UNIVERSITÉ
de Cergy-Pontoise

THÈSE DE DOCTORAT DE
l'UNIVERSITÉ DE CERGY-PONTOISE

Mathématiques

Ecole doctorale Économie, Management, Mathématiques et Physique

Présentée par

Pierre HODARA

Pour obtenir le grade de
DOCTEUR de l'UNIVERSITÉ de CERGY-PONTOISE

Systèmes de neurones en interactions. Modélisation probabiliste et estimation.

Président du Jury :	**Paul Doukhan**, Université de Cergy-Pontoise
Rapporteurs :	**Florent Malrieu**, Université de Tours
	Patricia Reynaud-Bouret, Université de Nice Sophia-Antipolis
Examinateur :	**Antonio Galves**, Université de Sao Paulo
Directrices de thèse :	**Nathalie Krell**, Université de Rennes I
	Eva Löcherbach, Université de Cergy-Pontoise

à Florence

Remerciements

Mes remerciements et ma gratitude vont à mes directrices de thèse Eva Löcherbach et Nathalie Krell. Eva a su me proposer un sujet motivant, je la remercie chaleureusement pour sa disponibilité et sa bienveillance. C'est avec un réel plaisir que j'ai travaillé avec elle pendant ces 3 années. Nathalie s'est intégrée au cours de cette aventure avec un investissement n'ayant pour seule limite que mes réponses parfois tardives. Je la remercie pour sa patience, pour son accueil chaleureux à Rennes, et pour tous ses excellents conseils. Je les remercie également toutes les deux pour m'avoir permis de participer à de nombreuses conférences, sources d'enrichissement personnel et scientifique.

Mes remerciements vont ensuite à Patricia Reynaud-Bouret et à Florent Malrieu qui ont accepté d'être rapporteurs de cette thèse. Je remercie également Paul Doukhan et Antonio Galves pour avoir accepté d'être membre du jury et d'évaluer cette thèse.

Le Labex MME-DII (ANR11-LBX-0023-01) m'a apporté un soutient financier dans le cadre de cette thèse, j'en remercie les coordinateurs.

Je tiens à remercier Gabriel Lang pour toutes les discutions mathématiques dont il m'a fait profiter. C'est quelqu'un que j'apprécie particulièrement, notamment pour son humour absurde. Il n'est pas étranger à ma vocation de matheux.

Je remercie l'ensemble du personnel du département de mathématiques de Cergy au sein duquel j'ai pu apprécier une ambiance conviviale. Merci à tous les thésards, à Thibault, Ayman et Mouhamadou avec qui j'ai partagé mon bureau et des amitiés sincères, à Pierre-Damien, Jean-François, Alexandre Borgniet et Alexandre Martin qui ont su rendre encore plus savoureux les menus du restaurant universitaire, à Thomas pour les pauses cigarettes qui arrivaient toujours à point. Merci à Smail Alili, Philipe Barbe, Philipe Gravejat, Alexandre Mizrahi, Emmanuel Volte avec qui j'ai pu collaborer dans le cadre de mes missions d'enseignement. Merci à Linda et Caroline pour leur disponibilité et leur capacité à fournir systématiquement une réponse immédiate et efficace à toutes mes questions d'ordre administratif.

Je remercie Romain Azaïs d'avoir pris du temps à Nancy et au cours d'une promenade dans l'arrière pays Montpelliérain pour me faire partager son expérience sur des problèmes d'estimation.

Je tiens surtout à remercier Florence et ma famille pour leur soutien et leur amour qui m'ont permis d'arriver jusqu'ici.

Table des matières

Résumé

L'objet de ce travail est de proposer différents modèles pour décrire l'activité d'un réseau de neurones. L'activité d'un neurone peut se mesurer à travers son potentiel de membrane. L'évolution de celui-ci présente de petites variations avec de temps à autres des dépolarisations brèves et brutales appelées potentiels d'action. L'enregistrement des instants auxquels différents neurones émettent des potentiels d'action est nommé "train de potentiels d'action" et permet également de mesurer l'activité du réseau. On présente ici deux types de modèles.

Le premier modèle décrit uniquement les instants auxquels on observe un potentiel d'action. Le type de processus utilisé est un processus de Hawkes à mémoire d'ordre variable. Ce type de processus nous permet de prendre en compte certaines considérations biologiques comme le fait que le potentiel de membrane d'un neurone revient toujours au même état d'équilibre après l'émission d'un potentiel d'action, quelque soit son état avant celle-ci. Cela donne donc un processus de comptage dont l'intensité, pour chaque neurone, dépend du passé de tout le processus jusqu'au dernier instant de saut du neurone considéré. C'est ce qui donne le caractère non-Markovien avec une mémoire d'ordre variable pouvant éventuellement remonter infiniment loin dans le passé. On montre que pour ce modèle il existe une unique mesure stationnaire en proposant une construction graphique du processus. Ceci est obtenu avec deux jeux d'hypothèses correspondant à deux modèles distincts : un premier modèle faisant intervenir un phénomène de saturation concernant la stimulation d'un neurone par un autre, et un deuxième modèle où on impose une structure au graphe d'interaction entre les neurones qui donne lieu à une cascade de potentiels d'action.

Le deuxième type de modèle décrit l'évolution continue du potentiel de membrane entre les potentiels d'action, ces derniers étant modélisés par des sauts qui remettent à zero le potentiel de membrane du neurone émetteur et augmentent le potentiel de membrane des autres neurones d'une quantité inversement proportionnelle au nombre de neurones dans le réseau. L'intensité de saut pour chaque neurone dépend uniquement de son potentiel de membrane, ce qui contribue à donner un caractère Markovien au processus utilisé qui fait partie de la classe des processus Markovien déterministes par morceaux (PDMP). On s'intéresse pour ce modèle à l'estimation de la fonction de taux de saut qui donne l'intensité du processus en fonction du potentiel de membrane. Dans ce but on commence par montrer l'existence d'une mesure invariante possédant une densité régulière et on construit un estimateur à noyau possédant une vitesse de convergence optimale pour une classe de fonction Hölderienne donnée.

Chapitre 1

Présentation des travaux de thèse

1.1 Introduction

L'objet de cette thèse est l'étude de modèles pour un réseau de neurones en interactions.

Un neurone est une cellule dont la membrane contient des canaux ioniques permettant de forcer le passage des ions à travers la membrane dans un sens ou dans l'autre. Ceci permet de créer un potentiel de membrane. Ce potentiel de membrane suit une évolution continue ponctuée par des dépolarisations brèves et brutales appelées potentiels d'action. Lorsqu'un neurone émet un potentiel d'action, ses synapses en contact avec les autres neurones libèrent des neurotransmetteurs qui sont captés par les neurones post-synaptiques grâce à des récepteurs spécifiques. La fixation de ces neuro-transmetteurs sur les récepteurs d'un neurone conduisent à une modification de son potentiel de membrane, l'augmentant ou le diminuant selon le caractère stimulateur ou inhibiteur de la synapse. Plus le potentiel de membrane d'un neurone est élevé, plus celui-ci est susceptible d'émettre un potentiel d'action. Ainsi, lorsque deux neurones sont en contacts par le biais d'un synapse stimulatrice, l'émission d'un potentiel d'action par le neurone pré-synaptique peut donner lieu à une émission par le neurone post-synaptique. Par ailleurs, les membranes de deux neurones distincts peuvent être en contact direct, formant ce qu'on appelle une "gap-junction", ce qui permet à un potentiel d'action de se propager directement d'un neurone à un autre. Ceci tend à synchroniser l'évolution des potentiels de membranes.

L'évolution du potentiel de membrane d'un neurone dépend donc de celle des autres neurones avec lesquels il est en contact. Cependant, après l'émission d'un potentiel d'action, le potentiel de membrane du neurone revient toujours à la même valeur. Ceci a pour effet de créer un phénomène de perte de mémoire : les potentiels d'actions émis par des neurones pré-synaptiques n'ont plus d'influence sur le potentiel de membrane du neurone post-synaptique après que celui-ci ait émis à son tour. De plus, chaque neurone possède un état d'équilibre vers lequel son potentiel de membrane évolue naturellement en l'absence d'interactions. Ceci a également pour effet un phénomène de perte de mémoire : pour les mêmes stimulations reçues, le potentiel de membrane d'un neurone sera plus élevé si celles-ci sont récentes.

La modélisation d'un réseau de neurones d'un point de vue probabiliste a été étudiée de différents points de vues. Voir [GK02] pour une introduction sur différents modèles. Dans [Yag16], l'auteur considère un potentiel de membrane à valeurs dans un espace dénombrable et modélise les interactions à l'aide de probabilités de transition d'un état à un autre. Dans [CCDRB15], les auteurs modélisent l'activité du réseau par un processus

de comptage donnant les instants d'émission de potentiels d'action, et font le lien entre différents types de processus de comptages permettant cette modélisation.

Le point de départ de mon travail de thèse est l'article [GL13]. Dans cet article, les auteurs proposent un modèle pour les réseaux de neurones en temps discret. En collaboration avec Eva Löcherbach, j'ai proposé une extension en temps continu de ce modèle. On obtient dans [HL14] l'existence et l'unicité d'une version stationnaire du processus pour deux différents jeux d'hypothèses, supposant d'une part l'existence de seuils de saturation et d'autre part une structure d'interactions en cascade. On donne notamment une construction graphique du processus stationnaire.

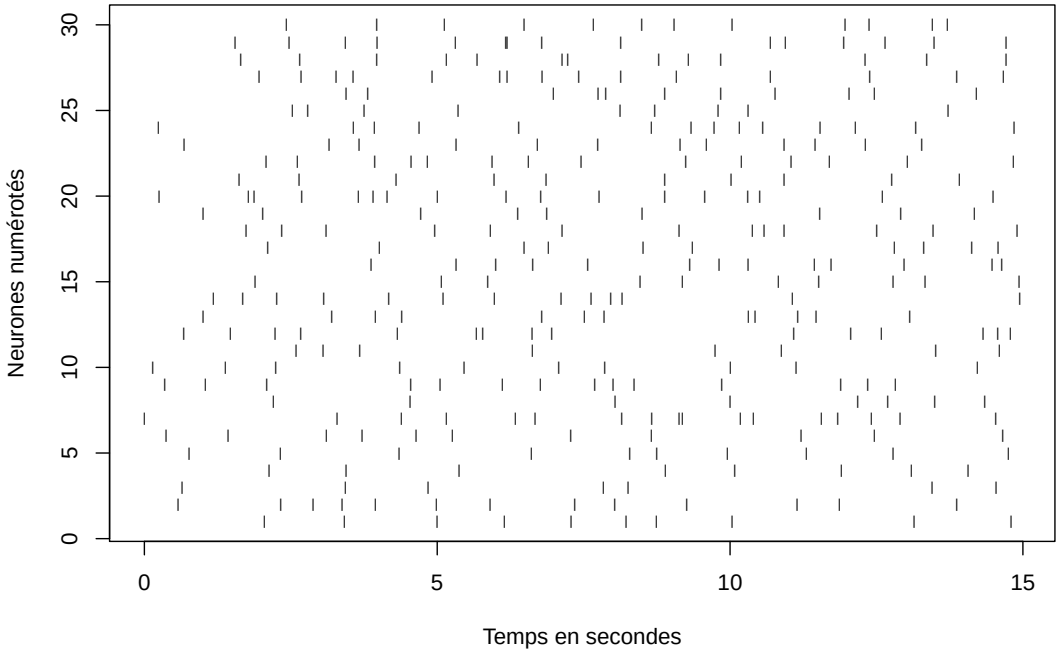
Par la suite, en collaboration avec Nathalie Krell et Eva Löcherbach j'ai travaillé sur l'estimation non-paramétrique du taux d'émission de potentiel d'action dans un modèle Markovien où le processus décrivant l'activité du réseau de neurones est un processus de Markov déterministe par morceaux (PDMP). Le taux d'émission de potentiels d'action correspond au taux de saut de notre PDMP pour lequel on construit un estimateur à noyau possédant une vitesse de convergence optimale pour l'erreur dans L^2.

1.2 Processus de Hawkes de dimension infinie à mémoire d'ordre variable.

1.2.1 Le modèle général.

L'activité d'un réseau de neurones peut se mesurer à travers un train de potentiels d'action, c'est à dire la donnée pour chaque neurone, de chaque instant auquel il émet un potentiel d'action.

Voici un exemple de train de potentiel d'action.



Ainsi, un processus de comptage est suffisant pour décrire une telle activité. Un

processus de comptage est caractérisé par son intensité. Dans la littérature différents modèles ont été proposés : des processus de Poisson homogènes ou inhomogènes ainsi que des processus de Hawkes, voir par exemple [RBRGTM14]. C'est ce dernier type de modèle que l'on retient pour modéliser l'activité du réseau de neurones dans le chapitre 2, car celui-ci permet de prendre en compte les interactions entre les neurones. Dans le cas le plus simple d'un processus de Hawkes N linéaire et univarié, l'intensité du processus à l'instant t s'écrit

$$\lambda(t) = \nu + \int_{-\infty}^{t^-} h(t-s)dN(s),$$

où ν est une constante qui donne l'intensité minimale du système, et h est une fonction de perte qui pondère chaque saut du processus en fonction du temps qui s'est écoulé depuis celui-ci. Dans les articles originaux [Haw72] et [HO74], introduisant le modèle, les auteurs considèrent des fonctions d'intensité linéaires. Des extensions au cas non-linéaire ont été étudiées dans [BM96], voir aussi [Mas98], où il est proposé une étude des propriétés de stabilité des processus de Hawkes multivariés non-linéaires. Les processus de Hawkes sont utilisés dans différents domaines d'application. Pour en citer quelques-uns, [DL16], [HRBR15] et [Che15] dans le cadre de trains de potentiels d'action de neurones, [RBS10] pour une application à l'analyse du génome. Dans un contexte différent, [JR15] obtiennent des théorème limites pour des processus de Hawkes appliqués à des modèles de finance. Pour une introduction générale aux processus de comptage, on incite le lecteur à consulter [DVJ03].

On étudie ici un processus de comptage infini-dimensionnel dont chaque coordonnée correspond à un neurone du réseau et dont les instants de sauts correspondent aux instants d'émission de potentiel d'action.

L'intensité de saut pour chaque coordonnée du processus dépend du passé de tout le processus jusqu'au dernier potentiel d'action émis par le neurone considéré (i.e. le dernier saut pour la coordonnée considérée). Ce choix permet de modéliser le fait que le potentiel de membrane d'un neurone après l'emission d'un potentiel d'action revient toujours au même potentiel d'équilibre. Ainsi, les potentiels d'actions émis par les autres neurones sont susceptibles de modifier le potentiel de membrane du neurone considéré, mais ces variations ne sont plus présentes dès lors que le neurone considéré émet lui-même un potentiel d'action.

A chaque instant t, l'intensité de saut d'un neurone i comptabilise les sauts du processus en les pondérant par le temps qui s'est écoulé entre leur apparition et l'instant t. Ceci permet de modéliser le phénomène de perte de mémoire.

De plus, afin de modéliser la nature et l'intensité de chaque interaction entre deux neurones on utilise des fonctions de poids synaptiques pour pondérer globalement l'influence de l'activité d'un neurone sur l'intensité de saut d'un autre.

Décrivons maintenant notre modèle plus en détail. On considère un ensemble dénombrable de neurones I. L'activité de chaque neurone $i \in I$ est décrite par un processus de comptage Z^i tel que pour chaque $-\infty < s < t < \infty$, $Z^i(]s,t])$ donne le nombre de potentiels d'action émis par le neurone i durant l'intervalle $]s,t]$. Ce processus de comptage $(Z^i, i \in I)$ est caractérisé par son intensité $(\lambda_t^i, i \in I)$ définie par la relation :

$$\mathbb{P}(Z^t \text{ saute durant } [\text{t , t + dt }] | \mathcal{F}_t) = \lambda_t^i dt, i \in I.$$

Ici, $\mathcal{F}_t$ désigne la tribu engendrée par $Z^i(]s,u]), s \leq u \leq t, i \in I$, et nous choisissons la forme suivante pour l'intensité du processus :

$$\lambda_t^i = \psi_i \left(\sum_{j \in I} h_{j \to i} \left(\int_{[L_t^i, t[} g_j(t-s) dZ_s^j \right) \right), \tag{1.2.1}$$

où $\psi_i : \mathbb{R} \to \mathbb{R}_+$ est la *fonction de taux de saut*, $\{h_{j\to i} : \mathbb{R} \to \mathbb{R}, i, j \in I\}$ une famille de *fonctions de poids synaptiques* modélisant l'influence d'un neurone j sur un neurone i, $g_j : \mathbb{R}_+ \to \mathbb{R}_+$ une *fonction de perte* décroissante, et

$$L_t^i = \sup\{s < t : Z^i([s]) > 0\} \tag{1.2.2}$$

le dernier instant de saut du neurone i avant le temps t, avec la convention $[s] := [s, s]$.

Les fonctions $h_{j\to i}$ sont choisies linéaires ou linéaires tronquées pour rendre compte du fait que le potentiel de membrane d'un neurone accumule les stimulations provenant des autres neurones. La fonction de perte g_j modélise la perte de mémoire, c'est à dire le fait qu'une stimulation aura un fort impact sur le potentiel de membrane au moment de son apparition, mais cet impact sera atténué avec le temps. Notons qu'une dynamique similaire peut également être considérée en se plaçant du point de vue du potentiel de membrane, comme par exemple dans [Ces11] pour un modèle en temps discret. Nous considérons les deux modèles suivants.

1.2.2 Modèles avec seuils de saturation.

On suppose qu'à chaque arrête dirigée $j \to i$ du graphe d'interaction est associée un seuil de saturation $K_{j\to i} > 0$ représentant le nombre maximal de potentiels d'action que la synapse $j \to i$ est capable de prendre en compte. On suppose que

$$h_{j\to i}(x) = W_{j\to i}(x \wedge K_{j\to i}), \tag{1.2.3}$$

où $W_{j\to i} \in \mathbb{R}$ est appelé *poids synaptique* du neurone j sur le neurone i. Les neurones j tels que $W_{j\to i} > 0$ sont dits *excitateurs* pour i, si $W_{j\to i} < 0$, alors j est dit *inhibiteur* pour i. De plus, on suppose que $g_j \equiv 1$ pour tous $j \in I$. On peut ainsi réécrire l'expression de l'intensité de la manière suivante :

$$\lambda_t^i = \psi_i \left(\sum_{j \in I} W_{j\to i}(Z^j(]L_t^i, t[) \wedge K_{j\to i}) \right). \tag{1.2.4}$$

On fait l'hypothèse suivante :

Hypothèse 1. *Pour tous $i \in I, W_{i\to i} = 0$ et*

$$\sup_{i \in I} \sum_j |W_{j\to i}| K_{j\to i} < \infty. \tag{1.2.5}$$

Cette hypothèse est faite pour garantir la non-explosion en temps fini du processus. On obtient dans le chapitre 3 le résultat de stationnarité suivant pour lequel on a besoin d'une hypothèse de sommabilité des interactions plus forte que (1.2.5).

Théorème 1 (Théorème 6 du chapitre 2)**.** *Il existe une unique mesure de probabilité $\mathbb{P}$ telle que sous $\mathbb{P}$, le processus de comptage canonique $\left(Z^i, i \in I\right)$ est stationnaire.*

1.2.3 Modèles de trains de potentiels d'action en cascade.

On décrit maintenant le second modèle de *trains de potentiels d'action en cascade*. On suppose que pour chaque $j \in I$, la fonction $g_j : \mathbb{R}_+ \to \mathbb{R}_+$ est mesurable et décroissante telle que $\int_0^{+\infty} g_j(x)dx < +\infty$. La fonction g_j modélise un phénomène de perte d'information. On suppose de plus que

$$h_{j\to i}(x) = W_{j\to i} \cdot x, \tag{1.2.6}$$

pour une famille de *poids synaptiques* $\{W_{j\to i} \in \mathbb{R}\}$ vérifiant la condition de sommabilité suivante

$$\sup_{i\in I} \sum_j |W_{j\to i}| < \infty. \tag{1.2.7}$$

Enfin, on impose la condition suivante sur le graphe d'interactions.

Hypothèse 2. *L'ensemble I des neurones est divisé en couches $(I_n)_{n\in\mathbb{Z}}$ telles qu'on ait la partition $I = \sqcup_{n\in\mathbb{Z}} I_n$. Pour chaque $n \in \mathbb{Z}$ et pour chaque $i \in I_n$, on suppose que*

$$\{j \in I, j \neq i : W_{j\to i} \neq 0\} \subset I_{n-1}. \tag{1.2.8}$$

De ce fait, un neurone ne reçoit d'information que de neurones présents dans la couche juste au-dessus de la sienne. Cette hypothèse ne s'applique pas à la structure du cerveau, mais on peut l'observer dans des tissus nerveux plus simples comme la rétine. Le processus de Hawkes que l'on considère ici est alors définit par son intensité donnée par

$$\lambda_t^i = \psi_i \left(\sum_{j\in I} W_{j\to i} \int_{[L_t^i, t[} g_j(t-s) dZ_s^j \right). \tag{1.2.9}$$

L'hypothèse (1.2.7) est analogue à l'hypothèse (1.2.5) du modèle avec seuils de saturation de la partie précédente. Elle garantit la non-explosion en temps fini du processus, mais pour établir le résultat de stationnarité suivant, on a besoin d'une hypothèse de sommabilité plus forte.

Théorème 2 (Théorème 13 du chapitre 2)**.** *Il existe une unique mesure de probabilité $\mathbb{P}$ telle que sous $\mathbb{P}$, le processus de comptage canonique $\left(Z^i, i \in I\right)$ est stationnaire.*

1.2.4 Décomposition de type Kalikow en environnement aléatoire.

Les résultats énoncés plus haut sont obtenus à l'aide d'une décomposition de type Kalikow. En effet, dans [GL13], le modèle est en temps discret et les auteurs montrent qu'à un instant donné, la probabilité d'apparition d'un potentiel d'action, qui dépend à priori du passé entier de tout le système, peut se décomposer de la manière suivante. On tire au hasard un voisinage spatio-temporel selon une certaine loi de probabilité μ, puis on définit une probabilité d'apparition d'un potentiel d'action qui, conditionnellement à la réalisation de cette loi, ne dépend que de l'état du processus à l'intérieur de ce voisinage. Concrètement, si on note $p_{(i,t)}(1|x)$ la probabilité d'apparition d'un potentiel d'action pour le neurone i à l'instant t sachant que le passé de tout le processus est donné par la configuration spatio-temporelle x, une décomposition de type Kalikow est une décomposition convexe du type

$$p_{(i,t)}(1|x) = \sum_{k\geq 0} \mu_{(i,t)}(k) p_{(i,t)}^{[k]}(1|x), \tag{1.2.10}$$

où les $\mu_{(i,t)}(k)$ sont des probabilités sur $\mathbb{N}$ et $p^{[k]}_{(i,t)}(1|x)$ est une probabilité qui ne dépend de la configuration x que sur un voisinage spatio-temporel de taille finie indicée par k. Il est toujours possible d'obtenir une telle décomposition dès lors que $x \to p_{(i,t)}(1|x)$ est continue, voir par exemple [FMMN00] et [CD01].

Cette décomposition permet une construction graphique du processus à l'aide d'un algorithme de simulation parfaite. En effet, on peut associer à chaque site (i,t) un clan d'ancêtre aléatoire correspondant à un voisinage spatio-temporel. On obtient ainsi un processus de branchement. On se place sous des hypothèses qui nous assurent que celui-ci s'éteint presque sûrement. Ceci signifie que l'état de chaque site (i,t) ne dépend que d'un nombre fini d'ancêtres qu'on peut alors simuler en commençant par les sites qui ne dépendent pas de la configuration (ceux pour lesquels $k = 0$).

Dans notre cas, le modèle est en temps continu et on obtient également une décomposition de type Kalikow, mais celle-ci a lieu en environnement aléatoire. L'intensité est bornée et on travaille conditionnellement à la réalisation d'une mesure de Poisson aléatoire dominante dont l'intensité correspond à l'intensité maximale de notre processus. Ceci nous donne des instants de sauts potentiels pour notre processus qu'on accepte ou qu'on rejette à l'aide d'un algorithme de simulation parfaite. Cependant, du fait de cet environnement aléatoire, les probabilités de la loi μ dans la décomposition de type Kalikow ne sont pas des constantes déterministes mais des variables aléatoires mesurables par rapport à la mesure de Poisson dominante. Ceci pose un problème technique qui nous a conduit à proposer deux hypothèses différentes donnant lieu à deux modèles différents.

Dans le premier modèle, on suppose qu'il y a une activité spontanée. Cela se traduit par une minoration de l'intensité et nous permet d'obtenir une décomposition de type Kalikow uniquement selon l'espace et pour laquelle la loi μ est indépendante de la mesure de Poisson dominante. Dans un second modèle, on impose une contrainte sur la structure du réseau.

1.3 Estimation non-parametrique du taux d'émission de potentiels d'action.

L'étude décrite ici est un travail en collaboration avec Nathalie Krell et Eva Löcherbach. On s'intéresse non plus aux seuls instants d'émission de potentiels d'action, mais à l'évolution du potentiel de membrane.

Pour l'étude de l'évolution du potentiel de membrane, différents modèles sont proposés dans la littérature. Le modèle "intègre et tire" modélise celui-ci comme étant la solution d'une équation différentielle stochastique dont les paramètres donnent la dynamique du modèle. Voir par exemple [DL07], [LST95], [LBD16], [IT14], [DIRT15b] ou [DIRT15a]. Le modèle de Hodgkin-Huxley, introduit dans [HH52], propose un système d'équations différentielles faisant intervenir des paramètres modélisant les canaux ioniques pour les ions Potassium et les ions Sodium. Voir [Izh06] pour une introduction générale pour ce modèle. Le modèle de FitzHugh-Nagumo, introduit dans [Fit61] et étudié par exemple dans [BL12] est une simplification du modèle de Hodgkin-Huxley.

Ces modèles permettent de prendre en compte différentes considérations biologiques sur les mécanismes régissant l'évolution du potentiel de membrane d'un neurone, mais ne permettent pas de prendre en compte les interactions entre différents neurones. On propose, dans le chapitre 3, un modèle plus simple mais multidimensionnel afin d'intégrer ces interactions à notre modèle. On se base sur la dynamique des processus de Hawkes pour construire un modèle décrivant l'évolution du potentiel de membrane qui est un

processus de Markov déterministe par morceaux (PDMP). En effet, si l'on pose

$$X_t^i = \sum_j W_{j\to i} \int_{[L_t^i, t[} g_j(t-s) dZ_s^j,$$

on peut interpréter X_t^i comme le potentiel de membrane du neurone i au temps t. En effet, ce processus intègre dans sa dynamique les potentiels d'action émis par les autres neurones et on sait que les potentiels d'action émis par les neurones pré-synaptiques influent sur le potentiel de membrane. D'autre part un neurone possédant un potentiel de membrane élevé est davantage susceptible d'émettre un potentiel d'action. Il est donc naturel de faire le parallèle entre le potentiel de membrane et ce processus, qui intervient dans l'intensité du processus de Hawkes aux travers d'une fonction de taux de saut ψ_i croissante. Sous certaines hypothèses sur g_j, ce processus est Markovien.

Ceci nous conduit à l'étude d'un processus qui suit une évolution déterministe ponctuée par des sauts intervenants à des instants aléatoires. Ce processus est à valeurs dans $[0, K]^N$, où N est le nombre de neurones dans le réseau et K est le potentiel de membrane maximal. Un saut du processus correspond à l'émission d'un potentiel d'action par un neurone. Lors d'un saut, le potentiel de membrane du neurone émetteur retombe à 0 et ceux des autres neurones sont augmentés de $\frac{1}{N}$. C'est un modèle plus simple que le précédent : on considère que chaque neurone stimule chacun des autres neurones et que ces stimulations sont toutes de même intensité, inversement proportionnelle au nombre de neurones. Le fait que le potentiel de membrane du neurone émetteur retombe à zero permet de modéliser le phénomène de perte de mémoire.

Ceci nous amène à étudier une classe de processus de Markov déterministes par morceaux (PDMP) particulière. Les processus de Markov déterministes par morceaux ont été introduits par Davis ([Dav84] et [Dav93]). Ils forment une famille de processus de Markov càdlàg suivant une évolution déterministe perturbée par des sauts aléatoires. Ils sont largement utilisés pour modéliser par exemple des phénomènes biologiques (voir par exemple [CDMR12], [Kre14] ou[PTW10]) et interviennent également dans le domaine de l'informatique avec le processus TCP (voir [BCG$^+$13]). Pour un aperçu global sur les PDMP, [ABG$^+$14] est une excellente référence. Notre processus $X_t = (X_t^1, ..., X_t^N)$, où chaque coordonnée représente le potentiel de membrane d'un neurone en particulier, possède la dynamique suivante. La dérive déterministe du PDMP attire le potentiel de membrane de chaque neurone vers un potentiel d'équilibre m avec une vitesse exponentielle de paramètre λ. Ceci permet de modéliser la synchronisation des potentiels de membranes due à la présence des "gap-junctions". De plus, un neurone avec un potentiel de membrane x "tire" (i.e. saute) avec une intensité $f(x)$. Lorsqu'un neurone tire (ceci correspond à l'emission d'un potentiel d'action), son potentiel de membrane retourne à 0, et celui de chaque autre neurone du réseau est augmenté de $\frac{1}{N}$. Plus concrètement, le générateur de notre processus X est donné pour toute fonction de test régulière $\varphi : \mathbb{R}_+^N \to \mathbb{R}$ par

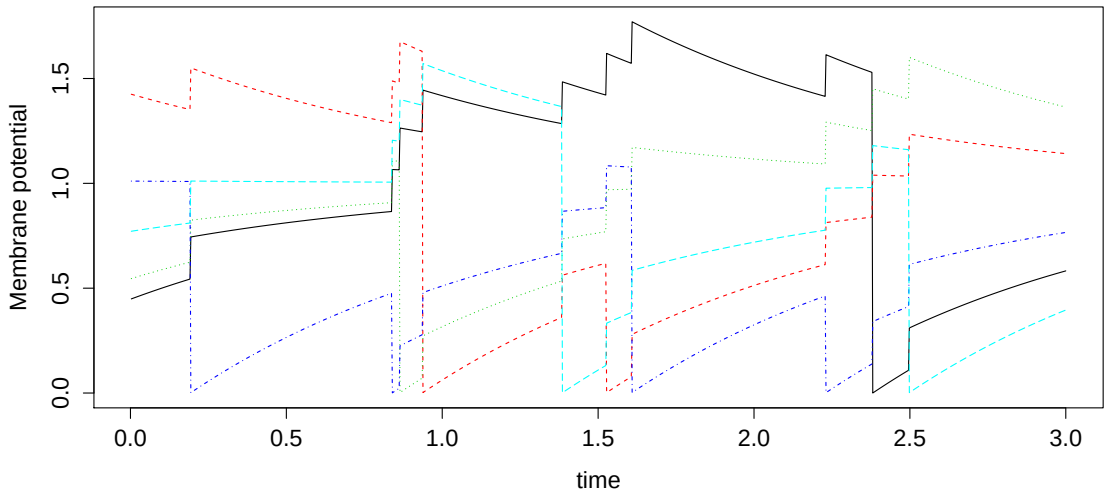
$$L\varphi(x) = \sum_{i=1}^N f(x_i) \left[\varphi(\Delta_i(x)) - \varphi(x)\right] - \lambda \sum_i \left(\frac{\partial \varphi}{\partial x_i}(x) \left[x_i - m\right] \right), \tag{1.3.11}$$

où

$$(\Delta_i(x))_j = \left\{ \begin{array}{ll} x_j + \frac{1}{N} & j \neq i \\ 0 & j = i \end{array} \right\}, j = 1, ..., N. \tag{1.3.12}$$

Fig. 1 est un exemple de trajectoires pour 5 neurones, avec les paramètres suivants : $f = Id,\ \lambda = 1,\ m = 1$, et $K = 2$.

fig. 1



Notre but est d'estimer la fonction de taux de saut f à l'aide d'un estimateur à noyau. Pour ce faire on introduit les notations suivantes.

Premièrement, on introduit les temps de saut pour le neurone i. On pose

$$T_0^i = 0, T_n^i = \inf\{t > T_{n-1}^i : X_{t-}^i > 0, X_t^i = 0\}, n \geq 1,$$

et on introduit ensuite les mesures de saut

$$\mu^i(ds, dy) = \sum_{n\geq 1} 1_{\{T_n^i < \infty\}} \delta_{(T_n^i, X^i_{T_n^i -})}(dt, dy), \quad \mu(dt, dx) = \sum_{i=1}^N \mu^i(ds, dx).$$

On pose de plus pour chaque $B \in \mathcal{B}([0, K])$,

$$\mu_t(B) = \mu([0, t], B).$$

La mesure μ^i est alors compensée par $\hat{\mu}^i(ds, dy) = f(X_s^i) ds \delta_{X_s^i}(dy)$, et le compensateur $\hat{\mu}_t$ de μ_t est donné par

$$\hat{\mu}_t(dy) = \int_0^t f(y)\eta(ds, dy), \text{ où } \eta(A \times B) = \int_A \left(\sum_{i=1}^N 1_B(X_s^i) \right) ds$$

est la mesure de temps d'occupation du processus X. On désigne par η_t la mesure $\eta([0, t], \cdot)$. Pour un noyau Q tel que

$$Q \in C_c(\mathbb{R}), \int_{\mathbb{R}} Q(y)dy = 1, \tag{1.3.13}$$

on définit un estimateur à noyau pour la fonction f évaluée au point a avec une fenêtre h, basé sur l'observation du processus X sur $[0, t]$, par

$$\hat{f}_{t,h}(a) = \frac{\int_0^t \int_{\mathbb{R}} Q_h(y - a)\mu(ds, dy)}{\int_0^t \int_{\mathbb{R}} Q_h(y - a)\eta(ds, dy)}, \text{ où } Q_h(y) := \frac{1}{h} Q\left(\frac{y}{h}\right) \text{ et } \frac{0}{0} := 0. \tag{1.3.14}$$

Cet estimateur est bien définit sur des évènements du type $A_{t,r} := \left\{ \frac{1}{Nt} \int_0^t \int_{\mathbb{R}} Q_h(y-a)\eta(ds,dy) \geq r \right\}$, ce qui correspond au fait d'avoir un minimum d'observations autour de a. D'autre part, comme il s'agit d'un estimateur à noyau, on a besoin de stationnarité et de propriétés de régularité pour la mesure invariante. Cependant, l'attraction du potentiel de membrane vers un potentiel d'équilibre m ainsi que le retour à 0 du potentiel de membrane après l'émission d'un potentiel d'action provoque un manque de régularité autour de ces positions 0 et m. On se place donc sur un ensemble $S_{d,k} \subset [0,K]$ qui contient des valeurs éloignées des bornes et du potentiel d'équilibre m.

