



**HAL**  
open science

# Mean field analysis of multi-population neural networks with random synaptic weights and stochastic inputs

Jonathan Touboul, Olivier Faugeras, Bruno Cessac

► **To cite this version:**

Jonathan Touboul, Olivier Faugeras, Bruno Cessac. Mean field analysis of multi-population neural networks with random synaptic weights and stochastic inputs. [Research Report] RR-6454, 2008. inria-00258345v1

**HAL Id: inria-00258345**

**<https://inria.hal.science/inria-00258345v1>**

Submitted on 21 Feb 2008 (v1), last revised 4 Aug 2008 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

*Mean field analysis of multi-population neural  
networks with random synaptic weights and  
stochastic inputs*

Jonathan Touboul — Olivier Faugeras — Bruno Cessac

**N° 1234**

February 21, 2008

Thème BIO

*R*apport  
de recherche





## Mean field analysis of multi-population neural networks with random synaptic weights and stochastic inputs

Jonathan Touboul , Olivier Faugeras , Bruno Cessac

Thème BIO — Systèmes biologiques  
Projet Odysée \*

Rapport de recherche n° 1234 — February 21, 2008 — 40 pages

**Abstract:** In this article we are interested in the dynamical behavior of the mean field description of neural networks with random synaptic weights and stochastic inputs. Unlike other authors, we consider this problem as a *measure problem*. From this point of view we address the question of existence and uniqueness of solutions of dynamic mean field equations describing the behavior of the neural network in the large size limit. To this purpose, we introduce a new metric on the set of stochastic processes for which this set is complete. In this framework, we use the standard theory of fixed points in complete spaces to prove existence and uniqueness of solutions. This framework gives us a direct method to compute explicitly these solutions, and allows us to generalize previous results to more general models. We also consider the problem of existence and uniqueness of stationary solutions which have been previously studied essentially using local or numerical methods. We finally give some examples of results obtained by the numerical technique proposed.

**Key-words:** mean-field, existence, uniqueness, complete space, measure space, neural masses, multi-population neural networks, Jansen-Rit model

\* Odysée is a joint project between ENS Ulm - INRIA

## **Analyse champ-moyen pour des réseaux de neurones multi-populations à coefficients synaptiques aléatoires et entrées stochastiques**

**Résumé :** Dans cet article on s'intéresse au comportement dynamique d'une description en champ moyen de réseaux de neurones avec poids synaptiques aléatoires et entrées stochastiques. Contrairement aux autres auteurs, nous considérons ce problème comme un problème de mesure. De ce point de vue, nous posons la question de l'existence et unicité de solutions aux équations de champ moyen dynamique décrivant le comportement du réseau dans la limite du nombre de neurones tendant vers l'infini. Pour ce faire, nous introduisons une nouvelle métrique sur l'espace des processus stochastiques qui le rend complet. Dans ce cadre, nous utilisons la théorie standard de points fixes dans les espaces complets pour prouver l'existence et unicité des points fixes. Cette étude nous donne une méthode directe pour calculer numériquement ces solutions, et nous permet de généraliser les résultats existants. Nous considérons également le problème d'existence et unicité de solutions stationnaires. Nous concluons en donnant des exemples de résultats obtenus par la méthode numérique que nous explicitons dans le papier.

**Mots-clés :** champ moyen, existence, unicité, espaces complets, espace de mesures, masses neuronales, réseaux de neurones multi-populations, Jansen et Rit

## Introduction

Modeling neural activity at scales integrating the effect of thousands of neurons is of central importance for several reasons. On one hand, most imaging techniques are not able to measure individual neuron activity (“microscopic” scale), but are instead measuring mesoscopic effects resulting from the activity of several hundreds to several thousands of neurons. On the other hand, anatomical data reveals, in the cortex, the existence of structures, such as the cortical columns, with a diameter of about  $100\mu m$ , containing about one thousand of neurons belonging to a few different species. These columns have specific functions. For example, in the visual cortex V1, they respond to preferential orientations of bar-shaped visual stimuli. As a matter of fact, in this case, information treatment does not occur at the scale of individual neurons but rather corresponds to a mesoscopic scale integrating the collective dynamics of many interacting neurons. The description of this collective dynamics requires models which are different from individual neurons models. In particular, if the accurate description of one neuron requires “ $m$ ” parameters such as sodium, potassium, calcium conductances, membrane capacitance, etc... it is not necessarily true that an accurate model of a  $N$  neurons assembly requires at least  $Nm$  parameters. Indeed, if  $N$  is large enough one expects to have “averaging” effects such that the collective dynamics is well described by an effective mean-field, summarizing the effect of the interactions of a neuron with the other neurons, and depending on a few effective control parameters. This vision, inherited from statistical physics requires that the space scale is large enough to include a large number of microscopic components (here neurons) and small enough so that the region considered is homogeneous. This is the case of cortical columns for instance.

However, obtaining the equations of evolution of the effective mean-field from microscopic dynamics is far from being evident. In simple physical models this can be achieved via the law of large numbers and the central limit theorem, provided that time correlations decrease sufficiently fast. This type of approach has been generalized to fields such as quantum field theory or non equilibrium statistical mechanics. To the best of our knowledge, the idea of applying mean-field methods to neural networks dates back to Amari [1, 2]. In his approach, he uses an assumption that he called “local chaos hypothesis”, reminiscent of Boltzmann’s “molecular chaos hypothesis”, and postulating the vanishing of individual correlations between neurons, when the number of neurons  $N$  tends to infinity. Later on, Crisanti, Sompolinsky and coworkers [20] used a dynamic mean-field approach to conjecture the existence of chaos in an homogeneous neural network with random independent synaptic weights. This approach was formerly developed by Sompolinsky and coworkers for spin-glasses [21, 10, 11], where complex effects such as aging or coexistence of a diverging number of metastable states, renders the mean-field analysis delicate in the long time limit [15].

On the opposite, these effects do not appear in the neural network considered in [20] because the synaptic weights are independent [7] (and especially non symmetric, in opposition to spin glasses). In this case, the Amari approach and the dynamic mean field approach leads to the same mean-field equations. Later on, the non-rigorous mean-field equations derived by Sompolinsky and Zippelius [21] for spin-glasses, were rigorously obtained by Ben Arous and Guionnet [3, 4, 14]. The application of their method to a discrete time version of the neural network considered in [20] and in [17] was done by Moynot and Samuelides [18].

Mean-field methods are often used in neural network community but there are only a few results using dynamic mean-field method and even less rigorous results. The main advantage of dynamic mean-field techniques is that they allow one to consider neural networks where synaptic weights are random (and independent), while other methods such as the Fokker-Planck approach, used e.g. by Brunel and Hakim, assume constant synaptic weights [6]. This allows to stand genericity results about the dynamics according to the statistical parameters controlling the probability distribution of the synaptic weights [19]. This approach does not only provide the evolution of the “mean” activity of the network. It is an equation on the law of the mean-field, and hence also provides informations on the fluctuations and correlations. These correlations are of crucial importance as revealed in Sompolinsky and coworkers paper [20]. **[Indeed, the analysis of correlations allows one to discriminate between stationary and chaotic dynamics: reécriture de la phrase]**. However, this approach has also several drawbacks explaining why it is so seldom used. On one hand, the usual physicists’ method for obtaining these equations use a generating functional approach requiring heavy computations and some “art” and which sorely generalizes, for instance to models with several populations. What is less known, is that Amari’s approach allows their derivation in a more straightforward way, which can be rigorously stated in the large deviations approach used by Ben Arous and Guionnet [7, 3, 19]. Also, dynamics mean field equations characterize *in fine* a nonstationary process. It is then natural to search for stationary solutions. This simplifies consequently dynamic mean field equations reducing the problem to a set of differential equations (see section 4) but the price to pay, is that there is a non free parameter in the problem, the initial condition, that must be obtained via the investigation of the non stationary problem. Hence there it is not clear wether if there exists such a solution, and if it is the case, how to characterize this solution. To the best of our knowledge, this tricky question has been fully investigated in only one (unpublished) paper by Crisanti and coworkers [9]. In this paper, we investigate this question using a different and rigorous approach.

Thus, the scope of this paper is twofold. On one hand, we analyze rigorously the mean-field equations for a multi-populations of neurons<sup>1</sup>. These equations are obtained via the Amari approach and we do not discuss their rigorous derivation with large deviations or hydrodynamic limit techniques, since this is a straightforward extension of [3, 4, 14]. Instead, we focus on the question of existence and uniqueness of the solutions. A proof of existence and uniqueness based on the Wasserstein distance between processes had been provided by Ben-Arous and Guionnet in [3] in the case of a spin glass with Langevin dynamics. In the present paper we introduce a new distance on processes, and show how it naturally helps for solving the problem of existence and uniqueness of a solution for the mean field equation. We also use this norm to show the existence and uniqueness of solutions of the mean field equation starting from  $t_0 = -\infty$  corresponding to the physical process of removing transients and focusing on stationary solutions. Actually, we prove that such stationary solutions exist and we compare our approach to the one presented by Sompolinsky and coworkers [20, 9].

The motivation of this work is also to give an effective description of assemblies of neurons to get a better understanding of the neuronal assembly models or neural masses models. A typical

<sup>1</sup>Note that we treat in this paper the case of a neural network equation, but the present framework can be adapted to a wide range of implicit equations on the law of stochastic processes.

example is Jansen and Rit’s neural mass model [16]. Our contribution is to generalize the existing mathematical results and to give a tractable way of computing the dynamical mean-field model. Indeed, we show that an iterative method used in the proof of existence and uniqueness of solutions can be numerically implemented, giving access to the non stationary mean-field dynamics, for a multi-population neuronal network. As an example, we analyze the case treated by Sompolinsky and coworkers. The application of this method to cortical columns models will be scope of another paper, less mathematical and more oriented toward neuroscience community.

# 1 Mean-field equations for different multi-population neural network models

## 1.1 The general model

### 1.1.1 General framework

We consider a network composed of  $N$  neurons indexed by  $i \in \{1, \dots, N\}$  belonging to  $P$  populations  $\alpha \in \{1, \dots, P\}$ . Each neuron is defined by its membrane potential. This potential has a linear intrinsic dynamics only depending on the neuron’s population. Each neuron of the network receives external inputs, modeled in this paper by the sum of a deterministic external current depending on the population index of the neuron and of independant stochastic inputs whose probability distributions only depend on the population index. It also receives inputs from other neurons, via its activity, which is a sigmoidal transform of the membrane potential<sup>2</sup>.

We denote for  $\alpha \in \{1, \dots, P\}$  a population index,  $N_\alpha$  the number of neurons in this population. Hence we have  $N = \sum_{\alpha=1}^P N_\alpha$ . We are interested in the limit  $N \rightarrow \infty$ .

We assume that the proportions of neurons in each population are non-trivial, i.e. :

$$\lim_{N \rightarrow \infty} \frac{N_\alpha}{N} = n_\alpha \in (0, 1) \forall \alpha \in \{1, \dots, P\}$$

If it was not the case, then the population having a trivial proportion will not affect the global behavior of the system, and will not contribute to the mean field equation.

Each neuron  $i$  of population  $\alpha$  is connected to each neuron  $j$  of population  $\beta$  with the connectivity weight  $J_{i,j}$ , which is a random variable independent of all the other weights, having a Gaussian law that depend only on population pair  $\alpha, \beta$ , and which depend on  $N$ :

$$J_{i,j} \sim \mathcal{N}\left(\frac{\bar{J}_{\alpha\beta}}{N_\beta}, \frac{\sigma_{\alpha\beta}}{\sqrt{N_\beta}}\right)$$

Synaptic weights  $J_{i,j}$  are drawn at the beginning of the network evolution, and are then “frozen” during the evolution. The activity of a neuron is transmitted to its neighbors via its activity variable

---

<sup>2</sup> Note that the present work affords an extension where each neuron receives a randomly distributed input whose law depends on the neuron population only. The proof will be essentially the same, but the expressions more intricate.



which is a sigmoidal transform of the membrane potential, i.e a bounded increasing functions such that

$$\begin{cases} \lim_{x \rightarrow -\infty} S(x) = S_{\min} \\ \lim_{x \rightarrow \infty} S(x) = S_{\max} \end{cases}$$

We do not assume this function continuous. This allows us to treat the case of analogical neurons as well as spiking neurons (where the spikes are modelled by step sigmoidal functions). We assume that the sigmoids depend on the presynaptic and postsynaptic populations, to keep the model general.

### 1.1.2 Model's equations

The intrinsic dynamics of neurons is considered as linear, governed by a time-dependant non-diagonal locally bounded matrix<sup>3</sup>.

For the General Model of order 1 (GM1), we consider that the intrinsic dynamics governing the membrane potential is a differential equation of order 1 (e.g. Jansen's model with exponential PSP (8)). For the General Model of order  $k$  (GMk), the dynamics is governed by a stochastic differential equation of order  $k$  (e.g. Jansen's model (12)).

The neurons receive correlated Brownian inputs. This accounts for instance for the possible finite number of noise sources or for the case when the same noise source can affect many neurons in different populations.

Denote by  $K(N)$  the number of independent noise sources, modeled as standard Brownian motions, when the total number of neurons is  $N$ .

Let us now write the microscopic stochastic differential equations when the number of neurons is finite. We first consider the GM1. The time-dependant matrix driving the intrinsic dynamics of the network is denoted by  $\mathbf{L}^{(N)}(t) = (\mathbf{L}_{ij}^{(N)}(t))_{(i,j) \in \{1, \dots, N\}^2}$ . The local boundedness assumption means that  $\mathbf{L}^{(N)}(t)$  is bounded on any bounded time interval. Denote by  $\mathbf{W}^{(K(N))}$  the  $K(N)$ -dimensional Brownian motion modeling the sources of noise and  $\mathbf{\Lambda}^{(N)}(t)$  the  $N \times K(N)$  matrix representing the action of the noise sources on the membrane potentials of the neurons. Let us denote by  $\mathbf{V}^{(N)}(t)$  the vector of membrane potentials  $(V_1(t), \dots, V_N(t))^T$ , by  $\mathbf{J}^{(N)}$  the matrix  $(J_{ij})_{(i,j) \in \{1, \dots, N\}^2}$ , by  $S^{(N)}$  the nonlinear matricial function such that  $S^{(N)}(\mathbf{x}) = (S_{\alpha(i), \alpha(j)}(\mathbf{x}))_{i,j}$ , by  $\text{Diag}^{(N)}$  the linear operator :

$$\text{Diag}^{(N)} : \begin{cases} \mathbb{R}^{N \times N} & \mapsto \mathbb{R}^N \\ M = (M_{ij}) & \mapsto (M_{ii})_{i=1, \dots, N} \end{cases}$$

and by  $\mathbf{I}^{(N)}$  the vector of input currents.

The microscopic equation for the GM1 reads:

<sup>3</sup>This would be useful for instance for the complex interacting neuronal models, modeling the mixing of components (this is for instance the case of Jansen's  $z$  variable in equation (12)).

$$d\mathbf{V}_t^{(N)} = \left( \mathbf{L}^{(N)}(t) \cdot \mathbf{V}^{(N)} + \text{Diag}^{(N)}(\mathbf{J}^{(N)} \cdot S^{(N)}(\mathbf{V}^{(N)}(t))) + \mathbf{I}^{(N)}(t) \right) dt + \mathbf{\Lambda}^{(N)}(t) \cdot d\mathbf{W}_t^{(K(N))} \quad (1)$$

**Remark 1.** The interaction term  $\text{Diag}^{(N)}(\mathbf{J}^{(N)} \cdot S^{(N)}(\mathbf{V}^{(N)}(t)))$  reads for a given neuron  $i$  of class  $\alpha$ :  $\sum_{\beta=1}^P \sum_{j=1}^{N_\beta} J_{ij} S_{\alpha\beta}(V_j(t))$ .

