_t||_{-2,\alpha} = 0$. Thus, we have $(\eta, \Gamma) = (\hat{\eta}, \hat{\Gamma})$ in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha} \times \mathbb{R})$. □

We are now in position to conclude with the convergence of $(\eta^n, \Gamma^n)_{n\geq 1}$.

Theorem 4.5.12. *Under ($\mathcal{A}_{CLT}$), for any $\alpha > 1/2$, the sequence $(\eta^n, \Gamma^n)_{n\geq 1}$ converges in law in $\mathcal{D}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha} \times \mathbb{R})$ to the unique solution of the system (4.60)-(4.61) in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha} \times \mathbb{R})$.*

Proof. Since $(\eta^n, \Gamma^n)_{n\geq 1}$ is tight (Theorem 4.4.14 and Proposition 4.5.4), let (η, Γ) be a limit point. According to Theorem 4.5.6, (η, Γ) is a solution of the limit system (4.60)-(4.61) in $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha} \times \mathbb{R})$. Finally, the law of (η, Γ) is uniquely characterized by the limit system (Proposition 4.5.11 gives path-wise uniqueness and so Yamada-Watanabe theorem gives weak uniqueness by the same argument as [135, Theorem IX.1.7(i)]) and uniqueness of the limit law implies convergence of $(\eta^n, \Gamma^n)_{n\geq 1}$. □

Remark 4.5.13. *As mentioned in the introduction of the present chapter, considering processes over finite time horizons would have lead to equivalent results. This claim is based on the fact that the limit equation (4.60) is independent of the values of the test function φ outside the support K_t of η_t^n. Indeed,*

- *on the one hand, φ appears in the drift term, which is $\int_0^t \left\langle P_z, \frac{\partial \Psi}{\partial y}(\cdot, \overline{\gamma}(z)) R\varphi \right\rangle \Gamma_z dz$, evaluated against the measure P_z which is supported in K_t;*
- *on the other hand, the covariance structure of the Gaussian process W implies this independence property for $W_t(R\varphi)$.*

In that sense, the convergence stated for the whole positive time line $\mathbb{R}_+$ in Theorem 4.5.12 implies that the central limit theorem also holds true for the process $(\eta_t^n)_{0\leq t\leq \theta}$ as taking values in the dual of a standard Sobolev space of functions supported by K_θ. Conversely, the limit equation is consistent in time in the sense that one can recover our result by sticking together the CLTs obtained for the finite time horizon processes $(\eta_t^n)_{0\leq t\leq \theta}$.

4.6 Proofs

4.6.a) Proof of (4.20)

For simplicity, we show that, for every $m \leq n$, there exists a constant C which is independent of n, p and m such that

$$\mathbb{E}\left[\left(\frac{1}{m}\sum_{j=1}^{m}(\Delta_{t-}^{n,j})^p\right)^k\right] \leq C\left(\sum_{k'=1}^{k-1} m^{k'-k}\varepsilon_n^{(k',pk)}(t) + \varepsilon_n^{(k,p)}(t)\right), \tag{4.73}$$

from which (4.20) follows by choosing $m = \lfloor \frac{n}{k} \rfloor$.

Let us recall the multinomial formula using multi-indices $\mathbf{q} = (q_1, \ldots, q_m)$,

$$\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right)^k = \frac{1}{m^k}\sum_{|\mathbf{q}|=k}\binom{k}{\mathbf{q}}\prod_{i=1}^{m} x_i^{q_i},$$

where $|\mathbf{q}| = \sum_{i=1}^m q_i$. Denote by $k(\mathbf{q})$ the number of strictly positive indices in $\mathbf{q}$. Since the q_i's are integers, $|\mathbf{q}| = k$ implies $k(\mathbf{q}) \leq k$. First, let us remark that, for all $k' = 1, \ldots, k$, the number of multi-indices $\mathbf{q}$ such that $k(\mathbf{q}) = k'$ and $|\mathbf{q}| = k$ is bounded by $p(k',k)m^{k'}$ with $p(k',k) := \binom{k'-1}{k-1}$ being the number of partitions of k into exactly k' parts. Indeed, the vector consisting in the k' strictly positive indices forms a partition of k and there are at most $m^{k'}$ ways to complete it by $m - k'$ zeros to build a vector of length m.

Then, using the exchangeability of the processes $\Delta^{n,j}$, we have

- if $k(\mathbf{q}) = k$, then all the positive q_i's are equal to one and $\mathbb{E}[\prod_{i=1}^n((\Delta_{t-}^{n,i})^p)^{q_i}] = \varepsilon^{(k,p)}(t)$,
- if $k(\mathbf{q}) < k$, we can bound all the positive q_i's by k so that $\mathbb{E}[\prod_{i=1}^n((\Delta_{t-}^{n,i})^p)^{q_i}] \leq \varepsilon_n^{(k(\mathbf{q}),pk)}(t)$.

Hence, using that $\binom{k}{\mathbf{q}} \leq k!$, (4.73) holds with $C = \max_{k'=1,\ldots,k} p(k',k)k!$ for instance.

4.6.b) Proof of (4.21)

Let us first recall that $\xi_n^{(k)}(t) = \mathbb{E}\left[|\gamma_t^n - \overline{\gamma}(t)|^k\right]$ where γ_t^n and $\overline{\gamma}(t)$ are respectively defined below (4.6) and in (4.5). By convexity of the function $x \mapsto |x|^k$ (remind that $k \geq 2$), let us consider the decomposition

$$\xi_n^{(k)}(t) \leq 4^{k-1}\left(A^n(t) + B^n(t) + C^n(t) + D^n(t)\right) \tag{4.74}$$

where

$$\begin{cases} A^n(t) := \mathbb{E}\left[\left|\int_0^t h(t-z)\frac{1}{n}\sum_{j=1}^n(N^{n,j}(dz) - \lambda_z^{n,j}dz)\right|^k\right], \\ B^n(t) := \mathbb{E}\left[\left|\int_0^t h(t-z)\frac{1}{n}\sum_{j=1}^n(\Psi(S_{z-}^{n,j},\gamma_z^n) - \Psi(\overline{S}_{z-}^j,\gamma_z^n))dz\right|^k\right], \\ C^n(t) := \mathbb{E}\left[\left|\int_0^t h(t-z)\frac{1}{n}\sum_{j=1}^n(\Psi(\overline{S}_{z-}^j,\gamma_z^n) - \Psi(\overline{S}_{z-}^j,\overline{\gamma}(z)))dz\right|^k\right], \\ D^n(t) := \mathbb{E}\left[\left|\int_0^t h(t-z)\frac{1}{n}\sum_{j=1}^n(\overline{\lambda}_z^j - \overline{\lambda}(z))dz\right|^k\right]. \end{cases}$$

Recall that $\lambda_z^{n,j} = \Psi(S_{z-}^{n,j}, \gamma_z^n)$ and $\overline{\lambda}_z^j = \Psi(\overline{S}_{z-}^j, \overline{\gamma}(z))$.

- Study of $A^n(t)$. Fix t and consider the martingale $(M_x^{n,t})_{x\geq 0}$ defined, for all $x \geq 0$, by

$$M_x^{n,t} := \int_0^x h(t-z)\frac{1}{n}\sum_{j=1}^n (N^{n,j}(dz) - \lambda_z^{n,j}dz).$$

Its quadratic variation is $[M^{n,t}]_x = n^{-2}\int_0^x h(t-z)^2 \sum_{j=1}^n N^{n,j}(dz)$. Yet, Assumption $(\mathcal{A}_\infty^h)$ implies that

$$[M^{n,t}]_t \leq n^{-2} h_\infty(t)^2 \sum_{j=1}^n N_t^{n,j}.$$

Using the convexity of the power function (since $k/2 \geq 1$) and exchangeability, one has

$$\mathbb{E}\left[[M^{n,t}]_t^{k/2}\right] \lesssim_{(t,k)} n^{-k}\,\mathbb{E}\left[\Big|\sum_{j=1}^n N_t^{n,j}\Big|^{k/2}\right] \lesssim_{(t,k)} n^{-k}\, n^{k/2}\mathbb{E}\left[|N_t^{n,1}|^{k/2}\right].$$

Yet, the intensity of $N^{n,1}$ is bounded by $||\Psi||_\infty$ so $N^{n,1}$ is stochastically dominated by a Poisson process with intensity $||\Psi||_\infty$. Hence, $\mathbb{E}[|N_t^{n,1}|^{k/2}] \leq \mathbb{E}[\text{Poiss}(t||\Psi||_\infty)^{k/2}]$ where $\text{Poiss}(t||\Psi||_\infty)$ is a Poisson variable with parameter $t||\Psi||_\infty$. This last expectation is bounded uniformly in n by a locally bounded function of the time t. Then, Burkholder-Davis-Gundy inequality [148, p. 894] gives

$$A_1^n(t) = \mathbb{E}\left[|M_t^{n,t}|^k\right] \leq \mathbb{E}\left[[M^{n,t}]_t^{k/2}\right] \lesssim_{(t,k)} n^{-k/2}.$$

- Study of $B^n(t)$. Here, we use the fact that $S_{z-}^{n,j} = \overline{S}_{z-}^j$ with high probability and more precisely we recover the quantities $\varepsilon_n^{(k,p)}$ that we want to control. Using the convexity of the power function, Assumption $(\mathcal{A}_\infty^h)$ and denoting $x_z^j = \Psi(S_{z-}^{n,j}, \gamma_z^n) - \Psi(\overline{S}_{z-}^j, \gamma_z^n)$ we have

$$B^n(t) \leq h_\infty(t)^k t^{k-1}\int_0^t \mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n x_z^j\right|^k\right] dz.$$

Yet, $|x_z^j|$ is bounded by $||\Psi||_\infty \mathbb{1}_{\{S_{z-}^{n,j} \neq \overline{S}_{z-}^j\}} \leq ||\Psi||_\infty \Delta_{z-}^{n,j}$. Hence, using (4.73) with $p = 1$ and $m = n$,

$$\begin{aligned}
B^n(t) &\leq ||\Psi||_\infty^k h_\infty(t)^k t^{k-1}\int_0^t \mathbb{E}\left[\left|\frac{1}{n}\sum_{j=1}^n \Delta_{z-}^{n,j}\right|^k\right] dz \\
&\leq ||\Psi||_\infty^k h_\infty(t)^k C t^{k-1}\int_0^t \left(\sum_{k'=1}^{k-1} n^{k'-k}\varepsilon_n^{(k',k)}(z) + \varepsilon_n^{(k,1)}(z)\right) dz \\
&\leq ||\Psi||_\infty^k h_\infty(t)^k C t^k \left(\sum_{k'=1}^{k-1} n^{k'-k}\varepsilon_n^{(k',k)}(t) + \varepsilon_n^{(k,1)}(t)\right), \qquad (4.75)
\end{aligned}$$

where the last line comes from the fact that the $\varepsilon_n^{(k',k)}$'s are non-decreasing functions of t. Hence, $B_t^n \lesssim_{(t,k)} \sum_{k'=1}^{k-1} n^{k'-k}\varepsilon_n^{(k',k)}(t) + \varepsilon_n^{(k,1)}(t)$.

- Study of $C^n(t)$. Using the Lipschitz continuity of Ψ, Assumption $(\mathcal{A}_\infty^h)$, one has

$$C^n(t) \leq \text{Lip}(\Psi)^k h_\infty(t)^k t^{k-1}\int_0^t \mathbb{E}\left[|\gamma_z^n - \overline{\gamma}(z)|^k\right] dz \lesssim_{(t,k)} \int_0^t \xi_n^{(k)}(z)dz. \qquad (4.76)$$

- Study of $D^n(t)$. First remark that using Assumption $(\mathcal{A}^h_\infty)$, we have

$$D^n(t) \leq h_\infty(t)^k t^{k-1} \int_0^t \mathbb{E}\left[\left| \frac{1}{n} \sum_{j=1}^n \overline{\lambda}_z^j - \overline{\lambda}(z) \right|^k \right] dz.$$

Yet, the $\overline{\lambda}_z^j$'s are i.i.d. with mean $\overline{\lambda}(z)$ and they are bounded by $||\Psi||_\infty$. Hence, Rosenthal inequality [102] gives the existence of a constant $C(k)$ which depends only on k and $||\Psi||_\infty$ such that

$$\mathbb{E}\left[\left| \frac{1}{n} \sum_{j=1}^n \overline{\lambda}_z^j - \overline{\lambda}(z) \right|^k \right] \leq C(k) n^{-k/2}.$$

It then follows that $D^n(t) \lesssim_{(t,k)} n^{-k/2}$.

One deduces from the decomposition (4.74) and the four bounds on A^n, B^n, C^n and D^n that

$$\xi_n^{(k)}(t) \lesssim_{(t,k)} \left(n^{-k/2} + \sum_{k'=1}^{k-1} n^{k'-k} \varepsilon_n^{(k',k)}(t) + \varepsilon^{(k,1)}(t) \right) + \int_0^t \xi_n^{(k)}(z) dz,$$

and so Lemma 4.7.1 below gives the desired bound.

4.7 Lemmas

The following lemma is a generalization of the standard Grönwall lemma.

Lemma 4.7.1. *Let $f, g : \mathbb{R}_+ \to \mathbb{R}_+$ be two locally bounded non-negative measurable functions. Assume that for all $t \geq 0$,*

$$f(t) \lesssim_t g(t) + \int_0^t f(s) ds. \tag{4.77}$$

Then, for any $\theta \geq 0$, $\sup_{t\in[0,\theta]} f(t) \lesssim_\theta \sup_{t\in[0,\theta]} g(t)$.

Proof. For a fixed θ, Equation (4.77) implies that there exists a constant C such that for all $t \leq \theta$, $f(t) \leq C(\sup_{t\in[0,\theta]} g(t) + \int_0^t f(s)ds)$. Hence, standard Grönwall's inequality gives $\sup_{t\in[0,\theta]} f(t) \leq C \sup_{t\in[0,\theta]} g(t) e^{C\theta}$ which ends the proof. □

The next lemma proves continuity in time for the law of the age process associated with a point process. Its proof is similar to the proof of Lemma 3.7.2.

Lemma 4.7.2. *Assume that N admits the bounded $\mathbb{F}$-intensity λ_t and satisfy Assumption $(\mathcal{A}^{\zeta_{N_-}}_{u^{\text{in}},\infty})$. Denote by $(S_t)_{t\geq 0}$ its associated age process. Then, the law of S_t denoted by w_t is such that $t \mapsto w_t$ belongs to $\mathcal{C}(\mathbb{R}_+, \mathcal{W}_0^{-2,\alpha})$ for any $\alpha > 1/2$.*

Proof. This continuity result comes from the fact that the probability that N has a point in an interval goes to 0 as the size of the interval goes to 0. Fix $\alpha > 1/2$ and let t, t' be positive real numbers. First, remark that $S_{t+t'} = S_t + t'$ as soon as there is no point of N in the interval $[t, t + t']$ and so one has for all φ in $\mathcal{W}_0^{2,\alpha}$,

$$|\varphi(S_{t+t'}) - \varphi(S_t)| \leq ||D_{S_{t+t'},S_t}||_{-2,\alpha} ||\varphi||_{2,\alpha} + (|\varphi(S_{t+t'})| + |\varphi(S_t)|) \mathbb{1}_{N([t,t+t'])\neq 0}.$$

The bound obtained in Lemma 4.4.2 for the operator $D_{x,y}$ is too rough here. We need a finer bound: it holds that there exists a constant C such that $||D_{x,y}||_{-2,\alpha} \leq C|x-y|(1+\max(|x|^\alpha,|y|^\alpha))$. Indeed, by density, let us assume that φ is $\mathcal{C}^\infty$ with compact support and remark that

$$|\varphi(x)-\varphi(y)| \leq |x-y| \sup_{z,\,|z|\leq\max(|x|,|y|)} |\varphi'(z)| \leq |x-y|(1+\max(|x|^\alpha,|y|^\alpha))||\varphi||_{\mathcal{C}^{1,\alpha}}$$
$$\leq C|x-y|(1+\max(|x|^\alpha,|y|^\alpha))||\varphi||_{2,\alpha},$$

where we used (4.25) in the last inequality. Since $(\mathcal{A}^{\zeta_{N_-}}_{u^{\text{in}},\infty})$ is satisfied, $S_{t+t'}$ and S_t are upper bounded by $M_{T_0}+t+t'$ so that

$$\begin{cases} ||D_{S_t+t',S_t}||_{-2,\alpha} \leq Ct'(1+(M_{T_0}+t+t')^\alpha) \\ (|\varphi(S_{t+t'})|+|\varphi(S_t)|) \leq 2(1+(M_{T_0}+t+t')^\alpha)||\varphi||_{\mathcal{C}^{0,\alpha}}. \end{cases}$$

Hence, (4.25) gives

$$|\varphi(S_{t+t'})-\varphi(S_t)| \leq C(t'+\mathbb{1}_{N([t,t+t'])\neq 0})||\varphi||_{2,\alpha}.$$

Yet, $\mathbb{P}\left(N([t,t+t'])\neq 0\right) \leq \mathbb{E}\left[N([t,t+t'])\right] = \mathbb{E}[\int_t^{t+t'} \lambda_z dz]$ goes to 0 as t' goes to 0. The same argument for $t'<0$ gives continuity. □

The three lemmas below are used to get the limit equation satisfied by the fluctuations.

Lemma 4.7.3. *Let h be a locally bounded function and $(\varphi_t)_{t\geq 0}$ be a family of test functions in $\mathcal{W}_0^{2,\alpha}$ such that $t \mapsto ||\varphi_t||_{2,\alpha}$ is locally bounded. Then, $F : g \mapsto \int_0^t h(t-z)\langle g(z),\varphi_z\rangle dz$ is a mapping from $\mathcal{D}(\mathbb{R}_+,\mathcal{W}_0^{-2,\alpha})$ to $\mathcal{C}(\mathbb{R}_+,\mathbb{R})$ which is continuous at every point g_0 in $\mathcal{C}(\mathbb{R}_+,\mathcal{W}_0^{-2,\alpha})$.*

Proof. Let $(g_n)_{n\geq 1}$ be any sequence such that $g_n \to g_0$ for the Skorokhod topology. Since g_0 is continuous, the convergence also holds true for the local uniform topology [77, Proposition VI.1.17.]. We have for all $\theta \geq 0$,

$$\sup_{t\in[0,\theta]} |F(g_n)(t)-F(g_0)(t)| \leq \sup_{z\in[0,\theta]} h(z) \sup_{z\in[0,\theta]} ||g_n(z)-g_0(z)||_{-2,\alpha} \sup_{z\in[0,\theta]} ||\varphi_z||_{2,\alpha}. \quad (4.78)$$

Yet, the right hand side of (4.78) goes to 0 as n goes to infinity, which ends the proof. □

Lemma 4.7.4. *Assume that $(g^n)_{n\geq 1}$ converges to g for the Skorokhod topology in $\mathcal{D}(\mathbb{R}_+,\mathbb{R})$. If h satisfies $(\mathcal{A}^h_{\text{Höl}})$ and f is locally bounded, then*

$$\int_0^t h(t-z)f(z)g^n(z)dz \xrightarrow[n\to+\infty]{} \int_0^t h(t-z)f(z)g(z)dz,$$

as functions of t in $\mathcal{C}(\mathbb{R}_+,\mathbb{R})$ for the local uniform topology. In particular, the application F from $\mathcal{D}(\mathbb{R}_+,\mathbb{R})$ to $\mathcal{D}(\mathbb{R}_+,\mathbb{R})$ defined by

$$F(g)(t) := \int_0^t h(t-z)f(z)g(z)dz,$$

is continuous.

Proof. Let $c^n(t) := \int_0^t h(t-z)f(z)g^n(z)dz$ and $c(t) := \int_0^t h(t-z)f(z)g(z)dz$. Assume for a while that $h(0) = 0$ and extend the function h to the whole real line by setting 0 on the negative real numbers. Then, for all $t, \delta \geq 0$,

$$\begin{aligned} |c^n(t+\delta) - c^n(t)| &\leq \int_0^{t+\delta} |h(t+\delta-z) - h(t-z)||f(z)||g^n(z)|dz \\ &\leq (t+\delta)\text{Höl}(h) \sup_{z\in[0,t+\delta]} |f(z)| \sup_{z\in[0,t+\delta]} |g^n(z)|\delta^{\beta(h)}. \end{aligned}$$

Yet, since g^n is convergent, we have $\sup_{n\geq 1}\sup_{z\in[0,t+\delta]} |g^n(z)| < +\infty$ (see [77, Proposition VI.2.4.] for instance) which implies that for all $\theta \geq 0$,

$$\sup_{n\geq 1}\sup_{t\in[0,\theta]} |c^n(t+\delta) - c^n(t)| \to 0 \quad \text{as } \delta \to 0.$$

Hence, the sequence $(c^n)_{n\geq 1}$ is uniformly continuous. Moreover, for all $n \geq 1$, $c^n(0) = 0$ and the uniform continuity gives the uniform boundedness

$$\sup_{n\geq 1}\sup_{t\in[0,\theta]} |c^n(t)| < +\infty.$$

Then, Ascoli-Arzela theorem implies that the sequence $(c^n)_{n\geq 1}$ is relatively compact. It only remains to identify the limit for all $t \geq 0$. Yet, as a consequence of the dominated convergence and the fact that for almost every z, $g^n(z) \to g(z)$, we have $\int_0^t h(t-z)f(z)g^n(z)dz \to \int_0^t h(t-z)f(z)g(z)dz$.

Now, if $h(0) \neq 0$, one can use the following decomposition,

$$c^n(t) = \int_0^t (h(t-z) - h(0))f(z)g^n(z)dz + h(0)\int_0^t f(z)g^n(z)dz.$$

The first term is convergent thanks to what we have done in the case $h(0) = 0$ whereas the convergence of the second one is simpler and left to the reader. □

Chapter

5 Detection of dependence patterns with delay

Abstract. The Unitary Events (UE) method is a popular and efficient method used this last decade to detect dependence patterns of joint spike activity among simultaneously recorded neurons. The first introduced method is based on binned coincidence count [62] and can be applied on two or more simultaneously recorded neurons. Among the improvements of the methods, a transposition to the continuous framework has recently been proposed in [108] and fully investigated in [158] for two neurons. The goal of the present chapter is to extend this study to more than two neurons. The main result is the determination of the limit distribution of the coincidence count. This leads to the construction of an independence test between $L \geq 2$ neurons. Finally we propose a multiple test procedure via a Benjamini and Hochberg approach [11]. All the theoretical results are illustrated by a simulation study, and compared to the UE method proposed in [64]. Furthermore our method is applied on real data.

This chapter is the fruit of a collaboration with Thomas Laloë[1]. It is based on the study performed during the internship for the Master degree of the author of the manuscript (started in April 2013). Note that this chapter is slightly different from its corresponding article [30], published in *Biometrical Journal*. The main changes lie in Section 5.2 but the main results remain unchanged.

Contents

[1]Laboratoire J. A. Dieudonné, UMR CNRS 6621, Université Nice Sophia Antipolis, Parc Valrose 06108 Nice Cedex 2, France.

5.1 Introduction

The communication between neurons relies on their capacity to generate characteristic electric pulses called action potentials. These action potentials are usually assumed to be identical stereotyped events. Their time of occurrence (called spike) is considered as the relevant information. That is why the study of spike frequencies (firing rates) of neurons plays a key role in the comprehension of the information transmission in the brain [1, 57, 147]. Such neuronal signals are recorded from awake behaving animals by insertion of electrodes into the cortex to record the extracellular signals. Potential spike events are extracted from these signals by threshold detection and, by spike sorting algorithms, sorted into the spike signals of the individual single neurons. After this preprocessing, we dispose of sequences of spikes (called spike trains).

The analysis of spike trains has been an area of very active research for many years [20]. Although the rules underlying the information processing in the brain are still under burning debate, the detection of correlated firing between neurons is the objective of many studies in the recent years [46, 123, 144]. This synchronization phenomenon may take an important role in the recognition of sensory stimulus. In this article, the issue of detecting dependence patterns between simultaneously recorded spike trains is addressed. Despite the fact that some studies used to consider neurons as independent entities [10], many theoretical works consider the possibility that neurons can coordinate their activities [72, 117, 145, 160]. The understanding of this synchronization phenomenon [149] required the development of specific descriptive analysis methods of spike-timing over the last decades: cross-correlogram [118], gravitational clustering [56] or joint peristimulus time histogram (JPSTH, [3]). Following the idea that the influence of a neuron over others (whether exciting or inhibiting) results in the presence (or absence) of coincidence patterns, Grün and collaborators developed one of the most popular and efficient method used this last decade: the Unitary Events (UE) analysis method [62] and the corresponding independence test, which detects where dependence lies by assessing p-values (A Unitary Event is a spike synchrony that recurs more often than expected by chance). This method is based on a binned coincidence count that is unfortunately known to suffer a loss in synchrony detection, but this flaw has been corrected by the multiple shift coincidence count [66].

In order to deal with continuous time processes, a new method (Multiple Tests based on a Gaussian Approximation of the Unitary Events method), based on a generalization

of this count, the delayed coincidence count, has recently been proposed for two parallel neurons (Section 3.1 of [158]). The results presented in this article are in the lineage of this newest method and are applied on continuous point processes (random set of points which are modelling spike trains). Testing independence between real valued random variables is a well known problem, and various techniques have been developed, from the classical chi-square test to re-sampling methods for example. The interested reader may look at [91]. Some of these methods and more general surrogate data methods have been applied on binned coincidence count, since the binned process transforms the spike train in vectors of finite dimension. However, the case of point processes that are not preprocessed needs other tools and remains to study. Although the binned method can deal with several neurons (six simultaneously recorded neurons are analysed in [64]), both of the improvements (Multiple Shift and MTGAUE) can only consider pairs of neurons. Thus, our goal is to generalize the method introduced in [158] for more than two neurons. Unlike MTGAUE, our test is not designed to be performed on multiple time windows. However it can be multiple with respect to the different possible patterns composed from $n \geq 2$ neurons (see Section 5.5.c)).