Le fait que le compensateur de la mesure de saut $\mu_t(dy)$ soit donné par $\int_0^t f(y)\eta(ds,dy)$ fait apparaître une martingale. Ceci nous permet d'obtenir la convergence de notre estimateur à l'aide de théorèmes limites pour les martingales. On suppose que la fonction de taux de saut f appartient à la classe de fonctions Hölderiennes $H(\beta,F,L,f_{min})$ suivante d'ordre de régularité $\beta = k+\alpha$.

$$H(\beta,F,L,f_{min}) = \{f \in C^k(\mathbb{R}_+) : |\frac{d^l}{dx^l} f(x)| \leq F, \text{ for all } 0 \leq l \leq k, x \in [0,K],$$
$$f(x) \geq f_{min}(x) \text{ for all } x \in [0,K], \ |f^{(k)}(x) - f^{(k)}(y)| \leq L|x-y|^\alpha \text{ for all } x,y \in [0,K]\}. \tag{1.3.15}$$

On adopte les notations P_x^f et E_x^f afin de préciser la dépendance à la fonction de taux de saut inconnue f et au point de départ $x \in [0,K]^N$ du processus X.

On obtient alors les propriétés suivantes pour notre estimateur.

Théorème 3 (Théorèmes 32 et 33 du chapitre 2)**.** *Soit $f \in H(\beta,F,L,f_{min})$ et $Q \in C_c(\mathbb{R})$ tel que $\int_{\mathbb{R}} Q(y)y^j dy = 0$ pour tout $1 \leq j \leq k$, et $\int_{\mathbb{R}} |y|^\beta Q(y)dy < \infty$. Alors il existe $r^* > 0$ tel que pour tout $a \in S_{d,k}$, on ait*

(i) Pour l'estimateur à noyau (1.3.14) *de fenêtre $h_t = t^{-\frac{1}{2\beta+1}}$, pour tout $x \in [0,K]$,*

$$\limsup_{t\to\infty} \sup_{f\in H(\beta,F,L,f_{min})} t^{\frac{2\beta}{2\beta+1}} E_x^f \left[|\hat{f}_{t,h_t}(a) - f(a)|^2 | A_{t,r^*} \right] < \infty.$$

(ii) De plus, pour $h_t = o(t^{-1/(1+2\beta)})$, pour tout $f \in H(\beta,F,L,f_{min})$ et $a \in S_{d,k}$

$$\sqrt{th_t} \left(\hat{f}_{t,h_t}(a) - f(a) \right) \to \mathcal{N}(0, \Sigma(a))$$

faiblement sous P_x^f, où $\Sigma(a) = \frac{f(a)}{N\pi_1(a)} \int Q^2(y)dy$.

(iii) Pour tout $a \in S_{d,k}$ et $x \in [0,K]$, on a

$$\liminf_{t\to\infty} \inf_{\hat{f}_t} \sup_{f\in H(\beta,F,L,f_{min})} t^{\frac{2\beta}{1+2\beta}} E_x^f [|\hat{f}_t(a) - f(a)|^2] > 0, \tag{1.3.16}$$

où l'infimum est pris sur l'ensemble des estimateurs possibles $\hat{f}_t(a)$ pour $f(a)$.

On présente également deux résultats probabilistes dont on a besoin pour l'estimation. Le premier est le caractère récurrent positif au sens de Harris de notre PDMP.

Théorème 4 (Théorème 30 du chapitre 3). *Le processus X est récurrent positif au sens de Harris avec une unique mesure invariante π, i.e. pour tout $B \in \mathcal{B}([0,K]^N)$,*

$$\pi(B) > 0 \ \ implies\ P_x\left(\int_0^\infty 1_B(X_s)ds = \infty\right) = 1 \tag{1.3.17}$$

pour tout $x \in [0,K]^N$. De plus, il existe des constantes $C > 0$ et $\kappa > 1$ dépendant uniquement de la classe $H(\beta, F, L, f_{min})$), telle que

$$\sup_{f\in H(\beta,F,L,f_{min})} \|P_t(x,\cdot) - \pi\|_{TV} \leq C\kappa^{-t}. \tag{1.3.18}$$

Ce résultat est obtenu à l'aide d'une méthode de couplage empruntée à [DO14]. Notons que la constante C ne dépend pas de la position de départ x puisque l'espace d'état $[0,K]^N$ est compact.

Le deuxième est l'existence d'une fonction de densité régulière pour la mesure invariante.

Théorème 5 (Théorème 5 de [Löc16]). *On pose*

$$\pi_1 := \mathcal{L}_\pi(X_t^1),$$

i.e. $\int g d\pi_1 = E_\pi(g(X_t^1))$. Alors π_1 possède une densité π^1 bornée et continue par rapport à la mesure de Lebesgue sur $S_{d,k}$ De plus, $\pi^1 \in C^k(S_{d,k})$ et

$$\sup_{\ell\leq k, w\in S_{d,k}} |\pi_1^{(\ell)}(w)| + \sup_{w\neq w', w,w'\in S_{d,k}} \frac{\pi_1^{(k)}(w) - \pi_1^{(k)}(w')}{|w-w'|^\alpha} \leq C_F, \tag{1.3.19}$$

où la constante C_F dépend uniquement de la classe $H(\beta, F, L, f_{min})$.

Afin d'obtenir la vitesse de convergence pour notre estimateur, on commence par établir la propriété suivante.

Il existe une constante $r^* > 0$ telle que

$$\liminf_{t\to\infty} \inf_{f\in H(\beta,F,L,f_{min})} P_x^f\left(\frac{1}{Nt}\int_0^t\int_{\mathbb{R}} Q_h(y-a)\eta(ds,dy) \geq r^*\right) = 1. \tag{1.3.20}$$

Ceci nous assure qu'on a suffisamment d'observations autour de la position a et nous permet d'obtenir un contrôle de l'erreur d'estimation dans L^2 dès lors qu'on a un contrôle de la quantité suivante :

$$\left(\frac{1}{Nt}\int_{\mathbb{R}} Q_h(y-a)\eta_t(dy)\right)\left(\hat{f}_{t,h}(a) - f(a)\right).$$

Pour obtenir ce contrôle, on effectue la décomposition suivante :

$$\begin{aligned}
&\left(\frac{1}{Nt}\int_{\mathbb{R}} \frac{1}{h}Q\left(\frac{y-a}{h}\right)\eta_t(dy)\right)\left(\hat{f}_{t,h}(a) - f(a)\right) = \\
&\frac{1}{Nt}\int_{\mathbb{R}} Q_h(y-a)\mu_t(dy) - \frac{1}{Nt}\int_{\mathbb{R}} Q_h(y-a)f(y)\eta_t(dy) && (1.3.21)\\
&+ \frac{1}{Nt}\int_{\mathbb{R}} Q_h(y-a)\Big(f(y)-f(a)\Big)\eta_t(dy) - \int_{\mathbb{R}} Q_h(y-a)\Big(f(y)-f(a)\Big)\pi_1(dy) && (1.3.22)\\
&+ \int_{\mathbb{R}} Q_h(y-a)\Big(f(y)-f(a)\Big)\pi_1(dy). && (1.3.23)
\end{aligned}$$

Le terme (1.3.21) est une martingale car le compensateur de la mesure $\mu_t(dy)$ est exactement $f(y)\eta_t(dy)$. On utilise alors un théorème limite pour les martingales afin d'obtenir un contrôle.

Le contrôle du terme (1.3.22) est obtenu grâce à l'ergodicité du processus, car on a d'un côté la mesure d'occupation et de l'autre la mesure invariante. Pour appliquer le résultat d'ergodicité, on utilise la propriété de Markov. L'espérance conditionnelle par rapport au passé avant le temps s s'écrit donc comme l'espérance sur une probabilité ayant comme condition initiale le processus X_s. C'est pour ne plus avoir de dépendance en X_s que l'hypothèse d'espace d'état compact est nécessaire car elle nous permet d'obtenir une ergodicité uniforme sur l'espace d'état.

Enfin, pour le dernier terme (1.3.23), on utilise les propriétés du noyau Q et le fait que la densité de π_1 ainsi que la fonction f sont régulières.

On présente maintenant quelques résultats obtenus sur des simulations, pour différents taux de saut f. Les autres paramètres sont fixes : $N = 100,\ \lambda = 1,\ K = 2$ et $m = 1$. La dynamique du système est la même lorsque λ et f ont le même ratio. Autrement dit, des variations de λ et f conservant le même ratio entre les deux paramètres donnent la même loi pour le processus rééchelonné en temps. C'est pourquoi on fixe $\lambda = 1$ et on propose différents choix pour f. Le noyau Q utilisé ici est un noyau Gaussien tronqué de variance 1.

La fenêtre optimale $h_t = t^{-\frac{1}{2\beta+1}}$ depend de la régularité de f donnée par le paramètre β. C'est pourquoi on propose une fenêtre adaptative choisie grâce à une méthode de validation croisée.

On représente pour chaque choix de fonction de taux de saut f l'estimateur $\hat{f}$ et la distribution observée de X. Les figures 2, 3 et 4 correspondent respectivement aux définitions suivantes de f : $f(x) = x,\ f(x) = \log(x+1)$ et $f(x) = \exp(x) - 1$.

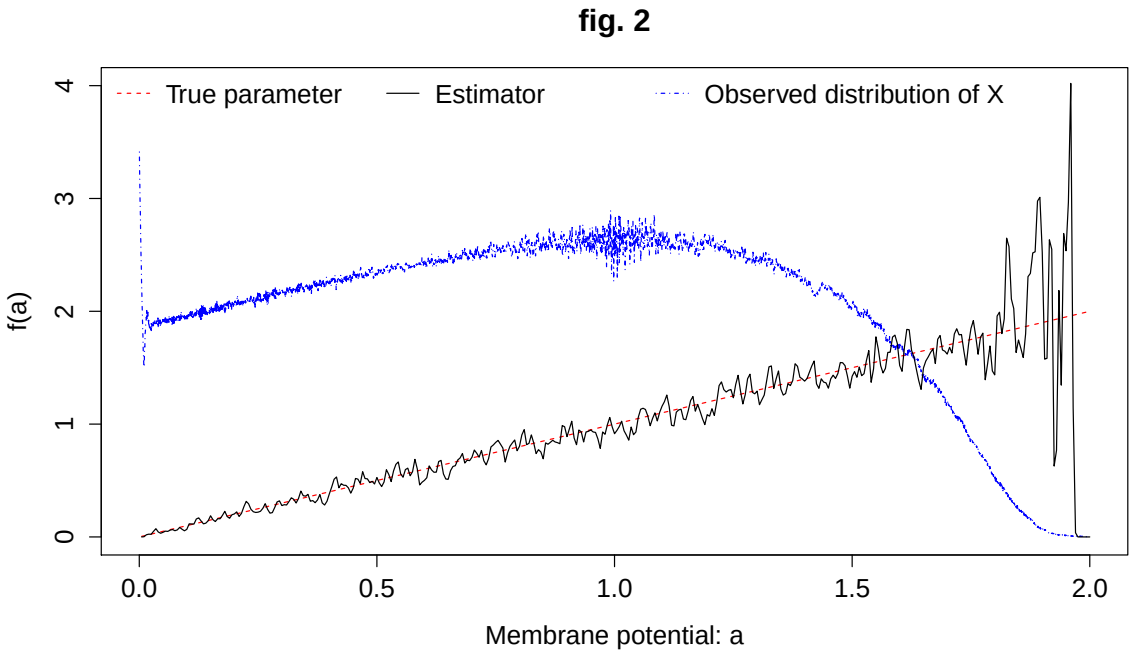


FIGURE 1.1 – Estimation of the intensity function $f(x) = x$

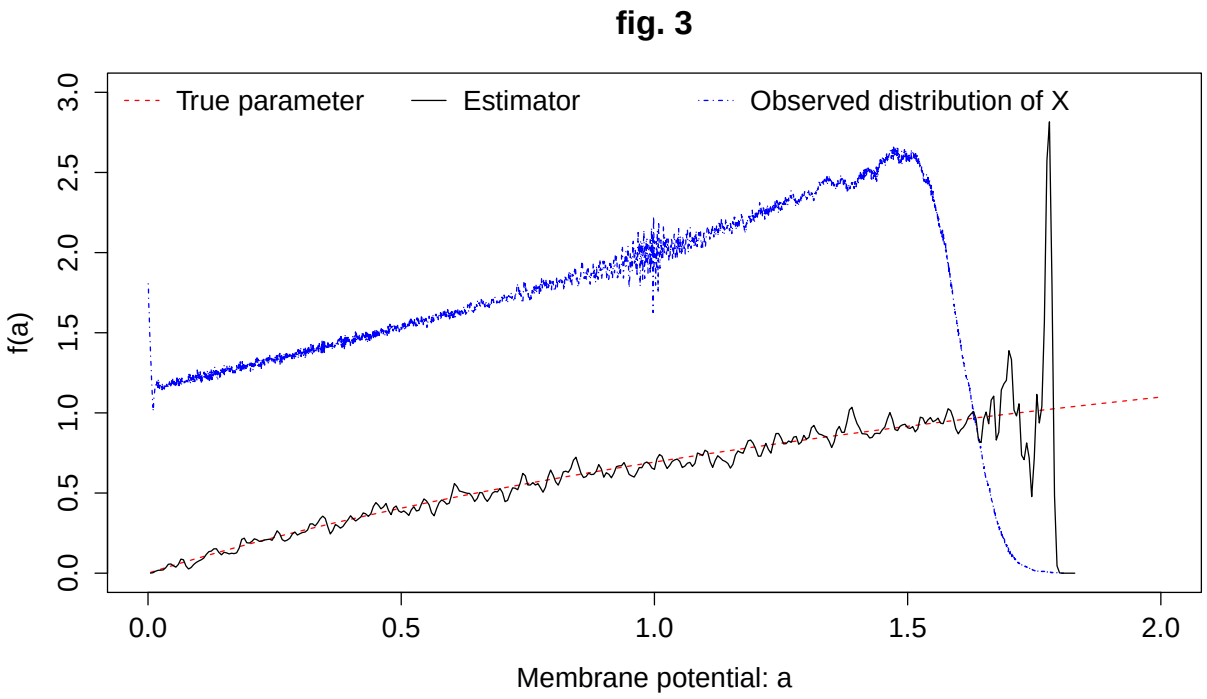


FIGURE 1.2 – Estimation of the intensity function $f(x) = \log(x + 1)$

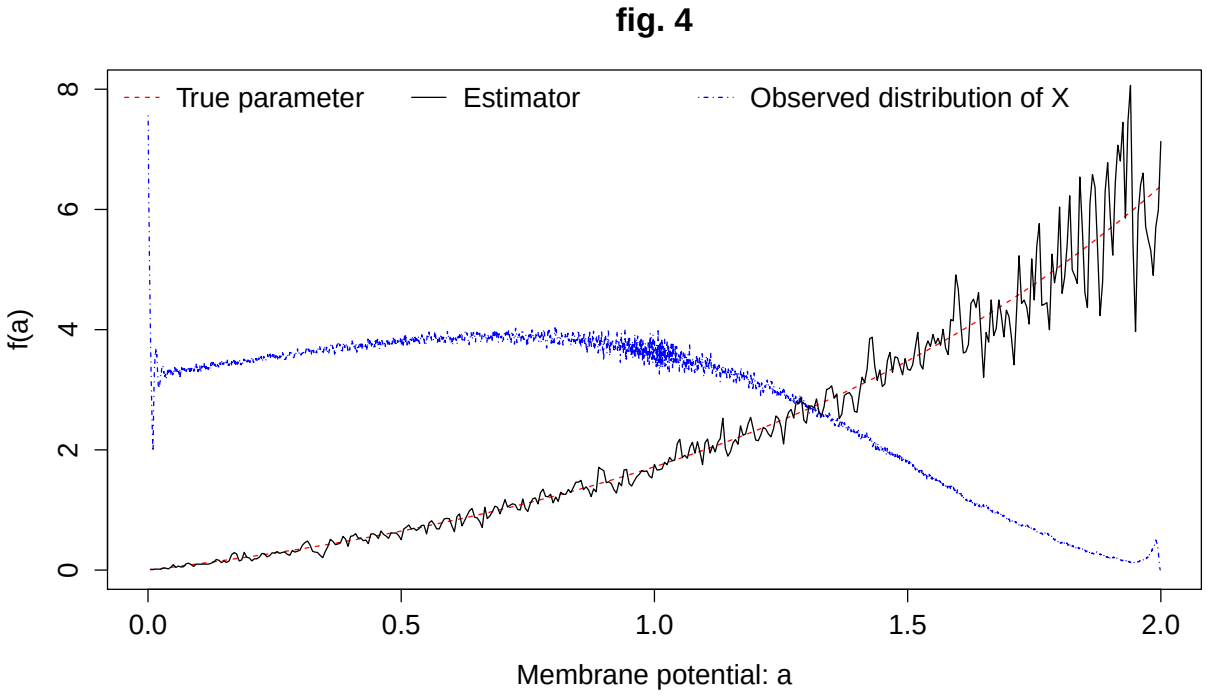


FIGURE 1.3 – Estimation of the intensity function $f(x) = \exp(x) - 1$

Pour conclure, ce modèle de PDMP peut être mis en relation avec le modèle utilisant des processus de Hawkes décrit précédemment. Le retour à 0 d'un neurone émettant un potentiel d'action modélise la perte de mémoire. Ainsi le niveau du potentiel de membrane à chaque instant ne dépend que des signaux émis par les autres neurones depuis le dernier instant où lui même a émis un potentiel d'action. Les poids synaptiques notés $W_{j\to i}$ pour le processus de Hawkes sont tous égaux à $\frac{1}{N}$ dans le cas du PDMP. Les fonctions d'intensité ψ_i pour le processus de Hawkes correspondent à la fonction f qui dans le cas du PDMP est la même pour tous les neurones. Cependant, ces simplifications sont uniquement faites dans le but d'alléger les calculs, mais une généralisation de notre modèle de PDMP peut être faite avec des poids synaptiques et des fonctions d'intensité différentiées. Dans les deux cas, les stimulations reçues par les autres neurones sont atténuées avec le temps par une fonction de perte, même si celle-ci diffère. Au lieu d'un phénomène de seuil de saturation ou d'une fonction décroissante et intégrable pour le processus de Hawkes, on a, dans le cas du PDMP, une attraction vers un potentiel d'équilibre m.

Le chapitre suivant a fait l'objet d'une soumission et est accepté pour parution dans *Advances in Applied Probability.*

Chapter 2

Hawkes processes with variable length memory and an infinite number of components

abstract In this paper, we propose a model for biological neural nets where the activity of the network is described by Hawkes processes having a variable length memory. The particularity of this paper is to deal with an infinite number of components. We propose a graphical construction of the process and build, by means of a perfect simulation algorithm, a stationary version of the process. To carry out this algorithm, we make use of a Kalikow-type decomposition technique.

Two models are described in this paper. In the first model, we associate to each edge of the interaction graph a saturation threshold that controls the influence of a neuron on another. In the second model, we impose a structure on the interaction graph leading to a cascade of spike trains. Such structures, where neurons are divided into layers, can be found in retina.

2.1 Introduction and main results

2.1.1 Motivation.

The aim of this paper is to give a model in continuous time for an infinite system of interacting neurons. Each neuron is represented by its spike train, i.e. the series of events of spiking over time. Since neural activity is continuously recorded in time, a time continuous description is natural. The model considered in this paper is an extension to the continuous time framework of a model which has been recently introduced by Galves and Löcherbach [GL13] in discrete time.

We consider a countable set of neurons I. The activity of each neuron $i \in I$ is described by a counting process Z^i where for any $-\infty < s < t < \infty$, $Z^i(]s,t])$ records the number of spikes of neuron i during the interval $]s,t]$. Under suitable assumptions, the sequence of counting processes $(Z^i, i \in I)$ is characterized by its intensity process $(\lambda_t^i, i \in I)$ which is defined through the relation

$$\mathbb{P}(Z^i \text{ has a jump in }]\mathrm{t} \ , \ \mathrm{t} + \mathrm{dt} \ ]|\mathcal{F}_t) = \lambda_t^i dt, i \in I.$$

Here, $\mathcal{F}_t$ is the sigma-field generated by $Z^i(]s,u]), s \leq u \leq t, i \in I$.

Our main motivation is to model neural nets. This naturally leads to the following choice of intensity processes:

$$\lambda_t^i = \psi_i \left(\sum_{j \in I} h_{j \to i} \left(\int_{[L_t^i, t[} g_j(t-s) dZ_s^j \right) \right), \tag{2.1.1}$$

where $\psi_i : \mathbb{R} \to \mathbb{R}_+$ is the *spiking rate function*, $\{h_{j\to i} : \mathbb{R}_+ \to \mathbb{R}, i, j \in I\}$ a family of *synaptic weight functions* modeling the influence of neuron j on neuron i, $g_j : \mathbb{R}_+ \to \mathbb{R}_+$ a non-increasing *decay function*, and

$$L_t^i = \sup\{s < t : Z^i([s]) > 0\} \tag{2.1.2}$$

the last spiking time before time t of neuron i (with the convention $[s] := [s, s]$).

The form (2.1.1) of our intensity process is close to the typical form of the intensity of a multivariate nonlinear Hawkes process. The original papers of Hawkes [Haw72] and Hawkes and Oakes [HO74], introducing the model, deal with linear intensity functions. Extensions to the nonlinear case have been considered by Brémaud and Massoulié [BM96], see also Massoulié [Mas98], who propose a study of the stability properties of multivariate nonlinear Hawkes processes. Hawkes processes have shown to be important in various fields of applications. To cite just a few, Hansen, Reynaud-Bouret and Rivoirard [HRBR15] and Chevallier [Che15] are excellent references proving the use of Hawkes processes as models of spike trains in neuroscience. Reynaud-Bouret and Schbath [RBS10] deal with an application to genome analysis. In a completely different context, Jaisson and Rosenbaum [JR15] obtain limit theorems for nearly unstable Hawkes processes in view of applications in financial price modeling. For a general introduction to Hawkes processes and their basic properties we refer the reader to Daley and Vere-Jones [DVJ03].

Our form of the intensity (2.1.1) differs from the classical Hawkes setting by its *variable memory structure* introduced through the term L_t^i. Hence the spiking intensity of a neuron only depends on its history up to its last spike time which is a biologically very plausible assumption on the memory structure of the process. Therefore, our model can be seen as a nonlinear multivariate Hawkes process where the number of components is infinite with a variable memory structure. The interactions between neurons are encoded through the synaptic weight functions $h_{j\to i}$ that we are going to specify below.

2.1.2 The setting.

We work on a filtered measurable space $(\Omega, \mathcal{A}, \mathbb{F})$ which we define as follows. We write $\mathbb{M}$ for the canonical path space of simple point processes given by

$$\mathbb{M} := \{\mathtt{m} = (t_n)_{n\in\mathbb{Z}} : t_0 \le 0 < t_1, t_n \le t_{n+1}, t_n < t_{n+1} \text{ if } t_n < +\infty \text{ or } t_{n+1} > -\infty; \lim_{n\to+\infty} t_n = +\infty, \lim_{n\to-\infty} t_n = -\infty\}.$$

For any $\mathtt{m} \in \mathbb{M}$, any $n \in \mathbb{Z}$, let $T_n(\mathtt{m}) = t_n$. We identify $\mathtt{m} \in \mathbb{M}$ with the associated point measure $\mu = \sum_n \delta_{T_n(\mathtt{m})}$ and put $\mathcal{M}_t := \sigma\{\mu(A) : A \in \mathcal{B}(\mathbb{R}), A \subset]-\infty, t]\}$, $\mathcal{M} = \mathcal{M}_\infty$. We will also systematically identify $\mathtt{m}$ with the associated counting process $\alpha(\mathtt{m})$, defined by $\alpha_0(\mathtt{m}) = 0$,

$$\alpha_t(\mathtt{m}) = \mu(]0, t]) \text{ if } t \ge 0, \alpha_t(\mathtt{m}) = -\mu(]t, 0]), \text{ if } t \le 0.$$

Finally we put $(\Omega, \mathcal{A}, \mathbb{F}) := (\mathbb{M}, \mathcal{M}, (\mathcal{M}_t)_{t\in\mathbb{R}})^I$. We write $(Z^i, i \in I)$ for the canonical multivariate point measure defined on Ω.

We specify the following parameters: a family of firing rate functions $\psi = \{\psi_i : \mathbb{R} \to \mathbb{R}_+, i \in I\}$, a family of synaptic weight functions $h = \{h_{j\to i} : \mathbb{R} \to \mathbb{R}, i, j \in I\}$, a family of functions $g = \{g_j : \mathbb{R}_+ \to \mathbb{R}_+, j \in I\}$, which are non-increasing. Recall the definition of the last spiking time of neuron i before time t, given in (2.1.2).

Definition 3. *A Hawkes process with variable length memory and an infinite number of interacting components with parameters (ψ, h, g) is a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A}, \mathbb{F})$ such that*

1. *$\mathbb{P}-$almost surely, for all $i \neq j$, Z^i and Z^j never jump simultaneously,*
2. *for all $i \in I$, the compensator of Z^i is given by $\nu^i(dt) = \lambda^i_t dt$, where*

$$\lambda^i_t = \psi_i \left(\sum_{j\in I} h_{j\to i} \left(\int_{[L^i_t, t[} g_j(t-s) dZ^j_s \right) \right).$$

2.1.3 Main results.

In the case where I is a finite set, under suitable assumptions on the parameters of the process, the existence and construction of $(Z^i, i \in I)$ is standard (see [BM96] and [DFH16]). In our case, however, the number of interacting components defining the process is infinite. In such a framework, Delattre, Fournier and Hoffmann [DFH16] prove pathwise existence and uniqueness of the processes, however without giving an explicit construction of the process. In the present paper, we show that – under suitable assumptions – a *graphical construction* is possible. This graphical construction does not only imply the existence but also the possibility of a perfect simulation of a stationary version of the process (i.e. a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A}, \mathbb{F})$, such that under $\mathbb{P}$, for all $u \in \mathbb{R}$ and all $i \in I$, the processes $Z^i(]u, u + t]), t \in \mathbb{R}$ are stationary). These results are achieved via a *Kalikow-type decomposition* in two types of models. The first model is a system containing a *saturation threshold* for any directed edge $i \to j$ in the interaction graph defined by the synaptic weights. The second model deals with a *cascade of spike trains*.

Kalikow-type decompositions are now largely used in the literature for perfect simulation issues and similar scopes. They have been considered first by Ferrari, Maass, Martínez and Ney [FMMN00] and in Comets, Fernández and Ferrari [CD01]. This type of technique was then studied in a series of papers for perfect simulation issues. See Galves and Löcherbach [GL13] for an application in the context of neural biological nets in discrete time. The decomposition that we use in the present paper is a non-trivial extension of the previous considerations to the framework of continuous time neural nets. In the case of our first model, it has to be achieved in a random environment.

To the best of our knowledge, the perfect simulation algorithm constructed in the present paper is the first result in this direction obtained for Hawkes processes with nonlinear intensity functions. The well known work of Møller and Rasmussen [MR05] on perfect simulation of Hawkes processes deals with linear intensity functions and exploits very heavily the underling branching structure. The precise form of our perfect simulation algorithm is given in Section 5.2.

2.1.4 Assumptions and notations.

Throughout this paper we suppose that the firing rate functions $\psi_i : \mathbb{R} \to \mathbb{R}_+$ are non-decreasing and bounded by a real number Λ_i. Introducing

$$\phi_i = \frac{\psi_i}{\Lambda_i}, \tag{2.1.3}$$

we assume

Assumption 4. *The family of functions $\{\phi_i\}_{i\in I}$, is equicontinuous, i.e. there exists a modulus of continuity $\omega : \mathbb{R}_+ \to \mathbb{R}_+$, satisfying $\omega(0) = 0$ and $\lim_{x\to 0} \omega(x) = 0$, such that for all $x, x' \in \mathbb{R}$, $i \in I$,*

$$|\phi_i(x) - \phi_i(x')| \leq \omega\left(|x - x'|\right). \tag{2.1.4}$$

Notice that we can always choose the function ω to be sub-additive.

The interactions between neurons are coded via the synaptic weight functions $h_{j\to i}$. For each neuron i, we define

$$\mathcal{V}_{i\to\cdot} = \{j \in I, j \neq i : h_{i\to j} \neq 0\},$$

where 0 denotes the constant function 0. As a consequence, $\mathcal{V}_{i\to\cdot}$ is the set of all neurons that are directly influenced by neuron i. In the same spirit, we put

$$\mathcal{V}_{\cdot\to i} = \{j \in I, j \neq i : h_{j\to i} \neq 0\}, \tag{2.1.5}$$

which is the set of neurons that have a direct influence on i. These sets may be finite or infinite.

In the following we introduce the two main types of models that we consider, firstly *models with saturation thresholds* and secondly *models with a cascade of spike trains.*

2.2 Models with saturation threshold.

2.2.1 Models with saturation thresholds.

We suppose that to each directed edge $j \to i$ is associated a saturation threshold $K_{j\to i} > 0$ representing the maximal number of spikes that the synapse $j \to i$ is able to support. We suppose that

$$h_{j\to i}(x) = W_{j\to i}(x \wedge K_{j\to i}), \tag{2.2.6}$$

where $W_{j\to i} \in \mathbb{R}$ is called the *synaptic weight* of neuron j on i. Moreover we suppose that $g_j \equiv 1$ for all $j \in I$ and write for short $g = \mathbf{1}$. Hence we can rewrite

$$\lambda_t^i = \psi_i\left(\sum_{j\in I} W_{j\to i}(Z^j(]L_t^i, t[) \wedge K_{j\to i})\right). \tag{2.2.7}$$

Introduce for any $i \in I$ a non-decreasing sequence $(V_i(k))_{k\geq 0}$ of finite subsets of I such that $V_i(0) = \emptyset$, $V_i(1) = \{i\}$, $V_i(k) \subset V_i(k+1)$, $V_i(k) \neq V_i(k+1)$ if $V_i(k) \neq \mathcal{V}_{\cdot\to i} \cup \{i\}$ and $\bigcup_k V_i(k) = \mathcal{V}_{\cdot\to i} \cup \{i\}$ (recall (2.1.5)).

We define, for all $k \geq 0$, $\partial V_i(k) := V_i(k+1) \setminus V_i(k)$, the border of the set $V_i(k)$ and make the following assumption.

Assumption 5. *For all $i \in I, W_{i\to i} = 0$ and*

$$\sup_{i\in I} \sum_j |W_{j\to i}| K_{j\to i} < +\infty. \tag{2.2.8}$$

The following theorem states that if any neuron has a sufficiently high spontaneous firing activity, then a unique stationary version of the Hawkes process with saturation threshold exists.

Theorem 6. *Grant Assumptions 4 and 5 and suppose that for all $i \in I$,*

$$\psi_i \geq \delta_i, \tag{2.2.9}$$

for some $\delta_i > 0$. Suppose moreover that

$$\sup_{i\in I} \left(\sum_{k\geq 1} \left[\left(\sum_{j\in V_i(k)} \frac{\Lambda_j - \delta_j}{\delta_i} \right) \bar{\mu}_i(k) \right] \right) < 1, \tag{2.2.10}$$

where

$$\bar{\mu}_i(k) := \omega \left(\sum_{j\in \partial V_i(k-1)} |W_{j\to i}| K_{j\to i} 1_{W_{j\to i}<0} \right) + \omega \left(\sum_{j\in \partial V_i(k-1)} |W_{j\to i}| K_{j\to i} 1_{W_{j\to i}>0} \right). \tag{2.2.11}$$

Then there exists a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$ such that under $\mathbb{P}$, the canonical process $\left(Z^i, i \in I\right)$ on Ω is a stationary nonlinear Hawkes process with variable length memory and an infinite number of interacting components, with parameters (ψ, h, g) where h is given by (2.2.6) and $g = \mathbf{1}$.

Remark 7. *(i) By sub-additivity of ω, the following condition*

$$\sup_{i\in I} \left(\sum_{k\geq 1} \left[\left(\sum_{j\in V_i(k)} \frac{\Lambda_j - \delta_j}{\delta_i} \right) \left(\sum_{j\in \partial V_i(k-1)} \omega(|W_{j\to i}| K_{j\to i}) \right) \right] \right) < 1 \tag{2.2.12}$$

implies condition (2.2.10).

(ii) The following additional assumptions on the parameters allows to work with a condition simpler than (2.2.10). *Suppose that $\delta_i = \delta > 0$ for all $i \in I$ and that a stronger summability than (2.2.8) holds:*

$$\sup_i \Lambda_i \sum_{k\geq 1} |V_i(k)| \sum_{j\in \partial V_i(k-1)} \omega(|W_{j\to i}| K_{j\to i}) < \infty. \tag{2.2.13}$$

By subadditivity of ω, the above condition (2.2.13) *implies* (2.2.10) *for δ sufficiently large.*

(iii) In the above model, there is a saturation threshold for each directed edge between two neurons. A different model would be a system where the saturation threshold concerns only the global input received by each neuron. This would amount to consider the following intensity

$$\lambda_t^i = \psi_i \left(K_i^- \vee \left(\left(\sum_{j\in I} W_{j\to i} Z^j \left( ]L_t^i, t[\right) \right) \wedge K_i^+ \right) \right),$$

where K_i^- and K_i^+ are global saturation thresholds respectively for inhibition and stimulation. Under obvious changes of condition (2.2.10), Theorem 6 remains also true in this framework.

