Stochastic mean-field dynamics and applications to life sciences

Paolo Dai Pra, Università di Padova

April 4, 2017

Contents

1 Introduction **1**

2 The prototypical model **2**

3 Propagation of chaos for interacting systems **4**
3.1 The microscopic model 4
3.2 The macroscopic limit 5
3.3 Well posedness of the microscopic model: Lipschitz conditions 5
3.4 Well posedness of the macroscopic limit 6
3.5 Propagation of chaos 9

4 Applications **10**
4.1 The stochastic Kuramoto model 10
4.2 Interacting Fitzhugh-Nagumo neurons 12
4.3 Interacting Hawkes processes 14

5 Further reading **15**
5.1 Long-time behavior of the microscopic system 15
5.2 Fluctuations 16
5.3 Critical fluctuations 16
5.4 Large Deviations 16
5.5 Generalizing network's microscopic geometry 16
5.6 Mean-field games 17

1 Introduction

Although we do not intend to give a general, formal definition, the stochastic mean-field dynamics we present in these notes can be conceived as random evolution of a system comprised by N interacting components which is: a) invariant in law for permutation of the components; b) the contribution of each component to the evolution of any other is of order $\frac{1}{N}$. The permutation invariance clearly does not allow any freedom in the choice of the geometry of the interaction; however, this is exactly the feature that makes these models analytically treatable, and therefore attractive for a wide scientific community.
Originally designed as toy models in Statistical Mechanics, the emergence of applications in which the interaction is typically of very long range and not determined by fundamental laws, have renewed the interest in models of this sort. Applications include, in particular, *Life Sciences* and *Social Sciences*.
The goal of these lectures is to

- review some of the basic techniques allowing to derive the macroscopic limit of a mean-field model, and provide quantitative estimates on the rate of convergence;
- illustrate, without technical details, some applications relevant to life sciences, in particular for what concerns the study of the properties of the macroscopic limit.

Mainly inspired by [43], we introduce the topic by some heuristics on a simple class of models

2 The prototypical model

Consider a system of N interacting diffusions on $\mathbb{R}^d$ solving the following system of SDE:

$$dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^N b(X_t^{i,N}, X_t^{j,N})dt + dW_t^i$$

where $b : \mathbb{R}^d \times \mathbb{R}^d$ is a Lipschitz function, $(W^i)_{i\geq 1}$ are independent Brownian motions, and we assume $(X_0^{i,N})_{i=1}^N$ are i.i.d square integrable random variables. In particular, the dynamical equation is well posed.
If we consider one component $X^{i,N}$, assume $X_0^{i,N} = X_0^i$ does not depend on N, let $N \to +\infty$ and "believe in laws of large numbers", it is natural to guess that $X^{i,N}$ converges, as $N \to +\infty$, to a limit process $\overline{X}^i$ solving

$$\begin{aligned} d\overline{X}_t^i &= \int b(\overline{X}_t^i, y)q_t(dy)dt + dW_t^i \\ \overline{X}_0^i &= X_0^i \end{aligned} \tag{2.1}$$

where $q_t = Law(\overline{X}_t^i)$. Once the nontrivial problem of well posedness of this las equation is settled, one aims at showing that for any given $T > 0$ and indicating by $X_{[0,T]} \in \mathcal{C}([0,T])$ the whole trajectory up to time T: for any $m \geq 1$

$$(X_{[0,T]}^{1,N}, X_{[0,T]}^{2,N}, \dots, X_{[0,T]}^{m,N}) \to (\overline{X}_{[0,T]}^1, \overline{X}_{[0,T]}^2, \dots, \overline{X}_{[0,T]}^m)$$

in distribution as $N \to +\infty$. Note that the components of the process $(\overline{X}_{[0,T]}^1, \overline{X}_{[0,T]}^2, \dots, \overline{X}_{[0,T]}^m)$ are independent. Thus, independence at time 0 propagates in time, at least in the macroscopic limit $N \to +\infty$. This property is referred to as *propagation of chaos*.
Propagation of chaos can be actually rephrased as a *Law of Large Numbers*. To this aim, given a generic vector $\underline{x} = (x_1, x_2, \dots, x_N)$, denote by $\rho_N(\underline{x}; dy) := \frac{1}{N}\sum_{i=1}^N \delta_{x_i}(dy)$ the corresponding empirical measure. The propagation of chaos property stated above, is equivalent to the fact that the sequence of empirical measures $\rho_N(\underline{X}_{[0,T]}^N)$ converges in distribution to $Q \in \mathcal{P}(\mathcal{C}([0,T]))$, where $\mathcal{P}(\mathcal{C}([0,T]))$ denotes the set of probabilities on $\mathcal{C}([0,T])$ provided with the topology of weak convergence and Q is the law of the solution of (2.1). This is established in the following result.

Proposition 2.1. *Let $(X^{i,N} : N \geq 1, 1 \leq i \leq N)$ be a triangular array of random variables taking values in a topological space E, such that for each N the law of $(X^{i,N})_{1\leq i\leq N}$ is symmetric (i.e. invariant by permutation of components). Moreover let $(\overline{X}^i)_{i\geq 1}$ be a i.i.d. sequence of E-valued random variables. Then the following statements are equivalent:*

(a) for every $m \geq 1$

$$(X^{1,N}, X^{2,N}, \dots X^{m,N}) \to (\overline{X}^1, \overline{X}^2, \dots, \overline{X}^m)$$

in distribution as $N \to +\infty$;

(b) the sequence of empirical measures $\rho_N(\underline{X}^N)$ converges in distribution to $Q := Law(\overline{X}^1)$ as $N \to +\infty$.

Proof. Denote by Q_N the joint law of $(X^{1,N}, X^{2,N}, \ldots X^{N,N})$ in E^N, and by $\Pi_m Q_N$ its projection on the first m components, i.e. the law of $(X^{1,N}, X^{2,N}, \ldots X^{m,N})$. The statements in (a) is equivalent to: for each $m \geq 1$

$$\Pi_m Q_N \to Q^{\otimes m} \tag{2.2}$$

weakly, where $Q^{\otimes m}$ is the m-fold product of Q.

(a) ⇒ (b).

To begin with, let $F : E \to \mathbb{R}$ be bounded and continuous. Writing $\langle F, \mu \rangle$ for $\int F d\mu$ and denoting by $\mathbb{E}^{Q_N}$ the expectation w.r.t. Q_N:

$$\begin{aligned}
\mathbb{E}^{Q_N}\left(\langle F, \rho_N(\underline{x}) - Q\rangle^2\right) &= \frac{1}{N^2}\sum_{i,j=1}^N \mathbb{E}^{Q_N}\left[F(x_i)F(x_j)\right] - \frac{2}{N}\langle F, Q\rangle \sum_{i=1}^N \mathbb{E}^{Q_N}[F(x_i)] + \langle F, Q\rangle^2 \\
&= \frac{1}{N}\mathbb{E}^{Q_N}[F^2(x_1)] + \frac{N-1}{N}\mathbb{E}^{Q_N}[F(x_1)F(x_2)] - 2\langle F, Q\rangle \mathbb{E}^{Q_N}[F(x_1)] + \langle F, Q\rangle^2,
\end{aligned}$$

where we have used the symmetry of Q_N. By Assumption (a) this last expression goes to zero as $N \to +\infty$.

Now, let $\Phi : \mathcal{P}(E) \to \mathbb{R}$ be continuous and bounded. By definition of weak topology, given $\epsilon > 0$ one can find $\delta > 0$ and $F_1, \ldots F_k : E \to \mathbb{R}$ bounded and continuos such that if

$$U := \{P \in \mathcal{P}(E) : |\langle P - Q, F_j\rangle| < \delta \text{ for } j = 1, \ldots, k\}$$

then $P \in U$ implies $|\Phi(P) - \Phi(Q)| < \epsilon$. Thus

$$\left|\mathbb{E}^{Q_N}[\Phi(\rho_N(\underline{x})] - \Phi(Q)\right| \leq \epsilon Q_N(\rho_N(\underline{x}) \in U) + \|\Phi\|_\infty Q_N(\rho_N(\underline{x}) \notin U).$$

Therefore, to show (b), i.e. $\left|\mathbb{E}^{Q_N}[\Phi(\rho_N(\underline{x})] - \Phi(Q)\right| \to 0$ for every Φ bounded and continuous, it is enough to show that

$$\lim_{N \to +\infty} Q_N(\rho_N(\underline{x}) \in U) = 0.$$

But, by what seen above and the Markov inequality,

$$Q_N(\rho_N(\underline{x}) \notin U) \leq \sum_{j=1}^k Q_N(|\langle \rho_N(\underline{x}) - Q, F_j\rangle| \geq \delta) \leq \sum_{j=1}^k \frac{\mathbb{E}^{Q_N}\left(\langle F_j, \rho_N(\underline{x}) - Q\rangle^2\right)}{\delta^2} \to 0.$$

(b) ⇒ (a).