In Section 5.2, we introduce the different notions of coincidence used through this article. In Section 5.3, a test is established and the asymptotic control of its false positive rate is proven. In Section 5.4 our test is confronted to the original UE method on simulated data and the accuracy of the Gaussian approximation is verified. In Section 5.5 the relevance of our method when our main theoretical assumptions are weakened is also empirically put on test. Section 5.6 presents an illustration on real data. All the technical proofs are given in the Appendix.

5.2 Notions of coincidence and original UE method

In order to detect synchronizations between the involved neurons, different notions of coincidence can be considered. Informally, there is a coincidence between neurons when they each emit a spike more or less simultaneously. This notion has already been used in UE methods [64] and is based on the following idea: a real dependency between $n \geq 2$ neurons should be characterized by an unusually large (or low) number of coincidence [61, 62, 158].

5.2.a) Original UE method

The UE method (see [62]) considers discretized spike trains at a resolution ℓ of typically 1 or 0.1 millisecond. Therefore, in the discrete-time framework, each trial consists of a set of n spike trains (one for each recorded neuron), each spike train being represented by a sequence of 0 and 1 of length S. Since it is quite unlikely that two spikes occur at exactly the same time at this resolution ℓ, spike trains are binned and clipped at a coarser level. More precisely for a fixed bin size $\Delta = d\ell$ (d being an integer), a new sequence of length S/d of 0 and 1 is associated to each spike train (1 if at least one spike occurs in the corresponding bin, 0 otherwise). For more precise informations on the binning procedure and the link with point processes we refer the interested reader to [158].

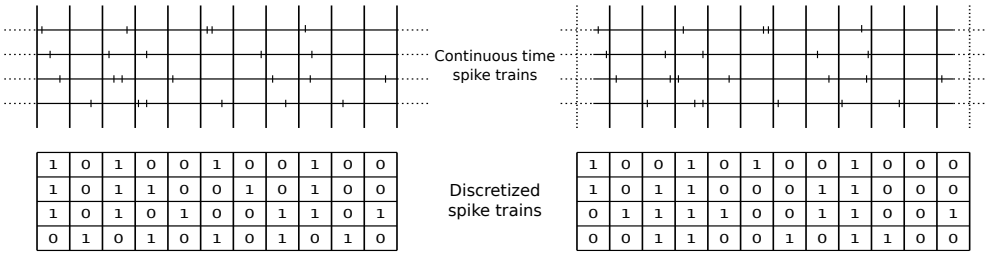
A constellation or pattern, denoted by w, is a vector of size n of 0 and 1 (see Figure 5.1.**A** or [64]). Of course, there are 2^n different constellations which are naturally in a

one-to-one correspondence with the subsets[2] of $\{1, ..., n\}$. Let us denote by $\mathscr{L}_w$ the subset corresponding to the constellation w (see Figure 5.1.**A**). The UE statistic associated to some constellation w consists in counting the number of occurrences of such w in the set of S/d vectors of size n. This number of occurrences being called *binned coincidence count*.

A : Simultaneously recorded neurons

Bin	1	2	3	...	(S/d)-1	S/d
Neuron 1	1	1	0	...	0	1
Neuron 2	0	1	0	...	1	1
Neuron 3	0	1	0	...	0	1
Neuron 4	0	0	1	...	1	1
$\mathcal{L}_w$	{1}	{1, 2, 3}	{4}	...	{2, 4}	{1, 2, 3, 4}

B : Discretization of spike trains



1	0	1	0	0	1	0	0	1	0	0
1	0	1	1	0	0	1	0	1	0	0
1	0	1	0	1	0	0	1	1	0	1
0	1	0	1	0	1	0	1	0	1	0

1	0	0	1	0	1	0	0	1	0	0	0
1	0	1	1	0	0	0	1	1	0	0	0
0	1	1	1	1	0	0	1	1	0	0	1
0	0	1	1	0	0	1	0	1	1	0	0

Figure 5.1: In **A**, 4 parallel binary processes of length S are displayed. At each time step, the constellation and its corresponding subset of $\{1, 2, 3, 4\}$ are given. For instance, the constellation associated to the first bins is the vector $(1, 0, 0, 0)$ and the corresponding subset is $\{1\}$. In **B**, illustration of the UE method with two different choices of bins of the same size (the results are different, for example the constellation full of 1s is present in the second case and not in the first one).

However, as shown in Figure 5.1.**B**, this method largely depends on the bin choice and it has been proven in [66] that this can lead in the case $n = 2$ to up to 60% of loss in detection when Δ is of the order of the range of interaction.

To detect dependency between neurons, two estimators of the expected coincidence count are compared. The first one is the empirical mean $\bar{m}_w$ of the number of occurrences of a given constellation w through M trials,

$$\bar{m}_w = \frac{1}{M} \sum_{k=1}^{M} m_w^{(k)},$$

where $m_w^{(k)}$ is the number of occurrences of w during the k^{th} trial. This estimator is consistent (that is, converges towards the expected value of the number of occurrences) even with dependency between the spike trains. The second one is consistent only under the independence hypothesis, and is given by

$$\hat{m}_{g,w} = \frac{S}{d} \prod_{l \in \mathscr{L}_w} \hat{p}_l \prod_{k \notin \mathscr{L}_w} (1 - \hat{p}_k), \tag{5.1}$$

where $\hat{p}_i$ is the empirical probability of finding a spike in a bin of neuron i.

[2]Here, the set $\{1, \ldots, n\}$ represents the recorded neurons numbered from 1 to n.

This enables the construction of the test described in [64] and based on the comparison between the statistic $M\bar{m}_w$ and a quantile of the Poisson distribution $\mathcal{P}(M\hat{m}_{g,w})$ where M is the number of trials. Most of the time only tests by upper values are computed [62, 64]. However, following the study of [158], we have decided to focus on symmetric tests. Hence, the symmetric test based on the UE method rejects the independence hypothesis when $\bar{m}_w$ is too different from $\hat{m}_{g,w}$. However, such a test necessarily makes mistakes. For example, a *false positive* corresponds to an incorrect rejection of the null hypothesis. Hence, an a priori upper bound on the false positive rate, that is the *significance level* (or just *level*), must be given in order to construct a decision rule. Grün [64] heuristically assumes that under the independence hypothesis, the Poisson distribution $\mathcal{P}(M\hat{m}_{g,w})$ is a good estimate of the coincidence count distribution. Therefore, the symmetric independence test with level α based on the UE method is governed by the following rule: if

$$M\bar{m}_w \geq q_{1-\alpha/2} \quad \text{or} \quad M\bar{m}_w \leq q_{\alpha/2},$$

where q_x is the x-quantile of the Poisson distribution $\mathcal{P}(M\hat{m}_{g,w})$, then the independence hypothesis is rejected.

Finally, let us mention that the theoretical framework of the UE method is the class of Bernoulli processes. The equivalent in the "continuous" framework is the Poisson process (as it can be seen in [158]). This leads to a different estimator of the expected coincidence count and a different test which are defined properly in Section 5.3.

5.2.b) Delayed coincidence

In this section, we focus on another coincidence count which deals with continuous data. It is based on the notion of delayed coincidence. In continuity with [158], we use a formalism based on point processes. Nevertheless, notice that the notion of delayed coincidence is pretty natural and was used in [108] or [13] with minor differences.

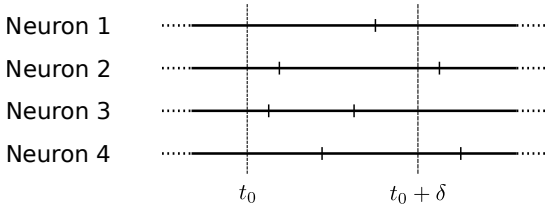
Considering $N_1, \ldots, N_n$, some point processes on $[a, b]$, and $\mathscr{L} \subset \{1, \ldots, n\}$ a set of indices of cardinal $L \geq 2$, the *delayed coincidence count* $X_{\mathscr{L}}$ (of delay $\delta < (b-a)/2$) over the neurons of subset $\mathscr{L}$ in the time window $[a, b]$ is given by

$$X_{\mathscr{L}} = X_{\mathscr{L}}(\delta) = \sum_{(x_1,\ldots,x_L)\in\prod_{l\in\mathscr{L}} N_l} \mathbb{1}_{\left|\max_{i\in\{1,\ldots,L\}} x_i - \min_{i\in\{1,\ldots,L\}} x_i\right| \leq \delta}. \tag{5.2}$$

The delayed coincidence count can be explained in the following way :

- Fix some duration parameter δ which is the equivalent of the bin size Δ,
- Count how many times each neuron in $\mathscr{L}$ spikes almost at the same time, modulo the delay δ.

Figure 5.2 below gives a graphical representation of delayed coincidences.



subset	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,3\}$	$\{2,4\}$	$\{3,4\}$	$\{1,2,3\}$	$\{1,2,4\}$	$\{1,3,4\}$	$\{2,3,4\}$	$\{1,..,4\}$
number	1	2	1	2	1	2	2	1	2	2	2

Figure 5.2: Four parallel spike trains are represented above. The table gives the number of coincidences, if one considers only the spikes occuring in the interval $[t_0, t_0 + \delta]$, of delay δ over the neurons of each subset $\mathscr{L}$ of $\{1, \ldots, 4\}$ with cardinal L greater than 2.

In practice, the algorithm used to count delayed coincidences uses a window of time length δ sliding from a to b the two ends of the analysed time interval. More precisely, assume that we fix a subset $\mathscr{L}$, at each step of the algorithm, the lower bound of the sliding window is a spike of a neuron in the subset $\mathscr{L}$ and we count the number of coincidences of delay δ occuring in the sliding window over the neurons in the subset $\mathscr{L}$. This last counting is done exactly like in Figure 5.2. A visualisation of the algorithm is given below.

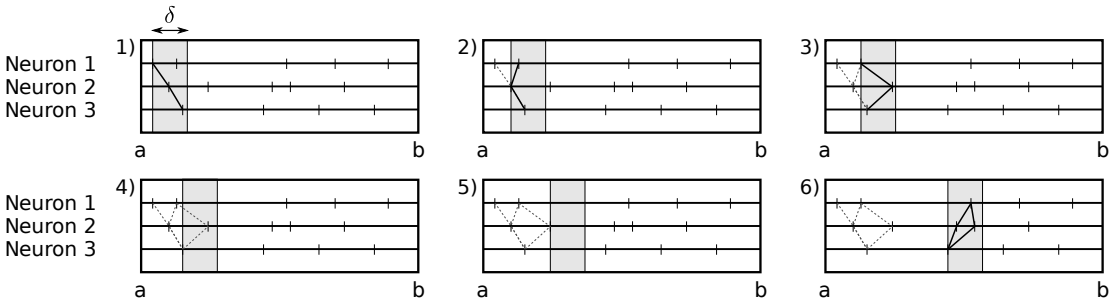


Figure 5.3: The six first steps in the dynamical computation of the delayed coincidence count. Here, there are 3 parallel time point processes. We consider the full pattern, i.e. $\mathscr{L} = \{1, 2, 3\}$. The grey rectangle represents the sliding time window of length δ. The bold lines denote the coincidence patterns counted at each step and the grey dashed ones denote the coincidence pattern which have been counted in any one of the previous steps. At each of steps 1,2 and 3, exactly one coincidence is counted. At steps 4 and 5, no coincidence is detected. And, at step 6, two coincidences are counted.

5.3 Study of the delayed coincidence count

Once the notion of coincidence is defined with respect to continuous data (Equation (5.2)), mathematical tools can be used to construct the desired independence test. The procedure is to provide the expected value and variance of the variable $X_{\mathscr{L}}$ in function of the firing rates. These computations classically imply a Gaussian approximation with respect to i.i.d trials. Unfortunately the firing rates are usually unknown. Thus the final step is to replace the firing rates by their estimator to compute the estimated expected value and variance. This plug-in procedure is known to change the underlying distribution. As in [158], the delta method provides the exact nature of this change.

In the continuous framework, a sample is composed of M observations of $N_1, \cdots, N_n$ which are the point processes associated to the spike trains of n neurons on a window $[a, b]$. The goal is to answer the following question:

Given $\mathscr{L}$ a subset of $\{1, \ldots, n\}$, are the processes $N_l,\ l \in \mathscr{L}$ independent?

To do this, a statistical test comparing the two hypotheses

$$\begin{cases} (\mathcal{H}_0) & \text{The processes } N_l,\ l \in \mathscr{L} \text{ are independent;} \\ (\mathcal{H}_1) & \text{The processes } N_l,\ l \in \mathscr{L} \text{ are not independent;} \end{cases}$$

is proposed.

In this section our test and its asymptotic relevance are introduced. First, let us present and discuss our main assumptions which are the same as in [158].

Assumption A1. *$N_1, \ldots, N_n$ are Poisson processes.*

This assumption can be resumed to an assumption of independence of a point process with respect to itself over the time, as Bernoulli processes in discrete settings.

Assumption A2. *The Poisson processes $N_1, \ldots, N_n$ are homogeneous on $[a, b]$.*

Assumption A2 may also appear very restrictive. But once again Bernoulli processes considered in [66, 64] have the same drawback. Moreover, if necessary, one can partition $[a, b]$ in smaller intervals on which A2 is satisfied. For more precise informations on Poisson processes we refer the interested reader to [86].

These assumptions are necessary in this work in order to obtain an explicit form for the expected number of coincidences (and its variance). Note that there exist some surrogate methods in the literature for which there is no need of a model on the data (see [63, 94] for a review). In particular two kind of methods are commonly used: dithering methods (involving random shifts of individual spikes [154, 95], or random shifts of patterns of spikes [69]), and trial-shuffling methods [124, 125]. However, they are based on binned coincidence count, and there is no equivalent, up to our knowledge, with a delayed coincidence count, due to serious computational issues. Alternative works have also been done in the Bayesian paradigm [6]. However, as announced in the introduction, we empirically show in Section 5.5 that the assumptions can be weakened. In particular, point processes admitting refractory periods can be taken into account. Thus, a nice perspective of this work could be to derive theoretical results with these weakened assumptions.

5.3.a) Asymptotic properties

In order to build our independence test, one needs to understand the behaviour of the number of coincidence $X_{\mathscr{L}}$ under the independence hypothesis $\mathcal{H}_0$. In particular, the expected value and the variance of $X_{\mathscr{L}}$ are computed here. In a general point processes framework, these computations are impossible. This is why some restrictive assumptions are needed, such as A1, A2, or the independence of the processes, as done in the original UE method where independent Bernoulli processes have been considered.

Theorem 5.3.1. *Let $\mathscr{L}$ and $X_{\mathscr{L}}$ be defined as previously. Assume assumptions A1 and A2 and denote by $\lambda_1, \dots, \lambda_n$ the respective intensities of $N_1, \dots, N_n$. Under hypothesis $\mathcal{H}_0$, the expected value and the variance of the number of coincidences $X_{\mathscr{L}}$ are given by:*

$$m_{0,\mathscr{L}} := \mathbb{E}\left[X_{\mathscr{L}}\right] = \left(\prod_{l \in \mathscr{L}} \lambda_l\right) I(L,0)$$

and

$$\mathbf{V}\mathrm{ar}(X_{\mathscr{L}}) = m_{0,\mathscr{L}} + \sum_{k=1}^{L-1} \left(\sum_{\substack{\mathscr{J} \subset \mathscr{L} \\ \#\mathscr{J} = k}} \prod_{j \in \mathscr{J}} \lambda_j^2 \prod_{l \notin \mathscr{J}} \lambda_l \right) I(L,k),$$

where the $I(L,k)$ are given by Proposition 5.3.2 below.

The proof relies on the calculus of the moments of a sum over a Poisson Process and is given in Appendix 5.8. The integral $I(L,k)$ can be seen as the contribution of a subset of k neurons to the number of coincidences between the L neurons.

Proposition 5.3.2. *For $b > a \geq 0$ and $0 < \delta < b - a$, define for every k in $\{0, \dots, L\}$*

$$I(L,k) = \int_{[a,b]^{L-k}} \left(\int_{[a,b]^k} \mathbb{1}_{\left| \max_{i \in \{1,\dots,L\}} x_i - \min_{i \in \{1,\dots,L\}} x_i \right| \leq \delta} \, dx_1 \dots dx_k \right)^2 dx_{k+1} \dots dx_L,$$

where the convention $\int_{[a,b]^0} f(x)\,dx = f(x)$ is set. Then, for $L \geq 2$, and k in $\{0, \dots, L-1\}$,

- $I(L,L) = L^2 (b-a)^2 \delta^{2L-2} - 2L(L-1)(b-a)\delta^{2L-1} + (L-1)^2 \delta^{2L}$,
- $I(L,k) = f(L,k)(b-a)\delta^{L+k-1} - h(L,k)\delta^{L+k}$,
  where $f(L,k) = \dfrac{k(k+1) + L(L+1)}{L-k+1}$,
  and $h(L,k) = \dfrac{-k^3 + k^2(2+L) + k(5+2L-L^2) + L^3 + 2L^2 - L - 2}{(L-k+2)(L-k+1)}$.

Once the behaviour of $X_{\mathscr{L}}$ under $\mathcal{H}_0$ is known, the method to construct an independence test is straight-forward. Suppose that M independent and identically distributed (i.i.d.) trials are given. Denote $N_i^{(k)}$ the spike train corresponding to the neuron i during the k^{th} trial. As for the UE method, the idea is to compare two estimates of the expectation of $X_{\mathscr{L}}$. The first one is the empirical mean of $X_{\mathscr{L}}$:

$$\bar{m}_{\mathscr{L}} = \frac{1}{M} \sum_{k=1}^{M} X_{\mathscr{L}}^{(k)}, \tag{5.3}$$

where $X_{\mathscr{L}}^{(k)}$ is the delayed coincidence count during the k^{th} trial. This estimate converges even if the processes are not independent. More precisely the following asymptotic result is given by the Central Limit Theorem

$$\sqrt{M}\left(\bar{m}_{\mathscr{L}} - \mathbb{E}\left[X_{\mathscr{L}}\right]\right) \xrightarrow[M \to \infty]{\mathcal{D}} \mathcal{N}\left(0, \mathbf{V}\mathrm{ar}(X_{\mathscr{L}})\right),$$

where $\xrightarrow{\mathcal{D}}$ denotes the convergence of distribution and $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2.

The second estimate is given by Theorem 5.3.1. Indeed, under $\mathcal{H}_0$ the following equality holds

$$\mathbb{E}\left[X_{\mathscr{L}}\right] = m_{0,\mathscr{L}} = \left(\prod_{l\in\mathscr{L}} \lambda_l\right) I(L,0).$$

Replacing each spiking intensity λ_l by

$$\hat{\lambda}_l := \frac{1}{M(b-a)} \sum_{k=1}^{M} N_l^{(k)}\left([a,b]\right),$$

where $N_l^{(k)}\left([a,b]\right)$ denotes the number of spikes in $[a,b]$ for neuron l during the k^{th} trial, gives the following estimator,

$$\hat{m}_{0,\mathscr{L}} = \left(\prod_{l\in\mathscr{L}} \hat{\lambda}_l\right) I\left(L,0\right). \tag{5.4}$$

Note that $\bar{m}_{\mathscr{L}}$ is always consistent (that is, converges towards the true parameter) whereas $\hat{m}_{0,\mathscr{L}}$ is consistent under $\mathcal{H}_0$. This leads to the following independence test: the independence assumption is rejected when the difference between $\bar{m}_{\mathscr{L}}$ and $\hat{m}_{0,\mathscr{L}}$ is too large. More precisely, Theorem 5.3.3 gives the asymptotic behaviour of $\sqrt{M}\left(\bar{m}_{\mathscr{L}} - \hat{m}_{0,\mathscr{L}}\right)$ under $\mathcal{H}_0$.

Theorem 5.3.3. *Under the notations and assumptions of Theorem 5.3.1, and under $\mathcal{H}_0$, the following affirmations are true*

- *The following convergence of distribution holds:*

$$\sqrt{M}\left(\bar{m}_{\mathscr{L}} - \hat{m}_{0,\mathscr{L}}\right) \underset{M\to\infty}{\xrightarrow{\mathcal{D}}} \mathcal{N}\left(0,\sigma^2\right),$$

  with

$$\sigma^2 = \mathbf{V}\mathrm{ar}(X_{\mathscr{L}}) - (b-a)^{-1}\mathbb{E}\left[X_{\mathscr{L}}\right]^2 \left(\sum_{l\in\mathscr{L}} \lambda_l^{-1}\right).$$

- *Moreover, σ^2 can be estimated by*

$$\hat{\sigma}^2 = \hat{v}\left(X_{\mathscr{L}}\right) - (b-a)^{-1} I(L,L) \prod_{l\in\mathscr{L}} \hat{\lambda}_l^2 \left(\sum_{l\in\mathscr{L}} \hat{\lambda}_l^{-1}\right),$$

  where

$$\hat{v}(X_{\mathscr{L}}) = \hat{m}_{0,\mathscr{L}} + \sum_{k=1}^{L-1} \left(\sum_{\substack{\mathscr{J}\subset\mathscr{L} \\ \#\mathscr{J}=k}} \prod_{j\in\mathscr{J}} \hat{\lambda}_j^2 \prod_{l\notin\mathscr{J}} \hat{\lambda}_l \right) I(L,k),$$

  and the following convergence of distribution holds:

$$\sqrt{M}\frac{\left(\bar{m}_{\mathscr{L}} - \hat{m}_{0,\mathscr{L}}\right)}{\sqrt{\hat{\sigma}^2}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,1\right).$$

The proof of this theorem relies on a standard application of the delta method [27] and is given in Appendix 5.8.b). The delta method is useful in order to deal with the plug-in step, i.e. the substitution of the real parameters by the estimated ones.

Note that the results obtained in Theorems 5.3.1 and 5.3.3 are true for more general delayed coincidence counts. A more general result and its proof are given in Appendix. However when one considers more general ways to count coincidences the integrals $I(L,k)$ are harder to compute.

5.3.b) Independence test

The results obtained in Theorem 5.3.3 allow us to straightforwardly build a test for detecting a dependency between neurons:

Definition 5.3.4 (The GAUE test)**.** *For α in $]0,1[$, denote z_α the α-quantile of the standard Gaussian distribution $\mathcal{N}(0,1)$. Then the symmetric test of level α rejects $\mathcal{H}_0$ when $\bar{m}$ and $\hat{m}_{0,\mathscr{L}}$ are too different, that is when*

$$\left|\sqrt{M}\frac{(\bar{m}_{\mathscr{L}}-\hat{m}_{0,\mathscr{L}})}{\sqrt{\hat{\sigma}^2}}\right| > z_{1-\alpha/2}.$$

Note that once a subset is rejected by our test, one can determine if the dependency is rather excitatory or inhibitory according to the sign of $\bar{m}_{\mathscr{L}}-\hat{m}_{0,\mathscr{L}}$. If $\bar{m}_{\mathscr{L}}-\hat{m}_{0,\mathscr{L}}>0$ (*respectively* <0) then the dependency is rather excitatory (*respectively inhibitory*).

The result of a test may be wrong in two distinct manners. On the one hand, a false positive is an error in which the test is incorrectly rejecting the null hypothesis. On the other hand, a *false negative* is an error in which the test is incorrectly accepting the null hypothesis. The false positive (respectively negative) rate is the test's probability that a false positive (resp. negative) occurs. Usually, a theoretical control is given only for the false positive rate which is considered as the worst error. The following corollary is an immediate consequence of Theorem 5.3.3 and states the appropriateness of the GAUE test.

Corollary 5.3.5. *Under assumptions of Theorem 5.3.3, the test of level α presented in Definition 5.3.4 is asymptotically of false positive rate α. That is, the false positive rate of the test tends to α when the sample size M tends to infinity.*

5.4 Illustration Study: Poissonian Framework

In this section, an illustration of the previous theoretical results is given. To obtain a global evaluation of the performance of the different methods, some parameters can randomly fluctuate. More precisely, the following procedure is applied,

P

1. Generate a set of random parameters according to the appropriate Framework;
2. Use this set to generate M trials;
3. Compute the different statistics;
4. Repeat steps 1 to 3 a thousand times.

We begin by an illustration of the results of Theorem 5.3.3 and Corollary 5.3.5, and a comparison with the original UE method.

5.4.a) Illustration of the asymptotic properties

The control on the false positive rate of our test being only asymptotic, it is evaluated on simulations in this Section. Moreover, it is shown that our test is empirically *conservative*, that is, when constructed for a prescribed level, say α, the empirical false positive rate is less than α. We simulate independent Poisson processes under the following Framework ($\mathbf{F}_1$) :

$\mathbf{F}_1$

- the trial duration $(b-a)$ is randomly selected (uniform distribution) between 0.2s and 0.4s;
- the $n = 4$ neurons are simulated with different intensities. Each one is randomly selected (uniform distribution) between 8 and 20Hz;
- the set of tested neurons is given by $\mathscr{L} = \{1, 2, 3, 4\}$.

Moreover, we set once and for all $\delta = 0.01s$. Note that the dependence with respect to the parameter δ has been fully discussed in [5].