(iv) In [Mas98], the author obtains similar results in a slightly different context dealing with truly infinite memory models. The fact that we consider variable length memory Hawkes processes constitutes of course the main difference with respect to [Mas98]. As a matter of fact, due to this difference, instead of (2.2.9) and (2.2.10), [Mas98] requires that $\sum_{i\in I}\sum_{j\in I} W_{j\to i}K_{j\to i} < \infty$, which is a stronger assumption than (2.2.8). Finally, let us stress that our approach gives more than only the existence of a stationary solution. It actually gives a graphical construction of the stationary measure, which is not the case in [Mas98].

2.2.2 A Markovian description in terms of a coupled 'house of cards'-process.

As in the case of classical Hawkes processes with exponential loss functions, under suitable assumptions, an alternative description of $(Z^i, i \in I)$ via its intensity processes yields a Markovian description of the process. Throughout this subsection, we impose the summability assumption (2.2.8). Moreover, we suppose that

$$\sup_i \Lambda_i |\mathcal{V}|_{i\to\cdot} < \infty, \ \inf_i \delta_i = \delta > 0 \tag{2.2.14}$$

and that

$$\mathcal{V}_i := \mathcal{V}_{\cdot\to i} \cup \mathcal{V}_{i\to\cdot} \text{ are finite sets for all } i \in I. \tag{2.2.15}$$

$\mathcal{V}_i$ is the set of neurons that directly influence neuron i or that are directly influenced by it, i.e. the set of neurons j such that $W_{j\to i} \neq 0$ or $W_{i\to j} \neq 0$.

It is convenient to adopt a description of a process living on the set of directed edges $\mathcal{E} = \{j \to i, j \in \mathcal{V}_{\cdot\to i}, i \in I\}$. We write $e = j \to i \in \mathcal{E}$ for a directed edge and introduce for any such e the process $U_t(e)$ defined by

$$U_t(e) = Z^j\left(]L_t^i, t[\right), t \in \mathbb{R}. \tag{2.2.16}$$

With this point of view, the neural network is described by the process $(U_t(e), e \in \mathcal{E})_{t\in\mathbb{R}}$, taking values in $S := \mathbb{N}^{\mathcal{E}}$. Its dynamic is described by its generator defined by

$$\mathcal{G}f(\eta) = \sum_{i\in I} \psi_i \left(\sum_j W_{j\to i}(\eta(j\to i) \wedge K_{j\to i}) \right) \left[f\left(\eta + \Delta_i\eta\right) - f(\eta) \right],$$

where

$$(\Delta_i\eta)(k\to l) = \left\{ \begin{array}{ll} -\eta(k\to l) & \text{if } l = i, k \in \mathcal{V}_{\cdot\to i} \\ 1 & \text{if } k = i, l \in \mathcal{V}_{i\to\cdot} \\ 0 & \text{else} \end{array} \right\},$$

and where $f \in \mathcal{D}(\mathcal{G}) = \{f : |||f||| := \sum_{e\in\mathcal{E}} \Delta_f(e) < \infty\}$ with $\Delta_f(e) = \sup\{|f(\eta) - f(\zeta)| : \eta, \zeta \in S, \eta(e') = \zeta(e') \text{ for all } e' \neq e\}$.

Remark 8. *(i) Notice that a spike of neuron i does not only affect all neurons $j \in \mathcal{V}_{i\to\cdot}$ which receive an additional potential, but also all neurons $j \in \mathcal{V}_{\cdot\to i}$, since all $\eta(j \to i)$ are reset to 0 when a spike of neuron i occurs. It is for this reason that we call the above process a coupled 'house of cards'-process.*

(ii) We could also work with $\tilde{S} := \{\eta \in S : \eta(j \to i) \leq K_{j\to i} \ \forall e = (j \to i) \in \mathcal{E}\}$ *which is the state space of relevant configurations of the process. This would imply to redefine* $U_t(j \to i) := Z^j\left(]L_t^i, t[\right) \wedge K_{j\to i}$.

We introduce $T_i := \{e = j_1 \to j_2 \in \mathcal{E} : j_1 = i \text{ or } j_2 = i\}$.

By Theorem 3.9 of Chapter 1 of Liggett [Lig05], (2.2.13) together with (2.2.14) and (2.2.15) implies that $\mathcal{G}$ is the generator of a Feller process $(U_t(e), e \in \mathcal{E})_{t\in\mathbb{R}}$ on S.

Proof. Under the above conditions, the generator $\mathcal{G}$ can be rewritten as

$$\mathcal{G}f(\eta) = \sum_i c_{T_i}(\eta, d\zeta)[f(\eta_i^\zeta) - f(\eta)],$$

where

$$\eta_i^\zeta(e) = \left\{ \begin{array}{ll} \eta(e) & \text{if } e \notin T_i \\ \zeta(e) & \text{if } e \in T_i \end{array} \right\},$$

and where

$$c_{T_i}(\eta, d\zeta) = \psi_i(\sum_j W_{j\to i}(\eta(j \to i) \wedge K_{j\to i}))\delta_{\eta+\Delta_i\eta}(d\zeta).$$

A straightforward calculation shows that the quantity $c_{T_i} = \sup_\eta c_{T_i}(\eta, S)$ defined by Liggett [Lig05] in formula (3.3) of Chapter 1 can be upper bounded by $c_{T_i} \leq \Lambda_i$ and that

$$c_{T_i}(e) := \sup\{\|c_{T_i}(\eta, \cdot) - c_{T_i}(\zeta, \cdot)\|_{TV} : \eta(e') = \zeta(e') \text{ for all } e' \neq e\},$$

where $\|\cdot\|_{TV}$ denotes the total variation distance, can be controlled by

$$c_{T_i}(e) \leq \left\{ \begin{array}{ll} \Lambda_i\, \omega(|W_e|K_e) & \text{if } e = k \to i, \\ \Lambda_i & \text{if } e = i \to l, \\ 0 & \text{else} \end{array} \right\}.$$

Condition (2.2.13) together with (2.2.14) and (2.2.15) thus imply that M defined in formula (3.8) of Chapter 1 of Liggett [Lig05] can be controlled as

$$M = \sup_e \sum_{i:e\in T_i} \sum_{u\neq e} c_{T_i}(u) \leq 2 \sup_i \Lambda_i \left(\sum_k \omega(|W_{k\to i}|K_{k\to i}) + |\mathcal{V}|_{i\to\cdot} \right) < \infty,$$

and then Theorem 3.9 of Chapter 1 of Liggett [Lig05] allows to conclude. □

As a consequence, we can reformulate Theorem 6 in the following way.

Theorem 9. *Grant the assumptions of Theorem 6 and suppose moreover that (2.2.8) together with (2.2.13), (2.2.14) and (2.2.15) are satisfied. Then the process* $(U_t^e, e \in \mathcal{E})_{t\in\mathbb{R}}$ *is ergodic.*

Proof. The above Theorem 9 follows immediately from Theorem 6. □

Remark 10. *Let us compare the above result to the* $M < \varepsilon-$*criterion of Liggett [Lig05], Theorem 4.1 of Chapter 1. The quantity* M *has already been introduced above. Moreover,* ε *is defined by*

$$\varepsilon := \inf_{e\in\mathcal{E}} \inf_{\eta_1,\eta_2\in S:\eta_1(e')=\eta_2(e')\forall e'\neq e} \sum_{k:e\in T_k} \Big[c_{T_k}\left(\eta_1, \{\zeta : \zeta(e) = \eta_2(e)\}\right) + c_{T_k}\left(\eta_2, \{\zeta : \zeta(e) = \eta_1(e)\}\right) \Big].$$

The sufficient condition for ergodicity in Theorem 4.1 of Liggett [Lig05] is $M < \varepsilon$. Note that in our particular case we have $\varepsilon = 0$ which can easily be seen by considering η_1 and η_2 with $\eta_1(e) = 1$ and $\eta_2(e) = 3$ for some $e \in \mathcal{E}$. Consequently the sufficient condition $M < \varepsilon$ is not satisfied. However, Theorem 9 implies the ergodicity of the process without the condition $M < \varepsilon$.

2.3 A cascade of spike trains.

We now describe the second model that we consider in this paper, a *cascade of spike trains*. We suppose that for each $j \in I$, the function $g_j : \mathbb{R}_+ \to \mathbb{R}_+$ is measurable and non-increasing such that $\int_0^{+\infty} g_j(x)dx < +\infty$. The function g_j models a leak phenomenon. We suppose moreover that

$$h_{j\to i}(x) = W_{j\to i} \cdot x, \tag{2.3.17}$$

for a family of *synaptic weights* $\{W_{j\to i} \in \mathbb{R}\}$ satisfying the summability condition

$$\sup_{i\in I} \sum_j |W_{j\to i}| < \infty. \tag{2.3.18}$$

Neurons j such that $W_{j\to i} > 0$ are called *excitatory* for i, if $W_{j\to i} < 0$, then j is called *inhibitory* for i. Finally, we impose the following condition on the structure of interactions.

Assumption 11. *The set I of the neurons is divided into layers $(I_n)_{n\in\mathbb{Z}}$ such that we have the partition $I = \sqcup_{n\in\mathbb{Z}} I_n$. For each $n \in \mathbb{Z}$ and for each $i \in I_n$, we suppose that*

$$\mathcal{V}_{\cdot\to i} \subset I_{n-1}. \tag{2.3.19}$$

Therefore, a neuron only receives information from neurons in the layer just above itself. This assumption does not apply to the brain's structure which is far too complicated. But such a structure can be found in simpler nervous tissues like the retina. The nonlinear Hawkes process that we consider in this section is defined by its intensity given by

$$\lambda_t^i = \psi_i \left(\sum_{j\in I} W_{j\to i} \int_{[L_t^i, t[} g_j(t-s) dZ_s^j \right), \tag{2.3.20}$$

together with the assumption (2.3.19) on the structure of the interactions.

In order to state our result for this model, we need to strengthen Assumption 4.

Assumption 12. *The family $(\phi_i)_{i\in I}$ is uniformly Lipschitz continuous, i.e. there exists $\gamma > 0$, such that for all $x, x' \in \mathbb{R}, i \in I$,*

$$|\phi_i(x) - \phi_i(x')| \le \gamma |x - x'|. \tag{2.3.21}$$

Theorem 13. *Grant Assumptions 11 and 12 and suppose moreover that condition (2.3.18) is satisfied. If*

$$\sup_{i\in I} \left(\sum_{k\ge 1} \left[\left(k \left(\sum_{j\in V_i(k)} \Lambda_j \right) + 1 \right) \times \right.\right.$$
$$\left.\left. \left(\sum_{j\in V_i(k-1)} |W_{j\to i}| \Lambda_j \int_{k-1}^k g_j(s)ds + \sum_{j\in \partial V_i(k-1)} |W_{j\to i}| \Lambda_j \int_0^k g_j(s)ds \right) \right] \right) < \frac{1}{\gamma}, \tag{2.3.22}$$

where γ is given in Assumption 12, then there exists a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$ such that under $\mathbb{P}$, the canonical process $\left(Z^i, i \in I\right)$ on Ω is a stationary nonlinear Hawkes process with variable length memory and an infinite number of interacting components, with parameters (ψ, h, g) where h is given by (2.3.17).

Remark 14. *(i) A similar result still holds imposing only the weaker Assumption 4 instead of Assumption 12. However, in this case, the summability condition (2.3.22) has to be restated in terms of the modulus of continuity of the ϕ_i and turns out to be much more complicated than (2.3.22) above. In particular, it can not be expressed directly in terms of the parameters of the model. See Remark 28 following the proof of Theorem 13 for more details.*

(ii) Let us once more compare this result with Theorem 3 of [Mas98]. Instead of our conditions (2.3.22) and Assumption 11 on the structure of interactions, [Mas98] requires that

$$\sum_{i \in I} \sum_{j \in I} W_{j \to i} \int_0^{+\infty} t g_j(t) dt < \infty.$$

A direct comparison with (2.3.22) is difficult, but (2.3.22) seems to be slightly less restrictive.

(iii) We can restate our result in the framework of 1-dimensional Hawkes processes with the following assumptions: $I = \{i\}$, $W_{i \to i} \neq 0$, $\Lambda_i = \Lambda \in \mathbb{R}_+$ and allowing for a time dependence where L_t^i is replaced by $-\infty$ for all $t \in \mathbb{R}$. In this framework our condition (2.3.22) *reads as*

$$\sum_{k \geq 1} \left[(k\Lambda + 1) \int_{k-1}^{k} g(s) ds \right] < \frac{1}{\Lambda \gamma}.$$

This assumption is stronger than the following condition (4) of [BM96]

$$\int_0^{\infty} s g(s) ds < +\infty,$$

since condition (4) of [BM96] does not impose a condition on the value of the L^1-norm of g and on the Lipschitz constant γ. However, our method gives a graphical construction of the stationary measure by the mean of a perfect simulation algorithm. To the best of our knowledge this has been obtained in the literature only for linear Hawkes processes as in [MR05], whereas our process has a non-linear intensity.

We give an example in which condition (2.3.22) is satisfied in order to illustrate our result. For the sake of simplicity, we will assume that for all $j, \Lambda_j = \Lambda$ and $g_j = g$. We will also choose a simple structure for the network; each layer of neurons is identical and indexed by $\mathbb{Z}$ so that we can take $I = \mathbb{Z}^2$ with indexes $i = (i_1, i_2) \in \mathbb{Z}^2$ where i_1 denotes the 'site' of the neuron and i_2 the 'height' of its layer.

Example 15. *Suppose that $g(s) = e^{-as}$ for some $a > 0$. We assume that each neuron is influenced by an infinite number of neurons and that*

$$W_{(j_1, j_2) \to (i_1, i_2)} = \frac{W}{|j_1 - i_1|^{\beta}} \mathbf{1}_{j_2 = i_2 - 1} \mathbf{1}_{j_1 \neq i_1}$$

for some $\beta > 0$ and $W > 0$. Putting $V_{(i_1,i_2)}(k) := \{(j_1, i_2 - 1); 1 \leq |j_1 - i_1| \leq k\}$, we have for all neurons $j \in \partial V_i(k-1), W_{j\to i} = \frac{1}{k^\beta}$. Condition (2.3.22) can be rewritten as

$$\sum_{k\geq 2}\left[\Big(k\Lambda(2k+1)+1\Big)\left(\sum_{l=1}^{k-1}\frac{\Lambda}{l^\beta}e^{-ak}(e^a-1)+\frac{\Lambda}{k^\beta}(1-e^{-ak})\right)\right] < \frac{a}{2W\gamma}.$$

This sum is finite if and only if $\beta > 3$. In particular, for fixed a, Λ, γ and $\beta > 3$ there exists W^* such that $W < W^*$ implies (2.3.22).

Remark 16. *In the model of Section 2.2 the presence of spontaneous spikes guarantees that the neural network is always active. In the frame of cascades of spike trains, we do not impose the presence of spontaneous spikes. Thus we have to study the non-extinction of the process. Two cases have to be considered. Firstly, if for all $i \in I, \psi_i(0) = 0$, then $Z^i \equiv \mathbf{0}$ for all $i \in I$ is a possible stationary version of the Hawkes process with intensity (2.3.20). Since we are dealing with the uniqueness regime here, Theorem 13 implies that this is the unique stationary solution in this case. This situation is of course not of interest for us.*

The second case to be considered is that $\psi_i(0) > 0$ for at least one i. In this situation, the process does not go extinct as shows the following lemma. As a consequence, in this case, the only invariant measure of the process is not the trivial one.

Lemma 17. *If there exists $i \in I$ such that $\psi_i(0) > 0$, then for all $t \in \mathbb{R}$, we have that almost surely,*

$$\sum_{j\in\{i\}\cup\mathcal{V}_{\cdot\to i}} Z^j\left([t,+\infty)\right) > 0.$$

Proof. $\psi_i(0) > 0$ means that if the neuron i receives no information from the other neurons, it has a positive rate of fire. In other words, the only way for the neuron i to stay silent is to be regularly inhibited. In both situations we will have an activity in the network, either of neuron i itself or of its inhibitors. The idea of the proof is then to suppose that for all $j \in \mathcal{V}_{\cdot\to i}$, $Z^j\left([t,+\infty)\right) = 0$, and to prove that in this case, almost surely $Z^i\left([t,+\infty)\right) > 0$ using the hypothesis $\psi_i(0) > 0$ together with the continuity of ψ_i. □

Remark 18 (Some remarks on the extinction problem)**.** *We continue the discussion of Remark 16 concerning the extinction probability of the system in the case $\psi_i(0) = 0$ for all $i \in I$. As pointed out above, in this case, the only invariant measure of the system is the trivial one δ_0.*

This situation might be different when considering the process from a macroscopic point of view, reminiscent of what is called 'hydrodynamical limit behavior' in statistical physics. To be more precise, suppose that we observe a set of N neurons, $I = I^N = \{1, \ldots, N\}$, that $W_{j\to i} = \frac{1}{N}$ for all $i \neq j$, $g_j(t) = e^{-\alpha t}$ for some $\alpha > 0$ and that $\psi_i(x) = \psi(x)$ for all i. We suppose moreover that $\psi(0) = 0$ and that $\psi'(0) = \beta$. In this case, for fixed N, the process goes extinct almost surely, as has been shown in Theorem 2.3 of Duarte and Ost [DO14]. Let us now consider the macroscopic behavior as $N \to \infty$. It has been proved (see e.g. Fournier and Löcherbach [FL14]) that the system possesses the so-called 'propagation-of-chaos'-property. This means that in the large population limit, the neurons converge in law to independent and identically distributed copies of the same limit law. In this macroscopic limit, the longtime behavior of a typical neuron can be different from the one of a typical neuron in a finite size system. In particular, in Theorem 12 of

Duarte [Dua15] it has been shown that under the assumption that $\beta > \alpha$, the macroscopic system does not go extinct almost surely in the long run, although $\psi(0) = 0$. In other words, the trivial invariant measure is unstable in this case.

However, the point of view adopted in the present paper is a different one. We are looking at infinite systems of neurons, but at a microscopic level at which each neuron is observed at scale 1*. In this situation, the only way of preventing the system from going extinct is summarized in the above Lemma 17.*

2.4 A dominating Poisson Random Measure.

Recall that the firing rate functions ψ_i considered in this paper are bounded by a constant Λ_i for each $i \in I$. We will use this assumption to introduce a Poisson Random Measure $N(dt, di, dz)$ on $\mathbb{R} \times I \times [0,1]$ with intensity $dt \left(\sum_{i \in I} \Lambda_i \delta_i\right) dz$ on $\mathbb{R} \times I \times [0,1]$ dominating the process $\left(Z^i, i \in I\right)$. This allows to select the jump times of Z^i among those of N^i according to probabilities driven by the function ϕ_i (recall (2.1.3)). This leads to the following definition.

Definition 19. *A family $(Z^i, i \in I)$ of random point measures defined on $(\Omega, \mathcal{A}, \mathbb{F})$ is said to be a Hawkes process with variable length memory and an infinite number of interacting components with parameters (ψ, h, g) if almost surely, for all $i \in I$ and $C \in \mathcal{B}(\mathbb{R})$,*

$$Z^i(C) = \int_C \int_{\{i\}} \int_{[0,1]} 1_{\{z \leq \frac{1}{\Lambda_i} \psi_i \left(\sum_j h_{j \to i} \left(\int_{[L_t^i, t[} g_j(t-s) dZ_s^j \right) \right)\}} N(dt, di, dz). \tag{2.4.23}$$

According to Brémaud and Massoulié [BM96], see also Proposition 3 of Delattre, Fournier and Hoffmann [DFH16], a Hawkes process according to Definition 19 is a Hawkes process according to Definition 3 and vice versa.

Formula (2.4.23) implies that we can construct the process $(Z^i, i \in I)$ by a thinning procedure applied to the a priori family of dominating Poisson random measures $N^i(dt) = N(dt, \{i\}, [0,1])$ having intensity $\Lambda_i dt$ each. Since Z^i is a simple point measure, it is enough to define it through the times of its atoms. Each atom of Z^i must also be an atom of N^i since $Z^i \ll N^i$. In other words, the associated counting process $\alpha(Z^i)$ can only jump when the counting process $\alpha(N^i)$ jumps. We write $T_n^i, n \in \mathbb{Z}$, for the jump times of N^i. Fix a given jump $t = T_n^i$ of N^i. Then, conditionally on N, the probability that this time is also a jump time of Z^i is given by

$$\mathbb{P}(Z^i(\{t\}) = 1 | \mathcal{F}_t) = \phi_i \left(\sum_j W_{j \to i} \left(\int_{[L_t^i, t[} g_j(t-s) dZ_s^j \right) \right) =: p_{(i,t)}(1 | \mathcal{F}_t). \tag{2.4.24}$$

In other words, given that t is a jump time of N^i, $p_{(i,t)}(1|\mathcal{F}_t)$ is the probability that this jump is also a jump of Z^i. This probability depends on the past before time t of the process. In what follows we propose a decomposition of $p_{(i,t)}(1|\mathcal{F}_t)$ according to growing time-space neighborhoods that explore the *relevant past* needed in order to determine $p_{(i,t)}(1|\mathcal{F}_t)$. This decomposition is a Kalikow-type decomposition as considered first by Ferrari, Maass, Martínez and Ney [FMMN00] and in Comets, Fernández and Ferrari [CD01]. This type of technique was then considered in a series of papers for perfect simulation issues. See Galves and Löcherbach [GL13] for an application in the context of neural biological nets in discrete time. The decomposition that we consider here is a nontrivial extension of the previous considerations to the framework of continuous time

neural nets. In the case of Theorem 13, this decomposition has to be achieved in a random environment, where the environment is given by the a priori realization of the PRM N. We start with the proof of Theorem 6 which is conceptually simpler.

2.5 Proof of Theorem 6 .

2.5.1 Kalikow-type decomposition.

The condition (2.2.10) of Theorem 6 allows us to decompose the law of the conditional probability (2.4.24) according to space neighborhoods $V_i(k)$. This decomposition will be independent of the realization of the a priori PRM N. This will be crucial in the perfect simulation procedure described in the next subsection. We will work with $\tilde{S}$, the state space of relevant configurations of the process defined in Remark 8. For the convenience of the reader, we recall its definition here:

$$\tilde{S} := \{\eta \in S : \eta(j \to i) \leq K_{j\to i} \ \forall e = (j \to i) \in \mathcal{E}\}$$

and introduce the following notations. First,

$$r_i^{[0]}(1) = \inf_{\eta\in\tilde{S}} \phi_i \left(\sum_j W_{j\to i}\eta(j \to i) \right), \tag{2.5.25}$$

which is the minimal probability that neuron i spikes uniformly with respect to all possible configurations. Then we define

$$r_i^{[0]}(1) = \inf_{\eta\in\tilde{S}} \left(1 - \phi_i \left(\sum_j W_{j\to i}\eta(j \to i) \right) \right), \tag{2.5.26}$$

which is the minimal probability that neuron i does not spike. Next, for each $k \geq 1$ and each $\zeta \in \tilde{S}$, we define the set $D_i^k(\zeta)$ by

$$D_i^k(\zeta) := \{\eta \in \tilde{S} : \forall j \in V_i(k), \eta(j \to i) = \zeta(j \to i)\}$$

and put

$$r_i^{[k]}(1|\zeta) = \inf_{\eta\in D_i^k(\zeta)} \phi_i \left(\sum_j W_{j\to i}\eta(j \to i) \right), \quad r_i^{[k]}(0|\zeta) = \inf_{\eta\in D_i^k(\zeta)} \left(1 - \phi_i \left(\sum_j W_{j\to i}\eta(j \to i) \right) \right). \tag{2.5.27}$$

Finally, we define

$$\alpha_i(0) = r_i^{[0]}(1) + r_i^{[0]}(0), \quad \alpha_i(k) = \inf_{\zeta\in\tilde{S}} \left(r_i^{[k]}(1|\zeta) + r_i^{[k]}(0|\zeta) \right)$$

for all $k \geq 1$ and let

$$\mu_i(0) = \alpha_i(0) \text{ and } \mu_i(k) = \alpha_i(k) - \alpha_i(k-1)$$

for all $k \geq 1$.

Lemma 20. $(\mu_i(k))_{k\geq 0}$ *defines a probability on* $\mathbb{N}$.

Proof. Indeed, since $D_i^k(\zeta) \subset D_i^{k-1}(\zeta)$, we have $\mu_i(k) \geq 0$ for all $k \geq 0$. Therefore, all we have to show is that

$$\sum_{k\geq 0} \mu_i(k) = 1. \tag{2.5.28}$$

Note that $\sum_{k\geq 0} \mu_i(k) = \lim_{k\to+\infty} \left[\inf_{\zeta\in\tilde{S}} \left(r_i^{[k]}(1|\zeta) + r_i^{[k]}(0|\zeta)\right)\right]$. Hence it is sufficient to show that for all $\zeta \in \tilde{S}$,

$$\lim_{k\to+\infty} \left(r_i^{[k]}(1|\zeta) + r_i^{[k]}(0|\zeta)\right) = 1.$$

For all $k \geq 0$, we have

$$r_i^{[k]}(1|\zeta)+r_i^{[k]}(0|\zeta) = 1-\left[\sup_{\eta\in D_i^k(\zeta)} \phi_i\left(\sum_j W_{j\to i}\eta(j\to i)\right) - \inf_{\eta\in D_i^k(\zeta)} \phi_i\left(\sum_j W_{j\to i}\eta(j\to i)\right)\right].$$

Using that ϕ_i is increasing and uniformly continuous with modulus ω, we deduce that

$$\begin{aligned}
&\sup_{\eta\in D_i^k(\zeta)} \phi_i\left(\sum_j W_{j\to i}\eta(j\to i)\right) - \inf_{\eta\in D_i^k(\zeta)} \phi_i\left(\sum_j W_{j\to i}\eta(j\to i)\right) \\
&\qquad \leq \omega\left(\sup_{\eta\in D_i^k(\zeta)} \left(\sum_j W_{j\to i}\eta(j\to i)\right) - \inf_{\eta\in D_i^k(\zeta)} \left(\sum_j W_{j\to i}\eta(j\to i)\right)\right) \\
&\qquad\qquad \leq \omega\left(\sum_{j\notin V_i(k)} |W_{j\to i}|\, K_{j\to i}\right).
\end{aligned}$$

Now, taking the limit as k tends to $+\infty$ and taking into account condition (2.2.8), we obtain (2.5.28) as desired. □

Let us come back to the conditional probability $p_{(i,t)}(1|\mathcal{F}_t)$ introduced in (2.4.24). The history is realized only through the effected choices of acceptance or rejection of jumps of the a priori PRM N. Therefore we introduce the time grid $\mathcal{G} = \{(i, T_n^i), i \in I\}$. Any realization of the Hawkes process, conditionally with respect to the PRM N, can be identified with an element of $X := \{0,1\}^{\mathcal{G}}$. We write $x = (x^i)_{i\in I}$ for elements of X, where $x^i = (x^i(T_n^i))_{n\in\mathbb{Z}}$. Elements $x \in X$ can be interpreted as point measures. The object of our study is

$$p_{(i,t)}(1|x) = \phi_i\left(\sum_j W_{j\to i}\left(x^j\left([L_t^i(x), t[\right) \wedge K_{j\to i}\right)\right).$$

The following proposition establishes a Kalikow-type decomposition for $p_{(i,t)}(\cdot|x)$ with respect to growing neighborhoods of $\mathcal{V}_{\cdot\to i}$.

Proposition 21. *Grant Assumption 4 and assume that (2.2.8) is satisfied. Fix $t = T_n^i$ for some $n \in \mathbb{Z}$ and $i \in I$. Then there exists a family of conditional probabilities $(p_{(i,t)}^{[k]}(\cdot|x))_{k\geq 0}$ on $\{0,1\}$ satisfying the following properties.*

1. *For all $a \in \{0,1\}, p_{(i,t)}^{[0]}(a|x) := \frac{r_i^{[0]}(a)}{\mu_i(0)}$ does not depend on the configuration x.*

2. *For all* $a \in \{0,1\}$, $k \geq 1$, $X \ni x \mapsto p^{[k]}_{(i,t)}(a|x)$ *depends only on the variables* $x^j : j \in V_i(k)$.

3. *For all* $x \in X$, $k \geq 1$, $p^{[k]}_{(i,t)}(1|x) \geq 0$ *and* $p^{[k]}_{(i,t)}(1|x) + p^{[k]}_{(i,t)}(0|x) = 1$.

4. *For all* $x \in X$, *we have the following convex decomposition.*

$$p_{(i,t)}(a|x) = \sum_{k\geq 0} \mu_i(k) p^{[k]}_{(i,t)}(a|x), \tag{2.5.29}$$

Proof. We identify a configuration $x \in X$ with an element x_t of $\tilde{S}$ for any $t \in \mathbb{R}$ by introducing

$$x_t(j \to i) = x^j([L^i_t, t[) \wedge K_{j\to i},$$

where x^j is interpreted as point measure. Note that $p_{(i,t)}(a|x)$ only depends on x_t. We have

$$p_{(i,t)}(a|x) = p_{(i,t)}(a|x_t) = r^{[0]}_i(a) + \sum_{k=1}^{N} \Delta^{[k]}_i(a|x_t) + \left(p_{(i,t)}(a|x_t) - r^{[N]}_i(a|x_t)\right), \tag{2.5.30}$$

where $\Delta^{[k]}_i(a|x_t) := r^{[k]}_i(a|x_t) - r^{[k-1]}_i(a|x_t)$. We start by showing that the last term in (2.5.30) tends to 0 as $N \to \infty$. Indeed, rewriting the definitions, we have

$$\left|p_{(i,t)}(a|x_t) - r^{[N]}_i(a|x_t)\right| = \sup_{z\in D^N_i(x_t)} \left|\phi_i\left(\sum_j W_{j\to i} x_t(j\to i)\right) - \phi_i\left(\sum_j W_{j\to i} z_{j\to i}\right)\right|.$$

Using the uniform continuity of ϕ_i and the definition of $D^N_i(x_t)$, we get

$$\left|p_{(i,t)}(a|x_t) - r^{[N]}_i(a|x_t)\right| \leq \sup_{z\in D^N_i(x_t)} \left[\omega\left(\sum_j |W_{j\to i}||x_t(j\to i) - z_{j\to i}|\right)\right]$$
$$\leq \sup_{z\in \tilde{S}} \left[\omega\left(\sum_{j\notin V_i(N)} |W_{j\to i}||x_t(j\to i) - z_{j\to i}|\right)\right] \leq \omega\left(\sum_{j\notin V_i(N)} |W_{j\to i}| K_{j\to i}\right) \underset{N\to+\infty}{\longrightarrow} 0.$$

Taking the limit when N tends to $+\infty$ in (2.5.30), we obtain $p_{(i,t)}(a|x_t) = r^{[0]}_i(a) + \left(\sum_{k\geq 1} \Delta^{[k]}_i(a|x_t)\right)$. Now we put for $k \geq 1$

$$\tilde{\mu}_i(k, x_t) := \sum_a \Delta^{[k]}_i(a|x_t) \quad \text{and} \quad \tilde{p}^{[k]}_{(i,t)}(a|x) = \frac{\Delta^{[k]}_i(a|x_t)}{\tilde{\mu}_i(k, x_t)}, \tag{2.5.31}$$

where we define $\tilde{p}^{[k]}_{(i,t)}(a|x_t)$ in an arbitrary way if $\tilde{\mu}_i(k, x_t) = 0$. In this way we can write $\Delta^{[k]}_i(a|x_t) = \tilde{\mu}_i(k, x_t)\tilde{p}^{[k]}_{(i,t)}(a|x_t)$ and therefore

$$p_{(i,t)}(a|x_t) = \mu_i(0)\, p^{[0]}_{(i,t)}(a) + \sum_{k=1}^{\infty} \tilde{\mu}_i(k, x_t)\, \tilde{p}^{[k]}_{(i,t)}(a|x_t). \tag{2.5.32}$$

This decomposition is not yet the one announced in the proposition since the $\tilde{\mu}_i(k, x_t)$ depend on the configuration x_t. The weights $\mu_i(k)$ are already defined and they do not

depend on the configuration. So we have to define new probabilities $p^{[k]}_{(i,t)}$, based on the previously defined $\tilde{p}^{[k]}_{(i,t)}$.