For the GMk, the  $N$ -neurons network will be described by the  $N$  membrane potential  $V_1(t), \dots, V_N(t)$  and the  $N(k-1)$  derivatives of order  $\{1, \dots, k-1\}$  of these membrane potential  $V_1^{(1)}, \dots, V_N^{(1)}, \dots, V_1^{(k-1)}, \dots, V_N^{(k-1)}$ . Let us denote by  $\mathbf{V}$  the vector containing all these processes:

$$\mathbf{V}(t) = \left[ V_1, V_2, \dots, V_N, V_1^{(1)}, \dots, V_N^{(1)}, \dots, V_1^{(k-1)}, \dots, V_N^{(k-1)} \right]^T$$

We denote the partition of  $\mathbb{R}^{kN}$  in the  $N$ -dimensional vector spaces corresponding to the consecutive derivatives of  $\tilde{\mathbf{V}} = (V_1, \dots, V_N)$ :  $\mathbb{R}^{Nk} = E^{(0)} \oplus \dots \oplus E^{(k-1)}$ , and use a block decomposition in this partition:  $\mathbf{V} = (\tilde{\mathbf{V}}, \tilde{\mathbf{V}}^{(1)}, \dots, \tilde{\mathbf{V}}^{(k-1)})$ .

The new linear term will be modelled as a time-dependent matrix  $L(t)$  of dimension  $Nk \times Nk$ . External inputs and inputs coming from other neurons are only linked with the neuron's membrane potential, and only feed block component  $\tilde{\mathbf{V}}^{(k-1)}$ . The equation of the  $d$ th differential of the membrane potential,  $d < k-1$ , reads:

$$\frac{d\tilde{\mathbf{V}}^{d-1}}{dt} = \tilde{\mathbf{V}}^{(d)} \quad (2)$$

The equation governing the  $(k-1)$ th differential of the membrane potential has a linear part possibly depending on all the components of vector  $\mathbf{V}$ , and receives both external inputs and the activity of its neighbors. Keeping the same notations as before for the inputs and denoting by  $\mathbf{L}_{k-1}^{(N)}$  the linear matrix describing the action of neurons' membrane potential and their derivatives on the  $(k-1)$ th derivative of  $\tilde{\mathbf{V}}$ , the microscopic equation can be written as follows:

$$d\tilde{\mathbf{V}}^{(k-1)} = \left( \mathbf{L}_{k-1}^{(N)}(t) \cdot \mathbf{V} + \text{Diag}^{(N)}(\mathbf{J}^{(N)} \cdot S^{(N)}(\tilde{\mathbf{V}}(t))) + \mathbf{I}^{(N)}(t) \right) dt + \mathbf{\Lambda}^{(N)}(t) \cdot d\mathbf{W}_t^{(K(N))} \quad (3)$$

Hence the full equation on  $\mathbf{V}$  can be written in a simple way using block matrices.

We are interested in the limit law when  $N \rightarrow \infty$  of the vector  $V^{(N)}$  under the law of the connectivities, which we call the mean field limit. Mean field equations can be obtained easily under the local chaos hypothesis [7, 8, 18, 12], assuming that in the limit  $N \rightarrow \infty$ , the neural network behaves as if the  $(V_i)_{i \in \{1, \dots, N\}}$  were asymptotically independent and also independent of

the synaptic weights  $J_{i,j}$ . The main advantage of this rough approximation is that it gives the correct description of the asymptotic process. Indeed, the mean field equations 4 below can be rigorously proved using large deviations methods [3, 4, 14].

To be able to give an effective description of the network population by population, we need the neurons in each population to be interchangeable, i.e. to have the same probability distribution under the joint law of the multidimensional Brownian motion and the connectivity weights.

To achieve this condition for the leak matrix, a sufficient condition is that for any  $t$  the sum of the coefficients of the matrix  $\mathbf{L}(t)$  over the neurons belonging to a given population are the same for all the neurons of the same population and whatever the network size (we recall that for the GMk, we formally consider membrane potential's derivatives as neural populations). This means that there exists a matrix  $\mathbf{L}(t) = (\mathbf{L}_{\alpha\beta}(t))_{(\alpha,\beta) \in \{1, \dots, P\}^2}$  such that for all  $\alpha, \beta \in \{1, \dots, P\}$  and all neurons  $i$  of class  $\alpha$ ,<sup>4</sup>

$$\sum_{j; \alpha(j)=\beta} \mathbf{L}_{ij}^{(N)}(t) = \mathbf{L}_{\alpha\beta}(t),$$

To achieve this same condition for the noise sources, the columns of the matrix  $\mathbf{\Lambda}^{(N)}(t)$  must be permutations of the coefficients that keep the populations unchanged to keep the same noise correlation matrix between any two neurons of different population. Moreover, the lines of the matrix  $\mathbf{\Lambda}^{(N)}(t)$  are assumed to be permutations of the same pattern whatever the network size for any neuron belonging to the same class to ensure that each neuron receives the same total input in law.

Moreover, we want this law to be unchanged for  $N$  big enough, hence we assume that the number of non-uniformly zero coefficients in the matrix  $\mathbf{\Lambda}^{(N)}(t)$  for each population keeps constant when  $N$  increases. We denote this number by  $K_{\min}$  the minimal number of Brownian motions we need to use, i.e. the minimal number of independent Brownian motions used to describe the evolution of a given set of  $P$  neurons belonging to different populations.

Under these assumption, we obtain for the GM1 the following MFE for the vector  $\mathbf{V}(t) = (V_{\alpha}(t))_{\alpha=1, \dots, P}$  (see [14]):

$$d\mathbf{V}_t = \left( \mathbf{L}(t) \cdot \mathbf{V}(t) + \mathbf{U}_t^{\mathbf{V}} \cdot \mathbf{1} + \mathbf{I}(t) \right) dt + \mathbf{\Lambda}(t) \cdot d\mathbf{W}_t \quad (4)$$

where  $\mathbf{U}_t^{\mathbf{V}}$  is a matrix whose components are independent Gaussian processes  $U_{\alpha,\beta}^{\mathbf{V}}$  statistically independent of  $V$  of mean  $\bar{J}_{\alpha\beta} m_{\alpha,\beta}(t)$  where  $m_{\alpha,\beta}(t) = \mathbb{E}[S_{\alpha,\beta}(V_{\beta}(t))]$  and of covariance:

$$\text{Cov}\left(U_{\alpha,\beta}^{\mathbf{V}}(t), U_{\gamma,\delta}^{\mathbf{V}}(s)\right) =: \mathbb{1}_{\alpha=\gamma; \beta=\delta} \sigma_{\alpha\beta}^2 \Delta_{\alpha,\beta}(t, s),$$

where  $\Delta_{\alpha,\beta}(t, s) := \mathbb{E}\left[S_{\alpha,\beta}(V_{\beta}(t))S_{\alpha,\beta}(V_{\beta}(s))\right]$ .

We denoted by  $\mathbf{1}$  is the vector of  $\mathbb{R}^P$  with each component equal to 1,  $\mathbf{I}$  the input current in each population,  $\mathbf{\Lambda}(t)$  is the  $P \times K_{\min}$  matrix accounting for the noisy inputs, and  $\mathbf{W}$  a  $K_{\min}$ -dimensional Brownian motion.

<sup>4</sup>This condition is quite strong but it allows us to withstand all classical neuronal models, and hence it will be the one used in the paper. It can probably be lighten with an increase of complexity unnecessary in the present paper

**Remark 2.** This limit can be seen as a sort of generalized central limit theorem. Nevertheless, the rigorous proof of the convergence in law of the process to this limit is quite intricate (see [14]). In this paper, we admit these equations which are generally used in this field we are interested in characterizing the solutions of this equation.

For the GMk, the Mean Field Equation (MFE) can be written as equation (4) in the space  $\mathbb{R}^{kP}$ .

We can formally integrate the equation using the flow of the equation which we call Time-Ordered Exponential (TOE)  $E[L](t_0, t)$  in reference to the Dyson chronological series widely used in physics (see appendix A), and we obtain, since we assumed  $L$  locally bounded, an implicit representation of  $V$ . For the GM1, this equation reads:

$$\begin{aligned} \mathbf{V}(t) = E[L](t_0, t)\mathbf{V}(t_0) + \int_{t_0}^t E[L](s, t) \cdot (\mathbf{U}_s^{\mathbf{V}} \cdot \mathbf{1} + \mathbf{I}(s)) ds \\ + \int_{t_0}^t E[L](s, t) \cdot \mathbf{\Lambda}(s) d\mathbf{W}_s \end{aligned} \quad (5)$$

Similarly, the GMk can be written with the same formal integration. Let us denote by  $\tilde{\mathbf{U}}_s^{\mathbf{V}}$  vector with components in  $E^{(k-1)}$  are equal to  $\mathbf{U}_s^{\mathbf{V}} \cdot \mathbf{1}$  and whose other components are 0 and  $\tilde{\mathbf{I}}(s)$  the vector whose components in  $E^{k-1}$  are  $\mathbf{I}(s)$  and whose other components are 0. The solution of the GMk reads (see appendix A):

$$\begin{aligned} \tilde{\mathbf{V}}(t) = E[L](t_0, t)\tilde{\mathbf{V}}(t_0) + \int_{t_0}^t E[L](s, t) \cdot (\tilde{\mathbf{U}}_s^{\mathbf{V}} + \tilde{\mathbf{I}}(s)) ds \\ + \int_{t_0}^t E[L](s, t) \cdot \mathbf{\Lambda}(s) d\mathbf{W}_s \end{aligned} \quad (6)$$

**Remark 3.** The TOE operator is a generalization of the matrix exponential. Indeed, in the case of the GM1, if we assume that the matrix  $L$  has the commutativity property  $L(s) \cdot L(t) = L(t) \cdot L(s)$  (for instance when  $L$  does not depend on time), the implicit equation governing the law of  $V$  depending on the process  $\mathbf{U}_t^{\mathbf{V}}$  reads:

$$\begin{aligned} \mathbf{V}(t) = e^{\int_{t_0}^t \mathbf{L}(s) ds} \cdot \left[ \mathbf{V}(t_0) + \int_{t_0}^t e^{-\int_{t_0}^s \mathbf{L}(u) du} \cdot (\mathbf{U}_s^{\mathbf{V}} \cdot \mathbf{1} + \mathbf{I}(s)) ds \right. \\ \left. + \int_{t_0}^t e^{-\int_{t_0}^s \mathbf{L}(u) du} \cdot \mathbf{\Lambda}(s) d\mathbf{W}_s \right] \end{aligned} \quad (7)$$

See appendix A for a short study of this operator.

We now introduce a simpler model which is quite frequently used in the description on neural networks. All along the paper, we will prove our mathematical results for the general model whenever possible, and explain what it means for the simpler model. We think that this duality provides

the best understanding of the general results we prove. Table 1 summarizes the differences between the three models we will use.

## 1.2 Example: Jansen-Rit neural mass model

One of the motivations of this study is to characterize the global behavior of an assembly of neurons. The aim is to get a better understanding of non-invasive cortical signals like EEG or MEG. This is why we are interested to give a microscopic interpretation in terms of neuronal populations of the equations. One of the classical models of neural masses is Jansen and Rit mass model [16]. In this model, three neuronal populations interact (see figure 1).

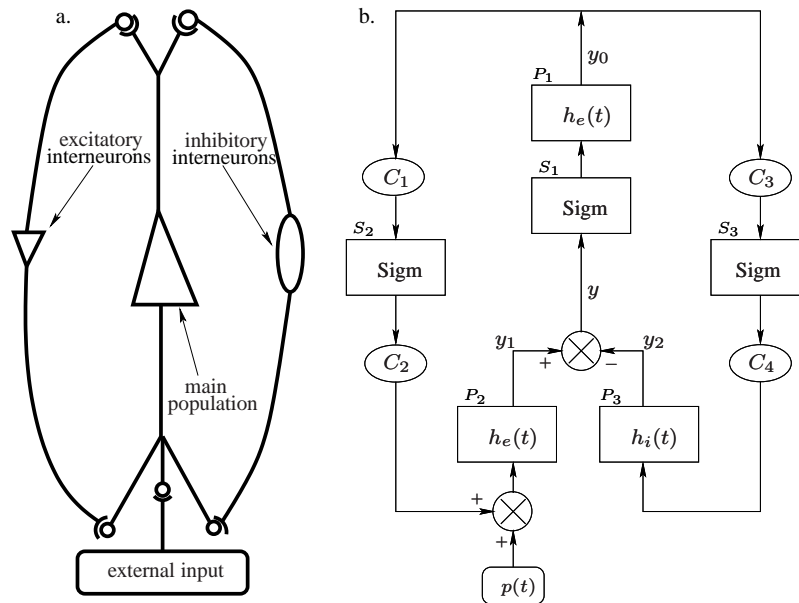


Figure 1: a. Neural mass model of a cortical unit: a population of pyramidal cells interacts with two populations of inter-neurons: an excitatory one and an inhibitory one. b. Block representation of the model. The  $h$  boxes account for the synaptic integration between neuronal populations. Sigm boxes simulate cell bodies of neurons by transforming the membrane potential of a population into an output firing rate. The constants  $C_i$  model the strength of the synaptic connections between populations.

The model features a population of pyramidal neurons (central part of figure 1.a.) that receive excitatory and inhibitory feedback from local inter-neurons and an excitatory input from neighboring cortical units and sub-cortical structures such as the thalamus. Actually the excitatory feedback must be considered as coming from both local pyramidal neurons and genuine excitatory interneurons like

spiny stellate cells. The excitatory input is represented by an arbitrary average firing rate  $p(t)$  that can be stochastic (accounting for a non specific background activity) or deterministic, accounting for some specific activity in other cortical units. The transfer functions  $h_e$  and  $h_i$  of figure 1 are sigmoidal transforms which convert the average firing rate describing the input to a population into an average excitatory or inhibitory post-synaptic potential (EPSP or IPSP). It corresponds to the synaptic integration.

The simplest model of synaptic integration is a first-order integration, which yields to exponential post-synaptic potentials:

$$h(t) = \begin{cases} Ae^{-\alpha t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

where  $\alpha$  is the time constant of the synaptic integration and  $A$  the synaptic efficiency.

This model gives rise to what we will call the first-order Jansen's model:

$$\begin{cases} \dot{y}_2 &= -\beta y_2(t) + \beta BS(C_3 y_0(t)) \\ \dot{y}_1 &= -\alpha y_1(t) + A\alpha \{p(t) + C_2 S(C_1 y_0(t))\} \\ \dot{y}_0 &= -\alpha y_0(t) + A\alpha S(y_1(t) - y_2(t)) \end{cases} \quad (8)$$

where the Sigmoid function  $S$  is given by

$$S(v) = \frac{\nu_{\max}}{1 + e^{r(v_0 - v)}}$$

where  $\nu_{\max}$  is the maximum firing rate and  $v_0$  is a voltage reference.