Considering M independent trials of n point processes, the asymptotic (with respect to M) of the delayed coincidence count is studied. To this aim, we use a Monte Carlo method following the procedure **P** presented at the beginning of Section 5.4. On each simulation, M independent trials are generated and the statistic $S_i = \sqrt{M}\left(\bar{m}_{\mathscr{L},i} - \hat{m}_{0,i}\right)/\sqrt{\hat{\sigma}_i^2}$ (for i from 1 to 1000) is computed. Theorem 5.3.3 tells us that the random variables S_i should be asymptotically distributed as the standard Gaussian distribution. Thus, we plot (Figure 5.4.A) the Kolmogorov distance $KS(F_{M,1000}, F)$ between the empirical distribution function over the 1000 repetition $F_{M,1000}$ and the standard Gaussian distribution function F:

$$KS(F_{M,1000}, F) = \sup_x |F_{M,1000}(x) - F(x)|.$$

Usually, a test of level $\alpha = 0$ always accepts, whereas a test of level $\alpha = 1$ always rejects. Hence, there is a critical value (depending on the observations, and called *p-value*)

for which the test decision passes from acceptance to rejection. If the false positive rate of a test of level α is exactly α for all α in $[0, 1]$, which should asymptotically be the case according to Corollary 5.3.5, then one can prove that the corresponding p-value is uniformly distributed on $[0, 1]$ under the null hypothesis. Thus, the evolution (with respect to M) of the Kolmogorov distance between the empirical distribution function of the obtained p-values (with our test and the one given by the UE method) and the uniform distribution function is plotted for symmetric tests (See Figure 5.4.B). It appears that the rate of convergence of the empirical distribution function of the p-values is faster for our test than the one given by the UE method.

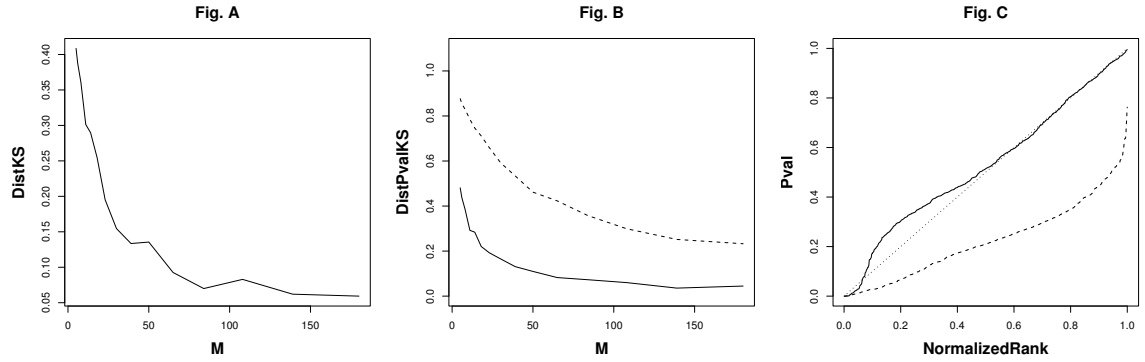


Figure 5.4: Under Framework $\mathbf{F}_1$ (Independence assumption). **Figure A.** Evolution of the Kolmogorov distance (in function of the number of trials) averaged on 1000 simulations between the empirical distribution function of the test statistics and the standard Gaussian distribution function. **Figure B.** Evolution of the Kolmogorov distance averaged on 1000 simulations between the empirical distribution function of the p-values and the uniform distribution function with respect to the number of trials. The plain line stands for our test and the dashed line for the original UE one. **Figure C.** Graphs of the sorted 1000 p-values (for 50 trials) in function of their normalized rank under $\mathcal{H}_0$. The plain line stands for our test, the dashed line for the original UE one and the dotted line for the uniform distribution function.

From Figures 5.4.A and B, it seems reasonable to consider, for our test, sample sizes M greater than 50. Indeed, one sees that the distribution of our statistic is then almost Gaussian and the distribution of the p-values almost uniform (as expected under the null hypothesis). Thus, in order to describe more precisely what happens, we plot in Figure 5.4.C the sorted p-values in function of their normalized rank for $M = 50$. Note that if the curve of sorted p-values is below (respectively above) the diagonal, then the observed p-values are globally smaller (respectively greater) than they should be under $\mathcal{H}_0$. Our test seems to be conservative except for big or very small p-values. The problem induced by this non conservativeness for very small p-values is detailed at the end of Section 5.5. On the other side, the false positive rate observed for the UE test is too high. For example, we see in the figure that the UE test with a theoretical test level of 5% rejects almost 20% of the cases.

5.4.b) Parameter Scan

Here is illustrated the influence of the parameters λ (the firing rate) and $b-a$ (the trial duration). We plot in Figure 5.5 the evolution (with respect to M) of the Kolmogorov distance between the empirical distribution function of the obtained p-values (with our test and the one given by the UE method) and the uniform distribution function.

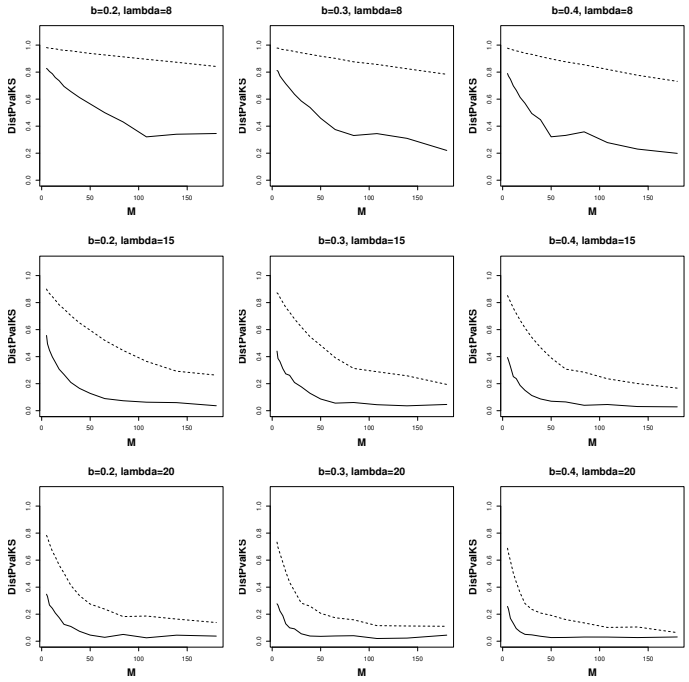


Figure 5.5: Under Framework $\mathbf{F}_1$ (Independence assumption). Evolution of the Kolmogorov distance averaged on 1000 simulations between the empirical distribution function of the p-values and the uniform distribution function with respect to the number of trials. The plain line stands for our test and the dashed line for the original UE one. Each plot stands for different values of λ and b (we set $a = 0$ so that b gives the trial duration $b-a$). From top to bottom λ takes the values 8, 15 and 20Hz. From left to right, b takes the values 0.2, 0.3 and 0.4s.

First of all, note that, if the Kolmogorov distance between the empirical distribution function of the obtained p-values and the uniform distribution function tends to 0, then it means that the false positive rate of the test of level α tends to α when the sample size tends to infinity. As predicted by Corollary 5.3.5, the Kolmogorov distance between the empirical distribution function of the obtained p-values with our test tends to 0 fast enough if λ is not too small ($\lambda \geq 15$Hz). The test induced by the UE method seems to share the same asymptotic behaviour, but with a slower rate of convergence. Finally, it seems that our method performs better than the UE method in all the configurations of parameters.

To describe more precisely what happens, we plot in Figure 5.6 the sorted p-values in function of their normalized rank (for $M = 50$). As expected in regard of Figure 5.5, the plain line sticks to the first diagonal when the parameters are large enough, since, in those cases, the KS distance between the empirical distribution function of the p-values of our test and the uniform distribution function was already small for $M = 50$. However, the test induced by the UE method does not respect the prescribed level even when the parameters

are large. Indeed, the dashed line remains under the diagonal in all cases. Thus, even if the asymptotic level of the UE method is good, in the practical cases where the sample size is small, the false positive rate is not guaranteed.

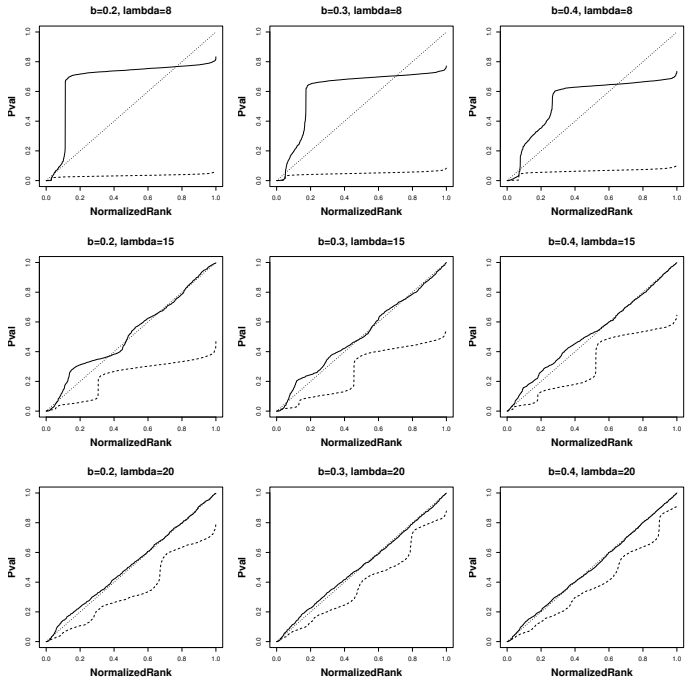


Figure 5.6: Under Framework $\mathbf{F}_1$ (Independence assumption). Graphs of the sorted 1000 p-values (for 50 trials) in function of their normalized rank under $\mathcal{H}_0$. The plain line stands for our test, the dashed line for the original UE one and the dotted line for the uniform distribution function. Each plot stands for different values of λ and b (we set $a = 0$ so that b gives the trial duration $b - a$). From top to bottom λ takes the values 8, 15 and 20Hz. From left to right, b takes the values 0.2, 0.3 and 0.4s.

5.4.c) Illustration of the true positive rate

First, let us note that the *true positive rate* of a test is the test's probability of correctly rejecting the null hypothesis. No theoretical result on this rate can be obtained from Theorem 5.3.3 who deals only with the false positive rate. So, in order to evaluate the true positive rate of the test, we simulate a sample which is dependent and check how many times the test rejects $\mathcal{H}_0$.

To obtain dependent Poisson processes an injection model inspired by the one used in [64, 66] or [158] is used. Consider independent homogeneous Poisson processes $\hat{N}_1, \ldots, \hat{N}_n$, drawn according to Framework $\mathbf{F}_1$. Then, simulate an other Poisson process (according to the same framework but independent from the previous ones) $\tilde{N}$, with an intensity of 0.3Hz, which is injected to every neuron. Thus our sequence of dependent Poisson processes is given by

$$N_i = \hat{N}_i \cup \tilde{N}.$$

This new framework ($\mathbf{F}_1$ completed by the injection) is referred as Framework $\mathbf{F}_2$ and described below :

- the trial duration $(b-a)$ is randomly selected (uniform distribution) between 0.2s and 0.4s;
- the $n = 4$ neurons are simulated with different intensities. Each one is randomly selected (uniform distribution) between 8 and 20Hz. The points of the auxiliary Poisson process $\tilde{N}$ (with intensity 0.3Hz) are injected to every neurons;
- the set of tested neurons is given by $\mathscr{L} = \{1, 2, 3, 4\}$.

} $\mathbf{F}_2$

Note that this injection model can only model excess of coincidences and not lack of coincidences. In the injection model used in [66], a small jitter is applied before injection to mimic temporal imprecision of the synchronous event. In our Poissonian framework this jitter cannot be performed in a similar way. Indeed, this jitter does not preserve the stationariness of the Poisson process near the edges. Although some other more elaborate injection models are available in the Poissonian framework, we do not use one of them here because their translation in the discrete time framework is not clear.

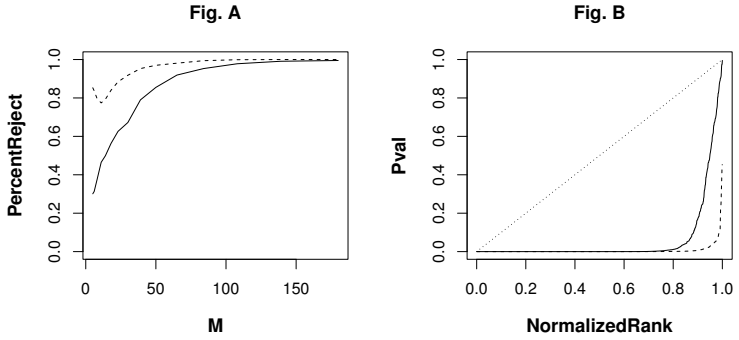


Figure 5.7: Under Framework $\mathbf{F}_2$ (Dependence assumption). **Figure A.** Illustration of the true positive rate of the test, for a theoretical test level of 5%. The curves represent the evolution, with respect to the number of trials, of the true positive rate (averaged on 1000 simulations). The plain line stands for our test and the dashed line for the original UE one. **Figure B.** Graphs of the sorted 1000 p-values for dependent Poisson processes (50 trials). The plain line stands for our test, the dashed line for the original UE one and the dotted line for the uniform distribution function.

For a fixed theoretical level of $\alpha = 5\%$, Figure 5.7.A illustrates the true positive rate of the two tests in function of the number of trials M. Then Figure 5.7.B represents the p-values as a function of their normalized rank, for $M = 50$. As in Figure 5.4.C, the lower the curve is, the greater the observed frequency of rejection of the null hypothesis. The true positive rate is higher for the UE method for small sample sizes, but this is at the price of an undervalued theoretical level. Indeed, we saw previously (Section 5.4.a)) that the UE test gives too much false positives (for $M = 50$, 20% of rejection under the null hypothesis with a theoretical test level of 5%).

5.5 Illustration Study: Non-Poissonian framework

In this section, a more neurobiologically realistic framework than the Poisson one is considered. Indeed, it is interesting to see if our test is still reliable when the Poisson framework

is not valid any-more. Our test is confronted to multivariate Hawkes processes, which can be simulated thanks to Ogata's thinning method [110] inspired by [92]. The use of Hawkes processes in neurobiology was first introduced in [31]. With the development of simultaneous neuron recordings there is a recent trend in favour of Hawkes processes for modelling spike trains ([88, 119, 120, 123, 158]). Furthermore, Hawkes processes have passed some goodness-of-fit tests on real data [136]. In this model, interaction between two neurons can be easily and in a more realistic way inserted. This is one of the reasons of this trend. Note that the homogeneous Poisson process is a particular case of Hawkes processes, with no interaction between neurons.

A counting process N is characterized by its conditional intensity λ_t which is related with the local probability of finding a new point given the past. (Informally, the quantity $\lambda_t dt$ gives the probability that a new point on N appears in $[t, t+dt]$ given the past). The process $(N^i)_{i=1\dots n}$ is a multivariate Hawkes process if there exist some functions $(h_{ij})_{i,j=1\dots n}$ (called interaction functions) and some positive constants $(\mu_i)_{i=1\dots n}$ (spontaneous intensities) such that, for all $j = 1, \dots, n$, λ^j given by

$$\lambda_t^j = \max\left(0, \mu_j + \sum_{i=1}^{n} \int_{s<t} h_{ij}\,(t-s)\; N^i\,(ds)\right)$$

is the intensity of the point process N^j, where $N^i(ds)$ is the point measure associated to N^i, that is $N^i(ds) = \sum_{T \in N^i} \delta_T(ds)$ where δ_T is the Dirac measure at point T.

The functions h_{ij} represent the influence of neuron i over neuron j in terms of spiking intensity. This influence can be either exciting ($h \geq 0$) or inhibiting ($h \leq 0$). For example, suppose that $h_{ij} = \beta \mathbb{1}_{[0,x]}$. If $\beta > 0$ (*respectively* $\beta < 0$) then the apparition of a spike on N^i increases (*respectively decreases*) the probability to have a spike on N^j during a short period of time (namely x): neuron i excites (*respectively inhibits*) neuron j. The processes N^i for $i = 1, \dots, n$ are independent if and only if $h_{ij} = 0$ for all $i \neq j$.

Note also that the self-interaction functions h_{jj} can model refractory periods, making the Hawkes model more realistic than Poisson processes, even in the independence case. In particular when $h_{jj} = -\mu_j \mathbb{1}_{[0,x]}$, all the other interaction functions being null, the n-dimensional process is composed by n independent Poisson processes with dead time x, modelling strict refractory periods of length x [133].

All the following tests are computed according to the Framework $\mathbf{F}_3$ below:

$\mathbf{F}_3$:
- the trial duration of $b-a$ is randomly selected (uniform distribution) between 0.2 and 0.4s;
- the $n = 4$ neurons are simulated with spontaneous intensity $\mu_1, \dots, \mu_4$ randomly selected (uniform distribution) between 8 and 20Hz;
- the non-positive auto interaction functions are given by $h_{i,i} = -\mu_i \mathbb{1}_{[0,0.003s]}$;
- the set of tested neurons is given by $\mathscr{L} = \{1, 2, 3, 4\}$.

We also performed a parameter scan. However, since the results are equivalent to those obtained in the Poissonian framework, they are not presented here.

5.5.a) Illustration of the level

Before all, one wants to know if Theorem 5.3.3 and Corollary 5.3.5 are still reliable for Hawkes processes. Thus as in section 5.4, Figure 5.8.A shows the evolution of the KS distance between $F_{M,1000}$ and F. Then as in Section 5.4, we look at the KS distance between the empirical distribution function of the p-values and the uniform distribution function to see if one can trust the level of the different tests (Figure 5.8.B). These two figures are pretty similar to Figures 5.4.A and B (Poissonian case), but with a slightly slower convergence rate (with respect to M). Finally, Figure 5.8.C plays the same role as Figure 5.4.C and presents the sorted p-values in function of their normalized rank (for $M = 50$). Again, the results are comparable to those obtained in the Poissonian case: our test is rather conservative whereas the UE test rejects too many cases.

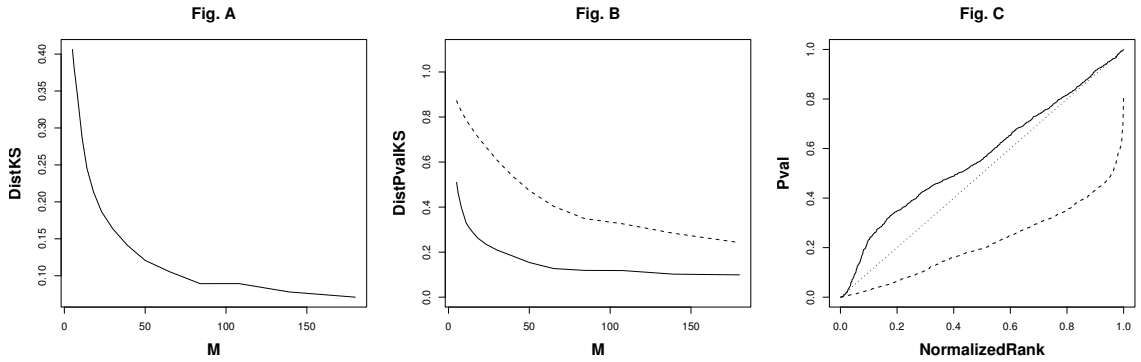


Figure 5.8: Under Framework $\mathbf{F}_3$ (Independence assumption). **Figure A.** Evolution of the Kolmogorov distance (in function of the number of trials) averaged on 1000 simulations between the empirical distribution function of the test statistics and the standard Gaussian distribution function. **Figure B.** Evolution of the Kolmogorov distance averaged on 1000 simulations between the empirical distribution function of the p-values and the uniform distribution function with respect to the number of trials. The plain line stands for our test and the dashed line for the original UE one. **Figure C.** Graphs of the sorted 1000 p-values (for 50 trials) in function of their normalized rank under $\mathcal{H}_0$. The plain line stands for our test, the dashed line for the original UE one and the dotted line for the uniform distribution function.

5.5.b) Illustration of the true positive rate

As said previously, it is more realistic to introduce dependency between Hawkes processes than Poisson processes. Still considering Framework $\mathbf{F}_3$, interaction functions $h_{i,j} = \beta \mathbb{1}_{[0,0.005s]}$, β being randomly selected between 20 and 30Hz, are added. More precisely, we add five interaction functions: $h_{1,3}$, $h_{2,3}$, $h_{1,4}$, $h_{2,4}$ and $h_{3,4}$ (summarized in Figure 5.9). Moreover, the auto interactions are updated to preserve strict refractory periods : $h_{i,i} = -(\mu_i + m_i.\beta)\mathbb{1}_{[0,0.003s]}$, where m_i is the number of neurons exciting neuron i (for example, $m_4 = 3$). This new framework ($\mathbf{F}_3$ completed by the five interaction function) is referred as Framework $\mathbf{F}_4$ and described below :

- the trial duration of $b-a$ is randomly selected (uniform distribution) between 0.2 and 0.4s;
- the $n = 4$ neurons are simulated with spontaneous intensity $\mu_1, \ldots, \mu_4$ randomly selected (uniform distribution) between 8 and 20Hz;
- the non-positive auto interaction functions are given by $h_{i,i} = -(\mu_i + m_i.\beta)\mathbb{1}_{[0,0.003s]}$ (the m_i's are defined above);
- the strength of interaction β is uniformly selected between 20 and 30Hz. There are five non-null interaction functions ($h_{1,3}$, $h_{2,3}$, $h_{1,4}$, $h_{2,4}$ and $h_{3,4}$) which are all equal to $x \mapsto \beta\mathbb{1}_{x\in[0,0.005s]}$;
- the set of tested neurons is given by $\mathscr{L} = \{1, 2, 3, 4\}$.

} $\mathbf{F}_4$

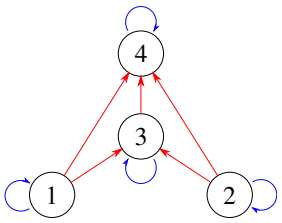


Figure 5.9: Local independence graph. An arrow means a non null interaction function. Blue arrow means inhibition and red arrow means excitation.

As previously, we first provide an illustration of the true positive rate of the two tests, associated to a theoretical level of 5%, in function of M (Figure 5.10.A). Then Figure 5.10.B represents the p-values in function of their normalized rank, for $M = 50$. The difference between the true positive rates is smaller than in the Poissonian Case.

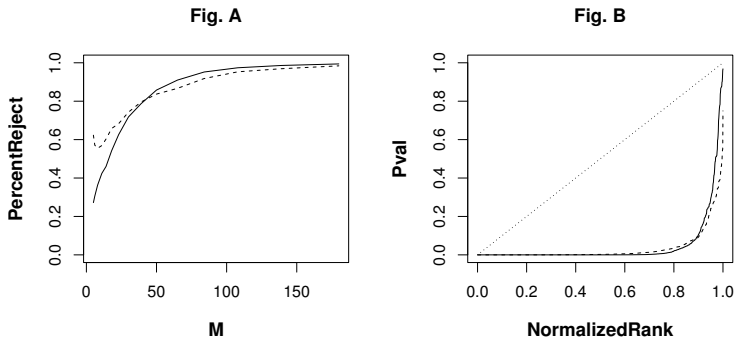


Figure 5.10: Under Framework $\mathbf{F}_4$ (Dependence assumption, see Figure 5.9). **Figure A.** Illustration of the true positive rate of the test, for a theoretical test level of 5%. The curves represent the evolution, with respect to the number of trials, of the true positive rate (averaged on 1000 simulations). The plain line stands for our test and the dashed line for the original UE one. **Figure B.** Graphs of the sorted 1000 p-values for 50 trials. The plain line stands for our test, the dashed line for the original UE one and the dotted line for the uniform distribution function.

5.5.c) Multiple pattern test

In the original MTGAUE method, a multiple testing procedure is applied with respect to 1900 sliding time windows. In our framework, we cannot guarantee the relevance of the multiple test with this high order of multiplicity. This is due to the default of the Gaussian approximation and, more precisely, to the excess of very small p-values as noted in Section 5.4.a). But, we are able to propose a multiple testing procedure with respect to the different possible patterns. For example, with four neurons there are eleven different possible patterns, which gives a much lower order of multiplicity. So, the multiple test over all the eleven sub-pattern of two, three or four neurons is presented here.

In multiple testing, the notion of false positive rate is not relevant. The closest notion might be the *Family-Wise Error Rate* (FWER) which is the probability to wrongly reject at least one of the tests. This error rate can be controlled using Bonferroni's method but it is too restrictive, in particular when the number K of tests involved is too large. One popular way to deal with multiple testing is the Benjamini-Hochberg procedure [11] which ensures a control of the *False Discovery Rate* (FDR). False discoveries cannot be avoided but it is not a problem if the ratio of F_p (the number of false positives detections) divided by R (the total number of rejects) is controlled. Therefore, the FWER and the FDR are mathematically defined by $\text{FWER} = \mathbb{P}(F_p > 0)$ and $\text{FDR} = \mathbb{E}\left[F_p/R\ \mathbb{1}_{R>0}\right]$.

Note that in the full independent case, the FWER and the FDR are equal. The following procedure, due to Benjamini and Hochberg ensures a small FDR over K tests:

1. Fix a level q ($q = 5\%$ for example);
2. Denote by $(P_1, \dots, P_K)$ the p-values obtained for all considered tests;
3. Order them in increasing order and denote the increasing vector $(P_{(1)}, \dots, P_{(K)})$;
4. Note k_0 the largest k such that $P_{(k)} \leq kq/K$;
5. Then, reject all the tests corresponding to p-values smaller than $P_{(k_0)}$.

The theoretical result of [11] ensures that if the p-values are upper bounded by a uniform distribution and independently distributed under the null hypothesis, then the procedure guarantees a FDR less than q. The main drawback of this procedure in our case is that one needs to compute p-values that are very small when K is large. For example, if $K \geq 50$ and $q = 5\%$, the upper bound given by kq/K can be smaller than 0.001 and as noted in Section 5.4.a) the empirical frequency of very small p-values is greater than expected and therefore the uniform upper bound of the p-values is not guaranteed in our case. However, only 11 tests are considered here and the procedure still returns reliable results.

We perform 1000 simulations and count how many times each test rejects the independence. The results, obtained for $M = 50$, are presented in Figure 5.11. The results show that our test detects all patterns except $\{1, 2\}$. This is consistent with the considered framework ($\mathbf{F}_4$) since we simulate connections between all pairs of neurons except $\{1, 2\}$. The U.E. test essentially detects the patterns $\{2, 3, 4\}$, $\{1, 3, 4\}$, $\{1, 2, 3, 4\}$ and to a lesser extent $\{1, 2, 4\}$ and $\{1, 2, 3\}$. Moreover, it misses all the pairs.