To define the new probabilities $p^{[k]}_{(i,t)}$, we introduce $\alpha_i(k, x_t) := \sum_{l=1}^{k} \tilde{\mu}_i(l, x_t)$. For each $k \geq 0$ let l' and l be indexes such that

$$\alpha_i(l'-1, x_t) < \alpha_i(k-1) \leq \alpha_i(l', x_t) < \ldots < \alpha_i(l, x_t) < \alpha_i(k) \leq \alpha_i(l+1, x_t).$$

We then decompose the interval $]\alpha_i(k-1), \alpha_i(k)]$ in the following way.

$$]\alpha_i(k-1), \alpha_i(k)] =]\alpha_i(k-1), \alpha_i(l', x_t)] \cup \left(\bigcup_{m=l'+1}^{l}]\alpha_i(m-1, x_t), \alpha_i(m, x_t)] \right) \cup]\alpha_i(l, x_t), \alpha_i(k)].$$

We define $p^{[k]}_i$ on each of the intervals of this partition. On the first interval $]\alpha_i(k-1), \alpha_i(l', x_t)]$, we use $\tilde{p}^{[l']}_i$, on each interval $]\alpha_i(m-1, x_t), \alpha_i(m, x_t)]$, we use $\tilde{p}^{[m]}_i$, and on $]\alpha_i(l, x_t), \alpha_i(k)]$, we use $\tilde{p}^{[l+1]}_i$. This leads, for each $k \geq 0$, to the following definition.

$$\begin{aligned} p^{[k]}_{(i,t)}(a|x_t) = \sum_{-1=l' \leq l}^{k-1} & \mathbf{1}_{\{\alpha_i(l'-1,x_t) < \alpha_i(k-1) \leq \alpha_i(l',x_t)\}} \mathbf{1}_{\{\alpha_i(l,x_t) < \alpha_i(k) \leq \alpha_i(l+1,x_t)\}} \\ & \left[\frac{\alpha_i(l', x_t) - \alpha_i(k-1)}{\mu_i(k)} \, \tilde{p}^{[l']}_{(i,t)}(a|x) + \sum_{m=l'+1}^{l} \frac{\tilde{\mu}_i(m, x_t)}{\mu_i(k)} \, \tilde{p}^{[m]}_{(i,t)}(a|x) \right. \\ & \left. + \frac{\alpha_i(k) - \alpha_i(l, x_t)}{\mu_i(k)} \, \tilde{p}^{[l+1]}_{(i,t)}(a|x_t) \right]. \quad (2.5.33) \end{aligned}$$

It can easily be verified that with this definition, we obtain the announced decomposition. □

2.5.2 Perfect simulation

In this section we show how to construct the stationary nonlinear Hawkes process with saturation threshold by a *perfect simulation procedure*, based on an a priori realization of the processes $(N^i, i \in I)$. Condition (2.2.9) allows us to decompose the Poisson process N^i of intensity Λ_i as

$$N^i = \widehat{N}^i + \widetilde{N}^i,$$

where $\widehat{N}^i$ and $\widetilde{N}^i$ are independent Poisson processes with respective intensities δ_i and $\Lambda_i - \delta_i$. Conditionally on these processes, we can characterize the process Z^i by the element $x \in X = \{0,1\}^{\mathcal{G}}$ recording the times and the indices of the neuron for which we have accepted a jump. All jumps of $\widehat{N}^i$ are also jumps of Z^i; we call them the *spontaneous spikes*. They appear at a jump time T^i_n of N^i with probability $d_i := \frac{\delta_i}{\Lambda_i}$. Moreover, any jump $\widetilde{T^i_n}$ of $\widetilde{N}^i$ will be a jump of Z^i with probability

$$\frac{1}{1-d_i} \left[\phi_i \left(\sum_j W_{j \to i} \left(Z^j \left([L^i_t, t[\right) \wedge K_{j \to i} \right) \right) - d_i \right],$$

and we have to decide for each $i \in I$ and each time $\widetilde{T^i_n}$ whether this jump is accepted or not. This acceptance/rejection procedure will be achieved by means of the Kalikow-type decomposition and gives rise to a perfect-simulation algorithm that we are going to introduce now.

Fix $i \in I$ and $t = \widetilde{T}^i_n$. Our first algorithm describes the random space-time subset of neurons and their associated spiking times that can possibly have an influence on the acceptance or rejection of a spike of neuron i at time t.

0. In the following steps of this algorithm, we will have to simulate $\widehat{N}^j$ and $\widetilde{N}^j$ for some sites $j \in I$ on bounded intervals of time. We introduce for each $j \in I$ the sets $\widehat{S}^j$ and $\widetilde{S}^j$ that will contain the intervals of time where the processes $\widehat{N}^j$ and $\widetilde{N}^j$ have already been simulated at the current step of the algorithm. We initialize these sets so that for all $j \in I$, $\widehat{S}^j = \widetilde{S}^j = \emptyset$.

1. We simulate a random variable $K^{(i,t)} \in \mathbb{N}$ such that

$$\mathbb{P}(K^{(i,t)} = k) = \mu_j(k), \text{ for all } k \geq 0.$$

2. We determine $\widehat{L}^i_t = \widehat{T}^i_{\widehat{N}^i_{t-}}$ in the following way. We simulate a random variable τ following an exponential law of parameter δ_i and introduce $\widehat{S}^{(i,t)}_{max} := \sup\{s \in \widehat{S}^i \cap]-\infty, t]\}$ with the convention $\sup(\emptyset) = -\infty$.

   If $\tau \leq t - \widehat{S}^{(i,t)}_{max}$, we put $\widehat{L}^i_t = t - \tau$.

   If $\tau > t - \widehat{S}^{(i,t)}_{max}$, we choose for $\widehat{L}^i_t$ the biggest jump time of $\widehat{N}^i$ in $\widehat{S}^i \cap]-\infty, t]$.

   In both cases, we update $\widehat{S}^i$, which in an algorithmic way can be written $\widehat{S}^i \leftarrow \widehat{S}^i \cup [\widehat{L}^i_t, t[$.

3. For each $j \in V_i\left(K^{(i,t)}\right)$, we simulate $\widetilde{N}^j$ on the time interval $[\widehat{L}^i_t, t[\backslash \widetilde{S}^j$ and update $\widetilde{S}^j$ such that $\widetilde{S}^j \leftarrow \widetilde{S}^j \cup [\widehat{L}^i_t, t[$.

4. We introduce the set

$$C^{(i,t)}_1 := \left\{(j, \widetilde{T}^j_l) \in I \times]-\infty, t[: j \in V_i\left(K^{(i,t)}\right), \widetilde{T}^j_l \in [\widehat{L}^i_t, t[\right\}, \tag{2.5.34}$$

   which we call "first generation of the clan of ancestors" of element (i,t). Here, by convention, $C^{(i,t)}_1 = \emptyset$ if $K^{(i,t)} = 0$.

5. If $C^{(i,t)}_1 \neq \emptyset$, then we iterate the above procedure and simulate for each element of $C^{(i,t)}_1$ a new clan of ancestors with steps (1) to (4). So we put for any $n \geq 2$,

$$C^{(i,t)}_n = \left(\bigcup_{(j,s) \in C^{(i,t)}_{n-1}} C^{(j,s)}_1 \right) \setminus \left(C^{(i,t)}_1 \cup \ldots \cup C^{(i,t)}_{n-1} \right), \tag{2.5.35}$$

   which is the "$n-$th generation of the clan of ancestors" of (i,t).

Remark 22. *From a practical point of view, it is impossible to simulate once and for all an infinity of independent Poisson processes on an infinite interval of time. This is why we detail this algorithm simulating step by step the processes $\widehat{N}^i$ and $\widetilde{N}^i$ when it is needed on finite space-time neighbourhoods. However, we can simplify this algorithm assuming from a theoretical point of view that $\widehat{N}^i$ and $\widetilde{N}^i$ are simulated once and for all on $\mathbb{R}$ for all $i \in I$.*

Notice that if $K^{(i,t)} = 0$ in step (1), then the algorithm stops immediately. Also if for all $j \in V_i(k)$, for $k = K^{(i,t)}$, we have $\widetilde{N}^j([\widehat{L}_t^i, t[) = 0$, then $C_1^{(i,t)} = \emptyset$, and the algorithm stops after one step. Else we introduce

$$N^{Stop} := \min\{n : C_n^{(i,t)} = \emptyset\}$$

the number of steps of the algorithm, where $\min \emptyset := \infty$. The set $\mathcal{C}^{(i,t)} := \bigcup_{n=1}^{N^{Stop}} C_n^{(i,t)}$ contains all non-spontaneous spikes which have to be accepted or not and whose possible presence has an influence on the acceptance/rejection decision of (i,t) itself. We will show below that under the conditions of Theorem 6, $N^{Stop} < \infty$ almost surely.

Once the clans of ancestors are determined, we can realize the acceptance/rejection procedure of the elements in these clans in a second algorithm which is a forward procedure going from the past to the present. We start with the sites for which this decision can be made independently from anything else. During the algorithm the set of all sites for which a decision has already been achieved will then progressively be updated.

0. At the initial stage of the algorithm, the set of sites for which the acceptance/rejection decision can be achieved is initialized by

$$D^{(i,t)} := \left\{(j,s) \in \mathcal{C}^{(i,t)}, C_1^{(j,s)} = \emptyset\right\}.$$

   The sites within this set are precisely those for which the decision can be made independently from anything else.

1. For each $(j,s) \in D^{(i,t)}$, we simulate, according to the probabilities $\frac{1}{1-d_j}\left(p_{(j,s)}^{[0]}(a) - d_j\right)$, the state of this site.

2. For any $(j,s) \in \mathcal{C}^{(i,t)}$ with $C_1^{(j,s)} \subset D^{(i,t)}$, we then decide, according to the probabilities $\frac{1}{1-d_j}\left(p_{(j,s)}^{[k]}(a|x) - d_j\right)$ with $k = K^{(j,s)}$, to accept or to reject the presence of a spike of neuron j at time s. This is possible since $p_{(j,s)}^{[k]}(a|x)$ depends on the configuration $x \in X$ only through the sites in $C_1^{(j,s)}$ whose states have already been determined since $C_1^{(j,s)} \subset D^{(i,t)}$. In other words, even if the configuration x is not determined in all sites of $\mathcal{G}$, the sites in $D^{(i,t)}$ were simulated in the previous step and this is sufficient to determine $p_{(j,s)}^{[k]}(a|x)$.

   Then we update $D^{(i,t)}$ in the following way:

$$D^{(i,t)} \longleftarrow D^{(i,t)} \bigcup \left\{(j,s) \in \mathcal{C}^{(i,t)}, C_1^{(j,s)} \subset D^{(i,t)}\right\}.$$

3. The update of $D^{(i,t)}$ allows to repeat the previous step until $(i,t) \in D^{(i,t)}$.

Once we have assigned a decision to the element (i,t) itself, our *perfect simulation* algorithm stops. Of course, the whole procedure makes sense only if $N^{Stop} < +\infty$ a.s. which we will prove now. For that sake, we define $C_1^{(i,t)}(k)$ to be the clan of ancestors of element (i,t) of size k, i.e.

$$C_1^{(i,t)}(k) := \left\{(j,s) \in I \times \mathbb{R} : j \in V_i(k), \exists n \in \mathbb{Z} \text{ s.t. } s = \widetilde{T}_n^j, s \in \left[\widehat{L}_t^i; t\right)\right\},$$

with the convention $C_1^{(i,t)}(0) := \emptyset$ and put $M^{(i,t)} := \sum_{k\geq 1} \left|C_1^{(i,t)}(k)\right| \mu_i(k)$ which is the conditional expectation, conditionally on the realization of the PRM N, of $\left|C_1^{(i,t)}\right|$. In order to prove that $N^{Stop} < +\infty$ a.s., we compare the process $|C_n^{(i,t)}|$ with a branching process of reproduction mean M defined by $M := \sup_{i\in I, t\in\mathbb{R}} \mathbb{E}(M^{(i,t)})$. We will prove that the parameters of this branching process are such that $M < 1$, i.e. it goes extinct a.s. implying that $N^{Stop} < \infty$ almost surely.

Writing $\mathbb{E}^N$ for the conditional expectation with respect to N, we obtain the following recurrence.

$$\mathbb{E}^N(|C_n^{(i,t)}||C_{n-1}^{(i,t)}) \leq \sum_{(j,s)\in C_{n-1}^{(i,t)}} \sum_{k=1}^{+\infty} \sharp\left(C^{(j,s)}(k)\backslash C_{n-1}^{(i,t)}\right)\mu_i(k). \tag{2.5.36}$$

Here we use $\sum_{k=1}^{+\infty} \sharp\left(C^{(j,s)}(k)\backslash C_{n-1}^{(i,t)}\right)\mu_i(k)$ instead of $M^{(j,s)}$ in order to use the independence of $1_{(j,s)\in C_{n-1}^{(i,t)}}$ and $\sum_{k=1}^{+\infty} \sharp\left(C^{(j,s)}(k)\backslash C_{n-1}^{(i,t)}\right)\mu_i(k)$. This independence is due to the properties of the Poisson Random Measure associated with N and the fact that $C_{n-1}^{(i,t)}$ and $C_{(j,s)}(k)\backslash C_{n-1}^{(i,t)}$ are disjoint. It will allow us to claim

$$\mathbb{E}\left(\sum_{(j,s)} 1_{(j,s)\in C_{n-1}^{(i,t)}} \sum_{k=1}^{+\infty} \sharp\left(C^{(j,s)}(k)\backslash C_{n-1}^{(i,t)}\right)\mu_i(k)\right) \leq \sum_{(j,s)} \mathbb{E}(1_{(j,s)\in C_{n-1}^{(i,t)}})\mathbb{E}\left(M^{(j,s)}\right). \tag{2.5.37}$$

In order to prove the above inequality in a rigorous way, we study the transition operator governing the evolution (2.5.35). This leads to the definition of the following operator. We fix a neuron $i \in I$, a time $t = \widetilde{T}_n^i$ and put

$$Q\left((i,t),\bullet\right) = \sum_{k\geq 0} \mu_i(k) \int_0^{+\infty} dt_1 \left(\delta_i e^{-\delta_i t_1}\right) \left[\sum_{j\in V_i(k)} \left(\sum_{n_j\geq 0} e^{-(\Lambda_j-\delta_j)t_1} \frac{(\Lambda_j-\delta_j)^{n_j}}{n_j!} \int_{[t-t_1,t]^{n_j}} ds_1^j ... ds_{n_j}^j \delta_{\{(j,s_l^j):l=1,...,n_j\}}\right)\right].$$

$Q\left((i,t),dC_1\right)$ is the law of $C_1^{(i,t)}$. In the above definition, k is the "size" of the neighborhood $V_i(k)$ which is simulated according to the probabilities $(\mu_i(k))_{k\in\mathbb{N}}$; t_1 is the time between t and $\widehat{L}_t^i$ which is an exponential random variable of parameter δ_i; n_j is the Poisson random variable $\widetilde{N}^j\left(\left[\widehat{L}_t^i, t\right]\right)$ of parameter $(\Lambda_j - \delta_j)\, t_1$; and $\left(s_l^j\right)_{l\in\{1,...,n_j\}}$ is the family of jump times in $\left[\widehat{L}_t^i, t\right]$ of $\widetilde{N}^j$, which, conditionally on the event $\{\widetilde{N}^j\left(\left[\widehat{L}_t^i, t\right]\right) = n_j\}$, are uniform random variables on $\left[\widehat{L}_t^i, t\right] = [t-t_1, t]$.

The definition of the transition kernel $Q(C_1, dC_2)$ defining the law of $C_2^{(i,t)}$ knowing $C_1^{(i,t)}$) is more complicated since C_2 is not the result of independent simulations for each element of C_1.[1] But $Q(C_1, dC_2)$ can be upper bounded (in the sense of inclusion of the

[1]Indeed, the jump times of a process $\widetilde{N}^j$ simulated for the clan of ancestors of an element (i_1, t_1) of C_1 have to be simulated once and for all, and re-used for the determination of the clan of ancestors of an element (i_2, t_2) with $j \in V_{i_1}(k_1) \cap V_{i_2}(k_2)$. Similarly, the jump times of $\widehat{N}^i$ have to be simulated once and for all in order to determine all the last spontaneous spiking times for all $(i,t) \in C_1$.

simulated sets) by

$$\prod_{(j,s)\in C_1} Q((j,s), dC^{(j,s)})\delta_{(\cup_{(j,s)\in C_1} C^{(j,s)})}(dC_2).$$

This upper bound simulates more random variables than necessary leading to bigger clans of ancestors. Since we are only interested in obtaining upper-bounds on the number of elements in the clans of ancestors, we will therefore work with this upper bound and obtain

$$\begin{aligned}\mathbb{E}\left(\sum_{(j,s)} 1_{(j,l)\in C_{n-1}^{(i,t)}} \sum_{k=1}^{+\infty} \sharp\left(C^{(j,s)}(k)\backslash C_{n-1}^{(i,t)}\right)\mu_i(k)\right)\\ \leq \int_{(i,t)^c} Q((i,t),dC_1)\int_{C_1^c} Q(C_1,dC_2)...\int_{\left(\cup_{\kappa=1}^{n-2}C_\kappa\right)^c} Q(C_{n-2},dC_{n-1})\\ \left[\sum_{(j,s)\in C_{n-1}}\int_{\left(\cup_{\kappa=1}^{n-1}C_\kappa\right)^c} Q((j,s),dC)|C|\right]. \quad (2.5.38)\end{aligned}$$

Observe that

$$\int_{\left(\cup_{\kappa=1}^{n-1}C_\kappa\right)^c} Q((j,s),dC)|C| \leq \sum_{k\geq 0}\mu_j(k)\int_0^{+\infty} dt_1\left[\left(\delta_j e^{-\delta_j t_1}\right)\sum_{a\in V_j(k)}\left(\left(\Lambda_a-\delta_a\right)t_1\right)\right].$$

The right term of the above inequality being exactly $\mathbb{E}\left(M^{(j,s)}\right)$, this proves claim (2.5.37).

As a consequence, we can compare the process $|C_n^{(i,m)}|$ with a branching process. It remains to verify that this branching process gets extinct almost surely in finite time. We will prove that actually

$$M = \sup_{i\in I, t\in\mathbb{R}} \mathbb{E}(M^{(i,t)}) < 1.$$

Proposition 23. *For all $i \in I$ and all $k \geq 1$, we have,*

$$\mu_i(k) \leq \bar{\mu}_i(k). \quad (2.5.39)$$

We recall here the definition of $\bar{\mu}_i(k)$ in (2.2.11):

$$\bar{\mu}_i(k) := \omega\left(\sum_{j\in\partial V_i(k-1)} |W_{j\to i}|K_{j\to i}1_{W_{j\to i}<0}\right) + \omega\left(\sum_{j\in\partial V_i(k-1)} |W_{j\to i}|K_{j\to i}1_{W_{j\to i}>0}\right)$$

Proof. The proof follows directly from the fact that ϕ_i is increasing and uniformly continuous with modulus of continuity ω. For more details, we refer the reader to the proof of proposition 27 below which is similar (yet more difficult due to the fact that the decomposition considered there takes place in a random environment).

□

We are now able to finish the proof of Theorem 13. We have the following first upper-bound for $M^{(i,t)}$ thanks to the inequality (2.5.39) : $M^{(i,t)} \leq \sum_{k\geq 1} \left|C^{(i,t)}(k)\right| \bar{\mu}_i(k)$. Consequently, by definition of $C^{(i,t)}(k)$,

$$\mathbb{E}\left(M_{(i,t)}\right) \leq \sum_{k\geq 1} \mathbb{E}\left(\sum_{j\in V_i(k)} \widetilde{N}^j\left(\left[\widehat{L}_t^i, t\right]\right)\right) \bar{\mu}_i(k) \leq \sum_{k\geq 1}\left[\left(\sum_{j\in V_i(k)} \frac{\Lambda_j - \delta_j}{\delta_i}\right)\bar{\mu}_i(k)\right].$$

Assumption (2.2.10) of Theorem 6 implies that $M = \sup_i \mathbb{E}\left(M_{(i,t)}\right) < 1$. Consequently,

$$\mathbb{E}(|C_n^{(i,t)}|) \leq M^n \to 0 \text{ when } n \to \infty,$$

implying that the process $\left(C_n^{(i,t)}\right)_{n\in\mathbb{N}}$ goes extinct almost surely in finite time. As a consequence, $N^{Stop} < \infty$ almost surely and the perfect simulation algorithm stops after a finite number of steps. This achieves the proof of the construction of the process.

□

2.6 Proof of Theorem 13 .

2.6.1 Some comments

We now give the proof of Theorem 13. As in the proof of Theorem 6 we use a Kalikow-type decomposition of the transition probabilities (2.4.24). We will use the same notations for objects with slightly different definitions but playing the same role in the proof, in order to avoid too heavy notations.

The main difference between the two models is that in the first one, no leakage term is present, whereas the second model contains a leakage term through the functions g_j. Moreover, in the first model, the presence of thresholds allows us to obtain a Kalikow-type decomposition according to space (and not to time) with probabilities $\mu_i(k)$ that are deterministic and do not depend on the realization of the Poisson Random Measure N. This is crucial because it gives us an independence argument in order to obtain statement (2.5.37). However, the statement of Theorem 6 is at the cost of two assumptions. The first one is the presence of spontaneous spikes due to condition (2.2.9). The second assumption is the existence of the thresholds $K_{j\to i}$.

In the second model, we will introduce a space-time decomposition in random environment with probabilities $\mu_{(i,t)}(k)$ that are $\sigma(N)$-measurable random variables. The independence argument leading to (2.5.37) will here be ensured by the condition (2.3.19) that we impose on the structure of the neural network.

2.6.2 Kalikow-type decomposition

In this section, the Kalikow-type decomposition will take place in a random environment depending on the realization of the Poisson Random Measure N. We will consider space-time neighborhoods $\mathbb{V}_t^i(k) := V_i(k) \times [t-k, t[$.

We work with the state space $X = \{0,1\}^{\mathcal{G}}$ where $\mathcal{G}$ is the time grid $\{(i, T_n^i), (i,n) \in I \times \mathbb{Z}\}$. Note that each element x of X can be interpreted as a discrete measure dominated by dN. We fix a time $t = T_n^i$ and we introduce the following notations

$$r_{(i,t)}^{[0]}(1) = \inf_{x\in X} \phi_i\left(\sum_j W_{j\to i} \int_{L_t^i(x)}^t g_j(t-s)dx_j(s)\right), \tag{2.6.40}$$

where $dx_j = \sum_{m\in\mathbb{Z}} x\left(j, T_m^j\right) d\delta_{T_m^j}$, $L_t^i(x) = \sup\{T_m^i < t : x(i, T_m^i) = 1\}$, and

$$r_{(i,t)}^{[0]}(0) = \inf_{x\in\mathcal{S}} \left(1 - \phi_i\left(\sum_j W_{j\to i} \int_{L_t^i(x)}^t g_j(t-s)dx_j(s)\right)\right). \tag{2.6.41}$$

Now fix $x \in X$ and define for each $k \geq 1$ the set $D_{(i,t)}^k(x)$ by $D_{(i,t)}^k(x) := \{z \in \mathcal{S} : z\left(\mathbb{V}_t^i(k)\right) = x\left(\mathbb{V}_t^i(k)\right)\}$ and put

$$r_{(i,t)}^{[k]}(1|x) = \inf_{z\in D_{(i,t)}^k(x)} \phi_i\left(\sum_j W_{j\to i} \int_{L_t^i(z)}^t g_j(t-s)dz_j(s)\right), \tag{2.6.42}$$

$$r_{(i,t)}^{[k]}(0|x) = \inf_{z\in D_{(i,t)}^k(x)} \left(1 - \phi_i\left(\sum_j W_{j\to i} \int_{L_t^i(z)}^t g_j(t-s)dz_j(s)\right)\right). \tag{2.6.43}$$

We define $\alpha_{(i,t)}(0) := \mu_{(i,t)}(0) := r_{(i,t)}^{[0]}(1)+r_{(i,t)}^{[0]}(0)$, $\alpha_{(i,t)}^{[k]} := \inf_{x\in\mathcal{S}}\left(r_{(i,t)}^{[k]}(1|x) + r_{(i,t)}^{[k]}(0|x)\right)$, and finally, for all $k \geq 1$ $\mu_{(i,t)}(k) := \alpha_{(i,t)}^{[k]} - \alpha_{(i,t)}^{[k-1]}$. These definitions are similar to those of Section 4.1 and play the same role in the proof. Note that in this section we have a dependence on time denoted by the index t.

Lemma 24. *$\left(\mu_{(i,t)}(k)\right)_{k\geq 0}$ defines a probability on $\mathbb{N}$.*

Proof. The proof is similar to the proof of Lemma 20. □

Remark 25. *Instead of condition (2.3.22), we can work with the following condition:*

$$\lim_{k\to+\infty} \sum_{j\in V_i(k)} |W_{j\to i}|\Lambda_j \int_k^{+\infty} |g_j(s)|\, ds + \sum_{j\notin V_i(k)} |W_{j\to i}|\Lambda_j \int_0^{+\infty} |g_j(s)|\, ds = 0, \tag{2.6.44}$$

which is a necessary but not sufficient condition for condition (2.3.22).

Using these probabilities $\left(\mu_{(i,t)}(k)\right)_{k\geq 0}$, we obtain a Kalikow-type decomposition for

$$p_{(i,t)}(1|x) = \phi_i\left(\sum_j W_{j\to i} \int_{L_t^i(x)}^t g_j(t-s)dx_j(s)\right).$$

Recall that $\mathcal{F}_t$ is the sigma-field generated by $Z^i(]s,u]), s \leq u \leq t, i \in I$, and that we work on the probability space $(\Omega, \mathcal{A}, \mathbb{F})$ defined in Section 2.1.2.

Proposition 26. *Assume that conditions (2.3.21) and (2.6.44) are satisfied. Then there exists a family of conditional probabilities $\left(p_{(i,t)}^{[k]}(a|x)\right)_{k\geq 0}$ satisfying the following properties.*

1. *For all $a \in \{0,1\}, p_{(i,t)}^{[0]}(a|x) := \frac{r_{(i,t)}^{[0]}(a)}{\mu_{(i,t)}(0)}$ does not depend on the configuration x.*
2. *For all $a \in \{0,1\}$, $k \geq 1$, $\mathcal{S} \ni x \mapsto p_{(i,t)}^{[k]}(a|x)$ depends only on the variables $\left(x(j, T_n^j) : (j, T_n^j) \in \mathbb{V}_t^i(k)\right)$.*

3. *For all* $x \in \mathcal{S}$, $k \geq 1$, $p^{[k]}_{(i,t)}(1|x) \in [0,1]$, $p^{[k]}_{(i,t)}(1|x) + p^{[k]}_{(i,t)}(0|x) = 1$.

4. *For all* $a \in \{0,1\}, x \in \mathcal{S}, p^{[k]}_{(i,t)}(a|x)$ *and* $\mu_{(i,t)}(k)$ *are* $\mathcal{F}_t -$ *measurable random variables.*

5. *For all* $x \in \mathcal{S}$, *we have the following convex decomposition*

$$p_{(i,t)}(a|x) = \sum_{k \geq 0} \mu_{(i,t)}(k) p^{[k]}_{(i,t)}(a|x), \tag{2.6.45}$$

Proof. The proof follows the lines of the proof of Proposition 21. □

2.6.3 Perfect simulation

The idea for the perfect simulation algorithm is the same as in Section 2.5.2, but here we use a decomposition in space and time. We work conditionally on the realization of the Poisson Random Measure N and consider, for each $k \geq 1$, the space time neighborhood $\mathbb{V}^i_t(k)$. Define the clan of ancestors of element (i,t) of size k by

$$C_1^{(i,t)}(k) := \mathcal{G} \cap \mathbb{V}^i_t(k),$$

where $\mathcal{G}$ is the time grid $\{(i, T^i_n), (i,n) \in I \times \mathbb{Z}\}$ and where by convention $C_1^{(i,t)}(0) := \emptyset$. We choose as before i.i.d. random variables $K^{(i,t)} \in \mathbb{N}$ which are attached to each site $(i,t) \in \mathcal{G}$, chosen according to

$$\mathbb{P}(K^{(i,t)} = k) = \mu_{(i,t)}(k), \text{ for all } k \geq 0.$$

These random variables allow us to define the clans of ancestors $(C_n^{(i,t)})_n \subset I \times]-\infty, t[$ as follows. We put $C_1^{(i,t)} := C_1^{(i,t)}\left(K^{(i,t)}\right)$ and

$$C_n^{(i,t)} := \left(\bigcup_{(j,s) \in C_{n-1}^{(i,t)}} C_1^{(j,s)} \right) \setminus \left(C_1^{(i,t)} \cup \ldots \cup C_{n-1}^{(i,t)} \right).$$

As before, we have to prove that the process $|C_n^{(i,t)}|$ converges almost surely to 0 as n tends to $+\infty$. For this sake, we compare the process $|C_n^{(i,t)}|$ with a branching process of reproduction mean (depending on space and time)

$$M^{(i,t)} := \sum_{k \geq 1} \left| C_1^{(i,t)}(k) \right| \mu_{(i,t)}(k).$$

Recall that $\mathbb{E}^N$ denotes the conditional expectation with respect to N, clearly,

$$\mathbb{E}^N(|C_n^{(i,t)}||C_{n-1}^{(i,t)}) \leq \sum_{(j,s) \in C_{n-1}^{(i,t)}} M^{(j,s)}. \tag{2.6.46}$$

Now, the structure that we imposed on the neural network in (2.3.19) implies that for neurons $i \in I_l$ and $j \in I_m$ the event $\{(j,s) \in C_n^{(i,t)}\}$ is empty if $l \neq m+n$. Moreover, this events depends only of realizations for neurons in layers I_p with $m \leq p < l$ whereas

$M^{(j,s)}$ depends on realizations for neurons in the layer I_{m-1}. Consequently $1_{(j,s)\in C_n^{(i,t)}}$ is independent of $M^{(j,s)}$, which allows us to write

$$\mathbb{E}\left(\sum_{(j,s)} 1_{(j,s)\in C_n^{(i,t)}} M^{(j,s)}\right) = \sum_{(j,s)} \mathbb{E}(1_{(j,s)\in C_n^{(i,t)}})\mathbb{E}(M^{(j,s)}). \tag{2.6.47}$$

The next step in the proof is to show that

$$\sup_{i\in I} \mathbb{E}^N(M^{(i,t)}) < 1.$$

As in Section 2.5.2, we start with a proposition that gives an upper bound for the probabilities $\left(\mu_{(i,t)}(k)\right)_{k\geq 1}$.

Proposition 27.

$$\mu_{(i,t)}(k) \leq \gamma\left(\sum_{j\in V_i(k-1)} |W_{j\to i}| \int_{t-k}^{t-k+1} g_j(t-s)dN_s^j + \sum_{j\in \partial V_i(k-1)} |W_{j\to i}| \int_{t-k}^{t} g_j(t-s)dN_s^j\right). \tag{2.6.48}$$

Proof. Using the definition of $\mu_{(i,t)}(k)$,

$$\mu_{(i,t)}(k) = \inf_{x\in\mathcal{S}}\left(r_{(i,t)}^{[k]}(1|x) + r_{(i,t)}^{[k]}(0|x)\right) - \inf_{x\in\mathcal{S}}\left(r_{(i,t)}^{[k-1]}(1|x) + r_{(i,t)}^{[k-1]}(0|x)\right).$$

Fix $\varepsilon > 0$ and let $u \in \mathcal{S}$ be such that

$$r_{(i,t)}^{[k-1]}(1|u) + r_{(i,t)}^{[k-1]}(0|u) \leq \inf_{x\in\mathcal{S}}\left(r_{(i,t)}^{[k-1]}(1|x) + r_{(i,t)}^{[k-1]}(0|x)\right) + \varepsilon,$$

then

$$\mu_{(i,t)}(k) \leq \left(r_{(i,t)}^{[k]}(1|u) + r_{(i,t)}^{[k]}(0|u)\right) - \left(r_{(i,t)}^{[k-1]}(1|u) + r_{(i,t)}^{[k-1]}(0|u)\right) + \varepsilon.$$

Here we can assume, without loss of generality that $L_t^i(u) = -\infty$. Indeed if $L_t^i(u) > -\infty$, let u' be such that $L_t^i(u') = -\infty, u'\left(]-\infty; L_t^i(u)]\right) = 0$ and $u'\left(]L_t^i(u); t]\right) = u\left(]L_t^i(u); t]\right)$, then u' and u are two equivalent configurations in terms of acceptance/rejection decision of the site (i,t). Then

$$\begin{aligned}
&\mu_{(i,t)}(k) - \varepsilon \\
&\leq \inf_{z\in D_{(i,t)}^k(u)} \phi_i\left(\sum_j W_{j\to i}\int_{L_t^i(z)}^t g_j(t-s)dz_j(s)\right) - \inf_{z\in D_{(i,t)}^{k-1}(u)} \phi_i\left(\sum_j W_{j\to i}\int_{L_t^i(z)}^t g_j(t-s)dz_j(s)\right) \\
&+ \sup_{z\in D_{(i,t)}^{k-1}(u)} \phi_i\left(\sum_j W_{j\to i}\int_{L_t^i(z)}^t g_j(t-s)dz_j(s)\right) - \sup_{z\in D_{(i,t)}^{k}(u)} \phi_i\left(\sum_j W_{j\to i}\int_{L_t^i(z)}^t g_j(t-s)dz_j(s)\right)
\end{aligned}$$

Using condition (2.1.4) and the fact that ϕ_i is non-decreasing, we obtain

$$\begin{aligned}
&\mu_{(i,t)}(k) - \varepsilon \\
&\leq \gamma\left(\inf_{z\in D_{(i,t)}^k(u)} \sum_j W_{j\to i}\int_{L_t^i(z)}^t g_j(t-s)dz_j(s) - \inf_{z\in D_{(i,t)}^{k-1}(u)} \sum_j W_{j\to i}\int_{L_t^i(z)}^t g_j(t-s)dz_j(s)\right) \\
&+\gamma\left(\sup_{z\in D_{(i,t)}^{k-1}(u)} \sum_j W_{j\to i}\int_{L_t^i(z)}^t g_j(t-s)dz_j(s) - \sup_{z\in D_{(i,t)}^{k}(u)} \sum_j W_{j\to i}\int_{L_t^i(z)}^t g_j(t-s)dz_j(s)\right)
\end{aligned}$$

Now we will simplify this expression detailing the configurations that realize the extrema. In order to reach a lower-bound, we have to fix z such that $L_t^i(z) = -\infty$ (we can do this since $L_t^i(u) = -\infty$) and whenever we have the choice for z we also have to fix $z = 1$ if the corresponding $W_{j\to i}$ is negative, else we have to fix $z = 0$. We do the opposite choice in order to reach an upper-bound. After factorization, we obtain the announced upper-bound (2.6.48)

$$\mu_{(i,t)}(k) \leq \gamma \left(\sum_{j\in V_i(k-1)} |W_{j\to i}| \int_{t-k}^{t-k+1} g_j(t-s)dN_s^j + \sum_{j\in \partial V_i(k-1)} |W_{j\to i}| \int_{t-k}^{t} g_j(\bar{t}-s)dN_s^j \right),$$

thanks to the simplification $\sum_j W_{j\to i}1_{W_{j\to i}>0} - \sum_j W_{j\to i}1_{W_{j\to i}<0} = \sum_j |W_{j\to i}|$. □

Now, we will use the stationarity of the PRM ; this will allow us to omit the dependence in time (by fixing $t = 0$) since we are only interested in the expectation. We have the following first upper-bound for $M^{(i,t)}$ thanks to the inequality (2.6.48)

$$\frac{M^{(i,t)}}{\gamma} \leq \sum_{k\geq 1} \left|C_{(i,t)}(k)\right| \left(\sum_{j\in V_i(k-1)} |W_{j\to i}| \int_{k-1}^{k} g_j(s)dN_j(s) \; + \sum_{j\in \partial V_i(k-1)} |W_{j\to i}| \int_{0}^{k} g_j(s)dN_j(s) \right). \tag{2.6.49}$$

The difficulty to calculate the expectation of this upper-bound is that N_s^j is present in each term of the product so that these terms are not independent. However, we have independence whenever the indexes denoting the neurons are different or when the intervals of time that we consider are disjoint. We will therefore decompose the sums in the previous expression in order to isolate the products of non-independent terms. Then we calculate separately the expectations of these terms. For example, we have:

$$\mathbb{E}\left[\left(\int_0^k dN_s^j \right) \left(\int_0^k g_j(s)dN_s^j \right) \right] = \Lambda_j \left(k\Lambda_j + 1\right) \left(\int_0^k g_j(s)ds \right). \tag{2.6.50}$$

This result allows to decompose the above expectation as the sum of the covariance and the product of expectations. We can make such a decomposition for each product of non independent terms in (2.6.49). Consequently, the expectation of the upper bound in (2.6.49) can be written as the sum of the covariances of the non-independent terms and the products of expectations of all the terms. After factorization, we finally obtain:

$$\frac{\mathbb{E}\left(M^{(i,t)}\right)}{\gamma} \leq \sum_{k\geq 1} \left[\left(k \left(\sum_{j\in V_i(k)} \Lambda_j \right) + 1 \right) \left(\sum_{j\in V_i(k-1)} |W_{j\to i}|\Lambda_j \int_{k-1}^{k} g_j(s)ds \right.\right.$$
$$\left.\left. + \sum_{j\in \partial V_i(k-1)} |W_{j\to i}|\Lambda_j \int_{0}^{k} g_j(s)ds \right) \right]. \tag{2.6.51}$$

Now we can use the assumption (2.3.22) of Theorem 13 in order to deduce that $M := \sup_i \mathbb{E}\left(M^{(i,t)}\right) < 1$. Consequently, we have

$$\mathbb{E}(|C_n^{(i,t)}|) \leq M^n \to 0 \text{ when } n \to \infty.$$

This ensures that the process $\left(C_n^{(i,t)}\right)_{n\in\mathbb{N}}$ goes extinct a.s., or in other words that $N^{Stop} < \infty$ a.s. Consequently the perfect simulation algorithm ends in a finite time and this achieves the proof of the construction of the process.