To better reproduce the synaptic integration, in order to reproduce well the characteristics of real EPSPs and IPSPs, van Rotterdam and colleagues [22] introduced a second order PSP:

$$h(t) = \begin{cases} \alpha\beta t e^{-\beta t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

solution of the differential equation  $\ddot{y}(t) = \alpha\beta x(t) - 2\beta\dot{y}(t) - \beta^2 y(t)$ . With this type of synaptic integration, we obtain the system:

$$\ddot{y}_2 = Bb\alpha_4 CS(\alpha_3 C y_0) - 2b\dot{y}_2 - b^2 y_2 \quad (9)$$

$$\ddot{y}_1 = Aa\alpha_2 C(p + S(\alpha_1 C y_0)) - 2a\dot{y}_1 - a^2 y_1 \quad (10)$$

$$\ddot{y}_0 = AaS(y_1 - y_2) - 2a\dot{y}_0 - a^2 y_0 \quad (11)$$

where  $A$  and  $B$  control the strength of the postsynaptic potentials,  $a$  and  $b$  their time constant and the  $C$ 's the strength of the interactions between the populations. This equation can be written in the simpler way with a change of variables and we get:

$$\begin{cases} \dot{y}_0(t) &= y_3(t) \\ \dot{y}_3(t) &= AaS[x] - 2ay_3(t) - a^2y_0(t) \\ \dot{y}_1(t) &= y_4(t) \\ \dot{y}_4(t) &= Aa\{p(t) + C_2S[C_1y_0(t)]\} - 2ay_4(t) - a^2y_1(t) \\ \dot{x}(t) &= z(t) \\ \dot{z}(t) &= Aa\{p(t) + C_2S[C_1y_0(t)]\} - 2ay_4(t) - a^2y_1(t) \\ &\quad - BbC_4S[C_3y_0(t)] - 2b(y_4(t) - z(t)) - b^2(y_1(t) - x(t)). \end{cases} \quad (12)$$

Note that considering synaptic integrations of higher order gives better models for cortical columns. For instance, Grimbert and Faugeras [13] showed that some bifurcations can appear in the second-order Jansen's model giving rise to epileptic oscillations and alpha activity, and these oscillations do not exist in the first order Jansen's model.

### 1.3 Example: The Simple Model (SM)

In the Simple Model (SM), each neuron's membrane potential decreases exponentially to its rest value if it receives no input, with a time constant  $\tau_\alpha$  depending only on the population it belongs to. The noise is simply modeled by independent Brownian motions whose standard deviation is the same for all neurons belonging to a given population.

Hence the dynamics of a given neuron  $i$  from population  $\alpha$  of the network reads:

$$dV_i^{(N)}(t) = \left[ -\frac{V_i^{(N)}(t)}{\tau_\alpha} + \sum_{\beta=1}^P \sum_{j=1}^{N_\beta} J_{i,j} S_{\alpha,\beta}(V_j^{(N)}(t)) + I_\alpha(t) \right] dt + s_\alpha dW_{(i)}(t) \quad (13)$$

Under the local chaos hypothesis, we can identify the limit process. This process is given by an implicit equation on the law of the limit process. We can prove under these assumption that if  $i_1, \dots, i_P$  are  $P$  neurons in each population (neuron  $i_\alpha$  belong to the population  $\alpha$ ), then the sequence of processes  $\left( (V_{i_1}^{(N)}(t), \dots, V_{i_P}^{(N)}(t))_{t \geq 0}^T \right)_{N \geq 1}$  converges in law to the process  $(V_1(t), \dots, V_P(t))_{t \geq 0}$  solution of the equation:

$$dV_\alpha(t) = \left( -\frac{V_\alpha(t)}{\tau_\alpha} + \sum_{\beta=1}^P U_{\alpha,\beta}^V(t) + I_\alpha(t) \right) dt + s_\alpha dW_\alpha(t) \quad \forall \alpha \in \{1, \dots, P\} \quad (14)$$

where the processes  $(W_\alpha(t))_{t \geq 0}$  are independent standard Brownian motions, where  $(U_{\alpha,\beta}^V(t); \alpha, \beta \in \{1, \dots, P\})_t$  is a diagonal Gaussian process statistically independent of  $V$  of mean  $\bar{J}_{\alpha\beta} m_{\alpha,\beta}(t)$

Model	Order of differentiation	Noise Model	Linear term
Simple Model (SM)	1	Independent inputs	Constant diagonal
General Model order 1 (GM1)	1	Correlated inputs	Time-dependent Non diagonal
General Model order $k$ (GMk)	$k \geq 1$	Correlated inputs	Time-dependent Non diagonal

Table 1: The different characteristics of the models we will study

where  $m_{\alpha,\beta}(t) = \mathbb{E}[S_{\alpha,\beta}(V_\beta(t))]$  and of covariance:

$$\text{Cov}\left(U_{\alpha,\beta}^V(t), U_{\gamma,\delta}^V(s)\right) =: \mathbb{1}_{\alpha=\gamma;\beta=\delta} \sigma_{\alpha\beta}^2 \Delta_{\alpha,\beta}(t, s),$$

where  $\Delta_{\alpha,\beta}(t, s) := \mathbb{E}\left[S_{\alpha,\beta}(V_\beta(t))S_{\alpha,\beta}(V_\beta(s))\right]$ .

This equation can be integrated implicitly and we obtain the following integral representation of the process  $V_\alpha(t)$ :

$$V_\alpha(t) = e^{-(t-t_0)/\tau_\alpha} V_\alpha(t_0) + \int_{t_0}^t e^{-(t-s)/\tau_\alpha} \left( \sum_{\beta=1}^P U_{\alpha,\beta}^V(s) + I_\alpha(s) \right) ds + s_\alpha \int_{t_0}^t e^{-(t-s)/\tau_\alpha} dW_\alpha(s) \quad (15)$$

where  $t_0$  is the initial time. It is an implicit equation on the probability distribution of  $V$ .

**Remark 4.** The case of (13) is a subcase of (1). We can easily check that the assumptions are satisfied. Indeed, the condition to be satisfied for the leak matrix is clearly satisfied. The condition for the Brownian motions is also satisfied and the minimal number of Brownian motions is  $P$ . Hence in all the paper we will mainly work on this last equations (4) and (7) which are more general than (14) and (15). All the properties we will prove for the (GM) will hence be valid for the (SM). Note also that Jansen's model with exponential synapses (8) is a particular case of GM1 and Jansen's model with a second order integration 12 is a particular case of GMk with  $k = 2$  and  $P = 3$ .

## 2 Existence and uniqueness of solutions in finite time

The mean field equation (6) we obtained is an implicit equation of the stochastic process  $(V(t))_{t \geq t_0}$ . We prove in this section that under some mild assumptions this implicit equation has a unique solution.



This solution is a fixed point in the set of  $kP$ -dimensional processes endowed with a given metric. We therefore consider the set  $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP}))$  of  $kP$ -dimensional stochastic processes and introduce a distance such that this set is complete. With this distance we prove that there exists a unique solution of the mean field equation.

We denote by  $\mathcal{X}$  the set of random variables (r.v.) with values in  $\mathbb{R}^{kP}$ .

## 2.1 A distance on $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP}))$

**Definition 2.1** (SupVar distance). The *SupVar distance* on the measure space  $\mathcal{M}_1^+(C([t_0, t], \mathbb{R}^{kP}))$  is defined by:

$$d_t(X, Y) := \sup_{s \in [t_0, t]} d_{\text{var}}(X_s, Y_s) \quad (16)$$

where  $d_{\text{var}}(X, Y)$  is the variational distance [5] on  $\mathcal{X}$ , defined by:

$$d_{\text{var}}(X, Y) := \sup \left\{ |\mathbb{E}[f(X) - f(Y)]| ; f : \mathbb{R}^{kP} \mapsto \mathbb{R} \right. \\ \left. \text{measurable and bounded such that } \|f\|_\infty \leq 1 \right\} \quad (17)$$

It is easy to show that this defines a distance. In the next proposition we prove a number of its properties that are used in the sequel.

**Proposition 2.1** (Properties of the variational distance). Let  $X, Y \in \mathcal{X}$ . The variational distance satisfies the following properties:

- (i). For all  $S : \mathbb{R}^{kP} \mapsto \mathbb{R}^{kP}$  measurable function,  $d_{\text{var}}(S(X), S(Y)) \leq d_{\text{var}}(X, Y)$ .
- (ii). Assume that the laws of the r.v.s  $X$  and  $Y$  have a density with respect to a probability measure  $\lambda$ , and let us denote by  $f_X$  and  $f_Y$  these densities. Then we have:

$$d_{\text{var}}(X, Y) = \int_{\mathbb{R}^P} |f_X - f_Y| d\lambda \quad (18)$$

- (iii). Let  $Z \in \mathcal{X}$  be a r.v. independent of  $X$  and  $Y$ , then

$$d_{\text{var}}(X + Z, Y + Z) = d_{\text{var}}(X, Y).$$

*Proof.* (i). We have:

$$\begin{aligned} d_{\text{var}}(S(X), S(Y)) &= \sup \left\{ |\mathbb{E}[f(S(X)) - f(S(Y))]| ; f : \mathbb{R}^{kP} \mapsto \mathbb{R}, \|f\|_\infty \leq 1 \right\} \\ &\leq \sup \left\{ |\mathbb{E}[f(X) - f(Y)]| ; f : \mathbb{R}^{kP} \mapsto \mathbb{R}, \|f\|_\infty \leq 1 \right\} \\ &\leq d_{\text{var}}(X, Y) \end{aligned}$$

since the set of functions  $\{f \circ S; f : \mathbb{R}^{kP} \mapsto \mathbb{R} \text{ measurable and bounded by } 1\}$  is a subset of  $\{f; f : \mathbb{R}^{kP} \mapsto \mathbb{R} \text{ measurable and bounded by } 1\}$ .

(ii). Let  $\Phi : \mathbb{R}^P \mapsto \mathbb{R}$  measurable bounded by 1. We have:

$$\begin{aligned} |\mathbb{E}[\Phi(X) - \Phi(Y)]| &= \left| \int (f_X - f_Y)\Phi d\lambda \right| \\ &\leq \int |f_X - f_Y| d\lambda \end{aligned}$$

since  $\|\Phi\|_\infty \leq 1$ , hence  $d_{\text{var}}(X, Y) \leq \|f_X - f_Y\|_{\mathbb{L}^1(\lambda)}$ . This bound is reached for  $\Phi = \text{sign}(f_X - f_Y)$  which is a measurable function bounded by 1.

(iii). To prove the last property we use the property (ii). Indeed, let  $\lambda$  be a measure dominating the laws of  $X, Y$  and  $Z$  (it is always possible to find one). Denote by  $f_X, f_Y, f_Z$  the densities of  $X, Y, Z$  with respect to  $\lambda$ . The density of the sum of two independent r.v. is the convolution of the density of each r.v. Hence we have:

$$\begin{aligned} d_{\text{var}}(X + Z, Y + Z) &= \int |f_X * f_Z - f_Y * f_Z| d\lambda \\ &= \int |(f_X - f_Y) * f_Z| d\lambda \\ &= \int |f_Z| d\lambda \int |f_X - f_Y| d\lambda \\ &= d_{\text{var}}(X, Y) \end{aligned}$$

since  $\int |f_Z| d\lambda = 1$  because it is a probability distribution. □

**Proposition 2.2.** The space  $(\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP})), d_t)$  is a complete metric space.

*Proof.* First of all, we prove that the set  $\mathcal{X}$  equipped with the variational distance is complete. To this purpose, let  $(\mu_n)_{n \geq 0}$  be a Cauchy sequence of measures and  $\lambda = \sum_{n=0}^{\infty} \frac{\mu_n}{2^n}$ . This measure dominates every  $\mu_n$  and hence every  $\mu_n$  has a density with respect to  $\lambda$ . Let  $f_n$  be the density of  $\mu_n$  with respect to  $\lambda$ .

Proposition 2.1(ii) ensures us that  $(f_n)$  is a Cauchy sequence in  $\mathbb{L}^1(\lambda)$  which is complete, hence the sequence of densities converges in this space to a function  $f \in \mathbb{L}^1(\lambda)$ . Let  $\mu := f d\lambda$ . We have  $d_{\text{var}}(\mu_n, \mu) = \|f_n - f\|_{\mathbb{L}^1(\lambda)} \xrightarrow{n \rightarrow \infty} 0$ , hence  $\mu_n \rightarrow \mu$  for the variational distance,  $(\mathcal{X}, d_{\text{var}})$  is a complete metric space.

The set  $(\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP})), d_T)$  is isomorphic to the set  $\mathbb{L}^\infty([t_0, T], (\mathcal{X}, d_{\text{var}}))$ , hence  $(\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP})), d_T)$  is also complete, as an  $L^\infty$  set of functions with values in a complete space. □

In the next two lemmas we compute the variational distance between two Gaussian random variables. These results are essential in the proof of existence and uniqueness of a solution of the mean field equation. We first compute an explicit bound in the case of unidimensional Gaussian r.v. and give a closed-form bound, and then generalize to the case of multivariate Gaussian random variables.

**Lemma 2.3** (Variational distance between two unidimensional Gaussian r.v.). Let us consider  $G_i := \mathcal{N}(\mu_i, \sigma_i)$  two unidimensional Gaussian random variables of mean  $\mu_i$  and standard deviation  $\sigma_i$ ,  $i = a, b$ . Let  $\mu_1 = \min(\mu_a, \mu_b)$ ,  $\mu_2 = \max(\mu_a, \mu_b)$ ,  $\sigma_1 = \min(\sigma_a, \sigma_b)$  and  $\sigma_2 = \max(\sigma_a, \sigma_b)$ . Assume  $\sigma_1 > 0$ . Their variational distance  $d_{\text{var}}(G_a, G_b)$  satisfies the following inequality:

$$d_{\text{var}}(G_a, G_b) \leq |\mu_a - \mu_b| K_m(\mu_1, \mu_2, \sigma_1, \sigma_2) + |\sigma_a - \sigma_b| S_m(\mu_1, \mu_2, \sigma_1, \sigma_2), \quad (19)$$

where

$$K_m(\mu_1, \mu_2, \sigma_1, \sigma_2) = \frac{\sigma_2}{\sigma_1^3} |\mu_1 - \mu_2| + \frac{\sigma_2^2}{\sigma_1^3} \sqrt{\frac{2}{\pi}} + \frac{\sigma_2 \mu_2}{\sigma_1^3}$$

$$S_m(\mu_1, \mu_2, \sigma_1, \sigma_2) = \frac{\sigma_2}{\sigma_1^4} ((\mu_1 - \mu_2)^2 + 2\sigma_2^2) + 3\sqrt{\frac{2}{\pi}} \frac{\sigma_2^3}{\sigma_1^4},$$

*Proof.* A Gaussian random variable has a density with respect to Lebesgue's measure, hence we use the equality between the variational norm on random variables and the  $\mathbb{L}^1(\lambda)$  norm on the densities, where  $\lambda$  stands here for the Lebesgue's measure:

$$d_{\text{var}}(G_a, G_b) = \int_{-\infty}^{\infty} \left| \frac{e^{-\frac{(x-\mu_a)^2}{2\sigma_a^2}}}{\sqrt{2\pi}\sigma_a} - \frac{e^{-\frac{(x-\mu_b)^2}{2\sigma_b^2}}}{\sqrt{2\pi}\sigma_b} \right| dx$$

We use the mean value theorem and compute the partial derivatives of the density with respect to  $\mu$  and  $\sigma$ . Let us note  $P(\mu, \sigma, x) := \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}$ . We have:

$$|P(\mu_a, \sigma_a, x) - P(\mu_b, \sigma_b, x)| \leq \sup_{\mu \in [\mu_1, \mu_2], \sigma \in [\sigma_1, \sigma_2]} \left| \frac{\partial P(\mu, \sigma, x)}{\partial \mu} \right| |\mu_a - \mu_b|$$

$$+ \sup_{\mu \in [\mu_1, \mu_2], \sigma \in [\sigma_1, \sigma_2]} \left| \frac{\partial P(\mu, \sigma, x)}{\partial \sigma} \right| |\sigma_a - \sigma_b|$$

Since

$$\frac{\partial P(\mu, \sigma, x)}{\partial \mu} = \frac{(x - \mu)}{\sigma^2} P(\mu, \sigma, x),$$

we find an upperbound for the partial derivative sharply bounding each term of the expression:

$$\sup_{\mu \in [\mu_1, \mu_2], \sigma \in [\sigma_1, \sigma_2]} \left| \frac{\partial P(\mu, \sigma, x)}{\partial \mu} \right| \leq \begin{cases} \left| \frac{x - \mu_2}{\sigma_1^2} \right| \frac{e^{-(x-\mu_1)/2\sigma_2^2}}{\sqrt{2\pi}\sigma_1} & x \in (-\infty, (\mu_1 + \mu_2)/2] \\ \left| \frac{x - \mu_1}{\sigma_1^2} \right| \frac{e^{-(x-\mu_2)/2\sigma_2^2}}{\sqrt{2\pi}\sigma_1} & x \in [(\mu_1 + \mu_2)/2, \infty) \end{cases}$$

which we note  $M(x)$

Similarly, we obtain:

$$\frac{\partial P(\mu, \sigma, x)}{\partial \sigma} = \frac{(x - \mu)^2 - \sigma^2}{\sigma^3} P(\mu, \sigma, x),$$

and hence the following upperbound

$$\sup_{\mu \in [\mu_1, \mu_2], \sigma \in [\sigma_1, \sigma_2]} \left| \frac{\partial P(\mu, \sigma, x)}{\partial \sigma} \right| \leq \begin{cases} \frac{(x - \mu_2)^2 + \sigma_2^2}{\sigma_1^3} \frac{e^{-(x - \mu_1)/2\sigma_2^2}}{\sqrt{2\pi}\sigma_1} & x \in (-\infty, (\mu_1 + \mu_2)/2] \\ \frac{(x - \mu_1)^2 + \sigma_2^2}{\sigma_1^3} \frac{e^{-(x - \mu_2)/2\sigma_2^2}}{\sqrt{2\pi}\sigma_1} & x \in [(\mu_1 + \mu_2)/2, \infty) \end{cases}$$

which we note  $S(x)$ .

