



HAL
open science

Étude asymptotique d'un réseau neuronal: le modèle de mémoire associative de Hopfield

Franck Vermet

► **To cite this version:**

Franck Vermet. Étude asymptotique d'un réseau neuronal: le modèle de mémoire associative de Hopfield. Mathématiques [math]. Université Rennes 1, 1994. Français. NNT: . tel-00598243

HAL Id: tel-00598243

<https://theses.hal.science/tel-00598243>

Submitted on 5 Jun 2011

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.

N° d'Ordre : 1087

THÈSE

Présentée

DEVANT L'UNIVERSITÉ DE RENNES I

pour obtenir

le grade de DOCTEUR DE L'UNIVERSITÉ DE RENNES I

Mention Mathématiques et Applications

par

Franck VERMET

Institut de Recherche Mathématique de Rennes

U.F.R. de Mathématiques

**Étude asymptotique d'un réseau neuronal : le modèle
de mémoire associative de Hopfield**

Soutenue le 28 janvier 1994 devant la Commission d'Examen

COMPOSITION DU JURY :

| | | |
|-----|----------------|-------------------------|
| Pr. | F. Comets | Rapporteur et Président |
| Pr. | J. Mémin | Examineur |
| Pr. | D. Petritis | Directeur |
| Pr. | P. Picco | Rapporteur |
| Pr. | B. Zegarlinski | Examineur |

Préface

Je tiens à remercier MM. F.Comets, J.Memin, P.Picco et B.Zegarliniski qui m'ont fait l'honneur de participer au jury.

Je tiens aussi à remercier mon directeur de thèse, D.Petritis, qui m'a aidé tout au long de ces deux années de recherche, d'un point de vue scientifique, mais également pour tous les aspects matériels. Je lui suis sincèrement très reconnaissant pour les conseils qu'il m'a donnés et les services qu'il m'a toujours rendus avec la plus grande sympathie.

Table des matières

| | |
|--|-----------|
| Préface | 3 |
| 1 Introduction | 1 |
| 2 Introduction au calcul neuronal | 5 |
| 2.1 Description des réseaux formels de neurones | 5 |
| 2.2 Description mathématique de la mémoire associative | 7 |
| 3 Étude de la dynamique séquentielle déterministe de Hopfield | 11 |
| 3.1 Définitions | 11 |
| 3.2 Étude asymptotique de la stabilité des images originales et de la convergence directe | 14 |
| 3.2.1 Stabilité des images originales | 14 |
| 3.2.2 Convergence directe des images originales bruitées | 16 |
| 3.2.3 Convergence directe des “mosaïques aléatoires” | 19 |
| 3.3 Étude des fluctuations du hamiltonien | 23 |
| 3.3.1 Mémoire associative avec erreurs tolérées | 23 |
| 3.3.2 Existence d’autres minima locaux : les images combinées | 24 |
| 3.3.3 Convergence presque sûre de $\frac{1}{N}H_{N,p}^S(\underline{\xi}, \cdot)$ pour certaines configurations dépendant des images originales | 31 |
| 4 Étude de la dynamique parallèle déterministe de Little | 41 |

| | | |
|----------|--|-----------|
| 4.1 | Définitions | 41 |
| 4.2 | Étude asymptotique de la stabilité et de la convergence directe | 43 |
| 5 | Étude de la dynamique stochastique | 47 |
| 5.1 | Description à volume fini | 47 |
| 5.1.1 | La dynamique séquentielle de Glauber | 48 |
| 5.1.2 | La dynamique parallèle de Glauber | 52 |
| 5.2 | Description à volume infini | 53 |
| 6 | Étude de l'énergie libre associée au modèle de Hopfield | 61 |
| 6.1 | Définition de l'énergie libre | 61 |
| 6.2 | Étude de l'énergie libre "annealed" | 63 |
| 6.3 | Convergence presque sûre de l'énergie libre et de son espérance, pour $\frac{p}{N} \rightarrow 0$ | 67 |
| 6.4 | Propriété de self-averaging de l'énergie libre pour toute température et $\frac{p}{N} \rightarrow \alpha \geq 0$ | 83 |
| 7 | Conclusion | 93 |
| | Bibliographie | 96 |

Chapitre 1

Introduction

La *mémoire associative* est la capacité de retrouver de l'information à partir d'une connaissance partielle de cette information.

Dans les ordinateurs classiques, chaque donnée est enregistrée à une adresse précise, et obtenir de nouveau cette donnée requiert la connaissance exacte de cette adresse. Bien entendu, le processus de mémorisation dans le cerveau ne peut fonctionner de cette manière. Si on cherche le nom d'une personne, il est inutile de savoir qu'il s'agit du 3743ème nom appris depuis notre naissance. Par contre, la donnée de divers indices concernant cette personne peut faire ressurgir au niveau conscient le nom cherché. Il est clair que les associations jouent un rôle fondamental dans les processus cognitifs, que ce soit au niveau de la mémorisation ou du souvenir.

De plus, on sait maintenant que telle information stockée dans le cerveau ne l'est pas au niveau de quelques neurones, à l'image d'un nombre stocké sur quelques bits d'une mémoire informatique. La mémorisation se fait de manière délocalisée, à grande échelle au niveau neuronal, et il semble que ce soit les connexions interneuronales, appelées synapses, qui "contiennent" l'information. Cette idée fut développée par Hebb [23], sous le nom de plasticité synaptique. Il postulait alors que les potentiels synaptiques peuvent être ajustés en fonction du degré d'activité des neurones.

Rappelons que les neurones communiquent entre eux via les axones, terminés par les synapses. L'axone peut être dans deux états : il propage, ou ne propage pas un potentiel d'action, sous forme de signal chimique, jusqu'à la liaison synaptique. L'arrivée d'un signal au niveau du neurone postsynaptique déclenche la sécrétion de neurotransmetteurs, engendrant la pénétration d'un courant ionique dans ce neurone. Ce courant caractérise l'efficacité synaptique qui peut être excitatrice (dépolarisant la membrane du neurone postsynaptique), ou inhibitrice (hyperpolarisant la membrane postsynaptique). Ainsi, dans chaque petit intervalle de temps, chaque neurone somme les différents potentiels lui parvenant, et en dessous d'un certain seuil (environ -60mV), se décharge violemment sous la forme d'un potentiel d'action, de force et de durée constante, envoyé aux

axones.

Cette description très simplifiée nous permet de comprendre le mécanisme décrit par Hebb : une synapse active, déclenchant de manière répétitive l'activation de son neurone postsynaptique, voit son efficacité augmenter, tandis que les autres voient la leur diminuer. Ce phénomène de plasticité synaptique joue un rôle dominant dans les processus complexes d'apprentissage et de mémorisation.

Les premiers modèles mathématiques de réseaux neuronaux sont dûs entre autres à Mac Culloch et Pitts (1943), et Rosenblatt (1960). Pour une description de ces modèles, dont le célèbre Perceptron de Rosenblatt, le lecteur pourra consulter plusieurs ouvrages [4, 37, 25, 8]. Tous les modèles de réseaux de neurones sont trop simplistes pour pouvoir décrire le fonctionnement du cerveau. Néanmoins, même en ne reproduisant que grossièrement le comportement des neurones biologiques, on peut construire des systèmes ayant des propriétés très intéressantes, et de nombreuses applications en informatique.

Ainsi le modèle proposé par Hopfield en 1982 [26] reproduit-il un processus de mémoire associative. Des images sont mémorisées par un réseau formel, de manière délocalisée, au niveau des poids synaptiques d'interaction. Hopfield fut le premier à introduire la notion de fonction énergie sur l'espace des configurations possibles, et à considérer les images mémorisées comme des attracteurs stables pour une dynamique. Je souhaite, par cette thèse contribuer modestement à la compréhension de ce modèle. En fait, nous étudions deux modèles, nommés modèle de Hopfield, et modèle de Little, et correspondant respectivement à une dynamique séquentielle et parallèle. Étant construits de manière analogue, nous verrons qu'ils ont un comportement très similaire.

Il me paraît judicieux de commencer cette étude par un chapitre d'introduction au "calcul neuronal". Le chapitre suivant est consacré à la description mathématique du phénomène de mémoire associative. Ensuite, nous présentons le modèle de Hopfield, tel qu'il fut originalement défini, et après avoir redémontré les propriétés de stabilité vérifiées par ce modèle, nous montrons que pour certaines configurations, suffisamment proches d'une image mémorisée, le processus de mémoire associative se réalise parfaitement en une seule étape de la dynamique, sous certaines hypothèses sur le nombre p d'images mémorisées. Cela nous conduit à étudier la structure des minima de la fonction énergie associée à la dynamique séquentielle de Hopfield ; après avoir rappelé les résultats obtenus par C.Newman [38], nous montrons la convergence presque sûre du hamiltonien normalisé vers des constantes bien déterminées, pour les images originales et des combinaisons de ces images.

Le chapitre 4 concerne les propriétés de stabilité et de convergence directe du modèle de Little : ce modèle est très proche de celui de Hopfield, puisque les efficacités synaptiques sont les mêmes ; il en diffère seulement par la dynamique d'évolution, qui est parallèle, contrairement à la dynamique séquentielle du modèle de Hopfield. Cette approche peut présenter un intérêt en informatique, pour la parallélisation des calculs.

Du fait de l'existence de nombreux points stables parasites (qui sont des minima locaux

du hamiltonien), il est nécessaire de considérer une dynamique stochastique, que nous présentons au chapitre 5. En autorisant avec une probabilité non nulle des transitions vers des états inaccessibles par la dynamique déterministe, on espère pouvoir quitter les états parasites stables, et converger vers les configurations de plus basse énergie (dynamique de Glauber). Malheureusement, il est facile de montrer que pour un réseau de taille fini, il s'agit d'une chaîne de Markov irréductible, convergeant pour toute condition initiale vers une unique probabilité invariante, décrivant la situation à l'équilibre : toute mémorisation est alors impossible. Il faut donc considérer la limite thermodynamique de ce système, i.e. le réseau indéfini par \mathbf{N} (ensemble des entiers naturels), et nous présentons le formalisme des états de Gibbs introduit notamment par R.Minlos et D.Ruelle. Nous citons alors les résultats dûs à A.Bovier, V.Gaynard et P.Picco [10] concernant l'existence des mesures extrémales pour $\frac{p}{N} \rightarrow 0$, où p est le nombre d'images et N la taille du réseau.

Le chapitre 6 est consacré à l'étude asymptotique de l'énergie libre associée au modèle de Hopfield : pour $\frac{p}{N} \rightarrow 0$, cette variable aléatoire converge presque sûrement et dans \mathbf{L}^2 vers une constante : la limite thermodynamique de l'énergie libre associée au modèle de Curie-Weiss, ainsi que son espérance ; et elle vérifie la propriété de self-averaging pour $\frac{p}{N} \rightarrow \alpha \geq 0$, i.e. sa variance converge vers 0. Pour obtenir ces résultats, nous redémontrons précisément des résultats dûs à H.Koch, issus d'un article dont les démonstrations présentent quelques lacunes : nous montrons notamment par une représentation en graphes une majoration intéressante concernant l'espérance de la trace des puissances de la matrice (aléatoire) de recouvrement.

Enfin, au chapitre 7, nous concluons cette thèse en traitant des questions ouvertes et de certaines extensions possibles du modèle de Hopfield.