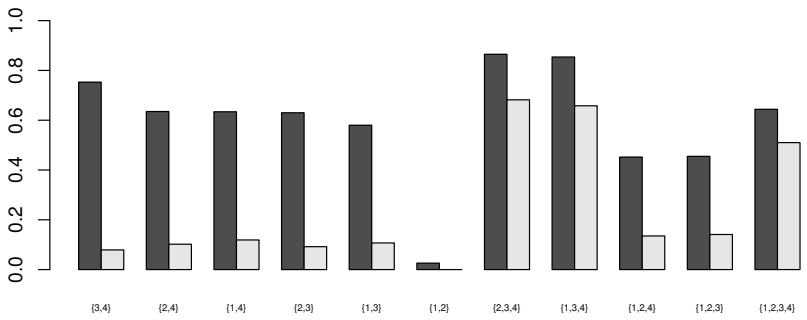


Figure 5.11: Under Framework $\mathbf{F}_4$ (Dependence assumption, see Figure 5.9). Frequency of dependence detections (1000 simulations) for each pattern. Grey for our test, white for the original UE method.

5.6 Illustration on real data

After validating our test on simulations, we apply our method on real data and show results in agreement with classical knowledge on those data.

5.6.a) Description of the data

The data set considered here is the same as in [158] and previous experimental studies [61, 141, 142]. The following description of the experiment is copied from Section 4.1 of [158]. These data were collected on a 5-year-old male Rhesus monkey who was trained to perform a delayed multi-directional pointing task. The animal sat in a primate chair in front of a vertical panel on which seven touch-sensitive light-emitting diodes were mounted, one in the center and six placed equidistantly (60 degrees apart) on a circle around it. The monkey had to initiate a trial by touching and then holding with the left hand the central target. After a fix delay of 500ms, the preparatory signal (PS) was presented by illuminating one of the six peripheral targets in green. After a delay of either 600ms (with probability 0.3) or 1200ms (with probability 0.7), it turned red, serving as the response signal and pointing target. Signals recorded from up to seven micro-electrodes (quartz insulated platinum-tungsten electrodes, impedance: 2-5MΩ at 1000Hz) were amplified and band-pass filtered from 300Hz to 10kHz. Using a window discriminator, spikes from only one single neuron per electrode were then isolated. Neuronal data along with behavioural events (occurrences of signals and performance of the animal) were stored on a PC for off-line analysis with a time resolution of 10kHz. The idea of the analysis is to detect some conspicuous patterns of coincident spike activity appearing during the response signal in the case of a long delay (1200ms). Therefore, we only consider trials where the response signal is indeed occurring after a long delay.

5.6.b) The test

We have at hand the following data: spike trains associated to four neurons (35 trials by neurons). We consider two sub windows: one between 300ms and 500ms (i.e. before the

preparatory signal), the other between 1100ms and 1300ms (i.e. around the expected signal). Our idea is that more synchronisation should be detected during the second window. Moreover, we do not only want to test if the four considered neurons are independent (that is perform our test on the complete pattern $\{1, 2, 3, 4\}$). Indeed one can be interested in knowing if neurons in some sub-patterns (for example $\{1, 2\}$ or $\{1, 3, 4\}$ are independent. That is why we use the multiple pattern test procedure defined at the end of Section 5.5 to test all the eleven subsets (of at least two neurons) of the four considered neurons are tested. Thus we use the Benjamini-Hochberg procedure (presented in the previous section) for $K = 22$ tests. Moreover, we took several values for the delay δ between 0.01s and 0.025s and the results remained stable.

The results are presented in Figure 5.12. Note that we saw in sections 5.4 and 5.5 that our test is too conservative even for small number of trials. This ensures that the theoretical level of our test can be trusted. We see that synchronizations between the subsets $\{3, 4\}$ and $\{1, 3, 4\}$ appear in the second window. These results suggest that neurons 1, 3 and 4 belong to a neuronal assembly which is formed around the expected signal. This is in agreement with more quantitative results on those data [61, 158].

Figure 5.12: Evolution of the synchronization between neurons. The lines indicate the subset for which our test detects dependence. Here we detect an excess of coincidences between neurons $\{1, 3, 4\}$ and $\{3, 4\}$

5.7 Conclusion

This chapter generalizes the statistical study of the delayed coincidence count performed in [158] to more than two neurons. This delayed coincidence count leads to an independence test for point processes which are commonly used to model spike trains.

Under the hypothesis that the point processes are homogeneous Poisson processes, the expectation and variance of the delayed coincidence count can be computed (Theorem 5.3.1), and then a test with prescribed asymptotic level is built (Theorem 5.3.3). A simulation study allows us to confirm our theoretical results and to state the empirical validity of our test with a relaxed Poisson assumption. Indeed, we considered Hawkes processes which are a more realistic model of spike trains. The simulation study gives good results, even for small sample size. This allows us to use our test on real data, in order to highlight the emergence of a neuronal assembly involved at some particular time of the experiment.

We achieved the full generalization of the single test procedure introduced in [158]. However, we could not achieve the multiple time windows testing procedure mainly because of the default of Gaussian approximation concerning extreme values of the test statistics. More precisely, very small p-values are not distributed as expected. In particular, as noted at the end of Section 5.4.a), when the sample size M is moderate ($M = 50$), our test returns too many very small p-values. In [158], the MTGAUE method is applied simultaneously on 1900 sliding windows. In the present work, in order to apply multiple testing both with respect to the sliding time windows and the subsets, the total number of tests is even larger. Indeed, for each sliding window, there are $2^n - n - 1$ tests to perform, where n is the number of recorded neurons. As said at the end of Section 5.5 this would lead to extremely small p-values, for which our test is less reliable.

Even if our test remains empirically reliable under a non Poissonian framework, it could be therefore of interest to explore surrogate data method such as trial-shuffling [124]. A very recent work based on permutation approach for delayed coincidence count with $n = 2$ neurons [5] is a first step in this direction but needs to be generalized to more than 2 neurons.

5.8 Proofs

As said in Section 5.3.a), we prove more general results than Theorems 5.3.1 and 5.3.3. Considering $N_1, \dots, N_n$, some point processes on $[a,b]$ and $\mathscr{L} \subset \{1,\dots,n\}$ a set of indices with cardinal $L \geq 2$, we prove the same kind of results with any coincidence function $c(x_1,\dots,x_L)$ with value either 0 or 1 satisfying Definition 5.8.1 below.

Definition 5.8.1.

1. *A coincidence function is a function* $c : [a,b]^L \to \{0,1\}$ *which is symmetric.*

2. *Let* $(x_1,\dots,x_L) \in \prod_{l\in\mathscr{L}} N_l$ *be a* L*-tuple with a spiking time of every neuron of the subset* $\mathscr{L}$*. Say that* $(x_1,\dots,x_L)$ *is a coincidence if and only if* $c(x_1,\dots,x_L) = 1$*.*

3. *Given* c *a coincidence function, we define* $X_{\mathscr{L}}$ *the number of coincidences on* $[a,b]$ *by:*
$$X_{\mathscr{L}} = \sum_{(x_1,\dots,x_L)\in\prod_{l\in\mathscr{L}} N_l} c(x_1,\dots,x_L).$$

4. *Define*
$$\forall k \in \{0,\dots,L\},\ I(L,k) = \int_{[a,b]^{L-k}} \left(\int_{[a,b]^k} c(x_1,\dots,x_L)\, dx_1 \dots dx_k \right)^2 dx_{k+1}\dots dx_L$$
where the convention $\int_{[a,b]^0} f(x)\,dx = f(x)$ *is set.*

5.8.a) Proof of Theorem 5.3.1

Theorem 5.8.2. *Under assumptions and notations of Definition 5.8.1, if* $N_1,\dots,N_n$ *are some independent homogeneous Poisson processes on* $[a,b]$ *with intensities* $\lambda_1,\dots,\lambda_n$*, the expected value and the variance of the number of coincidences* $X_{\mathscr{L}}$ *are given by:*
$$m_{0,\mathscr{L}} := \mathbb{E}[X_{\mathscr{L}}] = \left(\prod_{l\in\mathscr{L}} \lambda_l\right) I(L,0) \tag{5.5}$$
and
$$\mathbf{V}\mathrm{ar}(X_{\mathscr{L}}) = m_{0,\mathscr{L}} + \sum_{k=1}^{L-1} \left(\sum_{\substack{\mathscr{J}\subset\mathscr{L} \\ \#\mathscr{J}=k}} \prod_{j\in\mathscr{J}} \lambda_j^2 \prod_{l\notin\mathscr{J}} \lambda_l \right) I(L,k). \tag{5.6}$$

Proof. By definition,
$$\mathbb{E}[X_{\mathscr{L}}] = \mathbb{E}\left[\sum_{(x_1,\dots,x_L)\in\prod_{l\in\mathscr{L}} N_l} c(x_1,\dots,x_L) \right].$$

Using the fact that $N_1, \cdots, N_n$ are independent homogeneous Poisson processes with respective intensities $\lambda_1, \cdots, \lambda_n$ one can prove (see [37]) that

$$\mathbb{E}\left[X_{\mathscr{L}}\right] = \left(\prod_{l\in\mathscr{L}} \lambda_l\right) \int\limits_{[a,b]^L} c\left(x_1, \ldots, x_L\right)\ dx_1 \ldots dx_L = \left(\prod_{l\in\mathscr{L}} \lambda_l\right) I(L,0).$$

For sake of simplicity, the variance is computed in the simpler case where $\mathscr{L} = \{1, ..., L\}$, the generalization being pretty clear. In order to simplify, we use the integral form of the coincidence count, i.e.

$$X_{\mathscr{L}} = \int\limits_{[a,b]^L} c\left(x_1, \ldots, x_L\right)\ dN_1(x_1) \ldots dN_L(x_L)$$

where $dN_1, \ldots, dN_L$ are the point measures associated to $N_1, \ldots, N_L$. Thanks to Fubini Theorem we have

$$\mathbb{E}\left[X_{\mathscr{L}}^2\right] = \mathbb{E}\left[\int\limits_{[a,b]^{2L}} c\left(x_1, \ldots, x_L\right) c\left(y_1, \ldots, y_L\right)\ \prod_{l=1}^{L} dN_l\left(x_l\right) dN_l\left(y_l\right)\right]. \tag{5.7}$$

Then, let us define

$$[a,b]^{(1)} = \left\{(x,y) \in [a,b]^2 \mid x = y\right\} \text{ and } [a,b]^{(2)} = [a,b]^2 \setminus [a,b]^{(1)}. \tag{5.8}$$

Now, let us see that $[a,b]^{2L} = \left([a,b]^2\right)^L = \sqcup_{\varepsilon\in\{1,2\}^L} \left(\prod_{l=1}^{L} [a,b]^{(\varepsilon_l)}\right)$ where ε_l denotes the l-th coordinate of ε. Using this decomposition and Equation (5.7), it is clear that

$$\mathbb{E}\left[X_{\mathscr{L}}^2\right] = \sum_{\varepsilon\in\{1,2\}^L} A_\varepsilon, \tag{5.9}$$

where for all ε in $\{1,2\}^L$,

$$A_\varepsilon = \mathbb{E}\left[\int\limits_{\prod_{l=1}^{L} [a,b]^{(\varepsilon_l)}} c\left(x_1, \ldots, x_L\right) c\left(y_1, \ldots, y_L\right)\ \prod_{l=1}^{L} dN_l\left(x_l\right) dN_l\left(y_l\right)\right].$$

For every $p = 1, \ldots, L$, let $\varepsilon^{(p)} = (1, \ldots, 1, 2, \ldots, 2)$ where the number of 1's in $\varepsilon^{(p)}$ is exactly p. Properties of the moment measure of Poisson processes (see [37] or [86] in a more simplified framework) lead to

$$\begin{aligned} A_{\varepsilon^{(p)}} &= \mathbb{E}\left[\int\limits_{\left([a,b]^{(1)}\right)^p} \int\limits_{\left([a,b]^{(2)}\right)^{L-p}} c\left(x_1, \ldots, x_L\right) c\left(y_1, \ldots, y_L\right)\ \prod_{l=1}^{L} dN_l\left(x_l\right) dN_l\left(y_l\right)\right] \\ &= \prod_{l=1}^{p} \lambda_l \prod_{j=p+1}^{L} \lambda_j^2 \int\limits_{[a,b]^p} \left(\int\limits_{[a,b]^{2(L-p)}} c\left(t_1, \ldots, t_p, x_{p+1}, \ldots, x_L\right)\right. \\ &\qquad \left. c\left(t_1, \ldots, t_p, y_{p+1}, \ldots, y_L\right) \prod_{k=p+1}^{L} dx_k dy_k\right) dt_1 \ldots dt_p. \end{aligned} \tag{5.10}$$

For fixed $(t_1, \dots, t_p)$ one can apply Fubini Theorem to the inner integral which leads to:

$$\begin{aligned} A_{\varepsilon^{(p)}} &= \prod_{l=1}^{p} \lambda_l \prod_{j=p+1}^{L} \lambda_j^2 \int_{[a,b]^p} \left(\int_{[a,b]^{2(L-p)}} c\left(t_1, \dots, t_p, t_{p+1}, \dots, t_J\right) dt_{p+1} \dots dt_L \right)^2 dt_1 \dots dt_p. \\ &= \prod_{l=1}^{p} \lambda_l \prod_{j=p+1}^{L} \lambda_j^2 \; I\left(L, L-p\right). \end{aligned}$$

by definition of $I(L, L-p)$.

For more general vectors ε in $\{1,2\}^L$, let us note p the occurrence count of 1 in the vector ε and I_ε (respectively J_ε) the set of indices of the coordinates of ε equal to 1 (respectively 2). Then, using the symmetry of the coincidence function c, one can easily deduce from the computation of $A_{\varepsilon^{(p)}}$ that

$$A_\varepsilon = \prod_{i \in I_\varepsilon} \lambda_i \prod_{j \in J_\varepsilon} \lambda_j^2 \; I\left(L, L-p\right). \tag{5.11}$$

From (5.9) and (5.11), one deduces

$$\mathbb{E}\left[X_{\mathscr{L}}^2\right] = \sum_{p=0}^{L} \left(\sum_{\substack{\mathscr{J} \subset \mathscr{L} \\ \#\mathscr{J}=p}} \prod_{j \in \mathscr{J}} \lambda_j \prod_{l \notin \mathscr{J}} \lambda_l^2 \right) I(L, L-p).$$

Note that $I(L,L) = I(L,0)^2$ by definition, so the case $p=0$ in the sum corresponds to

$$\prod_{l \in \mathscr{L}} \lambda_l^2 \; I\left(L,L\right) = \prod_{l \in \mathscr{L}} \lambda_l^2 \; I\left(L,0\right)^2 = \mathbb{E}\left[X_{\mathscr{L}}\right]^2.$$

Moreover, the case $p=L$ in the sum corresponds to $\prod_{l \in \mathscr{L}} \lambda_l \; I\left(L,0\right) = \mathbb{E}\left[X_{\mathscr{L}}\right]$. So, we have

$$\mathbb{E}\left[X_{\mathscr{L}}^2\right] = \mathbb{E}\left[X_{\mathscr{L}}\right]^2 + \mathbb{E}\left[X_{\mathscr{L}}\right] + \sum_{p=1}^{L-1} \left(\sum_{\substack{\mathscr{J} \subset \mathscr{L} \\ \#\mathscr{J}=p}} \prod_{j \in \mathscr{J}} \lambda_j \prod_{l \notin \mathscr{J}} \lambda_l^2 \right) I(L, L-p),$$

and (5.6) clearly follows by defining the variable $k = L - p$.

□

Theorem 5.3.1 is a direct consequence of Theorem 5.8.2 since the function $c_\delta : [a,b]^L \to \{0,1\}$ defined by

$$c_\delta(x_1, \dots, x_n) = \mathbb{1}_{\left| \max\limits_{i \in \{1,\dots,L\}} x_i - \min\limits_{i \in \{1,\dots,L\}} x_i \right| \leq \delta}, \; 0 < \delta < \frac{b-a}{2} \tag{5.12}$$

satisfies Definition 5.8.1.

5.8.b) Proof of Theorem 5.3.3

Theorem 5.8.3. *Under Notations and Assumptions of Theorem 5.8.2, the two following affirmations are valid:*

- *The following convergence of distribution holds:*

$$\sqrt{M}\left(\bar{m}_{\mathscr{L}} - \hat{m}_{0,\mathscr{L}}\right) \xrightarrow[M\to\infty]{\mathcal{D}} \mathcal{N}\left(0, \sigma^2\right),$$

where

$$\sigma^2 = \mathbf{V}\mathrm{ar}(X_{\mathscr{L}}) - (b-a)^{-1}\mathbb{E}\left[X_{\mathscr{L}}\right]^2 \left(\sum_{l\in\mathscr{L}} \lambda_l^{-1}\right).$$

- *Moreover, σ^2 can be estimated by*

$$\hat{\sigma}^2 = \hat{v}\left(X_{\mathscr{L}}\right) - (b-a)^{-1} I(L,L) \prod_{l\in\mathscr{L}} \hat{\lambda}_l^2 \left(\sum_{k\in\mathscr{L}} \hat{\lambda}_k^{-1}\right),$$

where

$$\hat{v}(X_{\mathscr{L}}) = \hat{m}_{0,\mathscr{L}} + \sum_{k=1}^{L-1} \left(\sum_{\substack{\mathscr{J}\subset\mathscr{L} \\ \#\mathscr{J}=k}} \prod_{j\in\mathscr{J}} \hat{\lambda}_j^2 \prod_{l\notin\mathscr{J}} \hat{\lambda}_l \right) I(L,k),$$

and

$$\sqrt{M}\frac{\left(\bar{m}_{\mathscr{L}} - \hat{m}_{0,\mathscr{L}}\right)}{\sqrt{\hat{\sigma}^2}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,1\right).$$

Proof. For sake of simplicity, the result is proven in the simpler case where $\mathscr{L} = \{1, ..., L\}$, the generalization being pretty clear. An application of the Central Limit Theorem leads to:

$$\frac{1}{\sqrt{M}} \sum_{k=1}^{M} \left[\begin{pmatrix} X_{\mathscr{L}}^{(k)} \\ N_1^{(k)}\left([a,b]\right) \\ \vdots \\ N_L^{(k)}\left([a,b]\right) \end{pmatrix} - \begin{pmatrix} \mathbb{E}\left[X_{\mathscr{L}}\right] \\ \lambda_1(b-a) \\ \vdots \\ \lambda_L(b-a) \end{pmatrix} \right] \xrightarrow{\mathcal{D}} \mathcal{N}_{L+1}\left(\mathbf{0}, \Gamma\right),$$

where $\mathcal{N}_{L+1}\left(\mathbf{0}, \Gamma\right)$ is the multivariate Gaussian distribution with $L+1$-dimensional mean vector $\mathbf{0}$ and covariance matrix Γ defined by:

$$\Gamma = \begin{pmatrix} \mathbf{V}\mathrm{ar}(X_{\mathscr{L}}) & \mathbb{E}\left[X_{\mathscr{L}}\right] & \cdots & \mathbb{E}\left[X_{\mathscr{L}}\right] \\ \mathbb{E}\left[X_{\mathscr{L}}\right] & \lambda_1(b-a) & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ \mathbb{E}\left[X_{\mathscr{L}}\right] & 0 & 0 & \lambda_L(b-a) \end{pmatrix}$$

The matrix is obtained using the fact that the processes N_l, $l \in \mathscr{L}$ are independent and

from the following computation for all j_0 in $\mathscr{L}$,

$$\begin{aligned}
\mathbb{E}[X_{\mathscr{L}} N_{j_0}([a,b])] &= \mathbb{E}\Big[\int_{[a,b]^{L+1}} c(x_1,\dots,x_L)\; dN_1(x_1)\dots dN_L(x_L) dN_{j_0}(y)\Big] \\
&= \mathbb{E}\Big[\int_{[a,b]^{L-1}} \Big(\int_{[a,b]^{(2)}} c(x_1,...,x_L)\; dN_{j_0}(x_{j_0}) dN_{j_0}(y)\Big) \prod_{l\in\mathscr{L}, l\neq j_0} dN_j(x_j)\Big] \\
&\quad + \mathbb{E}\Big[\int_{[a,b]^{L-1}} \Big(\int_{[a,b]^{(1)}} c(x_1,...,x_L)\; dN_{j_0}(x_{j_0}) dN_{j_0}(y)\Big) \prod_{l\in\mathscr{L}, l\neq j_0} dN_j(x_j)\Big] \\
&= \lambda_{j_0}\Big(\prod_{l\in\mathscr{L}} \lambda_l\Big) \int_{[a,b]\times[a,b]^L} c(x_1,\dots,x_L)\; dx_1\dots dx_L dy + \Big(\prod_{l\in\mathscr{L}} \lambda_l I(L,0)\Big) \\
&\quad \text{(using the same arguments as for (5.10))} \\
&= \lambda_{j_0}(b-a)\Big(\prod_{l\in\mathscr{L}} \lambda_l\Big) I(L,0) + \Big(\prod_{l\in\mathscr{L}} \lambda_l\Big) I(L,0) \\
&= \mathbb{E}[N_{j_0}([a,b])]\mathbb{E}[X_{\mathscr{L}}] + \mathbb{E}[X_{\mathscr{L}}],
\end{aligned}$$

where $[a,b]^{(2)}$ and $[a,b]^{(1)}$ are defined by (5.8) in the previous proof. Define

$$g : (x, u_1, \dots, u_L) \mapsto x - (b-a)^{-L} I(L,0) \prod_{l=1}^{L} u_l,$$

and remark that:

$$\begin{cases} g\left(\frac{1}{M}\sum_{k=1}^{M} X_{\mathscr{L}}^{(k)}, \frac{1}{M}\sum_{k=1}^{M} N_1^{(k)}([a,b]), \dots, \frac{1}{M}\sum_{k=1}^{M} N_L^{(k)}([a,b])\right) = \bar{m}_{\mathscr{L}} - \hat{m}_{0,\mathscr{L}}, \\ g\left(\mathbb{E}[X_{\mathscr{L}}], \lambda_1(b-a), \dots, \lambda_L(b-a)\right) = 0 \quad \text{(thanks to Theorem 5.8.2).} \end{cases}$$

So we have

$$\begin{aligned}
\sqrt{M}\left(\bar{m}_{\mathscr{L}} - \hat{m}_{0,\mathscr{L}}\right) = \sqrt{M}\Bigg[& g\left(\frac{1}{M}\sum_{k=1}^{M} X_{\mathscr{L}}^{(k)}, \frac{1}{M}\sum_{k=1}^{M} N_1^{(k)}([a,b]), \dots, \frac{1}{M}\sum_{k=1}^{M} N_L^{(k)}([a,b])\right) \\
& - g\left(\mathbb{E}[X_{\mathscr{L}}], \lambda_1(b-a), \dots, \lambda_L(b-a)\right)\Bigg].
\end{aligned}$$

And the delta method [27] gives the following convergence of distribution,

$$\sqrt{M}\left(\bar{m}_{\mathscr{L}} - \hat{m}_{0,\mathscr{L}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, {}^tD\Gamma D\right),$$

where D is the gradient of the function g at the point $\left(\mathbb{E}[X_{\mathscr{L}}], \lambda_1(b-a), \dots, \lambda_L(b-a)\right)$ i.e.

$$D = \begin{pmatrix} 1 \\ -\lambda_1^{-1}\mathbb{E}[X_{\mathscr{L}}](b-a)^{-1} \\ \vdots \\ -\lambda_L^{-1}\mathbb{E}[X_{\mathscr{L}}](b-a)^{-1} \end{pmatrix}.$$

So,

$$
\begin{aligned}
{}^tD\Gamma D &= {}^tD \begin{pmatrix} \mathbf{V}\mathrm{ar}(X_{\mathscr{L}}) - (b-a)^{-1}\mathbb{E}\left[X_{\mathscr{L}}\right]^2 \left(\sum_{l\in\mathscr{L}} \lambda_l^{-1}\right) \\ \mathbb{E}\left[X_{\mathscr{L}}\right] - \mathbb{E}\left[X_{\mathscr{L}}\right] \\ \vdots \\ \mathbb{E}\left[X_{\mathscr{L}}\right] - \mathbb{E}\left[X_{\mathscr{L}}\right] \end{pmatrix} \\
&= \mathbf{V}\mathrm{ar}(X_{\mathscr{L}}) - (b-a)^{-1}\mathbb{E}\left[X_{\mathscr{L}}\right]^2 \left(\sum_{l\in\mathscr{L}} \lambda_l^{-1}\right),
\end{aligned}
$$

which proves the first part of the Theorem 5.8.3.

To get the second part, it suffices to apply Slutsky lemma [27] since the $\hat{\lambda}_l$'s are consistent. □

Once again, Theorem 5.3.3 is a direct consequence of Theorem 5.8.3 since the function $c_\delta : [a,b]^L \to \{0,1\}$ defined by (5.12) satisfies Definition 5.8.1.

5.8.c) Proof of Proposition 5.3.2

Here we compute

$$
I(L,k) = \int_{[a,b]^{L-k}} \left(\int_{[a,b]^k} \mathbb{1}_{|\max(\vee x_i, \vee y_i) - \min(\wedge x_i, \wedge y_i)| \le \delta}\, dx_1 \dots dx_k \right)^2 dy_1 \dots dy_{L-k}
$$

where $\wedge x_i = \min\{x_i, i \in \{1,\dots,k\}\}$, $\vee x_i = \max\{x_i, i \in \{1,\dots,k\}\}$ and respectively for $\wedge y_i$ and $\vee y_i$. Let us fix some $(y_1,\dots,y_{L-k})$ in $[a,b]^{L-k}$ and compute the inner integral

$$
\Sigma(y_1,\dots,y_{L-k}) = \int_{[a,b]^k} \mathbb{1}_{|\max(\vee x_i, \vee y_i) - \min(\wedge x_i, \wedge y_i)| \le \delta}\, dx_1 \dots dx_k.
$$

In order to do that let us decompose the integral with respect to the following conditions on $(x_1,\dots,x_k)$:

1. if $\wedge x_i > \wedge y_i$ and $\vee x_i > \vee y_i$, denote the integral A;
2. if $\wedge x_i < \wedge y_i$ and $\vee x_i < \vee y_i$, denote the integral B;
3. if $\wedge x_i > \wedge y_i$ and $\vee x_i < \vee y_i$, denote the integral C;
4. if $\wedge x_i < \wedge y_i$ and $\vee x_i > \vee y_i$, denote the integral D.