Finally we have:

$$d_{\text{var}}(G_a, G_b) \leq |\mu_a - \mu_b| \int_{\mathbb{R}} M(x) dx + |\sigma_a - \sigma_b| \int_{\mathbb{R}} S(x) dx$$

The two integrals can be computed in closed form and upperbounded:

$$\begin{aligned} \int_{\mathbb{R}} M(x) dx &\leq \frac{\sigma_2}{\sigma_1^3} |\mu_1 - \mu_2| + \frac{\sigma_2^2}{\sigma_1^3} \sqrt{\frac{2}{\pi}} + \frac{\sigma_2 \mu_2}{\sigma_1^3} \\ \int_{\mathbb{R}} S(x) dx &\leq \frac{\sigma_2}{\sigma_1^4} ((\mu_1 - \mu_2)^2 + 2\sigma_2^2) + 3\sqrt{\frac{2}{\pi}} \frac{\sigma_2^3}{\sigma_1^4} \end{aligned}$$

□

**Lemma 2.4** (Variational distance between multivariate Gaussian r.v.). Let  $X = \mathcal{N}(\mu^X, (\Sigma^X)^{-1})$  and  $Y = \mathcal{N}(\mu^Y, (\Sigma^Y)^{-1})$  be two multivariate non-degenerate Gaussian random variables in  $\mathbb{R}^P$ . Let  $\mu_1$  and  $\mu_2$  be respectively the minimal and maximal values of the coordinates of the mean vectors of  $X$  and  $Y$  and define  $\Sigma_{\max}$  and  $\Sigma_{\min}$  the matrix respectively of general term  $\max(\Sigma_{ij}^X, \Sigma_{ij}^Y)$  and  $\min(\Sigma_{ij}^X, \Sigma_{ij}^Y)$ .

There exists a constant  $K(\mu_1, \mu_2, \Sigma_{\min}, \Sigma_{\max})$  such that the variational distance between  $X$  and  $Y$  satisfies the inequality:

$$d_{\text{var}}(X, Y) \leq K(\mu_1, \mu_2, \Sigma_{\min}, \Sigma_{\max}) \left( \sup_{i=1 \dots P} |\mu_i^X - \mu_i^Y| + \sup_{i,j=1 \dots P} |\Sigma_{ij}^X - \Sigma_{ij}^Y| \right)$$

*Proof.* We note

$$P(x, \mu, \Sigma) = \frac{\sqrt{\det(\Sigma)}}{(2\pi)^{P/2}} e^{-\frac{1}{2}(x - \mu)^T \Sigma (x - \mu)}$$

the density of a  $P$ -dimensional Gaussian variable with mean vector  $\mu$  and covariance matrix  $\Sigma^{-1}$ . We need to compute  $\int_{\mathbb{R}^P} |P(x, \mu^X, \Sigma^X) - P(x, \mu^Y, \Sigma^Y)| dx$ . The zeroth order Taylor expansion with integral remainder can be written

$$P(x, \mu^X, \Sigma^X) - P(x, \mu^Y, \Sigma^Y) = \left( \int_0^1 \frac{\partial P}{\partial \mu}(x, \mu^Y + \zeta(\mu^X - \mu^Y), \Sigma^X) d\zeta \right) \cdot (\mu^X - \mu^Y) + \left( \int_0^1 \frac{\partial P}{\partial \Sigma}(x, \mu^X, \Sigma^Y + \zeta(\Sigma^X - \Sigma^Y)) d\zeta \right) \cdot (\Sigma^X - \Sigma^Y),$$

where the  $\cdot$ s indicate the obvious inner products. Let us consider first the term with the partial derivative with respect to the mean. Using standard calculus, we find that

$$\frac{\partial P}{\partial \mu}(x, \mu^Y + \zeta(\mu^X - \mu^Y), \Sigma^X) = P(x, \mu^Y + \zeta(\mu^X - \mu^Y), \Sigma^X) \Sigma^X (x - \mu^Y - \zeta(\mu^X - \mu^Y))$$

hence

$$\left| \left( \int_{\mathbb{R}^P} \int_0^1 \frac{\partial P}{\partial \mu}(x, \mu^Y + \zeta(\mu^X - \mu^Y), \Sigma^X) dx d\zeta \right) \cdot (\mu^X - \mu^Y) \right| \leq \frac{1}{2} (\mu^X - \mu^Y)^T \Sigma^X (\mu^X - \mu^Y) + \int_{\mathbb{R}^P} \int_0^1 P(x, \mu^Y + \zeta(\mu^X - \mu^Y), \Sigma^X) |(\mu^X - \mu^Y)^T \Sigma^X (x - \mu^Y)| dx d\zeta$$

We use the upper bound

$$|(\mu^X - \mu^Y)^T \Sigma^X (x - \mu^Y)| \leq \|\Sigma^X\| \|\mu^X - \mu^Y\| \|x - \mu^Y\|,$$

where  $\|\cdot\|$  is the Euclidean norm, and

$$\|\mu^X - \mu^Y\| \leq \sqrt{P} |\mu^X - \mu^Y|_\infty,$$

to obtain

$$\left| \left( \int_0^1 \frac{\partial P}{\partial \mu}(x, \mu^Y + \zeta(\mu^X - \mu^Y), \Sigma^X) d\zeta \right) \cdot (\mu^X - \mu^Y) \right| \leq \sqrt{P} \|\Sigma^X\| \left( \frac{1}{2} \|\mu^X - \mu^Y\| + K_1 \right) |\mu^X - \mu^Y|_\infty,$$

where  $K_1$  is given by

$$K_1 = \int_{\mathbb{R}^P} \int_0^1 P(x, \mu^Y + \zeta(\mu^X - \mu^Y), \Sigma^X) \|x - \mu^Y\| dx d\zeta.$$

An upperbound  $L_1$  for  $K_1$  can be found as follows. If  $\mu^X = \mu^Y$  then

**[Dans ce cas, ca fait 0 de toutes facons, puisque Taylor ne fait pas intervenir de terme en la difference de  $\mu$ s]**

$$L_1 = \int_{\mathbb{R}^P} P(x, \mu^Y, \Sigma^X) \|x - \mu^Y\| dx.$$

else

$$\begin{aligned} P(x, \mu^Y + \zeta(\mu^X - \mu^Y), \Sigma^X) &\leq \\ \frac{(2\pi)^{P/2}}{\sqrt{\det(\Sigma^X)}} P(x, \mu^Y, \Sigma^X) P(\mu^X, \mu^Y, \Sigma^X) e^{\zeta(x - \mu^Y)\Sigma^X(\mu^X - \mu^Y)} &\leq \\ P(x, \mu^Y, \Sigma^X) e^{\zeta(x - \mu^Y)\Sigma^X(\mu^X - \mu^Y)} & \end{aligned}$$

Therefore

$$\int_0^1 P(x, \mu^Y + \zeta(\mu^X - \mu^Y), \Sigma^X) d\zeta \leq \frac{P(x, \mu^Y, \Sigma^X)}{(x - \mu^Y)\Sigma^X(\mu^X - \mu^Y)} \left(1 - e^{(x - \mu^Y)\Sigma^X(\mu^X - \mu^Y)}\right),$$

and

$$L_1 = \int_{\mathbb{R}^P} \frac{P(x, \mu^Y, \Sigma^X) \|x - \mu^Y\|}{(x - \mu^Y)\Sigma^X(\mu^X - \mu^Y)} \left(1 - e^{(x - \mu^Y)\Sigma^X(\mu^X - \mu^Y)}\right) dx$$

The reader can verify that the integral of the righthand side exists. It is a function of  $\mu^X$ ,  $\mu^Y$  and  $\Sigma^X$ .

We next consider the term with the partial derivative with respect to the covariance matrix. We obtain:

$$\frac{\partial P(x, \mu, \Sigma)}{\partial \Sigma_{ij}} = \frac{1}{2} \left( \frac{1}{\det(\Sigma)} \frac{\partial \det(\Sigma)}{\partial \Sigma_{ij}} - (x - \mu)^T \frac{\partial \Sigma}{\partial \Sigma_{ij}} (x - \mu) \right) P(x, \mu, \Sigma)$$

We use the fact that  $\Sigma$  is a symmetric matrix, and hence that  $(x - \mu)^T \frac{\partial \Sigma}{\partial \Sigma_{ij}} (x - \mu) = 2(x - \mu)_i (x - \mu)_j$  and we recall that the differential of the determinant of a function is given by Jacobi's formula:  $d(\det(\Sigma)) = \text{Trace}(\text{Adj}(\Sigma) \cdot d\Sigma)$  where  $\text{Adj}(\Sigma)$  is the adjugate matrix of  $\Sigma$ :

$$\frac{\partial \det(\Sigma)}{\partial \Sigma_{ij}} = 2(\text{Adj}(\Sigma))_{ij}$$

We write next

$$\begin{aligned} &\left| \int_{\mathbb{R}^P} \left( \int_0^1 \frac{\partial P}{\partial \Sigma} (x, \mu^X, \Sigma^Y + \zeta(\Sigma^X - \Sigma^Y)) d\zeta \right) \cdot (\Sigma^X - \Sigma^Y) dx \right| \leq \\ &\int_{\mathbb{R}^P} \int_0^1 P(x, \mu^X, \Sigma^Y + \zeta(\Sigma^X - \Sigma^Y)) |(x - \mu^X)^T (\Sigma^X - \Sigma^Y) (x - \mu^X)| dx d\zeta + \\ &\int_0^1 \frac{1}{\det(\Sigma^Y + \zeta(\Sigma^X - \Sigma^Y))} \left| \sum_{i,j} (\text{Adj}(\Sigma^Y + \zeta(\Sigma^X - \Sigma^Y)))_{ij} ((\Sigma^X)_{ij} - (\Sigma^Y)_{ij}) \right| d\zeta \end{aligned}$$

We next use the fact that the function  $\zeta \rightarrow \log \det(\zeta \Sigma^X + (1 - \zeta) \Sigma^Y)$  is concave on  $[0, 1]$ :

$$\log \det(\zeta \Sigma^X + (1 - \zeta) \Sigma^Y) \geq \zeta \log \det(\Sigma^X) + (1 - \zeta) \log \det(\Sigma^Y), \zeta \in [0, 1]$$

Therefore

$$\frac{1}{\det(\Sigma^Y + \zeta(\Sigma^X - \Sigma^Y))} \leq \frac{1}{(\det(\Sigma^X))^\zeta (\det(\Sigma^Y))^{(1-\zeta)}}$$

Let  $\sigma_{\min}^X$  (respectively  $\sigma_{\min}^Y$ ) be the smallest eigenvalue of  $\Sigma^X$  (respectively of  $\Sigma^Y$ ) and  $\sigma$  the smallest of  $\sigma_{\min}^X$  and  $\sigma_{\min}^Y$ :

$$\frac{1}{\det(\Sigma^Y + \zeta(\Sigma^X - \Sigma^Y))} \leq \frac{1}{\sigma^P}$$

We now consider the term  $(Adj(\Sigma^Y + \zeta(\Sigma^X - \Sigma^Y)))_{ij}$ . It is a polynomial of degree  $P - 1$  in the variable  $\zeta$  whose coefficients are polynomials of degree  $P - 1$  in the variables  $(\Sigma^X)_{kl}$  and  $(\Sigma^Y)_{kl}$ ,  $k, l \neq i, j$ . Let  $\Sigma_{\max} = \max_{i,j} (|\Sigma_{ij}^X|, |\Sigma_{ij}^Y|)$ , we have

$$\int_0^1 \frac{1}{\det(\Sigma^Y + \zeta(\Sigma^X - \Sigma^Y))} \left| \sum_{i,j} (Adj(\Sigma^Y + \zeta(\Sigma^X - \Sigma^Y)))_{ij} ((\Sigma^X)_{ij} - (\Sigma^Y)_{ij}) \right| d\zeta \leq \frac{f(\Sigma_{\max})}{\sigma^P} |\Sigma^X - \Sigma^Y|_\infty,$$

where  $f$  is a polynomial with positive coefficients.

We have

$$\int_{\mathbb{R}^P} \int_0^1 P(x, \mu^X, \Sigma^Y + \zeta(\Sigma^X - \Sigma^Y)) |(x - \mu^X)^T (\Sigma^X - \Sigma^Y) (x - \mu^X)| \leq \int_{\mathbb{R}^P} \int_0^1 P(x, \mu^X, \Sigma^Y + \zeta(\Sigma^X - \Sigma^Y)) \|x - \mu^X\|^2 dx d\zeta |\Sigma^X - \Sigma^Y|_\infty$$

Putting all this together we obtain

$$\left| \int_{\mathbb{R}^P} \left( \int_0^1 \frac{\partial P}{\partial \Sigma} (x, \mu^X, \Sigma^Y + \zeta(\Sigma^X - \Sigma^Y)) d\zeta \right) \cdot (\Sigma^X - \Sigma^Y) dx \right| \leq (K_2 + \frac{f(\Sigma_{\max})}{\sigma^P}) |\Sigma^X - \Sigma^Y|_\infty,$$

where

$$K_2 = \int_{\mathbb{R}^P} \int_0^1 P(x, \mu^X, \Sigma^Y + \zeta(\Sigma^X - \Sigma^Y)) \|x - \mu^X\|^2 dx d\zeta.$$

An upperbound  $L_2$  for  $K_2$  can be found as follows. If  $\Sigma^X \neq \Sigma^Y$  then

$$L_2 = K_2 = \int_{\mathbb{R}^P} P(x, \mu^X, \Sigma^Y) \|x - \mu^X\|^2 dx,$$

else [**Meme remarque**]

$$P(x, \mu^X, \Sigma^Y + \zeta(\Sigma^X - \Sigma^Y)) = \sqrt{\frac{\det(\Sigma^Y + \zeta(\Sigma^X - \Sigma^Y))}{\det(\Sigma^Y)}} P(x, \mu^X, \Sigma^Y) e^{\zeta(x - \mu^X)^T (\Sigma^X - \Sigma^Y)(x - \mu^X)}$$

$\det(\Sigma^Y + \zeta(\Sigma^X - \Sigma^Y))$  is positive and can be bounded by a positive function  $g(\Sigma_{\max})$ , therefore

$$L_2 = \sqrt{\frac{g(\Sigma_{\max})}{\sigma^P}} \int_{\mathbb{R}^P} \frac{P(x, \mu^X, \Sigma^Y)}{(x - \mu^X)^T (\Sigma^X - \Sigma^Y)(x - \mu^X)} \left(1 - e^{(x - \mu^X)^T (\Sigma^X - \Sigma^Y)(x - \mu^X)}\right) dx.$$

The reader can verify that the integral in the righthand side exists. It is a function of  $\mu^X$ ,  $\Sigma^X$  and  $\Sigma^Y$ .