Chapitre 2

Introduction au calcul neuronal

2.1 Description des réseaux formels de neurones

La première modélisation mathématique d'un neurone est due à Mac Culloch et Pitts, en 1943[32] : un neurone est une variable binaire $n_i(t)$ pouvant prendre les deux valeurs 0, ou 1 à chaque instant t . Et à l'instant $(t + 1)$, le neurone est défini par :

$$n_i(t + 1) = \Theta \left(\sum_{j, j \neq i} w_{ij} n_j(t) - \mu_i \right)$$

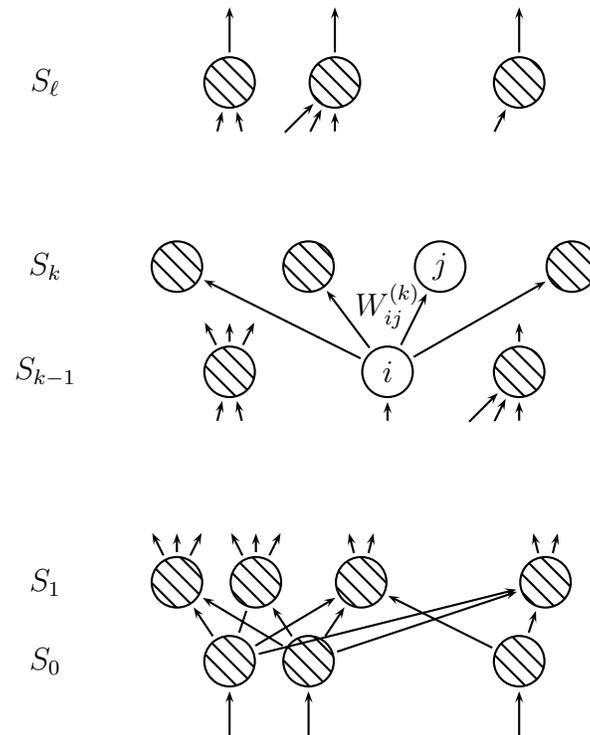
où $\Theta(x) = 1$, si $x > 0$, et 0 sinon.

μ_i est le seuil d'activation du neurone,

w_{ij} représente l'efficacité synaptique reliant le neurone j au neurone i , et peut être positif ou négatif, selon qu'il s'agisse d'une synapse excitatrice ou inhibitrice.

La généralisation de cette idée simple mais performante a conduit notamment à l'élaboration du Perceptron, puis des réseaux multicouches. Ces réseaux sont caractérisés par une relation de type entrée-sortie : il existe une couche S_0 de neurones recevant un stimulus, et une couche S_l "formulant" la réponse associée, calculée par le réseau.

Exemple de réseau multicouches avec $(l - 1)$ couches cachées :



Ces réseaux permettent par exemple de calculer les fonctions booléennes, et possèdent la capacité de généralisation : après une période d'apprentissage supervisée sur des exemples connus, durant laquelle les poids synaptiques sont ajustés pour obtenir un couple entrée-sortie correct, le réseau fournit une réponse exacte pour de nouvelles données. Un exemple d'application intéressant est la synthèse vocale : des réseaux sont capables de parler correctement une langue après s'être exercés sur une liste variée de mots. L'objet de cette thèse étant l'étude des modèles de mémoire associative, nous renvoyons le lecteur à la liste des ouvrages déjà cités pour une description mathématique des réseaux en couches [4, 37, 25, 8].

Pour le modèle de mémoire associative de Hopfield, il n'est plus nécessaire de considérer de tels réseaux. Tous les neurones jouent un rôle identique : le stimulus d'entrée et la réponse en sortie sont portés par tout le réseau, à des instants différents. Aussi ce modèle peut-il être étudié comme un système de particules en interaction.

Dans le cas de systèmes de particules sur un graphe, si l'interaction entre deux sites dépend de la distance les séparant, le comportement varie selon la structure géométrique (dimension, nombre de voisins). Dans le cas du graphe complet, correspondant au réseau neuronal étudié, la structure géométrique ne joue aucun rôle. Puisque nous supposons que chaque neurone interagit avec tout autre, leur disposition spatiale est totalement arbitraire, du moins pour l'étude théorique.

Nous allons maintenant définir plus précisément les objets nécessaires à l'étude mathématique du phénomène de mémoire associative.

Le *réseau neuronal* est un graphe simple orienté $G = (S, J)$, dont les sommets i de S sont appelés *neurones*, et les arêtes (i, j) *synapses*.

Les neurones sont numérotés de 1 à N , où $N = \text{card}(S)$.

L'*espace des états* d'un neurone est $\{-1, +1\}$.

La valeur du neurone i , noté x_i , est aussi appelé *spin* au *site* i .

L'ensemble $X_N = \{\underline{x} : \{1, \dots, N\} \longrightarrow \{-1, +1\}\}$ est *l'espace des configurations*.

Nous munissons X_N de la *distance de Hamming*, définie par :

$$d_H(\underline{x}, \underline{y}) = \frac{1}{4} \sum_{i=1}^N (x_i - y_i)^2, \text{ pour } \underline{x}, \underline{y} \in X_N.$$

Cette distance évalue le nombre de sites, où \underline{x} et \underline{y} diffèrent.

À chaque synapse (i, j) est attribué un nombre réel : *l'efficacité synaptique*, notée J_{ij} .

Dans la pratique, i.e. en informatique, un état de l'espace des configurations peut par exemple être assimilé à une image en noir et blanc, codée sous la forme de N bits valant ± 1 . C'est uniquement à ce niveau qu'intervient la disposition géométrique des neurones.

2.2 Description mathématique de la mémoire associative

Nous souhaitons conférer au réseau la capacité de stocker de l'information, et de la restituer, à partir d'une connaissance partielle de cette information.

L'information à stocker se présente sous la forme de p vecteurs binaires $(\underline{\xi}^\mu)^{\mu=1, \dots, p}$, de taille N , i.e. p configurations possibles du réseau : $\underline{\xi}^\mu = (\xi_i^\mu)_{i=1, \dots, N}$, avec $\xi_i^\mu \in \{-1, +1\}$.

Les configurations $\underline{\xi}^\mu$, $\mu = 1, \dots, p$, sont appelées *images originales*.

$$\begin{aligned} \text{Notation : } \underline{\xi}_i &= (\xi_i^\mu)^{\mu=1, \dots, p}, \text{ pour } i \in \{1, \dots, N\} \\ \underline{\xi}^\mu &= (\xi_i^\mu)_{i=1, \dots, N}, \text{ pour } \mu \in \{1, \dots, p\} \\ \underline{\xi} &= (\underline{\xi}^\mu)^{\mu=1, \dots, p} \end{aligned}$$

Notons que p peut dépendre de N .

Nous considérons des réseaux de grande taille, et le cas où p est fixé, quand N augmente, n'est pas vraiment intéressant d'un point de vue pratique.

Supposons désormais que les spins $(\xi_i^\mu)_{i=1, \dots, N}^{\mu=1, \dots, p}$ sont pN variables aléatoires indépendamment identiquement distribuées relativement à un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$, avec $\mathbf{P}[\xi_i^\mu = 1] =$

$$\mathbf{P} [\xi_i^\mu = -1] = 1/2.$$

Cela se justifie par le fait que les images mémorisées doivent être 2 à 2 suffisamment distinctes, pour pouvoir être différenciées.

Une méthode simple de mémoire associative consiste à stocker les p vecteurs dans des “mémoires”. Lorsqu’on présente une nouvelle configuration \underline{x} à l’ordinateur, il calcule les p valeurs $d_H(\underline{\xi}^\mu, \underline{x})$, pour $\mu = 1, \dots, p$. L’image $\underline{\xi}^\mu$ minimisant la distance de Hamming est l’image associée à \underline{x} . Cette procédure peut prendre un temps très long, pour un réseau de grande taille stockant beaucoup d’images; cependant dans l’article [36], B.Montgomery et V.Vijaya Kumar construisent un algorithme simple accomplissant cette tâche (“*Direct Storage Nearest-Neighbor Algorithm*”). Soit A la matrice à p lignes et N colonnes, telle que

$$A^T = [\underline{\xi}^1, \dots, \underline{\xi}^p],$$

où $\underline{\xi}^\mu$ est le vecteur colonne $(\xi_1^\mu, \dots, \xi_N^\mu)^T$. Soit $\underline{x} = (x_1, \dots, x_N)^T$. Pour trouver l’indice ν tel que $\underline{\xi}^\nu$ minimise $\{d_H(\underline{\xi}^\mu, \underline{x})\}_{\mu=1, \dots, p}$, on calcule

$$\underline{z} = A \underline{x}.$$

Alors

$$z_\mu = (\underline{\xi}^\mu)^T \underline{x} = a_\mu - d_\mu,$$

$$\text{où} \begin{cases} a_\mu \text{ est le nombre de composantes où } \underline{\xi}^\mu \text{ et } \underline{x} \text{ coïncident,} \\ d_\mu = d_H(\underline{\xi}^\mu, \underline{x}) = \frac{1}{2}(N - z_\mu). \end{cases}$$

Donc ν est l’indice tel que z_ν soit la plus grande composante de \underline{z} . Cet algorithme est sûr, contrairement au modèle de Hopfield, pour lequel il est facile de trouver des contrexemples, tels que la mémoire associative ne fonctionne pas (même si le nombre d’images à mémoriser est petit devant la taille du réseau : trois vecteurs de taille N arbitrairement grande [36]). Cela justifie d’ailleurs que l’on choisisse des v.a. indépendantes (condition d’orthogonalité), et que pour le modèle de Hopfield, la majorité des propriétés n’est vérifiée que pour un ensemble $\tilde{\Omega} \subset \Omega$, de probabilité asymptotiquement proche de 1. Cependant, comme le soulignent B.Montgomery et V.Vijaya Kumar, l’algorithme DSNN n’est pas un modèle de mémoire associative biologiquement valide : ils ne se soucient absolument pas de l’aspect de la modélisation neuronale, à l’origine du concept de mémoire associative. En ce sens, le modèle de Hopfield est beaucoup plus riche. Il est, par exemple, plus tolérant aux fautes.

Nous allons maintenant développer l’approche due à Hopfield [26] : l’expression “retrouver de l’information” est assimilée à une *dynamique d’évolution* sur l’espace des configurations, définie par une application $\mathcal{U} : X_N \longrightarrow X_N$. On souhaite que la dynamique conduise à $\underline{\xi}^\mu$, et s’y arrête, pour toute configuration initiale $\underline{x}(t=0)$ suffisamment proche de $\underline{\xi}^\mu$.

Pour une évolution à temps discret, l'état du réseau à l'instant $(t + 1)$, noté $\underline{x}(t + 1) = (x_1(t + 1), \dots, x_N(t + 1))$, est déterminé par

$$\underline{x}(t + 1) = \mathcal{U}(\underline{x}(t)), \quad t \in \mathbf{N}.$$

L'application \mathcal{U} induit ainsi une évolution temporelle *déterministe* de la configuration.

Au chapitre V, nous étudions une évolution *stochastique* générée par une chaîne de Markov sur X_N . L'aléa se situe alors à deux niveaux : les variables ξ_i^μ , ainsi que la dynamique, sont aléatoires relativement à l'espace $(\Omega, \mathcal{F}, \mathbf{P})$.

