Probab. Theory Relat. Fields 123, 41–75 (2002)
Digital Object Identifier (DOI) 10.1007/s004400100182

Olivier Moynot · Manuel Samuelides

Large deviations and mean-field theory for asymmetric random recurrent neural networks

Received: 10 January 2000 / Revised version: 15 June 2001 /
Published online: 13 May 2002 –

Abstract. In this article, we study the asymptotic dynamics of a noisy discrete time neural network, with random asymmetric couplings and thresholds. More precisely, we focus our interest on the limit behaviour of the network when its size grows to infinity with bounded time. In the case of gaussian connection weights, we use the same techniques as Ben Arous and Guionnet (see [3]) to prove that the image law of the distribution of the neurons' activation states by the empirical measure satisfies a temperature free large deviation principle. Moreover, we prove that if the connection weights satisfy a general condition of domination by gaussian tails, then the distribution of the activation potential of each neuron converges weakly towards an explicit gaussian law, the characteristics of which are contained in the mean-field equations stated by Cessac-Doyon-Quoy-Samuelides (see [4–6]). Furthermore, under this hypothesis, we obtain a law of large numbers and a propagation of chaos result. Finally, we show that many classical distributions on the couplings fulfill our general condition. Thus, this paper provides rigorous mean-field results for a large class of neural networks which is currently investigated in neural network literature.

1. Introduction

The dynamics of large random neural networks and their relations to particle systems as spin glasses has been investigated by numerous biologists, physicists and mathematicians in the recent past. An important scope of this research is to obtain the convergence of the distribution of the neuron activation potentials to a gaussian law when the size of the system grows to infinity. The equations describing this limit law are called the mean-field equations.

Amari can be considered as the initiator of the mean-field theory of random recurrent neural networks. In [1], he stated a kind of central limit property which allows to compute the empirical mean of a function of individual neuron state from the empirical mean and the empirical variance of neuron state. His justification relies on a so-called "Boltzmann property" concerning the asymptotic independance of the individual activation potentials. Then, he managed to derive a mean-field equation in a continuous-time framework. Later, Amari *et al.* [2] focused their

O. Moynot: Laboratoire de Statistiques et de Probabilités, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France

M. Samuelides: Office National d'Etudes et de Recherches Aérospatiales, 2 av. E. Belin, 31055 Toulouse Cedex, France. e-mail: Manuel.Samuelides@wanadoo.fr

Mathematics Subject Classification (2000): Primary 60F10; Secondary 92B20, 82C44

interest on statistical neurodynamics in discrete time neural networks. Their paper contains some interesting neural models, either with several populations or including neurons with a refractory period. It presents useful concepts, such as a propagation of chaos definition, and a precise framework relative to some convergence properties in these networks. Moreover, their work is dedicated to the study of the asymptotic behaviour of the mean activity level for various well normalized connection weights. They obtained many convergence results, but unfortunately, their argument relies heavily on a normalization hypothesis of interaction randomness, namely the connection weights are supposed to have variance of order $\frac{1}{N^2}$, where N is the size of the network. They also noticed the importance of finite time bounds to derive asymptotic results when the size of the networks goes to infinity. Furthermore, Geman (see [12] and [13]) proved a law of large numbers and a central limit theorem, in particular cases, and notably for linear models with asymmetric couplings, which variance is of order $\frac{1}{N}$ and which might be not gaussian. Geman pointed out that the interesting case is the normalization condition of order $1/N$ for variance of the connection weights in a large random neural network but he failed to obtain the proof of mean-field equations in spite of "simulation evidence" for the dynamics of such networks.

Further, Sompolinsky *et al.* [17, 7] used non rigorous statistical physics methods to obtain the mean-field equations and to study the dynamic properties of continous-time networks in case of asymmetric interactions. Cessac *et al.* [6] used the same approach for discrete time models and numerically showed the general occurence of chaos by a quasi-periodicity route in large size networks. It is stated in [5] that this chaotic regime can be described in the thermodynamic limit by the mean-field equations. The mean field equations are considered in these papers and in others as a key result but a clear proof is still missing. Our purpose is to give a rigorous proof for these basic equations. Furthermore, the convergence obtained with the large deviation techniques is stronger than the previously stated law convergence. It allows us to use Borel-Cantelli's lemma to obtain almost sure convergence properties, and to infer our results about some more general large neural networks, the couplings of which are not necessarily gaussian.

The first part of our demonstration uses the method developed by Ben Arous and Guionnet in [3]. In that paper, they showed many results in a continous time spin glass context. Contrary to most of the physicists, and more recently mathematicians [15], who focused their interest on the symmetric models, Ben Arous and Guionnet considered here asymmetric interactions. Their couplings are gaussian and centered. They proved that π^N, the averaged law of the spin's empirical measure on path space of these dynamics satisfies a large deviations principle in the high temperature regime. The study of the rate function, which admits a unique minimum, and the tightness obtained by Guionnet [15] allowed them to compute the weak convergence of the law of every spin towards a measure given by an implicit equation. Thanks to this measure tightness, they did not have any temperature condition for their "propagation of chaos result", which is the mathematically rigorous version for the vanishing correlations of activations states in densely connected recurrent networks. They nevertheless keep this temperature condition to get a quenched law of large numbers for the empirical measure ([3], Theorem 2.8), the proof of which requires a Borel-Cantelli's argument and uses the exponentially fast convergence given by the large deviations principle.

As in [3], our whole proof is built on the effective presence of a gaussian noise, which is essential, although it might be very small. Although these equations might remain true without any noise, our demonstration uses a comparison between the global law of the neurons and the distribution of the noise, which is therefore necessary. In [14], a large deviation principle is proved with a jump noise which leads to Glauber-type dynamics. The consideration of jump noise may be interesting for investigating spiking neuron network dynamics. In this paper we consider analog neurons and a gaussian noise for synaptic summation. The presence of the sigmoid function and of not centered connection weights in our model leads to some technical difficulties, but does not have any influence on the general organization of the demonstration. Moreover, in our discrete time context, we have finite dimensional gaussian properties which lead to the exponential tightness of the equivalent of the family π^N. Therefore, we obtain a large deviation principle and a law of large numbers without any temperature condition. Furthermore, the unique minimum of the rate function is given explicitly, so that we can compute the weak convergence of the law of each activation potential towards a gaussian distribution, the characteristics of which are the mean-field equations.

Our most important result will be to extend these properties for connection weights which are not supposed to be gaussian. Then, although we will not get a large deviations principle, we will deduce from the gaussian case an exponentially fast convergence, and we will therefore be able to infer some almost sure convergence properties. We will also obtain the propagation of chaos and a complete proof of the mean-field equations for these networks. Yet, these results will only be obtained under a domination condition by a gaussian tail for the law of the connection weights. We will show further that (C) is fulfilled if the couplings distribution is a well normalized scaling law with bounded support, and in particular if it is uniform or discrete. Moreover, we will prove that this condition remains stable by mixing, so that it is satisfied if the connection weights are the products of two particular independent random variables, where the first one satisfies the condition, and the second one is a Bernoulli random variable.

Therefore, our model and our results apply to large size diluted neural networks, which are closer to the brain biology than the fully connected ones.

The presentation of the model and the most important results will be given in section 2.

In section 3, we will suppose that the connection weights are gaussian and prove the large deviations principle. Moreover, we will deduce that the minima of the rate function are the fixed points of an operator L. We will then show that L admits a unique fixed point.

In section 4, the connection weights are not supposed to be gaussian. We will compare the law of the activation potentials of the neurons with the equivalent distribution in the gaussian case. We will then use the exponential convergence obtained in section 3 and a Lindeberg's argument to deduce our convergence properties and the mean-field equations.

Lastly, in in the appendix, we will check the equivalence of the domination by gaussian tails with a more technical condition which is used in section 4 and we will show that these conditions are fulfilled for a large class of recurrent random neural network models.

2. The model and the main results

We begin by describing the dynamics of the neural networks model. It is a discrete-time dynamics which relies on the formal neuron modelization of Mac Culloch and Pitts. Let N be an integer greater than 1. Integer j between 1 and N labels the generic neuron of the network. According to the formal neuron model, the state of a single neuron j at time t is described by its activation potential which is a real variable $(u_j)_t$ and its activation state $(x_j)_t$ which is primarily a binary variable. The activation potential $(u_j)_t$ is the algebraic sum of the weighted activation states of input neurons at previous time $t-1$ and of a reference threshold θ_j.

$$(u_j)_t = \sum_i J_{ji}(x_i)_{t-1} + \theta_j$$

The weights J_{ji} of the sum are called the synaptic weights. In this paper, the set of variable (J_{ji}, θ_j) is fixed along time. They are called the quenched variable. The activation state of a Mac-Culloch and Pitts formal neuron is just given by thresholding the activation potential. The neuron is said to be active and it "discharges" when its activation potential is positive:

$$(x_j)_t = 1_{\mathbb{R}^+}[u_{j,t}]$$

We shall use the common averaging approximation of the discharge rate by smoothing the thresholding function and replacing the previous relation by

$$(x_j)_t = f((u_j)_t)$$

where f is a sigmoid function, $(x_j)_t$ may be interpreted as a mean discharge rate between time t and time $t+1$. (For more details on neural networks one may refer to the excellent engineering textbook [16]). Mathematically speaking, f may be any continuous increasing bijection of $\mathbb{R}$ onto $]0, 1[$. For instance, one may choose

$$f(u) = \frac{1 + \tanh(u)}{2}$$

Furthermore, we shall consider a discrete-time synaptic Gauss white noise $[(B_j)_t]$ contributing to the formation of the activation potential. The $(B_j)_t$ are random centered normal i.i.d. variable of standard deviation σ. We suppose that $\sigma > 0$. Our study does not take the case $\sigma = 0$ into account. Therefore the evolution equation from time $t-1$ to time t is

$$(X_j)_t = f\left(\sum_{i=1}^{N} J_{ji}(X_i)_{t-1} + (B_j)_t + \theta_j\right)$$

The activation potential of the neuron j at time t is denoted by $(U_j)_t = f^{-1}((X_j)_t)$. If μ is a probability law on $\mathbb{R}^n$, we shall note μf or by $\widetilde{\mu}$ to shorten notation its image law on $]0, 1[^n$ by the application which maps the vector $(u_j)_j$ into $(f(u_j))_j$. Since we are dealing with asymmetric random recurrent neural networks, we consider that for $1 \leq i, j \leq N$, the weights J_{ji} are i.i.d. random variables. Their common law is noted ν_J^N with expectation $\frac{\overline{J}}{N}$ and variance $\frac{J^2}{N}$. For some technical resaons we suppose in this paper that $J = 0 \Rightarrow \overline{J} = 0$. If $\overline{J} = J = 0$, we will note $\frac{\overline{J}}{J} = 0$. In a neural network context, it's important to consider not centered weights because it allows to anticipate the behavior of large assemblies of excitatory or inhibitory neurons. For some technical reasons, we suppose in this whole article that $J = 0 \Rightarrow \bar{J} = 0$. If $J = \bar{J} = 0$, we will note $\frac{\bar{J}}{J} = 0$. In a neural network context, it's important to consider non centered couplings, because it allows one to anticipate the behavior of large assemblies of excitatory or inhibitory neurons.

Moreover, we suppose that the law ν_J^N of these couplings satisfies the following technical condition, which is denoted by (C) in this whole article :

$$\exists a > 0, \exists D_0 > 0, \forall N \geq 1, \forall J_1, \text{ independent random variable with law } \nu_J^N,$$

$$E\left(exp[aN(J_1)^2]\right) \leq D_0$$

Notice here that this condition is satisfied if ν_J^N is a gaussian distribution, or more generally when its mass taken outside any compact is dominated by a gaussian law.