□

Remark 28. *As already mentioned in Remark 14 (i), Theorem 13 remains true under the weaker Assumption 4 imposing only equi-continuity instead of Lipschitz continuity imposed in Assumption 12. It is easy to adapt the proofs of Lemma 24 and Propositions 26 and 27 working with Assumption 4. However, the form of the upper-bound in Proposition 27, under Assumption 12, allows to compute an upper bound for* $E\left(M^{(i,t)}\right)$ *explicitly in terms of the parameters, as explained in formula (2.6.50). This is not possible imposing only Assumption 4.*

Acknowledgments

We thank an anonymous referee for helpful comments and suggestions. Moreover we thank A. Galves and P.A. Ferrari for many nice discussions on the subject.

This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01) and as part of the activities of FAPESP Research, Dissemination and Innovation Center for Neuromathematics (grant 2013/07699-0, S. Paulo Research Foundation).

Le chapitre suivant a fait l'objet d'une soumission et est accepté pour parution dans *Statistical Inference for Stochastic Processes.*

Chapter 3

Non-parametric estimation of the spiking rate in systems of interacting neurons.

abstract

We consider a model of interacting neurons where the membrane potentials of the neurons are described by a multidimensional piecewise deterministic Markov process (PDMP) with values in $\mathbb{R}^N$, where N is the number of neurons in the network. A deterministic drift attracts each neuron's membrane potential to an equilibrium potential m. When a neuron jumps, its membrane potential is reset to 0, while the other neurons receive an additional amount of potential $\frac{1}{N}$. We are interested in the estimation of the jump (or spiking) rate of a single neuron based on an observation of the membrane potentials of the N neurons up to time t. We study a Nadaraya-Watson type kernel estimator for the jump rate and establish its rate of convergence in L^2. This rate of convergence is shown to be optimal for a given Hölder class of jump rate functions. We also obtain a central limit theorem for the error of estimation. The main probabilistic tools are the uniform ergodicity of the process and a fine study of the invariant measure of a single neuron.

3.1 Introduction

This paper is devoted to the statistical study of certain Piecewise Deterministic Markov Processes (PDMP) modeling the activity of a biological neural network. More precisely, we are interested in estimating the the underlying jump rate of the process, *i.e.* the spiking rate function of each single neuron.

Piecewise Deterministic Markov Processes (PDMP's) have been introduced by Davis ([Dav84] and [Dav93]) as a family of càdlàg Markov processes following a deterministic drift with random jumps. PDMP's are widely used in probabilistic modeling of *e.g.* biological or chemical phenomena (see *e.g.* [CDMR12] or [PTW10], see [ABG$^+$14] for an overview). In the present paper, we study the particular case of PDMP's which are systems of interacting neurons. Building a model for the activity of a neural network that can fit biological considerations is crucial in order to understand the mechanics of the brain. Many papers in the literature use Hawkes Processes in order to describe the spatio-temporal dependencies which are typical for huge systems of interacting neurons, see [GL13], [HRBR15] and [**?**] for example. Our model can be interpreted as Hawkes

process with memory of variable length (see [GL16]); it is close to the model presented in [DO14]. It is of crucial interest for modern neuro-mathematics to be able to statistically identify the basic parameters defining the dynamics of a model for neural networks. The most relevant mechanisms to study are the way the neurons are connected to each other and the way that a neuron deals with the information it receives. In [DGLO16] and in [HRBR15], the authors build an estimator for the interaction graph, in discrete or in continuous time. In the present work, we assume that we observe a subsystem of neurons which are all interconnected and behaving in a similar way. We then focus on the estimation of the firing rate of a neuron within this system. This rate depends on the membrane potential of the neuron, influenced by the activity of the other neurons.

More precisely, we consider a process $X_t = (X_t^1, ..., X_t^N)$, where N is the number of neurons in the network and where each variable X_t^i represents the membrane potential of neuron i, for $1 \leq i \leq N$. Each membrane potential X_t^i takes values in a compact interval $[0, K]$, where 0 is interpreted as resting potential (corresponding to $\sim -90mV$ in real neurons) and where $K \sim 140mV$ (see *e.g.* [**?**]). This process has the following dynamic. A deterministic drift attracts the membrane potential of each neuron to an equilibrium potential $m \in \mathbb{R}_+$ with an exponential speed of parameter $\lambda \in \mathbb{R}_+$. Moreover, a neuron with membrane potential x "fires" (*i.e.*, jumps) with intensity $f(x)$, where $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a given intensity function. When a neuron fires, its membrane potential is reset to 0, interpreted as resting potential, while the membrane potentials of the other neurons are increased by $\frac{1}{N}$ until they reach the maximal potential height K.

The goal of this paper is to explore the statistical complexity of the model described above in a non-parametric setting. We aim at giving precise statistical characteristics (such as optimal rates of convergence, estimation procedures) such that we are able to compare systems of interacting neurons to benchmark non-parametric models like density estimation or nonlinear regression. More precisely, given the continuous observation [1] of the system of interacting neurons over a time interval $[0, t]$ (with asymptotics being taken as $t \to \infty$), we infer on the different parameters of the model which are: the equilibrium potential m, the speed of attraction λ and the spiking rate function f. Since in a continuous time setting, the coefficients λ and m are known (they can be identified by any observation of the continuous trajectory of a neuron's potential between two successive jumps), the *typical* problem is the estimation of the unknown spiking rate $f(\cdot)$.

Therefore we restrict our attention to the estimation of the unknown spiking rate $f(\cdot)$. We measure smoothness of the spiking rate by considering Hölder classes of possible shapes for the spiking rate and suppose that the spiking rate has smoothness of order β in a Hölder sense. To estimate the jump rate f in a position a, we propose a Nadaraya-Watson type kernel estimator which is roughly speaking of the form

$$\hat{f}_t(a) = \frac{\sharp \text{ spikes in positions in } B_h(a) \text{ during } [0,t]}{\text{occupation time of } B_h(a) \text{ during } [0,t]},$$

where $B_h(a)$ is a neighborhood of size h of the position a where we estimate the jump rate function f. A rigorous definition of the estimator is given in terms of the jump measure and an occupation time measure of the process X. The convergence of the estimator is

[1] A short remark concerning the continuous time observation scheme : Presumably, if we deal with discrete time samples, observed at sufficiently high frequency such that with huge probability at most one jump can take place during one sampling step, it would be possible to reconstruct the continuous trajectory of the process with hight probability and to perform our estimation procedure also in this frame.

implied by the fact that the compensator of the jump measure is the occupation time measure integrated against the jump rate function f, together with uniform ergodicity of the process. Assuming that the jump rate function f has smoothness of order β in a Hölder sense, we obtain the classical rate of convergence of order $t^{-\frac{\beta}{2\beta+1}}$ for the point-wise L^2-error of the estimator. This rate is shown to be optimal. We also state two important probabilistic tools that are needed in order to obtain the statistical results. The first one is the uniform positive Harris recurrence of process. The second one is the existence of a regular density function of the invariant measure of a single neuron.

In the literature, non-parametric estimation for PDMP's has already been studied, see for example [ADGP14] and, more particularly concerning the estimation of the jump rate, [AMG15]. On the contrary to these studies, the framework of the present work is more difficult for two reasons. The first reason is the fact that our process is multidimensional, presenting real interactions between the neurons. Of course, estimation problems for multidimensional PDMP's have already been studied. However, in all cases we are aware of, a so-called "many-to-one formula" (see [Kre14], see also [HO]) allows to express the occupation time measure of the whole system in terms of a single "typical" particle. This is not the case in the present paper – and it is for this reason that we have to work under the relatively strong condition of uniform ergodicity which is implied by compact state space – a condition which is biologically meaningful. The second, more important, reason is the fact that the transition kernel associated to jumps is degenerate. This is why the construction of our estimator is different from other constructions in previous studies. The degeneracy of the transition kernel also leads to real difficulties in the study of the regularity of the invariant density of a single neuron, see [Löc16] and the discussions therein.

In Section 3.2, we describe more precisely our model and state our main results. We first provide two probabilistic results necessary to prove the convergence of the estimator: firstly, the positive Harris recurrence of the process X in Theorem 30 and secondly the properties of the invariant measure in Theorem 31. The speed of convergence of our estimator is established in Theorem 32. Finally, Theorem 33 states that our speed of convergence is optimal for the point-wise L^2-error, uniformly in f. The key tool to prove this optimality is to study the asymptotic properties of the likelihood process for a small perturbation of the function f close to a.

The proofs of Theorems 30,32 and 33 are respectively given in Sections 3.3, 3.4 and 3.5. We refer the reader to [Löc16] for a proof of Theorem 31.

3.2 The model

3.2.1 The dynamics

Let $N > 1$ be fixed and $(N^i(ds, dz))_{i=1,\dots,N}$ be a family of *i.i.d.* Poisson random measures on $\mathbb{R}_+ \times \mathbb{R}_+$ having intensity measure $dsdz$. We study the Markov process $X_t = (X_t^1, \dots, X_t^N)$ taking values in $[0, K]^N$ and solving, for $i = 1, \dots, N$, for $t \geq 0$,

$$\begin{aligned} X_t^i &= X_0^i - \lambda \int_0^t (X_s^i - m)ds - \int_0^t \int_0^\infty X_{s-}^i 1_{\{z \leq f(X_{s-}^i)\}} N^i(ds, dz) \\ &\quad + \sum_{j \neq i} \int_0^t \int_0^\infty a_K(X_{s-}^i) 1_{\{z \leq f(X_{s-}^j)\}} N^j(ds, dz). \end{aligned} \tag{3.2.1}$$

In the above equation, $\lambda > 0$ is a positive number, m is the equilibrium potential value such that $0 < m < K$. Moreover, we will always assume that $K \geq \frac{2}{N}$. Finally, the functions $a_K : [0, K] \to [0, K]$ and $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$ satisfy (at least) the following assumption.

Assumption 29.
1. $a_K : [0, K] \to [0, \frac{1}{N}]$ *is non-increasing and smooth,* $a_K(x) = \frac{1}{N}$*, for all* $x < K - \frac{2}{N}$ *and* $a_K(x) < K - x$ *for all* $x \geq K - \frac{2}{N}$*.*
2. $f \in C^1(\mathbb{R}_+)$*,* f *is non-decreasing,* $f(0) = 0$*, and there exists* $f_{min} : \mathbb{R}_+ \mapsto \mathbb{R}_+$*, non-decreasing, such that* $f(x) \geq f_{min}(x) > 0$ *for all* $x > 0$*.*

All membrane potentials take values in $[0, K]$, where K is the maximal height of the membrane potential of a single neuron. 0 is interpreted as resting potential (corresponding to $\sim -90mV$ in real neurons) and $K \sim 140mV$ (see *e.g.* [**?**]). In (3.2.1), λ gives the speed of attraction of the potential value of each single neuron to an equilibrium value m. The function a_K denotes the increment of membrane potential received by a neuron when an other neuron fires. For neurons with membrane potential away from the bound K, this increment is equal to $\frac{1}{N}$. However, for neurons with membrane potential close to K, this increment may bring their membrane potential above the bound K. This is why we impose this dynamic close to the bound K.

In what follows, we are interested in the estimation of the intensity function f, assuming that the parameters K, f_{min} and a_K are known and that the function f belongs to a certain Hölder class of functions. The parameters of this class of functions are also supposed to be known. The assumption $f(0) = 0$ comes from biological considerations and expresses the fact that a neuron, once it has fired, has a refractory period during which it is not likely to fire.

The generator of the process X is given for any smooth test function $\varphi : [0, K]^N \to \mathbb{R}$ and $x \in [0, K]^N$ by

$$L\varphi(x) = \sum_{i=1}^{N} f(x_i) \left[\varphi(\Delta_i(x)) - \varphi(x)\right] - \lambda \sum_i \left(\frac{\partial \varphi}{\partial x_i}(x) \left[x_i - m\right] \right), \tag{3.2.2}$$

where

$$(\Delta_i(x))_j = \left\{ \begin{array}{ll} x_j + a_K(x_j) & j \neq i \\ 0 & j = i \end{array} \right\}. \tag{3.2.3}$$

The existence of a process X with such dynamics is ensured by an acceptance/rejection procedure that allows to construct solutions to (3.2.1) explicitly. More precisely, since each neuron spikes at maximal intensity $f(K)$, we can work conditionally on the realization of a Poisson process $\bar{N}$ with intensity $Nf(K)$. We construct the process X considering the jump times $\bar{T}_n$ of $\bar{N}$ as candidates for the jump times of X and accepting them with probability

$$\frac{\sum_{i=1}^{N} f\left(X^i_{\bar{T}_n -}\right)}{Nf(K)}.$$

It is then possible to construct a solution to (3.2.1) step by step, following the deterministic drift between the jump times of $\bar{N}$, and jumping according to this acceptance/rejection procedure. We refer the reader to Theorem 9.1 in chapter IV of [IW89] for a proof of the existence of the process $(X_t)_t$.

We denote by P_x the probability measure under which the solution $(X_t)_t$ of (3.2.1) starts from $X_0 = x \in [0, K]^N$. Moreover, $P_\nu = \int_{[0,K]^N} \nu(dx) P_x$ denotes the probability measure under which the process starts from $X_0 \sim \nu$. Figure 3.1 is an example of trajectory for $N = 5$ neurons, choosing $f = Id,\ \lambda = 1,\ m = 1$, and $K = 2$.

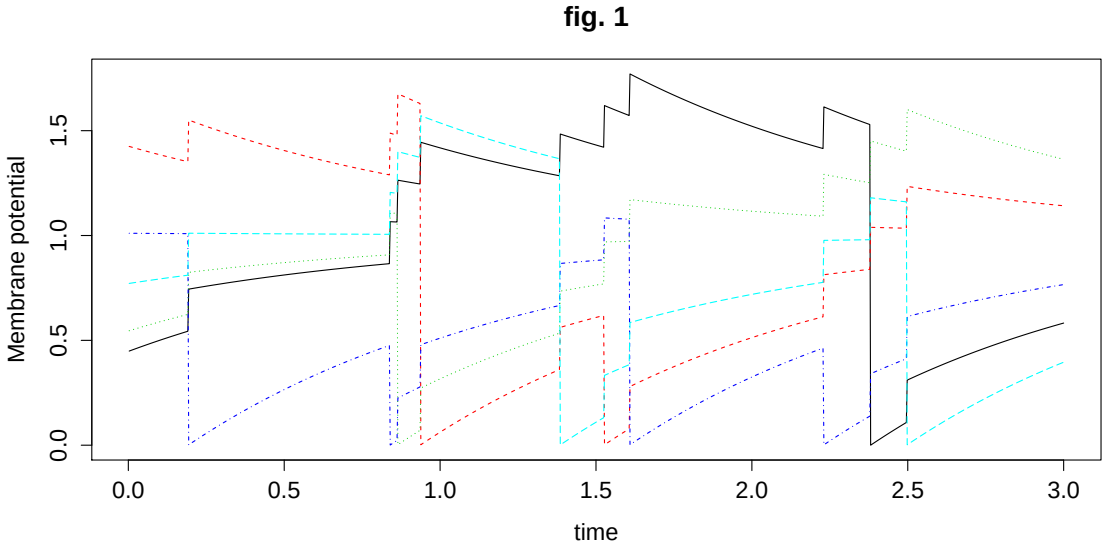


Figure 3.1: Trajectory of 5 neurons

The aim of this work is to estimate the unknown firing rate function f based on an observation of X continuously in time. Notice that for all $1 \leq i \leq N$, X^i reaches the value 0 only through jumps. Therefore, the following definition gives the successive spike times of the $i-$th neuron, $1 \leq i \leq N$. We put

$$T_0^i = 0, T_n^i = \inf\{t > T_{n-1}^i : X_{t-}^i > 0, X_t^i = 0\}, n \geq 1,$$

and introduce the jump measures

$$\mu^i(ds, dy) = \sum_{n \geq 1} 1_{\{T_n^i < \infty\}} \delta_{(T_n^i, X^i_{T_n^i -})}(dt, dy), \quad \mu(dt, dx) = \sum_{i=1}^N \mu^i(ds, dx).$$

By our assumptions, μ^i is compensated by $\hat{\mu}^i(ds, dy) = f(X_s^i) ds \delta_{X_s^i}(dy)$, and therefore the compensator $\hat{\mu}$ of μ is given by

$$\hat{\mu}(dt, dy) = f(y)\eta(dt, dy), \text{ where } \eta(A \times B) = \int_A \left(\sum_{i=1}^N 1_B(X_s^i) \right) ds$$

is the total occupation time measure of the process X.

We will also write $T_n, n \geq 0$, for the successive jump times of the process X, *i.e.*

$$T_0 = 0, T_n = \inf\{T_k^i : T_k^i > T_{n-1}, k \geq 1, 1 \leq i \leq N\}, n \geq 1.$$

For some kernel function Q such that

$$Q \in C_c(\mathbb{R}), \int_{\mathbb{R}} Q(y) dy = 1, \tag{3.2.4}$$

we define the kernel estimator for the unknown function f at a point a with bandwidth h, based on observation of X up to time t by

$$\hat{f}_{t,h}(a) = \frac{\int_0^t \int_{\mathbb{R}} Q_h(y-a)\mu(ds,dy)}{\int_0^t \int_{\mathbb{R}} Q_h(y-a)\eta(ds,dy)}, \text{ where } Q_h(y) := \frac{1}{h}Q\left(\frac{y}{h}\right) \text{ and } \frac{0}{0} := 0. \tag{3.2.5}$$

For h small, $\hat{f}_{t,h}(a)$ is a natural estimator for $f(a)$. Indeed, this expression as a ratio follows the intuitive idea to count the number of jumps that occurred with a position close to a and to divide by the occupation time of a neighborhood of a, which is natural to estimate an intensity function depending on the position a. More precisely, by the martingale convergence theorem, the numerator $\int_0^t \int_{\mathbb{R}} Q_h(y-a)\mu(ds,dy)$ should behave, for t large, as $\int_0^t \int_{\mathbb{R}} Q_h(y-a)f(y)\eta(ds,dy)$. But by the ergodic theorem,

$$\frac{\int_0^t \int_{\mathbb{R}} Q_h(y-a)f(y)\eta(ds,dy)}{\int_0^t \int_{\mathbb{R}} Q_h(y-a)\eta(ds,dy)} \to \frac{\pi_1(Q_h(\cdot-a)f)}{\pi_1(Q_h(\cdot-a))}$$

as $t \to \infty$, where π_1 is the stationary measure of each neuron X_t^i. Finally, if the invariant measure π_1 is sufficiently regular, then

$$\frac{\pi_1(Q_h(\cdot-a)f)}{\pi_1(Q_h(\cdot-a))} \to f(a)$$

as $h \to 0$.

We restrict our study to fixed Hölder classes of rate functions f. For that sake, we introduce the notation $\beta = k + \alpha$ for $k = \lfloor \beta \rfloor \in \mathbb{N}$ and $0 \leq \alpha < 1$. We consider the following Hölder class for arbitrary constants $F, L > 0$, and a function f_{min} as in Assumption 29.

$$\begin{aligned} H(\beta, F, L, f_{min}) = \{f \in C^k(\mathbb{R}_+) : |\frac{d^l}{dx^l} f(x)| \leq F, \text{ for all } 0 \leq l \leq k, x \in [0,K], \\ f(x) \geq f_{min}(x) \text{ for all } x \in [0,K], \ |f^{(k)}(x) - f^{(k)}(y)| \leq L|x-y|^\alpha \text{ for all } x, y \in [0,K]\}. \end{aligned} \tag{3.2.6}$$

3.2.2 Probabilistic results

In this Section, we collect important probabilistic results. We first establish that the process $(X_t)_{t \geq 0}$ is recurrent in the sense of Harris.

Theorem 30. *Grant Assumption 29. Then the process X is positive Harris recurrent having unique invariant probability measure π, i.e. for all $B \in \mathcal{B}([0,K]^N)$,*

$$\pi(B) > 0 \ \textit{ implies } P_x\left(\int_0^\infty 1_B(X_s)ds = \infty\right) = 1 \tag{3.2.7}$$

for all $x \in [0,K]^N$. Moreover, there exist constants $C > 0$ and $\kappa > 1$ which do only depend on the class $H(\beta, F, L, f_{min})$, but not on f, such that

$$\sup_{f \in H(\beta,F,L,f_{min})} \|P_t(x,\cdot) - \pi\|_{TV} \leq C\kappa^{-t}. \tag{3.2.8}$$

It is well-known that the behavior of a kernel estimator such as the one introduced in (3.2.5) depends heavily on the regularity properties of the invariant probability measure of the system. Our system is however very degenerate. Firstly, it is a piecewise deterministic Markov process (PDMP) in dimension N, with interactions between particles. Hence, no Brownian noise is present to smoothen things. Moreover, the transition kernels associated to the jumps of system (3.2.1) are highly degenerate (recall (3.2.3)). The transition kernel

$$K(x, dy) = \mathcal{L}(X_{T_1}|X_{T_1-} = x)(dy) = \sum_{i=1}^{N} \frac{f(x^i)}{\bar{f}(x)} \delta_{\Delta^i(x)}(dy)$$

with $\bar{f}(x) := \sum_{i=1}^{N} f(x^i)$ puts one particle (the one which is just spiking) to the level 0. As a consequence, the above transition does not create density – and it even destroys smoothness due to the reset to 0 of the spiking neuron. Finally, the only way that "smoothness" is generated by the process is the smoothness which is present in the "noise of the jump times" (which are basically of exponential density). For this reason, we have to stay away from the point $x = m$, where the drift of the flow vanishes. Moreover, the reset-to-0 of the spiking particles implies that we are not able to say anything about the behavior of the invariant density of a single particle in 0 (actually, near to 0) neither. Finally, we also have to stay strictly below the upper bound of the state space K. That is why we introduce the following open set $S_{d,\beta}$ given by

$$S_{d,\beta} := \{w \in [0, K] : \frac{\lfloor \beta \rfloor}{N} < w < K - \frac{\lfloor \beta \rfloor}{N}, |w - m| > d\}, \tag{3.2.9}$$

where β is the smoothness of the fixed class $H(\beta, F, L, f_{min})$ that we consider and where d is fixed such that $d > \frac{\lfloor \beta \rfloor + 2}{N}$. Notice that $S_{d,\beta}$ also depends on K, m and N which are supposed to be known. We are able to obtain a control of the invariant measure only on this set $S_{d,\beta}$. The dependence in β is due to the fact that the regularity of f is transmitted to the invariant measure by the means of successive integration by parts (see [Löc16] for more details).

We quote the following theorem from [Löc16].

Theorem 31. *(Theorem 5 of [Löc16])*
Suppose that $f \in H(\beta, F, L, f_{min})$. Let

$$\pi_1 := \mathcal{L}_\pi(X_t^1)$$

be the invariant measure of a single neuron, i.e. $\int g d\pi_1 = E_\pi(g(X_t^1))$. Then π_1 possesses a bounded continuous Lebesgue density π^1 on $S_{d,\beta}$ for any d such that $d > (\lfloor \beta \rfloor + 2)/N$, which is bounded on $S_{d,\beta}$, uniformly in $f \in H(\beta, F, L, f_{min})$. Moreover, $\pi^1 \in C^k(S_{d,\beta})$ and

$$\sup_{\ell \leq \lfloor \beta \rfloor, w \in S_{d,\beta}} |\pi_1^{(\ell)}(w)| + \sup_{w \neq w', w, w' \in S_{d,\beta}} \frac{\pi_1^{(\lfloor \beta \rfloor)}(w) - \pi_1^{(\lfloor \beta \rfloor)}(w')}{|w - w'|^\alpha} \leq C_F, \tag{3.2.10}$$

where the constant C_F depends on d and on the smoothness class $H(\beta, F, L, f_{min})$, but on nothing else.

3.2.3 Statistical results

We can now state the main theorem of our paper which describes the quality of our estimator in the minimax theory. We assume that m and λ are known and that f is the only parameter of interest of our model. We shall always write P_x^f and E_x^f in order to emphasize the dependence on the unknown f. Fix some $r > 0$ and some suitable point $a \in S_{d,\beta}$. For any possible rate of convergence $(r_t)_{t\geq 0}$ increasing to ∞ and for any process of $\mathcal{F}_t-$measurable estimators $\hat{f}_t$ we shall consider point-wise square risks of the type

$$\sup_{f\in H(\beta,F,L,f_{min})} r_t^2 E_x^f \left[|\hat{f}_t(a) - f(a)|^2 | A_{t,r}\right],$$

where

$$A_{t,r} := \left\{ \frac{1}{Nt} \int_0^t \int_{\mathbb{R}} Q_h(y-a)\eta(ds,dy) \geq r \right\}$$

is roughly the event ensuring that sufficiently many observations have been made near a, during the time interval $[0,t]$. We are able to choose r small enough such that

$$\liminf_{t\to\infty} \inf_{f\in H(\beta,F,L,f_{min})} P_x^f(A_{t,r}) = 1, \tag{3.2.11}$$

see Proposition 49 below.

Recall that the kernel Q is chosen to be of compact support. Let us write R for the diameter of the support of Q, therefore $Q(x) = 0$ if $|x| \geq R$. For any fixed $a \in S_{d,\beta}$, write $h_0 := h_0(a,R,\beta,d) := \sup\{h > 0 : B_{hR}(a) \subset S_{d/2,\beta}\}$. Here, $B_{hR}(a) = \{y \in \mathbb{R}_+ : |y-a| < hR\}$.

Theorem 32. *Let $f \in H(\beta,F,L,f_{min})$ and choose $Q \in C_c(\mathbb{R})$ such that $\int_{\mathbb{R}} Q(y)y^j dy = 0$ for all $1 \leq j \leq \lfloor\beta\rfloor$, and $\int_{\mathbb{R}} |y|^\beta Q(y)dy < \infty$. Then there exists $r^* > 0$ such that the following holds for any $a \in S_{d,\beta}, r \leq r^*$ and for any $h_t \leq h_0$.*

(i) For the kernel estimate (3.2.5) with bandwidth $h_t = t^{-\frac{1}{2\beta+1}}$, for all $x \in [0,K]$,

$$\limsup_{t\to\infty} \sup_{f\in H(\beta,F,L,f_{min})} t^{\frac{2\beta}{2\beta+1}} E_x^f \left[|\hat{f}_{t,h_t}(a) - f(a)|^2 | A_{t,r}\right] < \infty.$$

(ii) Moreover, for $h_t = o(t^{-1/(1+2\beta)})$, for every $f \in H(\beta,F,L,f_{min})$ and $a \in S_{d,\beta}$

$$\sqrt{th_t}\left(\hat{f}_{t,h_t}(a) - f(a)\right) \to \mathcal{N}(0,\Sigma(a))$$

weakly under P_x^f, where $\Sigma(a) = \frac{f(a)}{N\pi_1(a)} \int Q^2(y)dy$.

The next theorem shows that the rate of convergence achieved by the kernel estimate $\hat{f}_{t,t^{-1/(2\beta+1)}}$ is indeed optimal.

Theorem 33. *Let $a \in S_{d,\beta}$ and $x \in [0,K]$ be any starting point. Then we have*

$$\liminf_{t\to\infty} \inf_{\hat{f}_t} \sup_{f\in H(\beta,F,L,f_{min})} t^{\frac{2\beta}{1+2\beta}} E_x^f[|\hat{f}_t(a) - f(a)|^2] > 0, \tag{3.2.12}$$

where the infimum is taken over the class of all possible estimators $\hat{f}_t(a)$ of $f(a)$.

The proofs of Theorems 32 and 33 are given in Sections 3.4 and 3.5.

3.2.4 Simulation results

In this subsection, we present some results on simulations, for different jump rates f. The other parameters are fixed: $N = 100,\ \lambda = 1,\ K = 2$ and $m = 1$. The dynamics of the system are the same when λ and f have the same ratio. In other words, variations of λ and f keeping the same ratio between the two parameters lead to the same law for the process rescaled in time. This is why we fix $\lambda = 1$ and propose different choices for f. The kernel Q used here is a truncated Gaussian kernel with standard deviation 1.

We present for each choice of a jump rate function f the associated estimated function $\hat{f}$ and the observed distribution of X or more precisely of $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X^i$. Figures 2, 3 and 4 correspond respectively to the following definitions of f : $f(x) = x,\ f(x) = \log(x+1)$ and $f(x) = \exp(x) - 1$.

For Figures 3.2, 3.3 and 3.4, we fixed the length of the time interval for observations respectively to $t = 200$, 300 and 150. This allows us to obtain a similar number of jump for each simulation, respectively equal to 17324, 18579 and 21214. These simulations are realized with the software R.

The optimal bandwidth $h_t = t^{-\frac{1}{2\beta+1}}$ depends on the regularity of f given by the parameter β. Therefore, we propose a data-driven bandwidth chosen according to a Cross-validation procedure. For that sake, we define the sequence $(Z_k)_{k\in\mathbb{N}^*}$ by $Z_k^i = X^i_{T_k^i -}$ for all $1 \leq i \leq N$. For each $a \in [0, K]$ and each sample $Z = (Z_1, ..., Z_n)$, for $1 \leq \ell \leq n$ we define the random variable $\hat{\pi}_1^{\ell,n,h}(a)$ by

$$\hat{\pi}_1^{\ell,n,h}(a) = \frac{1}{(n-\ell)N} \sum_{k=\ell+1}^{n} \sum_{i=1}^{N} Q_h(Z_k^i - a).$$

$\hat{\pi}_1^{\ell,n,h}(a)$ can be seen as an estimator of the invariant measure π_1^Z of the discrete Markov chain.

We propose an adaptive estimation procedure at least for this simulation part. We use a Smoothed Cross-validation (SCV) to choose the bandwidth (see for example the paper of Hall, Marron and Park [**?**]), following ideas which were first published by Bowmann [**?**] and Rudemo [**?**]. As the bandwidth is mainly important for the estimation of the invariant probability π_1^Z, we use a Cross validation procedure for this estimation. More precisely, we use a first part of the trajectory to estimate $\widehat{\pi}_1^{\ell,n,h}$ and then another part of the trajectory to minimize the Cross validation $SCV(h)$ in h. In order to be closer to the stationary regime, we chose the two parts of the trajectory far from the starting time. Moreover we chose two parts of the trajectory sufficiently distant from each other. This is why we consider m_1, m_2 and ℓ such that $1 << m_1 \leq m_2 << \ell \leq n$.

We use the method of the least squares Cross validation and minimize

$$SCV(h) = \int \left(\widehat{\pi}_1^{\ell,n,h}(x)\right)^2 dx - \frac{2}{N(m_2 - m_1)} \sum_{k=m_1+1}^{m_2} \sum_{i=1}^{N} \widehat{\pi}_1^{\ell,n,h}(Z_k^i)$$

(where we have approximated the integral term by a Riemann approximation), giving rise to a minimizer $\hat{h}$. We then calculate the estimator $\hat{f}$ the long of the trajectory. In the next figure, we use this method to find the reconstructed f with an adaptive choice of h.