Définition 2.2.1 Une configuration \underline{y} de X_N est *stable* pour la dynamique induite par \mathcal{U} , si \underline{y} est un point fixe de \mathcal{U} , i.e. $\mathcal{U}(\underline{y}) = \underline{y}$.

Les points stables peuvent être interprétés comme des images stockées dans le réseau, et le processus d'attraction des vecteurs proches d'un point stable vers ce point comme un mécanisme de recouvrement d'information. Bien sûr, le cas le plus enviable est celui où les images originales elles mêmes constituent des points stables. Nous verrons aux chapitres III et IV que cette propriété peut être vérifiée, avec une probabilité arbitrairement proche de 1, pour N suffisamment grand, si leur nombre est borné par $\frac{N}{(4 \log N)}$.

On dit alors que les images ξ^μ , $\mu = 1, \dots, p$ sont *mémorisées sans erreur*.

Remarque : il nous arrivera parfois d'écrire "la propriété \mathcal{P} est vérifiée", par exemple, au lieu de "la propriété \mathcal{P} est vérifiée avec une probabilité arbitrairement proche de 1, pour N suffisamment grand", cela afin de rendre plus agréable la lecture de ce manuscrit. Il n'y a cependant aucun risque de confusion, ces résultats étant énoncés dans les propositions.

Si p devient plus important, en particulier $p = \alpha N$, où α est une constante, il n'est plus possible de retrouver exactement les images originales : on se rapproche du seuil de saturation du réseau. Toutefois, pour des valeurs raisonnables de α , il est possible de trouver des points stables au voisinage des images originales : il s'agit alors de *mémorisation avec un faible pourcentage d'erreurs tolérées* (Voir Chapitres III et IV).

La stabilité et la mémorisation partielle des images originales reposent entièrement et uniquement sur la définition de l'application $\mathcal{U} \equiv \mathcal{U}(\underline{\xi})$. En quelque sorte, c'est l'application $\mathcal{U}(\underline{\xi})$ qui "contient", et "rend" l'information manquante ou oubliée.

Nous allons maintenant décrire deux transformations : \mathcal{U}_S , et \mathcal{U}_P , induisant deux dynamiques d'évolution distinctes :

- * la dynamique séquentielle (de Hopfield),
- * la dynamique parallèle (de Little).

Chapitre 3

Étude de la dynamique séquentielle déterministe de Hopfield

3.1 Définitions

Les dynamiques séquentielles sont caractérisées par le fait qu'à un instant t donné, en un seul site, noté $b(t)$, la valeur du spin peut être modifiée.

Soit $T_N = \{t_0, t_1, \dots\}$ l'ensemble (supposé discret) des instants auxquels un spin peut être modifié.

$b : T_N \longrightarrow \{1, \dots, N\}$ est l' *application de balayage* du réseau, et on peut choisir :

- soit un *balayage séquentiel* : on pose simplement

$$b(t_0) = 1$$

$$b(t_i) = \left(b(t_{i-1}) \bmod N \right) + 1,$$

- soit un *balayage aléatoire* : $b(t)$ est choisi dans l'ensemble $\{1, \dots, N\}$, indépendamment de t , et des $b(u)$, $u < t$, selon la loi uniforme.

Nous considérons désormais que le balayage est séquentiel.

Nous définissons la matrice J des *efficacités synaptiques* entre neurones : pour $i, j \in \{1, \dots, N\}$,

$$\begin{cases} J_{ij} = \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu, \text{ si } i \neq j \\ J_{ii} = 0 \end{cases}$$

Remarques : - Cette matrice est symétrique.

- Dans la littérature, il est souvent fait référence aux *règles de Hebb* au sujet de cette matrice : si $\xi_i^\mu = \xi_j^\mu = 1$, alors les deux neurones sont actifs, et l'efficacité est renforcée de +1. Si un seul neurone est inactif (-1), l'efficacité est diminuée. Toutefois, l'analogie avec le mécanisme décrit par Hebb pour les neurones biologiques n'est que partielle : si les deux neurones sont inactifs, ($\xi_i^\mu = \xi_j^\mu = -1$), l'efficacité est augmentée de +1.

Convention : $\text{sgn}(x) = 1$, si $x \geq 0$, et -1 sinon.

La transformation $\mathcal{U}_S : X_N \longrightarrow X_N$, associée à la dynamique séquentielle déterministe, dépend de $\underline{\xi}$, et de la fonction de balayage.

Posons, pour $i \in \{1, \dots, N\}$,

$$\begin{aligned} \mathcal{U}_S^i(\underline{\xi}, \cdot) : X_N &\longrightarrow X_N \\ \underline{x} &\longmapsto \mathcal{U}_S^i(\underline{\xi}, \underline{x}) = (x'_1, \dots, x'_N), \text{ avec } \begin{cases} x'_j = x_j, \text{ si } j \neq i \\ x'_i = \text{sgn}\left(\sum_{j=1}^N J_{ij}x_j\right) \end{cases} \end{aligned}$$

Alors $\mathcal{U}_S(\underline{\xi}, \cdot)$ est défini par : $\mathcal{U}_S(\underline{\xi}, \underline{x}) = \mathcal{U}_S^N \circ \dots \circ \mathcal{U}_S^1(\underline{\xi}, \underline{x})$.

Partant d'une condition initiale $\underline{x}(t=0)$, la trajectoire induite par \mathcal{U}_S dans l'espace des configurations est alors définie par :

$$\underline{x}(t+1) = \mathcal{U}_S(\underline{\xi}, \underline{x}(t)).$$

Par analogie avec les systèmes physiques dont les états stables sont les minima de l'énergie, nous allons définir le *hamiltonien*, ou *fonction énergie*, associée au modèle séquentiel de Hopfield, telle que ses minima correspondent aux états d'équilibre.

$$\begin{aligned} H_{N,p}^S : \Omega \times X_N &\longrightarrow \mathbf{R} \\ (\omega, \underline{x}) &\longmapsto H_{N,p}^S(\underline{\xi}(\omega), \underline{x}) = -\frac{1}{N} \sum_{i,j=1}^N J_{ij}x_i x_j \\ &= -\frac{1}{N} \sum_{\substack{i,j=1 \\ i \neq j}}^N \sum_{\mu=1}^p \xi_i^\mu(\omega) \xi_j^\mu(\omega) x_i x_j \end{aligned}$$

$H_{N,p}^S(\underline{\xi}, \cdot)$ est une application sur X_N , dépendant des paramètres aléatoires $(\xi_i^\mu)_{i=1, \dots, N}^{\mu=1, \dots, p}$.

Propriété : la valeur du hamiltonien $H_{N,p}^S(\underline{\xi}, \cdot)$ décroît sur toute trajectoire induite par \mathcal{U}_S dans l'espace des configurations X_N .

En effet, pour tout $i \in \{1, \dots, N\}$,

$$\begin{aligned} H_{N,p}^S(\underline{\xi}, \mathcal{U}_S^i(\underline{x})) - H_{N,p}^S(\underline{\xi}, \underline{x}) &= -\frac{1}{N} \sum_{j=1}^N J_{ij} x_j \left[\operatorname{sgn}\left(\sum_{k=1}^N J_{ik} x_k\right) - x_i \right] \\ &= -\frac{1}{N} \left| \sum_{j=1}^N J_{ij} x_j \right| \left[1 - x_i \operatorname{sgn}\left(\sum_{k=1}^N J_{ik} x_k\right) \right] \leq 0. \end{aligned}$$

Proposition 3.1.1 *Pour la dynamique séquentielle déterministe de Hopfield,*

- un état stable est atteint pour toute condition initiale $\underline{x}(t=0)$.
- tout état stable est minimum local du hamiltonien $H_{N,p}^S(\underline{\xi}, \cdot)$.

Démonstration :

$H_{N,p}^S(\underline{\xi}, \cdot)$ est borné inférieurement : $\inf_{\underline{x} \in X_N} H_{N,p}^S(\underline{\xi}, \underline{x}) \geq -p(N-1)$,

et le nombre d'états est fini ($\operatorname{card}(X_N) = 2^N$), donc la propriété précédente implique qu'un minimum au moins local de $H_{N,p}^S(\underline{\xi}, \cdot)$ est atteint sur toute trajectoire.

Ce minimum correspond à un état stable si pour tout $i \in \{1, \dots, N\}$,

$$x_i(t) = \operatorname{sgn}\left(\sum_{j=1}^N J_{ij} x_j(t)\right)$$

À priori, cela n'est pas nécessairement le cas. On peut avoir $H_{N,p}^S(\underline{\xi}, \mathcal{U}_S^i(\underline{x})) - H_{N,p}^S(\underline{\xi}, \underline{x}) = 0$, avec $\sum_{j=1}^N J_{ij} x_j = 0$, et $1 - x_i \operatorname{sgn}\left(\sum_{j=1}^N J_{ij} x_j\right) \neq 0$.

Mais cela signifie : $x_i(t) = -1$, $x_i(t+1) = 1$, donc cette situation ne peut arriver que si le système est dans un état transitoire. Après un nombre fini de tels changements de -1 à $+1$ (au plus N), le système n'évolue plus, et $x_i(t) = \operatorname{sgn}\left(\sum_{j=1}^N J_{ij} x_j(t)\right)$, pour tout i . Finalement, un état stable est toujours atteint.

Réciproquement, tout point stable est minimum local de la fonction énergie. En effet, par définition, une configuration \underline{y} de X_N est minimum local de $H_{N,p}^S(\underline{\xi}, \cdot)$ si toute configuration \underline{y}' obtenue en modifiant une coordonnée de \underline{y} vérifie $H_{N,p}^S(\underline{\xi}, \underline{y}') \geq H_{N,p}^S(\underline{\xi}, \underline{y})$.

Soit i tel que $y'_i = -y_i$, et $y'_j = y_j$ pour $j \in \{1, \dots, N\} \setminus i$.

$$\text{Alors } H_{N,p}^S(\underline{\xi}, \underline{y}') - H_{N,p}^S(\underline{\xi}, \underline{y}) = \frac{2}{N} y_i \sum_{j=1}^N J_{ij} y_j.$$

Donc \underline{y} est minimum local si $y_i = \text{sgn}\left(\sum_{j=1}^N J_{ij}y_j\right)$, pour tout $i \in \{1, \dots, N\}$. Cette condition est vérifiée par les points stables. ♣

Remarques : 1. Pour $p = 1$, il est clair que $\underline{\xi}^1$ et $-\underline{\xi}^1 = (-\xi_1^1, \dots, -\xi_N^1)$ sont les seuls minima globaux de $H_{N,p}^S(\underline{\xi}, \cdot)$, et

$$H_{N,p}^S(\underline{\xi}, \underline{\xi}^1) = H_{N,p}^S(\underline{\xi}, -\underline{\xi}^1) = -N.$$

En outre, dès que $p \geq 2$, la situation n'est plus du tout triviale.

2. La remarque 1, ainsi que la proposition 3.1.1, justifient l'étude des fluctuations de $H_{N,p}^S(\underline{\xi}, \cdot)$, menée au paragraphe III.3.

3.2 Étude asymptotique de la stabilité des images originales et de la convergence directe

Les configurations *stables* pour \mathcal{U}_S sont les configurations \underline{y} de X_N vérifiant $\mathcal{U}_S(\underline{y}) = \underline{y}$.