We have therefore

$$d_{\text{var}}(X, Y) \leq \left(\frac{1}{2} |\Sigma^X|_{\infty} |\mu^X - \mu^Y| + L_1\right) |\mu^X - \mu^Y|_{\infty} + (L_2 + \frac{f(\Sigma_{\max})}{\sigma^P}) |\Sigma^X - \Sigma^Y|_{\infty}$$

□

**Remark 5.** It is important to note that in the constant we found in the previous lemma, the factor  $\frac{1}{\det(\Sigma)}$  appears in all the constants we can compute. Hence the non-degeneracy condition of the random variables is essential to have such a bound. It corresponds to the factor  $\frac{1}{\sigma_1}$  in the constant we made explicit in the lemma 2.3. To apply this lemma, we will need to ensure that all the coefficients of the mean and covariance matrix are bounded and that the determinant of the covariance matrix has a lower bound.

Armed with this definition and properties, we now prove that there exists a unique solution for the mean field equations.

## 2.2 Existence and uniqueness of solutions of the mean field equations

**Definition 2.2.** In the case of the GM1, we consider the map on the probability distribution  $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^P))$  defined by:

$$F_1 : \begin{cases} \mathcal{M}_1^+(C([t_0, T], \mathbb{R}^P)) & \mapsto \mathcal{M}_1^+(C([t_0, T], \mathbb{R}^P)) \\ (X_t)_{t \in [t_0, T]} & \mapsto F_1(X)_{t \in [t_0, T]} \end{cases} \quad (20)$$

such that :



$$F_1(X)_t = E[L](t_0, t) \cdot X_{t_0} + \int_{t_0}^t E[L](s, t) \cdot (\mathbf{U}_s^X \cdot \mathbf{1} + \mathbf{I}(s)) ds + \int_{t_0}^t E[L](s, t) \cdot \Lambda(s) d\mathbf{W}_s \quad (21)$$

where  $\mathbf{U}_s^X = (U_{\alpha, \beta}^X)_{\alpha, \beta}$  is the Gaussian diagonal process of parameters:

$$\begin{cases} \mathbb{E} [U_{\alpha, \beta}^X(t)] &= \bar{J}_{\alpha, \beta} \mathbb{E} [S_{\alpha, \beta}(X_t^\beta)] \\ \text{Cov} [U_{\alpha, \beta}^X(s), U_{\alpha, \beta}^X(t)] &= \sigma_{\alpha, \beta}^2 \mathbb{E} [S_{\alpha, \beta}(X_\beta(t)) S_{\alpha, \beta}(X_\beta(s))] \end{cases} \quad (22)$$

In the case of the GMk, we define can define function  $F_k$  such that

$$F_k(X)_t = E[L](t_0, t) \cdot X_{t_0} + \int_{t_0}^t E[L](s, t) \cdot (\tilde{\mathbf{U}}_s^X + \tilde{\mathbf{I}}(s)) ds + \int_{t_0}^t E[L](s, t) \cdot \Lambda(s) d\mathbf{W}_s$$

where  $\tilde{\mathbf{U}}_s^X$  and  $\tilde{\mathbf{I}}(s)$  are defined in section 1

Note that the mean of  $U_{\alpha, \beta}^X$  is contained in  $[\mu_1, \mu_2]$  where  $\mu_1 := -\max(|\bar{J}_{\alpha, \beta}|)$  and  $\mu_2 := -\mu_1$ . Its standard deviation is always lower than  $\sigma_2 := \max(\sigma_{\alpha, \beta})$ .

**Theorem 2.5** (Regularity for the GM1). *Let  $X$  and  $Y$  two processes of  $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^P))$  having the same initial condition at  $t = t_0$  and such that the related Gaussian processes  $U_{\alpha, \beta}^X(t)$  and  $U_{\alpha, \beta}^Y(t)$  are not degenerated and have their means bounded in  $[\mu_1, \mu_2]$ , standard deviations in  $[\sigma_1, \sigma_2]$  where  $\sigma_1 > 0$  for all  $t \in [t_0, T]$ .*

*Then there exist a constant  $K(\mu_1, \mu_2, \sigma_1, \sigma_2)$  such that :*

$$d_{\text{var}}(F_1(X)_t, F_1(Y)_t) \leq K(\mu_1, \mu_2, \sigma_1, \sigma_2) \int_{t_0}^t f(s, t) d_{\text{var}}(X_s, Y_s) ds, \quad (23)$$

where the function  $f(s, t)$  has the expression:

$$f(s, t) = e^{\|L\|_\infty(t-s)} + e^{\|L\|_\infty(t-s)} \int_{t_0}^t e^{\|L\|_\infty(t-u)} du.$$

where  $\|L\|_\infty$  is the operator norm of  $L$  induced by the uniform norm on  $\mathbb{R}^P$ .

*Proof.* We have, using the property 2.1:

$$d_{\text{var}}(F_1(X)_t, F_1(Y)_t) = d_{\text{var}}\left(\int_{t_0}^t E[L](s, t) (\mathbf{U}_s^X \cdot \mathbf{1}) ds, \int_{t_0}^t E[L](s, t) (\mathbf{U}_s^Y \cdot \mathbf{1}) ds\right)$$

The two processes inside the norm in the righthand side of the equality are Gaussian processes, whose mean is  $\int_{t_0}^t E[L](s, t) \left( \sum_{\beta=1}^P \bar{J}_{\alpha\beta} \mathbb{E} [S_{\alpha,\beta}(X_s^\beta)] \right)_{\alpha=1 \dots P} ds$  for the term corresponding to the process  $X$ , which we denote  $\mu^X$ . We can see that the components of this vector are contained in an interval  $[\mu_{\min}(\mu_1, \mu_2), \mu_{\max}(\mu_1, \mu_2)]$  where  $\mu_{\min}$  and  $\mu_{\max}$  only depend on the values of  $\mu_1$  and  $\mu_2$  and on the parameters of the function  $L(t)$ .

Its covariance matrix reads:

$$\begin{aligned} \Sigma^X &= \mathbb{E} \left[ \left( \int_{t_0}^t E[L](s, t) \left( U_s^X \cdot \mathbf{1} \right) ds \right) \cdot \left( \int_{t_0}^t E[L](s, t) \left( U_s^X \cdot \mathbf{1} \right) ds \right)^T \right] \\ &= \int_{t_0}^t \int_{t_0}^t E[L](s, t) \text{diag} \left( \sum_{\beta=1}^P \sigma_{\alpha\beta}^2 \mathbb{E} [S_{\alpha,\beta}(X_\beta(s)) S_{\alpha,\beta}(X_\beta(u))] \right), \alpha = 1 \dots P \\ &\quad E[L](u, t)^T ds du \end{aligned}$$

where we denoted by  $\text{diag}(x_1, \dots, x_n)$  the diagonal matrix whose  $i$ th diagonal coefficient is  $x_i$ . All the coefficients of this matrix are bounded by values only depending on  $\sigma_1$  and  $\sigma_2$  and the parameters of the matrix  $L(t)$ . Let us denote the two component per component extremal matrices  $\Sigma_{\min}(\sigma_1, \sigma_2)$  and  $\Sigma_{\max}(\sigma_1, \sigma_2)$ .

Hence we can use lemma 2.4 to bound this distance. We denote by  $\tilde{K}(\mu_1, \mu_2, \sigma_1, \sigma_2)$  the value of the constant found in lemma 2.4:

$$\tilde{K}(\mu_1, \mu_2, \sigma_1, \sigma_2) := K(\mu_{\min}(\mu_1, \mu_2), \mu_{\max}(\mu_1, \mu_2), \Sigma_{\min}(\sigma_1, \sigma_2), \Sigma_{\max}(\sigma_1, \sigma_2))$$

Let us now denote for a  $P \times P$  matrix  $A$ ,  $\|A\|_\infty$  the norm of  $A$  induced by the uniform norm on  $\mathbb{R}^P$  and by  $\|A\|_\infty$  the norm of  $A$  as an operator on the matrix  $P \times P$  induced by the uniform norm (see appendix B).

We have:

$$\begin{aligned}
d_{\text{var}}(F_1(X)_t, F_1(Y)_t) &\leq \tilde{K}(\mu_1, \mu_2, \sigma_1, \sigma_2) \left( \sup_{\alpha=1\dots P} |\mu_\alpha^X - \mu_\alpha^Y| + \sup_{\alpha,\beta=1\dots P} |\Sigma_{\alpha,\beta}^X - \Sigma_{\alpha,\beta}^Y| \right) \\
&= \tilde{K}(\mu_1, \mu_2, \sigma_1, \sigma_2) (\|\mu^X - \mu^Y\|_\infty + \|\Sigma^X - \Sigma^Y\|_\infty) \\
&\leq \tilde{K}(\mu_1, \mu_2, \sigma_1, \sigma_2) \\
&\quad \left( \int_{t_0}^t \|E[L](s, t)\|_\infty \left\| \left( \sum_{\beta=1}^P \bar{J}_{\alpha\beta} \mathbb{E}[S_{\alpha,\beta}(X_s^\beta) - S_{\alpha,\beta}(Y_s^\beta)] \right)_{\alpha=1\dots P}^T \right\|_\infty \right. \\
&\quad + \int_{t_0}^t \int_{t_0}^t \|E[L](s, t)\|_\infty \left\| \text{diag} \left( \sum_{\beta=1}^P \sigma_{\alpha\beta}^2 \mathbb{E}[S_{\alpha,\beta}(X_\beta(s)) S_{\alpha,\beta}(X_\beta(u)) \right. \right. \\
&\quad \left. \left. - S_{\alpha,\beta}(Y_\beta(s)) S_{\alpha,\beta}(Y_\beta(u)) \right], \alpha = 1 \dots P \right) E[L](u, t)^T \right\|_\infty ds du \Big) \\
&= \tilde{K}(\mu_1, \mu_2, \sigma_1, \sigma_2) \\
&\quad \left( \int_{t_0}^t \|E[L](s, t)\|_\infty \left\| \left( \sum_{\beta=1}^P \bar{J}_{\alpha\beta} \mathbb{E}[S_{\alpha,\beta}(X_s^\beta) - S_{\alpha,\beta}(Y_s^\beta)] \right)_{\alpha=1\dots P}^T \right\|_\infty \right. \\
&\quad + \int_{t_0}^t \int_{t_0}^t \|E[L](s, t)\|_\infty \|E[L](u, t)\|_\infty \left\| \text{diag} \left( \sum_{\beta=1}^P \sigma_{\alpha\beta}^2 \mathbb{E}[S_{\alpha,\beta}(X_\beta(s)) S_{\alpha,\beta}(X_\beta(u)) \right. \right. \\
&\quad \left. \left. - S_{\alpha,\beta}(Y_\beta(s)) S_{\alpha,\beta}(Y_\beta(u)) \right], \alpha = 1 \dots P \right) \right\|_\infty ds du \Big)
\end{aligned}$$

Now using the fact that

$$\begin{aligned}
&S_{\alpha,\beta}(X_\beta(s)) S_{\alpha,\beta}(X_\beta(u)) - S_{\alpha,\beta}(Y_\beta(s)) S_{\alpha,\beta}(Y_\beta(u)) \\
&= S_{\alpha,\beta}(X_\beta(s)) (S_{\alpha,\beta}(X_\beta(u)) - S_{\alpha,\beta}(Y_\beta(u))) + S_{\alpha,\beta}(Y_\beta(u)) (S_{\alpha,\beta}(X_\beta(s)) - S_{\alpha,\beta}(Y_\beta(s)))
\end{aligned}$$

and the fact that  $S_{\alpha,\beta}$  is bounded (we denote  $\|S_{\alpha,\beta}\|_\infty$  the supremum of the absolute value of  $S_{\alpha,\beta}$ ), we obtain:

$$\begin{aligned}
d_{\text{var}}(F_1(X)_t, F_1(Y)_t) &\leq \tilde{K}(\mu_1, \mu_2, \sigma_1, \sigma_2) \left( \sup_{\alpha \in \{1, \dots, P\}} \sum_{\beta=1}^P |\bar{J}_{\alpha\beta}| \int_{t_0}^t \|E[L](s, t)\|_{\infty} \|S_{\alpha\beta}\|_{\infty} d_{\text{var}}(X_s, Y_s) ds \right. \\
&\quad \left. + \sup_{\alpha \in \{1, \dots, P\}} \sum_{\beta=1}^P \sigma_{\alpha\beta}^2 \int_{t_0}^t \int_{t_0}^t \|E[L](s, t)\|_{\infty} \|E[L](u, t)\|_{\infty} \|S_{\alpha\beta}\|_{\infty}^2 (d_{\text{var}}(X_s, Y_s) + d_{\text{var}}(X_u, Y_u)) ds du \right) \\
&\leq \tilde{K}(\mu_1, \mu_2, \sigma_1, \sigma_2) \left( \sup_{\alpha} \sum_{\beta} \|S_{\alpha\beta}\|_{\infty} |\bar{J}_{\alpha\beta}| + 2 \sup_{\alpha} \sum_{\beta} \|S_{\alpha\beta}\|_{\infty}^2 |\sigma_{\alpha\beta}^2| \right) \\
&\quad \int_{t_0}^t \left( \|E[L](s, t)\|_{\infty} + \|E[L](s, t)\|_{\infty} \int_{t_0}^t \|E[L](u, t)\|_{\infty} du \right) d_{\text{var}}(X_s, Y_s) ds
\end{aligned}$$

We conclude using the result of proposition A.1 (iii) and the lemma B.1.  $\square$

**Corollary 2.6** (Regularity for the GMk). In the case of the GMk, the same result is valid.

*Proof.* In this case, standard deviation of the Gaussian processes  $\tilde{U}^X$  and  $\tilde{U}^Y$  are no more non-degenerate. Nevertheless, the degenerate components are the same in the two multivariate Gaussian processes. Hence the variational distance between these two processes at a given time  $s$  is equal to the variational distance between  $U_s^X$  and  $U_s^Y$ . Indeed, some components are Dirac distributions at 0 and hence for any measurable bounded function of  $\mathbb{R}^{kP}$  which are bounded by 1, we have:

$$\mathbb{E} \left[ \left| f(\tilde{U}_s^X) - f(\tilde{U}_s^Y) \right| \right] = \mathbb{E} \left[ \left| \hat{f}(U_s^X) - \hat{f}(U_s^Y) \right| \right]$$

where  $\hat{f}$  is the measurable function of  $\mathbb{R}^P$  such that for any vector  $\mathbf{x} = (x_1, \dots, x_P)$  we have  $\hat{f}(\mathbf{x}) = f(\bar{x})$  where  $\bar{x}$  is the vector of  $\mathbb{R}^{kP}$  having the values of  $\mathbf{x}$  in  $E^{k-1}$  and 0 otherwise. Hence the variational distance are equal since the two sets of functions are the same. Hence the result of theorem 2.5 holds for the GMk and we have:

$$\begin{aligned}
d_{\text{var}}(F_k(X)_t, F_k(Y)_t) &\leq K(\mu_1, \mu_2, \sigma_1, \sigma_2) \int_{t_0}^t f(s, t) d_{\text{var}}(X_s|_{E^{(k-1)}}, Y_s|_{E^{(k-1)}}) \\
&\leq K(\mu_1, \mu_2, \sigma_1, \sigma_2) \int_{t_0}^t f(s, t) d_{\text{var}}(X_s, Y_s)
\end{aligned}$$

$\square$

**Remark 6.** In the case of the diagonal process (14),  $\|E[L](t_0, t)\|_{\infty} = \max_{\alpha} e^{-(t-t_0)/\tau_{\alpha}}$ , and  $\|E[L](s, t)\|_{\infty} \|E[L](u, t)\|_{\infty} = \max_{\alpha} e^{(-2t+u+s)/\tau_{\alpha}}$ , values which could be found directly using the one-dimensional equations we obtain in this case.