It is enough to show that if $F_1, F_2, \ldots, F_m : E \to \mathbb{R}$ are bounded and continuous, then

$$\mathbb{E}^{Q_N}\left[F_1(x_1) \cdot F_2(x_2) \cdots F_m(x_m)\right] \to \prod_{j=1}^m \mathbb{E}^Q[F_j(x)] \tag{2.3}$$

Observe that

$$\begin{aligned}
&\left|\mathbb{E}^{Q_N}\left[F_1(x_1) \cdot F_2(x_2) \cdots F_m(x_m)\right] - \prod_{j=1}^m \mathbb{E}^Q[F_j(x)]\right| \\
&\qquad \leq \left|\mathbb{E}^{Q_N}\left[F_1(x_1) \cdot F_2(x_2) \cdots F_m(x_m)\right] - \mathbb{E}^{Q_N}\left[\prod_{j=1}^m \langle \rho_N(\underline{x}), F_j\rangle\right]\right| \\
&\qquad\qquad + \left|\mathbb{E}^{Q_N}\left[\prod_{j=1}^m \langle \rho_N(\underline{x}), F_j\rangle\right] - \prod_{j=1}^m \mathbb{E}^Q[F_j(x)]\right|
\end{aligned} \tag{2.4}$$

By (b), the last summand converges to 0. Using symmetry

$$\begin{aligned}
\mathbb{E}^{Q_N}\left[\prod_{j=1}^{m}\langle\rho_N(\underline{x}),F_j\rangle\right] &= \frac{1}{N^m}\mathbb{E}^{Q_N}\left[\sum_{\tau:\{1,\ldots,m\}\to\{1,\ldots,N\}}\prod_{j=1}^{m}F_j(x_{\tau(j)})\right] \\
&= \frac{D_{N,m}}{N^m}\mathbb{E}^{Q_N}\left[F_1(x_1)\cdot F_2(x_2)\cdots F_m(x_m)\right] \\
&\quad + \frac{1}{N^m}\mathbb{E}^{Q_N}\left[\sum_{\tau \text{ not injective}}\prod_{j=1}^{m}F_j(x_{\tau(j)})\right],
\end{aligned}$$

where $D_{N,m} = \frac{N!}{(N-k)!}$ is the number of injective functions $\{1,\ldots,m\}\to\{1,\ldots,N\}$. Since $\frac{D_{N,m}}{N^m}\to 1$, we obtain

$$\mathbb{E}^{Q_N}\left[\prod_{j=1}^{m}\langle\rho_N(\underline{x}),F_j\rangle\right]\to\mathbb{E}^{Q_N}\left[F_1(x_1)\cdot F_2(x_2)\cdots F_m(x_m)\right]$$

which, by (2.4), completes the proof.

In view of Proposition 2.1, the empirical measure at time t, $\rho_N(\underline{X}_t^N)$ converges in distribution to $q_t = Law(\overline{X}_t^1)$, for every $t\geq 0$. Moreover, being the law of the solution of (2.1), q_t solves the so-called *McKean-Vlasov equation*

$$\frac{\partial}{\partial t}q_t - \nabla\left[q_t\int b(\,\cdot\,,y)q_t(dy)\right]+\frac{1}{2}\Delta q_t = 0.$$

3 Propagation of chaos for interacting systems

3.1 The microscopic model

In this section we introduce a wide class of $\mathbb{R}^d$-valued interacting dynamics, which includes the prototypical model above. The main aim is to introduce *quenched disorder*, which accounts for inhomogeneities in the system, and jumps in the dynamics, which allows to include processes with discrete state space. The dynamics is determined by the following characteristics.

- "Local" parameters $(h_i)_{i=1}^N$, drawn independently from a distribution μ on $\mathbb{R}^{d'}$ *with compact support.*
- A *drift* $b(x_i,h_i;\rho_N(\underline{x},\underline{h}))$, where

  $$\rho_N(\underline{x},\underline{h}) = \frac{1}{N}\sum_{i=1}^{N}\delta_{(x_i,h_i)},$$

  and

  $$b:\mathbb{R}^d\times\mathbb{R}^{d'}\times\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^{d'})\to\mathbb{R}^d.$$

- A diffusion coefficient $\sigma(x_i,h_i;\rho_N(\underline{x},\underline{h}))$

  $$\sigma:\mathbb{R}^d\times\mathbb{R}^{d'}\times\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^{d'})\to\mathbb{R}^{d\times n},$$

  where n is the dimension of the driving Brownian Motion.
- A jump rate $\lambda(x_i,h_i;\rho_N(\underline{x},\underline{h}))$ with

  $$\lambda:\mathbb{R}^d\times\mathbb{R}^{d'}\times\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^{d'})\to[0,+\infty).$$

- A distribution for the jump $f(x_i, h_i; \rho_N(\underline{x}, \underline{h}); v)\alpha(dv)$ with

$$f : \mathbb{R}^d \times \mathbb{R}^{d'} \times \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^{d'}) \times [0,1] \to \mathbb{R}^d$$

  and $\alpha(dv)$ is a probability on $[0,1]$.

The dynamics could be introduced via generator and semigroup, but it will be convenient to use the language of Stochastic Differential Equations (SDE). So let $(W^i)_{i\geq 1}$ be a i.i.d. sequence of n-dimensional Brownian motions; $(N^i(dt, du, dv))_{i\geq 1}$ be i.i.d. Poisson random measures on $[0,+\infty) \times [0,+\infty) \times [0,1]$ with characteristic measure $dt \otimes du \otimes \alpha(dv)$. The microscopic model is given as solution of the SDE for *every given realization of the local parameters* (h_i):

$$\begin{aligned} X_t^{i,N} =& X_0^i + \int_0^t b\left(X_s^{i,N}, h_i, \rho(\underline{X}_s^N, \underline{h})\right) ds + \int_0^t \sigma\left(X_s^{i,N}, h_i, \rho(\underline{X}_s^N, \underline{h})\right) dW_s^i \\ &+ \int_{[0,t]\times[0,+\infty)\times[0,1]} f\left(X_{s^-}^{i,N}, h_i; \rho_N(\underline{X}_{s^-}^N, \underline{h}); \alpha\right) \mathbf{1}_{[0,\lambda(X_{s^-}^{i,N}, h_i, \rho(\underline{X}_{s^-}^N, \underline{h}))]}(u) N^i(ds, du, dv) \end{aligned} \tag{3.1}$$

It will be assumed, without further notice, that the initial states X_0^i are i.i.d., square integrable, independent of both the local parameters (h_i) and of the driving noises (W^i, N^i).

3.2 The macroscopic limit

At heuristic level it is not hard to identify the limit of a given component $X^{i,N}$ of (3.1) subject to a local field h. We omit the apex i on the process and of the driving noises

$$\begin{aligned} \overline{X}_t(h) =& \overline{X}_0 + \int_0^t b\left(\overline{X}_s(h), h, r_s\right) ds + \int_0^t \sigma\left(\overline{X}_s(h), h, r_s\right) dW_s \\ &+ \int_{[0,t]\times[0,+\infty)\times[0,1]} f\left(\overline{X}_{s^-}(h), h; r_s; \alpha\right) \mathbf{1}_{[0,\lambda(\overline{X}_{s^-}(h), h, r_s)]}(u) N(ds, du, dv) \end{aligned} \tag{3.2}$$

where $r_s = Law(\overline{X}_s(h)) \otimes \mu(dh)$. Choosing $\overline{X}_0 = X_0^i$, and driving noises W^i, N^i, we indicate by $\overline{X}^i$ the corresponding solution (3.2).