We prove in the appendix that (C) is equivalent to the apparently stronger condition (C') :

$\exists a > 0, \exists D_0 > 0, \forall N \geq 1, \forall k \leq N, \forall (J_1, .., J_k)$, independent random variables with law ν_J^N, $\forall (\lambda_1, .., \lambda_k) \in [0, 1]^k$,

$$E\left(exp[\frac{aN}{k}(\lambda_1 J_1 + .. + \lambda_k J_k)^2]\right) \leq D_0$$

For $1 \leq j \leq N$, we are considering the θ_j, which are independent, identically distributed random variables, with law $\mathcal{N}(\bar{\theta}, \tau^2)$. These variables represent thresholds in the activation dynamics of the neural network.

The stochastic specification of the network evolution will be completed by the definition of the initial law of the neural network activation state $[(X_j)_0]$. We suppose the random variables $(X_j)_0$ are i.i.d. with law μ_0. We note $(\Omega, \mathcal{A}, \gamma)$ the probability space of interest and we shall study the distribution of the random variables $(X_j)_t$ for $1 \leq j \leq N$ and for $O \leq t \leq T$ which represent the activation states of the neurons between time O and time T and which take their values in [0, 1].

Let $P = \mu_0 \otimes (\mathcal{N}(\bar{\theta}, \sigma^2 I_T) f^{-1})$ be the product measure of μ_0 by $\mathcal{N}(\bar{\theta}, \sigma^2 I_T)$ f^{-1}. Actually, P is the law on $(]0, 1[^{[0,T]})$ obtained by supposing that there is no interaction between the neurons (J_{ij}=0), and that the thresholds are constant and equal to their mean $\bar{\theta}$.

Q^N is the law of $X = ((X_j)_t)_{1 \leq j \leq N, 0 \leq t \leq T}$ under γ, and $\widetilde{Q}^N$ the law of $U = ((U_i)_t)_{1 \leq i \leq N, 0 \leq t \leq T}$.

We consider the empirical measure :

$$\hat{\mu}_N \;:\; ((]0,1[^{[0,T]})^N) \longrightarrow \mathcal{M}_1^+(]0,1[^{[0,T]})$$
$$\hat{\mu}_N(x) \;=\; \frac{1}{N}\sum_{i=1}^{N}\delta_{x_i}$$

where $\mathcal{M}_1^+(]0,1[^{[0,T]})$ represents the metric space made up of all the probabilities on $]0,1[^{\{0,\dots,T\}}$ endowed with the Vaserstein distance [11].

Our purpose is to study the averaged behaviour of this empirical measure under Q^N.

Let π^N be the image law of Q^N by $\hat{\mu}_N$. Thus, we have :

$$\forall B \in \mathcal{B}(\mathcal{M}_1^+(]0,1[^{[0,T]})),\; \pi^N(B) = Q^N(\hat{\mu}_N \in B)$$

We are now able to express the most important results :

Theorem 2.1. *We suppose here that the J_{ij} are gaussian. Then, the family π^N satisfies a strong large deviations principle when N grows to infinity. We note H its rate function.*

For $\mu \in \mathcal{M}_1^+(]0,1[^{[0,T]})$ and $1 \le t, s \le T$, we consider :

$$(K^\mu)_{t,s} = J^2 \int x_{t-1}x_{s-1}d\mu + \tau^2$$
$$c_t^\mu = \bar{J}\int x_{t-1}d\mu$$

Let $(G_t^\mu)_{1\le t\le T}$ be a random gaussian vector with covariance matrix K^μ and mean c^μ.

We prove that a non linear function L from $\mathcal{M}_1^+(]0,1[^{[0,T]})$to itself can be defined by :

$$\frac{dL(\mu)}{dP}(x) = \int exp[\frac{1}{\sigma^2}\sum_{t=1}^{T} G_t^\mu(f^{-1}(x_t) - \bar{\theta}) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(G_t^\mu)^2]d\gamma$$

Then, we have, if the J_{ij} are gaussian :

Theorem 2.2. *H admits a unique minimum Q, which is the unique fixed point of L. Q is given by $Q = L^T(\mu_0^{\otimes(T+1)})$. Moreover, for $0 \le t \le T$, if Q_t is the solution of the problem on [0, t], and $\mathcal{F}_t = \sigma(x_s)_{0\le s\le t}$, then $Q_t = Q_{/\mathcal{F}_t}$*

Notice here that L^T represents L iterated T times.

From now on, we are not supposing that the J_{ij} are gaussian any more.

Then we have the following important result :

Theorem 2.3. *We suppose that assumption (C) is satisfied by the couplings distribution. Let $\beta > 0$. Then :*

$$\limsup \frac{1}{N} ln \left(\pi^N(B(Q,\beta)^c)\right) < 0$$

In particular, π^N converges weakly towards δ_Q.

Note that Q is the limit law previously obtained in the gaussian case. This proves that whatever couplings we consider, π^N converges weakly towards the same limit. This is quite easily understandable, as the main argument of the demonstration is directly related to Lindeberg's theorem and therefore to central limit ideas.

The proof of this last theorem can be deduced from the large deviations principle that we obtained with gaussian couplings. This theorem gives an exponentially fast convergence, which enables us to infer many properties, as in [3].

The first one is a propagation of chaos result, deduced by using the symmetry properties of Q^N as in [19, 3].

Theorem 2.4. *$\forall k \in \mathbb{N}$, for any bounded continuous functions $f_1,...f_k$ from $(]0,1[^{[0,T]})$ to $\mathbb{R}$, we have :*

$$\lim_{N\to+\infty} \int [f_1(x_1)...f_k(x_k)]dQ^N = \prod_{i=1}^{k} \int f_i(x)dQ(x)$$

Moreover, we deduce the following law of large numbers. In our discrete time context, there is no large temperature/short time condition for this property :

Theorem 2.5. *For any $N \in \mathbb{N}$, let $(x_{k,N})_{1\le k\le N}$ be a family of random variables, chosen so that the law of $(x_{1,N},..x_{N,N})$ is Q^N. Let g be a bounded continuous function on $(]0,1[^{[0,T]})$. Then for γ-almost all ω,*

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{i=1}^{N} g(x_{i,N})(\omega) = \int g dQ$$

It is then possible to deduce the mean-field equations for the network dynamics in the asymptotics of large networks. Let $X = (X_t)_{0\le t\le T}$, be a process with law Q. One can deduce from 2.4 that every X_i converges weakly towards X. Let $\widetilde{Q}$ be the image law from Q by f^{-1}. Q is a probability on $(]0,1[^{[0,T]})$, while $\widetilde{Q}$ is a probability on $\mathbb{R}^{[0,T]}$.

Let $U = (U_t)_{0\le t\le T}$ be a process with law $\widetilde{Q}$. Each U_i converges weakly towards U. For $1 \le t \le T$. The mean of U_t in noted $\bar{\mu}(t)$, and its variance $\nu(t)$. For $1 \le t, s \le T$, let $\Delta(t,s)$ be the covariance between U_t and U_s. Moreover, for $0 \le t \le T, m(t)$ is the mean of X_t, and $q(t)$ its order-two moment. We finally note $d\lambda(h) = 1/\sqrt{(2\pi)}exp(-h^2/2)dh$. If $\bar{\mu} = (\bar{\mu}(t))_{1\le t\le T}$, we have :

Theorem 2.6. *1)*

$$Q = \mu_0 \otimes [\mathcal{N}(\bar{\mu}, \Delta) f]$$

2) Moreover :

$$m(0) = \int_0^1 x_0 d\mu_0(x_0)$$

$$q(0) = \int_0^1 (x_0)^2 d\mu_0(x_0)$$

for $1 \leq t \leq T$*,*

$$\bar{\mu}(t) = \bar{\theta} + \bar{J} m(t-1)$$

$$\nu(t) = \tau^2 + J^2 q(t-1) + \sigma^2$$

$$m(t) = \int_{-\infty}^{+\infty} f(\sqrt{\nu(t)} h + \bar{\mu}(t)) d\lambda(h)$$

$$q(t) = \int_{-\infty}^{+\infty} f^2(\sqrt{\nu(t)} h + \bar{\mu}(t)) d\lambda(h)$$

for $2 \leq t, t' \leq T$ *and* $t \neq t'$ *:*

$$\Delta(t, 1) = J^2 m(0) m(t-1) + \tau^2$$

$$\Delta(t, t') = J^2 C(t-1, t'-1) + \tau^2$$

where for $1 \leq t, t' \leq T-1$*,*

$$C(t, t') = \iint f\left(\frac{\sqrt{\nu(t)\nu(t') - \Delta^2(t, t')}}{\sqrt{\nu(t')}} h + \frac{\Delta(t, t')}{\sqrt{\nu(t')}} h' + \bar{\mu}(t)\right)$$

$$\times f(h'\sqrt{\nu(t')} + \bar{\mu}(t')) d\lambda(h) d\lambda(h')$$

One must remember that these mean-field equations, which were stated without noise in [6] and in [5], are used to draw a bifurcation diagram for mean-field long-time asymptotic regime.

3. The large deviations principle for gaussian couplings

We are supposing in this whole section that the connection weights are gaussian. Some results in this section are supported by proofs which are similar to the ones obtained by Ben Arous and Guionnet. We will often omit them. One might find the whole detailed proof in [8].

Definitions. We note:

$$I_T \text{ is the identity matrix of size } T \times T$$

$$A^\mu = K^\mu (\sigma^2 I_T + K^\mu)^{-1}$$

$$V(x) = (f^{-1}(x_t) - \bar{\theta})_{1 \leq t \leq T}$$

$$\Gamma_1(\mu) = -\tfrac{1}{2} ln[det(I_T + \tfrac{1}{\sigma^2} K^\mu)]$$

$$\Gamma_2(\mu) = \frac{1}{2\sigma^2} \int_{(]0,1[^{[\![0,T]\!]})} \left({}^t(V(x) - c^\mu) A^\mu (V(x) - c^\mu) + 2\, {}^tV(x) c^\mu \right) d\mu(x) - \frac{1}{2\sigma^2} ||c^\mu||^2$$

with

$$||c^\mu||^2 = \sum_{t=1}^{T} (c_t^\mu)^2$$

$$\Gamma = \Gamma_1 + \Gamma_2$$

We will note d_T, the Vaserstein distance on $(]0, 1[^{[0,T]})$. It is a distance which is compatible with the weak topology (see [11], theorem 2).

We have :

$$d_T(\mu, \nu) = \inf_{\xi \in E_{\mu,\nu}} \sqrt{\int \sup_{0 \le t \le T} |x_t - y_t|^2 d\xi(x, y)}$$

where the infimum is taken on the set $E_{\mu,\nu}$ of the laws ξ on $(]0, 1[^{[0,T]}) \times (]0, 1[^{[0,T]})$ with marginals μ and ν.

For $1 \le t \le T$, d_t is the Vaserstein distance on $\mathcal{M}_1^+(]0, 1[^{[0,t]})$. To simplify the notations, we will still denote by $d_t(\mu, \nu)$ the Vaserstein distance between the two marginals on $(]0, 1[^{[0,t]})$ of two given measures μ and ν in $\mathcal{M}_1^+(]0, 1[^{[0,T]})$.

3.1. Important preliminaries

We have the following first property :

Proposition 3.1. *We obtain :*

(a) Γ_2 *is well defined from* $\mathcal{M}_1^+(]0, 1[^{[0,T]})$ *into* $\mathbb{R} \cup \{+\infty\}$. *Moreover,* $\forall \mu \in \mathcal{M}_1^+(]0, 1[^{[0,T]})$,

$$\Gamma_2(\mu) \ge -T\bar{J}^2 \left(\frac{1}{2\sigma^2} + \frac{1}{2J^2} \right)$$

(b) $\forall \mu \in \mathcal{M}_1^+(]0, 1[^{[0,T]})$,

$$\Gamma_1(\mu) = ln\{\int exp[-\frac{1}{2\sigma^2} \sum_{t=1}^{T} (G_t^\mu - c_t^\mu)^2] d\gamma\}$$

(c) Γ_1 *is a bounded Lipschitz function.*

(d) We have the following expression :

$$\Gamma(\mu) = \int ln\{\int exp[\frac{1}{\sigma^2} \sum_{t=1}^{T} G_t^\mu (f^{-1}(x_t) - \bar{\theta}) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (G_t^\mu)^2] d\gamma\}\, d\mu(x)$$

The first step is to check (a).