Fix now a given initial time  $t_0$  and let  $V_0 \in \mathcal{X}$  a given random variable on  $\mathbb{R}^P$ . Let  $X$  a given stochastic process such that  $X(t_0) = V_0$ . We define the sequence of processes  $(X_n) \in \mathcal{M}_1^+(C([t_0, T], \mathbb{R}^P))$  by:

$$\begin{cases} X_0 & = X \\ X_{n+1} & = F_k(X_n) \end{cases} \quad (24)$$

**Theorem 2.7.** *We consider the mean field equation for the GMk. Let  $V_0$  be a random variable of  $\mathbb{R}^{kP}$  and assume that for all stochastic process  $X$  such that  $X_{t_0} = V_0$ , the standard deviation of  $U_{\alpha, \beta}^{X_n}(t)$  is greater than a given value  $\sigma_1 > 0$  for all  $n \in \mathbb{N}^*$ , all  $\alpha, \beta \in \{1, \dots, P\}$  and all  $t \geq t_0$ .*

*Then there exists a unique solution of the MFE with initial condition  $X_{t_0} = V_0$ .*

*Proof.* Note that  $F_k(X)_{t_0} = X_{t_0}$  by definition of  $F_k$ . We use theorem 2.5 and the estimate (23) to prove that under the assumption of the theorem the sequence  $(X_n)_{n \geq 0}$  is a Cauchy sequence.

Under the assumptions of the theorem, we are in the case of application of theorem 2.5. Let us denote by  $K$  the constant  $K(\mu_1, \mu_2, \sigma_1, \sigma_2)$ .

We also define  $\lambda := \sup_{t \in [t_0, T], s \leq t} f(s, t)$ , which is always finite since we assumed  $L$  locally bounded.

We have:

$$\begin{aligned} d_T(X_{n+1}, X_n) &= \sup_{t \in [t_0, T]} d_{\text{var}}(F_k(X_n)_t, F_k(X_{n-1})_t) \\ &\leq K\lambda \sup_{t \in [t_0, T]} \int_{t_0}^t d_{\text{var}}(X_n(s_1), X_{n-1}(s_1)) ds_1 \\ &\leq (K\lambda)^2 \sup_{t \in [t_0, T]} \int_{t_0}^t \int_{t_0}^{s_1} d_{\text{var}}(X_n(s_2), X_{n-1}(s_2)) ds_2 ds_1 \\ &\leq \dots \\ &\leq (K\lambda)^n \sup_{t \in [t_0, T]} \int_{t_0}^t \int_{t_0}^{s_1} \dots \int_{t_0}^{s_{n-1}} d_{\text{var}}(X_1(s_n), X_0(s_n)) ds_1 \dots ds_n \\ &\leq (K\lambda)^n d_T(X_1, X_0) \sup_{t \in [t_0, T]} \int_{t_0}^t \int_{t_0}^{s_1} \dots \int_{t_0}^{s_{n-1}} ds_1 \dots ds_n \\ &\leq \frac{(K\lambda(T - t_0))^n}{n!} d_T(X_1, X_0) \end{aligned} \quad (26)$$

Hence  $(X_n)$  is a Cauchy sequence of  $(\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP})), d_T)$ . Since  $(\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP})), d_T)$  is complete (see proposition 2.2), the sequence  $(X_n)_{n \geq 0}$  converges. The only limit possible of this sequence is a fixed point of  $F_k$ , hence we have existence of fixed points for  $F_k$ , hence of solutions of the mean field equation.

Uniqueness comes from equation (26). There exists  $N > 0$  such that  $\frac{(KT)^N}{N!} < 1$ . Assume that  $X$  and  $Y$  are two fixed points of  $F_k$  for the same initial condition. Since by assumption the constant  $K$  does not depend on the process we chose but only on the initial condition (see the assumption of the theorem), we have:

$$d_T(F_k^N(X), F_k^N(Y)) = d_T(X, Y) \leq \frac{(K(T-t_0))^N}{N!} d_T(X, Y) < d_T(X, Y)$$

which means that  $d_T(X, Y) = 0$  i.e.  $X = Y$ .

Hence we have proved the existence and uniqueness of the solution of the mean field equation (14). □

This theorem is quite general. We will now prove that under some very weak conditions the assumption of theorem 2.7 is fulfilled. For instance, for a Gaussian non-degenerate initial condition, we prove that there always exists a unique solution for the general problem (GM).

**Theorem 2.8.** *Let  $V_0$  be a non-degenerate diagonal Gaussian random process in  $\mathbb{R}^P$  and  $X_0$  a Gaussian r.v. such that  $X_0|_{E^{(0)}} = V_0$ . Then there exists a unique solution of the mean field equation (15).*

*Proof.* The only thing to prove is the assumption of theorem 2.7. In the case where  $V_0$  is a non-degenerate diagonal Gaussian random process, the image of  $F_k$  is contained in the set of Gaussian processes in  $\mathbb{R}^{kP}$ .

We prove here that the standard deviation of any  $F_k(X)$  such that  $X_{t_0}|_{E^{(0)}} = V_0$  has a strictly positive lower bound. This is true since, because of the independence between  $\mathbf{U}$  and  $X$ , the standard deviation is greater than the standard deviation of the first term:  $E[L](t_0, t) \cdot X_{t_0}$ , which is a linear invertible transformation of a non-degenerate Gaussian vector. Hence the resulting vector is also a non-degenerate Gaussian vector, and the standard deviation of each of its component is always strictly positive. Furthermore, since  $t \mapsto E[L](t_0, t)$  is continuous, the standard deviation of any component is lowerbounded by a strictly positive real  $S_1$  on the closed interval  $[t_0, T]$ .

Thus the standard deviation of  $U_{\alpha, \beta}^{F_k(X)}$  is always greater than

$$\sigma_1 := \min_{\alpha, \beta} \sigma_{\alpha, \beta} \inf_{\mu} \sqrt{\mathbb{E}[S_{\alpha, \beta}(G(\mu, S_1))^2]} > 0$$

since we focus on the point of mean where the function  $S_{\alpha, \beta}$  is minimal in absolute value.

The infimum in  $\mu$  is clearly reached as the infimum of a continuous function. It is strictly positive as an expectation of a positive non-trivial function under a probability which weights any non-empty interval.

Now we have a strictly positive lower bound  $\sigma_1$  for  $U_{\alpha, \beta}^X(t)$ . Thus theorem 2.7 applies and we have existence and uniqueness of a solution. □

### 2.3 Discussion

In this section we provide some statements on the results we have just proved.

First of all, note that the proof of theorem 2.7 is constructive, and hence gives us a way for computing the solution of the mean field equation, iterating the map  $F_k$  and starting from any initial process  $X$ , for instance a Gaussian process such as an Ornstein-Uhlenbeck.

Note that the existence and uniqueness is fulfilled whatever the initial time  $t_0$  and the final time  $T$ . Nevertheless, nothing is proved so far for solutions defined for any time. Part of this problem will be treated in section 3.

The only case when the condition of our theorems will not be fulfilled is the case when the whole mass of the initial condition is contained in an interval where the sigmoidal function vanishes (for instance for Heaviside functions).

Note that for this proof, we did not assume regularity on the sigmoidal functions nor anything on the form of the initial probability distribution  $V_0$ .

The case of the backward MFE can be treated the very same way. More precisely, let  $t_0 \in \mathbb{R}$  be the initial time and  $V_0$  a given random variable on  $\mathbb{R}^P$ . Under the conditions of theorem 2.7, there exist a unique solution of the backward mean field equation with the condition  $V_{t_0} = 0$  on any interval  $[S, t_0]$  for  $S < t_0$ .

An important remark is that this result holds for any sigmoidal transform  $S$  whatever its regularity. We only used the fact that it was measurable, increasing and bounded by one, which is always the case for the activity functions  $S$  in neuroscience. This is very interesting. For instance, our result applies for integrate-and-fire neurons for which the  $S_{\alpha,\beta}$ -functions for  $\alpha \neq \beta$  is simply a Heaviside function, and  $S_{\alpha,\alpha}$  accounts for the reset condition and the inter-population interactions (sharply speaking, we add the function  $(V_r - \theta)\text{Heaviside}(x - \theta)$  to the inter-population interaction function).

## 3 Existence and uniqueness of stationary solutions

So far, we have investigated the existence and uniqueness of mean field solutions for a given initial condition. We are now interested in solutions “starting from  $t_0 = -\infty$ ”. This terminology inspired from a physical point of view has a concrete mathematical meaning which we will make explicit. This problematic is quite interesting from a biological viewpoint. Indeed, it corresponds to a sort of weak stationarity, the “oblivion of initial conditions and transient states”.

These solutions are defined for any time and do not depend on their initial condition. They correspond to a generalized “ergodic case” when the system converges to a given measure. Nevertheless, these solutions are not necessarily stationary, and can depend on time, since inputs can themselves depend on time. Among these solutions, we will be particularly interested in the stationary ones. A stationary solution is a solution whose probability distribution does not change under the flow of the equation. These solutions have been already investigated by numerous authors such as [20], but their framework did not allow then to get any rigorous results. In this section we show that under a certain contraction condition there will exist a unique solution of this new equation, and prove

that stationary solutions also exist under the same condition and the assumption that the inputs are stationary.

**Remark 7.** The long-time mean-field description of a network is still a great endeavor in mathematics and statistical physics. In this section we formally take the mean-field equation we obtained and let  $t_0 \rightarrow -\infty$ . This way we obtain an equation which is the limit of the MFE when  $t_0 \rightarrow \infty$ . It means that we consider first the limit  $N \rightarrow \infty$  and then  $t_0 \rightarrow -\infty$ . These two limits do not necessarily commute.

### 3.1 Initial condition at minus infinity

Assume that there exists a stationary distribution for the equation:

$$d\mathbf{X}(t) = L(t)\mathbf{X}(t) dt + \Lambda(t) \cdot d\mathbf{W}_t.$$

This will be the case for instance if  $E[L](t_0, t)$  tends to a constant when  $t_0 \rightarrow -\infty$  and when the covariance of the process:

$$\int_{t_0}^t E[L](s, t)\Lambda(s)dW_s$$

converges to a given matrix function when  $t_0 \rightarrow -\infty$ . This is for instance the case for the (SM) where  $X$  is a  $P$ -dimensional stationary Ornstein-Uhlenbeck process with independent components.

We also assume that for all  $t \in \mathbb{R}$  we have:

$$\int_{-\infty}^t \|E[L](s, t)\|_{\infty} ds = M(t) < \infty$$

where  $\|\cdot\|_{\infty}$  is the operator norm induced by the uniform norm on  $\mathbb{R}^{kP}$ . This is true for instance for the simple model (SM), and in this case where  $M_1(t) = \max_{\alpha} \tau_{\alpha}$ , which is independent of  $t$ .

We call long time mean field equation (LTMFE) the implicit equation:

$$\mathbf{V}(t) = \int_{-\infty}^t E[L](s, t) \left( \tilde{\mathbf{U}}^{\mathbf{V}}(s) + \tilde{\mathbf{I}}(s) \right) ds + \mathbf{X}(t) \quad (27)$$

where  $\mathbf{X}$  is the stationary process we introduced and where  $\tilde{\mathbf{U}}^{\mathbf{V}}(s)$  is the Gaussian process introduced in the study of the GMk, i.e. whose components on  $E^{(k-1)}$  are  $\mathbf{U}^{\mathbf{V}}(t) \cdot \mathbf{1}$  where  $\mathbf{U}^{\mathbf{V}}(t) = (U_{\alpha, \beta}^{\mathbf{V}}(t); \alpha, \beta \in \{1, \dots, P\})_t$  is the diagonal Gaussian process of mean  $\sum_{\beta=1}^P \bar{J}_{\alpha\beta} m_{\beta}(t)$  where  $m_{\beta}(t) = \mathbb{E}[S_{\beta}(V_{\beta}(t))]$  and of covariance:

$$\text{Cov}\left(U_{\alpha, \beta}^{\mathbf{V}}(t), U_{\gamma, \delta}^{\mathbf{V}}(s)\right) =: \mathbb{1}_{\alpha=\gamma; \beta=\delta} \sum_{\beta=1}^P \sigma_{\alpha\beta}^2 \Delta_{\beta}(t, s),$$

where  $\Delta_{\alpha, \beta}(t, s) := \mathbb{E}\left[S_{\alpha, \beta}(V_{\beta}(t))S_{\alpha, \beta}(V_{\beta}(s))\right]$ , and 0 on  $E^{(0)} \oplus \dots \oplus E^{(k-2)}$ .



In this section we consider the set of processes  $\mathcal{M}_1^+(C((-\infty, T], \mathbb{R}^{kP}))$  equipped with the Sup-Var metric defined in equation (16).

As in section 2, we define the long term function  $F_{LT}$ :

$$F_{LT}(\mathbf{Y})_t = \int_{-\infty}^t E[L](s, t) \left( \mathbf{U}^{\mathbf{Y}}(s) \cdot \mathbf{1} + \mathbf{I}(s) \right) ds + \mathbf{X}(t).$$

Let us assume that the input current is bounded.

The mean of the processes in the image of  $F_{LT}$  are bounded. Indeed, we have:

$$\begin{aligned} \|\mathbb{E}[F_{LT}(X)]\|_{\infty} &= \left\| \int_{-\infty}^t E[L](s, t) \left[ \left( \sum_{\beta=1}^P \bar{J}_{\alpha\beta} \mathbb{E}[S_{\alpha,\beta}(X_{\beta}(s))] + I_{\alpha}(s) \right)_{\alpha=1\dots P}, \mathbf{0}_{E^{(0)} \oplus \dots \oplus E^{(k-2)}} \right] ds \right\|_{\infty} \\ &\leq M_1(t) \sup_{\alpha} \left[ \|I_{\alpha}\|_{\infty} + \sum_{\beta=1}^P |\bar{J}_{\alpha\beta}| \right] \end{aligned}$$

We denote by  $\mu(t)$  this bound. We can clearly see that it increases with  $t$ . Assume that for all  $\alpha, \beta \in \{1, \dots, P\}$  we have  $S_{\alpha,\beta}(-\mu(T)) > 0$ . Furthermore,  $F(X)$  is a Gaussian process, hence is symmetrical around its expectation. Therefore for all  $t \in (-\infty, T)$ , the standard deviation of  $U_{\alpha,\beta}^{F(X)}(t)$  is always greater than  $\min_{\alpha,\beta} \frac{\sigma_{\alpha,\beta}}{2} S_{\alpha,\beta}^2(-\mu(T))$  which we denote by  $\sigma_1$ . Hence the estimation (19) will apply in this case with this standard deviation lower bound. Let us denote by  $K$  the constant  $K(\mu_1, \mu_2, \sigma_1, \sigma_2)$  we defined in theorem 2.5, and where  $\mu_1, \mu_2$  and  $\sigma_2$  are the same that had been introduced in section 2.