3.3 Well posedness of the microscopic model: Lipschitz conditions

We now give conditions that guarantee well posedness of (3.1) and (3.2); they are far from being optimal, but allow a reasonable economy of notations. Weaker conditions can be found, for instance in [1]. It is useful to work with probability measures possessing mean value:

$$\mathcal{P}^1(\mathbb{R}^d) := \left\{ \nu \in \mathcal{P}(\mathbb{R}^d) : \int |x| \nu(dx) < +\infty \right\}$$

which is provided with the *Wasserstein metric*

$$d(\nu, \nu') := \inf \left\{ \int |x-y| \Pi(dx, dy) : \Pi \text{ has marginals } \nu \text{ and } \nu' \right\}.$$

- **[L1]** The function $b(x,h,r)$ and $\sigma(x,h,r)$, defined in $\mathbb{R}^d \times \mathbb{R}^{d'} \times \mathcal{P}^1(\mathbb{R}^d \times \mathbb{R}^{d'})$ are continuous, and globally Lipschitz in (x,r) uniformly in h.
- **[L2]** The Lipschitz condition of the jumps is slightly less obvious. We assume $f : \mathbb{R}^d \times \mathbb{R}^{d'} \times \mathcal{P}^1(\mathbb{R}^d \times \mathbb{R}^{d'}) \times [0,1] \to \mathbb{R}^d$ and $\lambda : \mathbb{R}^d \times \mathbb{R}^{d'} \times \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^{d'}) \to [0,+\infty)$ are continuous, and obey the following condition

$$\int \left| f(x,h,r,v) \mathbf{1}_{[0,\lambda(x,h,r)]}(u) - f(y,h,r',v) \mathbf{1}_{[0,\lambda(x,h,r)]}(u) \right| du \alpha(dv) \leq L \left[|x - x'| + d(r,r') \right] \tag{3.3}$$

  for all x, y, r, r', h.

Remark 3.1. The above assumptions imply that when one replaces r by the empirical measure $\rho_N(\underline{x}, \underline{h})$, one recovers a Lipschitz condition in $\underline{x}$. For instance, the function $b(x_i, h_i; \rho_N(\underline{x}, \underline{h}))$ is globally Lipschitz in $\underline{x}$ uniformly in $\underline{h}$.

Remark 3.2. Continuity, global Lipscitzianity and compactness of the support of μ imply the linear growth conditions

$$\begin{aligned}|b(x,h,r)| &\leq C\left[1+|x|+\int |y| r(dy,dh)\right]\\ |s(x,h,r)| &\leq C\left[1+|x|+\int |y| r(dy,dh)\right]\\ \int |f(x,h,r,v)|\lambda(x,h,r)\alpha(dv) &\leq C\left[1+|x|+\int |y| r(dy,dh)\right].\end{aligned} \tag{3.4}$$

Remark 3.3. Condition **L2** is satisfied if both f and λ are continuous, bounded and globally Lipschitz in x, r uniformly of the other variables. In the case f does not depend on x, r but on h, v only, unbounded Lipschitz jump rate λ can be afforded.

Using Remark 3.1, together with standard methods in stochastic analysis, one obtains the following result. A detailed proof can be found e.g. in [29].

Proposition 3.1. *Under* **L1** *and* **L2***, the system* (3.1) *admits a unique strong solution.*

3.4 Well posedness of the macroscopic limit

The proof of the convergence of one component of (3.1) toward a solution of (3.2) allows two alternative strategies. One consists in: (a) showing tightness of the sequence of microscopic processes; (b) showing that any limit point solves weakly (3.2); (c) showing that for (3.2) uniqueness in law holds true. We rather follow the following approach, which is somewhat simpler and allows for quantitative error estimates: (a) we show that (3.2) is well posed; (b) by a coupling argument we show L^1-convergence of one component of (3.1) to a solution of (3.2) driven by the *same noise.*

Proposition 3.2. *Under* **L1** *and* **L2***, the system* (3.2) *admits a unique strong solution.*

Proof. We sketch the proof of existence. We use a standard Picard iteration. Define $X_t^{(0)}(h) \equiv \overline{X}_0$ and

$$\begin{aligned}X_t^{(k+1)}(h) =& \overline{X}_0 + \int_0^t b\left(X_s^{(k)}(h), h, r_s^{(k)}\right) ds + \int_0^t \sigma\left(X_s^{(k)}(h), h, r_s^{(k)}\right) dW_s\\ &+ \int_{[0,t]\times[0,+\infty)\times[0,1]} f\left(X_{s^-}^{(k)}(h), h; r_s^{(k)}; \alpha\right) \mathbf{1}_{\left[0,\lambda\left(X_{s^-}^{(k)}(h),h,r_s^{(k)}\right)\right]}(u) N(ds,du,dv)\end{aligned} \tag{3.5}$$

where

$$r_s^{(k)} = Law\left(X_s^{(k)}(h)\right) \otimes \mu(dh).$$

We estimate

$$E_T^{(k)} := \int \mathbb{E}\left[\sup_{t\in[0,T]} \left|X_t^{(k+1)}(h) - X_t^{(k)}(h)\right|\right] \mu(dh). \tag{3.6}$$

If we use (3.5) and subtract the equations for $X^{(k+1)}$ and $X^{(k)}$, take the $\sup_{t\in[0,T]}$ and use the triangular inequality, we obtain the sum of three terms.

(A). The first term comes from the drift.

$$\begin{aligned}
&\sup_{t\in[0,T]}\left|\int_0^t b\left(X_s^{(k)}(h),h,r_s^{(k)}\right)ds-\int_0^t b\left(X_s^{(k-1)}(h),h,r_s^{(k-1)}\right)ds\right|\\
&\qquad\le \int_0^T \left|b\left(X_s^{(k)}(h),h,r_s^{(k)}\right)-b\left(X_s^{(k-1)}(h),h,r_s^{(k-1)}\right)\right|ds\\
&\qquad\le L\int_0^T\left(\left|X_s^{(k)}(h)-X_s^{(k-1)}(h)\right|+d(r_s^{(k)},r_s^{(k-1)})\right)ds\\
&\qquad\le L\int_0^T\left(\left|X_s^{(k)}(h)-X_s^{(k-1)}(h)\right|+\int\mathbb{E}\left|X_s^{(k)}(h')-X_s^{(k-1)}(h')\right|\mu(dh')\right)ds
\end{aligned}$$

where the inequality

$$d(r_s^{(k)},r_s^{(k-1)})\le\int\mathbb{E}\left|X_s^{(k)}(h')-X_s^{(k-1)}(h')\right|\mu(dh') \tag{3.7}$$

comes directly form the definition of the metric d, and we have used (**L1**). Averaging:

$$\begin{aligned}
&\int\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t b\left(X_s^{(k)}(h),h,r_s^{(k)}\right)ds-\int_0^t b\left(X_s^{(k-1)}(h),h,r_s^{(k-1)}\right)ds\right|\right]\mu(dh)\\
&\qquad\le 2L\int_0^T\int\mathbb{E}\left|X_s^{(k)}(h)-X_s^{(k-1)}(h)\right|\mu(dh)\le 2LTE_T^{(k-1)}.
\end{aligned}$$

(B). The second term comes from the diffusion coefficient.

$$\sup_{t\in[0,T]}\left|\int_0^t\sigma\left(X_s^{(k)}(h),h,r_s^{(k)}\right)ds-\int_0^t\sigma\left(X_s^{(k-1)}(h),h,r_s^{(k-1)}\right)dW_s\right|.$$

By the L^1 Burkholder-Davis-Gundy inequality (see e.g. [39])

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t\left[\sigma\left(X_s^{(k)}(h),h,r_s^{(k)}\right)-\sigma\left(X_s^{(k-1)}(h),h,r_s^{(k-1)}\right)\right]dW_s\right|\right]\\
&\qquad\le C\mathbb{E}\left[\left(\int_0^T\left|\sigma\left(X_s^{(k)}(h),h,r_s^{(k)}\right)-\sigma\left(X_s^{(k-1)}(h),h,r_s^{(k-1)}\right)\right|^2ds\right)^{\frac12}\right]\\
&\qquad\le CL\mathbb{E}\left[\left(\int_0^T\left(\left|X_s^{(k)}(h)-X_s^{(k-1)}(h)\right|+d(r_s^{(k)},r_s^{(k-1)})\right)^2ds\right)^{\frac12}\right]\\
&\qquad\le CL\sqrt{T}\mathbb{E}\left[\sup_{s\in[0,T]}\left(\left|X_s^{(k)}(h)-X_s^{(k-1)}(h)\right|+d(r_s^{(k)},r_s^{(k-1)})\right)ds\right]
\end{aligned}$$