We consider

$$h(x) = {}^t (V(x) - c^\mu) A^\mu (V(x) - c^\mu) + 2\, {}^t V(x) c^\mu$$

We study $h(x)$:

Let's diagonalize K^μ and A^μ in the same orthonormal basis. Let D and Δ be diagonal matrices, and O be orthogonal, so that :

$$K^\mu = O^{-1}\ D\ O$$

$$A^\mu = O^{-1}\ \Delta\ O$$

with

$$\Delta_{tt} = \frac{D_{tt}}{\sigma^2 + D_{tt}}$$

We set $W(x) = O.V(x)$ and we get:

$$h(x) = \sum_{t=1}^{T} \left(\Delta_{tt} W_t(x)^2 + 2(1 - \Delta_{tt}) W_t(x) (Oc^\mu)_t + \Delta_{tt} (Oc^\mu)_t^2 \right)$$

Therefore

$$h(x) = \sum_{\Delta_{tt} \neq 0} \Delta_{tt} \left(\left[W_t(x) + \frac{(Oc^\mu)_t (1 - \Delta_{tt})}{\Delta_{tt}} \right]^2 - \frac{(Oc^\mu)_t^2 (1 - 2\Delta_{tt})}{\Delta_{tt}^2} \right) + 2 \sum_{\Delta_{tt} = 0} W_t(x) (Oc^\mu)_t$$

Remark that

$$D_{tt} = (O K^\mu\, {}^t O)_{tt} \geq J^2 \int \left(\sum_{s=1}^{T} O_{ts} x_{s-1} \right)^2 d\mu$$

Cauchy-Schwartz inequality gives

$$(Oc^\mu)_t^2 \leq \frac{\bar{J}^2}{J^2} D_{tt}$$

Then, if $\Delta_{tt} = 0$, we obtain $D_{tt} = 0$ and therefore $(Oc^\mu)_t = 0$

Moreover, if $\Delta_{tt} \neq 0$,

$$\frac{(Oc^\mu)_t^2}{\Delta_{tt}} \leq \frac{\bar{J}^2 \sigma^2}{J^2 (1 - \Delta_{tt})}$$

We can thus infer the following inequality :

$$h(x) \geq -\left(\frac{T \bar{J}^2 \sigma^2}{J^2} \right)$$

This achieves the proof of property (a). Notice that the hypothesis ($J = 0 \Rightarrow \bar{J} = 0$) was necessary to obtain this property.

The following gaussian lemma is useful in this whole paper, and is of great help to prove the other assertions of 3.1 :

Lemma 3.2. *Let X be a gaussian vector taking its values into $\mathbb{R}^T$, with mean c and covariance K. Let $a \in \mathbb{R}^T$, and $b \in \mathbb{R}$. We suppose that all the eigenvalues α of K satisfy $\alpha b > -1$. We note*

$$A = K(I_T + bK)^{-1}$$

Then :

$$E[exp({}^t aX - \frac{b}{2}||X||^2)] = \frac{1}{\sqrt{det(I_T + bK)}}.exp\left({}^t ac - \frac{b}{2}||c||^2 + \frac{1}{2}\,{}^t(a - bc)A(a - bc)\right)$$

Thus, from this lemma, we can deduce assertions (b) and (d) of 3.1 through straightforward computations.

The demonstration of (c) is nearly the same as in lemma 3.3 of [3]. □

One consequence of this proposition is that Γ is well defined, taking its values into $\mathbb{R} \cup \{+\infty\}$. We are now able to define the function H, which will be the rate function of the large deviations principle :

For any μ in $\mathcal{M}_1^+(]0, 1[^{[0,T]})$, we note :

$$H(\mu) = I(\mu, P) - \Gamma(\mu)$$

where $I(\mu, P)$ represents the relative entropy with respect to P, i.e :

$$I(\mu, P) = \int ln(\frac{d\mu}{dP})\frac{d\mu}{dP}\,dP \quad if\,\mu \ll P$$

Otherwise, $I(\mu, P) = +\infty$.

We have the three following properties, the demonstrations of which are similar to the ones obtained in [3] (see lemma 3.3 and theorem 3.1) :

Proposition 3.3. *(a)* $\forall\mu \in \mathcal{M}_1^+(]0, 1[^{[0,T]})$, $\Gamma(\mu) \leq I(\mu, P)$*, i.e* $H \geq 0$.

(b) There exists real constants $e > 1$ and $f > 0$, such that

$$\forall\mu \in \mathcal{M}_1^+(]0, 1[^{[0,T]}), \Gamma(\mu) \leq \frac{(I(\mu, P) + f)}{e}$$

(c) H is lower semi-continuous.

The next lemma represents the center of the proof.

Lemma 3.4.
$$\frac{dQ^N}{dP^{\otimes N}} = exp\left(N\Gamma(\hat{\mu}_N(x))\right)$$

Demonstration. Remember that we denote by $\widetilde{P}^N(J,\theta)$ the conditional law of the $((U_i)_t)_{1\leq i\leq N, 0\leq t\leq T}$ for fixed $(J_{ij},\theta_j)_{1\leq i,j\leq N}$. The image law of P by f^{-1} is denoted by $\widetilde{P}$.

Note that:
$$\widetilde{Q}^N = \int_{\mathbb{R}^{N^2+N}} \widetilde{P}^N(J,\theta)d\nu(J,\theta)$$
where ν is the gaussian distribution of $(J_{ij},\theta_j)_{1\leq i,j\leq N}$.

For any $u \in (\mathbb{R}^{[0,T]})^N$, we are interested in
$$\frac{d\widetilde{P}^N(J,\theta)}{d\widetilde{P}^{\otimes N}}(u)$$

For $1\leq t\leq T$ and $1\leq j\leq N$, we consider :
$$(Y_j)_t = (B_j)_t + \bar{\theta}$$

To simplify the notations, we will denote the vector $((u_1)_0, ..., (u_N)_0)$ by $(u)_0$.
For any fixed $(J, \Theta, (u)_0)$ we consider $\Psi_{(J,\Theta,(u)_0)}$, defined from $\mathbb{R}^{NT}$ to $\mathbb{R}^{NT}$ by :
$$\Psi_{(J,\Theta,(u)_0)}((u_1)_1, ...(u_N)_T) = ((y_1)_1, ...(y_N)_T)$$
so that
$$(y_j)_t = (u_j)_t - \sum_{i=1}^N J_{ji} f((u_i)_{t-1}) - \theta_j - \bar{\theta}$$

Let g be a continuous bounded function.
We note :
$$S = E\left(g((U_1)_0, ..., (U_N)_T)/(J,\Theta)\right)$$
We have :
$$S = E\left(g\left((U_1)_0, ..., (U_N)_0, \Psi^{-1}_{(J,\Theta,(U)_0)}((Y_1)_1, ...(Y_N)_T)\right)/(J,\Theta)\right)$$
As $((Y_j)_t)_{1\leq j\leq N, 1\leq t\leq T}$ and $(U)_0$ are independent from (J,Θ), we obtain :
$$S = \int g\left((u)_0, \Psi^{-1}_{(J,\Theta,(u)_0)}((y_1)_1, .., (y_N)_T)\right)\prod_{i=1}^N\prod_{t=1}^T exp - \frac{1}{2\sigma^2}((y_i)_t - \bar{\theta})^2$$
$$\times \frac{d(y_1)_1..d(y_N)_T d\mu_0^{\otimes N}}{(\sigma\sqrt{2\pi})^{NT}}$$
Using the change of variables
$((y_1)_1, ...(y_N)_T) \to ((u_1)_1, ..(u_N)_T) = \Psi^{-1}_{J,\Theta,(u)_0}((y_1)_1, .., ((y_N)_T)$,

we get:

$$A = \frac{1}{(\sigma\sqrt{2\pi})^{NT}} \int_{\mathbb{R}^N} d\mu_0^{\otimes N}((u)_0) \left(\int_{\mathbb{R}^{NT}} g((u_1)_0, .., (u_N)_T) exp \right.$$
$$\left. \times \; (-\frac{1}{2}\Phi((u_1)_1, .., (u_N)_T)) du_1^1..du_T^N \right)$$

with

$$\Phi((u_1)_1, ..(u_N)_T) = \frac{1}{\sigma^2} \sum_{t=1}^{T} \sum_{j=1}^{N} \left((u_j)_t - \sum_{i=1}^{N} J_{ji} f((u_i)_{t-1}) - \theta_j - \bar{\theta} \right)^2$$

For any fixed u and j, we now consider :

$$(G_j)_t(f(u)) = \sum_{i=1}^{N} J_{ji} f((u_i)_{t-1}) + \theta_j$$

Therefore, we have :

$$\frac{d\widetilde{P}^N(J, \Theta)}{d\widetilde{P}^{\otimes N}}(u) = h(u)$$

where

$$h(u) = h((u_1)_0, ..(u_N)_T) = exp \frac{1}{\sigma^2} \left[\sum_{t=1}^{T} \sum_{j=1}^{N} \left(((u_j)_t - \bar{\theta})(G_j)_t - \frac{1}{2}((G_j)_t)^2 \right) \right]$$

It follows that :

$$\frac{d\widetilde{Q}^N}{d\widetilde{P}^{\otimes N}}(u) = E\left(h(J, \Theta, u)\right) = \int h(J, \Theta, u) d\nu(J, \Theta)$$

For any fixed j and u, one should notice that the distributions of the two gaussian vectors $G_j(u)$ and $G^{\hat{\mu}_N(f(u))}$ are the same (their means and covariances are equal).

Using the independence of the $(G_j)'s$, we deduce :

$$\frac{d\widetilde{Q}^N}{d\widetilde{P}^{\otimes N}}(u) = \prod_{j=1}^{N} E\left[exp \frac{1}{\sigma^2} \left(\sum_{t=1}^{T} G_t^{\hat{\mu}_N(f(u))}((u_j)_t - \bar{\theta}) - \frac{1}{2}(G_t^{\hat{\mu}_N(f(u))})^2 \right) \right]$$

Therefore :

$$\frac{d\widetilde{Q}^N}{d\widetilde{P}^{\otimes N}}(u) = exp\left(N\Gamma(\hat{\mu}_N(f(u)))\right)$$

Finally, we get

$$\frac{dQ^N}{dP^{\otimes N}} = exp\left(N\Gamma(\hat{\mu}_N(x))\right). \qquad \square$$

3.2. Proof of the large-deviations principle

We now need to prove that Γ is lower semi-continuous. Thanks to 3.1, we already know that Γ_1 is continuous, so that we only have to show that Γ_2 is lower semi-continuous which is given by the application of [3].

More precisely, for any $M > 0$, we consider

$$\Gamma_2^M(\mu) = \frac{1}{2\sigma^2}\int 1_{(||V(x)||\leq M)}\Big({}^t(V(x)-c^\mu)A^\mu(V(x)-c^\mu)+2\,{}^tV(x)c^\mu - ||c^\mu||^2 + d\Big)\,d\mu(x) - d$$

where

$$d = T\bar{J}^2\left(\frac{1}{2\sigma^2}+\frac{1}{2J^2}\right)$$

so that the function defined under the integral is always positive (see proposition 3.1(a)).