Nous dirons qu'une configuration \underline{y} *converge directement* vers une image $\underline{\xi}^\mu$, pour la dynamique séquentielle, si $\underline{x}(t=0) = \underline{y} \implies \underline{x}(t=1) \stackrel{\text{déf}}{=} \mathcal{U}_S(\underline{y}) = \underline{\xi}^\mu$.

3.2.1 Stabilité des images originales

Théorème 3.2.1 Soit $p = \frac{N}{\gamma \log N}$.

(i) si $\gamma > 6$, alors pour $N \rightarrow \infty$, les p images originales sont p.s. stables, i.e.

$$\mathbf{P} \left[\liminf_N \left\{ \bigcap_{\nu=1}^p [\mathcal{U}_S(\underline{\xi}^\nu) = \underline{\xi}^\nu] \right\} \right] = 1 \quad (3.1)$$

(ii) si $\gamma > 4$,

$$\mathbf{P} \left[\bigcap_{\nu=1}^p \left\{ \mathcal{U}_S(\underline{\xi}^\nu) = \underline{\xi}^\nu \right\} \right] = 1 - R_N \quad (3.2)$$

avec $\lim_{N \rightarrow \infty} R_N = 0$.

Le (ii) de ce théorème, dont nous donnons une preuve simple, est due à Mac Eliece et al. (voir l'article [33], où il est présenté de manière un peu différente).

La méthode utilisée pour montrer le théorème 3.2.1 repose essentiellement sur une estimation de grande déviation classique, appelée parfois *inégalité exponentielle de Markov*. Soit X une variable aléatoire quelconque ; alors :

$$\mathbf{P}\left[X > \epsilon\right] \leq \inf_{t>0} e^{-t\epsilon} \mathbf{E}e^{tX} \quad (3.3)$$

Bien sûr, cette majoration ne présente un intérêt que si $\mathbf{E}e^{tX} < \infty$, pour un $t > 0$.

Notation : appelons π_k l'application coordonnée, i.e.

$$\begin{aligned} \pi_k : \quad \mathbf{R}^N &\longrightarrow \mathbf{R} \\ \underline{x} = (x_1, \dots, x_N) &\longmapsto x_k \end{aligned}$$

Démonstration du théorème 3.2.1 : Posons $\Lambda_N = \{1, \dots, N\}$. Soit $\nu \in \{1, \dots, p\}$. Si $\underline{x}(0) = \underline{\xi}^\nu$, alors $\mathcal{U}_S^i(\underline{x}(0))$ a pour i ème coordonnée :

$$\text{sgn}\left(\sum_{j \in \Lambda_N \setminus i} \xi_j^\nu (\xi_j^\nu)^2 + \sum_{\substack{\nu=1, \dots, p \\ \mu \neq \nu}} \xi_i^\mu \sum_{j \in \Lambda_N \setminus i} \xi_j^\mu \xi_j^\nu\right) = \xi_i^\nu \text{sgn}\left(N - 1 + \xi_i^\nu \sum_{\substack{\mu=1, \dots, p \\ \mu \neq \nu}} \xi_i^\mu \sum_{j \in \Lambda_N \setminus i} \xi_j^\mu \xi_j^\nu\right)$$

Donc l'image $\underline{\xi}^\nu$ est stable si pour $i \in \Lambda_N$, le terme entre parenthèses est positif.

$$\text{Soit } A(N, p, i, \nu) = \left\{ \omega \in \Omega : N - 1 + \xi_i^\nu(\omega) \sum_{\substack{\mu=1, \dots, p \\ \mu \neq \nu}} \xi_i^\mu(\omega) \sum_{j \in \Lambda_N \setminus i} \xi_j^\mu(\omega) \xi_j^\nu(\omega) \leq 0 \right\}.$$

Nous avons :

$$\begin{aligned} \mathbf{P}\left[\bigcup_{\nu=1}^p \bigcup_{i=1}^N A(N, p, i, \nu)\right] &\leq \sum_{\nu=1}^p \sum_{i=1}^N \mathbf{P}\left[A(N, p, i, \nu)\right] \\ &\leq \sum_{\nu=1}^p \sum_{i=1}^N \inf_{t>0} e^{-t(N-1)} \mathbf{E} \exp\left(-t \xi_i^\nu \sum_{\substack{\mu=1, \dots, p \\ \mu \neq \nu}} \xi_i^\mu \sum_{j \in \Lambda_N \setminus i} \xi_j^\mu \xi_j^\nu\right) \end{aligned}$$

Notons $\mathbf{E}_{\underline{\xi}^\mu}$, respectivement $\mathbf{E}_{\xi_j^\mu}$, l'espérance conditionnelle par rapport à la tribu engendrée par les v.a. $(\xi_k^\lambda)_{k=1, \dots, N}^{\lambda=1, \dots, p; \lambda \neq \mu}$, respectivement $\left\{ (\xi_k^\lambda)_{k=1, \dots, N}^{\lambda=1, \dots, p; \lambda \neq \mu}, (\xi_k^\mu)_{k=1, \dots, N; k \neq j} \right\}$.

Par indépendance des v.a. $(\xi_i^\mu)_{i=1, \dots, N}^{\mu=1, \dots, p}$,

$$\begin{aligned}
\mathbf{E} \exp\left(-t \xi_i^\nu \sum_{\substack{\mu=1, \dots, p \\ \mu \neq \nu}} \xi_i^\mu \sum_{j \in \Lambda_N \setminus i} \xi_j^\nu \xi_j^\mu\right) &= \mathbf{E}_{\xi^\nu} \prod_{\substack{\mu=1, \dots, p \\ \mu \neq \nu}} \mathbf{E}_{\xi^\mu} \exp\left(-t \xi_i^\mu \sum_{j \in \Lambda_N \setminus i} \xi_j^\mu \xi_j^\nu\right) \\
&= \mathbf{E}_{\xi^\nu} \prod_{\substack{\mu=1, \dots, p \\ \mu \neq \nu}} \mathbf{E}_{\xi_i^\mu} \prod_{j \in \Lambda_N \setminus i} \mathbf{E}_{\xi_j^\mu} \exp(-t \xi_j^\mu) \\
&= \mathbf{E}_{\xi^\nu} \prod_{\substack{\mu=1, \dots, p \\ \mu \neq \nu}} \mathbf{E}_{\xi_i^\mu} (\cosh t)^{(N-1)} \\
&= (\cosh t)^{(N-1)(p-1)}
\end{aligned}$$

On sait que $\cosh(t) \leq \exp(\frac{t^2}{2})$, donc

$$\mathbf{P}\left[\bigcup_{\nu=1}^p \bigcup_{i=1}^N A(N, p, i, \nu)\right] \leq pN \inf_{t>0} \exp\left((N-1)\left(-t + (p-1)\frac{t^2}{2}\right)\right) = pN e^{-\frac{N-1}{2(p-1)}}$$

Par le lemme de Borel-Cantelli, (i) est vrai s'il existe un $N_0 \in \mathbf{N}$ tel que la série $\sum_{N \geq N_0} \mathbf{P}\left[\bigcup_{\nu=1}^p \bigcup_{i=1}^N A(N, p, i, \nu)\right]$ soit convergente. Or la série $\sum_{N \geq 2} \frac{1}{N(\log N)^2}$ converge : il suffit donc de choisir γ tel que, pour $p = \frac{N}{\gamma \log N}$, il existe un $N_0 \in \mathbf{N}$, avec

$$-\frac{N-1}{2(p-1)} + \log N + \log p + \log(N(\log N)^2) < 0,$$

pour $N \geq N_0$, ce qui s'écrit : $(3 - \frac{\gamma}{2}) \log N + R_N < 0$, où $\lim_{N \rightarrow \infty} \frac{R_N}{\log N} = 0$. Il faut donc prendre $\gamma > 6$.

Pour obtenir (ii), qui est le résultat de Mac Eliece, il suffit de choisir γ tel que :

$$\lim_{N \rightarrow \infty} pN \exp\left(-\frac{N-1}{2(p-1)}\right) = 0.$$

$\gamma > 4$ convient. ♣

3.2.2 Convergence directe des images originales bruitées

Dans l'article [33], les auteurs s'intéressent également à la convergence directe (pour la dynamique *parallèle*) des images perturbées aléatoirement (i.e. des images originales contenant un certain nombre d'erreurs aléatoires). Une image perturbée, obtenue à partir

de $\underline{\xi}^\nu$, et contenant ρN erreurs (où $0 \leq \rho \leq 1$, tel que ρN soit entier), peut être considérée comme un point de la sphère centrée en $\underline{\xi}^\nu$ de rayon ρN (pour la distance de Hamming). Nous noterons cette sphère $\mathcal{S}(\underline{\xi}^\nu, \rho N)$.

Mac Eliece et al. prouvent que si $p < (1 - 2\rho)^2 \frac{N}{4 \log N}$, $\rho < \frac{1}{2}$, un vecteur à une distance ρN d'une image originale arbitraire converge directement pour la dynamique parallèle vers cette image, avec une forte probabilité.

Les mêmes arguments que ceux développés pour le théorème 3.2.1 nous permettent d'établir simplement un résultat analogue.

Theorème 3.2.2 *Soit $\rho \in [0, \frac{1}{2}[$, et $p = (1 - 2\rho)^2 \frac{N}{\gamma \log N}$. Pour $\nu = 1, \dots, p$, soit $\underline{x}^\nu(0)$ un vecteur situé sur la sphère de Hamming centré en $\underline{\xi}^\nu$, de rayon ρN , i.e. $d_H(\underline{\xi}^\nu, \underline{x}^\nu(0)) = \rho N$. Alors :*

(i) si $\gamma > 6$,

$$\mathbf{P} \left[\liminf_N \left\{ \bigcap_{\nu=1}^p [\mathcal{U}_S(\underline{x}^\nu(0)) = \underline{\xi}^\nu] \right\} \right] = 1 \quad (3.4)$$

(ii) si $\gamma > 4$,

$$\mathbf{P} \left[\bigcap_{\nu=1}^p \{ \mathcal{U}_S(\underline{x}^\nu(0)) = \underline{\xi}^\nu \} \right] = 1 - R_N \quad (3.5)$$

avec $\lim_{N \rightarrow \infty} R_N = 0$.

Lemme 3.2.3 *Sous les mêmes hypothèses que pour le théorème 3.2.2, pour tout $i \in \{1, \dots, N\}$,*

$$\mathbf{P} \left[\pi_i \circ \mathcal{U}_S^i(\underline{x}^\nu(0)) = -\xi_i^\nu \right] \leq e^{-\frac{(N(1-2\rho)-1)^2}{(p-1)(N-1)}} \equiv K_{N,p}(\rho) \quad (3.6)$$

$K_{N,p}(\rho)$ est une fonction croissante en ρ .