Since we have partitioned $[a,b]^k$ up to a null measure set, we have $\Sigma(y_1,\dots,y_{L-k}) = A + B + C + D$. Let us show the following equations for all $k = 2,\dots,L-1$,

$$
A = \mathbb{1}_{|\vee y_i - \wedge y_i| \le \delta} \left[\left(\min(\delta, b - \wedge y_i)\right)^k - \left(\vee y_i - \wedge y_i\right)^k \right], \tag{5.13}
$$

$$
B = \mathbb{1}_{|\vee y_i - \wedge y_i| \le \delta} \left[\left(\min(\delta, \vee y_i - a)\right)^k - \left(\vee y_i - \wedge y_i\right)^k \right], \tag{5.14}
$$

$$
C = \mathbb{1}_{|\vee y_i - \wedge y_i| \le \delta} \left(\vee y_i - \wedge y_i\right)^k, \tag{5.15}
$$

$$
\begin{aligned}
D = \mathbb{1}_{|\vee y_i - \wedge y_i| \le \delta} \Big[& \left(\vee y_i - \wedge y_i\right)^k - \left(\min(\delta, \vee y_i - a)\right)^k \\
& + k\left(\min(\wedge y_i, b - \delta) - \max\left(\vee y_i - \delta, a\right)\right)\delta^{k-1} + \left(\max(\delta, b - \wedge y_i)\right)^k - \left(b - \wedge y_i\right)^k \Big]
\end{aligned} \tag{5.16}
$$

and

$$\Sigma\left(y_1,\dots,y_{L-k}\right)=\mathbb{1}_{|\vee y_i-\wedge y_i|\leq\delta}\left[(k+1)\,\delta^k+k\left(\min(\wedge y_i,b-\delta)-\max\left(\vee y_i,a+\delta\right)\right)\delta^{k-1}\right]. \tag{5.17}$$

Let us fix some k in $\{2,\dots,L-1\}$.

Proof of (5.13) To compute A, it is sufficient to consider the case when $x_1=\vee x_i$, provided a multiplication by k, hence

$$\begin{aligned}
A &= k\int_{x_1=\vee y_i}^{b}\left(\int_{[\wedge y_i,x_1]^{k-1}}\mathbb{1}_{|x_1-\wedge y_i|\leq\delta}\,dx_2\dots dx_k\right)dx_1\\
&= k\mathbb{1}_{|\vee y_i-\wedge y_i|\leq\delta}\int_{x_1=\vee y_i}^{\min(\wedge y_i+\delta,b)}\left(\int_{[\wedge y_i,x_1]^{k-1}}1\,dx_2\dots dx_k\right)dx_1\\
&= k\mathbb{1}_{|\vee y_i-\wedge y_i|\leq\delta}\int_{x_1=\vee y_i}^{\min(\wedge y_i+\delta,b)}\left(x_1-\wedge y_i\right)^{k-1}dx_1\\
&= \mathbb{1}_{|\vee y_i-\wedge y_i|\leq\delta}\left[\left(\min(\wedge y_i+\delta,b)-\wedge y_i\right)^k-\left(\vee y_i-\wedge y_i\right)^k\right]\\
&= \mathbb{1}_{|\vee y_i-\wedge y_i|\leq\delta}\left[\left(\min(\delta,b-\wedge y_i)\right)^k-\left(\vee y_i-\wedge y_i\right)^k\right].
\end{aligned}$$

Proof of (5.14) To calculate B, we use the same idea and consider the case when $x_1=\wedge x_i$, leading to

$$\begin{aligned}
B &= k\int_{x_1=\max(\vee y_i-\delta,a)}^{\wedge y_i}\left(\vee y_i-x_1\right)^{k-1}dx_1\\
&= \mathbb{1}_{|\vee y_i-\wedge y_i|\leq\delta}\left[\left(\vee y_i-\max\left(\vee y_i-\delta,a\right)\right)^k-\left(\vee y_i-\wedge y_i\right)^k\right]\\
&= \mathbb{1}_{|\vee y_i-\wedge y_i|\leq\delta}\left[\left(\min(\delta,\vee y_i-a)\right)^k-\left(\vee y_i-\wedge y_i\right)^k\right].
\end{aligned}$$

Proof of (5.15) This case is pretty clear.

$$C=\int_{[\wedge y_i,\vee y_i]^k}\mathbb{1}_{|\vee y_i-\wedge y_i|\leq\delta}\,dx_1\dots dx_k=\mathbb{1}_{|\vee y_i-\wedge y_i|\leq\delta}\left(\vee y_i-\wedge y_i\right)^k$$

Proof of (5.16) To calculate D, it is sufficient to consider the case when $x_1 = \wedge x_i$ and $x_2 = \vee x_i$, provided a multiplication by $k\,(k-1)$, hence

$$\begin{aligned}
D &= k\,(k-1) \int_{x_1=a}^{\wedge y_i} \int_{x_2=\vee y_i}^{b} \left(\int_{x_1}^{x_2} \mathbb{1}_{|x_2-x_1|\leq\delta}\, dx_3 \dots dx_k \right) dx_2 dx_1 \\
&= k\,(k-1)\, \mathbb{1}_{|\vee y_i - \wedge y_i|\leq\delta} \int_{x_1=\max(\vee y_i-\delta,a)}^{\wedge y_i} \int_{x_2=\vee y_i}^{\min(x_1+\delta,b)} (x_2-x_1)^{k-2}\, dx_2 dx_1 \\
&= k \mathbb{1}_{|\vee y_i - \wedge y_i|\leq\delta} \int_{x_1=\max(\vee y_i-\delta,a)}^{\wedge y_i} (\min(x_1+\delta,b)-x_1)^{k-1} - (\vee y_i - x_1)^{k-1}\, dx_1 \\
&= \mathbb{1}_{|\vee y_i - \wedge y_i|\leq\delta} \Big[(\vee y_i - \wedge y_i)^k - (\vee y_i - \max(\vee y_i - \delta, a))^k \\
&\quad + k\,(\min(\wedge y_i, b-\delta) - \max(\vee y_i - \delta, a))\,\delta^{k-1} + (b - \min(\wedge y_i, b-\delta))^k - (b-\wedge y_i)^k \Big] \\
&= \mathbb{1}_{|\vee y_i - \wedge y_i|\leq\delta} \Big[(\vee y_i - \wedge y_i)^k - (\min(\delta, \vee y_i - a))^k \\
&\quad + k\,(\min(\wedge y_i, b-\delta) - \max(\vee y_i - \delta, a))\,\delta^{k-1} + (\max(\delta, b - \wedge y_i))^k - (b-\wedge y_i)^k \Big].
\end{aligned}$$

Proof of (5.17) Remark that

$$(\min(\delta, b-\wedge y_i))^k + (\max(\delta, b-\wedge y_i))^k = \delta^k + (b-\wedge y_i)^k \tag{5.18}$$

and

$$\max(\vee y_i - \delta, a) = \max(\vee y_i, a+\delta) - \delta. \tag{5.19}$$

Gathering (5.13), (5.14), (5.15), (5.16), (5.18) and (5.19) gives (5.17). Hence, Equation (5.17) holds for every k in $\{2, \dots, L-1\}$.

Moreover, if $k = 0$, then $\Sigma(y_1, \dots, y_{L-k}) = \mathbb{1}_{|\vee y_i - \wedge y_i|\leq\delta}$ and, if $k = 1$, then $\Sigma(y_1, \dots, y_{L-k}) = \mathbb{1}_{|\vee y_i - \wedge y_i|\leq\delta}\,[\min(\wedge y_i + \delta, b) - \max(\vee y_i - \delta, a)]$. To summarize, Equation (5.17) holds for every k in $\{0, \dots, L-1\}$.

It remains to compute

$$I(L,k) = \int_{[a,b]^{L-k}} \Sigma(y_1, \dots, y_{L-k})^2\, dy_1 \dots dy_{L-k}.$$

In order to do that, let us decompose the integral with respect to the following conditions on $(y_1, \dots, y_{L-k})$:

1. if $\vee y_i < a + \delta$, then $\Sigma = \delta^{k-1}\,[\delta + k\,(\wedge y_i - a)]$ and the corresponding integral is denoted by Y;
2. if $\wedge y_i > b - \delta$, then $\Sigma = \delta^{k-1}\,[\delta + k\,(b - \vee y_i)]$ and the corresponding integral is denoted by Z;
3. if $\vee y_i > a + \delta$ and $\wedge y_i < b - \delta$, then $\Sigma = \mathbb{1}_{|\vee y_i - \wedge y_i|\leq\delta}\delta^{k-1}\,[(k+1)\,\delta - k\,(\vee y_i - \wedge y_i)]$ and the corresponding integral is denoted by W.

These three cases are distinct because $\delta < (b-a)/2$, so we have partitioned $[a,b]^{L-k}$ up to a null measure set and $I(L,k) = Y + Z + W$. Let us show the following equations for all $k = 0, \ldots, L-2$,

$$Y = Z = C(L,k)\,\delta^{L+k}, \tag{5.20}$$

$$W = f(L,k)\,(b-a)\,\delta^{L+k-1} - [f(L,k) + g(L,k)]\,\delta^{L+k}, \tag{5.21}$$

where

$$C(L,k) = (L-k)\,\frac{(k+1)^{L-k+2}}{k^{L-k}} \int_0^{\frac{k}{k+1}} t^{L-k-1}\,(1-t)^2\,dt, \tag{5.22}$$

$$f(L,k) = (L-k)\,(k+1)^2 - 2\,(L-k-1)\,k\,(k+1) + \frac{(L-k)\,(L-k-1)}{(L-k+1)}k^2$$

and

$$g(L,k) = (k+1)^2 - 2\frac{(L-k-1)\,k\,(k+1)}{(L-k+1)} + \frac{(L-k)\,(L-k-1)}{(L-k+1)\,(L-k+2)}k^2.$$

Let us fix some k in $\{0, \ldots, L-2\}$.

Proof of (5.20) To compute Y, it is sufficient to consider the case when $y_1 = \wedge y_i$, provided a multiplication by $(L-k)$, hence

$$\begin{aligned}
Y &= \int_{\vee y_i \le a+\delta} \Sigma(y_1, \ldots, y_{L-k})^2\,dy_1 \ldots dy_{L-k} \\
&= (L-k)\,\delta^{2k-2} \int_{y_1=a}^{a+\delta} \left(\int_{[y_1, a+\delta]^{L-k-1}} [\delta + k\,(y_1 - a)]^2\,dy_2 \ldots dy_{L-k} \right) dy_1 \\
&= (L-k)\,\delta^{2k-2} \int_{y_1=a}^{a+\delta} (a+\delta-y_1)^{L-k-1}\,[\delta + k\,(y_1-a)]^2\,dy_1.
\end{aligned}$$

Defining the variable $u = a + \delta - y_1$ leads to

$$\begin{aligned}
Y &= (L-k)\,\delta^{2k-2} \int_0^{\delta} u^{L-k-1}\,[\delta + k\,(\delta - u)]^2\,du \\
&= (L-k)\,\delta^{2k-2} \int_0^{\delta} u^{L-k-1}\,[(k+1)\,\delta - ku]^2\,du,
\end{aligned}$$

and by defining the variable $t = \frac{ku}{(k+1)\delta}$ we have

$$\begin{aligned}
Y &= (L-k)\,\delta^{2k-2} \int_0^{\frac{k}{k+1}} \left(\frac{(k+1)\,\delta t}{k} \right)^{L-k-1} (k+1)^2\,\delta^2\,(1-t)^2\,\frac{(k+1)\,\delta}{k} dt \\
&= (L-k)\,\delta^{L+k} \frac{(k+1)^{L-k+2}}{k^{L-k}} \int_0^{\frac{k}{k+1}} t^{L-k-1}\,(1-t)^2\,dt.
\end{aligned}$$

The computation of Z can be done in the same way by inverting the roles of a and b on the one hand and the roles of $\wedge y_i$ and $\vee y_i$ on the other hand. This leads to $Z = Y$ and Equation (5.20).

Proof of (5.21) To compute W, it is sufficient to consider the case when $y_1 = \wedge y_i$ and $y_2 = \vee y_i$, provided a multiplication by $(L-k)(L-k-1)$, hence

$$\begin{aligned} W &= (L-k)(L-k-1)\delta^{2k-2} \\ &\quad \int_{y_1=a}^{b-\delta} \int_{y_2=\max(y_1,a+\delta)}^{b} \left(\int \mathbb{1}_{|y_2-y_1|\leq\delta} \left[(k+1)\delta - k(y_2-y_1)\right]^2 dy_3 \dots dy_{L-k} \right) dy_2 dy_1 \\ &= (L-k)(L-k-1)\delta^{2k-2} \\ &\quad \int_{y_1=a}^{b-\delta} \int_{y_2=\max(y_1,a+\delta)}^{y_1+\delta} \Big((y_2-y_1)^{L-k-2} \left[(k+1)^2\delta^2 - 2k(k+1)\delta(y_2-y_1) \right. \\ &\qquad \left. + k^2(y_2-y_1)^2\right] \Big) dy_2 dy_1, \end{aligned}$$

which leads to

$$\begin{aligned} W &= \int_a^{b-\delta} \left\{ (L-k)\delta^{2k-2}(k+1)^2\delta^2 \left[\delta^{L-k-1} - (\max(y_1, a+\delta) - y_1)^{L-k-1}\right] \right\} \\ &\quad - \left\{ 2(L-k-1)\delta^{2k-2}k(k+1)\delta \left[\delta^{L-k} - (\max(y_1, a+\delta) - y_1)^{L-k}\right] \right\} \\ &\quad + \left\{ \frac{(L-k)(L-k-1)}{(L-k+1)}\delta^{2k-2}k^2 \left[\delta^{L-k+1} - (\max(y_1, a+\delta) - y_1)^{L-k+1}\right] \right\} dy_1 \\ &= W_1 + W_2, \end{aligned}$$

where W_1 (resp. W_2) denotes the integral between a and $a+\delta$ (resp. between $a+\delta$ and $b-\delta$). Let us denote

$$f(L,k) = (L-k)(k+1)^2 - 2(L-k-1)k(k+1) + \frac{(L-k)(L-k-1)}{(L-k+1)}k^2. \tag{5.23}$$

Then, on the one hand

$$\begin{aligned} W_1 &= \delta^{2k-2} \left[\int_a^{a+\delta} f(L,k)\delta^{L-k+1} dy_1 \right. \\ &\quad - (L-k)(k+1)^2\delta^2 \int_a^{a+\delta} (a+\delta-y_1)^{L-k-1} dy_1 \\ &\quad + 2(L-k-1)k(k+1)\delta \int_a^{a+\delta} (a+\delta-y_1)^{L-k} dy_1 \\ &\quad \left. - \frac{(L-k)(L-k-1)}{(L-k+1)}k^2 \int_a^{a+\delta} (a+\delta-y_1)^{L-k+1} dy_1 \right] \\ &= f(L,k)\delta^{L+k} - g(L,k)\delta^{L+k}, \end{aligned}$$

with

$$g(L,k) = (k+1)^2 - 2\frac{(L-k-1)k(k+1)}{(L-k+1)} + \frac{(L-k)(L-k-1)}{(L-k+1)(L-k+2)}k^2. \tag{5.24}$$

On the other hand,

$$\begin{aligned} W_2 &= \delta^{2k-2}\int_{a+\delta}^{b-\delta}\left[(L-k)(k+1)^2 - 2(L-k-1)k(k+1)\right. \\ &\qquad \left. + \frac{(L-k)(L-k-1)}{(L-k+1)}k^2\right]\delta^{L-k+1}\,dy_1 \\ &= (b-a-2\delta)\,f(L,k)\,\delta^{L+k-1} \\ &= f(L,k)(b-a)\,\delta^{L+k-1} - 2f(L,k)\,\delta^{L+k} \end{aligned}$$

where $f(L,k)$ is defined by (5.23). Then, (5.21) clearly follows from $W = W_1 + W_2$.

Moreover, if $k = L-1$, then Equations (5.20) and (5.21) are still valid. To summarize, Equations (5.20) and (5.21) hold true for every k in $\{0,\ldots,L-1\}$.

Gathering (5.20) and (5.21) yields

$$I(L,k) = f(L,k)(b-a)\,\delta^{L+k-1} - [f(L,k) + g(L,k) - 2C(L,k)]\,\delta^{L+k},$$

for every k in $\{0,\ldots,L-1\}$.

To conclude, the integral involved in (5.22) can be computed with respect to k and L in the following way,

$$\int_0^{\frac{k}{k+1}} t^{L-k-1}(1-t)^2\,dt = \left(\frac{k}{k+1}\right)^{L-k}\left[\frac{1}{L-k} - \frac{2k}{(k+1)(L-k+1)} + \frac{k^2}{(k+1)^2(L-k+2)}\right].$$

Moreover, in the result stated in Proposition 5.3.2 we just used the software Mathematica in order to simplify the expressions. These simplifications lead to

$$f(L,k) = \frac{k(k+1)+L(L+1)}{L-k+1}$$

and

$$\begin{aligned} h(L,k) &:= f(L,k) + g(L,k) - 2C(L,k) \\ &= \frac{-k^3 + k^2(2+L) + k(5+2L-L^2) + L^3 + 2L^2 - L - 2}{(L-k+2)(L-k+1)}. \end{aligned}$$

Chapitre

6 Conclusion et perspectives

Ce travail se focalise sur l'étude théorique des processus ponctuels temporels. Les résultats théoriques démontrés dans ce manuscrit trouvent également leur place dans le cadre de la modélisation de neurones, et plus précisément des trains de spikes des neurones, par des processus ponctuels. On retrouve ici trois grands domaines de recherche des Mathématiques : équations aux dérivées partielles (EDP), probabilités et statistique. Nous proposons, pour chacun des domaines cités précédemment, un résumé de cette thèse adapté au point de vue adopté.

EDP. Un système d'équations aux dérivées partielles est particulièrement étudié dans ce manuscrit : c'est le système structuré en âge, noté (PPS), introduit dans [114] pour décrire la dynamique d'un groupe de neurones. D'une part, nous montrons dans le chapitre 2 qu'il représente, à quelques modifications près, la dynamique moyenne d'un neurone. D'autre part, les résultats du chapitre 3 permettent de voir le système (PPS) comme décrivant la dynamique d'un réseau de neurones en interaction dans la limite d'une grande population en interaction de type champ-moyen. Les fluctuations autour de cette dynamique asymptotique sont étudiées dans le chapitre 4. Notons également qu'un résultat d'existence/unicité des solutions de (PPS) est donné dans le chapitre 3 sous d'autres hypothèses que dans l'article originel [114].

Probabilités. Dans la première partie de cette thèse, nous étudions le processus d'âge associé à un processus ponctuel temporel. Plus précisément, nous cherchons à mieux connaître sa dynamique et à caractériser sa distribution. Dans un cas particulier (processus de renouvellement), ce processus d'âge est markovien et la dynamique de sa distribution est caractérisée par l'équation de Fokker-Planck, écrite sous la forme d'un système d'EDP dans ce manuscrit. En général, le processus d'âge n'est pas markovien mais il est possible d'écrire un système d'EDP vérifié par sa distribution bien que celui-ci soit plus complexe. D'autre part, les chapitres 3 et 4 étudient respectivement le premier ordre (loi des grands nombres) et le deuxième ordre (théorème central limite) de l'approximation de processus de Hawkes dépendants de l'âge en interaction de type champ-moyen dans la lignée de [43]. D'un point de vue probabiliste, le chapitre 5 donne l'approximation gaussienne d'une certaine fonctionnelle de processus de Poisson indépendants.

Statistique. Le chapitre 5 se propose de répondre à la problématique statistique suivante : comment tester l'indépendance entre deux, ou plus, processus de Poisson. Cette question est intéressante dans le cadre de la détection de synchronisations entre neurones. Nous répondons à cette problématique par la mise en œuvre d'une procédure de test dont le contrôle asymptotique (en la taille de l'échantillon) du niveau du test est assuré.

Toutefois, de nombreuses questions restent ouvertes et relèvent également de l'un des trois domaines cités ci-dessus. Des perspectives sont listées ci-dessous en respectant l'ordre des chapitres concernés.

Étude théorique et statistique d'EDP. Comme expliqué dans le chapitre 2, l'étude théorique du système (PPS) introduit dans [114] est bien connue, mais ce n'est pas le cas pour les systèmes d'EDP introduits dans le chapitre 2 : la généralisation du système (PPS) avec k âges, i.e. le système (2.27)-(2.29), et le système fermé obtenu dans le cas du processus de Hawkes linéaire, i.e. (2.36)-(2.37). La mise en évidence d'une convergence exponentielle vers un état d'équilibre ou de solutions périodiques sont des pistes de recherche intéressantes du point de vue de la modélisation.

De plus, dans le cadre de la confrontation de ces modèles avec des données réelles, la problématique de l'estimation des paramètres (taux de décharge) de ces EDP est centrale. Citons par exemple l'estimation du taux de division dans les équations de croissance/-fragmentation qui sont également des EDP de type équation de transport (voir la thèse d'Olivier [112]).

Étude couplée des modèles microscopiques et macroscopiques. Nous avons montré dans le chapitre 3 que le système (PPS) est la limite en champs moyen de processus de Hawkes dépendants de l'âge. La présence d'états d'équilibre et de solutions périodiques a été relevée pour ce système (PPS) pour des choix spécifiques des coefficients [114, 115]. Malheureusement, ces choix des coefficients ne rentrent pas dans le cadre d'hypothèses du chapitre 3. Il serait intéressant de mettre en évidence des conditions sur les paramètres des modèles telles que la présence d'états d'équilibres ou de solutions périodiques soit avérée et que le lien entre modèles microscopique et macroscopique soit théoriquement justifié. En effet, cela permettrait en particulier d'identifier les caractéristiques de la dynamique microscopique qui génèrent de tels phénomènes macroscopiques.

Deux angles d'attaque sont possibles :

- souligner d'autres choix de paramètres pour lesquels apparaissent des phénomènes macroscopiques intéressants,
- ou bien affaiblir les hypothèses du chapitre 3.

Casser l'indépendance asymptotique. Nous montrons une propagation du chaos dans le chapitre 3. Cela implique une indépendance asymptotique entre les neurones modélisés. D'un point de vue neurobiologique, cette indépendance semble irréaliste. Nous avons par exemple relevé des synchronisations entre neurones dans le chapitre 5 à certains moments précis au cours d'une tâche.

D'une part, cette indépendance peut être brisée en considérant des poids synaptiques corrélés [48]. D'autre part, la dépendance observée entre les neurones enregistrés par les biologistes peut être la conséquence d'un biais d'échantillonnage. Pour modéliser ce biais, il faudrait étudier ce que devient la propriété de propagation du chaos si l'on regarde des neurones en particulier (par exemple ceux qui ont les plus fortes interactions) plutôt que de regarder des neurones pris au hasard.

Processus de Hawkes localement stationnaires. Récemment, le modèle des processus de Hawkes (linéaires) localement stationnaires a été introduit dans [143]. C'est-à-dire

que la fonction d'interaction h ainsi que le taux d'apparition spontanée μ dépendent également du temps. D'un point de vue neurobiologique, son utilisation permet de prendre en compte l'aspect non-stationnaire d'un réseau de neurones.

Dans la continuité du chapitre 3, il serait intéressant de voir si l'approximation de champ-moyen tient toujours en faisant varier (sur une dynamique lente) la fonction d'interaction.

Dans le cadre de la modélisation de l'apprentissage d'un réseau de neurones, il serait aussi intéressant de pouvoir imprimer aux fonctions d'interaction une dynamique stochastique qui dépende des spikes précédents et plus particulièrement des synchronisations précédentes. Ici, les questions concernent la dynamique en temps long de ce réseau. Est-ce qu'un réseau d'interaction non trivial émerge de cette dynamique ?

Procédure statistique bi-échelle. Dans toute la première partie de cette thèse, nous étudions le lien entre deux échelles de la modélisation de l'activité électrique des neurones. De nombreuses questions se posent sur l'adéquation de ces modèles aux données enregistrées par les biologistes qui comportent au moins deux échelles : le potentiel de champ local et les trains de spikes. La construction de procédures de test d'adéquation aux données pour les deux échelles de modélisation prend donc tout son sens et le théorème central limite démontré dans le chapitre 4 pourrait en être la pierre angulaire.

Détection non-paramétrique de synchronisations. La problématique de la détection de synchronisations dans l'activité neuronale est abordée dans le chapitre 5. Nous la réduisons à la détection de dépendance entre processus de Poisson homogènes. Or, cette hypothèse de stationnarité et d'indépendance temporelle est trop forte. Pour dépasser ce genre d'hypothèses coûteuses, une possibilité est de se tourner vers des tests non-paramétriques tels que ceux abordés dans la thèse d'Albert [4]. A l'heure actuelle, ces procédures ne peuvent gérer que deux neurones en même temps. La généralisation de ces méthodes à plus de deux neurones offre des perspectives intéressantes à la fois théoriques et pratiques.