First, for any $\mu \in \mathcal{M}_1^+(]0,1[^{[0,T]})$, we have

$$|(K^\mu)_{ts} - (K^\nu)_{ts}| \leq 2J^2 d_{T-1}(\mu,\nu)$$

This implies that $\mu \to K^\mu$ and therefore $\mu \to A^\mu$ are Lipschitz functions. Moreover, $\mu \to c^\mu$ is Lipschitz too.

We now deduce that Γ_2^M is continuous :

Let $\mu \in \mathcal{M}_1^+(]0,1[^{[0,T]})$ be a given probability and $\mu_n \in \mathcal{M}_1^+(]0,1[^{[0,T]})$ be supposed to converge weakly towards μ.

We consider :

$$\phi(\mu,x) = \frac{1}{2\sigma^2}1_{(||V(x)||\leq M)}\Big({}^t(V(x)-c^\mu)A^\mu(V(x)-c^\mu)+2\,{}^tV(x)c^\mu - ||c^\mu||^2 + d\Big)$$

Thus, we have :

$$|\Gamma_2^M(\mu)-\Gamma_2^M(\mu_n)| \leq \left|\int\phi(\mu,x)d\mu - \int\phi(\mu,x)d\mu_n\right| + \left|\int\phi(\mu,x)d\mu_n - \int\phi(\mu_n,x)d\mu_n\right|$$

The weak convergence of μ_n towards μ implies that the first term converges to 0 when n grows to infinity. As $\mu \to A^\mu$ and $\mu \to c^\mu$ are lipschitz, we can compute that there exists a real constant F, which does not depend on x and n, such that :

$$|\phi(\mu,x)-\phi(\mu_n,x)| \leq F d_T(\mu_n,\mu)$$

We are now able to conclude that Γ_2^M is continuous.

Therefore, as for any $\mu \in \mathcal{M}_1^+(]0,1[^{[0,T]})$, $\Gamma_2^M(\mu)$ grows to $\Gamma_2(\mu)$ when M grows to infinity, we obtain that Γ_2 is lower semi-continuous. □

Hence :

Lemma 3.5. Γ *is lower semi-continuous*

It is from this property that we are going to infer the first part of the large deviations principle :

Proposition 3.6. *If* O *is an open set in* $\mathcal{M}_1^+(]0,1[^{[0,T]})$*, then :*

$$-\inf_O H \leq \liminf \frac{1}{N} ln\pi^N(O)$$

Proof. Let R^N be the image law of $P^{\otimes N}$ by $\hat{\mu}_N$. Sanov's theorem allows one to assert that the family R^N satisfies a large deviation principle with good rate function $\mu \to I(\mu, P)$. Because of lemma 3.4, we have :

$$\pi^N(O) = \int_O exp\,(N\Gamma(\mu))\,dR^N(\mu)$$

As Γ is lower semi-continuous, we are able to conclude the demonstration by using a Varadhan's argument (see [10], thm 2.1.7). □

Remark that the large deviations principle we are proving is derived from the one satisfied by the noise, the importance of which is thus underlined.

We now use the explicit expression of Γ to obtain the exponential tightness without any temperature condition. We first give a gaussian lemma which will be very useful :

Lemma 3.7. *Let* X *be a gaussian vector, with covariance* K *and mean* c*, taking its values in* $\mathbb{R}^T$*. Let otherwise* $a \in \mathbb{R}^T$*, and* $b \in \mathbb{R}^+$*. The eigenvalues of the matrix* K *are denoted by* $(\lambda_1, ..\lambda_T)$ *We consider* $\lambda = \sup|\lambda_i|$*.Then :*

$$E\left(exp({}^t aX - \frac{b}{2}||X||^2)\right) \leq exp({}^t ac - \frac{b}{2}||c||^2)exp\left(\frac{\lambda||a-bc||^2}{2(1+\lambda b)}\right)$$

Therefore, we are able to deduce that :

Lemma 3.8.

$$\exists\alpha > 1, \sup_N \left(\int exp(\alpha N\Gamma(\hat{\mu}_N))dP^{\otimes N}\right)^{\frac{1}{N}} < \infty$$

For $1 \leq j \leq N$, we note :

$$b_j(x) = \int exp\frac{1}{\sigma^2}\sum_{t=1}^{T}\left(G_t^{\hat{\mu}_N(x)}V_t(x_j) - \frac{1}{2}(G_t^{\hat{\mu}_N(x)})^2\right)d\gamma$$

The largest eigenvalue of matrix $K^{\hat{\mu}_N(x)}$ is smaller than $T(J^2+\tau^2)$.
We consider $\alpha > 1$, such that

$$\beta = \frac{\alpha T(J^2+\tau^2)}{\sigma^2 + T(J^2+\tau^2)} < 1$$

As a direct consequence of lemma 3.7, we get:

$$b_j(x) \leq exp\left[\frac{1}{2\alpha\sigma^2}\sum_{t=1}^{T}\Big(\beta(V_t(x_j) - c_t^{\hat{\mu}_N(x)})^2 + 2\alpha V_t(x_j)c_t^{\hat{\mu}_N(x)} - \alpha(c_t^{\hat{\mu}_N(x)})^2\Big)\right]$$

As $|c_t^{\hat{\mu}_N(x)}| \leq |\bar{J}|$, we deduce :

$$(b_j(x))^\alpha \leq exp\frac{1}{2\sigma^2}\sum_{t=1}^{T}\Big(\beta(V_t(x_j))^2 + 2(\alpha-\beta)|\bar{J}V_t(x_j)|\Big)$$

Therefore, we can find a real constant C such that :

$$\int exp(\alpha N\Gamma(\hat{\mu}_N(x)))dP^{\otimes N}(x) \leq \prod_{j=1}^{N}\prod_{t=1}^{T}\int exp\frac{1}{2\sigma^2}\Big((\beta-1)((u_j)_t)^2+C|(u_j)_t|\Big) \times\frac{d(u_j)_t}{\sigma\sqrt{2\pi}}$$

This achieves the proof. □

We obtain :

Proposition 3.9. *The family π^N is exponentially tight.*

We use the same ideas as in [3] (see page 471). We recall them here. Let $\alpha > 1$ be the real constant obtained in the previous lemma.

We consider $\delta = 1-\frac{1}{\alpha}$. Hölder's inequality gives : $\forall B \in \mathcal{B}(\mathcal{M}_1^+(]0,1[^{[0,T]}))$,

$$\pi^N(B) \leq R^N(B)^\delta.exp(\frac{CN}{\alpha})$$

We know by Sanovs' exponential tightness property that, for any $L > 0$, there is a compact set K_L such that :

$$\limsup_{n\to+\infty}\frac{1}{N}ln(R^N(K_L^c)) \leq -L$$

Therefore, for any $L > 0$, we just have to take $K_{\frac{L+C/\alpha}{\delta}}$ to obtain the tightness. □

Proposition 3.10. *For any compact set K of $\mathcal{M}_1^+(]0,1[^{[0,T]})$:*

$$\limsup\frac{1}{N}ln(\pi^N(K)) \leq -inf H$$

See Lemma 3.8 of [3] to get the ideas of the proof, or [8] to obtain all the details.

Gathering 3.8, 3.9 and 3.10, and remarking that H is a good rate function because of 3.3 (b), we can achieve the proof of Theorem 2.1. □

3.3. The minima of the rate function

We recall here that for any $\mu \in \mathcal{M}_1^+(]0,1[^{[0,T]})$, $L(\mu)$ is the probability defined on $\mathcal{M}_1^+(]0,1[^{[0,T]})$ by :

$$\frac{dL(\mu)}{dP}(x) = \int exp[\frac{1}{\sigma^2}\sum_{t=1}^{T} G_t^\mu(f^{-1}(x_t)-\bar{\theta}) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(G_t^\mu)^2]d\gamma$$

Note that L is well defined because Fubini's theorem and first integration with respect to P imply :

$$\int \frac{dL(\mu)}{dP}(x)dP(x) = \int exp\left(\frac{1}{2\sigma^2}\sum_{t=1}^{T}(G_t^\mu)^2\right).exp\left(\frac{-1}{2\sigma^2}\sum_{t=1}^{T}(G_t^\mu)^2\right)d\gamma = 1$$

Then:

Proposition 3.11.

$$H(Q) = 0 \Longleftrightarrow L(Q) = Q$$

As in theorem 5.1 of [3], the proof is based on the study of a variational equation around a given minimum of the function H.

We thus have to characterize the fixed points of L :

Proposition 3.12. *L admits a unique fixed point given explicitly by $Q = L^T(\mu_0^{\otimes(T+1)})$.*

In our discrete time context, the proof is much simpler than in [3].

It is essential to remark that the mean and covariance of G_t^μ only depend on the restriction of μ to $\mathcal{F}_{t-1}$. Therefore, let μ and ν be two probabilities in $\mathcal{M}_1^+(]0,1[^{[0,T]})$. Then, for $0 \leq t \leq T-1$:

$$\mu_{/\mathcal{F}_t} = \nu_{/\mathcal{F}_t} \Rightarrow L(\mu)_{/\mathcal{F}_{t+1}} = L(\nu)_{/\mathcal{F}_{t+1}}$$

It follows that if we consider the sequence of probability measures defined on $\mathcal{M}_1^+(]0,1[^{[0,T]})$ by $V_0 = \mu_0^{\otimes(T+1)}$, and $V_{t+1} = L(V_t)$, we are able to infer a step to step convergence of V_t towards its limit Q. More precisely, we observe that

$$(V_0)_{/\mathcal{F}_0} = L(V_0)_{/\mathcal{F}_0} = \mu_0$$

We deduce that $\forall t \geq T$, $V_t = V_T = Q$, and that $\forall t \leq T$, $(V_T)_{/\mathcal{F}_t} = (V_t)_{/\mathcal{F}_t}$.

This gives the existence of a fixed point for L. The uniqueness of this fixed point is based on the following argument : if μ and ν are two fixed points of L, we have

$$L(\mu)_{/\mathcal{F}_0} = L(\nu)_{/\mathcal{F}_0} = \mu_0$$

Therefore, $L^{T+1}(\mu) = L^{T+1}(\nu)$, and then $\mu = \nu$.

If we note L_t the operator associated to the same problem on $[0,t]$, with $t \leq T$, we have $L_t(\mu_{/\mathcal{F}_t}) = L(\mu)_{/\mathcal{F}_t}$ $\forall \mu \in \mathcal{M}_1^+(]0,1[^{[0,T]})$. The proof of theorem 2.2 is thus achieved. □

4. Convergence properties for general couplings

In this whole section, we will consider more general connection weights satisfying the condition (C) of domination by a gaussian tail. It must be remembered that Q^N is the global law of the activation potentials of the neurons. We note Q_0^N the distribution associated to the gaussian couplings.

Our purpose is to compare the respective densities of Q^N and Q_0^N relatively to $P^{\otimes N}$. We prove that these two densities are close enough for deducing the result of theorem 2.3 from the large deviation principle obtained for gaussian connection weights.

More precisely, let $\nu(J, \Theta)$ be the distribution of the couplings and the thresholds.

The calculations made in the proof of lemma 3.4 give that :

$$\frac{dQ^N}{dP^{\otimes N}} = \prod_{j=1}^{N} a_j(x)$$

with

$$a_j(x) = exp\left(\frac{1}{2\sigma^2}\sum_{t=1}^{T} V_t(x_j)^2\right)\int exp\left(\frac{-1}{2\sigma^2}\sum_{t=1}^{T}[(G_j)_t(x) - V_t(x_j)]^2\right)d\nu(J, \Theta)$$

where

$$(G_j)_t(x) = \sum_{i=1}^{N} J_{ji}(x_i)_{t-1} + \theta_j$$

If the couplings are gaussian, one can use lemma 3.2 to calculate this expression. Thus, we have :

$$\frac{dQ_0^N}{dP^{\otimes N}} = \prod_{j=1}^{N} b_j(x)$$

$$b_j(x) = \frac{1}{\sqrt{det(I + \frac{K^{\hat{\mu}_N(x)}}{\sigma^2})}} exp\left(\frac{1}{2\sigma^2}\phi(x)\right)$$

with

$$\phi(x) =^t (V(x_j) - c^{\hat{\mu}_N(x)})(A^{\hat{\mu}_N(x)} - I_T)(V(x_j) - c^{\hat{\mu}_N(x)}) + {}^tV(x_j)V(x_j)$$

Remark that a_j and b_j depend on N, but we will neglect it in the notations in order to be simpler.