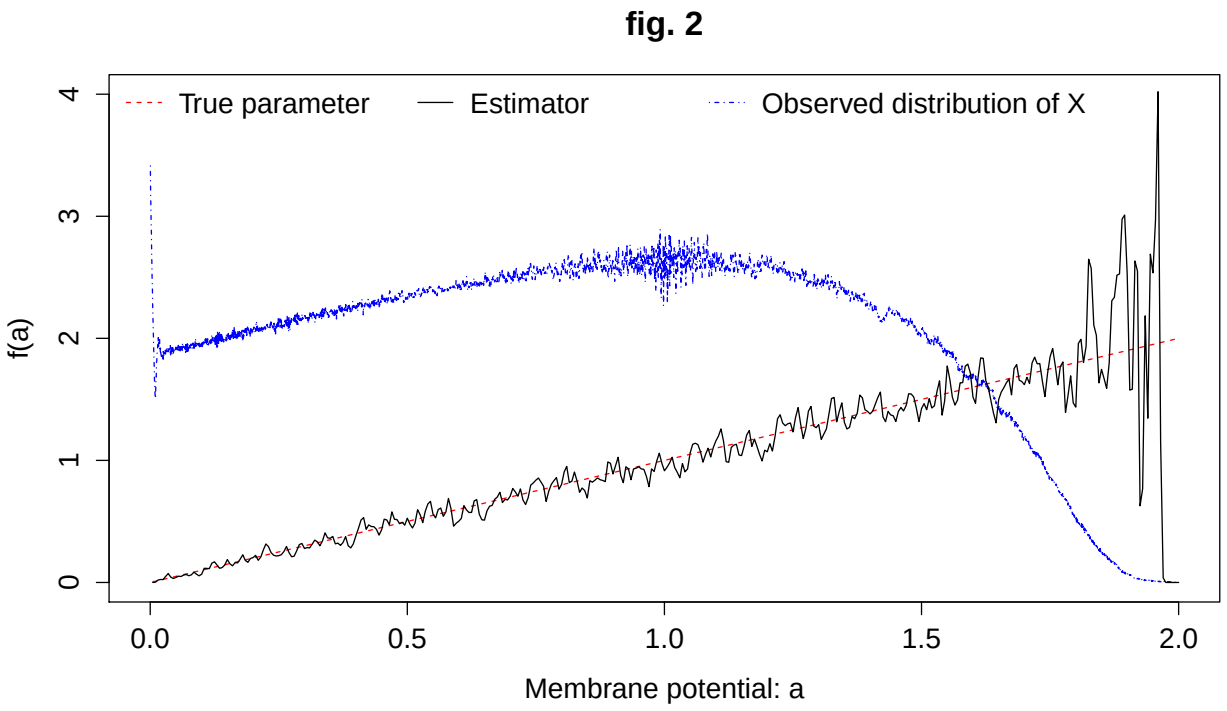


Figure 3.2: Estimation of the intensity function $f(x) = x$

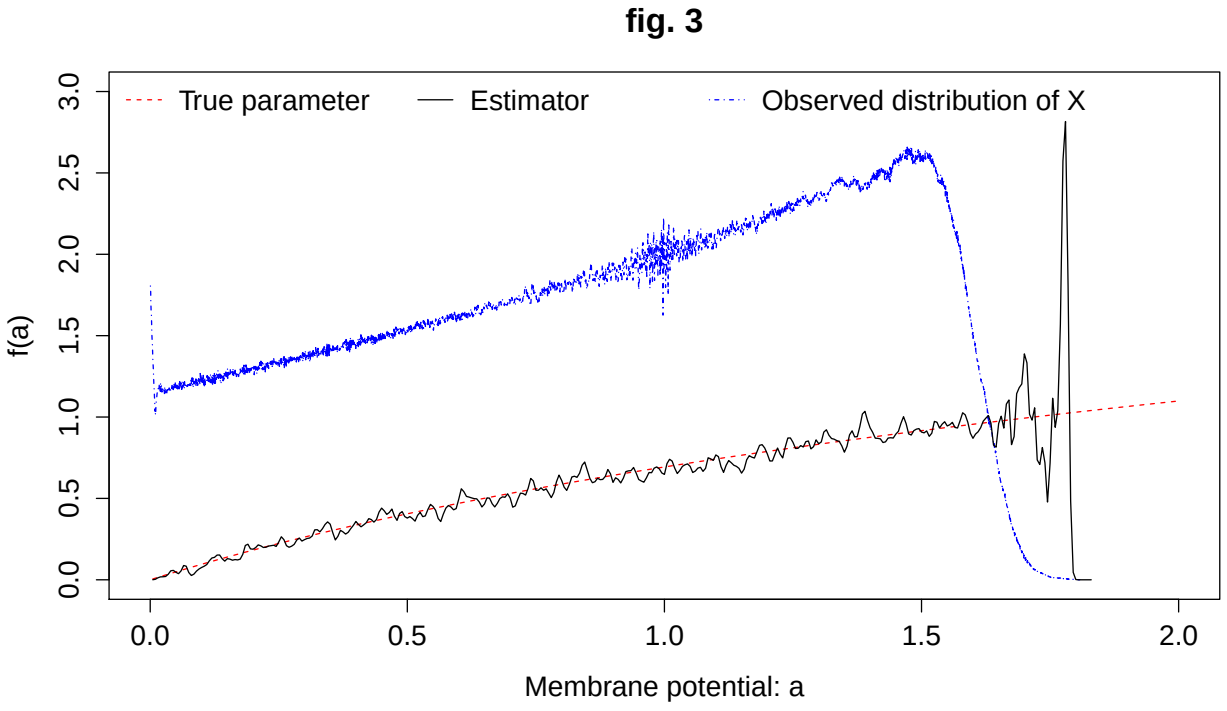


Figure 3.3: Estimation of the intensity function $f(x) = \log(x + 1)$

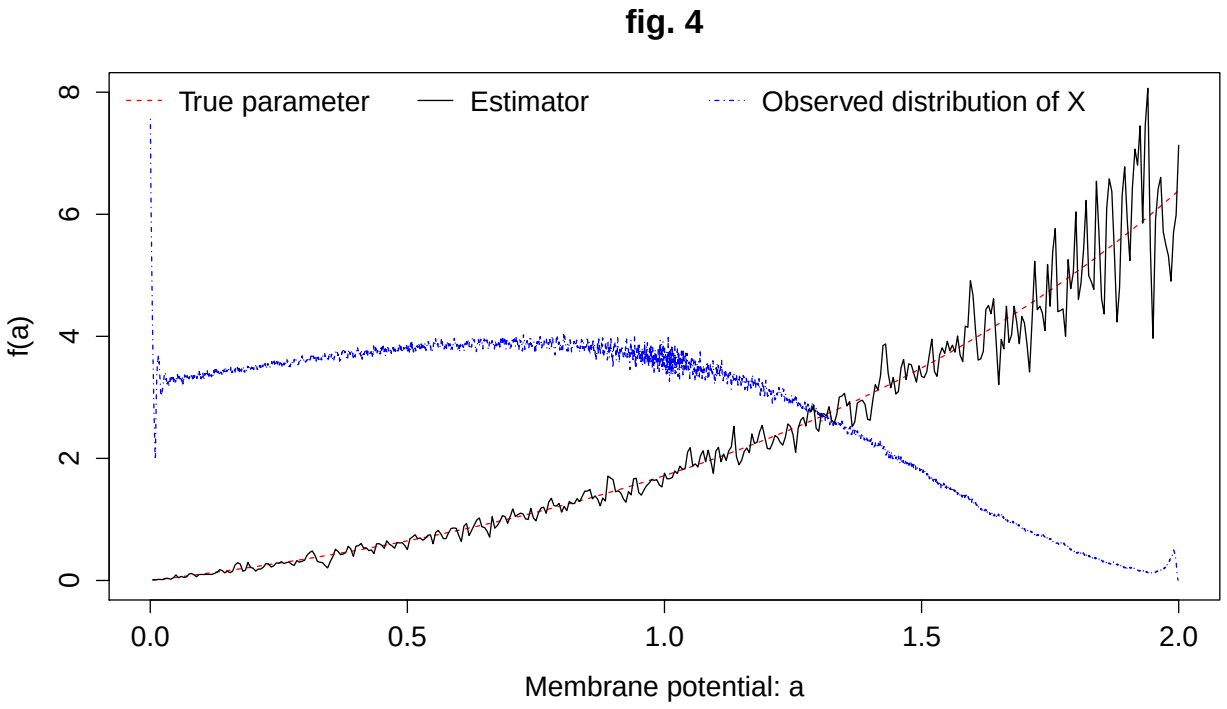


Figure 3.4: Estimation of the intensity function $f(x) = \exp(x) - 1$

As expected, we can see that the less observations we have, the worse is our estimator. Note that close to 0 the observed density of X explodes. This was also expectable due to the reset to 0 of the jumping neurons. Moreover, the simulations show a lack of regularity of the observed density close to m, which is consistent with our results, but this does not seem to affect the quality of the estimator.

3.3 Harris recurrence of X and speed of convergence to equilibrium – Proof of Theorem 30

In this section, we give the proof of Theorem 30 and show that the process $(X_t)_{t\geq 0}$ is positive recurrent in the sense of Harris. We follow a classical approach and prove the existence of regeneration times. This is done in the next subsection and follows ideas given in Duarte and Ost [DO14].

3.3.1 Regeneration

The main idea of proving a regeneration property of the process is to find some uniform "noise" for the whole process on some "good subsets" of the state space. Since the transition kernel associated to the jumps of our process is not creating any density (and actually destroys it for the spiking neurons which are reset to 0), the only source of noise is given by the random times of spiking. These random times are then transported through the deterministic flow $\gamma_{s,t}(v) = (\gamma_{s,t}(v^1), \dots, \gamma_{s,t}(v^N))$, which is given for any

starting configuration $v \in [0,K]^N$ by

$$\gamma_{s,t}(v^i) = e^{-\lambda(t-s)}v^i + (1-e^{-\lambda(t-s)})m,\ 0 \le s \le t,\ \gamma_t(v^i) := \gamma_{0,t}(v^i). \tag{3.3.13}$$

The key idea of what follows – which is entirely borrowed from [DO14] – is the following.

Write $I_n, n \ge 1$, for the sequence giving the index of the spiking neuron at time T_n, *i.e.* $I_n = i$ if and only if $T_n = T_k^i$ for some $k \ge 1$. It is clear that in order to produce an absolute continuous law with respect to the Lebesgue measure on $[0,K]^N$, we need at least N jumps of the process. On any event of the type $\{T_1 = t_1, I_1 = i_1, \dots, T_N = t_N, t_N < t < T_{N+1}, I_N = i_N\}$, it is possible to write the position of the process at time t as a concatenation of the deterministic flows given by

$$\Gamma_{(t_1,\dots,t_N,i_1,\dots,i_N)}(t,v) = \gamma_{t_N,t}(\Delta_{i_N}(\gamma_{t_{N-1},t_N}(\Delta_{i_{N-1}}(\dots\Delta_{i_1}(\gamma_{0,t_1}(v)))))). \tag{3.3.14}$$

Proving absolute continuity amounts to prove that the determinant of the Jacobian of the map $(t_1,\dots,t_N) \to \Gamma_{(t_1,\dots,t_N,i_1,\dots,i_N)}(t,v)$ does not vanish. For general sequences of $(i_1,\dots,i_N)$, this will not be true (think *e.g.* of the sequence $(i_1 = \dots = i_N = 1)$).

The main idea is however to consider the sequence $i_1 = 1, i_2 = 2, \dots, i_N = N$ and to use the *regeneration property of spiking*, *i.e.* the fact that the neuron k spiking at time t_k is reset to zero at time t_k. In this case, for all later times, its position does not depend on $t_1,\dots,t_{k-1}$ any more. In other words, the Jacobian of $\Gamma_{(t_1,\dots,t_N,1,\dots,N)}(t,v)$ is a diagonal matrix, and all we have to do is to control that all diagonal elements do not vanish. The second idea is to linearize the flow, *i.e.* to consider the flow during very short time durations, and to use that, just after spiking, each diagonal element is basically of the form

$$\frac{\partial \gamma_{s,t}(0)}{\partial s} \sim -\lambda m, \quad \text{as } t-s \to 0.$$

The important fact here is that the absolute value of the drift term of the deterministic flow of one neuron is strictly positive when starting from the initial value 0.

In the following, this idea is made rigorous. Our proof follows the approach given in Section 4 of [DO14]. We fix $\varepsilon > 0$ and put

$$A_\varepsilon = \{i\varepsilon - \varepsilon/4 < T_i < i\varepsilon, i = 1,\dots,N\}$$

and

$$S = \{I_1 = 1, I_2 = 2, \dots, I_N = N\}$$

which is the event that all N neurons have spiked in the fixed order given by their numbers, *i.e.* neuron 1 spikes first, then neuron 2, then 3, and so on. We introduce

$$u^* = \left(\frac{N-1}{N}, \frac{N-2}{N}, \dots, \frac{1}{N}, 0\right)$$

which would be the position of neurons after N spikes and on the event S, if $\lambda = 0$ (here, we suppose w.l.o.g. that $K > 1 + \frac{1}{N}$).

Now we fix any initial configuration $v \in [0,K]^N$ and introduce the sequence of configurations $v(k), 0 \le k \le N$, given by $v(0) = v$, $v_k(k) = 0$ and

$$v_i(k) = \left\{ \begin{array}{ll} \frac{k-i}{N}, & i < k \\ \underbrace{a_K \circ \dots \circ a_K}_{k \text{ times}}(v_i), & i > k \end{array} \right\}. \tag{3.3.15}$$

Notice that $\underbrace{a_K \circ \ldots \circ a_K}_{k \text{ times}}(v_i) = v_i + \frac{k}{N}$ if $v_i < K - \frac{2+k}{N}$. Notice also that $v(N) = u^*$.

We cite the following lemma from [DO14].

Lemma 34 (Lemma 4.1 of [DO14]). *If $X_0 = u \in B_\delta(v)$, then on the event $A_\varepsilon \cap S$, we have for all $1 \leq k \leq N$,*
(i) $X_i(T_k) = v_i(k) + \sum_{r=i+1}^{k} \lambda(T_r - T_{r-1}) d_i(r-1) + R_{\delta\varepsilon}(T_1^k, u) + R_{\varepsilon^2}(T_1^k, u)$, if $i < k$,

(ii) $X_i(T_k) = v_i(k) + \sum_{r=i}^{k} \lambda(T_r - T_{r-1}) d_i(r-1) + R_\delta(u) + R_{\delta\varepsilon}(T_1^k, u) + R_{\varepsilon^2}(T_1^k, u)$, if $i > k$,

(iii) $\bar{X}^N(T_k) = \bar{v}(k) + R_\delta(u) + R_\varepsilon(T_1^k, u)$, if $k < N$, and $\bar{X}^N(T_N) = \bar{u}^* + R_\varepsilon(T_1^N, u)$.
Here, $d_i(r) = m - v_i(r)$ and $T_1^k = (T_1, \ldots, T_k)$. Moreover, the remainder functions are of order

$$R_{\delta\varepsilon}(T_1^k, u) = O(\delta\varepsilon), R_{\varepsilon^2}(T_1^k, u) = O(\varepsilon^2), R_\delta(u) = O(\delta), \ldots,$$

and all partial derivatives are of order either δ or ε, uniformly in v.

Remark 35. *Our model is slightly different from the model in [DO14]: instead of an attraction to the empirical mean of the system, we have an attraction to a fixed equilibrium value m. This leads to our definition of $d_i(r)$ which is slightly different from the one used in [DO14].*

Corollary 36 ([DO14], Corollary 3). *Put $t^* = N\varepsilon$. Then we have on $A_\varepsilon \cap S$,*

$$X_i(t^*) = u_i^* + \lambda(t^* - T_N) d_i^* + \sum_{r=i+1}^{N} \lambda(T_r - T_{r-1}) d_i(r-1) + R_{\delta\varepsilon}(T_1^N, u) + R_{\varepsilon^2}(T_1^N, u), \tag{3.3.16}$$

where $d_i^* = m - u_i^*$.

We put as in [DO14] $\gamma^0(t_1^N) = (\gamma_1^0(t_1^N), \ldots, \gamma_N^0(t_1^N))$, where

$$\gamma_i^0(t_1^N) := u_i^* + \lambda(t^* - t_N) d_i^* + \sum_{r=i+1}^{N} \lambda(t_r - t_{r-1}) d_i(r-1), 1 \leq i \leq N.$$

Hence $\gamma_i^0(t_1^N)$ models how the N successive jump times $t_1 < t_2 < \ldots < t_N$ are mapped, through the deterministic flow, into a final position at time t^*- on the event $\{T_1 = t_1, \ldots, T_N = t_N\} \cap A_\varepsilon \cap S$. In order to control how the law of the N successive jump times $t_1, \ldots, t_N$ is transported through this flow, we calculate the partial derivatives of γ^0 with respect to $t_i, 1 \leq i \leq N$. One sees immediately that

$$\frac{\partial \gamma_i^0}{\partial t_k} = 0, k < i, \quad \frac{\partial \gamma_i^0}{\partial t_i} = -\lambda m, 1 \leq i \leq N,$$

whence

Corollary 37 (Corollary 4 of [DO14]). *For each $u \in B_\delta(v)$, the determinant of the Jacobian of the map $\{i\varepsilon - \varepsilon/4 < t_i < i\varepsilon, i = 1, \ldots, N\} \ni t_1^N \mapsto \gamma^0(t_1^N) + R_{\delta\varepsilon}(t_1^N, u) + R_{\varepsilon^2}(t_1^N, u)$ is given by*

$$\lambda^N m^N + R_\varepsilon(t_1^N, u) + R_\delta(t_1^N, u)$$

which is different from zero for ε and δ small enough, for all $u \in B_\delta(v)$.

As in Proposition 4.1 of [DO14], we now have two important conclusions from the above discussion.

Proposition 38. *There exists $\delta^* > 0$ and $\varepsilon > 0$, such that for $t^* = N\varepsilon$,*

$$P_{t^*}(x, \cdot) \geq \eta_1 1_{B_{\delta^*}(u^*)}(x)\nu, \tag{3.3.17}$$

where ν is a probability measure and $\eta_1 \in]0, 1[$.

The lower bound (3.3.17) is a local Doeblin condition, and its proof is given in Proposition 4.1 of [DO14]. We call $B_{\delta^*}(u^*)$ a regeneration set: if the process visits this regeneration set, then after a time t^* there is a probability η_1 that the law of the process is independent from its initial position $x \in B_{\delta^*}(u^*)$.

To be able to make use of the local Doeblin condition, we have to be sure that the process actually does visit the regeneration set $B_{\delta^*}(u^*)$. This is granted by the following result.

Proposition 39. *There exist $\varepsilon > 0$ and $\eta_2 > 0$ such that*

$$\inf_{f \in H(\beta, F, L, f_{min})} \inf_{v \in [0,K]^N} P_{t^*}(v, B_{\delta^*}(u^*)) \geq \eta_2,$$

for $t^* = N\varepsilon$.

Proof. By (3.3.16), there exists ε such that for all $v \in [0, K]^N$, we have that $X(t^*) \in B_{\delta^*}(u^*)$ on $A_\varepsilon \cap S$, when $X(0) = v$. Hence

$$P_{t^*}(v, B_{\delta^*}(u^*)) \geq P_v(A_\varepsilon \cap S).$$

Recalling (3.3.13), we then obtain

$$\begin{aligned} P_v(A_\varepsilon \cap S) = \int_{\varepsilon - \varepsilon/4}^{\varepsilon} & f(\gamma_{0,t_1}(v^1)) e^{-\int_0^{t_1} \bar{f}(\gamma_{0,s}(v))ds} dt_1 \\ & \int_{2\varepsilon - \varepsilon/4}^{2\varepsilon} f(\gamma_{t_1,t_2}(v(1)^2)) e^{-\int_{t_1}^{t_2} \bar{f}(\gamma_{t_1,s}(v(1)))ds} dt_2 \dots \\ & \int_{N\varepsilon - \varepsilon/4}^{N\varepsilon} f(\gamma_{t_{N-1},t_N}(v(N-1)^N)) e^{-\int_{t_{N-1}}^{t_N} \bar{f}(\gamma_{t_{N-1},s}(v(N-1)))ds} dt_N, \end{aligned}$$

where $\bar{f}(x) = \sum_{i=1}^N f(x^i)$, where the sequence $v(1), \dots, v(N-1)$ is given as in (3.3.15). Since by assumption $v \in [0, K]^N$, it is immediate to see that $\gamma_{s,t}(v(k)^i) \leq C$, for a constant C, for all $0 \leq s \leq t \leq t^*$, for all $k \leq N$ and for all $i \leq N$. Moreover,

$$\gamma^1_{0,t_1}(v^1) \geq (1 - e^{-\lambda \frac{3}{4}\varepsilon})m > 0, \text{ on } t_1 \geq \frac{3}{4}\varepsilon,$$

and since f is non decreasing, satisfying $f(x) \geq f_{min}(x)$ and $\|f\|_\infty \leq F$, this implies that

$$f(\gamma_{0,t_1}(v^1)) e^{-\int_0^{t_1} \bar{f}(\gamma_{0,s}(v))ds} \geq f((1 - e^{-\lambda \frac{3}{4}\varepsilon})m) e^{-Nt_1 f(C)} \geq f_{min}\left((1 - e^{-\lambda \frac{3}{4}\varepsilon})m\right) e^{-Nt_1 F},$$

on $t_1 \geq \frac{3}{4}\varepsilon$. Similar arguments show that all consecutive terms are strictly lower bounded uniformly in $f \in H(\beta, F, L, f_{min})$ as well. As a consequence,

$$P_v(A_\varepsilon \cap S) \geq \left(\frac{\varepsilon}{4} f_{min}\left((1 - e^{-\frac{3}{4}\lambda\varepsilon})m\right) e^{-t^* N F}\right)^N > 0,$$

which concludes the proof. □

Remark 40. *In the proof of Proposition 4.1 of [DO14], the authors have no need to obtain (3.3.17) uniformly in $f \in H(\beta, F, L, f_{min})$. However, it is easy to see that we can rewrite their proof using the bounds for $f \in H(\beta, F, L, f_{min})$ appearing in the proof of Proposition 39 above. As a consequence, we obtain*

$$\inf_{f \in H(\beta,F,L,f_{min})} P_{t^*}(x, \cdot) \geq \eta_1 1_{B_{\delta^*}(u^*)}(x)\nu, \tag{3.3.18}$$

for some $\eta_1 > 0$.

Once we dispose of the uniform local Doeblin condition (3.3.18) and of the control given in Proposition 39, it is classical, using regeneration arguments, to show that the process is recurrent in the sense of Harris.

3.3.2 Harris recurrence and invariant measure

Using the regeneration procedure, we can prove that the process X is positive Harris recurrent. We denote by $\|\cdot\|_{TV}$ the total variation distance, *i.e.* $\|\nu_1 - \nu_2\|_{TV} = \sup_{B \in \mathcal{B}([0,K]^N)} |\nu_1(B) - \nu_2(B)|$, for any two probability measures ν_1, ν_2 on $([0, K]^N, \mathcal{B}([0, K]^N))$.

We first show that the process is indeed Harris. For that sake, define the sequence of stopping times $(\tilde{S}_n)_{n\in\mathbb{N}}$

$$\tilde{S}_1 := \inf\{t > 0 : X_t \in B_{\delta^*}(u^*)\},$$

and for all $n \geq 1$,

$$\tilde{S}_{n+1} := \inf\{t > \tilde{S}_n + t^* : X_t \in B_{\delta^*}(u^*)\}.$$

Let $(U_n)_{n\in\mathbb{N}}$ be a sequence of *i.i.d.* uniform random variables on $[0, 1]$, which are independent of the process X. Then, working conditionally on the realization of $(U_n)_{n\in\mathbb{N}}$, we define the sequence $(S_n)_{n\in\mathbb{N}}$ and the sequence $(R_n)_{n\in\mathbb{N}}$ of regeneration times as follows.

$$S_1 := \inf\{\tilde{S}_n : U_n < \eta_1\}, R_1 := S_1 + t^*,$$

and for all $n \geq 1$,

$$S_{n+1} := \inf\{\tilde{S}_k > S_n : U_k < \eta_1\}, R_{n+1} := S_{n+1} + t^*,$$

where η_1 is given in (3.3.18).

Remark 41. *(3.3.18) allows us to construct the process $(X_t)_{t\geq 0}$ on a bigger probability space in such a way that for all $n, X_{R_n} \sim \nu$ and $(X_{R_n+t})_{t\in\mathbb{R}^+}$ is independent from $\mathcal{F}_{S_n-}$. This construction is known as Nummelin splitting, we refer the interested reader to Chapter 6 of Löcherbach (2013) [Löc15].*

Lemma 42. *For all $x \in [0, K]^N, E_x(R_1) < \infty$ and $E_x(R_2 - R_1) < \infty$.*

The proof of this lemma is postponed to the next subsection where we prove a stronger result. Now the following result implies that our process is actually positive Harris recurrent.

Proposition 43. *X is Harris recurrent with invariant probability measure π which is given by*

$$\pi(B) := \frac{1}{E_x(R_2 - R_1)} E_x\left(\int_{R_1}^{R_2} 1_B(X_s)ds\right).$$

Proof. Fix $B \in \mathcal{B}([0,K]^N)$ and define the process A_t by

$$A_t := \int_0^t 1_B(X_s)ds.$$

Assume that $\pi(B) > 0$, then, according to the definition of Harris recurrence, it is enough to show that for all $x, \lim_{t\to+\infty} \frac{A_t}{t} > 0$.

We denote by $\tilde{N}_t$, $\tilde{N}_t^e$ and $\tilde{N}_t^o$ the counting processes respectively associated with the sequences of stopping times $(R_n)_{n\in\mathbb{N}^*}, (R_{2n})_{n\in\mathbb{N}^*}$ and $(R_{2n+1})_{n\in\mathbb{N}}$:

$$\tilde{N}_t := \sum_{n=1}^{\infty} 1_{R_n\le t}, \quad \tilde{N}_t^e := \sum_{n=1}^{\infty} 1_{R_{2n}\le t} \text{ and } \tilde{N}_t^o := \sum_{n=0}^{\infty} 1_{R_{2n+1}\le t}.$$

For all t we have $\tilde{N}_t = \tilde{N}_t^e + \tilde{N}_t^o$ and

$$\frac{A_t}{t} = \frac{1}{t}\left(\int_0^{R_1} 1_B(X_s)ds + \sum_{n=1}^{\tilde{N}_t}\int_{R_n}^{R_{n+1}} 1_B(X_s)ds - \int_{R_{\tilde{N}_t}}^{t} 1_B(X_s)ds\right)$$
$$= \frac{1}{t}\left(\int_0^{R_1} 1_B(X_s)ds + \sum_{k=1}^{\tilde{N}_t^e}\int_{R_{2k}}^{R_{2k+1}} 1_B(X_s)ds + \sum_{k=1}^{\tilde{N}_t^o}\int_{R_{2k-1}}^{R_{2k}} 1_B(X_s)ds - \int_{R_{\tilde{N}_t}}^{t} 1_B(X_s)ds\right).$$

When t goes to ∞, we obtain, using Lemma 42 to deal with the first and the last terms,

$$\lim_{t\to+\infty} \frac{A_t}{t}$$
$$= \lim_{t\to+\infty} \frac{\tilde{N}_t}{t}\frac{1}{\tilde{N}_t}\left(\tilde{N}_t^e\left(\frac{1}{\tilde{N}_t^e}\sum_{k=1}^{\tilde{N}_t^e}\int_{R_{2k}}^{R_{2k+1}} 1_B(X_s)ds\right) + \tilde{N}_t^o\left(\frac{1}{\tilde{N}_t^o}\sum_{k=1}^{\tilde{N}_t^o}\int_{R_{2k-1}}^{R_{2k}} 1_B(X_s)ds\right)\right).$$

The decomposition between even and odd regeneration times is used here to be able to apply the strong law of large numbers, based on Remark 41. In this way we obtain that

$$\lim_{t\to+\infty} \frac{1}{\tilde{N}_t^e}\sum_{k=1}^{\tilde{N}_t^e}\int_{R_{2k}}^{R_{2k+1}} 1_B(X_s)ds = \lim_{t\to+\infty} \frac{1}{\tilde{N}_t^o}\sum_{k=1}^{\tilde{N}_t^o}\int_{R_{2k-1}}^{R_{2k}} 1_B(X_s)ds = \pi(B) > 0 \text{ a.s.}$$

We can use the same decomposition to obtain that

$$\lim_{t\to+\infty} \frac{\tilde{N}_t}{t} = \frac{1}{E_x(R_2 - R_1)} \text{ a.s.}$$

Putting all together we have

$$\lim_{t\to+\infty} \frac{A_t}{t} = \frac{\pi(B)}{E_x(R_2 - R_1)}$$

and we can conclude the proof using Lemma 42 once again. □

Speed of convergence to equilibrium – Proof of (3.2.8) in Theorem 30

We now show how to couple two processes X and Y following the same dynamics (3.2.1) using Proposition 39 and the lower bound (3.3.18) of Remark 40. This coupling will give us a control of the distance in total variation between P_x and P_y, where x and y are the respective starting points of processes X and Y.

The coupling procedure consists in using the same realization of uniform random variables $(U_n)_{n\in\mathbb{N}}$ for both processes, relying on (3.3.18), when both processes X and Y are in the regeneration set $B_{\delta^*}(u^*)$ at the same time. More precisely, we let evolve X and Y independently up to the first time that they are both in the set $B_{\delta^*}(u^*)$. We introduce the sequence of stopping times

$$\bar{S}_1 = \inf\{t > 0 : (X_t, Y_t) \in B_{\delta^*}(u^*) \times B_{\delta^*}(u^*)\}$$

and

$$\bar{S}_n = \inf\{t > \bar{S}_{n-1} + t^* : (X_t, Y_t) \in B_{\delta^*}(u^*) \times B_{\delta^*}(u^*)\},\ n \geq 1.$$

Applying Proposition 39 to two independent processes X and Y, we obtain

$$\inf_{f\in H(\beta,F,L,f_{min})}\ \inf_{v_1,v_2\in[0,K]^N} P_{t^*}^{\otimes 2}((v_1,v_2), B_{\delta^*}(u^*)^2) \geq \eta_2^2. \tag{3.3.19}$$

As a consequence, $\bar{S}_n < \infty$ almost surely for all n, and $P_{(v_1,v_2)}(\bar{S}_1 > nt^*) \leq (1-\eta_2^2)^n$, *i.e.* $\bar{S}_1$ and $\bar{S}_{n+1} - \bar{S}_n$ possess exponential moments

$$E_{(v_1,v_2)}[e^{\alpha \bar{S}_1}] < \infty$$

uniformly in the starting configuration (v_1, v_2) for all $\alpha < \frac{-\ln(1-\eta_2^2)}{t^*}$.

We are now able to couple the processes X and Y. We work conditionally on the realization of a sequence of *i.i.d.* uniform random variables $(U_n)_{n\in\mathbb{N}}$ and define the coupling time τ by

$$\tau := \inf\{\bar{S}_n : U_n \leq \eta_1\} + t^*.$$

Using the regenerative construction described in the previous subsection based on (3.3.18), it is evident that X and Y can be constructed jointly in such a way that $X_\tau = Y_\tau \sim \nu$ and such that $X_t = Y_t$ for all $t \geq \tau$. Since τ is constructed by sampling within the sequence $(\bar{S}_n)_{n\in\mathbb{N}}$ at an independent geometrical time, it is immediate to see that there exists $\kappa > 1$ such that

$$\sup_{v_1,v_2\in[0,K]^N} E_{(v^1,v^2)}(\kappa^\tau) < +\infty. \tag{3.3.20}$$

Remark 44. *Notice that the regeneration time R_1 can be compared to τ and that $R_1 \leq \tau$. As a consequence,* (3.3.20) *implies a proof of Lemma 42.*

Since the two processes X and Y follow the same trajectory after time τ, we obtain the following classical upper bound on the total variation distance.

$$\|P_t(x,\cdot) - P_t(y,\cdot)\|_{TV} \leq P_{(x,y)}(\tau > t) \leq \kappa^{-t} E_{(x,y)}(\kappa^\tau). \tag{3.3.21}$$

Now putting $C := \sup_{x,y\in[0,K]^N} E_{(x,y)}(\kappa^\tau)$, the integration of (3.3.21) with respect to the invariant measure $\pi(dy)$ implies that

$$\sup_{x\in[0,K]^N} \|P_t(x,\cdot) - \pi\|_{TV} \leq C\kappa^{-t}.$$

This finishes the proof of Theorem 30. □

3.3.3 Estimates on the invariant density of a single particle

We start with some simple preliminary estimates. Recall that

$$\mu(ds,dx)=\sum_n \delta_{T_n}(ds)\delta_{X_{T_n-}}(dx)$$

denotes the jump measure of the system, with compensator

$$\hat{\mu}(ds,dx)=\bar{f}(X_s)ds\delta_{X_s}(dx),\ \text{with}\ \bar{f}(x)=\sum_{i=1}^N f(x^i).$$

Let $Z_k=X_{T_k-}, k\geq 1$, be the jump chain. Then the following holds.

Proposition 45. *$(Z_k)_k$ is Harris recurrent with invariant measure given by*

$$\pi^Z(g)=\frac{1}{\pi(\bar{f})}\pi(\bar{f}g),$$

for any $g:\mathbb{R}_+^N\to\mathbb{R}$ measurable and bounded.