Averaging over h and using (3.7) as before, we obtain

$$\int\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t\left[\sigma\left(X_s^{(k)}(h),h,r_s^{(k)}\right)-\sigma\left(X_s^{(k-1)}(h),h,r_s^{(k-1)}\right)\right]dW_s\right|\right]\mu(dh)\le 2CL\sqrt{T}E_T^{(k-1)}.$$

(C). Finally, we have the term coming from the jumps.

$$\begin{aligned}
&\sup_{t\in[0,T]}\left|\int_{[0,t]\times[0,+\infty)\times[0,1]}f\left(X_{s^-}^{(k)}(h),h;r_s^{(k)};v\right)\mathbf{1}_{\left[0,\lambda\left(X_{s^-}^{(k)}(h),h,r_s^{(k)}\right)\right]}(u)N(ds,du,dv)\right.\\
&\left.-\int_{[0,t]\times[0,+\infty)\times[0,1]}f\left(X_{s^-}^{(k-1)}(h),h;r_s^{(k-1)};v\right)\mathbf{1}_{\left[0,\lambda\left(X_{s^-}^{(k-1)}(h),h,r_s^{(k-1)}\right)\right]}(u)N(ds,du,dv)\right|
\end{aligned} \tag{3.8}$$

Let

$$F_s^k := f\left(X_{s^-}^{(k)}(h), h; r_s^{(k)}; v\right) \mathbf{1}_{\left[0,\lambda\left(X_{s^-}^{(k)}(h),h,r_s^{(k)}\right)\right]}(u).$$

Since N is a positive measure, (3.8) is bounded above by,

$$\int_0^T |F_s^k - F_s^{k-1}| N(ds,du,dv) = \int_0^T |F_s^k - F_s^{k-1}| ds du \alpha(dv) + \int_0^T |F_s^k - F_s^{k-1}| \tilde{N}(ds,du,dv), \quad (3.9)$$

where $\int_0^T |F_s^k - F_s^{k-1}| \tilde{N}(ds,du,dv)$ has mean zero, since $dsdu\alpha(dv)$ is the compensator of $N(ds,du,dv)$. Thus averaging, we are only left with the term $\int_0^T |F_s^k - F_s^{k-1}| dsdu\alpha(dv)$, which is dealt with using **(L2)**, and gives an upper bound similar of that of part **(A)**.
Summing up the contributions of **(A)**, **(B)** and **(C)**, we get, for a sufficiently large constant C,

$$E_T^{(k)} \le C(T + \sqrt{T}) E_T^{(k-1)}.$$

We now observe that the processes $X^{(k)}$, $k \ge 0$, $h \in \mathbb{R}^{d'}$ are progressively measurable for the filtration generated by the initial condition and the driving noise W, N, and satisfy

$$\int \mathbb{E}\left[\sup_{t\in[0,T]} \left|X_t^{(k)}(h)\right|\right] \mu(dh) < +\infty.$$

This can be seen by induction on k, replicating the steps above but using, rather than the Lipschitz conditions, the linear growth conditions (3.4). If we denote by $\mathcal{M}$ the space of progressively measurable, *cadlag*, $\mathbb{R}^d$ valued processes such that

$$\|X\| := \mathbb{E}\left[\sup_{t\in[0,T]} |X_t|\right] < +\infty,$$

and we take T sufficiently small, we have shown that

$$\sum_k \int \|X^{(k)}(h)\| \mu(dh) < +\infty,$$

and therefore for all h in a set F of μ-full measure

$$\sum_k \|X^{(k)}(h)\| < +\infty.$$

The norm $\|\cdot\|$ is not complete in $\mathcal{M}$, as the sup-norm is not complete in the space of cadlag functions. To get a complete metric, we replace the distance in sup-norm by the Skorohod distance d_S (see [5]), i.e.

$$D_S(X,Y) := \mathbb{E}\left[d_S(X,Y)\right].$$

Since the Skorohod distance is dominated by the distance in sup-norm, a Cauchy sequence for $\|\cdot\|$ is also Cauchy for the metric D_S. Thus, the limit $\overline{X}(h)$ of the sequence $X^{(k)}(h)$ can be defined for all $h \in F$, where F is a set of measure one for μ, and it is not hard to show (using also Proposition 2.1) that (3.2) holds for the limit. $\overline{X}(h)$ can be then easily defined for $h \notin F$ just by imposing that (3.2) holds.
This establishes existence of solution in $\mathcal{M}$ for T small. Since the condition on T does not involve the initial condition, the argument can be iterated on adjacent time intervals, obtaining a solution on any time interval.
Establishing uniqueness would actually be easy by using similar arguments. For us it is not actually needed, as uniqueness will follow from the convergence result in next section (Theorem 3.1).

Remark 3.4. It is more customary to use L^2 norms rather that L^1 norms for constructing solutions to SDE. The main difference is in **(C)**, where we estimate (3.8). When estimating the mean of the *square* of (3.9), the martingale contributes with

$$\int_0^T |F_s^k - F_s^{k-1}|^2 ds du \alpha(dv).$$

To complete the argument one needs a Lipschitz condition of the form

$$\int \left| f(x,h,r,v)\mathbf{1}_{[0,\lambda(x,h,r)]}(u) - f(y,h,r',v)\mathbf{1}_{[0,\lambda(y,h,r')]}(u)\right|^2 du\alpha(dv) \le L\left[|x-x'|^2 + d_2^2(r,r')\right], \tag{3.10}$$

where, in the whole argument, the distance

$$d_2(\nu,\nu') := \left(\inf\{\int |x-y|^2 \Pi(dx,dy) : \Pi \text{ has marginals } \nu \text{ and } \nu'\}\right)^{\frac{1}{2}}$$

would be used. The Lipschitz condition (3.10) is harder to check then (3.3), for the simple reason that "squaring an indicator function does not produce any square".

3.5 Propagation of chaos

Theorem 3.1. *Suppose conditions* **L1** *and* **L2** *hold. For $i \ge 1$ denote by $\overline{X}^i(h)$ the solution of* (3.2) *with the local parameter h and the same initial condition X_0^i of* (3.1). *Then for each i and $T > 0$*

$$\lim_{N\to+\infty} \int \mathbb{E}\left[\sup_{t\in[0,T]} \left| X_t^{i,N} - \overline{X}_t^i(h_i)\right|\right] \mu^{\otimes N}(d\underline{h}) = 0$$

where $\mu^{\otimes N}$ is the N-fold product of μ.

Proof. As in the proof of Proposition 3.2 we subtract the two equations for $X^{i,N}$ and $\overline{X}^i$. Using the triangular inequality, we estimate $\sup_{t\in[0,T]}\left|X_t^{i,N} - \overline{X}_t^i(h_i)\right|$ as sum of three terms, corresponding respectively to drift, diffusion and jumps. In this proof we only show how to deal with the drift term. The other two terms, involving stochastic integrals, are reduced to terms with Lebesgue time integrals as in the proof of Proposition 3.2, and then are estimated as the drift term.
We therefore give estimates for

$$\begin{aligned}\int \mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t b\left(X_s^{i,N}, h_i, \rho(\underline{X}_s^N, \underline{h})\right) ds - \int_0^t b\left(\overline{X}_s^i(h_i), h_i, r_s\right) ds\right|\right] \mu^{\otimes N}(d\underline{h}) \\ \le \int \mathbb{E}\left[\int_0^T \left| b\left(X_s^{i,N}, h_i, \rho(\underline{X}_s^N, \underline{h})\right) - b\left(\overline{X}_s^i(h_i), h_i, r_s\right)\right|\right] \mu^{\otimes N}(d\underline{h})\end{aligned} \tag{3.11}$$

By **(L1)**

$$\left| b\left(X_s^{i,N}, h_i, \rho(\underline{X}_s^N, \underline{h})\right) - b\left(\overline{X}_s^i(h_i), h_i, r_s\right)\right| \le L\left[\left|X_s^{i,N} - \overline{X}_s^i(h_i)\right| + d\left(\rho(\underline{X}_s^N, \underline{h}), r_s\right)\right]. \tag{3.12}$$

Now,

$$d\left(\rho(\underline{X}_s^N, \underline{h}), r_s\right) \le d\left(\rho(\underline{X}_s^N, \underline{h}), \rho(\underline{\overline{X}}_s, \underline{h})\right) + d\left(\rho(\underline{\overline{X}}_s, \underline{h}), r_s\right). \tag{3.13}$$