Démonstration du lemme 3.2.3 Il existe un sous-ensemble $\mathcal{A} \equiv \mathcal{A}(\underline{x}^\nu(0)) \subset \Lambda_N$, $\text{card}(\mathcal{A}) = \rho N$, et tel que

$$\begin{cases} x_i^\nu(0) = \xi_i^\nu, \text{ pour } i \in \Lambda \setminus \mathcal{A} \\ x_i^\nu(0) = -\xi_i^\nu, \text{ pour } i \in \mathcal{A} \end{cases}$$

Alors :

$$\begin{aligned}\pi_i \circ \mathcal{U}_S^i(\underline{x}^\nu(0)) &= \operatorname{sgn} \left[\sum_{j \in (\Lambda_N \setminus \mathcal{A}) \setminus i} J_{ij} \xi_i^\nu + \sum_{j \in \Lambda_N \setminus i} J_{ij} (-\xi_i^\nu) \right] \\ &= \xi_i^\nu \operatorname{sgn} \left[N(1-2\rho) - \epsilon_i + \xi_i^\nu \sum_{j \in \Lambda_N \setminus i} \sum_{\substack{\mu=1, \dots, p \\ \mu \neq \nu}} \xi_i^\mu \xi_j^\mu \epsilon_j \xi_j^\nu \right],\end{aligned}$$

$$\text{où } \epsilon_j = \begin{cases} 1, & \text{si } j \notin \mathcal{A} \\ -1, & \text{si } j \in \mathcal{A} \end{cases}$$

Soit

$$B(N, p, i, \nu, A) = \left\{ \omega \in \Omega : \xi_i^\nu(\omega) \sum_{\substack{j \in \Lambda_N \\ j \neq i}} \sum_{\substack{\mu=1, \dots, p \\ \mu \neq \nu}} \xi_i^\mu(\omega) \xi_j^\mu(\omega) \epsilon_j \xi_j^\nu(\omega) + N(1-2\rho) - \epsilon_i \geq 0 \right\}$$

En reprenant les calculs de la démonstration du théorème 3.2.1, nous obtenons :

$$\mathbf{P} \left[B(N, p, i, \nu, A)^C \right] \leq \inf_{t>0} e^{-t((1-2\rho)-\epsilon_i)+(p-1)(N-1)\frac{t^2}{2}} \leq K_{N,p}(\rho). \clubsuit$$

Démonstration du théorème 3.2.2 : Nous avons

$$\begin{aligned}\mathbf{P} \left[\bigcap_{\nu=1}^p \left\{ \mathcal{U}_S(\underline{x}^\nu(0)) = \underline{\xi}^\nu \right\} \right] &= \mathbf{P} \left[\bigcap_{\nu=1}^p \left\{ \mathcal{U}_S^N \circ \dots \circ \mathcal{U}_S^1(\underline{x}^\nu(0)) = \underline{\xi}^\nu \right\} \right] \\ &= \mathbf{P} \left[\bigcap_{\nu=1}^p \left\{ \pi_1 \circ \mathcal{U}_S^1(\underline{x}^\nu(0)) = \xi_1^\nu, \pi_2 \circ \mathcal{U}_S^2(\underline{z}_1^\nu) = \xi_2^\nu, \dots, \pi_N \circ \mathcal{U}_S^N(\underline{z}_{N-1}^\nu) = \xi_N^\nu \right\} \right],\end{aligned}$$

$$\text{où } \underline{z}_i^\nu = (\xi_1^\nu, \dots, \xi_i^\nu, x^\nu(0)_{i+1}, \dots, x^\nu(0)_N)$$

Donc

$$\mathbf{P} \left[\bigcup_{\nu=1}^p \left\{ \mathcal{U}_S(\underline{x}^\nu(0)) \neq \underline{\xi}^\nu \right\}^C \right] \leq \sum_{\nu=1}^p \sum_{i=1}^N \mathbf{P} \left[\pi_i \circ \mathcal{U}_S^i(\underline{z}_{i-1}^\nu) = -\xi_i^\nu \right] \leq pN K_{N,p}(\rho),$$

d'après le lemme précédent. Par le lemme de Borel-Cantelli, (i) est vérifiée si : $\exists N_0 > 0, \forall N \geq N_0, \log K_{N,p}(\rho) + \log N + \log p + \log(N(\log N)^2) < 0$.

Posons $p = \frac{N}{\tilde{\gamma} \log N}$. Il faut prendre $\tilde{\gamma} > \frac{1}{6(1-2\rho)^2}$.

(ii) est vérifiée si : $\lim_{N \rightarrow \infty} pN K_{N,p}(\rho) = 0$. $\tilde{\gamma} > \frac{1}{4(1-2\rho)^2}$ convient. \clubsuit

Pour obtenir une estimation concernant l'ensemble des vecteurs de $\bigcup_{\nu=1}^p \mathcal{S}(\underline{\xi}^\nu, \rho N)$, il nous faut écrire :

$$\mathbf{P} \left[\bigcup_{\substack{K \subset \Lambda_N \\ \text{card}(K) = \rho N}} \bigcup_{i=1}^N \bigcup_{\nu=1}^p A(N, p, i, \nu, K)^C \right] \leq C_N^{\rho N} p N \mathbf{P} \left[A(N, p, i, \nu, K)^C \right],$$

qui ne peut converger vers 0 avec la majoration obtenue pour $\mathbf{P} \left[A(N, p, i, \nu, K)^C \right]$. On peut donc se demander si les images originales constituent effectivement des attracteurs directs de **tous** les vecteurs de $\bigcup_{\nu=1}^p \mathcal{S}(\underline{\xi}^\nu, \rho N)$.

Comme le soulignent J.Komlos et R.Paturi [28], la réponse est négative :

- il est possible de choisir certains états, contenant seulement $\mathcal{O}(\sqrt{N})$ erreurs, qui ne convergent pas directement vers l'image originale. Il faut donc distinguer les erreurs aléatoires des erreurs volontairement défavorables !

- Montgomery et Vijaya Kumar [36] ont observé qu'il est impossible d'avoir un rayon de convergence proche de $\frac{1}{2}$: $\rho > \frac{1}{8}$ est déjà impossible, même pour p petit.

3.2.3 Convergence directe des “mosaïques aléatoires”

Nous allons maintenant nous intéresser à d'autres types de configurations présentant la propriété de convergence directe avec une forte probabilité. Les mosaïques aléatoires constituées de fragments de diverses images originales convergent vers une image originale, si celle-ci est prédominante. Plus précisément :

Theorème 3.2.4 Soient $(\xi_i^{p+1})_{i=1, \dots, N}$ N v.a.i.i.d., indépendantes des v.a. $(\xi_i^\mu)_{i=1, \dots, N}^{\mu=1, \dots, p}$, et de même loi que ξ_1^1 . (ξ_1^{p+1} joue le rôle d'un “bruit” étranger aux images originales).

Soient $n_1, \dots, n_{p+1} \in \mathbf{N}$, avec $0 = n_0 < n_1 \leq n_2 \leq \dots \leq n_p \leq n_{p+1} = N$, et $(\mathbf{I}_\lambda)_{\lambda=1, \dots, p+1}$ une partition de Λ_N avec, pour $\lambda \in \{1, \dots, p+1\}$, $\text{card}(I_\lambda) = n_\lambda - n_{\lambda-1}$. Pour $\nu \in \{1, \dots, p\}$, soit $\underline{x}^\nu(0)$ la configuration définie par :

$$\begin{cases} x_i^\nu(0) = \xi_i^\nu, & \text{pour } i \in \mathbf{I}_1 \\ x_i^\nu(0) = \xi_i^{a^\nu(\mu)}, & \text{pour } i \in \mathbf{I}_\mu \ (\mu = 2, \dots, p) \\ x_i^\nu(0) = \xi_i^{p+1}, & \text{pour } i \in \mathbf{I}_{p+1}, \end{cases}$$

où a^ν est la transposition de $\{1, \dots, p\}$ telle que $a^\nu(1) = \nu$, $a^\nu(\nu) = 1$.

Posons $p = \frac{N}{\gamma \log N}$, $n_1 = \gamma_1 N$, $n_p = \gamma_p N$, où $\gamma > 0$, $0 < \gamma_1 \leq \gamma_p \leq 1$.

Soit $\epsilon > 0$ arbitrairement petit,

(i) Si $\gamma_1 > \frac{1}{2}(\gamma_p + \epsilon)$, et $\gamma > \frac{6}{(2\gamma_1 - \gamma_p - \epsilon)^2}$, i.e. $\gamma_1 > \frac{(\gamma_p + \epsilon)}{2} + \frac{1}{2}\sqrt{\frac{6}{\gamma}}$, alors

$$\mathbf{P} \left[\liminf_N \left\{ \bigcap_{\nu=1}^p [\mathcal{U}_S(\underline{x}^\nu(0)) = \underline{\xi}^\nu] \right\} \right] = 1 \quad (3.7)$$

(ii) Si $\gamma_1 > \frac{1}{2}(\gamma_p + \epsilon)$, et $\gamma > \frac{4}{(2\gamma_1 - \gamma_p - \epsilon)^2}$, i.e. $\gamma_1 > \frac{(\gamma_p + \epsilon)}{2} + \frac{1}{2}\sqrt{\frac{1}{\gamma}}$, alors

$$\mathbf{P} \left[\bigcap_{\nu=1}^p [\mathcal{U}_S(\underline{x}^\nu(0)) = \underline{\xi}^\nu] \right] = 1 - R_N \quad (3.8)$$

où $\lim_{N \rightarrow \infty} R_N = 0$.

Remarques : 1. Il suffit de montrer le théorème pour la “mosaïque ordonnée” (voir lemme 3.2.5). En effet, le choix des indices, et l’ordre dans lequel apparaissent les différentes composantes de chaque $\underline{\xi}^\mu$ ne jouent aucun rôle. Seules importent les valeurs n_1, \dots, n_p , ou plus exactement le nombre de fois où apparaît chaque image $\underline{\xi}^\mu$ (par une de ses coordonnées) dans le vecteur $\underline{x}(0)$.

2. De même que pour le théorème 3.2.2, les estimations ne sont pas suffisantes pour obtenir un résultat analogue valide uniformément pour tous les choix possibles de telles configurations.

3. Ce théorème est en quelque sorte une extension du théorème 3.2.1, puisque le cas $n_1 = n_2 = \dots = n_p = N$ correspond à la stabilité des images originales $\underline{\xi}^\nu, \nu = 1, \dots, p$.

Lemme 3.2.5

Soit $\underline{x}(0)$ tel que
$$\begin{cases} x_i(0) = \xi_i^1, & \text{pour } 1 \leq i \leq n_1 \\ x_i(0) = \xi_i^\mu, & \text{pour } n_{\mu-1} < i \leq n_\mu \ (\mu = 2, \dots, p) \\ x_i(0) = \xi_i^{p+1}, & \text{pour } n_p < i \leq N \end{cases}$$

Soit

$$A(N, n \equiv (n_1, \dots, n_p), i, p) = \{\omega \in \Omega : \pi_i \circ \mathcal{U}_S^i(\underline{\xi}(\omega), \underline{x}(0)) = \xi_i^1\}.$$

Alors

$$\mathbf{P} \left[A(N, n, i, p)^C \right] \leq \exp \left(-\frac{1}{2} \frac{(2n_1 - n_p - 1)^2}{p(N-1) - n_p + 2} \right) \equiv K_{N,p}(n), \text{ si } n_1 > \frac{1}{2}(n_p + 1)$$

$K_{N,p}(n)$ est décroissante en la variable n_1 .