**Theorem 3.1.** *Assume that  $K(M(T) + M(T)^2) < 1$ . Then there exist a unique solution to the mean field equation (27).*

*Proof.* We prove that the function  $F_{LT}$  is contracting, which in a complete space implies that there exists a unique fixed point which we can construct via the iteration of the function  $F_{LT}$ .

We have for all  $t \in (-\infty, T]$ :

$$\begin{aligned} d_{\text{var}}(F_{LT}(X_t), F_{LT}(Y_t)) &\leq K \int_{-\infty}^t f(s, t) d_{\text{var}}(X_s, Y_s) ds \\ &\leq K d_T(X, Y) \int_{-\infty}^t f(s, t) ds \\ &\leq K d_T(X, Y) \int_{-\infty}^t \left( \|E[L](s, t)\|_{\infty} + \|E[L](s, t)\|_{\infty} \int_{t_0}^t \|E[L](u, t)\|_{\infty} du \right) ds \\ &\leq K(M(T) + M(T)^2) d_T(X, Y) \end{aligned}$$

The righthand term does not depend on  $t$ , hence the inequality is still true for the supremum of the lefthand term on  $t$ . Hence we have:

$$d_T(F_{LT}(X), F_{LT}(Y)) \leq K(M_1(T) + M_2(T)^2)d_T(X, Y) \quad (28)$$

And hence under the assumption of the theorem, the operator  $F_{LT}$  is contracting on the complete space  $\mathcal{M}_1^+(C((-\infty, T], \mathbb{R}))$  together with the distance  $d_T$ , hence admits a unique fixed point. Hence there exists a unique solution of the LTMFE (27).  $\square$

Note that this unique solution is not necessarily stationnary. For instance when the input currents  $I_\alpha$  are not constant, then the solution will not be stationnary.

### 3.2 Stationary solutions

In this section, we are interested in the stationary solutions. We prove some general theorems on the existence and apply it to the (SM) (14). Stationary solutions are processes whose law is invariant under the flow of the equation. Since the long term mean-field solution is necessarily a Gaussian process, one just needs to check that the mean of the process is constant in time and that its covariance matrix of the membrane potential process  $C(s, t)$  only depends on  $(t - s)$ .

First of all, as we already noted, it is clear that if we want the process to be stationary, it is necessary to assume that the currents  $I_\alpha$  do not depend on time. Another important assumption will be that the TOE  $E[L](t, s)$  is a function of  $(t - s)$ . In this case, the function  $M(t)$  is constant. We show that under these two assumptions, there exist stationary solutions of the MFE.

**Lemma 3.2.** Assume that the currents  $I_\alpha$  do not depend on time and that  $E[L](t, s)$  is a function of  $(t - s)$ . Then the set of stationary processes is invariant under the action of  $F_{LT}$ .

*Proof.* Let  $Z$  be a stationary process. We show that  $Y := F_{LT}(Z)$  is also stationary, proving that its mean is constant and its variance is a function of  $(t - s)$ . Let us denote by  $\mu_\alpha^Z(t)$  the mean of the process  $Z_\alpha(t)$  and by  $C_\alpha^Z(t - s)$  its covariance. The mean of the process  $U_{\alpha, \beta}^Z$  reads:

$$m_{\alpha, \beta}^Z(t) = \frac{1}{\sqrt{2\pi C_\alpha^Z(0)}} \int_{\mathbb{R}} S_{\alpha\beta}(x) e^{\frac{(x - \mu_\alpha^Z)^2}{2C_\alpha^Z(0)}} dx$$

and hence does not depends on time. We denote by  $\bar{U}^Z$  the expectation of the vector  $\tilde{U}$ . Similarly, its covariance function is a function of  $(t - s)$  and reads:

$$\int_{\mathbb{R}^2} S_{\alpha\beta}(x) S_{\alpha\beta}(y) \exp\left(-\frac{1}{2} \begin{pmatrix} x - \mu_\alpha^Z \\ y - \mu_\alpha^Z \end{pmatrix}^T \begin{pmatrix} C_\alpha^Z(0) & C_\alpha^Z(t - s) \\ C_\alpha^Z(t - s) & C_\alpha^Z(0) \end{pmatrix}^{-1} \begin{pmatrix} x - \mu_\alpha^Z \\ y - \mu_\alpha^Z \end{pmatrix}\right) dx dy$$

which we denote by  $\Delta_\alpha^Z(t - s)$ . Hence  $U$  is stationary and we denote by  $C_U(t - s)$  its covariance function.

Hence the expectation of  $Y^\alpha(t)$  reads:

$$\begin{aligned}\mu^Y(t) &= \mathbb{E}[F_{\text{LT}}(Z)] \\ &= \mathbb{E}[X_t] + \mathbb{E}\left[\int_{-\infty}^t E[L](s, t) (\mathbf{I} + \tilde{U}(s)) ds\right] \\ &= \mathbb{E}[X_t] + \int_{-\infty}^t E[L](t-s) (\mathbf{I} + \mathbb{E}[\tilde{U}(s)]) ds\end{aligned}$$

We just proved that  $\mathbb{E}[\tilde{U}(s)]$  is independent of  $s$  and hence obtain:

$$\mu^Y(t) = \mathbb{E}[X_t] + \int_{-\infty}^0 E[L](u) du (\mathbf{I} + \bar{U}^Z)$$

which is independent of  $t$  since the process  $X$  is stationary by assumption.

Similarly, we compute the covariance function and check that it can be written as a function of  $(t-s)$ . Indeed, it reads:

$$\begin{aligned}C^Y(t, s) &= \mathbb{E}[(Y_t - \mu^Y(t)) \cdot (Y_s - \mu^Y(s))^T] \\ &= \int_{-\infty}^t \int_{-\infty}^s E[L](t-u) \text{Cov}(U_u^Y, U_v^Y) E[L](s-v)^T du dv + \text{Cov}(X_t, X_s) \\ &= \int_{-\infty}^0 \int_{-\infty}^0 E[L](u) C_U(u-v+t-s) E[L](v)^T du dv + \text{Cov}(X_t, X_s)\end{aligned}$$

which is clearly a function of  $(t-s)$ . Hence  $Y$  is a stationary process, hence the theorem is proved.  $\square$

**Theorem 3.3.** *Assume that  $I_\alpha$  is constant in and that  $K(M(T) + M^2(T)) < 1$ . Then the unique solution of the LTMFE is stationary.*

*Proof.* Under the assumption  $K(M(T) + M^2(T)) < 1$ , theorem 3.1 ensures us that there exists a unique solution. Let now  $Z$  be a given stationary process (for instance a stationary Ornstein-Uhlenbeck process). The sequence of iterates  $(Z_n)_{n \geq 0}$  of  $F_{\text{LT}}$  on  $Z$  is a sequence of stationary processes. As we already proved, this sequence is a Cauchy sequence, hence converges, and the only limit possible is the unique fixed point. On the other hand, since for all  $n$ ,  $Z_n$  is stationary, the limit is necessarily stationary, Hence the unique fixed point is a stationary process.  $\square$

## 4 Numerical Results

The theorem we proved in the previous sections gives us a very natural way to get the solutions of the dynamical mean-field equation. Indeed, we have proved that the iteration of the maps  $F$  of

$F_{LT}$  (under a certain assumption) starting from any initial condition converged to the solution of the MFE. In this section, we will always choose as initial condition a Gaussian process. Hence our theorem 2.8 will apply, since in our simulations we start from a finite time initial condition. Another interest for taking Gaussian initial condition is the compactness of the description. Indeed, we can describe our sequence with only two functions: its mean and its covariance. We obtain our solution constructing the sequence of means and variances attached to the sequence (24) starting from any given process.

In this section we present our simulation strategy and some numerical results compared with the results of Sompolinsky, Crisanti, Sommers [20], and then to a two-populations neuronal network with a negative feedback loop.

#### 4.1 Simulation algorithm

In this section we explain our simulation algorithm in the case of the simple model (SM) (14). As noticed, when considering Gaussian processes, we just need to consider the sequences on means and standard deviations. Hence we write the effect of the iteration of the function  $F$  on the mean and the standard deviation of a Gaussian process. Let  $X$  be a Gaussian process of mean  $\mu^X$  and covariance  $C^X$ ,  $Y = F(X)$  the Gaussian process image of  $X$  by  $F$  and denote  $\mu^Y$  its mean and  $C^Y$  its covariance function. We have:

$$\mu_\alpha^Y(t) = \mu_\alpha^X(0)e^{-t/\tau_\alpha} + \int_0^t e^{-(t-s)/\tau_\alpha} \left( \sum_{\beta=1}^P \sigma_{\alpha\beta} m_{\alpha,\beta}^X(s) + I_\alpha(s) \right) ds$$

where  $m_{\alpha,\beta}^X(s) = \mathbb{E}[S_{\alpha,\beta}(X_\beta(t))]$ .

In our algorithm, the numerical computation of the sequence of means has been done using a standard discretization of the integral, with a time step quite smaller than  $\tau_\alpha$  and than the characteristic time of variation for  $I_\alpha$ .

The covariance function of  $Y$  reads:

$$C_\alpha^Y(t, s) = e^{-(t+s)/\tau_\alpha} \left[ v_\alpha^X(0) + \frac{\tau_\alpha s_\alpha^2}{2} \left( e^{\frac{2s}{\tau_\alpha}} - 1 \right) + \sum_{\beta=0}^P \sigma_{\alpha\beta} \int_0^t \int_0^s e^{(u+v)/\tau_\alpha} \Delta_\beta^X(u, v) dudv \right]$$

where  $v_\alpha^X(t) = Cov(X_\alpha(t), X_\alpha(t))$  and  $\Delta_\beta^X(s, t) = \mathbb{E}[S_{\alpha,\beta}(X_\beta(t))S_{\alpha,\beta}(X_\beta(s))]$ .

Its computation requires a more specific treatment since  $\Delta_\beta^X(u, v)$  depend on all  $C_\beta^X(u, v)$ , with  $u \in [0, t]$ ,  $v \in [0, s]$ . To compute this value efficiently we developed a triangular finite-differences integration scheme.

#### 4.2 Mean field equations in the case Sompolinsky, Crisanti, Sommers

In this section we compare our approach to the works of Sompolinsky and colleagues [20]. First of all, we introduce their approach and state the main results they get with their method. We then show some numerical experiments illustrating the results they obtained in their article.

#### 4.2.1 Theoretical framework.

Sompolinsky et al. [20] have studied the following case of the simple model (SM) with one population ( $P = 1$ ), a centered sigmoidal function  $S(x) = \tanh(gx)$ , centered connections  $\bar{J} = 0$  of standard deviation  $\sigma = 1$  and no input ( $I = 0, s = 0$ ).

They assume that the MFE converges to a stationary solution and they found that the mean of the stationary process is zero and that  $C_\alpha(t, s) \equiv C(t - s) = C(x)$  obey the second order differential equation:

$$\frac{d^2 C}{dx^2} = -\frac{\partial V}{\partial C} \quad (29)$$

which corresponds to the motion of a particle in a potential well. Nevertheless, it is not really an ordinary differential equation, since the initial condition  $C(0)$  depends on the whole trajectory of the solution we are searching for.

Indeed,  $V$  has the following Taylor expansion ( $V$  is odd because  $S$  is even):

$$V(C) = \frac{\lambda}{2}C^2 + \frac{\gamma}{4}C^4 + O(C^6) \quad (30)$$

where  $\lambda = (-1 + g^2 J^2 \langle f' \rangle_{C(0)}^2)$  and  $\gamma = \frac{1}{6} J^2 g^6 \langle f^{(3)} \rangle_{C(0)}^2, \langle \phi \rangle_{C(0)}^2$  being the average value of  $\phi$  under the Gaussian distribution with mean zero and variance  $C(0)$ . Therefore, the determination of  $V$  requires the knowledge of  $C(0)$ , hence requires to have solved the stationary equations!

Nevertheless, they deal with this problem considering that  $C(0)$  acts as a control parameter. Essentially, if  $\lambda > 0$  corresponding to  $g^2 J^2 \langle f' \rangle_{C(0)}^2 < 1$  then the dynamical system (29) has a unique stationary solution corresponding to a minimum of the large deviation functional. This is  $C(t) = 0$ . This has to be interpreted as a stable fixed point dynamics for the neural network. On the other hand if  $g^2 J^2 \langle f' \rangle_{C(0)}^2 \geq 1$  the unique stationary solution corresponding to a minimum of the large deviation functional is the homoclinic trajectory in (29) connecting the point  $C(0) = C^* > 0$  where  $V$  vanishes to the point  $C = 0$ . This solution is interpreted by the authors as a chaotic solution in the neural network.

The equation for the homoclinic solution is easily found using energy conservation and the fact that  $V(C(0)) = 0$  and  $\frac{dV}{dC}(C(0)) = 0$ . One finds:

$$u = \frac{dC}{dx} = -\sqrt{-V(C)}.$$

At the fourth order in the Taylor expansion of  $V$  this gives

$$C(x) = \frac{\sqrt{\frac{-2\lambda}{\gamma}}}{\cosh(\sqrt{-\frac{\lambda}{2}}x)} \quad (31)$$

### 4.2.2 Results.

The model they study is a particular case of our simple model. Indeed, the only thing which differs from our analysis is the fact that his sigmoidal function is not always positive. Nevertheless, it fits in our framework if we change the connectivity function  $S(x) = \tanh(gx)$  by  $\tanh(gx) + 1$ , and the input current (which is 0 in this case) by the random input current  $I_\alpha = -\sum_{\beta=1}^P J_{\alpha\beta}$ . As already mentioned, this case is a straightforward application of our results in section 2. Hence we are ensured to find a solution to our problem. This solution will be approximatively stationary if the initial time  $t_0$  is small enough.

As a benchmark of our numerical procedure we have iterated the evolution function  $\mathcal{F}$  defined in section 2 with these new parameters, for various initial trajectories. We expect to find two regimes. In one case the correlation function is identically zero in the stationary regime, for sufficiently small  $g$  values or for a sufficiently small  $v(t=0)$  (trivial case). The other case corresponds to a regime where  $C(x=0) > 0$  and  $C(x) \rightarrow 0$  has  $x \rightarrow \infty$  ( $x = t - s$ ) (“chaotic” case). This regime requires that  $g$  is sufficiently large and that  $v(t=0)$  is large too.

We have first considered the trivial case corresponding to small  $g$  values. We took  $g = 0.5$  and  $T = 5$ . We have considered as initial conditions an Ornstein-Uhlenbeck process (corresponding to having no interactions in the system), with  $s_\alpha = 0.1$ , with  $\mu_\alpha(0)$  random variable uniformly distributed in  $[-1, 1]$  and  $v_\alpha(0)$  random variable uniformly distributed in  $[0, 1]$ . This is used to initialize the procedure, and in particular to match the condition  $v(t)v(s) - C(t, s)^2 \geq 0$ . Then, during the iterations, we set  $s_\alpha = 0$  in order to match the conditions imposed by the authors. The method converges relatively fast. First, the mean converges to zero, while the variance  $C(t, t)$  decreases (exponentially fast) in time to a constant value corresponding to a stationary regime. This value decreases between consecutive iterations. This is consistent with the theoretical expectation that  $C(t, t) = 0$  in the stationary regime of the trivial case. Finally, the covariance  $C(t - s, s)$  stabilizes to a curve that does not depend on  $s$  and the stationary value (large  $t - s$ ) converges to zero.