This section is divided in two subsections. The first one is dedicated to prove theorem 4.1, which consists in proving conclusion of theorem 2.3 if four hypothesis are checked. In the second part, we show that if the couplings fulfill assumption (C) of section 2, then the four conditions of the theorem are satisfied. Thus, theorem 2.3 is completely proved.

4.1. An exponential convergence result

In subsection 4.2, we will check that theorem 4.1 applies to probabilities Q^N and Q_0^N defined above. As this theorem remains true under more general hypothesis, we prove it here for a larger class of probabilities. We voluntarily choose to keep the same notations for Q^N and Q_0^N to simplify the reader's understanding.

Thus, in this subsection, let Q^N satisfy :

$$\frac{dQ^N}{dP^{\otimes N}} = \prod_{j=1}^{N} a_j(x)$$

where P is the law defined in section 2, and where the notations omit the dependance of a_j on N. Moreover, let Q_0^N be another family of probabilities on $((]0,1[^{[0,T]})^N)$, the image of which by the empirical measure satisfies a large deviation principle when N grows to infinity. We suppose that the associated rate function H is good and admits a unique minimum Q such that $H(Q) = 0$. We have

$$\frac{dQ_0^N}{dP^{\otimes N}} = \prod_{j=1}^{N} b_j(x)$$

π^N is the image law of Q^N by the empirical measure.

We suppose that (a_j) and (b_j) satisfy the four following assumptions :

(H1) $\exists A, B > 0, \forall N \geq 1, \forall j \in \{1,..,N\}, \forall x \in ((]0,1[^{[0,T]})^N)$,

$$a_j(x) \geq A.exp(-B \sum_{t=1}^{T} |V_t(x_j)|)$$

with

$$V_t(x_j) = f^{-1}((x_j)_t) - \bar{\theta}$$

(H2) $\exists \lambda < 1, \exists C > 0, \forall N \geq 1, \forall j \in \{1,..,N\}, \forall x \in ((]0,1[^{[0,T]})^N)$,

$$a_j(x) \leq Cexp\left(\frac{1}{2\sigma^2} \sum_{t=1}^{T} [\lambda (V_t(x_j))^2 + C|V_t(x_j)|]\right)$$

(H3) $\forall \eta > 0, \exists \alpha > 0, \forall N \geq 1, \forall k \leq N$, if $\frac{k}{N} \leq \alpha$ then

$\forall s$, injection from $\{1,..,k\}$ into $\{1,..,N\}, \forall j \notin \{s(1),..,s(k)\}, \exists \tilde{a}_j(x)$, which only depends on $(x_i)_{i \neq s(1),..,s(k)}, \forall x \in ((]0,1[^{[0,T]})^N)$:

$$\sup\left(\frac{a_j(x)}{\tilde{a}_j(x)}, \frac{\tilde{a}_j(x)}{a_j(x)}\right) \leq (1+\eta) exp \frac{\eta}{2\sigma^2} \left(\sum_{t=1}^{T} (V_t(x_j))^2\right)$$

(H4) $\exists D > 0, \forall \eta > 0, \forall \beta > 0, \exists N_0, \forall N \geq N_0, \forall j \in \{1,..,N\}, \forall x \in ((]0,1[^{[0,T]})^N)$

$$\frac{a_j(x)}{b_j(x)} \leq (1+\eta) exp(\frac{\eta}{2\sigma^2} \left(\sum_{t=1}^{T} (V_t(x_j))^2\right) + \beta exp\left(\frac{1}{2\sigma^2} \sum_{t=1}^{T} [(V_t(x_j))^2 + D|V_t(x_j)|]\right)$$

The first and second assumptions are useful to control the a_j's. The third hypothesis means that a_j does not much depend on a small set of x_i's, such that $i \neq j$. The fourth condition means that the quotient between a_j and b_j is close to one, it is a contiguity relation between probability measures Q^N and Q_0^N. From these four hypothesis, we can't deduce that π^N and π_0^N are exponentially equivalent (what would imply that π_0^N satisfies a large deviations principle, see [9]). But we are able to derive the following weaker property :

Theorem 4.1. *We suppose that the four conditions (H1), (H2), (H3) and (H4) given above are satisfied. Then :* $\forall \delta > 0, \exists b > 0, \exists N_0, \forall N \geq N_0,$

$$\pi^N(B(Q,\delta)^c) \leq exp(-bN)$$

Proof. We have

$$\pi^N(B(Q,\delta)^c) = \int \mathbf{1}_{\hat{\mu}_N(x) \in (B(Q,\delta)^c)} \frac{dQ^N}{dQ_0^N} dQ_0^N$$

Let's fix $q \in \mathbb{R}$, such that $1 < q < \frac{3}{2}$ and $\lambda + 2(q-1) < 1$, where λ is defined in assumption (H2).

Hölder's inequality gives, for $\frac{1}{p} + \frac{1}{q} = 1$:

$$\pi^N(B(Q,\delta)^c) \leq \left[\int (\frac{dQ^N}{dQ_0^N})^q dQ_0^N\right]^{\frac{1}{q}} .[Q_0^N(\hat{\mu}_N(x) \in (B(Q,\delta)^c)]^{\frac{1}{p}}$$

As the rate function H of the large deviation principle is good, we know that $\exists b' > 0, \exists N_1, \forall N \geq N_1,$

$$Q_0^N(\hat{\mu}_N(x) \in (B(Q,\delta)^c) \leq exp(-Nb')$$

Thus, for $N \geq N_1$,

$$\pi^N(B(Q,\delta)^c) \leq exp(-\frac{Nb'}{p}).\left[\int (\frac{dQ^N}{dQ_0^N})^{q-1} dQ^N\right]^{\frac{1}{q}} \tag{1}$$

Our purpose is now to control Z_N, defined by

$$Z_N = \int (\frac{dQ^N}{dQ_0^N})^{q-1} dQ^N$$

Let $\eta > 0, \beta > 0$. Because of (H4), we know that for N large enough,

$$Z_N \leq \int \prod_{j=1}^{N} \left((1+\eta)exp[\frac{\eta}{2\sigma^2} \sum_{t=1}^{T} V_t(x_j)^2] + \beta exp(\frac{1}{2\sigma^2} \sum_{t=1}^{T} [(V_t(x_j))^2 \right.$$
$$\left. + D|V_t(x_j)|]\right)^{q-1} dQ^N$$

Therefore, we have, by using Cauchy-Schwarz inequality and that $2(q-1) < 1$, which implies that for any positive reals y and z, $(y+z)^{2(q-1)} \leq y^{2(q-1)} + z^{2(q-1)}$:

$$Z_N \leq (1+\eta)^{N(q-1)}.(W_N)^{\frac{1}{2}}.(Y_N)^{\frac{1}{2}} \tag{2}$$

where

$$W_N = \int \prod_{j=1}^{N} exp[\frac{(q-1)\eta}{\sigma^2} \sum_{t=1}^{T} V_t(x_j)^2] dQ^N \tag{3}$$

$$Y_N = \int \prod_{j=1}^{N} \left(1 + \beta^{2(q-1)} exp(\frac{q-1}{\sigma^2} \sum_{t=1}^{T} [(V_t(x_j))^2 + D|V_t(x_j)|]) \right) dQ^N \tag{4}$$

The rest of the proof consists in controlling W_N and Y_N. To control W_N, we split the integral in two parts and get :

$$W_N \leq exp\left(\frac{N}{\sigma^2}\sqrt{\eta}(q-1)\right) + E_N \tag{5}$$

$$E_N = \int \prod_{j=1}^{N} exp[\frac{(q-1)\eta}{\sigma^2} \sum_{t=1}^{T} V_t(x_j)^2] \mathbf{1}_{\sum_{j=1}^{N} \sum_{t=1}^{T} (V_t(x_j)^2) > \frac{N}{\sqrt{\eta}}} dQ^N$$

Let $\xi < 1$, and $\eta_0 > 0$, such that $\lambda + 2(q-1)\eta_0 \leq \xi$.
We then use the second hypothesis (H2) of the theorem to deduce :
$\forall \eta \leq \eta_0$,

$$E_N \leq C^N \int \prod_{j=1}^{N} exp\left(\frac{1}{2\sigma^2} \sum_{t=1}^{T} [(\xi - 1)(u_j)_t^2 + C|(u_j)_t|] \right) \mathbf{1}_{\sum_{j=1}^{N} \sum_{t=1}^{T} (u_j)_t^2 > \frac{N}{\sqrt{\eta}}} \frac{d(u_1)_1..d(u_N)_T}{(\sigma\sqrt{2\pi})^{NT}}$$

Thus, $\exists C_1$, real constant, such that :

$$E_N \leq C_1^{NT}.exp(\frac{(\xi-1)N}{4\sigma^2\sqrt{\eta}})$$

Therefore, coming back to equation (5), we have, for η small enough :

$$W_N \leq exp\left(2N\sqrt{\eta}(q-1)\right) + C_1^{NT}.exp(\frac{(\xi-1)N}{4\sigma^2\sqrt{\eta}}) \tag{6}$$

We now study Y_N. For any $k \leq N$, let I_N^k be the set of injective applications from $\{1, .., k\}$ into $\{1, .., N\}$.

We develop the product of equation (4)

$$Y_N = 1 + \sum_{k=1}^{N} \frac{\beta^{2k(q-1)}}{k!} \sum_{s \in I_N^k} O_{s,k}$$

$$O_{s,k} = \int \prod_{j=1}^{k} exp\left(\frac{q-1}{\sigma^2}\sum_{t=1}^{T}[(V_t(x_{s(j)}))^2 + D|V_t(x_{s(j)})|]\right) dQ^N$$

Our purpose is now to control $O_{s,k}$ for fixed s and k by C_2^k, where C_2 is a real constant.

Let now $\alpha > 0$ be defined as in the third hypothesis (H3) of the theorem. We choose α smaller than η. Then :

-if $\frac{k}{N} > \alpha$.

The second condition of the theorem and the property $\lambda + 2(q-1) < 1$ allow us to write that there is a constant C_3 such that $O_{s,N} \leq C_3^N$, and therefore we have:

$$O_{s,k} \leq O_{s,N} \leq (C_3^{\frac{1}{\alpha}})^k \tag{7}$$

-if $\frac{k}{N} \leq \alpha$

$$O_{s,k} = \int \prod_{j=1}^{k} exp\left(\frac{q-1}{\sigma^2}\sum_{t=1}^{T}[(V_t(x_{s(j)}))^2 + D|V_t(x_{s(j)})|]\right) \prod_{j=1}^{N} a_j(x) dP^{\otimes N}(x)$$

For any $j \notin \{s(1), .., s(k)\}$, let $\tilde{a}_j$ be defined as in the third hypothesis.

For any $j \in \{s(1), .., s(k)\}$, we use the second condition (H2) to bound a_j.