We consider the two summands in the r.h.s. of (3.13) separately. By definition of the metric $d(\cdot,\cdot)$

$$d\left(\rho(\underline{X}_s^N, \underline{h}), \rho(\underline{\overline{X}}_s, \underline{h})\right) \le \frac{1}{N}\sum_{j=1}^N \left|X_s^{j,N} - \overline{X}_s^j\right|,$$

so, by symmetry,

$$\int \mathbb{E}\left[d\left(\rho(\underline{X}_s^N,\underline{h}),\rho(\underline{\overline{X}}_s,\underline{h})\right)\right]\mu^{\otimes N}(d\underline{h}) \leq \int \mathbb{E}\left[\left|X_s^{i,N}-\overline{X}_s^i(h_i)\right|\right]\mu^{\otimes N}(d\underline{h}). \tag{3.14}$$

For the second summand in (3.13) we observe that, under $\mathbb{P}\otimes\mu^{\otimes\infty}$, the random variables $\left(\overline{X}_s^i(h_i),h_i\right)$ are i.i.d. with law $r_s \in \mathcal{P}(\mathbb{R}^{d+d'})$. By a recent version of the Law of Large Number ([26], Theorem 1), there exists a constant $C>0$, only depending on d and d', and $\gamma>0$ (any $\gamma<\frac{1}{d+d'}$ does the job) such that

$$\int \mathbb{E}\left[d\left(\rho(\underline{\overline{X}}_s,\underline{h}),r_s\right)\right]\mu^{\otimes N}(d\underline{h}) \leq \frac{C}{N^\gamma}. \tag{3.15}$$

Inserting what obtained in (3.12), (3.13) and (3.14) in (3.11) we get for some $C>0$, which may also depend on T,

$$\begin{aligned}\int \mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t b\left(X_s^{i,N},h_i,\rho(\underline{X}_s^N,\underline{h})\right)ds - \int_0^t b\left(\overline{X}_s^i(h_i),h_i,r_s\right)ds\right|\right]\mu^{\otimes N}(d\underline{h})\\ \leq C\int \mathbb{E}\left[\int_0^T\left|X_s^{i,N}-\overline{X}_s^i(h_i)\right|\right]\mu^{\otimes N}(d\underline{h})+\frac{C}{N^\gamma}.\end{aligned}$$

Dealing similarly with all terms arising in $\sup_{t\in[0,T]}\left|X_t^{i,N}-\overline{X}_t^i(h_i)\right|$, if we set

$$E_t := \int \mathbb{E}\left[\sup_{s\in[0,t]}\left|X_s^{i,N}-\overline{X}_s^i(h_i)\right|\right]\mu^{\otimes N}(d\underline{h})$$

we obtain

$$E_t \leq C\int_0^t E_s ds + \frac{C}{N^\gamma},$$

which, by Gromwall's Lemma and the fact that $E_0=0$ yields

$$E_T \leq \frac{C_T}{N^\gamma}$$

for some T-dependent constant C_T, and this complete the proof.

4 Applications

In this section we review some classes of models that are relevant for life sciences. Some key results will be stated, but no proofs are given.

4.1 The stochastic Kuramoto model

Synchronization phenomena leading to macroscopic rhythms are ubiquitous in science. Most (ab)used examples include

- applauses;
- flashing fireflies;
- protein concentration within cells in a multicellular system (reprissilators).

In these examples the systems are comprised by many units, each unit tending to behave periodically. Under circumstances depending on how units communicates, oscillation may *synchronize*, producing macroscopic pulsing. The (stochastic) Kuramoto model ([32]) is perhaps the most celebrated stylized model to capture this behavior.
In the Kuramoto model units are *rotators*, i.e. the state variable is an angle. Denoting by $X^{i,N}$ the angular variable (*phase*) of the i-th rotator, with $i = 1, 2, \dots, N$, the evolution is given by

$$dX_t^{i,N} = h_i dt + \frac{\theta}{N} \sum_{j=1}^{N} \sin\left(X_t^{j,N} - X_t^{i,N}\right) dt + dW_t^i. \tag{4.1}$$

Here h_i is the characteristic angular velocity of the i-th rotator. The effect of the interaction term is to favor phases to stay close. We assume the h_i's are i.i.d., drawn from a distribution μ on $\mathbb{R}$ with compact support. By possibly adding a constant speed rotation, there is no further loss of generality to assume that μ has mean zero. We further assume μ is symmetric, i.e. invariant by reflection around zero.
Clearly all results in Section 3 apply, and we get the following macroscopic limit:

$$d\overline{X}_t(h) = h dt + \theta \int \sin(y - \overline{X}_t) q_t(dy; h') \mu(dh') dt + dW_t, \tag{4.2}$$

where $q_t(dy; h')$ is the law of $\overline{X}_t(h')$. The flow of measures $q_t(\,\cdot\,, h)$ solves (indeed in the classical sense for the density w.r.t. the Lebesgue measure)

$$\frac{\partial}{\partial t} q_t(x; h) = \frac{1}{2} \frac{\partial^2}{\partial x^2} q_t(x; h) - \frac{\partial}{\partial x} \left[\left(h + \theta r_{q_t} \sin(\varphi_{q_t} - x)\right) q_t(x, h) \right] =: \mathcal{M}[q_t](h), \tag{4.3}$$

where

$$r_{q_t} e^{i\varphi_{q_t}} := \int e^{ix} q_t(dx; h) \mu(dh).$$

Equation (4.3) describes the collective behavior of the system of rotators. r_{q_t} captures the degree of synchronization of the system: $r_{q_t} = 0$ indicates total lack of synchronization, while a perfectly synchronized systems has $r_{q_t} = 1$.
One is interested in the long time behavior of solutions of (4.3), in particular stable equilibria. Note that, since the model is rotation invariant, if $q(x; h)$ solves $\mathcal{M}[q] = 0$, then also $q(x + x_0; h)$ does; thus there is no loss of generality in looking for equilibria satisfying $\varphi_q = 0$.
The proof of the following statement can be found in [7].

Theorem 4.1. *q^* is a solution of $\mathcal{M}[q] = 0$ with $\varphi_{q^*} = 0$ if and only if it is of the form*

$$q^*(x; h) = (Z_*)^{-1} \cdot e^{2(hx + \theta r_* \cos x)} \left[e^{4\pi h} \int_0^{2\pi} e^{-2(hx + \theta r_* \cos x)} dx \right.$$
$$\left. + (1 - e^{4\pi h}) \int_0^x e^{-2(hy + \theta r_* \cos y)} dy \right], \tag{4.4}$$

where Z_* is a normalization factor and r_* satisfies the consistency relation

$$r_* = \int e^{ix}\, q_*(x, h)\, \mu(dh)\, dx\,. \tag{4.5}$$

$r_* = 0$ is a solution of (4.5), and it corresponds to the incoherent *solution*

$$q^*(x; h) \equiv \frac{1}{2\pi},$$

i.e. the phases of the rotators are uniformly distributed on the torus.

Linear stability of the incoherent solution depends in a highly nontrivial way on θ and on the distribution μ of the local parameters. It is rather well understood in some special cases ([7, 8, 19]).