Proof. Let g be a bounded test function. We have to prove that

$$\frac{1}{n}\sum_{k=1}^n g(Z_k)\to\pi^Z(g)$$

as $n\to\infty$, P_x-almost surely, for any fixed starting point $x\in[0,K]^N$. But

$$\frac{1}{n}\sum_{k=1}^n g(Z_k)=\frac{1}{n}\sum_{k=1}^n g(X_{T_k-}),$$

and, putting $N_t=\sup\{n:T_n\leq t\}$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n g(X_{T_k-})=\lim_{t\to\infty}\frac{t}{N_t}\frac{1}{t}\sum_{k=1}^{N_t} g(X_{T_k-})=\lim_{t\to\infty}\frac{t}{N_t}\frac{1}{t}\int_0^t\int_{\mathbb{R}_+^N} g(x)\mu(ds,dx).$$

By the law of large numbers, $N_t/t\to\int\bar{f}(x)\mu(dx)=\mu(\bar{f})$, and this convergence holds almost surely. Moreover,

$$\frac{1}{t}\int_0^t\int_{\mathbb{R}_+^N} g(x)\mu(ds,dx)=\frac{1}{t}M_t+\frac{1}{t}\int_0^t\int_{\mathbb{R}_+^N} g(x)\hat{\mu}(ds,dx), \tag{3.3.22}$$

where $M_t=\int_0^t\int g(x)[\mu(ds,dx)-\hat{\mu}(ds,dx)]$. Then M_t is in $\mathcal{M}_{loc}^{2,d}$, the set of all locally square integrable purely discontinuous martingales, with predictable quadratic covariation process

$$<M>_t=\int_0^t g^2(X_s)\bar{f}(X_s)ds \tag{3.3.23}$$

where

$$\frac{<M>_t}{t}\to\pi(g^2\bar{f})$$

almost surely, as $t\to\infty$. By the martingale convergence theorem, see *e.g.* Jacod-Shiryaev (2003) [JS87], chapter VIII, Corollary 3.24 , $t^{-1/2}M_t$ converges in law to a normal distribution. As a consequence, $M_t/t\to 0$ almost surely.

We now treat the second term in (3.3.22). By the ergodic theorem for integrable additive functionals,

$$\frac{1}{t}\int_0^t \int_{\mathbb{R}_+^N} g(x)\hat{\mu}(ds,dx) = \frac{1}{t}\int_0^t g(X_s)\bar{f}(X_s)ds \to \pi(\bar{f}g),$$

and this finishes the proof. □

Exchangeability of the invariant measure We denote by $X^i : \mathbb{R}_+^N \to \mathbb{R}_+, (x^1,\dots,x^N) \mapsto x^i$ the $i-$th coordinate map.

Proposition 46. *For all* $1 \le i \le N$, $\mathcal{L}_\pi(X^i) = \mathcal{L}_\pi(X^1)$.

Proof. Fix an initial configuration $x = (x^1, x^1, \dots, x^1) \in [0,K]^N$ consisting of N particles which are all in the same position. Let $\check{g} : \mathbb{R}_+ \to \mathbb{R}$ be a bounded test function and introduce $g(x) := \check{g}(x^1)$, *i.e.* g depends only on the first coordinate. By the ergodic theorem,

$$\frac{1}{t}\int_0^t g(X_s)ds \to \int_{\mathbb{R}} \check{g} d\mathcal{L}_\pi(X^1)$$

P_x-almost surely.

Now, introduce the system $Y_t = (Y_t^1,\dots,Y_t^N)$ given by $Y_t^k = X_t^k$ for all $k \neq 1, i$ and $Y_t^1 = X_t^i$, $Y_t^i = X_t^1$. Since the generator of X is invariant under permutations, $(Y_t)_{t\ge0} \overset{\mathcal{L}}{=} (X_t)_{t\ge0}$. In particular,

$$\int_{\mathbb{R}} \check{g} d\mathcal{L}_\pi(X^1) = \lim_{t\to\infty}\frac{1}{t}\int_0^t g(X_s)ds = \lim_{t\to\infty}\frac{1}{t}\int_0^t g(Y_s)ds$$

On the other hand,

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t g(Y_s)ds = \lim_{t\to\infty}\frac{1}{t}\int_0^t \check{g}(X_s^i)ds = \int_{\mathbb{R}} \check{g} d\mathcal{L}_\pi(X^i),$$

and this finishes the proof. □

We are now going to study the support properties of the invariant measure of a single neuron. For that sake define for all $x \in [0,K]$, $b(x) := \lambda(x-m)$ and recall that $\gamma_t(x^i) \in \mathbb{R}_+$ denotes the solution of $d\gamma_t(x^i) = -b(\gamma_t(x^i))dt$, given by

$$\gamma_t(x^i) = e^{-\lambda t}x^i + (1-e^{-\lambda t})m.$$

Moreover, for $x = (x^1,\dots,x^N)$, $\gamma_t(x) = (\gamma_t(x^1),\dots,\gamma_t(x^N))$. Finally, let

$$K(x,dy) = \sum_{i=1}^N f(x^i)\delta_{\Delta^i(x)}(dy), \quad H_f^x(t) = e^{-\int_0^t \bar{f}(\gamma_s(x))ds},$$

where $\bar{f}(\gamma_s(x)) = \sum_{i=1}^N f(\gamma_s(x^i))$ and where $\Delta^i(x)$ was defined in (3.2.3) before.

We will use the change of variable, for a fixed value of y,

$$z = \gamma_t(y^1), dz = -b(z)dt = -\lambda(z-m)dt \tag{3.3.24}$$

and denote by $\kappa_y(z)$ the inverse function of $t \to \gamma_t(y^1)$.

These definitions permit to obtain an expression of π_1.

Proposition 47. *For all $z \in S_{d,k}$, we have*

$$\pi_1(z) = \int \pi(dx) \int K(x, dy) H_f^y(\kappa_y(z)) \frac{1_{I_y^m}(z)}{|b(z)|}. \tag{3.3.25}$$

Here the notation I_y^m denotes either $]y, m[$ if $y < m$ or $]m, y[$ if $m < y$.

Proof. We have, by Proposition 45,

$$\pi(G) = E_\pi(G(X_t)) = E_\pi(\bar{f}(X_t) G(X_t) \frac{1}{\bar{f}(X_t)}) = \pi(\bar{f}) E_{\pi^Z}\left(G(Z_n)\frac{1}{\bar{f}(Z_n)}\right).$$

We use that $\pi^Z = \mathcal{L}(X_{T_2-} | X_{T_1-} \sim \pi^Z)$. Then we obtain

$$\begin{aligned}
\pi(G) &= \pi(\bar{f})\pi^Z\left(\frac{G(Z_n)}{\bar{f}(Z_n)}\right) \\
&= \int_{\mathbb{R}_+^N} \bar{f}(x)\pi(dx) \sum_{i=1}^N \frac{f(x^i)}{\bar{f}(x)} \int_0^\infty \bar{f}(\gamma_t(\Delta_i(x))) \\
&\qquad e^{-\int_0^t \bar{f}(\gamma_s(\Delta_i(x)))ds} G(\gamma_t(\Delta_i(x))) \frac{1}{\bar{f}(\gamma_t(\Delta_i(x)))} dt \\
&= \int_{\mathbb{R}_+^N} \pi(dx) \int K(x, dy) \int_0^\infty H_f^y(t) G(\gamma_t(y)) dt.
\end{aligned} \tag{3.3.26}$$

Now, let $g \in C_c^\infty(S_{d,\beta})$ be a smooth test function having compact support in $S_{d,\beta}$. Using (3.3.26), we obtain

$$\pi_1(g) = \int_{\mathbb{R}_+^N} \pi(dx) \int K(x, dy) \int_0^\infty H_f^y(t) g(\gamma_t(y^1)) dt. \tag{3.3.27}$$

Then, with the change of variable announced in (3.3.24), we can rewrite (3.3.27) in the following way:

$$\pi_1(z) = \int \pi(dx) \int K(x, dy) H_f^y(\kappa_y(z)) \frac{1_{I_y^m}(z)}{|b(z)|}.$$

□

3.3.4 Support of the invariant measure

Proposition 48. *For all $y \in]0, K[$, all $\delta > 0$, we have*

$$\inf_{f \in H(\beta, F, L, f_{min})} \pi_1\left(B_\delta(y)\right) > 0.$$

Proof. Fix $y \in]0, K[$ and let $k \in \mathbb{N}$ and $s \in [0, \frac{1}{N}[$ be such that $y = \frac{k}{N} + s$.

We define the time t_s such that $m\left(1 - e^{-\lambda t_s}\right) = s$ and consider, for a fixed $\varepsilon > 0$, the following events:

$$A_y = \left\{\frac{\varepsilon}{2} < T_1 < \varepsilon; t_s + (i-1)\varepsilon < T_i < t_s + (i-1)\varepsilon + \frac{\varepsilon}{2} \ \forall i = 2, \ldots, k+1\right\}$$

and

$$S_y = \{I_1 = 1, I_2 \neq 1, \ldots, I_{k+1} \neq 1\}.$$

The idea of the proof is that the event $A_y \cap S_y$ leads the neuron 1 to a position close to y after a time $t_y := t_s + k\varepsilon$:

At time T_1 the neuron 1 jumps so that its position is reset to 0, the time t_s is defined such that at time T_{2-} the position of neuron 1 is close to s, then in an interval of time short enough for the deterministic drift to be insignificant, we impose that the other neurons jump k times so that at time T_{k+1}, the position of neuron 1 is indeed close to y.

In other words we can use similar arguments to the ones used in the proof of Lemma 34 to obtain that, for all $x \in [0,K]^N$, if $X_0 = x$, then on the event $A_y \cap S_y$, we have $X^1_{T_{k+1}} = y + O(\varepsilon)$, and we can choose ε such that $X^1_{T_{k+1}} \in B_\delta(y)$.

Now we have to prove that

$$\inf_{f\in H(\beta,F,L,f_{min})} \inf_{x\in[0,K]^N} P_x(A_y \cap S_y) > 0,$$

which can be done as in the proof of proposition 39.

Finally, integrating this result against the measure π gives us the conclusion of the proof. □

We can now obtain (3.2.11) as corollary of the following Proposition.

Proposition 49. *We have that*

$$r^* := \inf_{a\in S_{d,\beta}} \inf_{f\in H(\beta,F,L,f_{min})} \pi_1(a) > 0, \tag{3.3.28}$$

and for all $x \in [0,K]^N$,*and for all* $r \le r^*$,

$$\liminf_{t\to\infty} \inf_{f\in H(\beta,F,L,f_{min})} P^f_x(A_{t,r}) = 1. \tag{3.3.29}$$

Proof. Recalling the construction of π_1 in (3.3.25), we have

$$\begin{aligned}\pi_1(a) &= \int \pi(dx) \int K(x,dy) H^y_f(\kappa_y(a)) \frac{1_{I^m_y}(a)}{|b(a)|} \\ &= \int \pi(dx) \int \left(\sum_{i=1}^N f(x^i)\delta_{\Delta^i(x)}(dy)\right) \exp\left(-\int_0^{\kappa_y(a)} \bar f(\gamma_u(x))du\right) \frac{1_{I^m_y}(a)}{|b(a)|}.\end{aligned}$$

To obtain a lower bound uniform in f of this expression we use again the bounds of the class of function $H(\beta,F,L,f_{min})$:

$$\forall f \in H(\beta,F,L,f_{min}),\ f_{min}(x) \le f(x) \le F.$$

Doing this, we will also need an upper bound for $\kappa_y(a)$. This is possible due to the term $1_{I^m_y}(a)$: since y is such that $a \in I^m_y \cap S_{d,\beta}$, the flow starting from y can reach a in a finite time, even if we consider the worst cases where $y = 0$ or K.

Thanks to Proposition 48, we have $\pi_1\left(\{y : a \in I^m_y\}\right) > 0$ implying that the integration of $1_{I^m_y}(a)$ against the measure $\pi_1(dy)$ is not 0. Finally, due to the definition of $S_{d,\beta}$, we have no problem to obtain this lower bound uniformly in $a \in S_{d,\beta}$, and this finishes the proof of (3.3.28).

(3.3.29) is obtained easily from (3.3.28) thanks to the ergodic theorem: we have

$$\lim_{t\to+\infty} \frac{1}{Nt} \int_0^t \int_{\mathbb{R}} Q_h(y-a)\eta(ds,dy) = \pi_1(a),$$

(recall that $\int_{\mathbb{R}} Q(x)dx = 1$), which concludes the proof. □

3.4 proof of theorem 32

3.4.1 Convergence of the estimator

We now study the speed of convergence of our estimator. First we have the following classical kernel approximation:

Proposition 50. *For any Hölder function g of order $\beta = k + \alpha$ satisfying*

$$\sup_{w \neq w'} \left| g^{(k)}(w) - g^{(k)}(w') \right| \leq C_g |w - w'|^\alpha \tag{3.4.30}$$

for some constant C_g and for a kernel Q as in Theorem 32, we have:

$$\left| \int_{\mathbb{R}} Q_h(y-a) g(y) dy - g(a) \right| \leq \frac{C_g \parallel Q \parallel_{L^1} R^\beta}{k!} h^\beta,$$

where we recall that R is the diameter of the support of Q and where $\parallel Q \parallel_{L^1}$ denotes the L^1-norm of Q.

Proof. Using the property $\int_{\mathbb{R}} Q(x)dx = 1$ and the change of variable $x = \frac{y-a}{h}$, we obtain

$$\int_{\mathbb{R}} Q_h(y-a) g(y) dy - g(a) = \int_{\mathbb{R}} Q(x) \left(g(a + xh) - g(a) \right) dx.$$

Then, a Taylor-Lagrange expansion of the function g gives us

$$\int_{\mathbb{R}} Q_h(y-a) g(y) dy - g(a) = \int_{\mathbb{R}} Q(x) \left(\sum_{l=1}^{k} \frac{g^{(l)}(a)}{l!} (xh)^l + \frac{g^{(k)}(z) - g^{(k)}(a)}{k!} (xh)^k \right) dx,$$

for some $z \in]a, a + xh[\cup]a + xh, a[$. By the assumptions of Theorem 32, $\int_{\mathbb{R}} Q(y) y^j dy = 0$ for all $1 \leq j \leq k$. Then condition (3.4.30) allows to conclude. □

Fix $a \in S_{d,\beta}$ and define, for all $t \in \mathbb{R}^+$, $\tilde{\mu} = \mu - \hat{\mu}$ the centered jump measure.

Proposition 51. *Under the conditions of Theorem 32, there exists a constant C_1 depending only on β, F, L, N, f_{min} and Q, such that for all $f \in H(\beta, F, L, f_{min})$, for all $x \in [0, K]$ and for a bandwidth of the form $h = h_t = t^{-\alpha}$ for some $0 < \alpha < 1$,*

$$E_x \left[\left(\frac{1}{Nt} \int_{[0,t]} \int_{\mathbb{R}} Q_h(y-a) \tilde{\mu}(ds, dy) \right)^2 \right] \leq \frac{C_1}{ht}. \tag{3.4.31}$$

Proof. We start working under the invariant regime in the first part of the proof, *i.e.* we will work under E_π. In a second time we will use Theorem 30 to obtain the result for any starting point $x \in [0, K]^N$.

We use the properties of the compensator $\hat{\mu}_t$ and its explicit expression to write

$$\begin{aligned} E_\pi \left[\left(\frac{1}{Nt} \int_{[0,t]} \int_{\mathbb{R}} Q_h(y-a) \tilde{\mu}(ds, dy) \right)^2 \right] &= \frac{1}{(Nt)^2} E_\pi \left[\int_0^t \int_{\mathbb{R}} \left(Q_h(y-a) \right)^2 \hat{\mu}(ds, dy) \right] \\ &= \frac{1}{(Nt)^2} E_\pi \left[\int_0^t \int_{\mathbb{R}} \left(Q_h(y-a) \right)^2 f(y) \eta(ds, dy) \right]. \end{aligned}$$

Now, since we are in the invariant regime, we can use the density of the invariant measure of a single particle (recall Theorem 31) to obtain

$$E_\pi\left[\left(\frac{1}{Nt}\int_{[0,t]}\int_{\mathbb{R}} Q_h(y-a)\tilde{\mu}(ds,dy)\right)^2\right] = \frac{1}{Nt}\int_{\mathbb{R}}\left(Q_h(y-a)\right)^2 f(y)\pi_1(y)dy.$$

Our aim is to obtain a control of $\int_{\mathbb{R}} hf(y)(Q_h(y-a))^2\pi_1(y)dy$ independently of h. To do this we use the change of variable $x=\frac{y-a}{h}$ and write

$$E_\pi\left[\left(\frac{1}{Nt}\int_{[0,t]}\int_{\mathbb{R}} Q_h(y-a)\tilde{\mu}(ds,dy)\right)^2\right] = \frac{1}{Nht}\int_{\mathbb{R}} Q^2(x)f(a+xh)\pi_1(a+xh)dx.$$

This yields

$$E_\pi\left[\left(\frac{1}{Nt}\int_{[0,t]}\int_{\mathbb{R}} Q_h(y-a)\tilde{\mu}(ds,dy)\right)^2\right] \leq \frac{F}{Nht}\|Q\|_{L^2}^2 \sup_{x\in S_{d/2,k}}\pi_1(x). \tag{3.4.32}$$

This result holds in stationary regime, but thanks to the exponential speed of convergence of Theorem 30, we can obtain it for any starting point $x\in[0,K]^N$ as we are going to show now. For that sake we fix the bandwidth h in function of t so that this speed of convergence depends only on t. For the moment, we will assume that h is of the form

$$h_t := t^{-\alpha} \tag{3.4.33}$$

for some constant $\alpha\in]0,1[$. As in the beginning of the proof, we can write

$$E_x\left[\left(\frac{1}{Nt}\int_{[0,t]}\int_{\mathbb{R}} Q_h(y-a)\tilde{\mu}(ds,dy)\right)^2\right] = \frac{1}{(Nt)^2}E_x\left[\int_0^t\int_{\mathbb{R}}\left(Q_h(y-a)\right)^2 f(y)\eta(ds,dy)\right].$$

Now, we have the following decomposition

$$\begin{aligned}
&E_x\left[\left(\frac{1}{Nt}\int_{[0,t]}\int_{\mathbb{R}} Q_h(y-a)\tilde{\mu}(ds,dy)\right)^2\right]\\
&= \frac{1}{(Nt)^2}E_x\left[\sum_{i=1}^N\int_0^t\left(Q_h(X_s^i-a)\right)^2 f(X_s^i)ds\right] - \frac{1}{(Nt)^2}E_\pi\left[\sum_{i=1}^N\int_0^t\left(Q_h(X_s^i-a)\right)^2 f(X_s^i)ds\right]\\
&\qquad + \frac{1}{(Nt)^2}E_\pi\left[\sum_{i=1}^N\int_0^t\left(Q_h(X_s^i-a)\right)^2 f(X_s^i)ds\right].
\end{aligned}$$

The last term is controlled by (3.4.32). We will deal with the difference in the second line using Theorem 30 as follows: for all $p\in]0,1-\alpha[$, we have

$$\begin{aligned}
&\frac{1}{(Nt)^2}E_x\left[\sum_{i=1}^N\int_0^t\left(Q_h(X_s^i-a)\right)^2 f(X_s^i)ds\right] - \frac{1}{(Nt)^2}E_\pi\left[\sum_{i=1}^N\int_0^t\left(Q_h(X_s^i-a)\right)^2 f(X_s^i)ds\right]\\
&= \frac{1}{(Nt)^2}E_x\left[\sum_{i=1}^N\int_0^{t^p}\left(Q_h(X_s^i-a)\right)^2 f(X_s^i)ds\right] - \frac{1}{(Nt)^2}E_\pi\left[\sum_{i=1}^N\int_0^{t^p}\left(Q_h(X_s^i-a)\right)^2 f(X_s^i)ds\right]\\
&\qquad + \frac{1}{(Nt)^2}\sum_{i=1}^N\int_{t^p}^t\Big(E_x\left[\left(Q_h(X_s^i-a)\right)^2 f(X_s^i)\right] - E_\pi\left[\left(Q_h(X_s^i-a)\right)^2 f(X_s^i)\right]\Big)ds.
\end{aligned}$$

To conclude, we use the upper bounds $\| Q \|_\infty$ and F for Q and f to control the second line and we use Theorem 30 to control the last term. As a consequence,

$$\left| \frac{1}{(Nt)^2} E_x \left[\sum_{i=1}^N \int_0^t \left(Q_h(X_s^i - a) \right)^2 f(X_s^i) ds \right] - \frac{1}{(Nt)^2} E_\pi \left[\sum_{i=1}^N \int_0^t \left(Q_h(X_s^i - a) \right)^2 f(X_s^i) ds \right] \right|$$
$$\leq \frac{F \| Q \|_\infty^2}{Nh^2 t^2} \left(2t^p + C \int_{t^p}^t \kappa^{-s} ds \right) = \frac{F \| Q \|_\infty^2}{Nh^2 t^2} \left(2t^p + C \frac{\kappa^{-t^p} - \kappa^{-t}}{\ln(\kappa)} \right) = \frac{1}{ht} \mathcal{O}\left(\frac{t^p}{ht} \right).$$

Now recall that $h = h_t = t^{-\alpha}$ by (3.4.33) and that $p \in]0, 1-\alpha[$. Thus

$$\frac{1}{(Nt)^2} E_x \left[\sum_{i=1}^N \int_0^t \left(Q_h(X_s^i - a) \right)^2 f(X_s^i) ds \right] - \frac{1}{(Nt)^2} E_\pi \left[\sum_{i=1}^N \int_0^t \left(Q_h(X_s^i - a) \right)^2 f(X_s^i) ds \right] = o\left(\frac{1}{ht} \right),$$

which allows to conclude. □

Proposition 51 will help us to control the numerator of our estimator. We want to establish the same kind of result for the denominator and this leads to the following proposition:

Proposition 52. *For all $a \in S_{d,\beta}$, define*

$$\tilde{Q}_{h,f}(y) := Q_h(y - a)\Big(f(y) - f(a)\Big) - \pi_1\Big(Q_h(\cdot - a)\Big(f(\cdot) - f(a)\Big)\Big). \tag{3.4.34}$$

Under the conditions of Theorem 32, there exists a constant C_2 depending only on β, F, L, N, f_{min} and Q, such that for all $f \in H(\beta, F, L, f_{min})$, for all $x \in [0, K]$ and for a bandwidth of the form $h = h_t = t^{-\alpha}$ for some $0 < \alpha < 1$,

$$E_x \left[\left(\frac{1}{Nt} \int_0^t \int_{\mathbb{R}} \tilde{Q}_{h,f}(y) \eta(ds, dy) \right)^2 \right] \leq \frac{C_2}{t} h^{2(1 \wedge \beta) - 1}. \tag{3.4.35}$$

Proof. As in the preceding proof we start by working in the stationary regime, *i.e.* under E_π.

$$E_\pi \left[\left(\frac{1}{Nt} \int_0^t \int_{\mathbb{R}} \tilde{Q}_{h,f}(y) \eta(ds, dy) \right)^2 \right]$$
$$\leq \frac{2}{(Nt)^2} E_\pi \left[\int_0^t \int_{\mathbb{R}} \left| \tilde{Q}_{h,f}(x) \right| \eta(ds, dx) \left| E_\pi \left(\int_s^t \int_{\mathbb{R}} \tilde{Q}_{h,f}(y) \eta(du, dy) \Big| \mathcal{F}_s \right) \right| \right]. \tag{3.4.36}$$

We deal with the conditional expectation using the Markov property and write

$$E_\pi \left(\int_s^t \int_{\mathbb{R}} \tilde{Q}_{h,f}(y) \eta(du, dy) \Big| \mathcal{F}_s \right)$$
$$= E_{X_s} \left(\int_0^{t-s} \sum_{i=1}^N \tilde{Q}_{h,f}(X_u^i) du \right) = \int_0^{t-s} \sum_{i=1}^N E_{X_s} \left(\tilde{Q}_{h,f}(X_u^i) \right) du.$$

Now going back to the definition of $\tilde{Q}_{h,f}$, we can use Theorem 30 and write

$$E_{X_s} \left(\tilde{Q}_{h,f}(X_u^i) \right) = E_{X_s} \Big(Q_h(X_u^i - a) \left(f(X_u^i) - f(a) \right) \Big) - \pi_1 \Big(Q_h(\cdot - a) \left(f(\cdot) - f(a) \right) \Big)$$
$$\leq \frac{C}{h} (F \vee L)(Rh)^{1 \wedge \beta} \| Q \|_\infty \kappa^{-u},$$

due to the assumption (3.2.6) on the Hölder space containing f. (Recall that R is the diameter of the support of Q.) The integrability of the function $u \to \kappa^{-u}$allows to deduce from this that

$$\left|E_\pi\left(\int_s^t \int_{\mathbb{R}} \tilde{Q}_h(y-a)\eta(du,dy)\middle|\mathcal{F}_s\right)\right| \leq \frac{N\tilde{C}}{h}(F\vee L)(Rh)^{1\wedge\beta} \parallel Q \parallel_\infty$$

for some constant $\tilde{C}$. Taking this result into account in (3.4.36), we obtain

$$E_\pi\left[\left(\frac{1}{Nt}\int_0^t\int_{\mathbb{R}} \tilde{Q}_{h,f}(y)\eta(ds,dy)\right)^2\right] \leq \frac{2\tilde{C}}{Nht^2}(F\vee L)(Rh)^{1\wedge\beta} \parallel Q \parallel_\infty E_\pi\left[\int_0^t\int_{\mathbb{R}}\left|\tilde{Q}_{h,f}(x)\right|\eta(ds,dx)\right].$$

The end of the proof is similar to the one of Proposition 51: the fact that we are in the invariant regime allows to use the density of the invariant measure of a single particle and its control given by Theorem 31. Then we use the same change of variable $x = \frac{y-a}{h}$ to obtain

$$E_\pi\left[\left(\frac{1}{Nt}\int_0^t\int_{\mathbb{R}} \tilde{Q}_{h,f}(y)\eta(ds,dy)\right)^2\right] \leq \frac{4\tilde{C}}{ht}(F\vee L)^2(Rh)^{2(1\wedge\beta)} \parallel Q \parallel_\infty \parallel Q \parallel_{L^1} \sup_{x\in S_{d/2,k}} \pi_1(x).$$

This result is established under the invariant regime, but we are able to extend it to any starting point $x \in [0,K]^N$, using the same trick as the one in the proof of Proposition 51. This finishes the proof. □

3.4.2 Proof of Theorem 32, (i)

Introducing

$$D^{t,h} = \frac{1}{Nt}\int_{\mathbb{R}} \frac{1}{h} Q\left(\frac{y-a}{h}\right)\eta_t(dy),$$

we have

$$\begin{aligned} D^{t,h}(\hat{f}_{t,h}(a) - f(a)) &= \frac{1}{Nt}\int_{[0,t]}\int_{\mathbb{R}} Q_h(y-a)\mu(ds,dy) - f(a)D^{t,h} \\ &= \frac{1}{Nt}\int_{[0,t]}\int_{\mathbb{R}} Q_h(y-a)\tilde{\mu}(ds,dy) + \frac{1}{Nt}\int_0^t\int_{\mathbb{R}} \frac{1}{h}Q\left(\frac{y-a}{h}\right)(f(y)-f(a))\,\eta(ds,dy). \end{aligned}$$

With the definition of $\tilde{Q}_{h,f}$ in (3.4.34), we have the following decomposition:

$$\begin{aligned} &D^{t,h}(\hat{f}_{t,h}(a) - f(a)) \\ &= \frac{1}{Nt}\int_{[0,t]}\int_{\mathbb{R}} Q_h(y-a)\tilde{\mu}(ds,dy) + \frac{1}{Nt}\int_0^t\int_{\mathbb{R}} \tilde{Q}_{h,f}(y)\eta(ds,dy) + \pi_1\Big(Q_h(\cdot-a)\,(f(\cdot)-f(a))\Big). \end{aligned} \tag{3.4.37}$$

The first two terms of the previous sum are controlled respectively by Propositions 51 and 52. We deal with the third term using Proposition 50 as follows:

$$\begin{aligned} &\pi_1\Big(Q_h(\cdot-a)\,(f(\cdot)-f(a))\Big) = \int_{\mathbb{R}} Q_h(y-a)\,(f(y)-f(a))\,\pi_1(y)dy \\ &= \left(\int_{\mathbb{R}} Q_h(y-a)\,(f(y)\pi_1(y) - f(a)\pi_1(a))\,dy\right) + f(a)\left(\int_{\mathbb{R}} Q_h(y-a)\,(\pi_1(a)-\pi_1(y))\,dy\right). \end{aligned}$$

Both functions π_1 and $f\pi_1$ are Hölder of order β (recall Theorem 31) and we can apply Proposition 50 to each of the last two terms, using the upper bound F for $f(a)$.

Putting all together in (3.4.37), we have

$$\| D^{t,h}\left(\hat{f}_{t,h}(a) - f(a)\right) \|_{L^2(P_x^f)} \leq \sqrt{\frac{C_1}{ht}} + \sqrt{\frac{C_2}{ht}} h^{1\wedge\beta} + C_3 h^{\beta}, \tag{3.4.38}$$

with constants C_1, C_2 and C_3 depending only on β, F, L, f_{min} and Q. As in the proof of Proposition 52, we will fix the dependence in t of h putting $h_t := t^{-\alpha}$ and choosing $\alpha \in]0,1[$ to obtain an optimal speed of convergence.

This leads to the choice $\alpha := \frac{1}{2\beta+1}$ and $h = h_t = t^{-\frac{1}{2\beta+1}}$ which gives us

$$\| D^{t,h_t}\left(\hat{f}_{t,h_t}(a) - f(a)\right) \|_{L^2(P_x^f)} \leq C(\beta, F, L, f_{min}, Q) t^{-\frac{\beta}{2\beta+1}}.$$

To finish the proof of Theorem 32, we have to work conditionally on the event $A_{t,r}$, for $r \leq r^*$, on which we have $D^{t,h} \geq r$.

$$\begin{aligned} E_x\left[\left(\hat{f}_{t,h_t}(a) - f(a)\right)^2 \Big| A_{t,r}\right] &= \frac{1}{P_x\left(A_{t,r}\right)} E_x\left[\left(\hat{f}_{t,h_t}(a) - f(a)\right)^2 1_{A_{t,r}}\right] \\ &\leq \frac{1}{r^2 P_x\left(A_{t,r}\right)} \| D^{t,h_t}\left(\hat{f}_{t,h_t}(a) - f(a)\right) \|^2_{L^2(P_x^f)} \leq \frac{C(\beta, F, L, f_{min}, Q)^2 t^{-\frac{2\beta}{2\beta+1}}}{r^2 P_x\left(A_{t,r}\right)}, \end{aligned}$$

and the conclusion follows thanks to (3.3.29). □

3.4.3 Proof of Theorem 32 (ii):

The proof relies on the martingale convergence theorem given in Corollary 3.24 of [JS87] chapter VIII. We use the following decomposition

$$D^{t,h}(\hat{f}_{t,h}(a) - f(a)) = \frac{1}{N\sqrt{th}} M^{t,h} + \frac{1}{Nt} \int_0^t \int_{\mathbb{R}} Q_h(y-a)\left(f(y) - f(a)\right) \eta(du, dy), \tag{3.4.39}$$

where

$$M^{t,h} := \frac{1}{\sqrt{th}} \int_{[0,t]} \int_{\mathbb{R}} Q\left(\frac{y-a}{h}\right) \tilde{\mu}(ds, dy).$$

We define for all $t \in \mathbb{R}_+$

$$(M^t)_s := \frac{1}{\sqrt{th}} \int_{[0,ts]} \int_{\mathbb{R}} Q\left(\frac{y-a}{h}\right) \tilde{\mu}(du, dy)$$

and show that the Assumption 3.23 of [JS87] chapter VIII is satisfied for this sequence of processes. Therefore, we have to study, for all $\varepsilon > 0$ and all $s \in \mathbb{R}_+$, the limit of

$$\frac{1}{th} \int_0^{ts} \sum_{i=1}^N f\left(X_u^i\right) Q^2\left(\frac{X_u^i - a}{h}\right) 1_{\left\{\frac{1}{\sqrt{th}} Q\left(\frac{X_u^i - a}{h}\right) > \varepsilon\right\}} du$$

as t goes to $+\infty$. Since Q is bounded and $\lim_{t\to+\infty} th_t = +\infty$, there exists t_0 such that for all $t > t_0$, $1_{\left\{\frac{1}{\sqrt{th}} Q\left(\frac{X_u^i - a}{h}\right) > \varepsilon\right\}} = 0$. Consequently, the above limit is 0 and Assumption 3.23 of [JS87] chapter VIII is indeed satisfied.

Moreover,

$$\left\langle M^t, M^t \right\rangle_s = \frac{1}{th} \int_0^{ts} \int_{\mathbb{R}} Q^2 \left(\frac{y-a}{h} \right) f(y) \eta(du, dy).$$

Since our process is positive Harris recurrent, by the ergodic theorem, we have the following proposition.

Proposition 53. *$\left\langle M^t, M^t \right\rangle_s$ converges in P_x-Probability as t goes to $+\infty$ to*

$$Nsf(a)\pi_1(a) \int Q^2(x)dx \text{ a.s.}$$

Proof. Since our process is positive Harris recurrent, f being continuous and Q with compact support, we have

$$\lim_{t \to +\infty} E_x \left[\left(\frac{1}{th} \int_0^{ts} \int_{\mathbb{R}} Q^2 \left(\frac{y-a}{h} \right) f(y) \eta(du, dy) - \frac{N}{th} \int_0^{ts} \int_{\mathbb{R}} Q^2 \left(\frac{y-a}{h} \right) f(y) \pi_1(y) dy \right)^2 \right] = 0.$$

Then the result is obtained by continuity of π_1 and f on $S_{d,k}$. □

Consequently, Corollary 3.24 of [JS87] chapter VIII with $s = 1$ gives us the weak convergence of $M^{t,h}$ to $\mathcal{N}\left(0, Nf(a)\pi_1(a) \int Q^2(x)dx\right)$.