Appendix

A Supplementary material

A.1 Thinning procedure

A.1.a) Brief history

This section gives some history about the key representation of point processes used throughout this manuscript that is the thinning procedure. The idea of this representation comes from an article written by Lewis and Shedler in 1979 [92]. In this article, they first describe the procedures to simulate inhomogeneous Poisson processes that were commonly used at that time: time-rescaling of an homogeneous Poisson process, generation of the inter-events intervals, simulation of a Poisson variable plus an order statistics. All these procedures need either numerical integration or simulation of a Poisson variable that were known to suffer from computational complexity. That is why they introduced the thinning procedure which does not need any of those two features. Their result is the following.

Say you want to simulate a Poisson process N with intensity $\lambda(t)$ and that you are given a simulation of a Poisson process N^* with intensity $\lambda^*(t)$ such that for all $t \geq 0$, $\lambda^*(t) \geq \lambda(t)$. Of course we assume that the simulation of N^* is simpler than the simulation of N, for instance think of it as an homogeneous Poisson process. If each point T^* of N^* is deleted (independently) with probability[1] $1 - \lambda(T^*)/\lambda^*(T^*)$, then the remaining points form a realization of N.

A few years later, Ogata generalized this procedure to general multivariate point processes and not only Poisson processes. In [110], he proves the following result.

Say you want to simulate a multivariate point process $(N^i)_{i=1,\dots,m}$ with respective predictable intensities λ_t^i and that you are given a simulation of a multivariate point process $(N^{*,i})_{i=1,\dots,m}$ with respective predictable intensities $\lambda_t^{*,i}$ such that for all $t \geq 0$ and $i = 1, \dots, m$, $\lambda_t^{*,i} \geq \lambda_t^i$. Of course, once again, we assume that the simulation of the $N^{*,i}$, $i = 1, \dots, n$, is simpler than the simulation of the N^i, for instance take the $N^{*,i}$ as point processes with piecewise constant conditional intensity. Then, for any $i = 1, \dots, m$, if each point $T^{*,i}$ of $N^{*,i}$ is deleted (independently) with probability $1 - \lambda^i_{T^{*,i}}/\lambda^{*,i}_{T^{*,i}}$, then the remaining points form a realization of N^i.

This thinning procedure can also be generalized to spatial point processes. Once again, but in higher dimension, the simplest example is the case where you want to simulate an inhomogeneous spatial Poisson process thanks to the (simpler) simulation of a dominating homogeneous spatial Poisson process [107].

[1] Since T^* is a point of N^*, the probability that $\lambda^*(T^*) = 0$ is null.

A.1.b) On its applications

This simulation procedure still is computationally efficient and commonly used at the time this manuscript is written. Moreover the seminal idea of Lewis and Shedler can also be used for theoretical purposes by providing a representation of general one dimensional point processes thanks to two-dimensional homogeneous Poisson processes as stated in [17] or Theorem 2.8.11 (the heuristic behind this result is thoroughly explained in Section 2.3.c) so it is not reproduced here). The advantage relies on the fact that homogeneous Poisson processes are simpler and that their independence and stationarity assumptions are really convenient for many purposes. As theoretical applications of the thinning procedure, let us mention:

- a perfect simulation algorithm for linear Hawkes processes [105] (refinement of the standard simulation using the branching decomposition of linear Hawkes processes),
- the stability of non linear Hawkes processes [17],
- the coupling between mean-field interacting Hawkes processes and their limit counterpart ([43] or Chapter 3).

More precisely, Theorem 2.8.11 means that for every predictable intensity, one can construct a point process with this given intensity from a bivariate Poisson process. This result admits a converse given below. Its statement is inspired by [17, Lemma 4]

Theorem A.1.1 (Inversion Theorem). *Let $N = \{T_n\}_{n\geq 1}$ be a non explosive point process on $\mathbb{R}_+$ with $(\mathcal{F}_t)$-predictable intensity λ_t. Let $\{U_n\}_{n\geq 1}$ be a sequence of independent identically distributed random variables with uniform distribution on $[0,1]$. Moreover, suppose that they are independent of $\mathcal{F}_\infty$. Denote $\mathcal{G}_\infty = \sigma(U_n, n \geq 1)$. Let $\hat{N}$ be an homogeneous Poisson process with intensity 1 on $\mathbb{R}_+^2$ independent of[2] $\mathcal{F}_\infty \vee \mathcal{G}_\infty$. Define a point process $\bar{N}$ on $\mathbb{R}_+^2$ by*

$$\bar{N}((a,b]\times A) = \sum_{n\geq 1} \mathbb{1}_{(a,b]}(T_n)\,\mathbb{1}_A(\lambda_{T_n}U_n) + \int_{(a,b]}\int_{A\setminus[0,\lambda_t]} \hat{N}(dt\times dz)$$

for every $0 \leq a < b$ and A in $\mathcal{B}(\mathbb{R}_+)$.

Then, $\bar{N}$ is an homogeneous Poisson process on $\mathbb{R}_+^2$ with intensity 1 with respect to the filtration $\mathbb{H} := (\mathcal{H}_t)_{t\geq 0} = \left(\mathcal{F}_t \vee \mathcal{F}_t^{\hat{N}}\right)_{t\geq 0}$. Furthermore, the point process resulting from Theorem 2.8.11 with $\Pi = \bar{N}$ is exactly N.

Remark A.1.2. *Theorem A.1.1 implies that for any point process N which admits an intensity on $\mathbb{R}_+$, one can construct (on a possibly enlarged probability space) a bivariate Poisson process $\bar{N}$ such that N is the result of the thinning procedure of Theorem 2.8.11 with $\Pi = \bar{N}$. The heuristic lying behind Theorem A.1.1 is the following: on the one hand, the Poisson process $\bar{N}$ with intensity 1 on $\mathbb{R}_+^2$ can be taken as any Poisson process outside the strip where the thinning takes place, that is the random strip $\{(t,z), 0 \leq z \leq \lambda_t, t \geq 0\}$. On the other hand, inside this strip, $\bar{N}$ is obtained by lifting any point T_n of N uniformly in the interval $[0, \lambda_{T_n}]$ (by the means of the mark $Z_n = \lambda_{T_n}U_n$).*

[2]For any two σ-algebras $\mathcal{F}$ and $\mathcal{G}$, $\mathcal{F}\vee\mathcal{G}$ denotes the join α-algebra, that is the σ-algebra generated by $\mathcal{F}\cup\mathcal{G}$.

Due to the lack, up to our knowledge, of a complete proof of Theorem A.1.1 in the literature, we give a proof below.

Proof. Since $\mathcal{F}_\infty$ and $\mathcal{F}_\infty^{\hat{N}}$ are independent, N admits λ_t as a $\mathbb{H}$-intensity. Similarly, since $\mathcal{F}_\infty^{\hat{N}}$ and $\mathcal{F}_\infty$ are independent, $\hat{N}$ is an homogeneous Poisson process on $\mathbb{R}_+^2$ with intensity 1 with respect to the filtration $\mathbb{H}$.

For every integer k, let us define $\bar{N}^{(k)}$ by

$$\bar{N}^{(k)}(A) = \int_A \bar{N}(dt \times dz),$$

for every A in $\mathcal{B}(\mathbb{R}_+ \times [0,k])$ and let us show that for every $k \in \mathbb{N}$, $\bar{N}^{(k)}$ is a point process on $\mathbb{R}_+$ marked in $[0,k]$ with intensity kernel $1.dz$ with respect to $\mathbb{H}$.

It suffices to show that for every A_k in $\mathcal{B}([0,k])$, $\bar{N}_t^{(k)}(A_k)$ admits $\mathcal{L}_1(A_k)$ as a $\mathbb{H}$-intensity where $\mathcal{L}_1$ is the uni-dimensional Lebesgue measure. Fix k and let $\hat{N}^{(k)}$ denote the restriction of $\hat{N}$ to the points whose ordinate is smaller than k. We have $\bar{N}_t^{(k)}(A_k) = P_t^1 + P_t^2$ with

$$P_t^1 := \sum_{n\geq 1} \mathbb{1}_{[0,t]}(T_n)\, \mathbb{1}_{A_k}(\lambda_{T_n} U_n) \text{ and } P_t^2 := \int_0^t \int_{A_k \setminus [0,\lambda_u]} \hat{N}^{(k)}(du \times dz).$$

- Study of P^1. Let C_t be a non-negative $\mathbb{H}$-predictable process. Using the independence of $\mathcal{G}_\infty$ with respect to $\mathcal{H}_\infty$ and the fact that for all u in $[0,1]$, $\mathbb{1}_{A_k}(\lambda_t u)$ is $\mathbb{H}$-predictable, we have

$$\begin{aligned}
\mathbb{E}\left[\int_0^\infty C_t P^1(dt)\right] &= \mathbb{E}\left[\sum_{n\geq 1} C_{T_n} \mathbb{1}_{A_k}(\lambda_{T_n} U_n)\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{n\geq 1} C_{T_n} \mathbb{1}_{A_k}(\lambda_{T_n} U_n) \,\middle|\, \mathcal{H}_\infty\right]\right] \\
&= \mathbb{E}\left[\sum_{n\geq 1} C_{T_n} \int_0^1 \mathbb{1}_{A_k}(\lambda_{T_n} u) du\right] = \mathbb{E}\left[\int_0^\infty C_t \int_0^1 \mathbb{1}_{A_k}(\lambda_t u) du N(dt)\right] \\
&= \int_0^1 \mathbb{E}\left[\int_0^\infty C_t \mathbb{1}_{A_k}(\lambda_t u) \lambda_t dt\right] du = \mathbb{E}\left[\int_0^\infty C_t \int_0^{\lambda_t} \mathbb{1}_{A_k}(z) dz dt\right] \\
&= \mathbb{E}\left[\int_0^\infty C_t \mathcal{L}_1(A_k \cap [0,\lambda_t]) dt\right].
\end{aligned}$$

Hence, P^1 admits $\mathcal{L}_1(A_k \cap [0,\lambda_t])$ as a $\mathbb{H}$-intensity.

- Study of P^2. Since the intensity λ_t is predictable, for some fixed z in $[0,k]$,

$$\{(u,\omega) \in \mathbb{R}_+ \times \Omega \; tq \; z \leq \lambda_u(\omega)\} \in \mathcal{P}(\mathbb{H}),$$

where $\mathcal{P}(\mathbb{H})$ denotes the predictable σ-algebra associated with $\mathbb{H}$ (see [16, page 8]). Let us denote

$$\Gamma = \{(u,\omega,z) \in \mathbb{R}_+ \times \Omega \times E_k \, tq \; z \in A_k \setminus [0,\lambda_u(\omega)]\}.$$

We have

$$\Gamma = \bigcap_{n\in\mathbb{N}^*} \bigcup_{q\in\mathbb{Q}_+} \{(u,\omega) \in \mathbb{R}_+ \times \Omega \; tq \; q \leq \lambda_u(\omega)\} \times \left(A \cap \left[q + \frac{1}{n}, k\right]\right) \in \tilde{\mathcal{P}}_k(\mathbb{H}),$$

where $\tilde{\mathcal{P}}_k(\mathbb{H}) = \mathcal{P}(\mathbb{H}) \otimes \mathcal{B}([0,k])$ is the associated marked predictable σ-algebra. Hence, applying the Integration Theorem for marked point processes [16, VIII C.4.], one can deduce that P^2 admits $\mathcal{L}_1(A_k \setminus [0,\lambda_t])$ as a $\mathbb{H}$-intensity.

Finally, summing the two previous steps, $\bar{N}_t^{(k)}(A_k)$ admits $\mathcal{L}_1(A_k)$ as a $\mathbb{H}$ intensity. By definition of the intensity kernel, $\bar{N}^{(k)}$ is a point process on $\mathbb{R}_+$ marked in $[0,k]$ with intensity kernel $1.dz$ with respect to the filtration $\mathbb{H}$, i.e. $\bar{N}^{(k)}$ is a $\mathbb{H}$-Poisson process with intensity 1 on $\mathbb{R}_+ \times [0,k]$.

It remains to show that $\bar{N}$ is a $\mathbb{H}$)-Poisson process with intensity 1 on $\mathbb{R}_+^2$. Yet a Poisson process is characterized by its local behaviour since compact intervals generate the Borel σ-algebra. □

A.2 Solution of a PDE in the weak sense

We give here some insights about the necessity of the notion of weak solution to partial differential equation. This section is partially inspired by [121, Chapter 6]. Consider the didactic example of the constant speed transport equation:

$$\frac{\partial}{\partial t}u(t,x) + c\frac{\partial}{\partial x}u(t,x) = 0, \tag{A.1}$$

with $t \in \mathbb{R}$, $x \in \mathbb{R}$ and the initial condition $u(0,x) = u^{\text{in}}(x)$. To mean something in a standard (strong) way, we seek for solutions u that are continuously differentiable that is $u \in \mathcal{C}^1(\mathbb{R}^2)$. In particular, one needs a differentiable initial condition $u^{\text{in}} \in \mathcal{C}^1(\mathbb{R})$. In this case, the (strong) solution, which belongs to $\mathcal{C}^1(\mathbb{R}^2)$, is explicit and given by $u(t,x) = u^{\text{in}}(x-ct)$: that is the simplest application of the method of the characteristics. Here, the characteristics are the affine lines $\{(t,x) \in \mathbb{R}^2,\ x - ct = x_0\}$ for all x_0 in $\mathbb{R}$ and the solution is indeed constant along the characteristics.

Can we weaken the assumption $u^{\text{in}} \in \mathcal{C}^1(\mathbb{R})$? Indeed, if the initial condition is not differentiable then the strong formulation of (A.1) is compromised since $\frac{\partial}{\partial t}u$ and $\frac{\partial}{\partial x}u$ are, a priori, not well defined as functions. We need another formulation in order to weaken the sense of being a solution of (A.1). Yet this weaker formulation must be compatible with the stronger one when the required assumptions are fulfilled. This can be achieved using the heuristics of the integration by parts formula.

Assume for now that u is a solution in $\mathcal{C}^1(\mathbb{R})$. Let $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be a compactly supported smooth function (test function), multiply (A.1) by $\varphi(t,x)$ and integrate it for t and x in $\mathbb{R}$. We have,

$$\int_{\mathbb{R}^2} \frac{\partial}{\partial t}u(t,x)\varphi(t,x)dtdx + c\int_{\mathbb{R}^2} \frac{\partial}{\partial x}u(t,x)\varphi(t,x)dtdx = 0.$$

Since φ is compactly supported, the boundary terms are null and the integration by parts formula yields

$$-\int_{\mathbb{R}^2} u(t,x)\frac{\partial}{\partial t}\varphi(t,x)dtdx - c\int_{\mathbb{R}^2} u(t,x)\frac{\partial}{\partial x}\varphi(t,x)dtdx = 0. \tag{A.2}$$

Remark that the derivatives of u are not involved in this last equation. This justifies the following definition.

Definition A.2.1. *If (A.2) is satisfied for all test functions φ, then u is said to be a solution of (A.1) in the weak sense.*

Remark A.2.2. *The definition above does not characterize the initial condition u^{in}. To do so, one can divide the set of times into the set of positive times and negative times and write two weak equations like (A.2) as it is done in [121, Definition 6.1.]*

A standard (strong) solution of (A.1) is also a solution in the weak sense. Conversely, a solution in the weak sense that is furthermore continuously differentiable is a strong solution. However, there exist some weak solutions that are not strong solutions. For instance, $u(t,x) = |x - ct|$ is a solution in the weak sense (associated with the initial condition $u^{\text{in}} = |\cdot|$) but it is not everywhere differentiable.

Only functions are written above for simplicity but this theory applies to more general objects such as distributions (continuous linear forms acting on the test function space, see Schwartz reference book [146]). In particular, the measures studied in this thesis are often considered, in a natural way, as living in dual functional spaces.

A.3 On Riesz representation theorem

In Chapter 4, the fluctuation process $(\eta_t^n)_{t\geq 0}$ is considered as a stochastic process taking values in the dual space of some Hilbert space of test functions. One could use the Riesz representation theorem to consider the fluctuation process as taking values in a functional space, hence a simpler space than the natural distributional space in which it belongs. However, this option comes with the major issue that it is not consistent with the natural embeddings existing between test function spaces as explained below, with a didactic example.

Consider $L^2 = L^2(\mathbb{R})$ and $W^1 = W^1(\mathbb{R})$, the latter being the space of functions f that belongs to L^2 as well as their first order derivative. Denote by $\mathcal{R}_0$ the Riesz-isomorphism from L^2 to its dual $(L^2)^*$ and by $\mathcal{R}_1$ the Riesz-isomorphism from W^1 to its dual $(W^1)^*$. Moreover, denote by i and i^* the following canonical continuous embeddings:

$$i : W^1 \hookrightarrow L^2 \text{ and } i^* : (L^2)^* \hookrightarrow (W^1)^*.$$

Say you want to regard a distribution w in $(L^2)^*$ either in its natural space $(L^2)^*$ or in $(W^1)^*$ through the embedding i^*. You may as well consider the representation of w via the Riesz representation theorem. To keep some coherence through the representation, it requires the following consistency property: $(\mathcal{R}_0 \circ i \circ \mathcal{R}_1^{-1} \circ i^*)(w) = w$. Equivalently, we need the following diagram to commute.

$$\begin{array}{ccc} (L^2)^* & \xleftarrow{\mathcal{R}_0} & L^2 \\ \downarrow{\scriptstyle i^*} & & \uparrow{\scriptstyle i} \\ (W^1)^* & \xleftarrow{\mathcal{R}_1} & W^1 \end{array}$$

Yet this diagram do not commute. Obviously, starting from f in $L^2 \setminus W^1$, we have $(\mathcal{R}_1^{-1} \circ i^* \circ \mathcal{R}_0)(f) \in W^1$ and so

$$(i \circ \mathcal{R}_1^{-1} \circ i^* \circ \mathcal{R}_0)(f) \neq f.$$

The issue is even stronger. Let f be a compactly supported function of class $\mathcal{C}^\infty$. Obviously, f belongs to W^1 and L^2. Hence, $w := i^* \circ \mathcal{R}_0 \circ i(f)$ is the following mapping:

$$w : \varphi \in W^1 \mapsto < \varphi, f >_{L^2} = \int_{\mathbb{R}} \varphi(x) f(x) dx.$$

Consequently, $g := \mathcal{R}_1^{-1}(w)$ is the function in W^1 satisfying

$$\forall \varphi \in W^1, \quad \int_{\mathbb{R}} \varphi(x) f(x) dx =< \varphi, g >_{W^1} = \int_{\mathbb{R}} \varphi(x) g(x) dx + \int_{\mathbb{R}} \varphi'(x) g'(x) dx.$$

This condition does not imply $f = g$. Indeed, assume that $f = g$, then, for any compactly supported test functions φ of class $\mathcal{C}^\infty$, we have, thanks to the integration by parts formula,

$$\int_{\mathbb{R}} \varphi(x) f(x) dx = \int_{\mathbb{R}} \varphi(x) f(x) dx + \int_{\mathbb{R}} \varphi(x) f''(x) dx,$$

which is clearly false in general.

A.4 Hilbert space valued stochastic processes

Let us denote by H a Hilbert space, equipped with the inner scalar product $<\cdot, \cdot>$, by H^* its dual space. To avoid confusion, let us mention that the notation $\langle w, h \rangle$ is used to denote the dual action of any w in H^* on any h in H. Let us denote by $\mathcal{C}(\mathbb{R}_+, H)$ (respectively $\mathcal{D}(\mathbb{R}_+, H)$) the space of continuous functions (resp. the space of càdlàg functions) from $\mathbb{R}_+$ to H. This section gathers some results on H-valued stochastic processes that are, in our case, random variables taking values in either $\mathcal{C}(\mathbb{R}_+, H)$ or $\mathcal{D}(\mathbb{R}_+, H)$.

We start with martingale properties for H-valued processes. We refer to [36] for the definition of an H-valued martingale and we mention the following equivalent definition [128, Remark 2.2.5]:

> Let $(M_t)_{t\geq 0}$ be such that $\mathbb{E}[||M_t||_H] < +\infty$ for all $t \geq 0$. Then, $(M_t)_{t\geq 0}$ is an $\mathbb{F}$-martingale if and only if $(\langle w_k, M_t \rangle)_{t\geq 0}$ is an $\mathbb{F}$-martingale for all $k \geq 1$, where $(w_k)_{k\geq 1}$ is an arbitrary orthonormal basis of H^*.

Definition A.4.1 ([103] or [36]). *Let $(M_t)_{t\geq 0}$ be an H-valued square integrable $\mathbb{F}$-martingale. We denote by $(\ll M \gg_t)_{t\geq 0}$ its Doob-Meyer process, that is the unique (up to indistinguishability) $\mathbb{F}$-predictable process with values in $L(H^*, H)$ such that: for all w_1, w_2 in H^*,*

$$\left(\langle w_1, M_t \rangle \, \langle w_2, M_t \rangle - \langle w_1, \ll M \gg_t w_2 \rangle \right)_{t\geq 0}$$

is a real valued $\mathbb{F}$-martingale.

Remark A.4.2. *Since the Doob-Meyer process takes values in a space of linear mappings, it is still a quite complex object. It is often summarized by its trace in H, that is the real valued process given by*

$$\mathrm{Tr} \ll M \gg_t = \sum_{k\geq 1} \langle w_k, \ll M \gg_t w_k \rangle$$

where $(w_k)_{k\geq 1}$ is an arbitrary orthonormal basis of H^*.

Then, we stress the fact that the finite dimensional distributions of a stochastic process characterize its distribution. Even if this result is standard for processes with values in $\mathbb{R}$ or $\mathbb{R}^d$, it is not so easy to find a clear statement for Hilbert space valued processes. Hence, we give here an extension of [77, Lemma 3.19].

Proposition A.4.3. *Let us denote by $\mathcal{L}(X)$ the distribution of a random variable X. Let D be a dense subset of $\mathbb{R}_+$, and $(X_t)_{t\geq 0}$, $(Y_t)_{t\geq 0}$ be two processes with trajectories in $\mathcal{D}(\mathbb{R}_+, H)$ satisfying*

$$\forall k \in \mathbb{N}^*, \forall (t_1, \ldots, t_k) \in D^k, \quad \mathcal{L}(X_{t_1}, ..., X_{t_k}) = \mathcal{L}(Y_{t_1}, ..., Y_{t_k}). \tag{A.3}$$

Then, $\mathcal{L}((X_t)_{t\geq 0}) = \mathcal{L}((Y_t)_{t\geq 0})$ as laws of random variables taking values in $\mathcal{D}(\mathbb{R}_+, H)$ endowed with the Skorokhod topology.

Proof. The idea is to reduce the problem to the finite dimensional case of processes in $\mathcal{D}(\mathbb{R}_+, \mathbb{R}^d)$ given by [77, Lemma 3.19] and then to use some approximation argument.

Let F be a bounded continuous mapping from $\mathcal{D}(\mathbb{R}_+, H)$ to $\mathbb{R}$. It suffices to prove that $\mathbb{E}[F((X_t)_{t\geq 0})] = \mathbb{E}[F((Y_t)_{t\geq 0})]$. Let $(h_k)_{k\geq 1}$ be an orthonormal basis of H. For all $p \geq 1$, we construct a canonical bounded continuous mapping from $\mathcal{D}(\mathbb{R}_+, \mathbb{R}^p)$ to $\mathbb{R}$. Let us denote by $R_p : \mathbb{R}^p \to H$ the canonic realisation of a vector of $\mathbb{R}^p$ as an element of H, that is

$$R_p(x_1, \ldots, x_p) := \sum_{k=1}^{p} x_k h_k.$$

This mapping is Lipschitz continuous with constant 1 from $\mathbb{R}^p$ equipped with the Euclidean norm to H. In particular, it is uniformly continuous, so the mapping

$$(x_1(t), \ldots, x_p(t))_{t\geq 0} \in \mathcal{D}(\mathbb{R}_+, \mathbb{R}^p) \mapsto (R_p(x_1(t), \ldots, x_2(t)))_{t\geq 0} \in \mathcal{D}(\mathbb{R}_+, H)$$

is continuous with respect to the Skorokhod topology. Then, the mapping F_p, defined by $F_p((x_1(t), \ldots, x_p(t))_{t\geq 0}) := F((R_p(x_1(t), \ldots, x_2(t)))_{t\geq 0})$ is continuous from $\mathcal{D}(\mathbb{R}_+, \mathbb{R}^p)$ to $\mathbb{R}$. Let us denote by $\pi_p : H \to \mathbb{R}^p$ the vector of coordinates smaller than p, that is $\pi_p(h) := (< h, h_1 >, \ldots, < h, h_p >)$. Notice that $\pi_p \circ R_p$ is the identity mapping from $\mathbb{R}^p$ to $\mathbb{R}^p$ and that for all h in H, $R_p \circ \pi_p(h)$ converges to h as p tends to $+\infty$. Then, (A.3) implies that $\mathcal{X}_p := (\pi_p(X_t))_{t\geq 0}$ and $\mathcal{Y}_p := (\pi_p(Y_t))_{t\geq 0}$ are càdlàg processes with values in $\mathbb{R}^p$ with the same finite dimensional distributions. Hence, [77, Lemma 3.19] implies that $(\pi_p(X_t))_{t\geq 0}$ and $(\pi_p(Y_t))_{t\geq 0}$ have the same distribution on $\mathcal{D}(\mathbb{R}_+, \mathbb{R}^p)$ and in particular,

$$\mathbb{E}\left[F_p(\mathcal{X}_p)\right] = \mathbb{E}\left[F_p(\mathcal{Y}_p)\right]. \tag{A.4}$$

Now, notice that $F_p(\mathcal{X}_p) = F((R_p \circ \pi_p(X_t))_{t\geq 0})$ and that $(R_p \circ \pi_p(X_t))_{t\geq 0} \to (X_t)_{t\geq 0}$ for the local uniform topology and consequently for the Skorokhod topology. Finally, this implies that $\mathbb{E}\left[F((X_t)_{t\geq 0})\right] = \lim_{p\to+\infty} \mathbb{E}\left[F_p(\mathcal{X}_p)\right]$ by dominated convergence, so taking the limit in (A.4) gives

$$\mathbb{E}\left[F((X_t)_{t\geq 0})\right] = \mathbb{E}\left[F((Y_t)_{t\geq 0})\right]$$

which ends the proof. □

Finally, we give some informations about Gaussian variables and Gaussian processes with values in Hilbert spaces. Most of what follows is taken from [36]. A random variable X with values in H is called Gaussian if for arbitrary h in H, the real valued random variable $< X, h >$ is Gaussian. A Gaussian random variable X admits a mean and a covariance in the following sense: there exist m in H and a symmetric non-negative continuous mapping Q from H to H such that

$$\begin{cases} \forall h \in H, \quad \mathbb{E}\left[< X, h >\right] =< m, h >, \\ \forall h_1, h_2 \in H, \quad \mathbb{E}\left[< X, h_1 >< X, h_2 >\right] - < m, h_1 >< m, h_2 >=< Qh_1, h_2 > . \end{cases}$$

Then, m and Q are respectively called the mean and covariance operator of X. Notice that, by generalization, a random variable X with values in H^k is called Gaussian if for arbitrary $h_1, \dots, h_k$ in H, the real valued random vector $(< X, h_1 >, \dots, < X, h_k >)$ is Gaussian.