Theorem 4.2. *Denote by*

$$\theta_c = \left[\int \frac{\mu(dh)}{1+4h^2} \right]^{-1}. \tag{4.6}$$

(a) Suppose μ is unimodal, *i.e. it has a (even) density decreasing on $(0, +\infty)$. Then the incoherent solution is linearly stable if and only if $\theta < \theta_c$. At θ_c one (circle of)* synchronized *solution (i.e. with $r_q > 0$ bifurcates for the incoherent solution.*

(b) Suppose $\mu = \frac{1}{2}\left(\delta_{-h_0} + \delta_{h_0}\right)$ for some $h_0 > 0$. Then the incoherent solution is linearly stable if and only if $\theta < \theta_c \wedge 2$. For $\theta_c < 2$ at $\theta = \theta_c$ one (circle of) synchronized *solution (i.e. with $r_q > 0$ bifurcates. For $\theta_c > 2$ (which occurs for h_0 sufficiently large), at $\theta = 2$ the incoherent solution loses stability via a Hopf bifurcation: it is believed, but not rigorously proved, that stable time-periodic solutions emerge.*

It is not true in general that when the incoherent solution is stable then it is unique. It is believed it is in the unimodal case, but proved ether for θ small, or up to the critical point if μ is sufficiently concentrated around zero [36]. In the binary case, for certain values of the parameters it is known that there are values of θ smaller that the critical value for which *two distinct* circles of synchronized solutions exists [36].
In general, when the support of μ is contained in a sufficiently small interval, then synchronized solutions exist if and only if $\theta > \theta_c$, are unique up to rotation, and are linearly stable ([4, 27])

4.2 Interacting Fitzhugh-Nagumo neurons

Designed as reduction of more realistic models (e.g. the Hodgkin-Huxley model), the Fitzhugh-Nagumo model describes the evolution of the membrane potential x_t of a neuron through the following differential equation

$$\begin{aligned} \dot{x}_t &= x_t - \frac{1}{3}x_t^3 + y_t + I_t^{ext} \\ \dot{y}_t &= \epsilon(a + bx_t - \gamma y_t) \end{aligned} \tag{4.7}$$

where

- y_t is a *recovery variable* obtained by reduction of other variables;
- I_t^{ext} is the input current, assumed to be random and stationary. Without loss of generality, choosing a properly, we can assume I_t^{ext} has mean zero.
- b is the interaction strength between x and y, $\gamma \geq 0$ is a dissipation parameter. a is a kinetic parameter related with input current and synaptic conductance.

The parameter ϵ can be used to separate the time scales of the evolutions of the two variables. In what follows we assume $dI_t^{ext} = \sigma dW_t$ for a Brownian motion W.
To begin with, consider the equation in absence of randomness in the input current ($\sigma = 0$), and set $b = -1$, $\gamma = 0$ to make the analysis simpler. In this case (4.7) has a unique equilibrium in $(a, -a + a^3/3)$, which is globally stable for $|a| < 1$, is has a Hopf bifurcation at $|a| = 1$ and a stable periodic orbit emerges for $|a| > 1$. Thus, the system can be excited by the input, producing, at least for appropriate choice of the parameters, rapid variations of the potential (*spikes*) which occur periodically.
There are various ways to make several neurons interact in a network, even within the mean-field scheme, depending of how we model synapsis (see [2]). The simplest, corresponding to electrical

synapsis, leads to the following system. $X^{i,N}$ denotes the membrane potential of the i-th neuron. The local parameter h_i may be interpreted as the *macroscopic location* of the neuron, or its *type*.

$$\begin{aligned} dX_t^{i,N} &= \left(X_t^{i,N} - \frac{1}{3}(X_t^{i,N})^3 + Y_t^{i,N} \right) dt + \frac{1}{N} \sum_{j=1}^{N} J(h_i, h_j) \left(X_t^{i,N} - X_t^{j,N} \right) dt + \sigma dW_t^i \\ dY_t^{i,N} &= \epsilon(h_i) \left[a(h_i) + b(h_i) X_t^{i,N} - \gamma(h_i) Y_t^{i,N} \right] dt, \end{aligned} \tag{4.8}$$

where the coupling parameters $J(h_i, h_j)$ tune the interaction petween pairs of neurons.
The model become more interesting if one introduces a delay τ in the transmission of informations between different neurons:

$$\begin{aligned} dX_t^{i,N} &= \left(X_t^{i,N} - \frac{1}{3}(X_t^{i,N})^3 + Y_t^{i,N} \right) dt + \frac{1}{N} \sum_{j=1}^{N} J(h_i, h_j) \left(X_t^{i,N} - X_{t-\tau(h_i,h_j)}^{j,N} \right) dt + \sigma dW_t^i \\ dY_t^{i,N} &= \epsilon(h_i) \left[a(h_i) + b(h_i) X_t^{i,N} - \gamma(h_i) Y_t^{i,N} \right] dt. \end{aligned} \tag{4.9}$$

Delay makes a bit more painful the well posedness analysis for both the model and its macroscopic limit, but for propagation of chaos the same proof carries through (see [45] for details), giving the following macroscopic limit

$$\begin{aligned} d\overline{X}_t(h) &= \left(\overline{X}_t(h) - \frac{1}{3}\overline{X}_t^3(h) + \overline{Y}_t(h) + \int J(h, h') \left(\overline{X}_t(h) - y \right) q_{t-\tau(h,h')}(dy; h') \mu(dh') \right) dt + \sigma dW_t \\ d\overline{Y}_t(h) &= \epsilon(h)(a(h) + b(h)\overline{X}_t(h) - \gamma(h)\overline{Y}_t(h))dt, \end{aligned} \tag{4.10}$$

where $q_t(dx; h)$ denotes the law of $\overline{X}_t(h)$. Not much is known at this level of generality, so we consider the simplest, homogeneous case in which h is constant, $\gamma = 0$, $b = -1$ which gives

$$\begin{aligned} d\overline{X}_t &= \left[\overline{X}_t) - \frac{1}{3}\overline{X}_t^3 + \overline{Y}_t + J(\overline{X}_t - \mathbb{E}(\overline{X}_{t-\tau})) \right] dt + \sigma dW_t \\ d\overline{Y}_t &= \epsilon(a - \overline{X}_t)dt \end{aligned} \tag{4.11}$$

A further simplification consists in letting the noise going to zero, in both the diffusion and the initial condition. Note the the noise is *essential* to prove propagation of chaos, so this must be meant as a limiting procedure at macroscopic level. We obtain the deterministic system with delay

$$\begin{aligned} \dot{x}_t &= x_t - \frac{1}{3}x_t^3 + y_t + J(x_t - x_{t-\tau}) \\ \dot{y} &= \epsilon(a - x_t). \end{aligned} \tag{4.12}$$

This system has been extensively studies in [31]. Here we assume $J \geq 0$

- The point $(a, -a + a^3/3)$ is still the unique fixed point, and it is stable for $|a| > \sqrt{1 + 2J}$ and unstable for $|a| < 1$, no matter what τ is.
- For $1 < \tau < \sqrt{1 + 2J}$ loss of stability via a Hopf bifurcation can be obtained by increasing τ: *interaction and transmission delay may produce oscillations even if single neurons are in the stability region.*

Does noise play any role in exciting the neuronal network?

This question has only partial and non-rigorous answers (see e.g [34]). Consider the simplified system (4.11) and remove the delay.

$$
\begin{aligned}
d\overline{X}_t &= \left[\overline{X}_t) - \frac{1}{3}\overline{X}_t^3 + \overline{Y}_t + J(\overline{X}_t - \mathbb{E}(\overline{X}_t))\right] dt + \sigma dW_t \\
d\overline{Y}_t &= \epsilon(a - \overline{X}_t)dt
\end{aligned} \tag{4.13}
$$

Some indications on the behavior of this system, confirmed by numerical simulations, are obtained via the following heuristic argument. For a similar model details can be found in [17]

- Writing down the equation for the moments of $(\overline{X}_t, \overline{Y}_t)$ and *pretending* the system is Gaussian, we get at formal level a closed equation for the means and the covariance matrix.
- This equation corresponds to a truly Gaussian process $(\tilde{X}, \tilde{Y})$, which can be shown to be a good approximation of $(\overline{X}, \overline{Y})$ for σ small.