We deal with the second term of (3.4.39) as in the previous subsection and obtain

$$\left\| \frac{1}{Nt} \int_0^t \int_0^K Q_h(y-a) \left(f(y) - f(a) \right) \eta(du, dy) \right\|_{L^2(P_x^f)} \leq \sqrt{\frac{C_2}{ht}} h^{1 \wedge \beta} + C_3 h^\beta.$$

Therefore, when t goes to $+\infty$, (3.4.39) gives us the following weak convergence:

$$\sqrt{th_t} D^{t,h_t} (\hat{f}_{t,h_t}(a) - f(a)) \longrightarrow \mathcal{N}\left(0, \frac{f(a)\pi_1(a)}{N} \int Q^2(x)dx \right),$$

since $h_t = o(t^{-1/(1+2\beta)})$.

Finally, we deal with the additive functional D^{t,h_t} using the ergodic theorem. Recall that

$$D^{t,h} = \frac{1}{Nt} \int_0^t \int_{\mathbb{R}} \frac{1}{h} Q \left(\frac{y-a}{h} \right) \eta(ds, dy).$$

Thanks to (3.3.28), $\pi_1(a) > 0$, and the ergodic theorem gives us the almost sure convergence to $\pi_1(a)$ (since $\int Q(x)dx = 1$), which allows us to conclude. □

3.5 Proof of Theorem 33

The proof of Theorem 33 follows closely the proof of Theorem 8 of Hoffmann and Olivier (2015) [HO], going back to similar ideas developed in [HHL02]. Let $h_t = t^{-\frac{1}{2\beta+1}}$ and fix any test rate function $f_0 \in H(\beta, F - \delta, L - \delta, f_{min})$, for some fixed $\delta \in]0, F \wedge L[$. Then, as in [HO], we define a perturbation f_t of f_0 by

$$f_t(x) = f_0(x) + b h_t^{\beta+1} \chi_{h_t}(x - a),$$

where $b > 0$ is a positive constant, $\chi \in C_c(\mathbb{R}_+, \mathbb{R}_+)$ is a positive kernel function of compact support included in $[-1, 1]$ such that $\chi(0) = 1$, $\chi(x) \leq 1$ for all x and

$$\chi_{h_t}(x) = \frac{1}{h_t} \chi(\frac{x}{h_t}). \tag{3.5.40}$$

Notice that the first l derivatives of χ_{h_t} are of order $h_t^{-(l+1)}$, therefore the factor $h_t^{\beta+1}$ implies that $f_t \in H(\beta, F, L, f_{min})$, if we choose b sufficiently small. An important point in the above choice of f_t is that

$$f_t(a) - f_0(a) = bh_t^\beta = bt^{-\frac{\beta}{2\beta+1}}, \tag{3.5.41}$$

since $\chi(0) = 1$.

In the following, we shall write shortly $\mathbb{P}_0 := (P_x^{f_0})_{|\mathcal{F}_t}$ and $\mathbb{P}_t := (P_x^{f_t})_{|\mathcal{F}_t}$ for the associated probability measures in restriction to $\mathcal{F}_t$. The following lower bound is by now classical. For any fixed constant $C > 0$, using Markov's inequality and denoting by $L_t^{f_t/f_0} = \frac{d\mathbb{P}_0}{d\mathbb{P}_t}$ the likelihood ratio of $\mathbb{P}_0$ with respect to $\mathbb{P}_t$, on $\mathcal{F}_t$,

$$\begin{aligned}
&\sup_{f\in H(\beta,F,L,f_{min})} t^{\frac{2\beta}{1+2\beta}} E_x^f[|\hat{f}_t(a) - f(a)|^2] \\
&\geq t^{\frac{2\beta}{1+2\beta}} \left[\frac{1}{2}\mathbb{E}_0[|\hat{f}_t(a) - f_0(a)|^2] + \frac{1}{2}\mathbb{E}_t[|\hat{f}_t(a) - f_t(a)|^2]\right] \\
&\geq \frac{C^2}{2}\left[\mathbb{P}_0\left(t^{\frac{\beta}{1+2\beta}}|\hat{f}_t(a) - f_0(a)| \geq C\right) + \mathbb{P}_t\left(t^{\frac{\beta}{1+2\beta}}|\hat{f}_t(a) - f_t(a)| \geq C\right)\right] \\
&= \frac{C^2}{2}\left[\mathbb{P}_0\left(t^{\frac{\beta}{1+2\beta}}|\hat{f}_t(a) - f_0(a)| \geq C\right) + \mathbb{E}_0\left(L_t^{f_t/f_0} 1_{\{t^{\frac{\beta}{1+2\beta}}|\hat{f}_t(a)-f_t(a)|\geq C\}}\right)\right].
\end{aligned}$$

Now,

$$t^{\frac{\beta}{1+2\beta}}[|\hat{f}_t(a) - f_0(a)| + |\hat{f}_t(a) - f_t(a)|] \geq t^{\frac{\beta}{1+2\beta}}|f_0(a) - f_t(a)| \geq b,$$

which is due to (3.5.41). As a consequence, if we choose $C = b/2$, then

$$1_{\{t^{\frac{\beta}{1+2\beta}}|\hat{f}_t(a)-f_0(a)|\geq C\}} + 1_{\{t^{\frac{\beta}{1+2\beta}}|\hat{f}_t(a)-f_t(a)|\geq C\}} \geq 1,$$

in particular,

$$1_{\{t^{\frac{\beta}{1+2\beta}}|\hat{f}_t(a)-f_t(a)|\geq C\}} \geq 1_{\{t^{\frac{\beta}{1+2\beta}}|\hat{f}_t(a)-f_0(a)|<C\}}.$$

We conclude that

$$\begin{aligned}
&\sup_{f\in H(\beta,F,L,f_{min})} t^{\frac{2\beta}{1+2\beta}} E_x^f[|\hat{f}_t(a) - f(a)|^2] \\
&\geq \frac{b^2}{8}\mathbb{E}_0\left[1_{\{t^{\frac{\beta}{1+2\beta}}|\hat{f}_t(a)-f_0(a)|\geq\frac{b}{2}\}} + L_t^{f_t/f_0} 1_{\{t^{\frac{\beta}{1+2\beta}}|\hat{f}_t(a)-f_0(a)|<\frac{b}{2}\}}\right] \\
&\geq \frac{b^2}{8}e^{-s}\mathbb{P}_0(L_t^{f_t/f_0} \geq e^{-s}),
\end{aligned}$$

for any $s > 0$. Therefore, in order to achieve the proof of Theorem 33, it suffices to show that

$$\limsup_{t\to\infty} \mathbb{E}_0[|\log L_t^{f_t/f_0}|] < \infty. \tag{3.5.42}$$

Indeed, we can deduce from (3.5.42) the following statements:

$$\begin{aligned}
\exists M, \forall t, \mathbb{E}_0\left(|\log L_t^{f_t/f_0}|\right) \leq M, \\
\exists M, \forall t, \mathbb{P}_0\left(\log L_t^{f_t/f_0} < -2M\right) \leq \frac{1}{2}, \\
\exists s, \forall t, \mathbb{P}_0\left(\log L_t^{f_t/f_0} \geq -s\right) \geq \frac{1}{2}.
\end{aligned}$$

Recall that by construction, $f_t \geq f_0$. Moreover, since the support of χ is included in $[-1,1]$, $f_t(y) \neq f_0(y)$ implies $y \in J_t := [a-h_t, a+h_t]$. Now, Theorem 3.5 of Löcherbach (2002) [Löc02], applied to the particular case without branching, shows that $\mathbb{P}_0$ and $\mathbb{P}_t$ are equivalent on $\mathcal{F}_t$, with density

$$\log L_t^{f_t/f_0} = \int_0^t \int_{J_t} \log(\frac{f_t}{f_0}(y))\mu(ds,dy) - \int_{J_t} \left(\frac{f_t}{f_0} - 1\right)(y) f_0(y)\eta_t(dy). \tag{3.5.43}$$

We now proceed exactly as in [HHL02], proof of Lemma 11. The $\mathbb{P}_0-$ martingale part within (3.5.43) is given by

$$\int_0^t \int_{J_t} \left(\frac{f_t}{f_0} - 1\right)(\mu - \hat{\mu}^{f_0})(ds,dy),$$

where $\hat{\mu}^{f_0}(ds,dy) = \sum_{i=1}^N f_0(X_s^i)\delta_{X_s^i}(dy)ds$ is the $\mathbb{P}_0-$compensator of μ. Its angle bracket is

$$\begin{aligned} b^2 h_t^{2\beta+2} \int_{J_t} \left(\frac{\chi_{h_t}^2(y-a)}{f_0(y)}\right)\eta_t(dy) &\leq \frac{b^2}{\inf_{y\in J_t}(f_0(y))} t^{-\frac{2\beta}{2\beta+1}} h_t^2 \int_{J_t} \chi_{h_t}^2(y-a)\eta_t(dy) \\ &\leq \frac{b^2}{\inf_{y\in J_t}(f_0(y))} t^{-\frac{2\beta}{2\beta+1}} \eta_t(J_t) = \frac{b^2}{\inf_{y\in J_t}(f_0(y))} t^{\frac{1}{2\beta+1}} \frac{1}{t}\eta_t(J_t), \end{aligned}$$

since $\chi(\cdot) \leq 1$, by definition of χ_{h_t} (recall (3.5.40)). All other terms in (3.5.43) are treated exactly as in [HHL02]. Therefore, it only remains to show that

$$\limsup_{t\to\infty} \mathbb{E}_0\left(\frac{1}{th_t}\eta_t(J_t)\right) < \infty. \tag{3.5.44}$$

We apply once more Theorem 30 and rewrite

$$\begin{aligned} \mathbb{E}_0(\eta_t(J_t)) = \int_0^t E_x^{f_0}(\bar{1}_{J_t}(X_s))ds &= \int_0^t E_x^{f_0}\left(\bar{1}_{J_t}(X_s) - \pi^{f_0}(\bar{1}_{J_t})\right))ds + t\pi^{f_0}(\bar{1}_{J_t}) \\ &\leq CN\int_0^t \kappa^{-s}ds + t\pi^{f_0}(\bar{1}_{J_t}) \leq CN\frac{1}{\ln\kappa} + t\pi^{f_0}(\bar{1}_{J_t}) \\ &= CN\frac{1}{\ln\kappa} + Nt\int_{J_t}\pi_1^{f_0}(y)dy, \end{aligned}$$

where $\bar{1}_{J_t}(x) := \sum_{i=1}^N 1_{J_t}(x^i)$ for $x \in \mathbb{R}^N$, and $\pi_1^{f_0}(y)$ denotes the Lebesgue density of $\pi_1^{f_0}$, which exists on J_t by choice of a, for t sufficiently large. Using the change of variables $z = (y-a)/h_t$, we obtain

$$\mathbb{E}_0(\eta_t(J_t)) \leq CN\frac{1}{\ln\kappa} + Nth_t\int_{-1}^1 \pi_1^{f_0}(a+h_tz)dz \leq CN\frac{1}{\ln\kappa} + 2Nth_t \sup_{x\in B_{h_t(a)}} \pi_1^{f_0}(x),$$

which implies finally (3.5.44) by Theorem 31. □

Acknowledgments

We thank an anonymous referee for helpful comments and suggestions. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01), as part of the Agence Nationale de la Recherche PIECE 12-JS01-0006-01 and as part of the activities of FAPESP Research, Dissemination and Innovation Center for Neuromathematics (grant 2013/07699-0, S. Paulo Research Foundation).

Appendices

Annexe A

Calcul d'espérance

On détaille ici le calcul permettant d'obtenir (2.6.51) à partir de (2.6.49) dans le chapitre 2, partie 2.6.3. On pose :

$$A_i = \sum_{k\geq 1} \left|C_{(i,t)}(k)\right| \left(\sum_{j\in V_i(k-1)} |W_{j\to i}| \int_{k-1}^{k} g_j(s) dN_j(s) + \sum_{j\in \partial V_i(k-1)} |W_{j\to i}| \int_{0}^{k} g_j(s) dN_j(s) \right)$$

On décompose $V_i(k)$ en $V_i(k-1)$ plus le bord pour isoler les termes non indépendants

$$A_i = \sum_{k\geq 0} \left[\left(\sum_{j\in \partial V_i(k-1)} \int_0^k dN^j + \sum_{j\in V_i(k-1)} \int_0^k dN^j \right) \sum_{j\in V_i(k-1)} |W_{j\to i}| \int_{k-1}^{k} g_j(s) dN_j(s) \right.$$
$$\left. + \sum_{j\in \partial V_i(k-1)} |W_{j\to i}| \int_{0}^{k} g_j(s) dN_j(s) \right]$$

On continue, chaque fois que c'est nécéssaire, de décomposer selon le temps ou l'espace pour isoler les termes non indépendants. Ce faisant on calcule explicitement l'espérance des produits indépendants par le produit des espérances, ces termes sont déplacés en début d'expression pour garder les termes non indépendants en fin d'expression :

$$E(A_i) = \sum_{k\geq 0} \left[k \left(\sum_{j\in V_i(k-1)} \Lambda_j \right) \left(\sum_{j\in \partial V_i(k-1)} |W_{j\to i}| \Lambda_j \int_0^k g_j(s) ds \right) \right.$$
$$+ k \left(\sum_{j\in \partial V_i(k)} \Lambda_j \right) \left(\sum_{j\in V_i(k-1)} |W_{j\to i}| \Lambda_j \int_{k-1}^{k} g_j(s) ds \right)$$
$$+ E\left[\left(\sum_{j\in \partial V_i(k-1)} \int_0^k dN^j \right) \left(\sum_{j\in \partial V_i(k-1)} |W_{j\to i}| \int_0^k g_j(s) dN_j(s) \right) \right]$$
$$\left. + E\left[\left(\sum_{j\in V_i(k-1)} \int_0^{k-1} dN^j + \sum_{j\in V_i(k-1)} \int_{k-1}^{k} dN^j \right) \sum_{j\in V_i(k-1)} |W_{j\to i}| \int_{k-1}^{k} g_j(s) dN_j(s) \right] \right]$$

$$
\begin{aligned}
E(A_i) = \sum_{k\geq 0} \Bigg[& k \left(\sum_{j\in V_i(k-1)} \Lambda_j \right) \left(\sum_{j\in \partial V_i(k-1)} |W_{j\to i}| \Lambda_j \int_0^k g_j(s) ds \right) \\
& + k \left(\sum_{j\in \partial V_i(k)} \Lambda_j \right) \left(\sum_{j\in V_i(k-1)} |W_{j\to i}| \Lambda_j \int_{k-1}^k g_j(s) ds \right) \\
& + (k-1) \left(\sum_{j\in V_i(k-1)} \Lambda_j \right) \left(\sum_{j\in V_i(k-1)} |W_{j\to i}| \Lambda_j \int_{k-1}^k g_j(s) ds \right) \\
& + \sum_{j,j'\in \partial V_i(k-1): j\neq j'} k \Lambda_{j'} \Lambda_j |W_{j\to i}| \int_0^k g_j(s) ds \\
& + \sum_{j,j'\in V_i(k-1): j\neq j'} \Lambda_j \Lambda_{j'} |W_{j\to i}| \int_{k-1}^k g_j(s) ds \\
& + \sum_{j\in \partial V_i(k-1)} E\left[\left(\int_0^k dN^j \right) \left(|W_{j\to i}| \int_0^k g_j(s) dN_j(s) \right) \right] \\
& + \sum_{j\in V_i(k-1)} E\left[\left(\int_{k-1}^k dN^j \right) |W_{j\to i}| \int_{k-1}^k g_j(s) dN_j(s) \right] \Bigg].
\end{aligned}
$$

Le calcul des deux derniers termes qui sont non-indépendants permet d'écrire l'espérance du produit, comme le produit des espérances plus la covariance. Cela donne

$$
E\left[\left(\int_0^k dN^j \right) \left(|W_{j\to i}| \int_0^k g_j(s) dN_j(s) \right) \right] = \Lambda_j (k\Lambda_j + 1) |W_{j\to i}| \int_0^k g_j(s) ds,
$$

et

$$
E\left[\left(\int_{k-1}^k dN^j \right) |W_{j\to i}| \int_{k-1}^k g_j(s) dN_j(s) \right] = \Lambda_j (\Lambda_j + 1) |W_{j\to i}| \int_{k-1}^k g_j(s) ds.
$$

Cela permet de regrouper tous les produits d'espérance ensemble et de retrouver ainsi (2.6.51).

Annexe B

Mesure de Poisson dominante

Dans la partie 2.4 du chapitre 2, on a défini (voir la définition 19) le processus Z à partir d'une mesure de Poisson aléatoire $N(dt, di, dz)$ sur $\mathbb{R} \times I \times [0,1]$ d'intensité $dt \left(\sum_{i \in I} \Lambda_i \delta_i\right) dz$ sur $\mathbb{R} \times I \times [0,1]$ de la manière suivante :

$$Z^i(C) = \int_C \int_{\{i\}} \int_{[0,1]} 1_{\{z \leq \frac{1}{\Lambda_i} \psi_i \left(\sum_j h_{j \to i} \left(\int_{[L^i_t, t[} g_j(t-s) dZ^j_s \right) \right)\}} N(dt, di, dz). \tag{B.0.1}$$

Cela correspond à la construction annoncée et le lemme suivant nous permet d'affirmer que le processus ainsi construit a l'intensité voulue.

Lemma 54. *Soit E un espace mesurable et μ une mesure de Poisson aléatoire sur $\mathbb{R} \times E$ d'intensité $\nu = dt \otimes \bar{\nu}$. Alors, pour toute fonction $f(t,x,\omega)$ $\mathcal{F}_t$-prévisible,*

$$E\left(\int_t^{t+h} \int_E f(s,x,\omega) \mu(ds, dx) / \mathcal{F}_t \right) = E\left(\int_t^{t+h} \int_E f(s,x,\omega) ds\, \bar{\nu}(dx) / \mathcal{F}_t \right).$$

Démonstration. On commence par démontrer le lemme pour une fonction f du type :

$$f(t,x,\omega) = \sum_{n=1}^{N} \phi_n(x,\omega) 1_{]T_n, T_{n+1}]}(t)$$

avec

$$\phi_n(x,\omega) = \sum_{k=1}^{K} \phi_{n,k}(\omega) 1_{B_{n,k}}(x)$$

où $\phi_{n,k}$ est $\mathcal{F}_{T_n}$-mesurable. On conclut ensuite grâce à un argument de densité.

$$
\begin{aligned}
E\left(\int_t^{t+h}\int_E f(s,x,\omega)\mu(ds,dx)/\mathcal{F}_t\right) &= \sum_{n=1}^{N} E\left(\int_{T_n\vee t}^{T_{n+1}\wedge(t+h)}\int_E \phi_n(x,\omega)\mu(ds,dx)/\mathcal{F}_t\right)\\
&= \sum_{n=1}^{N} E\left(\int_E \phi_n(x,\omega)\mu\left(]T_n\vee t, T_{n+1}\wedge(t+h)],dx\right)/\mathcal{F}_t\right)\\
&= \sum_{n=1}^{N}\sum_{k=1}^{K} E\left(\phi_{n,k}(\omega)\mu\left(]T_n\vee t, T_{n+1}\wedge(t+h)], B_{n,k}\right)/\mathcal{F}_t\right)\\
&= \sum_{n=1}^{N}\sum_{k=1}^{K}\Big(\phi_{n,k}(\omega)\nu\left(]T_n\vee t, T_{n+1}\wedge(t+h)], B_{n,k}\right)1_{t\geq T_n}\\
&\quad + E\left[E\left(\phi_{n,k}(\omega)\mu\left(]T_n\vee t, T_{n+1}\wedge(t+h)], B_{n,k}\right)/\mathcal{F}_{T_n}\right)/\mathcal{F}_t\right]1_{t<T_n}\Big)\\
&= \sum_{n=1}^{N}\sum_{k=1}^{K}\Big(E\left(\phi_{n,k}(\omega)\nu\left(]T_n\vee t, T_{n+1}\wedge(t+h)], B_{n,k}\right)/\mathcal{F}_t\right)1_{t\geq T_n}\\
&\quad + E\left[\phi_{n,k}(\omega)\mu\left(]T_n\vee t, T_{n+1}\wedge(t+h)], B_{n,k}\right)/\mathcal{F}_t\right]1_{t<T_n}\Big)\\
&= E\left(\int_t^{t+h}\int_E f(s,x,\omega)ds\,\bar{\nu}(dx)/\mathcal{F}_t\right)
\end{aligned}
$$

□

Annexe C

Application du corollaire 3.24 de [JS87] chapitre VIII

On détaille ici l'étude de la suite de processus

$$(M^t)_s := \frac{1}{\sqrt{th}} \int_{\mathbb{R}} Q\left(\frac{y-a}{h}\right) \tilde{\mu}_{ts}(dy)$$

dans la preuve du théorème 32 (ii) partie **??**.

On cherche à appliquer le Corollaire 3.24 de [JS87] chapitre VIII.

La mesure de saut associée au processus M^t est donnée par la proposition 1.16 de [JS87] chapitre II :

$$\mu^{M^t}(\omega; du, dx) := \sum_{\tau} 1_{\{\Delta M^t_\tau \neq 0\}}\, \delta_{(\tau, \Delta M^t_\tau(\omega))}(du, dx).$$

Les instants de saut du processus M^t sont du type $\frac{T^i_n}{t}$ où T^i_n est un instant de saut donné par la mesure μ^i. La taille d'un saut à un instant $\sigma = \frac{T^i_n}{t}$ pour un certain neurone i et pour un certain $n \in \mathbb{N}$ est $\frac{1}{\sqrt{th}} Q\left(\frac{X^i_{t\sigma-} - a}{h}\right)$.

On peut ainsi réécrire la mesure de saut μ^{M^t} de la manière suivante :

$$\mu^{M^t}(du, dx) = \sum_{i=1}^{N} \sum_{n \geq 1} 1_{\{T^i_n < \infty\}} \delta_{\left(\frac{T^i_n}{t}, \frac{1}{\sqrt{th}} Q\left(\frac{X^i_{T^i_n -} - a}{h}\right)\right)}(du, dx).$$

Comme pour chaque i les instants de saut T^i_n apparaissent avec une intensité donnée en fonction de la position X^i par la fonction f, le compensateur de cette mesure est :

$$\nu^t(du, dx) = \sum_{i=1}^{N} f(X^i_{tu}) t du \delta_{\left(\frac{1}{\sqrt{th}} Q\left(\frac{X^i_{tu} - a}{h}\right)\right)}(dx).$$

Ainsi,

$$\nu^t_s(dx) = \int_0^s \sum_{i=1}^{N} f(X^i_{tu}) t du \delta_{\left(\frac{1}{\sqrt{th}} Q\left(\frac{X^i_{tu} - a}{h}\right)\right)}(dx) = \int_0^{ts} \sum_{i=1}^{N} f(X^i_u) du \delta_{\left(\frac{1}{\sqrt{th}} Q\left(\frac{X^i_u - a}{h}\right)\right)}(dx).$$

En conséquence, en reprenant les notations de [JS87] pour l'hypothèse 3.23 du théorème 3.22 chapitre VIII, on obtient

$$|x|^2 1_{\{|x|>\varepsilon\}} * \nu_s^t = \frac{1}{th} \int_0^{ts} \sum_{i=1}^{N} f\left(X_u^i\right) Q^2\left(\frac{X_u^i - a}{h}\right) 1_{\left\{\frac{1}{\sqrt{th}} Q\left(\frac{X_u^i - a}{h}\right) > \varepsilon\right\}} du.$$

Comme montré dans la partie **??**, ceci converge presque sûrement vers 0, et l'hypothèse 3.23 du chapitre VIII de [JS87] est donc vérifiée.

Le processus M^t étant défini comme une intégrale contre une mesure aléatoire compensée, le théorème 1.33 du chapitre II de [JS87] nous donne

$$\left\langle M^t, M^t \right\rangle_s = \frac{1}{th} \int_0^{ts} \int_{\mathbb{R}} Q^2\left(\frac{y-a}{h}\right) f(y) \eta(du, dy).$$

La proposition **??** nous donne la convergence en probabilité de $\left\langle M^t, M^t \right\rangle_s$ vers $Nsf(a)\pi_1(a) \int Q^2(x)dx$ lorsque t tend vers $+\infty$. Ceci nous permet de conclure en appliquant le Corollaire 3.24 de [JS87] chapitre VIII.

Bibliographie

[ABG+14] Romain Azaïs, Jean-Baptiste Bardet, Alexandre Génadot, Nathalie Krell, and Pierre-André Zitt. Piecewise deterministic Markov process—recent results. In *Journées MAS 2012*, volume 44 of *ESAIM Proc.*, pages 276–290. EDP Sci., Les Ulis, 2014.

[ADGP14] Romain Azaïs, François Dufour, and Anne Gégout-Petit. Non-parametric estimation of the conditional distribution of the interjumping times for piecewise-deterministic Markov processes. *Scand. J. Stat.*, 41(4) :950–969, 2014.

[AMG15] R. Azaïs and A. Muller-Gueudin. Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic markov processes. 2015. Available on http ://arxiv.org/abs/1506.07722.

[BCG+13] J.-B. Bardet, A. Christen, A. Guillin, F. Malrieu, and P.-A. Zitt. Total variation estimates for the tcp process. *Electronic Journal of Probability*, 18(10) :1–21, 2013.

[BL12] N. Berglund and D. Landon. Mixed-mode oscillations and interspike interval statistics in the stochastic fitzhugh-nagumo model. *Nonlinearity*, 25 :2303–2335, 2012.

[BM96] Pierre Brémaud and Laurent Massoulié. Stability of nonlinear Hawkes processes. *Ann. Probab.*, 24(3) :1563–1588, 1996.

[CCDRB15] Julien Chevallier, María José Cáceres, Marie Doumic, and Patricia Reynaud-Bouret. Microscopic approach of a time elapsed neural model. *Mathematical Models and Methods in Applied Sciences*, 25(14) :2669–2719, 2015.

[CD01] Francis Comets and Amir Dembo. Ordered overlaps in disordered mean-field models. *Probab. Theory Related Fields*, 121(1) :1–29, 2001.

[CDMR12] A. Crudu, A. Debussche, A. Muller, and O. Radulescu. Convergence of stochastic gene networks to hybrid piecewise deterministic processes. *Ann. Appl. Probab.*, 22(5) :1822–1859, 2012.

[Ces11] B. Cessac. A discrete time neural network model with spiking neurons : Ii : Dynamics with noise. *Journal of Mathematical Biology*, 62 :863–900, 2011.

[Che15] Julien Chevalier. Mean-field limit of generalized hawkes processes. 2015. Preprint at http ://arxiv.org/abs/1510.05620.

[Dav84] M. H. A. Davis. Piecewise-deterministic Markov processes : a general class of nondiffusion stochastic models. *J. Roy. Statist. Soc. Ser. B*, 46(3) :353–388, 1984. With discussion.

[Dav93] M. H. A. Davis. *Markov models and optimization*, volume 49 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1993.

[DFH16] Sylvain Delattre, Nicolas Fournier, and Marc Hoffmann. Hawkes processes on large networks. *Ann. Appl. Probab.*, 26(1) :216–261, 2016.

[DGLO16] A. Duarte, A. Galves, E. Löcherbach, and G. Ost. Absolute continuity of the invariant measure in piecewise deterministic markov processes having degenerate jumps. 2016. Available on https ://arxiv.org/abs/1604.00419.

[DIRT15a] F. Delarue, J. Inglis, S. Rubenthaler, and E. Tanré. Global solvability of a networked integrate-and-fire model of mckean-vlasov type. 2015. To appear in Ann. Appl. Probab., available on https ://arxiv.org/abs/1211.0299.

[DIRT15b] F. Delarue, J. Inglis, S. Rubenthaler, and E. Tanré. Particle systems with a singular mean-field self-excitation. application to neuronal networks. *Stoch. Proc. Appl.*, 125 :2451–2492, 2015.

[DL07] S. Ditlevsen and P. Lansky. Parameters of stochastic diffusion processes estimated from observations of first-hitting times : Application to the leaky integrate-and-fire neuronal model. *Phys. Rev. E*, 76, 2007.

[DL16] S. Ditlevsen and E. Löcherbach. Multiclass oscillating systems of interacting neurons. 2016. To appear in Stoch. Proc. Appl, available on http ://arxiv.org/abs/1512.00265.

[DO14] A. Duarte and G. Ost. A model for neural activity in the absence of external stimuli. 2014. Available on http ://arxiv.org/abs/1410.6086.

[Dua15] Aline Duarte. Stochastic models in neurobiology : from a multiunitary regime to eeg data. 2015. PhD thesis, Sao Paulo.

[DVJ03] 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.

[Fit61] R. FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1 :445–466, 1961.

[FL14] N. Fournier and E. Löcherbach. On a toy model for interacting neurons. 2014. To appear in *Annales I.H.P.* See also http ://arxiv.org/abs/1410.3263.

[FMMN00] Pablo A. Ferrari, Alejandro Maass, Servet Martínez, and Peter Ney. Cesàro mean distribution of group automata starting from measures with summable decay. *Ergodic Theory Dynam. Systems*, 20(6) :1657–1670, 2000.

[GK02] W. Gerstner and W. Kistler. *Spiking Neuron Models : An Introduction.* Cambridge University Press, 2002.

[GL13] A. Galves and E. Löcherbach. Infinite systems of interacting chains with memory of variable length—a stochastic model for biological neural nets. *J. Stat. Phys.*, 151(5) :896–921, 2013.

[GL16] A. Galves and E. Löcherbach. Modelling networks of spiking neurons as interacting processes with memory of variable length. 2016. Available on http ://arxiv.org/abs/1502.06446.

[Haw72] A. G. Hawkes. Spectra of some mutually exciting point processes with associated variables. In *Stochastic point processes : statistical analysis, theory, and applications (Conf., IBM Res. Center, Yorktown Heights, N.Y., 1971)*, pages 261–271. Wiley-Interscience, New York, 1972.

[HH52] A. L. Hodgkin and A. F. Huxsley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4) :500–544, 1952.

[HHL02] R. Höpfner, M. Hoffmann, and E. Löcherbach. Non-parametric estimation of the death rate in branching diffusions. *Scand. J. Statist.*, 29(4) :665–692, 2002.

[HL14] P. Hodara and E. Löcherbach. Hawkes processes with variable length memory and an infinite number of components. 2014. Available on http ://arxiv.org/abs/1410.5235.

[HO] M. Hoffmann and A. Olivier. Nonparametric estimation of the division rate of an age dependent branching process. To appear in *Stochastic Processes and their Applications* see also http ://arxiv.org/abs/1412.5936.

[HO74] Alan G. Hawkes and David Oakes. A cluster process representation of a self-exciting process. *J. Appl. Probability*, 11 :493–503, 1974.

[HRBR15] Niels Richard Hansen, Patricia Reynaud-Bouret, and Vincent Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1) :83–143, 2015.

[IT14] 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. 2014. Available on http ://arxiv.org/abs/1409.8221.

[IW89] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library.* North-Holland Publishing Co., Amsterdam ; Kodansha, Ltd., Tokyo, second edition, 1989.

[Izh06] E. Izhikevich. *Dynamical systems in neuroscience : the geometry of excitability and bursting.* MIT Press, 2006.

[JR15] Thibault Jaisson and Mathieu Rosenbaum. Limit theorems for nearly unstable Hawkes processes. *Ann. Appl. Probab.*, 25(2) :600–631, 2015.

[JS87] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1987.

[Kre14] N. Krell. Statistical estimation of jump rates for a specific class of piecewise deterministic markov processes. 2014. Available on http ://arxiv.org/abs/1406.2845.

[LBD16] K. Li, C. Bundesen, and S. Ditlevsen. Responses of leaky integrate-and-fire neurons to a plurality of stimuli in their receptive fields. *The Journal of Mathematical Neuroscience*, 6(8), 2016.

[Lig05] Thomas M. Liggett. *Interacting particle systems*. Classics in Mathematics. Springer-Verlag, Berlin, 2005. Reprint of the 1985 original.

[Löc02] E. Löcherbach. Likelihood ratio processes for Markovian particle systems with killing and jumps. *Stat. Inference Stoch. Process.*, 5(2) :153–177, 2002.

[Löc15] E. Löcherbach. Ergodicity and speed of convergence to equilibrium for diffusion processes. 2015. Unpublished note, available on http ://eloecherbach.u-cergy.fr/cours.pdf.

[Löc16] E. Löcherbach. Absolute continuity of the invariant measure in piecewise deterministic markov processes having degenerate jumps. 2016. Available on https ://arxiv.org/abs/1601.07123.

[LST95] P. Lansky, L. Sacerdote, and F. Tomassetti. On the comparison of feller and ornstein-uhlenbeck models for neuronal activity. *Biological Cybernetics*, 73 :457–465, 1995.

[Mas98] Laurent Massoulié. Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic Process. Appl.*, 75(1) :1–30, 1998.

[MR05] Jesper Møller and Jakob G. Rasmussen. Perfect simulation of Hawkes processes. *Adv. in Appl. Probab.*, 37(3) :629–646, 2005.

[PTW10] K. Pakdaman, M. Thieullen, and G. Wainrib. Fluid limit theorems for stochastic hybrid systems with application to neuron models. *Adv. in Appl. Probab.*, 42(3) :761–794, 2010.

[RBRGTM14] Patricia Reynaud-Bouret, Vincent Rivoirard, Franck Grammont, and Christine Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *Journal of Mathematical Neuroscience*, page 4 :3, April 2014.

[RBS10] Patricia Reynaud-Bouret and Sophie Schbath. Adaptive estimation for Hawkes processes ; application to genome analysis. *Ann. Statist.*, 38(5) :2781–2822, 2010.

[Yag16] K. Yaginuma. A stochastic system with infinite interacting components to model the time evolution of the membrane potentials of a population of neurons. 2016. To appear in Journal of Statistical Physics, available on https ://arxiv.org/abs/1505.00045.