Démonstration : Nous avons

$$\begin{aligned}\pi_i \circ \mathcal{U}_S^i(\underline{\xi}, \underline{x}(0)) &= \operatorname{sgn}\left(\sum_{\mu=1}^p \xi_i^\mu \sum_{j \in \Lambda_N \setminus i} \xi_j^\mu x_j(0)\right) \\ &= \xi_i^1 \operatorname{sgn}\left(n_1 - \epsilon_i + \sum_{\substack{j > n_1 \\ j \neq i}} \xi_j^1 x_j(0) + \xi_i^1 \sum_{\mu=2}^p \xi_i^\mu \sum_{j \in \Lambda_N \setminus i} \xi_j^\mu x_j(0)\right),\end{aligned}$$

$$\text{avec } \epsilon_i = \begin{cases} 1, & \text{si } i \leq n_1 \\ 0, & \text{sinon.} \end{cases}$$

D'où :

$$\mathbf{P}\left[A(N, n, i, p)^C\right] \leq \inf_{t>0} e^{-t(n_1 - \epsilon_i)} \mathbf{E} \exp t \left\{ \sum_{\substack{j > n_1 \\ j \neq i}} \xi_j^1 x_j(0) + \xi_i^1 \sum_{\mu=2}^p \xi_i^\mu \sum_{j \in \Lambda_N \setminus i} \xi_j^\mu x_j(0) \right\}.$$

L'indépendance des v.a. $(\xi_i^\mu)_{i=1, \dots, N}^{\mu=1, \dots, p+1}$ permet de calculer cette espérance :

Supposons que $i \leq n_1$. Les variables $(\xi_j^1)_{j > n_1}$ figurent uniquement dans le premier terme, qui devient :

$$\begin{aligned}\mathbf{E} \exp\left(-t \sum_{j > n_1} \xi_j^1 x_j(0)\right) &= \prod_{j > n_1} \cosh(tx_j(0)), \quad \text{car } x_j(0) \neq \xi_j^1 \\ &= (\cosh t)^{(N - n_1)}\end{aligned}$$

Il reste à intégrer le second terme. Nous adoptons la notation suivante : $\mathbf{E}_{\underline{\xi}_j}$ désigne l'espérance conditionnelle par rapport à la tribu engendrée par les v.a. $(\xi_i^\mu)_{i=1, \dots, N; i \neq j}^{\mu=1, \dots, p}$.

$$\begin{aligned}& \mathbf{E} \exp\left\{-t \xi_i^1 \sum_{\mu=2}^p \xi_i^\mu \left(\sum_{\substack{j \leq n_1 \\ j \neq i}} \xi_j^\mu \xi_j^1 + \sum_{j > n_1} \xi_j^\mu x_j(0)\right)\right\} \\ &= (\cosh t)^{(p-1)(n_1-1)} \mathbf{E} \exp\left\{-t \xi_i^1 \sum_{\mu=2}^p \xi_i^\mu \sum_{j > n_1} \xi_j^\mu x_j(0)\right\} \\ &= (\cosh t)^{(p-1)(n_1-1)} \mathbf{E} \left\{ \prod_{n_1 < j \leq n_2} \mathbf{E}_{\underline{\xi}_j} \exp(-t \xi_i^1 \xi_j^2 \sum_{\mu=2}^p \xi_i^\mu \xi_j^\mu) \dots \right. \\ & \quad \left. \dots \prod_{n_p < j \leq N} \mathbf{E}_{\underline{\xi}_j} \exp(-t \xi_i^1 \xi_j^{p+1} \sum_{\mu=2}^p \xi_i^\mu \xi_j^\mu) \right\}\end{aligned}$$

$$\begin{aligned}
 \text{Pour } \nu = 2, \dots, p, \quad \mathbf{E}_{\xi_j} \exp(-t\xi_i^1 \xi_j^\nu \sum_{\mu=2}^p \xi_i^\mu \xi_j^\mu) &= \exp(-t\xi_i^1 \xi_i^\nu) \mathbf{E}_{\xi_j} \exp(-t\xi_i^1 \xi_j^\nu \sum_{\substack{\mu=2 \\ \mu \neq \nu}}^p \xi_i^\mu \xi_j^\mu) \\
 &= \exp(-t\xi_i^1 \xi_i^\nu) (\cosh t)^{p-2}. \\
 \text{Pour } \nu = p+1, \quad \mathbf{E}_{\xi_j} \exp\left(-t\xi_i^1 \xi_j^{p+1} \sum_{\mu=2}^p \xi_i^\mu \xi_j^\mu\right) &= (\cosh t)^{p-1},
 \end{aligned}$$

donc le second terme s'écrit :

$$\begin{aligned}
 &(\cosh t)^{(p-1)(n_1-1)} (\cosh t)^{(p-2)(n_p-n_1)} (\cosh t)^{(p-1)(N-n_p)} \\
 &\quad \times \mathbf{E} \exp\left(-t\xi_i^1 \left((n_2 - n_1)\xi_i^2 + (n_3 - n_2)\xi_i^3 + \dots + (n_p - n_{p-1})\xi_i^p\right)\right) \\
 &= (\cosh t)^{N(p-1)-n_p-p+n_1+1} \prod_{j=2}^p \cosh[t(n_j - n_{j-1})]
 \end{aligned}$$

En utilisant les majorations suivantes :

$$\prod_{j=2}^p \cosh[t(n_j - n_{j-1})] \leq \exp\left(t \sum_{j=2}^p (n_j - n_{j-1})\right) = \exp t(n_p - n_1)$$

et

$$(\cosh t)^{p(N-1)-n_p+1} \leq \exp\left\{(p(N-1) - n_p + 1) \frac{t^2}{2}\right\},$$

on obtient :

$$\begin{aligned}
 \mathbf{P}\left[A(N, n, i, p)^C\right] &\leq \inf_{t>0} \exp\left\{-t(2n_1 - n_p + 1) + (p(N-1) - n_p + 1) \frac{t^2}{2}\right\} \\
 &= \exp\left\{-\frac{1}{2} \frac{(2n_1 - n_p + 1)^2}{p(N-1) - n_p + 1}\right\}, \text{ si } n_1 > \frac{1}{2}(n_p - 1).
 \end{aligned}$$

Si maintenant on suppose que $i > n_1$, on obtient de même :

$$\mathbf{P}\left[A(N, n, i, p)^C\right] \leq \exp\left\{-\frac{1}{2} \frac{(2n_1 - n_p - 1)^2}{p(N-1) - n_p + 2}\right\}, \text{ si } n_1 > \frac{1}{2}(n_p + 1). \clubsuit$$

Démonstration du théorème 3.2.4 : Soit $\underline{x}(0)$ comme dans le lemme 3.2.5.

$$\begin{aligned} & \mathbf{P}\left[\mathcal{U}_S(\underline{x}(0)) = \underline{\xi}^1\right] \\ &= \mathbf{P}\left[\pi_1 \circ \mathcal{U}_S^1(\underline{x}(0)) = \xi_1^1, \pi_2 \circ \mathcal{U}_S^2(\underline{z}_1) = \xi_2^1, \dots, \pi_N \circ \mathcal{U}_S^N(\underline{z}_{N-1}) = \xi_N^1\right], \\ & \text{où } \underline{z}_i = (\xi_1^1, \dots, \xi_2^1, \pi_{i+1} \circ \underline{x}(0), \dots, \pi_N \circ \underline{x}(0)). \end{aligned}$$

Donc :

$$\begin{aligned} \mathbf{P}\left[\bigcup_{\nu=1}^p \left\{\mathcal{U}_S(\underline{x}^\nu(0)) = -\underline{\xi}^\nu\right\}^C\right] &\leq p \mathbf{P}\left[\left\{\mathcal{U}_S(\underline{x}(0)) = -\underline{\xi}^1\right\}^C\right] \\ &\leq p \sum_{i=1}^N \mathbf{P}\left[\pi_i \circ \mathcal{U}_S^i(\underline{z}_{i-1}) = -\xi_i^1\right] \\ &\leq pN K_{N,p}(n), \text{ par le lemme 3.2.5.} \end{aligned}$$

Par le lemme de Borel-Cantelli, (i) est vérifiée s'il existe un entier N_0 , tel que $\sum_{N>N_0} pN K_{N,p}(n) < \infty$. Et (ii) est vérifiée si $\lim_{N \rightarrow \infty} pN K_{N,p}(n) = 0$.

On en déduit les conditions sur γ, γ_1 , et γ_p . ♣

3.3 Étude des fluctuations du hamiltonien

Au paragraphe III.1, nous avons montré que le hamiltonien, en tant que fonction de la configuration du réseau aux instants entiers successifs, décroît avec le temps pour la dynamique considérée. Et les états stables atteints sont des minima au moins locaux de $H_{N,p}^S(\underline{\xi}, \cdot)$.

Il est donc intéressant d'étudier les fluctuations de $H_{N,p}^S(\underline{\xi}, \cdot)$:

- les configurations $\underline{\xi}^\mu$ sont elles des minima locaux, ou globaux, de $H_{N,p}^S(\underline{\xi}, \cdot)$?
- existe-t-il d'autres minima ?

3.3.1 Mémoire associative avec erreurs tolérées

En 1988, C.Newman a montré le résultat suivant, par une méthode de grandes déviations [38].

Theorème 3.3.1 (Newman) *Il existe un nombre $\alpha_c > 0$ tel que pour chaque $\alpha \leq \alpha_c$, $p = \alpha N$, il existe $\delta \in]0, \frac{1}{2}[$, $\epsilon > 0$, $\epsilon' \equiv f(\alpha, \delta, \epsilon) > 0$, tels que*

$$\lim_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow \alpha}} e^{\epsilon' N} \left(1 - \mathbf{P} \left[\bigcap_{\mu=1}^p \bigcap_{\underline{y} \in \mathcal{S}(\underline{\xi}^\mu, \delta N)} \left\{ H_{N,p}^{\mathcal{S}}(\underline{\xi}, \underline{y}) > H_{N,p}^{\mathcal{S}}(\underline{\xi}, \underline{\xi}^\mu) + \epsilon N \right\} \right] \right) = 0 \quad (3.9)$$

En particulier, le lemme de Borel-Cantelli implique :

$$\mathbf{P} \left[\liminf_N \left\{ \bigcap_{\mu=1}^p \bigcap_{\underline{y} \in \mathcal{S}(\underline{\xi}^\mu, \delta N)} \left\{ H_{N,p}^{\mathcal{S}}(\underline{\xi}, \underline{y}) > H_{N,p}^{\mathcal{S}}(\underline{\xi}, \underline{\xi}^\mu) + \epsilon N \right\} \right\} \right] = 1 \quad (3.10)$$

Le théorème de Newman signifie que pour presque tout choix d'images mémorisées $\{\underline{\xi}^1, \dots, \underline{\xi}^p\}$, chacune de ces images $\underline{\xi}^\mu$ est entourée d'une barrière énergétique. Cela implique en particulier qu'il existe au moins un minimum local dans la boule de centre $\underline{\xi}^\mu$, de rayon δN . Toutefois, le minimum n'est pas nécessairement $\underline{\xi}^\mu$!

Remarque : la propriété de mémoire associative sans erreur avec attraction directe n'est vérifiée que dans le cas où $p \leq Cte \frac{N}{\log N}$ (voir paragraphe 2). Et Mac Eliece donne des arguments heuristiques suggérant que le nombre d'images mémorisées ne doit pas dépasser cet ordre de grandeur, si on autorise une attraction en plusieurs étapes, mais sans erreur, vers une de ces images.