We are thus obtaining that the integral is bounded by the product of two terms, and therefore get :

$$O_{s,k} \leq (1+\eta)^N.F_N.G_N \tag{8}$$

$$F_N = C^k \int \prod_{j=1}^{k} exp\frac{1}{2\sigma^2}\left(\sum_{t=1}^{T}[(\lambda + 2(q-1))(V_t(x_{s(j)}))^2 + (2D(q-1)+C)\right.$$
$$\left.|V_t(x_{s(j)})|]\right) dP^{\otimes N}(x)$$

$$G_N = \int \prod_{j \notin \{s(1),..,s(k)\}} \tilde{a}_j(x) exp\frac{\eta}{2\sigma^2}\sum_{t=1}^{T}(V_t(x_j))^2 dP^{\otimes N}(x)$$

As $(\lambda + 2(q-1)) < 1$, it is clear that there is a constant C_4, such that

$$F_N \leq C_4^k$$

We use (H3) to replace $\tilde{a}_j$ by a_j, and then (H1) to introduce the missing (a_j)'s and to recover dQ^N :

We thus have:

$$G_N \leq \frac{(1+\eta)^N}{A^k} \int \left(\prod_{j=1}^{N} exp(\frac{\eta}{\sigma^2}\sum_{t=1}^{T}(V_t(x_j))^2)\right).$$
$$\times \left(\prod_{j=1}^{k} exp(B\sum_{t=1}^{T}|(V_t(x_{s(j)}))|)\right) dQ^N$$

Therefore:

$$G_N \leq \frac{(1+\eta)^N}{A^k}(I_N + J_N)$$

$$I_N = \int exp[\frac{\eta}{\sigma^2}\sum_{j=1}^{N}\sum_{t=1}^{T} V_t(x_j)^2 + B\sum_{j=1}^{k}\sum_{t=1}^{T}|V_t(x_{s(j)})|]\mathbf{1}_{\sum_{j=1}^{N}\sum_{t=1}^{T}(V_t(x_j)^2)\leq\frac{N}{\sqrt{\eta}}} dQ^N$$

$$J_N = \int exp[\frac{\eta}{\sigma^2}\sum_{j=1}^{N}\sum_{t=1}^{T} V_t(x_j)^2 + B\sum_{j=1}^{k}\sum_{t=1}^{T}|V_t(x_{s(j)})|]\mathbf{1}_{\sum_{j=1}^{N}\sum_{t=1}^{T}(V_t(x_j)^2)>\frac{N}{\sqrt{\eta}}} dQ^N$$

As $\alpha \leq \eta$ and

$$\sum_{j=1}^{k}\sum_{t=1}^{T}|V_t(x_{s(j)})| \leq \left(\sum_{j=1}^{N}\sum_{t=1}^{T} V_t(x_j)^2\right)^{\frac{1}{2}} .\sqrt{kT}$$

we obtain

$$I_N \leq expN[\frac{\sqrt{\eta}}{\sigma^2} + B\sqrt{T}\eta^{\frac{1}{4}}]$$

Moreover, one can bound J_N the same way as E_N. Therefore, coming back to equation (8), we obtain that there is a constant C_5 such that for η small enough and for N large enough :

$$O_{s,k} \leq (\frac{C_4}{A})^k.(1+\eta)^{2N}.\left(expN[\frac{\sqrt{\eta}}{\sigma^2} + B\sqrt{T}\eta^{\frac{1}{4}}] + C_5^N exp(\frac{(\xi-1)N}{4\sigma^2\sqrt{\eta}})\right)$$

Let

$$C_\alpha = max(C_3^{\frac{1}{\alpha}}, \frac{C_4}{A})$$

We can thus deduce that

$$Y_N \leq (1+\beta^{2(q-1)}C_\alpha)^N.(1+\eta)^{2N}.\left(expN[\frac{\sqrt{\eta}}{\sigma^2} + B\sqrt{T}\eta^{\frac{1}{4}}] + C_5^N exp(\frac{(\xi-1)N}{4\sigma^2\sqrt{\eta}})\right)$$

Thanks to equation (2) and (6), it's now possible to bound Z_N. We choose η and then β small enough to get, for N large enough :

$$Z_N \leq exp\left(\frac{b'Nq}{2p}\right)$$

Let $b = \frac{b'}{2p}$. We deduce from (1) that

$$\pi^N(B(Q,\delta)^c) \leq exp(-bN) \qquad \square$$

4.2. The four conditions of Theorem 4.1

This subsection is dedicated to check the four assumptions of the theorem 4.1 if the couplings satisfy (C). As a matter of fact we will replace condition (C) by the equivalent condition (C'). The equivalence of (C) and (C') is shown in the appendix. We choose a in (C') such that :

$$\int exp(a\theta_j^2)d\nu < +\infty \tag{9}$$

Note that

$$a_j(x) = exp\left(\frac{1}{2\sigma^2}\sum_{t=1}^{T} V_t(x_j)^2\right)\int exp\left(\frac{-1}{2\sigma^2}\sum_{t=1}^{T}[(G_j)_t(x) - V_t(x_j)]^2\right)d\nu(J,\Theta)$$

Then hypothesis (H1) is a direct consequence of Jensen's inequality.

We are now going to check (H2) :

We study

$$r_j(x) = \int exp\left(\frac{-1}{2\sigma^2}\sum_{t=1}^{T}[(G_j)_t(x) - V_t(x_j)]^2\right)d\nu(J,\Theta)$$

Remark that if $h_1, .., h_T$ are integrable real valued functions such that $\forall t \leq T, 0 \leq h_t \leq 1$, and if ν is a probability measure, Cauchy-Schwartz inequality gives :

$$\int \prod_{t=1}^{T} h_t d\nu \leq \prod_{t=1}^{T}\left(\int (h_t)^T d\nu\right)^{\frac{1}{2T}}$$

Thus:

$$r_j(x) \leq \prod_{t=1}^{T}\left[\int exp\left(\frac{-T}{2\sigma^2}[(G_j)_t(x) - V_t(x_j)]^2\right)d\nu\right]^{\frac{1}{2T}}$$

First, we suppose that $V_t(x_j) \geq 0$.

Then:

$$[V_t(x_j) - (G_j)_t(x)]^2 \geq \frac{(V_t(x_j))^2}{4}\mathbf{1}_{2(G_j)_t < V_t(x_j)}$$

Therefore,

$$\int exp\left(\frac{-T}{2\sigma^2}[(G_j)_t(x) - V_t(x_j)]^2\right)d\nu \leq exp(\frac{-T}{8\sigma^2}V_t(x_j)^2) + \nu[exp(\frac{a}{2}(G_j)_t)^2$$
$$\geq exp(\frac{aV_t(x_j)^2}{8})]$$

We obtain the same result if $V_t(x_j) \leq 0$.

Moreover, we easily get

$$((G_j)_t)^2 \leq 2[\sum_{i=1}^{N} J_{ji}(x_i)_{t-1}]^2 + 2\theta_j^2$$

so that we are able to obtain from (C) and (9) a constant D_1 such that

$$\int exp\left(\frac{-T}{2\sigma^2}[(G_j)_t(x) - V_t(x_j)]^2\right) d\nu \leq exp(\frac{-T}{8\sigma^2}V_t(x_j)^2) + D_1 exp(-\frac{a}{8}V_t(x_j)^2)$$

therefore, we can bound $r_j(x)$ and achieve the proof. □

We are now going to focus our interest on the checking of the third condition (H3) :

Let $s \in I_N^k$ and $j \notin \{s(1), ..s(k)\}$.

Let us split $(G_j)_t(x)$ into two parts, and set :

$$(G_j^1)_t(x) = \sum_{i \notin \{s(1),..s(k)\}} J_{ji}(x_i)_{t-1} + \theta_j$$

$$(G_j^2)_t(x) = \sum_{i \in \{s(1),..s(k)\}} J_{ji}(x_i)_{t-1} = (G_j)_t(x) - (G_j^1)_t(x)$$

Let

$$\tilde{a}_j(x) = \int exp\left(\frac{1}{\sigma^2}\sum_{t=1}^{T}(G_j^1)_t V_t(x_j) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}[(G_j^1)_t]^2\right) d\nu(J, \Theta)$$

We use Hölder to get :

$$\tilde{a}_j(x) \leq (K_j)^{\frac{1}{p}}.(L_j)^{\frac{1}{q}} \tag{10}$$

where $\frac{1}{p} + \frac{1}{q}$.

$$K_j = \int exp\left(\frac{p}{\sigma^2}\sum_{t=1}^{T}(G_j)_t V_t(x_j) - \frac{p}{2\sigma^2}\sum_{t=1}^{T}[(G_j)_t]^2\right) d\nu(J, \Theta)$$

$$L_j = \int exp\left(\frac{-q}{\sigma^2}\sum_{t=1}^{T}(G_j^2)_t V_t(x_j) + \frac{q}{2\sigma^2}\sum_{t=1}^{T}[2(G_j)_t(G_j^2)_t - (G_j^2)_t^2]\right) d\nu(J, \Theta)$$

Thus, we have

$$K_j \leq exp\left(\frac{p}{2\sigma^2}\sum_{t=1}^{T}V_t(x_j)^2\right).\int exp\left(\frac{-p}{2\sigma^2}\sum_{t=1}^{T}[(G_j)_t(x) - V_t(x_j)]^2\right) d\nu(J, \Theta)$$

We can deduce that

$$K_j \leq exp\left(\frac{p-1}{2\sigma^2}\sum_{t=1}^{T}V_t(x_j)^2\right).a_j(x)$$

If we apply (H1) :

$$K_j^{\frac{1}{p}} \leq a_j(x).\frac{1}{A^{\frac{1}{q}}} exp\left(\frac{1}{2q\sigma^2}\sum_{t=1}^{T}[V_t(x_j)]^2 + \frac{B}{q}\sum_{t=1}^{T}|V_t(x_j)|\right)$$

Moreover, let $\varepsilon > 0$. For any $a, b > 0$, $ab \leq \frac{a^2}{2\varepsilon} + \frac{\varepsilon b^2}{2}$. Therefore:

$$L_j \leq exp\left(\frac{q\varepsilon}{2\sigma^2}\sum_{t=1}^{T} V_t(x_j)^2\right) M_j$$

where :

$$M_j = \int \left(exp\frac{q\varepsilon}{2\sigma^2}\sum_{t=1}^{T}[(G_j)_t]^2\right).\left(exp\frac{q}{\sigma^2\varepsilon}\sum_{t=1}^{T}[(G_j^2)_t]^2\right)d\nu$$

Thus, from Cauchy-Schwartz, we deduce :

$$M_j = \left(\int exp\frac{q\varepsilon}{\sigma^2}\sum_{t=1}^{T}[(G_j)_t]^2 d\nu\right)^{\frac{1}{2}}.\left(\int exp\frac{2q}{\sigma^2\varepsilon}\sum_{t=1}^{T}[(G_j^2)_t]^2 d\nu\right)^{\frac{1}{2}}$$

We choose

$$q = \frac{a\sigma^2}{4T}.\sqrt{\frac{N}{k}}$$

$$\epsilon = \frac{a\sigma^2}{4qT}$$

where a is the real number defined in condition (C).

Remark that if $h_1, .., h_T$ are integrable real valued functions such that $\forall t \leq T, h_t \geq 1$, and if ν is a probability measure, Cauchy-Schwartz inequality gives :

$$\int \prod_{t=1}^{T} h_t d\nu \leq \prod_{t=1}^{T}\left(\int (h_t)^{2T} d\nu\right)^{\frac{1}{T}}$$

Hence :

$$M_j \leq \prod_{t=1}^{T}\left[\left(\int exp(\frac{a}{2}(G_j)_t^2)d\nu\right).\left(\int exp[\frac{aN}{k}(G_j^2)_t^2]d\nu\right)\right]^{\frac{1}{2T}}$$

Then condition (C) implies that there is a constant D_2 such that

$$L_j \leq exp\left(\frac{a}{4T}\sum_{t=1}^{T} V_t(x_j)^2\right) D_2$$

Therefore, on can deduce from equation (10) that for any $\eta > 0$, if $\frac{k}{N}$ is small enough (so that q is large enough), :

$$\tilde{a}_j(x) \leq a_j(x).(1+\eta).exp(\eta \sum_{t=1}^{T} (V_t(x_j))^2)$$

The same method can be used to bound $\frac{a_j(x)}{\tilde{a}_j(x)}$. □

We study the fourth hypothesis (H4) :

Let $(\tilde{J}_{ij})_{1\leq i,j\leq N}$ be gaussian, independent random variables, with mean $\frac{\bar{J}}{N}$, and variance $\frac{J^2}{N}$. These variables are also supposed to be independent from the θ_j's and the J_{ij}'s. The distribution of θ_j is denoted by ν_θ, and the jointed law of the couplings $(J_{ij}, \tilde{J}_{ij})$ is $\nu_{(J,\tilde{J})}$.