An H-valued stochastic process $(X_t)_{t\geq 0}$ is called Gaussian if, for any k in $\mathbb{N}^*$ and arbitrary non-negative times $t_1, \dots, t_k$, the H^k-valued random variable $(X_{t_1}, \dots, X_{t_k})$ is Gaussian. Equivalently, a process is Gaussian if and only if, for any k in $\mathbb{N}^*$ and arbitrary $h_1, \dots, h_k$ in H, the real-valued process $(< X_t, h_1 >, \dots, < X_t, h_k >)_{t\geq 0}$ is Gaussian.

Under regularity assumptions that are assumed here, a Gaussian process admits a continuous version (Kolmogorov continuity theorem [36, Proposition 3.15]). Furthermore, its distribution is characterized by its mean process and its covariance.

Proposition A.4.4. *Let $(X_t)_{t\geq 0}$ be a Gaussian process with trajectories in $\mathcal{C}(\mathbb{R}_+, H)$. Then, its distribution is uniquely characterized by its mean process*

$$\forall h \in H, \quad \left(\mathbb{E}\left[< X, h >\right]\right)_{t\geq 0},$$

and its covariance

$$\forall t_1, t_2 \geq 0, \forall h_1, h_2 \in H, \quad \mathbb{E}\left[< X_{t_1}, h_1 >< X_{t_2}, h_2 >\right]. \tag{A.5}$$

Proof. Obvisouly, the finite dimensional distributions of $(X_t)_{t\geq 0}$ are characterized by its mean process and its covariance. Furthermore, since we assume that $(X_t)_{t\geq 0}$ is continuous, Proposition A.4.3 gives the result. □

Remark A.4.5. *Since H and its dual H^* are isometrically isomorphic (Riesz representation theorem), the covariance of a Gaussian process is also characterized by*

$$\forall t_1, t_2 \geq 0, \forall w_1, w_2 \in H^*, \quad \mathbb{E}\left[\langle w_1, X_{t_1}\rangle \langle w_2, X_{t_2}\rangle\right],$$

in comparison with (A.5).

To conclude this section, let us state here the result which gives the existence of the Gaussian process W introduced in Chapter 4.

Proposition A.4.6. *Let H and H^+ be two Hilbert spaces such that $H \hookrightarrow_{H.S} H^+$ and $(V_t)_{t\geq 0}$ be a familly of bilinear forms on H. Assume that there exists a locally bounded function C such that*

$$\forall h \in H, \ \forall \theta \geq 0, \quad \sup_{t\in[0,\theta]} V_t(h,h) \leq C(\theta)||h||^2_{H^+}. \tag{A.6}$$

Then, for all $\theta \geq 0$, there exists a Gaussian process $(X_t)_{t\in[0,\theta]}$ with mean zero and covariance given by

$$\forall t_1, t_2 \leq \theta, \ \forall h_1, h_2 \in H, \quad \mathbb{E}\left[< X_{t_1}, h_1 >< X_{t_2}, h_2 >\right] = \int_0^{t_1\wedge t_2} V_t(h_1, h_2)dt. \tag{A.7}$$

The existence of W then follows from the fact that the familly $(V_t)_{t\geq 0}$ of bilinear forms on $W_0^{2,\alpha}$, given by

$$\forall t \geq 0, \forall \varphi_1, \varphi_2 \in W_0^{2,\alpha}, \quad V_t(\varphi_1, \varphi_2) = \int_0^{+\infty} \varphi_1(s)\varphi_2(s)\Psi(s, \overline{\gamma}(t))u(t,s)ds,$$

satisfies (A.6). Indeed, Proposition A.4.6 applied with $H = \mathcal{W}_0^{2,\alpha}$ and $H^+ = \mathcal{W}_0^{2,\alpha+1}$ gives the existence of a Gaussian process $(X_t)_{t\in[0,\theta]}$. Finally, the Gaussian process W with values in $\mathcal{W}_0^{-2,\alpha}$ is the image of X (with values in $\mathcal{W}_0^{2,\alpha}$) by Riesz representation.

Proof. The idea of the proof is similar to one of the different proofs of the existence of the Brownian motion: construct a probability space such that the space of L^2 random variables is isometrically isomorphic to the space of L^2 functions in time.

For any $\theta \geq 0$, let us denote $H_\theta := W^1([0,\theta], H)$, that is the space of functions with values in H which belongs to L^2 as well as their first derivative, and $Q_\theta : (H_\theta)^2 \to \mathbb{R}$ defined by, for all h^1, h^2 in H_θ,

$$Q_\theta(h^1, h^2) := \int_0^\theta V_t(h^1(t), h^2(t))dt.$$

Let $(h^k)_{k\geq 1}$ be an orthonormal basis of H and $(\rho^j)_{j\geq 1}$ be an orthonormal basis of $W^1([0,\theta],\mathbb{R})$. In the following, we hence consider the orthonormal basis of H_θ given by $(h^k \otimes \rho^j)_{k,j\geq 1}$.

Let us show that Q_θ is trace class, that is $\operatorname{Tr} Q_\theta < +\infty$. For all k, j, we have

$$\begin{aligned} Q_\theta(h^k \otimes \rho^j, h^k \otimes \rho^j) &= \int_0^\theta V_t(\rho^j(t)h^k, \rho^j(t)h^k)dt = \int_0^\theta \rho^j(t)^2 V_t(h^k, h^k)dt \\ &\leq C(\theta)||h^k||^2_{H^+}||\rho^j||^2_{L^2}. \end{aligned}$$

Hence, $\operatorname{Tr} Q_\theta \leq C(\theta)\sum_{k\geq 1}||h^k||^2_{H^+} + \sum_{j\geq 1}||\rho^j||^2_{L^2([0,\theta])}$ which is finite since $H \hookrightarrow_{H.S} H^+$ and $W^1 \hookrightarrow_{H.S} L^2$.

Since Q_θ is trace class, there exists a probabiliy space $(\Omega, \mathcal{F}, \mathbb{P})$ and H_θ-valued Gaussian variable, denoted by G_θ, with zero mean and covariance operator given by Q_θ. Let us denote, for all $\theta \geq 0$, $\Phi_\theta : W^1([0,\theta],\mathbb{R}) \to L^2(\Omega, \mathcal{F}, \mathbb{P}; H)$ defined by, for all ρ in $W^1([0,\theta],\mathbb{R})$,

$$\Phi_\theta(f) := \sum_{k\geq 1} < G_\theta, f \otimes h^k > h^k.$$

Let us verify that $\Phi_\theta(f)$ is a L^2 random variable. We have,

$$\mathbb{E}\left[||\Phi_\theta(f)||^2\right] = \mathbb{E}\left[\sum_{k\geq 1} | < G_\theta, f \otimes h^k > |^2\right] = \sum_{k\geq 1} Q_\theta(f \otimes h^k, f \otimes h^k)$$

where we used Tonelli's theorem and the definition of G_θ. Then, the same kind of computations as we have done above yields

$$\sum_{k\geq 1} Q_\theta(f \otimes h^k, f \otimes h^k) \leq C(\theta)\sum_{k\geq 1}||h^k||^2_{H^+}||f||^2_{L^2} < +\infty,$$

which gives that $\Phi_\theta(f)$ is a L^2 random variable. Furthermore, the inequality above implies that Φ_θ is uniformly continuous and can be extended as a mapping from $L^2([0,\theta],\mathbb{R})$ to $L^2(\Omega, \mathcal{F}, \mathbb{P}; H)$: approximate any f in L^2 by a sequence $(f_n)_{n\geq 1}$ of functions in W^1 and use the uniform continuity to prove that the sequence $(\Phi_\theta(f_n))_{n\geq 1}$ is Cauchy in $L^2(\Omega, \mathcal{F}, \mathbb{P}; H)$.

It only remains to check that the process $(X_t)_{t\in[0,\theta]}$ defined by $X_t := \Phi_\theta(\mathbb{1}_{[0,t]})$ is indeed a Gaussian process with zero mean and covariance given by (A.7). This is left to the reader. □

Annexe

B Développement

Cette thèse s'inscrit dans l'étude théorique des processus ponctuels temporels (sous-ensembles aléatoires discrets de $\mathbb{R}$) tant d'un point de vue probabiliste que statistique et puise sa motivation dans l'utilisation de ces processus pour modéliser l'activité électrique des neurones. Nous cherchons en particulier à comprendre les liens qui peuvent exister entre ces processus ponctuels, qui modélisent les neurones individuellement, et des équations aux dérivées partielles (EDP), qui modélisent l'activité globale/moyennée sur un grand nombre de neurones.

Contexte biologique. Les neurones, cellules de base du système nerveux, forment un réseau qui permet de transporter de l'information, depuis des récepteurs jusqu'au système cognitif par exemple, sous la forme d'un signal électro-chimique (influx nerveux). Le substrat électrique de ce transport d'information est constitué par les *potentiels d'action*. Un potentiel d'action correspond à une dépolarisation brève et stéréotypée du potentiel de la membrane plasmique d'un neurone : on dit alors que le neurone *décharge*, et le nombre moyen de décharge par unité de temps est appelé *taux de décharge*.

L'information nécessaire à la compréhension du réseau formé par les neurones peut être réduite à la séquence des temps d'occurrence de ces potentiels d'action également appelée *train de spikes* (un *spike* correspond à un temps de décharge du neurone). Outre les interactions qui existent au sein de ce réseau, il faut également noter la présence, à l'échelle d'un neurone individuel, du phénomène de *période réfractaire* : un neurone ne peut pas décharger deux fois dans un trop court délai (de l'ordre de la milliseconde).

Plusieurs échelles de modélisation sont possibles. Il existe naturellement plusieurs échelles d'étude du système nerveux, depuis le niveau moléculaire (canaux ioniques) jusqu'à un niveau macroscopique (régions corticales). Dans ce manuscrit, les deux échelles suivantes sont étudiées : l'échelle (microscopique) des neurones décrits par des processus ponctuels temporels (ces derniers modélisent les trains de spikes) et l'échelle (macroscopique) du réseau neuronal tout entier dont la dynamique est décrite par un système d'équations aux dérivées partielles (EDP) structuré en âge.

Processus ponctuels : notations. Rappelons ici quelques notations utilisées dans ce manuscrit. Nous nous intéressons à des processus ponctuels, généralement notés N, sur $\mathbb{R}$ muni de sa tribu borélienne $\mathcal{B}(\mathbb{R})$. Pour tout A dans $\mathcal{B}(\mathbb{R})$, nous notons $N(A)$ le cardinal de $N \cap A$. De plus, nous notons $N(dt)$ la mesure ponctuelle associée à N, i.e. la mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour toute fonction mesurable positive f, $\int_{\mathbb{R}} f(t)N(dt) = \sum_{i\in\mathbb{Z}} f(T_i)$.

Dans ce manuscrit, nous nous intéressons plus particulièrement au comportement de N pour les temps strictement positifs : nous supposerons que la trace de N sur les temps négatifs est caractérisée par sa loi notée ζ_{N_-}. Ainsi, nous nous intéressons de manière équivalente au processus ponctuel N ou au *processus de comptage* $(N_t)_{t\geq 0}$ associé, défini

par $N_t := N(]0,t])$. Il est commun de simplifier l'étude d'un processus de comptage à celle de son intensité stochastique notée $(\lambda_t)_{t\geq 0}$. De manière informelle, $\lambda_t dt$ représente la probabilité que le processus N admette un nouveau point dans l'intervalle $[t, t+dt]$ sachant $\mathcal{F}_{t-}$, et ce conditionnement permet en particulier de modéliser des interactions (attirance, répulsion) entre les différents points du processus. Pour finir, mentionnons également les processus d'âge (standard et prévisible) associés à un processus ponctuel N, respectivement notés $(S_t)_{t\geq 0}$ et $(S_{t-})_{t\geq 0}$. Ils décrivent le temps écoulé depuis le dernier point du processus : voir (1.1) et (1.2) pour une définition formelle.

Processus ponctuels : état de l'art. Compte tenu de notre motivation biologique, nous nous intéressons plus particulièrement à des processus ponctuels utilisés pour modéliser des trains de spikes. Ainsi, nous discutons des processus de Poisson [40, 66, 158], de renouvellement, de Wold [89, 127], et de Hawkes [31, 70]. Une attention toute particulière est portée sur les processus de Hawkes compte tenu de leur usage simple et efficace pour modéliser des neurones en interaction. Rappelons ici qu'un processus de Hawkes multivarié est un processus ponctuel multivarié $(N^1, \dots, N^n)$ dont les intensités respectives pour $i = 1, \dots, n$ sont données par

$$\lambda_t^i = \Phi_i \left(\sum_{j=1}^{n} \int_0^{t-} h_{j\to i}(t-z) N^j(dz) \right),$$

où les Φ_i et les $h_{j\to i}$ sont des fonctions réelles. Ici, l'intensité stochastique (taux de décharge du neurone i) au temps t dépend de tous les spikes de tous les neurones avant le temps t. La fonction $h_{j\to i}$ décrit l'influence d'une décharge du neurone j sur le potentiel de membrane, et donc sur le taux de décharge, du neurone i.

Les propriétés théoriques/pratiques des processus ponctuels en général, et des processus de Hawkes en particulier, ont été largement étudiées dans la littérature : stationnarité [17], stabilité, asymptotique sur la dimension n [43], asymptotique en temps long [163], simulation [110], estimation [139], tests d'hypothèses [53]. Ce manuscrit propose d'étudier :

1. l'asymptotique sur la dimension n de processus de Hawkes multivariés,

2. un test d'indépendance entre plusieurs processus de Poisson.

EDP structurée en âge : notations. À un niveau macroscopique, nous nous concentrons sur le point de vue proposé par Pakdaman, Perthame et Salort dans une série d'articles [114, 115, 116]. Ils étudient le système d'EDP structuré en âge suivant :

$$\begin{cases} \dfrac{\partial n\,(t,s)}{\partial t} + \dfrac{\partial n\,(t,s)}{\partial s} + p\,(s, X\,(t))\, n\,(t,s) = 0, \\ m\,(t) := n\,(t,0) = \displaystyle\int_0^{+\infty} p\,(s, X\,(t))\, n\,(t,s)\, ds. \end{cases} \tag{PPS}$$

Ici, $n(t,s)$ représente la densité de probabilité de trouver un neurone (typique du réseau) d'âge s au temps t, l'âge d'un neurone étant le délai depuis sa dernière décharge. Bien entendu, la définition de l'âge d'un neurone correspond à la définition du processus d'âge associé à un processus ponctuel dès l'instant où ce dernier modélise le train de spikes dudit neurone. La fonction p représente le taux de décharge d'un neurone. Notons que dépendance de p par rapport à l'âge s permet de modéliser le phénomène de période réfractaire (par

exemple, $p(s,x) = \mathbb{1}_{s\geq\delta}$). De plus, la fonction p dépend également de la variable $X(t)$ représentant l'activité globale du réseau au temps t, sous la forme

$$X(t) := \int_0^t d(t-z)n(z,0)dz, \tag{B.1}$$

pour d une certaine fonction de délai. Cette intégrale de convolution modélise le temps de propagation de la décharge d'un neurone au reste du réseau.

EDP structurée en âge : état de l'art. Un phénomène de relaxation à l'équilibre a été mis en évidence pour la solution de (PPS) dans [114]. Plus précisément, un critère qualitatif sur la fonction de délai d peut être donné : si $||d||_{L^1} \approx 0$ ou bien $||d||_{L^1} \approx +\infty$, alors la solution de l'équation converge vers un état stationnaire. Entre ces deux régimes, des solutions numériques semblent présenter des oscillations. De manière plus précise, des solutions périodiques de (PPS) sont exhibées (de manière explicite) dans [115] pour des formes très particulières du taux de décharge p.

Bien que le système (PPS) présente des caractéristiques qualitatives intéressantes du point de vue de la modélisation de réseaux de neurones, seule une approche heuristique est invoquée dans [114] pour justifier son introduction. Une partie de ce manuscrit se propose donc de justifier, de manière rigoureuse, la dérivation du système (PPS) à partir de modèles microscopiques de neurones en interaction. Notons que cette problématique est très similaire à celle abordée dans [130].

Liens entre processus ponctuels et EDP : première approche. Cette première approche consiste à voir la densité $n(t,s)$ du système (PPS) comme l'espérance de son pendant aléatoire. Nous supposerons que la dynamique stochastique sous-jacente est caractérisée par un processus ponctuel N, d'intensité $(\lambda_t)_{t\geq 0}$, qui modélise le train de spikes d'un neurone typique du réseau.

Dans le chapitre 2 de ce manuscrit, nous étudions la loi de probabilité de l'âge S_t associé à un processus ponctuel N. Plus précisément, nous cherchons à caractériser sa dynamique, en fonction du temps t, au moyen d'une EDP qui se rapproche de (PPS).

Nous construisons dans un premier temps, de manière ad hoc, une mesure (aléatoire) en temps t et en âge s, notée $U(dt,ds)$, qui se veut être l'analogue aléatoire de $n(t,s)$. La procédure de thinning, présentée dans la section 1.1.d), permet d'écrire un système d'équations aux dérivées partielles stochastiques vérifié par la mesure U (voir la proposition 2.4.1). De plus, sous certaines hypothèses d'intégrabilité sur l'intensité $(\lambda_t)_{t\geq 0}$, l'espérance de la mesure U est bien définie et nous la notons u : on s'attend à ce que la mesure $u(dt,ds)$ soit l'analogue de la densité $n(t,s)$.

Le théorème 2.4.4 montre que u satisfait le système d'équations aux dérivées partielles (déterministe) suivant :

$$\begin{cases} \dfrac{\partial}{\partial t}u\left(dt,ds\right) + \dfrac{\partial}{\partial s}u\left(dt,ds\right) + \rho_{\lambda,\zeta_{N_-}}(t,s)u\left(dt,ds\right) = 0, \\ u\left(dt,0\right) = \displaystyle\int_0^{+\infty} \rho_{\lambda,\zeta_{N_-}}(t,s)u\left(t,ds\right)\, dt, \end{cases} \tag{B.2}$$

où $\rho_{\lambda,\zeta_{N_-}}(t,s) := \mathbb{E}[\lambda_t | S_{t-} = s]$. En comparant les systèmes (PPS) et (B.2), il apparaît que $n(t,s)$ est remplacé par $u(dt,ds)$ et que le terme $p(s,X(t))$ est remplacé par $\rho_{\lambda,\zeta_{N_-}}(t,s)$.

Notons que $\rho_{\lambda,\zeta_{N_-}}$ dépend de l'intensité du processus ponctuel sous-jacent $(\lambda_t)_{t\geq 0}$ ainsi que de la condition initiale ζ_{N_-} (à savoir, la loi de $N \cap \mathbb{R}_-$). À notre connaissance, le coefficient $\rho_{\lambda,\zeta_{N_-}}(t,s)$ ne peut pas être explicité dans le cas général, mais :

- si l'intensité ne dépend que du temps t et du dernier point, i.e. $\lambda_t = f(t, S_{t-})$, alors, quelque soit ζ_{N_-}, $\rho_{\lambda,\zeta_{N_-}}(t,s) = f(t,s)$ et le système (B.2) est linéaire. Dans ce cas particulier, le système (B.2) est en fait l'équation de Fokker-Planck associée au processus markovien $(S_{t-})_{t\geq 0}$. Ainsi, le théorème 2.4.4 peut être vu comme la généralisation de l'équation de Fokker-Planck à des dynamiques sous-jacentes non markoviennes, ce qui est le cas pour les processus de Hawkes.

- De plus, dans le cas particulier où le processus ponctuel sous-jacent est un processus de Hawkes linéaire, sa structure de branchement (voir la section 2.8.c)) permet de mieux appréhender l'espérance conditionnelle $\mathbb{E}[\lambda_t | S_{t-} = s]$. La conséquence pour le système (B.2) est la suivante (voir la proposition 2.5.4) :

  > La fonction (de survie) v définie par $v(t,s) := \int_s^{+\infty} u(t, d\sigma)$ vérifie un système d'EDP fermé (ce qui n'est pas le cas pour le système (1.9)).

Liens entre processus ponctuels et EDP : seconde approche. Cette seconde approche est abordée dans les chapitres 3 et 4 de ce manuscrit. Elle consiste à étudier la limite de champ-moyen d'un réseau de processus de Hawkes dépendants de l'âge. Un réseau de processus de Hawkes dépendants de l'âge (ADHP) de paramètres (n, h, Ψ) est un processus ponctuel multivarié $(N^{n,i})_{i=1,..,n}$ dont les intensités sont respectivement

$$\lambda_t^{n,i} = \Psi\left(S_{t-}^{n,i}, \frac{1}{n}\sum_{j=1}^{n} \int_0^{t-} h(t-z) N^{n,j}(dz)\right), \tag{B.3}$$

où $(S_{t-}^{n,i})_{t\geq 0}$ est le processus d'âge prévisible associé à $N^{n,i}$. Notons que l'ajout de cette dépendance par rapport à l'âge permet de contrecarrer l'une des principales carences du processus de Hawkes dans son utilisation en neurosciences : l'absence d'une description simple du phénomène de période réfractaire.

Basé sur un argument de couplage, le théorème 3.4.1 montre que des ADHP peuvent être approchés (dans la limite $n \to +\infty$) par des processus ponctuels indépendants et identiquement distribués, notés $\overline{N}^i$ pour $i \geq 1$.

En plus des processus ponctuels, l'approximation des processus d'âge est également possible. En particulier, nous montrons la "loi des grands nombres" suivante : la mesure empirique des âges $\overline{\mu}_{S_t}^n := n^{-1}\sum_{i=1}^n \delta_{S_t^{n,i}}$ converge vers P_t (la loi de l'âge $\overline{S}_t^1$ associé à $\overline{N}^1$). De plus, cette loi de probabilité P_t permet de faire le lien avec l'EDP (PPS).

> Si la condition initiale $u(0,\cdot) = u^{\text{in}}$ est une densité, alors l'unique solution u du système suivant,
>
> $$\begin{cases} \dfrac{\partial u(t,s)}{\partial t} + \dfrac{\partial u(t,s)}{\partial s} + \Psi(s, X(t))\, u(t,s) = 0, \\ u(t,0) = \displaystyle\int_{s\in\mathbb{R}} \Psi(s, X(t))\, u(t,s)\, ds, \end{cases} \tag{B.4}$$
>
> où $X(t) = \int_0^t h(t-z)u(z,0)dz$, est telle que $u(t,\cdot)$ est la densité de P_t.

Notons que le système ci-dessus est identique au système (PPS) à condition de remplacer $u(t,s)$ par $n(t,s)$, Ψ par p et h par d.

Ayant obtenu un résultat de type "loi des grands nombres" pour la mesure empirique des âges, il est naturel de chercher le résultat de type "théorème central limite" associé. L'étude menée dans le chapitre 4 porte donc sur les "petites" fluctuations de $\overline{\mu}^n_{S_t}$ autour de sa limite P_t. Considérons le processus de fluctuation $(\eta^n_t)_{t\geq 0}$ défini par $\eta^n_t = \sqrt{n}(\overline{\mu}^n_{S_t} - P_t)$. En s'inspirant de travaux de Méléard et de ses co-auteurs [50, 82, 101], nous considérons η^n_t comme un élément du dual d'un espace de Sobolev à poids. En utilisant de la compacité et l'unicité du point limite, nous montrons la convergence en loi de la suite de processus $(\eta^n)_{n\geq 1}$ vers un processus gaussien (théorème 4.5.12). Pour finir, notons que ce résultat, bien que technique, peut permettre de dériver une version bruitée de (B.4) qui intègre le terme de fluctuation gaussienne et se trouve donc être plus fidèle à la dynamique de la mesure empirique $\overline{\mu}^n_{S_t}$.

Détection de synchronisations : problématique. Les synchronisations d'activité entre neurones permettent une optimisation de l'énergie nécessaire à la propagation de l'information dans le réseau neuronal et sont caractéristiques de certains moments précis d'une tâche sensorimotrice. L'utilisation de multi-électrodes permet l'enregistrement des trains de spikes correspondant à plusieurs neurones en simultané. La recherche de procédure statistiques permettant de traiter efficacement ce genre de données est florissante, et la dernière partie de ce manuscrit en fait partie. Nous construisons une procédure statistique qui puisse détecter des synchronisations d'activité entre plusieurs neurones.

Nous nous focalisons sur la méthode des *Unitary Events* (UE) introduite dans la thèse de Grün [62]. Cette méthode est basée sur l'idée qu'une dépendance excitatrice du neurone A vers le neurone B doit être caractérisée par une surabondance du motif suivant : spike du neurone A suivi (dans un délai très court de l'ordre de la milliseconde) d'un spike du neurone B. Si un tel motif est sur-représenté, il est alors qualifié de Unitary Event. Nous utilisons cette idée pour construire un test d'indépendance.