We did the same procedure for  $g = 5$  corresponding to the “chaotic” regime. The behavior is the same for  $\mu(t)$  but it is quite different for  $C$ . Indeed, while in the first case the stationary value of  $C(t, t)$  tends to zero with the number of iterations, in the chaotic case it stabilizes to a finite value. In the same way, the covariance  $C(t - s, s)$  stabilizes to a curve that does not depend on  $s$ . The shape of this curve can be extrapolated thanks to Sompolinsky et al. results. Indeed we observe a very good agreement with the theoretical predictions with a fit  $f_4(x) = \frac{a}{\cosh(b(x-\delta))}$ , corresponding to the fourth expansion of  $V$ . Using a 6-th order expansion of  $V(x) = \frac{a}{2}x^2 + \frac{b}{4}x^4 + \frac{c}{6}x^6$  gives a fit  $f_6(x) = \frac{\rho}{\cosh(\lambda(x-\delta))} \frac{1}{\sqrt{1+K^2 - \frac{1}{\cosh(\lambda(x-\delta))}}}$ , where  $\rho, K, \lambda$  are explicit functions of  $a, b, c$ , which provides a slightly better approximation.

## Conclusion

In this paper we present a new mathematical method to show existence and uniqueness of solution for the multi-population mean-field description of a neuronal network with random synaptic weights and

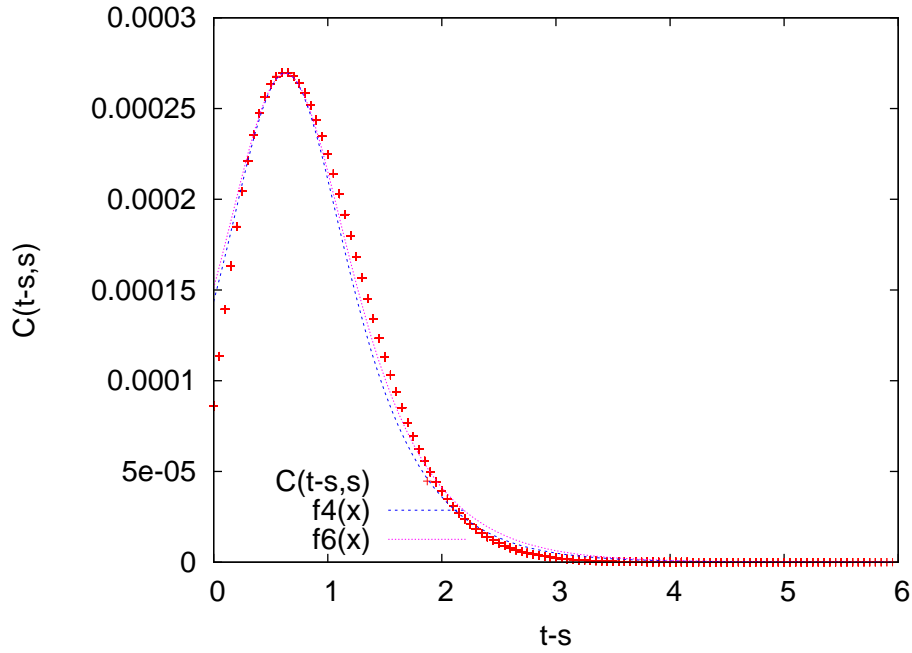


Figure 2: The chaotic case.  $C(t - s, s)$  versus  $t - s$  and fit with  $f_4(x)$  and  $f_6(x)$ .

stochastic inputs. This framework is large enough to include the classical neuronal network models from the simplest analogical model to the spiking neuron models, including multidimensional neuron models and correlated external inputs. We show under very weak conditions the existence and uniqueness of solutions for the mean-field equation, and provide a quite efficient way for computing the dynamical mean-field solution. We also provide conditions for the existence and uniqueness of stationary or pseudo-stationary solutions. We finally show that this approach can be implemented and we obtain numerical results showing a good agreement with some previous studies. The main claim of this article is that a mean-field description of a neuronal network has to be treated as a *global problem* on the probability distribution of the membrane potential, unlike previous studies. This framework enables us to investigate other types of neuronal or cortical area networks and their biological properties. Exploring these different types of behaviors and their differences with neural masses or neural fields equations is a very interesting problem and will be studied in future papers.

## A Time-ordered exponential operator

In this section we explain some properties of the Time-Ordered Exponential (TOE) operator we use all along the paper to integrate our functions. Let  $L(\cdot)$  be a given matrix function. In this appendix, we denote by  $E[L](t_0, t)$  the time-ordered exponential operator  $\mathcal{T} e^{\int_{t_0}^t L(s) ds}$  which we used in the paper and prove some basic and useful properties on this operator. Here we consider that we are in  $\mathbb{R}^P$ . Nevertheless, almost all the properties we prove here are valid in for operators in Banach spaces.

**Definition A.1.** The time-ordered exponential operation is defined as the unique solution of the linear equation:

$$\begin{cases} \frac{dE[L](t_0, t)}{dt} &= L(t)E[L](t_0, t) \\ E[L](t_0, t_0) &= Id \end{cases}$$

where  $Id$  is the identity matrix. It can also be written as the limit of the series:

$$E[L](t_0, t) = \sum_{n=0}^{\infty} \int_{t_0}^t \int_{t_0}^{t_1} \cdots \int_{t_0}^{t_{n-1}} L(t_1)L(t_2)\dots L(t_n) dt_1 \dots dt_n$$

with the convention that the first term of this series is  $Id$  under some loose assumptions on  $L$ . For instance,  $L$  locally bounded is enough to ensure that the series will converge.

In the field of physics, this solution is frequently written  $\mathcal{T} e^{\int_{t_0}^t L(s) ds}$  where  $\mathcal{T}$  is the *time-ordering* operator. The operator  $\mathcal{T}$  is linear and of uniform norm equal to 1. It operates on a product of operators ordering the time. More precisely, its effects reads for any  $n \in \mathbb{N}$ , any  $f_k : \mathbb{R} \mapsto \mathbb{R}$  operators and any times  $t_1, \dots, t_n$ :

$$\mathcal{T}[f_1(t_1)\dots f_n(t_n)] = f_{p_1}(t_{p_1})\dots f_{p_n}(t_{p_n})$$

where  $p : \{1, \dots, n\} \mapsto \{1, \dots, n\}$  is a permutation that orders the time, i.e. such that  $t_{p_1} \leq t_{p_2} \leq \dots \leq t_{p_n}$ .

To give a simple idea of this expansion, we can write formally:



$$\begin{aligned}
E[L](t_0, t) &= Id + \int_{t_0}^t L(t_1)E[L](t_0, t_1) dt_1 \\
&= Id + \int_{t_0}^t L(t_1)dt_1 + \int_{t_0}^t \int_{t_0}^{t_1} L(t_1)L(t_2)E[L](t_0, t_2) dt_2 dt_1 \\
&\vdots \\
&= \sum_{n=0}^N \int_{t_0}^t \int_{t_0}^{t_1} \cdots \int_{t_0}^{t_{n-1}} L(t_1)L(t_2)\dots L(t_n) dt_1 \dots dt_n \\
&\quad + \int_{t_0}^t \int_{t_0}^{t_1} \cdots \int_{t_0}^{t_{N-1}} L(t_1)L(t_2)\dots L(t_N)E[L](t_0, t_N) dt_1 \dots dt_N
\end{aligned}$$

The locally bounded condition ensures us that this series converges.

**Proposition A.1.** The following properties are true for the TOE:

- (i).  $E[L](t_0, t+s) = E[L](t_0, t) \cdot E[L](t, t+s)$
- (ii).  $E[L](t_0, t)$  is invertible of inverse  $E[L](t, t_0)$ .
- (iii). Assume that the operator  $L$  is bounded on  $[t_0, t]$  for the uniform norm on  $\mathbb{R}^P$ , and denote this bound by  $\|L\|$ . Then the operator  $E[L](t_0, t)$  is bounded and its norm uniform operator norm is bounded by  $e^{\|L\|(t-t_0)}$ .

*Proof.* The properties (i) and (ii) are directly linked with the property of group of the flow of a reversible ODE. The last property is a consequence of the series expansion. It can also be found as an application of Gronwald's theorem.  $\square$

**Theorem A.2** (Solution of an inhomogeneous linear SDE). *The solution of the inhomogeneous linear Stochastic Differential Equation:*

$$\begin{cases} dX_t &= (L(t)X(t) + I(t)) dt + \Lambda(s)dW_s \\ X_{t_0} &= V_0 \end{cases}$$

can be written using the TOE:

$$X_t = E[L](t_0, t)V_0 + \int_{t_0}^t E[L](s, t)I(s) ds + \int_{t_0}^t E[L](s, t)\Lambda(s)dW_s$$

*Proof.* We have using the standard theorems of stochastic integration existence and uniqueness of the solution of the given SDE. Hence we check that the solution we wrote is solution of our equation.

It is clear that  $X_{t_0} = V_0$ . We use Itô's formula for the product of two stochastic processes to differentiate the solution we give in the theorem:

$$\begin{aligned}
dX_t &= \left( L(t)E[L](t_0, t)V_0 + E[L](t, t)I(t) + \int_{t_0}^t L(t)E[L](s, t)I(s) ds \right) dt \\
&\quad + E[L](t, t)\Lambda(t)dW_t + \int_{t_0}^t L(t)E[L](s, t)\Lambda(s)dW_s dt \\
&= \left( L(t) \left[ E[L](t_0, t)V_0 + \int_{t_0}^t E[L](s, t)I(s) ds + \int_{t_0}^t E[L](s, t)\Lambda(s)dW_s \right] + I(t) \right) dt \\
&\quad + \Lambda(t)dW_t \\
&= (L(t)X(t) + I(t)) dt + \Lambda(t)dW_t
\end{aligned}$$

Hence the theorem is proved.  $\square$

## B Uniform operator norms

In this section we recall some definitions on the induced norms of matrix. We prove for instance the result we used that  $\|A\|_\infty = \|\|A\|\|_\infty$ .

**Definition B.1.** Let  $A$  be a matrix  $N \times N$  and  $\|\cdot\|$  a norm on  $\mathbb{R}^N$ . The operator norm of  $A$  induced by  $\|\cdot\|$  is defined by:

$$\|A\| = \sup\{\|Ax\| ; x \in \mathbb{R}^N \text{ and } \|x\| = 1\}$$

Consider the uniform norm on  $\mathbb{R}^N$ . It is easy to see that the uniform norm of  $A = (a_{ij})_{i,j}$  is:

$$\|A\|_\infty = \max_i \sum_{j=1}^N |a_{ij}|$$

**Lemma B.1.** The uniform norm of a matrix  $A$  as an operator on the vectors is equal to the uniform norm of  $A$  as an operator on the  $N \times N$  matrix.

*Proof.* Let  $M$  be a matrix such that  $\|M\|_\infty = \max_{ij} M_{ij} = 1$ . We have:

$$\begin{aligned}
\max_{ij} (AM)_{ij} &= \max_{ij} \sum_{k=1}^N A_{ik}M_{kj} \\
&\leq \max_{ij} \sum_k 1^N |A_{ik}| \\
&\leq \max_i \sum_j 1^N |A_{ij}|
\end{aligned}$$

and this upper bound is reached, hence the operator norm  $\|\|A\|\|_\infty$  is equal to the operator norm  $\|A\|_\infty$   $\square$

## References

- [1] S. Amari. Characteristics of random nets of analog neuron-like elements. *IEEE Trans. Syst. Man Cybernet. SMC-2*, pages 643–57, 1972.
- [2] S.I. Amari, K. Yoshida, and K.I. Kanatani. A Mathematical Foundation for Statistical Neurodynamics. *SIAM Journal on Applied Mathematics*, 33(1):95–126, 1977.
- [3] GB Arous and A. Guionnet. Large deviations for Langevin spin glass dynamics. *Probability Theory and Related Fields*, 102(4):455–509, 1995.
- [4] G.B. Arous and A. Guionnet. Symmetric Langevin Spin Glass Dynamics. *The Annals of Probability*, 25(3):1367–1422, 1997.
- [5] Patrick Billingsley. *Convergence of Probability Measures*. Wiley series in probability and statistics, 1999.
- [6] N Brunel and V Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comp.*, 11:1621–1671, 1999.
- [7] B. Cessac. Increase in complexity in random neural networks. *J. de Physique I (France)*, 5:409–432, 1995.
- [8] B. Cessac and M. Samuelides. From neuron to neural networks dynamics. *The European Physical Journal-Special Topics*, 142(1):7–88, 2007.
- [9] A. Crisanti, HJ. Sommers, and H. Sompolinsky. chaos in neural networks : chaotic solutions. 1990.
- [10] A. Crisanti and H. Sompolinsky. Dynamics of spin systems with randomly asymmetric bonds: Langevin dynamics and a spherical model. *Physical Review A*, 36(10):4922–4939, 1987.
- [11] A. Crisanti and H. Sompolinsky. Dynamics of spin systems with randomly asymmetric bounds: Ising spins and glauber dynamics. *Phys. Review A*, 37(12):4865, 1987.
- [12] E. Dauce, O. Moynot, O. Pinaud, and M. Samuelides. Mean-field Theory and Synchronization in Random Recurrent Neural Networks. *Neural Processing Letters*, 14(2):115–126, 2001.
- [13] F. Grimbert and O. Faugeras. Bifurcation analysis of Jansen’s neural mass model. *Neural Computation*, 18(12):3052–3068, December 2006.
- [14] A. Guionnet. Averaged and quenched propagation of chaos for spin glass dynamics. *Probability Theory and Related Fields*, 109(2):183–215, 1997.
- [15] A. Houghton, S. Jain, and A. P. Young. Role of initial conditions in the mean-field theory of spin-glass dynamics. *Phys. Rev. B*, 28(5):2630–2637, Sep 1983.
- [16] Ben H. Jansen and Vincent G. Rit. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics*, 73:357–366, 1995.
- [17] Molgedey L., Schuchardt J., and Schuster H.G. Suppressing chaos in neural networks by noise. *Physical Review Letters*, 69(26):3717–3719, 1992.
- [18] O. Moynot and M. Samuelides. Large deviations and mean-field theory for asymmetric random recurrent neural networks. *Probability Theory and Related Fields*, 123(1):41–75, 2002.
- [19] M. Samuelides and B. Cessac. Random recurrent neural networks. *European Physical Journal - Special Topics*, 142:7–88, 2007.
- [20] H. Sompolinsky, A. Crisanti, and HJ Sommers. Chaos in Random Neural Networks. *Physical Review Letters*, 61(3):259–262, 1988.
- [21] H. Sompolinsky and A. Zippelius. Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses. *Physical Review B*, 25(11):6860–6875, 1982.
- [22] A. van Rotterdam, F.H. Lopes da Silva, J. van den Ende, M.A. Viergever, and A.J. Hermans. A model of the spatial-temporal characteristics of the alpha rhythm. *Bulletin of Mathematical Biology*, 44(2):283–305, 1982.



---

Unité de recherche INRIA Sophia Antipolis  
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes  
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique  
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

---

Éditeur  
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)  
<http://www.inria.fr>  
ISSN 0249-6399