For any $1 \leq t \leq T$, we consider :

$$(\tilde{G}_j)_t = \sum_{i=1}^{N} \tilde{J}_{ji}(x_i)_{t-1} + \theta_j$$

Then,

$$\left| \frac{a_j(x)}{b_j(x)} - 1 \right| \leq N_j(x).O_j(x) \tag{11}$$

where

$$N_j(x) = \sqrt{det(I + \frac{K^{\hat{\mu}_N(x)}}{\sigma^2})}.exp(\frac{1}{2\sigma^2}\phi(x))$$

$$\phi(x) =^t (V(x_j) - c^{\hat{\mu}_N(x)})(A^{\hat{\mu}_N(x)} - I_T)(V(x_j) - c^{\hat{\mu}_N(x)}) + {}^tV(x_j)V(x_j)$$

$$O_j(x) = \left| \int \left[exp\left(\frac{-1}{2\sigma^2} \sum_{t=1}^{T} [(G_j)_t - V_t(x_j)]^2 \right) \right.\right.$$
$$\left.\left. -exp\left(\frac{-1}{2\sigma^2} \sum_{t=1}^{T} [(\tilde{G}_j)_t - V_t(x_j)]^2 \right) \right] d\nu_{(J,\tilde{J})} d\nu_\theta \right|$$

Recall here that the eigenvalues of $A^{\hat{\mu}_N(x)}$ are positive and bounded by 1.

Moreover, we have : - $\forall 1 \leq t \leq T$, $|c_t^{\hat{\mu}_N(x)}| \leq |\bar{J}|$,

- $\forall 1 \leq t, s \leq T$, $|K_{t,s}^{\hat{\mu}_N(x)}| \leq J^2 + \tau^2$.

It follows that we can find two real constants C_1 and C_2 such that :

$$N_j(x) \leq C_1 exp\left(\frac{1}{2\sigma^2} \sum_{t=1}^{T} [V_t(x_j)^2 + C_2|V_t(x_j)|] \right)$$

We shall now prove that $O_j(x)$ is uniformly small in x for N large enough.

We first get the following lemma. Its proof is close to the demonstration of Lindeberg's theorem (see [18]).

Lemma 4.2. *Let N be fixed, and let $Y_1, .., Y_N$ be N independent, centered, $\mathbb{R}^T$-valued random variables on a given probability space $(\Omega, \mathcal{A}, \nu)$. We consider:*

$$(M_i)_{ts} = E((Y_i)_t (Y_i)_s)$$

Let $\tilde{Y}_1, .., \tilde{Y}_N$ be independent, centered, $\mathbb{R}^\mathrm{T}$-valued gaussian random variables. We suppose that the covariance of $\tilde{Y}_i$ is M_i and that the $\tilde{Y}_i$'s are independent from the Y_i's. $\forall i \in \{1, .., N\}$, let $S_N = Y_1 + .. + Y_N$, and $\tilde{S}_N = \tilde{Y}_1 + .. + \tilde{Y}_N$.

Let $\Phi \in \mathcal{C}^3(\mathbb{R}^\mathrm{T}, \mathbb{R})$, whose derivatives of first, second and third order are uniformly bounded by a constant C_3.

If W is a $\mathbb{R}^\mathrm{T}$-valued vector, let $||W|| = \sum_{t=1}^{T} |W_t|$.

Then, if

$$O = \left| \int \Phi\left((S_N)_1, .., (S_N)_T\right) d\nu - \int \Phi\left((\tilde{S}_N)_1, .., (\tilde{S}_N)_T\right) d\nu \right|$$

we have :

$\forall \epsilon > 0$,

$$O \leq C_3 \left([\frac{\epsilon}{6} \sum_{i=1}^{N} \int ||Y_i||^2 d\nu] + \sum_{i=1}^{N} [\int ||\tilde{Y}_i||^3 d\nu + \int (||Y_i||^2 \mathbf{1}_{||Y_i|| > \epsilon}) d\nu] \right)$$

Proof. $\forall i \in \{1, .., N\}$, let

$$(U_i) = \sum_{k=1}^{i-1} Y_k + \sum_{k=i+1}^{N} \tilde{Y}_k$$

where the first sum is taken to be 0 if $i = 1$ and the second sum is 0 if $i = N$.

We then have easily

$$O \leq \sum_{i=1}^{N} O_i$$

with

$$O_i = \left| \int \Phi(U_i + Y_i) d\nu - \int \Phi(U_i + \tilde{Y}_i) d\nu \right|$$

We consider

$$R_i(\xi) = \Phi(U_i + \xi) - \Phi(U_i) - \sum_{t=1}^{T} \xi_t \frac{\partial \Phi}{\partial x_t}(U_i) - \frac{1}{2} \sum_{t=1}^{T} \sum_{s=1}^{T} \xi_t \xi_s \frac{\partial^2 \Phi}{\partial x_t \partial x_s}(U_i)$$

Therefore (because Y_i and $\tilde{Y}_i$ are centered, independent of U_i and have the same covariance matrix),

$$O_i = \left| \int R_i(Y_i) d\nu - \int R_i(\tilde{Y}_i) d\nu \right|$$

so that

$$O_i \leq |\int R_i(Y_i)d\nu| + |\int R_i(\tilde{Y}_i)d\nu|$$

Thanks to Taylor's theorem, we obtain :

$$|R_i(\xi)| \leq C_3 inf(||\xi||^2, \frac{||\xi||^3}{6})$$

therefore

$$O \leq C_3\left(\frac{1}{6}\sum_{i=1}^{N} E(||Y_i||^3\mathbf{1}_{||Y_i||\leq\epsilon}) + \sum_{i=1}^{N}[\frac{1}{6}E(||\tilde{Y}_i||^3) + E(||Y_i||^2\mathbf{1}_{||Y_i||>\epsilon})]\right)$$

We thus deduce the result of the lemma. □

Let

$$a_t = \theta_j + \frac{\bar{J}}{N}\sum_{i=1}^{N}(x_i)_{t-1} - V_t(x_j)$$

Let's define

$$\Phi(y_1, .., y_T) = \prod_{t=1}^{T}\phi(y_t + a_t)$$

where

$$\phi(z) = exp(\frac{-z^2}{2\sigma^2})$$

Notice that the derivatives of first, second and third order of function Φ are uniformly bounded by a real constant.

Let's define

$$(Y_i)_t = (J_{ji} - \frac{\bar{J}}{N})(x_i)_{t-1}$$

$$(\tilde{Y}_i)_t = (\tilde{J}_{ji} - \frac{\bar{J}}{N})(x_i)_{t-1}$$

We focus our interest on O_j and first integrate with respect to $\nu_{(J,\tilde{J})}$.

Then let $\epsilon > 0$. Lemma 4.2 gives that there is a constant C_3 such that:

$$O_j(x) \leq C_3\left([\frac{\epsilon}{6}\sum_{i=1}^{N}\int(||Y_i||^2)d\nu_{(J,\tilde{J})}] + \sum_{i=1}^{N}[\int(||\tilde{Y}_i||^3)d\nu_{(J,\tilde{J})} + \int(||Y_i||^2\mathbf{1}_{||Y_i||>\epsilon})d\nu_{(J,\tilde{J})}]\right)$$

We have :

$$\int(||Y_i||^2)d\nu_{(J,\tilde{J})} \leq \frac{J^2T^2}{N}$$

$$\int(||\tilde{Y}_i||^3)d\nu_{(J,\tilde{J})} \leq \int(|y|^3 exp - (\frac{y^2}{2})\frac{dy}{\sqrt{2\pi}}.\frac{J^3T^3}{N^{\frac{3}{2}}}$$

Moreover,

$$r_i(\epsilon) = \int (||Y_i||^2 \mathbf{1}_{||Y_i||>\epsilon}) d\nu_{(J,\tilde{J})} \leq \frac{2T^2}{N}[\frac{\bar{J}^2}{N} + \int (N||J_{ji}||^2 \mathbf{1}_{||Y_i||>\epsilon}) d\nu_{(J,\tilde{J})}]$$

If a is the real defined in condition (C), let $C_a = \sup_{x \geq 0} x^2 exp(-ax)$.
Hence, from Cauchy-Schwartz, we get:

$$r_i(\epsilon) \leq \frac{2T^2}{N}[\frac{\bar{J}^2}{N} + \left(\frac{JT}{\epsilon\sqrt{N}}\right).C_a \left(\int exp(aNJ_{ji}^2) d\nu_{J,\tilde{J}}\right)^{\frac{1}{2}}]$$

Therefore, using (C) and summing over i, we can conclude that for any $\beta > 0$, for N large enough, for any $x \in ((]0,1[^{[0,T]})^N)$, we have :

$$O_j(x) \leq \beta$$

We finally come back to equation (11) to conclude. □

This achieves the proof of Theorem 2.3.

Finally, we are able to study the limit probability Q and to prove the mean-field equations (theorem 2.6) which characterize this limit distribution.

We proved in proposition 3.12 that Q is the unique fixed point of L. We denote by Q' the restriction of Q to $\sigma(x_1, .., x_T)$. We then have $Q = \mu_0 \otimes Q'$. We consider $\tilde{Q}'$, the image of Q' by f.

We note $V(u) = (u_1 - \bar{\theta}, ..., u_T - \bar{\theta})$. Then, from the expression of L (see section 3), we deduce :

$$\frac{d\tilde{Q}'}{du} = (\frac{1}{\sqrt{2\pi}})^T \frac{1}{\sqrt{det(\sigma^2 I_T + K^Q)}} exp\chi(u)$$

with

$$\chi(u) = \frac{1}{2\sigma^2}\left({}^t(V(u) - c^Q)(A^Q - I_T)(V(u) - c^Q)\right)$$

One can deduce from this expression that $(U_t)_{1 \leq t \leq T}$ is a discrete time gaussian process, with mean $c^Q + \bar{\theta}$ and covariance matrix defined by

$$R = \sigma^2 (A^Q - I)^{-1}$$

By diagonalizing K^Q in an orthonormal basis, we can deduce, after some calculations, that :

$$R = K^Q + \sigma^2 I_T$$

This gives the mean-field equations. □

5. Appendix: about the condition (C) of domination by gaussian tails

This appendix will be devoted to the consequences of condition (C). First we will show that the technical condition (C') is the consequence of condition (C). Since condition (C') obviously implies condition (C), this proves the equivalence of condition (C) and condition (C'). Then we show that condition (C) is stable by barycentric combination. Eventually, we show it is checked by "natural" models with appropriate scaling. We prove here the equivalence between the two conditions (C) and (C').

We first give a sufficient condition on the couplings to satisfy (C') :

Lemma 5.1. *Let J_1 be a random variable with law ν_J^N. We suppose that :*
$\exists C_1 > 0, \forall N \geq 1, \forall l \geq 2,$

$$E\left(|J_1 - \frac{\bar{J}}{N}|^l\right) \leq \left(\frac{C_1}{\sqrt{N}}\right)^l . \frac{l!}{(E(\frac{l}{2}))!}$$

Then the technical condition (C') on the couplings is fulfilled.