Par contre, la mémoire associative avec erreurs, suggérée par le théorème de Newman, fonctionne pour un nombre d'images mémorisées proportionnel à la taille du réseau. Il faut également tenir compte d'un autre paramètre : $\delta(\alpha)$, représentant la fraction de bits erronés dans l'image retrouvée. Dans son article, Newman obtient les bornes rigoureuses suivantes : $\alpha_c > 0,056$; et $\delta(\alpha) < 0,012$ pour $\alpha \leq 0,056$.

Il est intéressant de noter que Amit, et al[6, 4], par des simulations numériques et des techniques de calcul non rigoureuses, ont obtenu la borne $\alpha_c \simeq 0,14$.

3.3.2 Existence d'autres minima locaux : les images combinées

Amit et al[5] ont découvert en 1985 l'existence d'autres minima locaux du hamiltonien $H_{N,p}^{\mathcal{S}}$: il s'agit des combinaisons d'un nombre **fini** (i.e. indépendant de N et p) d'images mémorisées. Plus précisément : notons $\underline{x}(\underline{d}) = (x_1(\underline{d}), \dots, x_N(\underline{d}))$,

$$\text{avec } x_i(\underline{d}) \stackrel{\text{déf}}{=} \text{sgn} \left(\sum_{\mu=1}^p d_\mu \xi_i^\mu \right)$$

et $\underline{d} = (d_1, \dots, d_p)$ est un vecteur de \mathbf{R}^p , vérifiant :

(i) \underline{d} n'a qu'un nombre fini (i.e. indépendant de N et p) de composantes non nulles.

$$(ii) \sum_{\mu=1}^p \epsilon_{\mu} d_{\mu} \neq 0, \forall \epsilon \in \{-1, +1\}^p.$$

$$(iii) \mathbf{E} \left[x_i(\underline{d}) \xi_i^{\mu} \right] = d_{\mu}, \forall \mu \in \{1, \dots, p\}.$$

Les $\underline{x}(\underline{d})$ sont appelées *images combinées*.

Voici quelques exemples, cités par Amit, de vecteurs de \mathbf{R}^p satisfaisant aux hypothèses (i), (ii), et (iii) :

- des vecteurs \underline{d} contenant un nombre impair de composantes identiques : $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0, 0, \dots)$, $(\frac{3}{8}, \frac{3}{8}, \frac{3}{8}, \frac{3}{8}, \frac{3}{8}, 0, 0, \dots)$. Nous désignerons par (\star) les vecteurs de ce type.

-certains vecteurs incluant des composantes distinctes : $(\frac{1}{2}, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, \dots)$.

Bien entendu, tout vecteur obtenu par permutation des composantes d'un vecteur vérifiant (i), (ii), et (iii) les vérifient également.

Lemme 3.3.2 *Les hypothèses (i) et (iii) impliquent :*

$$(a) \|\underline{d}\|^2 \stackrel{\text{déf}}{=} \left(\sum_{\mu=1}^p (d_{\mu})^2 \right) \leq 1, \text{ avec égalité si et seulement si } \underline{d} \text{ a une seule composante } d_{\nu} \text{ non nulle (dans ce cas, } \underline{x}(\underline{d}) = \pm \underline{\xi}^{\nu} \text{)}.$$

(b) Soit $\underline{x}(\underline{d})$ une image combinée du type (\star) , i.e. les D composantes non nulles de \underline{d} sont identiques, et D est impair. Alors :

$$\|\underline{d}\|^2 = D \left(2^{1-D} C_{d-1}^{\frac{D-1}{2}} \right)^2 \leq \frac{2}{\pi} \frac{D}{D-1} \exp \left\{ \frac{-18D+19}{6(D-1)(6D-5)} \right\}.$$

(c) Soit $D \in \mathbf{N}$ fixé. Alors, le nombre d'images combinées à D composantes est majoré par $(2^{D-1}p)^D$.

Démonstration : (a) l'égalité (iii) s'écrit sous forme vectorielle :

$$\underline{d} = \mathbf{E} \left[\underline{\xi}_1 \operatorname{sgn}((\underline{\xi}_1, \underline{d})) \right], \text{ où } (\cdot, \cdot) \text{ désigne le produit scalaire dans } \mathbf{R}^p.$$

$$\text{Donc } (\underline{d}, \underline{d}) = \mathbf{E} \left[|(\underline{\xi}_1, \underline{d})| \right] = \sum_{\underline{y} \in \{-1, +1\}^p} \frac{1}{2^p} |(\underline{y}, \underline{d})|.$$

Posons, pour $k \in \{1, \dots, q \equiv 2^p\}$, $a_k = \frac{1}{2^{\frac{p}{2}}}$, $b_k = \frac{1}{2^{\frac{p}{2}}} |(\underline{y}_k, \underline{d})|$, où $\underline{y}_1, \underline{y}_2, \dots, \underline{y}_q$ est une énumération arbitraire des vecteurs de $\{-1, +1\}^p$. On peut alors écrire :

$$\begin{aligned}
(\underline{d}, \underline{d}) &= \sum_{k=1}^q a_k b_k \leq \sqrt{\sum_{k=1}^q (a_k)^2 \sum_{k=1}^q (b_k)^2} \text{ (Inégalité de Cauchy)} \\
&= \sqrt{\sum_{k=1}^q (b_k)^2} = \sqrt{\sum_{k=1}^q \frac{1}{2^p} (y_k, \underline{d})^2} \\
&= \sqrt{\mathbf{E}[(\xi_1, \underline{d})^2]} \\
&= \sqrt{\sum_{\mu=1}^p (d_\mu)^2} \text{ (par indépendance des v.a. } (\xi_1^\mu)^{\mu=1, \dots, p}).
\end{aligned}$$

Nous avons montré : $\|\underline{d}\|^2 \leq \|\underline{d}\|$. Cela implique : $\|\underline{d}\| \leq 1$.

De plus, dans l'inégalité de Cauchy, l'égalité n'est obtenue que si $\underline{x} = \lambda \underline{y}$, où λ est une constante, ce qui signifie :

$$\left| \sum_{\mu=1}^p y_k^\mu d_\mu \right| = \lambda, \text{ pour tout vecteur } y_k \in \{-1, +1\}^p.$$

Le seul cas possible est celui où \underline{d} n'a qu'une composante non nulle, égale à ± 1 .

(b) Si toutes les composantes d_μ sont égales (appelons d la valeur commune), (iii) s'écrit :

$$d = \mathbf{E} \operatorname{sgn}\left(1 + \sum_{\mu=1}^{D-1} \xi_1^\mu\right)$$

Donc

$$\begin{aligned}
d &= 1 - 2 \mathbf{P}\left[\sum_{\mu=1}^{D-1} \xi_1^\mu < -1\right] = 1 - 2 \sum_{j=-D+1}^{-2} C_{D-1}^{\frac{D+j-1}{2}} \left(\frac{1}{2}\right)^{D-1} \\
&= 1 - 2^{-D+2} \left(\sum_{k=0}^{\frac{D-1}{2}} C_D^k - C_{D-1}^{\frac{D-1}{2}}\right) = 2^{-D+1} C_{D-1}^{\frac{D-1}{2}}
\end{aligned}$$

Donc

$$\|\underline{d}\|^2 = D \left(2^{-D+1} C_{D-1}^{\frac{D-1}{2}}\right)^2$$

En utilisant l'encadrement suivant, dû à H.Robbins (voir [17],chap.3) :

$$n^n e^{-n} \sqrt{2\pi n} e^{\frac{1}{12n+1}} < n! < n^n e^{-n} \sqrt{2\pi n} e^{\frac{1}{12n}},$$

pour majorer $C_{D-1}^{\frac{D-1}{2}} = \frac{(D-1)!}{\{(\frac{D-1}{2})!\}^2}$, on obtient (b).

(c) Soit D fixé. Supposons que $\underline{d} = (d_1, \dots, d_D, 0, 0, \dots)$.

$$\begin{aligned} \forall \nu \in \{1, \dots, D\}, d_\nu &= \mathbf{E} \left[\text{sgn} \left(d_\nu + \sum_{\substack{\mu=1, \dots, D \\ \mu \neq \nu}} d_\mu \xi_1^\mu \xi_1^\nu \right) \right] \\ &= \frac{1}{2^{D-1}} \sum_{\substack{(\epsilon_i = \pm 1) \\ i=1, \dots, D \\ i \neq \nu}} \text{sgn} \left(d_\nu + \sum_{\substack{\mu=1, \dots, D \\ \mu \neq \nu}} d_\mu \epsilon_\mu \right). \end{aligned}$$

Donc d_ν peut prendre 2^{D-1} valeurs possibles, et $\underline{d} = (d_1, \dots, d_D, 0, 0, \dots)$ $2^{D(D-1)}$ valeurs vectorielles possibles. On en déduit que le nombre d'images combinées est majoré (strictement) par $(2^{D-1}p)^D$. ♣

Remarque : il semble difficile d'étendre la méthode utilisée au (b) au cas général. Cela conduit en fait à l'étude d'une marche aléatoire sur \mathbf{R} , dont la longueur des sauts dépend de l'instant du saut.

Amit et al [6] ont étudié la stabilité de telles configurations par des techniques non rigoureuses (méthodes des répliques), et affirment que les configurations du type (\star) sont stables pour $p = \alpha N$, $\alpha \leq 0,03$. Ce résultat est partiellement confirmé par le théorème suivant, dû à Newman [38], qui montre que chaque $\underline{x}(\underline{d})$ est entouré d'une barrière énergétique, pour α suffisamment petit.

Théorème 3.3.3 (Newman) *Soit D un entier non nul fixé. Soit $\mathbf{I}(D)$ l'ensemble des vecteurs de \mathbf{R}^p , dont le nombre de composantes non nulles est D , et vérifiant (i), (ii), et (iii).*

Il existe $\alpha_c > 0$, tel que pour $\alpha < \alpha_c$, $p = \alpha N$, il existe $\delta \in]0, 1/2[$, $\epsilon > 0$, $\epsilon' = f(\alpha, \delta, \epsilon) > 0$ tel que

$$\lim_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow \alpha}} e^{\epsilon' N} \left(1 - \mathbf{P} \left[\bigcap_{\underline{d} \in \mathbf{I}(D)} \bigcap_{\underline{y} \in \mathcal{S}(\underline{x}(\underline{d}), \delta N)} \left\{ H_{N,p}^{\mathcal{S}}(\underline{\xi}, \underline{y}) > H_{N,p}^{\mathcal{S}}(\underline{\xi}, \underline{x}(\underline{d})) + \epsilon N \right\} \right] \right) = 0 \quad (3.11)$$

En particulier, le lemme de Borel-Cantelli implique :

$$\mathbf{P} \left[\liminf_N \bigcap_{\underline{d} \in \mathbf{I}(D)} \bigcap_{\underline{y} \in \mathcal{S}(\underline{x}(\underline{d}), \delta N)} \left\{ H_{N,p}^S(\underline{\xi}, \underline{y}) > H_{N,p}^S(\underline{\xi}, \underline{x}(\underline{d})) + \epsilon N \right\} \right] = 1 \quad (3.12)$$

Lorsque D dépend de p , $\underline{x}(\underline{d})$ dépend d'un nombre important d'images, et les arguments développés par Newman ne fonctionnent plus.