The evolution of the law of $(\tilde{X}, \tilde{Y})$ can be studied at least locally around the fixed point. It can be shown that for $|a| > 1$ but sufficiently close to 1, periodic solutions for the law of $(\tilde{X}_t, \tilde{Y}_t)$ emerge for *moderate* values of σ, i.e. within some interval $0 < \sigma_0 < \sigma < \sigma_1$: we therefore obtain *noise-induced* oscillations. It should be remarked noise-induced oscillations were pointed out in similar Gaussian models long time ago ([42])

4.3 Interacting Hawkes processes

The Fitzhugh-Nagumo model exhibits some qualitative features of neuronal dynamics, in particular excitability. Periodicity of spikes for a single neuron is however unrealistic: spike trains are more effectively modeled by point processes. An appropriate model in this context is obtained by using Hawkes processes ([14, 13, 15]).
Let $Z_t^{i,N}$ be the counting process that counts the spikes of neuron i, having local parameter (position, type...) h_i. It is assumed that $Z_t^{i,N}$ jumps with a *rate* $\lambda_i^N(t)$ of the form

$$
\lambda_i^N(t) = f\left(h_i; \frac{1}{N}\sum_{j=1}^N J(h_i, h_j)\int_{[0,t]} k(t-s)dZ_s^{j,N}\right)
$$

where $f(h; \cdot)$ is a positive, increasing function, and $k(\cdot)$ is a given positive function modeling the *memory* of the system, including possible transmission delay. If $J(h_i, h_j) > 0$ then spikes of neuron j tend to favor future spikes of neuron i (excitatory link), while the opposites holds true (inhibitory link) when $J(h_i, h_j) < 0$.
There are convenient choices for the kernel $k(\cdot)$ which allow a simple "Markovianization" of the system, namely the *Erlang kernels*: $k(r) = c\frac{r^m}{m!}e^{-\lambda r}$, $c, \lambda > 0$. Note that for $m \geq 1$ the function k attains its maximum at some positive $r^* = \tau$, producing a "smoothed" form of delay. For simplicity, we deal here with the case $k(r) = e^{-\lambda r}$, corresponding to no delay.
Define

$$
X_t^{i,N} := \int_{[0,t]} k(t-s)dZ_s^{i,N},
$$

the "discounted" number of spikes of neuron i before time t. The exponential form of $k(\cdot)$ yields

$$
\begin{aligned}
X_t^{i,N} &= -\lambda\int_0^t X_s^{i,N}ds + Z_t^{i,N} \\
&= -\lambda\int_0^t X_s^{i,N}ds + \int_{[0,t]} \mathbf{1}_{[0, f(h_i, \frac{1}{N}\sum_{j=1}^N J(h_i,h_j)X_{s-}^{j,N}]}(u)N^i(du, ds),
\end{aligned} \tag{4.14}
$$

where the N^i are i.i.d. Poisson random measures on $[0, +\infty) \times [0, +\infty)$ with characteristic measure $duds$. The system is therefore in the form seen in Section 3. Assuming $f(\cdot)$ is Lipschitz, propagation of chaos holds, and we obtain the macroscopic limit

$$\overline{X}_t(h) = -\lambda \int_0^t \overline{X}_s(h)ds + \int_{[0,t]} \mathbf{1}_{[0,f(h,\int J(h,h')\mathbb{E}[\overline{X}_{s-}(h')]\mu(dh'))]}(u)N(du,ds). \tag{4.15}$$

Letting $m_t(h) := \mathbb{E}[\overline{X}_t(h)]$, we obtain from (4.15) a closed equation for m_t:

$$\dot{m}_t(h) = -\lambda m_t(h) + f\left(h, \int J(h,h')m_s(h')\mu(dh')\right). \tag{4.16}$$

If the support of μ is finite, this is a finite dimensional dynamical system. A case considered recently [24] is that of the so-called *cyclic negative feedback systems.*

Theorem 4.3. *Suppose μ is supported on the discrete torus $\mathbb{Z}/n\mathbb{Z}$, $J(h,h') = 0$ unless $h' = h+1$ mod n. Set*

$$\delta := \prod_{h \in \mathbb{Z}/n\mathbb{Z}} J(h,h+1).$$

If $n \geq 3$, $\delta < 0$ and $|\delta|$ is large enough, then (4.16) has at least one stable periodic orbit. This orbit is unique for $n = 3$.

Although single neurons have no intrinsic tendency of spiking periodically, the collective spike train may be periodic if

- the macroscopic geometry of the network is circular;
- at macroscopic level there is an odd number of inhibitory links;
- the interaction is sufficiently strong.

5 Further reading

These notes on mean field models have been essentially dealing with propagation of chaos and, for what applications are concerned, with the analysis of the attractors of the macroscopic dynamics. We briefly mention here some further developments, well aware of being far from exhaustive.

5.1 Long-time behavior of the microscopic system

Theorem 3.1 states that if we *fix* the time interval $[0,T]$ then the microscopic and the macroscopic systems are close if N is large enough. How large, for a giver error threshold, might indeed depend on T. In other words for a given large N, this "closeness" might deteriorate as time increases: the long time behavior of the microscopic system is not necessarily reflected in the macroscopic one.
Whenever such "deterioration" *does not* occur, we say there is *uniform propagation of chaos.* One consequence of uniform propagation of chaos is that stationary measures for the microscopic system are close to products of stationary measures of the macroscopic one.
Uniform propagation of chaos has been proves in cases in which the microscopic process satisfies very strong ergodicity properties, see e.g. [38, 44, 23, 6, 46].
When uniform propagation of chaos fails, it is of interest to identify the time scale (possibly diverging with N) in which the limit macroscopic system still approximate the microscopic one, and determine the behavior beyond this time scale. In general this is a very delicate problem. Quite remarkable results for a class of system inspired by the Kuramoto model are obtained in [28].

5.2 Fluctuations

We have seen (Proposition 2.1) that propagation of chaos is equivalent to a Law of Large Numbers:

$$\rho_N(\underline{X}^N) = \frac{1}{N}\sum_{i=1}^{N}\delta_{X^{i,N}} \;\longrightarrow\; Q \tag{5.1}$$

as $N \to +\infty$, where Q is the law of the macroscopic dynamics. It is therefore natural to consider a corresponding Central Limit Theorem, describing *fluctuations* around the limit. In particular, one considers the distribution-valued process

$$\Phi_t^N := \sqrt{N}\left[\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^{i,N}} - q_t\right],$$

where q_t is the marginal of Q at time t. One can prove, with remarkable generality, that for any bounded time-interval $[0,T]$, the process Φ^N converges weakly to a distribution valued Gaussian process. Classical results in this direction can be found in [18, 19, 25, 30].
When quenched disorder is present, fluctuations of the disorder compete with state fluctuations, producing phenomena which are not seen if the disorder is averaged out; the dynamics of state fluctuations are different for different realizations of the disorder. Sharp results have been obtained for the Kuramoto model in [35].

5.3 Critical fluctuations

All examples we have treated undergo a *phase transition*: in the macroscopic dynamics, the stationary solution that is unique for small interaction, loses its stability as the interaction strength crosses a threshold, and is subject to bifurcation. At the critical point, the fluctuations process Φ^N defined above exhibit, if evaluated on certain observables, a peculiar space-time scaling, that typically leads to non-Gaussian fluctuations. Literature on this subject has a long history, going back to [21, 18]. Possible effects of quenched disorder are dealt with in [16]. Recently, examples of mean field dynamics in which critical fluctuations *self-organize*, i.e. do not require tuning parameters to critical values, are provided in [12]
The results just cited apply to cases in which the bifurcation at the critical point is of *pitchfork* type. Various interesting models, including the Kuramoto model with large quenched disorder, undergoes a Hopf bifurcation. Some indications on how critical fluctuations look like in this case can be found in [20].

5.4 Large Deviations

A refinement of the Law of Large Numbers in (5.1) different from the Central Limit Theorems consists in obtaining a Large Deviation Principle, i.e. the exponential decay in N of probabilities of the form

$$\mathbb{P}(\rho_N(\underline{X}^N) \in U) \quad \text{for } U \not\ni Q.$$

The case of mean-field interacting diffusions with a constant diffusion coefficient given by a multiple of the identity matrix dates back to [22] and [19], where spin-flip dynamics have also been dealt with. Large deviation principles for system as general as those in Section 3 of these note require more sophisticated tools, see [9].
In presence of quenched disorder, it would be desirable to obtain a Large deviation Principle that holds *for almost every* realization of the disorder. For interacting diffusions this is done in [37].

5.5 Generalizing network's microscopic geometry

In the models presented in these notes the quenched disorder is introduced via the local parameters h_i, one per each component. An interesting alternative way of introducing disorder is to associate it with *links*, i.e. to pairs of components. For instance, this would modify the prototypical model in Section 2 as

$$dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^{N} b(h_{ij}, X_t^{i,N}, X_t^{j,N})dt + dW_t^i$$

where the h_{ij} are random parameters, describing the microscopic architecture of the network. Model of this type, motivated by neurosciences, are dealt with in [40]. We remark that these models are reminiscent of mean field spin-glass dynamics (see e.g. [3]), but actually have very different nature: in spin glasses the contribution of each pair scales as $\frac{1}{\sqrt{N}}$ rather that as $\frac{1}{N}$; the thermodynamic limit for spin glasses is in general much harder to analyze, and the resulting dynamic behavior is quite different.