This means that if the moments of the connection weights decrease with N as fast as in the gaussian case, then condition (C') occurs.

Recall here (C') :
$\exists a > 0, \exists D_0 > 0, \forall N \geq 1, \forall k \leq N, \forall (J_1, .., J_k)$, independent random variables with law ν_J^N, $\forall(\lambda_1, .., \lambda_k) \in [0, 1]^k$,

$$E\left(exp[\frac{aN}{k}(\lambda_1 J_1 + .. + \lambda_k J_k)^2]\right) \leq D_0$$

Let $N \geq 1, k \leq N$. Let $J_1, .., J_k$ be independent random variables with law ν_J^N. Their common mean is $\frac{\bar{J}}{N}$ and their common variance $\frac{J^2}{N}$.

Remark that (C) is satisfied if $J = 0$. When $J \neq 0$, we define, for $1 \leq i \leq k$:

$$W_i = \frac{\sqrt{N}}{J}\left(J_i - \frac{\bar{J}}{N}\right)$$

so that $E(W_i) = 0$, $E(W_i^2) = 1$.

The condition (D) in proposition 2.7 gives a constant D_2, which does not depend on k and N, such that, for any $l > 2$, for $1 \leq i \leq k$:

$$E(|W_i^l|) \leq D_2^l . \frac{l!}{(E(\frac{l}{2}))!} \tag{12}$$

Let $(\lambda_1, .., \lambda_k) \in [0, 1]^k$.
It is sufficient to find $a > 0$ such that

$$S = E\left(exp[\frac{a}{k}(\lambda_1 W_1 + .. + \lambda_k W_k)^2]\right)$$

is uniformly bounded with regard to k and N.
Let $B_{k,p} = \{i = (i_1, ..i_k) \in (\mathbb{N} - \{1\})^k, i_1 + .. + i_k = 2p\}$.

As the first order moments of the W_i's are zero, we can deduce that :

$$S = \sum_{p=0}^{+\infty} \frac{(2p)!a^p}{k^p p!} \sum_{i \in B_{k,p}} \frac{\lambda_1^{i_1}..\lambda_k^{i_k}}{(i_1)!..(i_k)!} E(W_1^{i_1})..E(W_k^{i_k})$$

Thus (12) gives

$$S \leq \sum_{p=0}^{+\infty} \frac{D_2^{2p}.a^p(2p)!}{k^p p!} \sum_{i \in B_{k,p}} \left(\frac{1}{E(\frac{i_1}{2})!..E(\frac{i_k}{2})!}\right)$$

Let
$B_r = \{i = (i_1, ..i_k) \in B_{k,p}, i_1, .., i_r$ odd, $i_{r+1}, .., i_k$ even $\}$
Then

$$S \leq \sum_{p=0}^{+\infty} \frac{D_2^{2p}.a^p(2p)!}{k^p p!} \sum_{r=0}^{\inf(\frac{2p}{3},k)} C_k^r S_r$$

where

$$S_r = \sum_{i \in B_r} \frac{1}{E(\frac{i_1}{2})!..E(\frac{i_k}{2})!}$$

Notice that r is necessarily even, and consider the following change of index :
for $1 \leq p \leq \frac{r}{2}$, let $j_p = \frac{i_p - 3}{2}$.
For $\frac{r}{2} < p \leq r$, let $j_p = \frac{i_p - 1}{2}$.
For $r + 1 \leq p \leq k$, let $j_p = \frac{i_p}{2}$.
Therefore :

$$S_r \leq \sum_{j_1+..+j_k=p-r} \frac{1}{(j_1)!..(j_k)!}$$

$$S_r \leq \frac{k^{p-r}}{(p-r)!}$$

As

$$C_k^r \leq \frac{k^r}{r!}$$

we deduce

$$S \leq \sum_{p=0}^{+\infty} \frac{(2p)!D_2^{2p}.(2a)^p}{(p!)^2} = \frac{1}{\sqrt{1 - 2D_2^2 a}}$$

We are thus able to choose a small enough and to complete the proof of the lemma. □

We are now going to prove that (C) implies the hypothesis of the lemma.

If (C) is satisfied, it is possible to find a constant D_1 such that : $\forall N \geq 1$,

$$E\left(exp[\frac{a}{2}N(J_1-\frac{\bar{J}}{N})^2]\right) \leq D_1$$

We obtain :
$\forall N \geq 1, \forall k \geq 1$,

$$E\left(\frac{a^k N^k}{2^k k!}(J_1-\frac{\bar{J}}{N})^{2k}\right) \leq D_1$$

Therefore :

$$E\left((J_1-\frac{\bar{J}}{N})^{2k}\right) \leq \frac{D_1.2^k.(2k)!}{(aN)^k.k!}$$

As

$$E\left((J_1-\frac{\bar{J}}{N})^{2k+1}\right) \leq \sqrt{E\left(J_1-\frac{\bar{J}}{N})^{2k}\right)}.\sqrt{E\left(J_1-\frac{\bar{J}}{N})^{2(k+1)}\right)}$$

we deduce that conditions of lemma 5.1 are satisfied, and therefore that (C) implies (C'). □

Furthermore, we also have the following mixing stability property :

Proposition 5.2. *We suppose that condition (C) is fulfilled by couplings with respective laws ρ_1 and ρ_2. Let $\beta > 0$ and consider couplings with law $\rho = \beta\rho_1 + (1-\beta)\rho_2$. These connection weights satisfy (C'), and therefore (C).*

Proof. Let $N \geq 1, k \leq N$. Let $J_1, .., J_k$ be independent random variables with law ρ.

Then

$$E\left(exp[\frac{aN}{k}(\lambda_1 J_1 + .. + \lambda_k J_k)^2]\right) = \sum_{r=0}^{k} C_k^r \beta^r (1-\beta)^{k-r} Z_r$$

$$Z_r = \int exp\left(\frac{aN}{k}[\lambda_1 y_1 + .. + \lambda_r y_r + \lambda_{r+1} z_{r+1} + .. + \lambda_k z_k]^2\right) \prod_{i=1}^{r} d\rho_1(y_i)$$
$$\prod_{i=r+1}^{k} d\rho_2(z_i)$$

We suppose that (C) is satisfied by ρ_1 and ρ_2 with the same real number a_0. Then let $a = \frac{a_0}{2}$, and remark that :

$$Z_r \leq \int exp\left(\frac{aN}{r}[\sum_{i=1}^{r}\lambda_i y_i]^2\right)\prod_{i=1}^{r} d\rho_1(y_i).\int exp\left(\frac{aN}{k-r}[\sum_{i=r+1}^{k}\lambda_i z_i]^2\right)\prod_{i=r+1}^{k} d\rho_2(z_i)$$

we are therefore able to conclude. □

We are finishing this section by giving some concrete examples, among others, of couplings law that fulfill (C). The parameters are chosen so that the mean of the connection weights is $\frac{\bar{J}}{N}$ and their variance $\frac{J^2}{N}$.

1) The law ν_J^N is a uniform distribution on $[\frac{\bar{J}}{N} - \frac{J\sqrt{3}}{\sqrt{N}}, \frac{\bar{J}}{N} + \frac{J\sqrt{3}}{\sqrt{N}}]$.

2) The distribution ν_J^N is a $\frac{1}{2}(\delta_a + \delta_b)$, where $a = \frac{\bar{J}}{N} - \frac{J}{\sqrt{N}}$, and $b = \frac{\bar{J}}{N} + \frac{J}{\sqrt{N}}$

3) The couplings are the product of a gaussian random variable $\mathcal{N}(\frac{\bar{J}}{pN}, \frac{J^2}{pN} + \frac{\bar{J}^2(p-1)}{N^2p^2})$ and of a Bernoulli with parameter $p \in [0, 1]$.

This gives the mean-field equations for some diluted networks.

6. Conclusion and perspectives

Our work gives a rigorous basis to previous results obtained by statistical physicists about dynamics of large size recurrent neural networks. In the case of gaussian couplings, we proved a large deviations principle and deduced a law of large numbers, a propagation of chaos property and the mean-field equations which describe the limit behavior of the neurons.

An adapted generalization of Lindeberg's theorem allowed us to extend these results for some more general connection weights distribution, and in particular for discrete or uniform couplings and for diluted networks. It seems that this methodology can be used in a spin glass context, although it should be more technical. Remark, moreover, that obtaining our mean-field properties for non gaussian couplings is important in a Neural Network context : although our models are still far from biological realism, they try to reproduce certain features of the brain, which is a sparsely connected neural network. A next step would be to establish a proof for the dynamics of large size networks composed of spiking neurons.

As underlined in [3, 20], mathematicians have now obtained few results about symmetric or asymmetric neural networks and spin glasses. We are far from proving all the physicists assertions completely rigorously. As a matter of fact, our study is only completed for bounded time. The determination of the dynamics of large networks without any temporal limit is an open problem.

References

[1] Amari, S.: Characteristics of random nets of analog-like elements. *IEEE Trans. Syst. Man. Cyb.*, SMC-2(N5), 1972

[2] Amari, S., Yoshida K., Kanatani K.: A mathematical fundation for statistical neurodynamics. *SIAM Journal appl. math.*, 1977

[3] Ben Arous, G., Guionnet, A.: Large deviations for langevin spin glass dynamics. *PTRF*, **102**: 455–509 (1995)

[4] Cessac, B.: *Propriétés statistiques des dynamiques de réseaux neuromimétiques*. PhD thesis, Université Paul Sabatier, Toulouse, 1994

[5] Cessac, B.: Increase in complexity in random neural networks. *Journal de Physique I*, **5**: 409–432 (1995)

[6] Cessac, B., Doyon, B., Quoy, M., Samuelides, M.: Mean-field equations, bifurcation map and route to chaos in discrete time neural networks. *Physica D*, **74**: 24–44 (1994)
[7] Crisanti, A., Sommers, H.J., Sompolinsky, H.: chaos in neural networks: chaotic solutions *Preprint*, 1990
[8] Daucé, E., Moynot, O., Pinaud, O., Samuelides, M.: Etude de la dynamique de réseaux neuronaux aléatoires récurrents. ONERA/DTIM, 1999
[9] Dembo, A., Zeitouni, O.: *Large Deviations techniques ansd Applications*. jones and Bartlett Publishers, 1992
[10] Deuschel and Stroock. *Large Deviations*. Academic press, New York, 1989
[11] Dobrushin. Prescribing a system of random variables with conditional distributions. *Theor. Proba. Appl.*, 1970
[12] Geman, S.: Almost sure Stable oscillations in a large system of randomly coupled equations. *Siam J. Appl. Math.*, **42**: 695–703 (1982)
[13] Geman, S., Hwang, C.R.: A chaos hypothesis for some large systems of random equations. *Z. Wahr. Verw. Gebiete*, **60**: 291–314 (1982)
[14] Grunwald, M.: Sanov results for glauber spin-glasses. *Prob. Theory and Rel. Fields*, 1996
[15] Guionnet, A.: Averaged and quenched propagation of chaos for spin glass dynamics. *PTRF*, *109*: 183–215 (1997)
[16] Haykin, S.: *Neural Networks, a comprehensive foundation*. Prentice Hall, New York 1999
[17] Sompolinsky, H., Crisanti, A., Sommers, H.J.: Chaos in random neural networks. *Phys. Rev. Lett.*, **61**: 259–262 (1988)
[18] Stroock.: *Probability theory*. Cambridge University Press, (1993)
[19] Sznitman, A.S.: Equations de Boltzmann spatialement homogènes *Z.F. Wahrscheinlichkeitstheorie Verw. Gebiete*, 1984
[20] Talagrand, M.: The Sherrington-Kirkpatrick Model: a challenge for mathematicians. *Probability Theory and Related Fields*, 1998