Détection de synchronisations : construction d'un test d'indépendance. Nous généralisons (en partie) la méthode statistique introduite dans [158] au cas de $n \geq 2$ processus ponctuels représentant les trains de spikes de n neurones. Pour ce faire, nous introduisons une notion de coïncidence, indicateur statistique fortement inspiré de la notion de Unitary Event mais adapté au cadre des processus ponctuels. Le schéma de construction du test est par la suite relativement standard :

1. nous calculons l'espérance et la variance de notre indicateur statistique (le nombre de coïncidences) sous l'hypothèse que les trains de spikes sont des processus de Poisson homogènes indépendants,

2. l'utilisation du théorème central limite et une étape de *plug-in* permettent de déduire le comportement asymptotiquement gaussien de notre indicateur.

Ces deux étapes permettent de construire un test statistique dont le niveau asymptotique est contrôlé.

Pour finir, ces résultats théoriques sont illustrés via une étude par simulation, et les résultats de l'application de notre procédure statistique à des données réelles sont présentés.

Bibliographie

[1] M. Abeles. Quantification, smoothing, and confidence limits for single-units' histograms. *Journal of Neuroscience Methods*, 5(4) :317–325, 1982.

[2] R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[3] A. M. H. J. Aertsen, G. L. Gerstein, M. K. Habib, and G. Palm. Dynamics of Neuronal Firing Correlation : Modulation of "Effective Connectivity". *Journal of Neurophysiology*, 61 :900–917, 1989.

[4] M. Albert. *Tests d'indépendance par bootstrap et permutation : étude asymptotique et non-asymptotique. Application en neurosciences.* PhD thesis, Nice, 2015.

[5] M. Albert, Y. Bouret, M. Fromont, and P. Reynaud-Bouret. Bootstrap and permutation tests of independence for point processes. *Ann. Statist.*, 43(6) :2537–2564, 12 2015.

[6] E. W. Archer, I. M. Park, and J. W. Pillow. Bayesian entropy estimation for binary spike train data using parametric prior knowledge. In *Advances in Neural Information Processing Systems*, pages 1700–1708, 2013.

[7] E. Bacry, K. Dayri, and J.-F. Muzy. Non-parametric kernel estimation for symmetric Hawkes processes. Application to high frequency financial data. *The European Physical Journal B-Condensed Matter and Complex Systems*, 85(5) :1–12, 2012.

[8] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy. Scaling limits for Hawkes processes and application to financial statistics, Feb. 2012.

[9] P. Bao, H.-W. Shen, X. Jin, and X.-Q. Cheng. Modeling and predicting popularity dynamics of microblogs using self-excited Hawkes processes. *arXiv preprint arXiv:1503.02754*, 2015.

[10] H. B. Barlow. Single Units and Sensation : A Neuron Doctrine for Perceptual Psychology ? *Perception*, 1 :371–394, 1972.

[11] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B. Methodological*, 57(1) :289–300, 1995.

[12] P. Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[13] C. Borgelt and D. Picado-Muino. Finding frequent patterns in parallel point processes. In *Advances in Intelligent Data Analysis XII*, pages 116–126. Springer, 2013.

[14] M. Bossy, N. Champagnat, et al. Markov processes and parabolic partial differential equations. *Encyclopedia of Quantitative Finance*, pages 1142–1159, 2010.

[15] O. Boxma, D. Perry, W. Stadje, and S. Zacks. A markovian growth-collapse model. *Advances in applied probability*, pages 221–243, 2006.

[16] P. Brémaud. *Point processes and queues.* Springer-Verlag, New York, 1981. Martingale dynamics, Springer Series in Statistics.

[17] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *Ann. Probab.*, 24(3) :1563–1588, 1996.

[18] D. Brillinger, H. Bryant, and J. Segundo. Identification of synaptic interactions. *Biol. Cybernetics*, 22 :213–228, 1976.

[19] E. Brown, R. Barbieri, V. Ventura, R. Kass, and L. Frank. The time rescaling theorem and its application to neural spike train analysis. *Neural Computation*, 14(2) :325–346, 2002.

[20] E. N. Brown, R. E. Kass, and P. P. Mitra. Multiple neural spike train data analysis : state-of-the-art and future challenges. *Nature neuroscience*, 7(5) :456–461, 2004.

[21] N. Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, 8 :183–208, 2000.

[22] N. Brunel and V. Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural computation*, 11(7) :1621–1671, 1999.

[23] M. J. Cáceres, J. A. Carrillo, and B. Perthame. Analysis of nonlinear noisy integrate & fire neuron models : blow-up and steady states. *J. Math. Neurosci*, 1(7), 2011.

[24] M. J. Cáceres, J. A. Carrillo, and L. Tao. A numerical solver for a nonlinear Fokker-Planck equation representation of neuronal network dynamics. *J. Comp. Phys.*, 230 :1084–1099, 2011.

[25] M. J. Cáceres and B. Perthame. Beyond blow-up in excitatory integrate and fire neuronal networks : refractory period and spontaneous activity. *Journal of theoretical biology*, 350 :81–89, 2014.

[26] J. A. Cañizo, J. A. Carrillo, and S. Cuadrado. Measure solutions for some models in population dynamics. *Acta applicandae mathematicae*, 123(1) :141–156, 2013.

[27] G. Casella and R. Berger. *Statistical Inference*. Duxbury, 2002.

[28] J. Chevallier. Mean-field limit of generalized Hawkes processes. *arXiv preprint arXiv:1510.05620*, 2015.

[29] J. Chevallier, M. J. Cáceres, M. Doumic, and P. Reynaud-Bouret. Microscopic approach of a time elapsed neural model. *Mathematical Models and Methods in Applied Sciences*, 25(14) :2669–2719, 2015.

[30] J. Chevallier and T. Laloë. Detection of dependence patterns with delay. *Biometrical Journal*, 57(6) :1110–1130, 2015.

[31] E. Chornoboy, L. Schramm, and A. Karr. Maximum likelihood identification of neural point process systems. *Biological Cybernetics*, 59(4-5) :265–275, 1988.

[32] B. Cloez. Limit theorems for some branching measure-valued processes, Oct. 2012.

[33] A. Compte, N. Brunel, P. S. Goldman-Rakic, and X.-J. Wang. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex 10*, 10 :910–923, 2000.

[34] M. Costa. A piecewise deterministic model for prey-predator communities. *arXiv preprint arXiv:1503.02492*, 2015.

[35] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41) :15649–15653, 2008.

[36] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 152. Cambridge university press, 2014.

[37] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2003. Elementary theory and methods.

[38] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. II.* Probability and its Applications (New York). Springer, New York, second edition, 2008. General theory and structure.

[39] M. H. Davis. Piecewise-deterministic markov processes : A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 353–388, 1984.

[40] P. Dayan and L. F. Abbott. *Theoretical neuroscience*, volume 806. Cambridge, MA : MIT Press, 2001.

[41] F. Delarue, J. Inglis, S. Rubenthaler, and E. Tanré. Particle systems with a singular mean-field self-excitation. application to neuronal networks. *Stochastic Processes and their Applications*, 125(6) :2451–2492, 2015.

[42] F. Delarue, J. Inglis, S. Rubenthaler, E. Tanré, et al. Global solvability of a networked integrate-and-fire model of mckean–vlasov type. *The Annals of Applied Probability*, 25(4) :2096–2133, 2015.

[43] S. Delattre, N. Fournier, and M. Hoffmann. Hawkes processes on large networks. *Ann. Appl. Probab.*, 26(1) :216–261, 02 2016.

[44] S. Ditlevsen and E. Löcherbach. Multi-class oscillating systems of interacting neurons. *arXiv preprint arXiv:1512.00265*, 2015.

[45] S. Ditlevsen, A. Samson, et al. Estimation in the partially observed stochastic Morris–Lecar neuronal model with particle filter and stochastic approximation methods. *The Annals of Applied Statistics*, 8(2) :674–702, 2014.

[46] Y. Dong, S. Mihalas, F. Qiu, R. von der Heydt, and E. Niebur. Synchrony and the binding problem in macaque visual cortex. *Journal of vision*, 8(7) :30, 2008.

[47] M. Doumic, M. Hoffmann, N. Krell, L. Robert, et al. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli*, 21(3) :1760–1799, 2015.

[48] O. Faugeras and J. Maclaurin. Asymptotic description of stochastic neural networks. i. existence of a large deviation principle. *Comptes Rendus Mathematique*, 352(10) :841–846, 2014.

[49] O. Faugeras, J. Touboul, and B. Cessac. A constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs. *Frontiers in computational neuroscience*, 3, 2009.

[50] B. Fernandez and S. Méléard. A Hilbertian approach for fluctuations on the McKean-Vlasov model. *Stochastic Process. Appl.*, 71(1) :33–53, 1997.

[51] N. Fournier and A. Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, pages 1–32, 2014.

[52] N. Fournier and E. Löcherbach. On a toy model of interacting neurons. *arXiv preprint arXiv:1410.3263*, 2014.

[53] M. Fromont, B. Laurent, P. Reynaud-Bouret, et al. The two-sample problem for poisson processes : Adaptive tests with a nonasymptotic wild bootstrap approach. *The Annals of Statistics*, 41(3) :1431–1461, 2013.

[54] A. Galves and E. Löcherbach. Infinite systems of interacting chains with memory of variable length - a stochastic model for biological neural nets. *Journal of Statistical Physics*, 151(5) :896–921, 2013.

[55] A. Galves and E. Löcherbach. Modeling networks of spiking neurons as interacting processes with memory of variable length. *arXiv preprint arXiv:1502.06446*, 2015.

[56] G. L. Gerstein, A. M. Aertsen, et al. Representation of cooperative firing activity among simultaneously recorded neurons. *Journal of Neurophysiology*, 54(6) :1513–1528, 1985.

[57] G. L. Gerstein and D. H. Perkel. Simultaneously Recorded Trains of Action Potentials : Analysis and Functional Interpretation. *Science*, 164(3881) :828–830, 1969.

[58] G. Giacomin, E. Luçon, and C. Poquet. Coherence stability and effect of random natural frequencies in populations of coupled oscillators. *Journal of Dynamics and Differential Equations*, 26(2) :333–367, 2014.

[59] G. Giacomin, K. Pakdaman, X. Pellegrin, and C. Poquet. Transitions in active rotator systems : invariant hyperbolic manifold approach. *SIAM Journal on Mathematical Analysis*, 44(6) :4165–4194, 2012.

[60] R. D. Gill, N. Keiding, and P. K. Andersen. *Statistical models based on counting processes*. Springer, 1997.

[61] F. Grammont and A. Riehle. Spike synchronization and firing rate in a population of motor cortical neurons in relation to movement direction and reaction time. *Biological Cybernetics*, 88(5) :360–373, 2003.

[62] S. Grün. *Unitary joint events in multiple neuron spiking activity : detection, significance, and interpretation*. PhD thesis, 1996.

[63] S. Grün. Data-driven significance estimation for precise spike correlation. *Journal of Neurophysiology*, 101(3) :1126–1140, 2009.

[64] S. Grün, M. Diesmann, and A. Aertsen. Unitary Events in Multiple Single-Neuron Spiking Activity : I. Detection and Significance. *Neural Computation*, 14(1) :43–80, 2002.

[65] S. Grün, M. Diesmann, and A. Aertsen. Unitary Events in Multiple Single-Neuron Spiking Activity : II. Nonstationary Data. *Neural Computation*, 14(1) :81–119, 2002.

[66] S. Grün, M. Diesmann, F. Grammont, A. Riehle, and A. Aertsen. Detecting unitary events without discretization of time. *Journal of neuroscience methods*, 94(1) :67–79, 1999.

[67] G. Gusto and S. Schbath. FADO : A Statistical Method to Detect Favored or Avoided Distances between Occurrences of Motifs using the Hawkes' Model. *Statistical Applications in Genetics and Molecular Biology*, 4(1), 2005.

[68] N. R. Hansen, P. Reynaud-Bouret, and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1) :83–143, 2015.

[69] M. T. Harrison and S. Geman. A rate and history-preserving resampling algorithm for neural spike trains. *Neural Computation*, 21(5) :1244–1258, 2009.

[70] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1) :83–90, 1971.

[71] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, pages 493–503, 1974.

[72] D. O. Hebb. *The Organization of Behavior : A Neuropsychological Theory*. Wiley, New York, 1949.

[73] P. Hodara and E. Löcherbach. Hawkes processes with variable length memory and an infinite number of components. *arXiv preprint arXiv:1410.5235*, 2014.

[74] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4) :500–544, 1952.

[75] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical analysis and modelling of spatial point patterns*, volume 70. John Wiley & Sons, 2008.

[76] J. Inglis and D. Talay. Mean-field limit of a stochastic particle system smoothly interacting through threshold hitting-times and applications to neural networks with dendritic component. *SIAM Journal on Mathematical Analysis*, 47(5) :3884–3916, 2015.

[77] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[78] P. Jahn, R. Berg, J. Hounsgaard, and S. Ditlevsen. Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *Journal of Computational Neuroscience*, 31 :563–579, 2011.

[79] P. Jahn, R. W. Berg, J. Hounsgaard, and S. Ditlevsen. Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *Journal of computational neuroscience*, 31(3) :563–579, 2011.

[80] T. Jaisson and M. Rosenbaum. Limit theorems for nearly unstable hawkes processes. *The Annals of Applied Probability*, 25(2) :600–631, 2015.

[81] A. Joffe and M. Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Advances in Applied Probability*, pages 20–65, 1986.

[82] B. Jourdain and S. Méléard. Propagation of chaos and fluctuations for a moderate model with smooth initial data. In *Annales de l'IHP Probabilités et statistiques*, volume 34, pages 727–766, 1998.

[83] S. Jovanović, J. Hertz, and S. Rotter. Cumulants of hawkes point processes. *Physical Review E*, 91(4) :042802, 2015.

[84] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 1955, pages 171–197, 1954.

[85] Y. Y. Kagan. Statistical distributions of earthquake numbers : consequence of branching process. *Geophysical Journal International*, 180(3) :1313–1328, 2010.

[86] J. F. C. Kingman. *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1993. Oxford Science Publications.

[87] A. Klenke. *Probability theory : a comprehensive course*. Springer, 2007.

[88] M. Krumin, I. Reutsky, and S. Shoham. Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. *Frontiers in computational neuroscience*, 4, 2010.

[89] C. Lai. An example of Wold's Point Processes with Markov-Dependent Intervals. *Journal of Applied Probability*, 15(4) :748–758, 1978.

[90] F. Lavancier, J. Møller, and E. Rubak. Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 77(4) :853–877, 2015.

[91] E. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. Springer-Verlag New York Inc, 3rd edition, 2005.

[92] P. A. W. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3) :403–413, 1979.

[93] T. J. Liniger. *Multivariate hawkes processes*. PhD thesis, Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 18403, 2009, 2009.

[94] S. Louis, C. Borgelt, and S. Grün. Generation and selection of surrogate methods for correlation analysis. In *Analysis of Parallel Spike Trains*, pages 359–382. Springer, 2010.

[95] S. Louis, G. L. Gerstein, S. Grün, and M. Diesmann. Surrogate spike train generation through dithering in operational time. *Frontiers in computational neuroscience*, 4, 2010.

[96] E. Luçon and W. Stannat. Mean field limit for disordered diffusions with singular interactions. *The Annals of Applied Probability*, 24(5) :1946–1993, 2014.

[97] E. Luçon and W. Stannat. Transition from gaussian to non-gaussian fluctuations for mean-field diffusions in spatial interaction. *arXiv preprint arXiv:1502.00532*, 2015.

[98] L. Massoulié. Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic processes and their applications*, 75(1) :1–30, 1998.

[99] M. Mattia and P. Del Giudice. Population dynamics of interacting spiking neurons. *Phys. Rev. E*, 66 :051917, 2002.

[100] P. Mattila. *Geometry of sets and measures in Euclidean spaces : fractals and rectifiability*. Number 44. Cambridge University Press, 1999.

[101] S. Méléard. Asymptotic behaviour of some interacting particle systems ; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 42–95. Springer, Berlin, 1996.

[102] F. Merlevède, M. Peligrad, et al. Rosenthal-type inequalities for the maximum of partial sums of stationary processes and examples. *The Annals of Probability*, 41(2) :914–960, 2013.

[103] M. Métivier. *Semimartingales : a course on stochastic processes*, volume 2. Walter de Gruyter, 1982.

[104] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493), 2011.

[105] J. Møller and J. G. Rasmussen. Perfect simulation of hawkes processes. *Advances in Applied Probability*, pages 629–646, 2005.

[106] J. Møller and J. G. Rasmussen. Approximate simulation of hawkes processes. *Methodology and Computing in Applied Probability*, 8(1) :53–64, 2006.

[107] J. Moller and R. P. Waagepetersen. *Statistical inference and simulation for spatial point processes*. CRC Press, 2003.

[108] D. P. Muiño and C. Borgelt. Frequent item set mining for sequential data : Synchrony in neuronal spike trains. *Intelligent Data Analysis*, 18(6) :997–1012, 2014.

[109] I. Muni Toke and F. Pomponio. Modelling trades-through in a limited order book using hawkes processes. *Economics discussion paper*, (2011-32), 2011.

[110] Y. Ogata. On Lewis simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1) :23–30, 1981.

[111] Y. Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2) :379–402, 1998.

[112] A. Olivier. *Analyse statistique des modèles de croissance-fragmentation*. PhD thesis, Paris 9, 2015.

[113] A. Omurtag, K. B. W., and L. Sirovich. On the simulation of large populations of neurons. *J. Comp. Neurosci.*, 8 :51–63, 2000.

[114] K. Pakdaman, B. Perthame, and D. Salort. Dynamics of a structured neuron population. *Nonlinearity*, 23(1) :55, 2010.

[115] K. Pakdaman, B. Perthame, and D. Salort. Relaxation and self-sustained oscillations in the time elapsed neuron network model. *SIAM Journal on Applied Mathematics*, 73(3) :1260–1279, 2013.

[116] K. Pakdaman, B. Perthame, and D. Salort. Adaptation and fatigue model for neuron networks and large time asymptotics in a nonlinear fragmentation equation. *The Journal of Mathematical Neuroscience (JMN)*, 4(1) :1–26, 2014.

[117] G. Palm. Cell assemblies as a guideline for brain research. *Concepts in Neuroscience*, 1(1) :133–147, 1990.

[118] D. H. Perkel, G. L. Gerstein, and G. P. Moore. Neuronal Spike Trains and Stochastic Point Processes : II. Simultaneous Spike Trains. *Biophysical Journal*, 7(4) :419–440, 1967.

[119] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter. How structure determines correlations in neuronal networks. *Public Library of Science : Computational Biology*, 7(5) :e1002059, 2011.

[120] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter. Recurrent interactions in spiking networks with arbitrary topology. *Physical Review E*, 85(3) :031916, 2012.

[121] B. Perthame. *Transport equations in biology*. Springer Science & Business Media, 2006.

[122] J. Pham, K. Pakdaman, J. Champagnat, and J.-F. Vibert. Activity in sparsely connected excitatory neural networks : effect of connectivity. *Neural Networks*, 11(3) :415–434, 1998.

[123] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. Chichilnisky, and E. P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207) :995–999, 2008.

[124] G. Pipa, M. Diesmann, and S. Grün. Significance of joint-spike events based on trial-shuffling by efficient combinatorial methods. *Complexity*, 8(4) :79–86, 2003.

[125] G. Pipa and S. Grün. Non-parametric significance estimation of joint-spike events by shuffling and resampling. *Neurocomputing*, 52 :31–37, 2003.

[126] G. Pipa, S. Grün, and C. van Vreeswijk. Impact of Spike Train Autostructure on Probability Distribution of Joint Spike Events. *Neural Computation*, 25 :1123–1163, 2013.

[127] C. Pouzat and A. Chaffiol. Automatic spike train analysis and report generation. an implementation with r, {R2HTML} and {STAR}. *Journal of Neuroscience Methods*, 181(1) :119 – 144, 2009.

[128] C. Prévôt and M. Röckner. *A concise course on stochastic partial differential equations*, volume 1905. Springer, 2007.

[129] D. Purves, G. J. Augustine, D. Fitzpatrick, L. C. Katz, A.-S. LaMantia, J. O. McNamara, and S. M. Williams. Ion channels underlying action potentials. 2001.

[130] C. Quininao. A microscopic spiking neuronal network for the age-structured model. *arXiv preprint arXiv:1503.00115*, 2015.

[131] A. H. Rafati, F. Safavimanesh, K.-A. DORPH-PETERSEN, J. G. Rasmussen, J. Møller, and J. R. Nyengaard. Detection and spatial characterization of minicolumnarity in the human cerebral cortex. *Journal of microscopy*, 261(1) :115–126, 2015.

[132] R. Rebolledo. Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete*, 51(3) :269–286, 1980.

[133] I. C. Reimer, B. Staude, W. Ehm, and S. Rotter. Modeling and analyzing higher-order correlations in non-Poissonian spike trains. *Journal of neuroscience methods*, 208(1) :18–33, 2012.

[134] A. Renart, N. Brunel, and X.-J. Wang. Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks. In J. Feng, editor, *Computational Neuroscience : A comprehensive approach*. Chapman & Hall/CRC Mathematical Biology and Medicine Series, 2004.

[135] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion (Grundlehren der mathematischen Wissenschaften)*. Springer-Verlag, 3rd edition, 1999.

[136] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *The Journal of Mathematical Neuroscience (JMN)*, 4(1) :1–41, 2014.

[137] P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot. Inference of functional connectivity in Neurosciences via Hawkes processes. In *1st IEEE Global Conference on Signal and Information Processing*, 2013, Austin Texas.

[138] P. Reynaud-Bouret and E. Roy. Some non asymptotic tail estimates for hawkes processes. *Bulletin of the Belgian Mathematical Society-Simon Stevin*, 13(5) :883–896, 2007.

[139] P. Reynaud-Bouret, S. Schbath, et al. Adaptive estimation for hawkes processes ; application to genome analysis. *The Annals of Statistics*, 38(5) :2781–2822, 2010.

[140] M. G. Riedler, M. Thieullen, and G. Wainrib. Limit theorems for infinite-dimensional piecewise deterministic markov processes. applications to stochastic excitable membrane models. *Electron. J. probab*, 17(55) :1–48, 2012.

[141] A. Riehle, F. Grammont, M. Diesmann, and S. Grün. Dynamical changes and temporal precision of synchronized spiking activity in monkey motor cortex during movement preparation. *Journal of Physiology-Paris*, 94(5) :569–582, 2000.

[142] A. Riehle, F. Grammont, and W. A. MacKay. Cancellation of a planned movement in monkey motor cortex. *Neuroreport*, 17(3) :281–285, 2006.

[143] F. Roueff, R. Von Sachs, and L. Sansonnet. Time-frequency analysis of locally stationary hawkes processes. prepublication on HAL, 2015.

[144] A. Roy, P. N. Steinmetz, S. S. Hsiao, K. O. Johnson, and E. Niebur. Synchrony : a neural correlate of somatosensory attention. *Journal of neurophysiology*, 98(3) :1645–1661, 2007.

[145] Y. Sakurai. How do cell assemblies encode information in the brain ? *Neuroscience & Biobehavioral Reviews*, 23(6) :785–796, 1999.

[146] L. Schwartz. *Théorie des distributions.* Publications de l'Institut de Mathématique de l'Université de Strasbourg, No. IX-X. Nouvelle édition, entiérement corrigée, refondue et augmentée. Hermann, Paris, 1966.

[147] S. Shinomoto. Estimating the Firing Rate. In S. Grün and S. Rotter, editors, *Analysis of Parallel Spike Trains*, volume 7 of *Springer Series in Computational Neuroscience*, pages 21–35. Springer US, 2010.

[148] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*, volume 59. Siam, 2009.

[149] W. Singer. Synchronization of Cortical Activity and its Putative Role in Information Processing and Learning. *Annual Review of Physiology*, 55 :349–374, 1993.

[150] L. Sirovich, A. Omurtag, and K. Lubliner. Dynamics of neural populations : Stability and synchrony. *Network : Computation in Neural Systems*, 17 :3–29, 2006.

[151] G. Smith. Modeling the stochastic gating of ion channels. *Computational cell biology*, pages 285–319, 2002.

[152] H. Spohn. Kinetic equations from hamiltonian dynamics : Markovian limits. *Reviews of Modern Physics*, 52(3) :569, 1980.

[153] H. Spohn. *Large scale dynamics of interacting particles.* Springer Science & Business Media, 2012.

[154] E. Stark and M. Abeles. Unbiased estimation of precise temporal correlations between spike trains. *Journal of neuroscience methods*, 179(1) :90–100, 2009.

[155] A.-S. Sznitman. Équations de type de boltzmann, spatialement homogenes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 66(4) :559–592, 1984.

[156] A.-S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.

[157] W. Thompson. *Point process models with applications to safety and reliability.* Springer Science & Business Media, 2012.

[158] C. Tuleau-Malot, A. Rouis, F. Grammont, and P. Reynaud-Bouret. Multiple Tests Based on a Gaussian Approximation of the Unitary Events Method with delayed coincidence count. appearing in *Neural Computation*, 26 :7, 2014.

[159] V. Ventura, R. Carta, R. Kass, S. Gettner, and C. Olson. Statistical analysis of temporal evolution in single-neuron firing rates. *Biostatistics*, 3(1) :1–20, 2002.

[160] C. von der Malsburg. The Correlation Theory of Brain Function. Internal Report 81-2, Department of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany, 1981.

[161] G. Wainrib. *Randomness in neurons : a multiscale probabilistic analysis*. PhD thesis, PhD thesis, 2010.

[162] K. Yosida. *Functional analysis*, volume 123 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin-New York, sixth edition, 1980.

[163] L. Zhu. Central limit theorem for nonlinear hawkes processes. *Journal of Applied Probability*, 50(3) :760–771, 2013.

[164] L. Zhu. *Nonlinear Hawkes Processes*. PhD thesis, New York University, 2013.

Index