5.6 Mean-field games

In many applications, mainly in social science, collective dynamics are the result of a competitive optimization procedure involving several entities (players). Each player controls, to some extent, his own dynamics, and aims at maximizing his utility; he is therefore taking part to a *dynamic game*. Under symmetry conditions of the players, letting the number of players going to infinity, one expect to obtain a macroscopic game, called *mean field game*.
Introduced in the seminal paper [33], the theory of mean field games has had a tremendous development. The actual convergence of the microscopic dynamics to the mean-field game has been however left open for several years, and recently proved, under rather severe conditions, in [10]. Further developments, such as fluctuations around the limit and Large Deviations, have not appeared yet.

References

[1] L. Andreis, P. Dai Pra, M. Fischer. McKean-Vlasov limit for interacting systems with simultaneous jumps. Preprint, 2016.

[2] J. Baladron, D. Fasoli, O. Faugeras, & J. D. Touboul. Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. The Journal of Mathematical Neuroscience, 2(1), 10, 2012

[3] G. Ben Arous, A. Guionnet. Large deviations for Langevin spin glass dynamics. Probability Theory and Related Fields, 102(4), 455-509, 1995.

[4] L. Bertini, G. Giacomin, and K. Pakdaman. Dynamical aspects of mean field plane rotators and the Kuramoto model. J. Statist. Phys., 138:270-290, 2010.

[5] P. Billingsley, Convergence of probability measures. John Wiley & Sons, 2013.

[6] F. Bolley, I. Gentil, and A. Guillin, Uniform convergence to equilibrium for granular media, Archive for Rational Mechanics and Analysis, 208 (2013), pp. 429-445.

[7] L. L. Bonilla, J. C. Neu, and R. Spigler. Nonlinear stability of incoherence and collective synchronization in a population of coupled oscillators. J. Statist. Phys., 67(1-2):313?330, 1992.

[8] L. L. Bonilla, C. J. Pérez Vicente, and R. Spigler. Time-periodic phases in populations of nonlinearly coupled oscillators with bimodal frequency distributions. Phys. D, 113(1):79-97, 1998.

[9] A. Budhiraja, P. Dupuis, and M. Fischer, Large deviation properties of weakly interacting processes via weak convergence methods, Ann. Probab. Volume 40, Number 1 (2012), 1-435

[10] P. Cardaliaguet, F. Delarue, J.M. Lasry, P.-L. Lions. The master equation and the convergence problem in mean field games. arXiv preprint arXiv:1509.02505, 2015.

[11] P. Cattiaux, A. Guillin, and F. Malrieu, Probabilistic approach for granular media equations in the non-uniformly convex case, Probability theory and related fields, 140(1-2):19-40, 2008.

[12] R. Cerf, M. Gorny. A Curie-Weiss model of self-organized criticality. The Annals of Probability, 44(1), 444-478, 2016.

[13] J. Chevallier. Mean-field limit of generalized Hawkes processes. Available on http://arxiv.org/abs/1510.05620, 2015.

[14] J. Chevallier, MJ. Caceres, M. Doumic, P. Reynaud-Bouret. Microscopic approach of a time elapsed neural model. Math. Mod. & Meth. Appl. Sci., 25(14), 2669-2719, 2015.

[15] J. Chevallier, A. Duarte, E. Löcherbach, & G. Ost. Mean field limits for nonlinear spatially extended Hawkes processes with exponential memory kernels. arXiv preprint arXiv:1703.05031, 2017.

[16] F. Collet and P. Dai Pra. The role of disorder in the dynamics of critical fluctuations of mean field models. Electronic Journal of Probability, 26:1-40, 2012.

[17] F. Collet, P. Dai Pra, M. Formentin, Collective Periodicity in Mean-field Models of Cooperative Behavior, NoDEA Nonlinear Differential Equations Appl. 22, No.5, pp. 1461-1482, 2015.

[18] F. Comets and T. Eisele. Asymptotic dynamics, noncritical and critical fluctuations for a geometric long-range interacting model. Comm. Math. Phys., 118:531?567, 1988.

[19] P. Dai Pra and F. den Hollander. McKean-Vlasov limit for interacting random pro- cesses in random media. J. Statist. Phys., 84(3-4):735-772, 1996.

[20] P. Dai Pra, D. Tovazzi. The dynamics of critical fluctuations in asymmetric Curie-Weiss models. arXiv:1703.07572, 2017.

[21] D. A. Dawson, Critical dynamics and fluctuations for a mean-field model of cooperative behavior. J. Statist. Phys., 31(1):29-85, 1983.

[22] D.A. Dawson and J. Gartner, Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. Stochastics 20 (1987) pp. 247-308.

[23] P. Del Moral and E. Rio, Concentration inequalities for mean field particle models, Annals of Applied Probability, (2011).

[24] S. Ditlevsen, & E. Löcherbach. Multi-class oscillating systems of interacting neurons. Stochastic Processes and their Applications, in press.

[25] B. Fernandez and S. Méléard. A Hilbertian approach for fluctuations on the McKean- Vlasov model. Stochastic Process. Appl., 71(1):33-53, 1997.

[26] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. Probability Theory and Related Fields, 162(3-4):707-738, 2015

[27] G. Giacomin, E. Lucon, and C. Poquet. Coherence stability and effect of random natural frequencies in populations of coupled oscillators. Journal of Dynamics and Differential Equations 26.2: 333-367, 2014.

[28] G. Giacomin, C. Poquet, A. Shapira. Small noise and long time phase diffusion in stochastic limit cycle oscillators. arXiv preprint arXiv:1512.04436, 2015

[29] C. Graham. Nonlinear diffusion with jumps. *Annales de l'I.H.P. Probabilités et statistiques*, 28(3):393–402, 1992.

[30] B. Jourdain and S. Méléard. Propagation of chaos and fluctuations for a moderate model with smooth initial data. Ann. Inst. H. Poincar Probab. Statist., 34(6):727- 766, 1998.

[31] M. Krupa, & J. D. Touboul, J. D. Complex oscillations in the delayed FitzHugh-Nagumo equation. Journal of Nonlinear Science, 26(1), 43-81, 2016.

[32] Y. Kuramoto. Chemical oscillations, waves, and turbulence. Courier Dover Publications, 2003.

[33] J.M. Lasry, P.-L. Lions J.-M., Jeux achamp moyen i. le cas stationnaire. C.R. Acad. Sci. Paris, 343(9), 2006.

[34] B. Lindner, J. Garca Ojalvo, A. Neiman, and L. Schimansky-Geier. Effects of noise in excitable systems. Physics Reports, 392(6):321-424, 2004.

[35] E. Lucon. Quenched limits and fluctuations of the empirical measure for plane rotators in random media. Electronic Journal of Probability, 16:792-829, 2011.

[36] E. Lucon. Oscillateurs coupls, dsordre et synchronisation. Diss. Universit Pierre et Marie Curie-Paris VI, 2012.

[37] E. Lucon. Quenched large deviations for interacting diffusions in random media. arXiv preprint arXiv:1608.06442, 2016.

[38] F Malrieu, Logarithmic sobolev inequalities for some nonlinear pde?s, Stochastic Processes and their Applications, 95 (2001), pp. 109-132.

[39] P. Protter, Stochastic Differential Equations. In Stochastic Integration and Differential Equations (pp. 187-284). Springer Berlin Heidelberg, 2004.

[40] C. Quiñinao, J. Touboul. Limits and dynamics of randomly connected neuronal networks. Acta Applicandae Mathematicae, 136(1), 167-192, 2015

[41] J. Salhi, J. MacLaurin. On uniform propagation of chaos, https://arxiv.org/abs/1503.07807v1, 2015.

[42] M. Scheutzow, Noise can create periodic behavior and stabilize nonlinear diffusions, Stochastic Process. Appl. 20 (1985), 323-331.

[43] A.-S. Sznitman. Topics in propagation of chaos. In *Ecole d'Eté de Probabilités de Saint-Flour XIX—1989*, pages 165–251. Springer, 1991.

[44] A. Yu Veretennikov, On ergodic measures for mckean?vlasov stochastic equations, Monte-Carlo and Quasi-Monte-Carlo Methods, (2006), pp. 471-486.

[45] J. D. Touboul. Limits and dynamics of stochastic neuronal networks with random heterogeneous delays. Journal of Statistical Physics, 149(4), 569-597, 2012.

[46] J. Salhi, J. MacLaurin. On uniform propagation of chaos, https://arxiv.org/abs/1503.07807v1, 2015.