Dans l'article [28], Komlos et Paturi étudient notamment la stabilité des 2^p configurations $\underline{w}(\underline{d}) = \text{sgn}(\sum_{\mu=1}^p d_\mu \underline{\xi}^\mu)$, où \underline{d} décrit l'ensemble des vecteurs de $X_p = \{-1, +1\}^p$. Je pense que ce problème n'est pas complètement résolu. En effet, ils annoncent qu'il existe au moins c^p vecteurs \underline{d} de X_p (c constante < 2) tels que $\underline{w}(\underline{d})$ soit à une distance ϵN d'un vecteur stable (ϵ arbitrairement petit), avec une probabilité tendant vers 1, lorsque $N \rightarrow \infty$, et $p < \alpha_c(\epsilon)N$. Toutefois, la preuve de ce résultat requiert des estimations exponentielles qui ne sont pas montrées (ils utilisent la même méthode que Newman), et que nous n'avons pas réussi à obtenir.

L'étude de la localisation des vecteurs $\underline{w}(\underline{d})$, $\underline{d} \in X_p$, dans l'espace X_N est également un problème délicat. En un certain sens, on peut dire que ces images combinées sont "symétriques" : elles dépendent de toutes les images originales, sans qu'aucune ne soit privilégiée. Cela implique notamment, qu'avec probabilité 1, pour N suffisamment grand, les images combinées sont à une distance arbitrairement proche de $N/2$ de chaque image originale (voir proposition 3.3.4).

D'autre part, il est facile de voir que si \underline{d}_1 et \underline{d}_2 sont deux vecteurs de X_p ne différant que par un nombre fini de coordonnées, alors

$$\forall \rho > 0, \mathbf{P} \left[d_H(\underline{w}(\underline{d}_1), \underline{w}(\underline{d}_2)) > \rho N \right] \xrightarrow[N \rightarrow \infty]{} 0.$$

Il est beaucoup plus difficile de répondre dans le cas général : soit $\rho > 0$, et $d_H(\underline{d}_1, \underline{d}_2) = \delta(p, \rho)p$ (nous notons indifféremment d_H la distance de Hamming sur X_N et X_p). Pour quelles valeurs de $\delta(p, \rho)$ a-t-on :

$$\mathbf{P} \left[d_H(\underline{w}(\underline{d}_1), \underline{w}(\underline{d}_2)) < \rho N \right] \xrightarrow[N \rightarrow \infty]{} 1 ?$$

Ce problème, à ma connaissance non résolu, est essentiel pour évaluer le nombre de vecteurs stables, car deux vecteurs $\underline{w}(\underline{d}_1)$ et $\underline{w}(\underline{d}_2)$ suffisamment proches l'un de l'autre convergent vers le même état stable.

Proposition 3.3.4 *Supposons que $\lim_{N \rightarrow \infty} \frac{p}{N} < 1$, et soit $\rho \in]0, 1/2[$. Alors :*

$$\mathbf{P} \left[\liminf_N \bigcap_{\underline{d} \in X_p} \bigcap_{\nu=1}^p \left\{ \left\{ d_H(\underline{\xi}^\nu, \underline{w}(\underline{d})) \geq \rho N \right\} \cap \left\{ d_H(-\underline{\xi}^\nu, \underline{w}(\underline{d})) \geq \rho N \right\} \right\} \right] = 1 \quad (3.13)$$

Pour montrer ce résultat, nous avons besoin de l'estimation de grande déviation suivante :

Lemme 3.3.5 (Chernoff, 1952) *Soient p, q deux réels positifs tels que $p + q = 1$. Alors, pour $M \leq pm$:*

$$\sum_{k=0}^M C_m^k p^k q^{m-k} \leq \exp \left\{ (m - M) \log \left(\frac{mq}{m - M} \right) + M \log \left(\frac{mp}{M} \right) \right\} \quad (3.14)$$

La même estimation est vraie pour $\sum_{k=M}^m C_m^k p^k q^{m-k}$, si $M \geq pm$.

Démonstration du lemme : soient $(Y_i)_{i=1, \dots, m}$ m v.a.i.i.d. valant 0 ou 1, avec $\mathbf{P}[Y_1 = 0] = q$.

$$\begin{aligned} \text{Alors } \sum_{k=0}^M C_m^k p^k q^{m-k} &= \mathbf{P} \left[\sum_{i=1}^m Y_i \leq M \right] \\ &\leq \inf_{t>0} e^{tM} \mathbf{E} e^{-t \sum_{i=1}^m Y_i} = \inf_{t>0} e^{tM} (pe^{-t} + q)^m \end{aligned}$$

La fonction $f(t) = tM + m \log(pe^{-t} + q)$ atteint son minimum en $t_0 = \log\left(\frac{p(m-M)}{Mq}\right)$, d'où on déduit la majoration annoncée.

On obtient de même la deuxième estimation en écrivant

$$\sum_{k=M}^m C_m^k p^k q^{m-k} = \mathbf{P} \left[\sum_{i=1}^m Y_i \geq M \right] \leq \inf_{t>0} e^{-tM} \mathbf{E} e^{t \sum_{i=1}^m Y_i} \clubsuit$$

Démonstration de la proposition 3.3.4 : soit $\tilde{\rho}$ tel que $\tilde{\rho}N = \sup\{k \in \mathbf{N} : k < \rho N\}$. Nous avons :

$$\mathbf{P} \left[\bigcup_{\underline{d} \in X_p} \bigcup_{\nu=1}^p d_H(-\underline{\xi}^\nu, \underline{w}(\underline{d})) < \rho N \right] \leq \sum_{\underline{d} \in X_p} \sum_{\nu=1}^p \mathbf{P} \left[d_H(\underline{\xi}^\nu, \underline{w}(\underline{d})) < \rho N \right].$$

Et, par indépendance des v.a. $w_i(\underline{d})$, $i = 1, \dots, N$, nous avons

$$\mathbf{P} \left[d_H(\underline{\xi}^\nu, \underline{w}(\underline{d})) < \rho N \right] \leq \sum_{k=0}^{\tilde{\rho}N} C_N^k \left\{ \mathbf{P} \left[\xi_1^\nu = -w_1(\underline{d}) \right] \right\}^k \left\{ \mathbf{P} \left[\xi_1^\nu = w_1(\underline{d}) \right] \right\}^{N-k}. \quad (3.15)$$

Maintenant, supposons que $d_\nu = 1$.

$$\begin{aligned} \mathbf{P}\left[\xi_1^\nu = w_1(\underline{d})\right] &= \sum_{s=-p}^p \mathbf{P}\left[\sum_{\mu=1}^p d_\mu \xi_1^\mu = s, \xi_1^\nu = \text{sgn}(s)\right] \\ &= \sum_{s=-p}^{-1} \mathbf{P}\left[\sum_{\mu=1}^p d_\mu \xi_1^\mu = s, \xi_1^\nu = -1\right] + \sum_{s=0}^p \mathbf{P}\left[\sum_{\mu=1}^p d_\mu \xi_1^\mu = s, \xi_1^\nu = 1\right] \end{aligned}$$

Supposons p et s pairs. Alors :

$$\mathbf{P}\left[\sum_{\mu=1}^p d_\mu \xi_1^\mu = s, \xi_1^\nu = 1\right] = C_{p-1}^{(p+s)/2-1} \left(\frac{1}{2}\right)^p,$$

donc

$$\begin{aligned} \sum_{s=0}^p \mathbf{P}\left[\sum_{\mu=1}^p d_\mu \xi_1^\mu = s, \xi_1^\nu = 1\right] &= \sum_{s=0}^p C_{p-1}^{(p+s)/2-1} \left(\frac{1}{2}\right)^p \\ &= \frac{1}{4} + C_{p-1}^{p/2} \left(\frac{1}{2}\right)^p. \end{aligned}$$

De même,

$$\sum_{s=-p}^{-2} \mathbf{P}\left[\sum_{\mu=1}^p d_\mu \xi_1^\mu = s, \xi_1^\nu = -1\right] = \frac{1}{4}.$$

On en déduit :

$$\begin{cases} \mathbf{P}\left[\xi_1^\nu = w_1(\underline{d})\right] = \frac{1}{2} + C_{p-1}^{p/2} \left(\frac{1}{2}\right)^p = \frac{1}{2} + \frac{1}{\sqrt{2\pi p}} + R_p \\ \mathbf{P}\left[\xi_1^\nu = -w_1(\underline{d})\right] = \frac{1}{2} - C_{p-1}^{p/2} \left(\frac{1}{2}\right)^p = \frac{1}{2} - \frac{1}{\sqrt{2\pi p}} - R_p \end{cases},$$

où R_p est d'ordre $1/p$, pour $p \rightarrow \infty$. D'après (3.15), on a alors :

$$\mathbf{P}\left[d_H(\underline{\xi}^\nu, \underline{w}(\underline{d})) < \rho N\right] \leq \sum_{k=0}^{\bar{\rho}N} C_N^k \left(\frac{1}{2} - \frac{1}{\sqrt{2\pi p}} - R_p\right)^k \left(\frac{1}{2} + \frac{1}{\sqrt{2\pi p}} + R_p\right)^{N-k} \quad (3.16)$$

Si $d_\nu = -1$, on obtient par la même méthode :

$$\mathbf{P}\left[d_H(\underline{\xi}^\nu, \underline{w}(\underline{d})) < \rho N\right] \leq \sum_{k=0}^{\bar{\rho}N} C_N^k \left(\frac{1}{2} + \frac{1}{\sqrt{2\pi p}} + R_p\right)^k \left(\frac{1}{2} - \frac{1}{\sqrt{2\pi p}} - R_p\right)^{N-k} \quad (3.17)$$

Il existe un p_0 tel que $\tilde{\rho} \leq \frac{1}{2} - \frac{1}{\sqrt{2\pi p}} - R_p$, pour tout $p > p_0$, car $\rho < 1/2$. Le lemme 3.3.5 s'applique donc, et on obtient :

$$\mathbf{P} \left[\bigcup_{\underline{d} \in X_p} \bigcup_{\nu=1}^p d_H(\underline{\xi}^\nu, \underline{w}(\underline{d})) < \rho N \right] \leq \exp \left\{ Nh(\rho) + p \log 2 + \log p + c_N \right\},$$

$$\text{avec } h(\rho) = -\log 2 - (1 - \rho) \log(1 - \rho) - \rho \log \rho < 0, \quad \forall \rho \in]0, 1/2[, \\ \lim_{N \rightarrow \infty} \frac{c_N}{N} = 0.$$

Ceci est encore vrai pour p impair, et on obtient évidemment la même majoration pour $\mathbf{P} \left[\bigcup_{\underline{d} \in X_p} \bigcup_{\nu=1}^p d_H(-\underline{\xi}^\nu, \underline{w}(\underline{d})) < \rho N \right]$.

Par le lemme de Borel-Cantelli, on a alors le résultat annoncé, si $p < N$ à partir d'un certain rang. ♣

3.3.3 Convergence presque sûre de $\frac{1}{N} H_{N,p}^S(\underline{\xi}, \cdot)$ pour certaines configurations dépendant des images originales

Pour $p = 1$, les vecteurs $\underline{\xi}^1$ et $-\underline{\xi}^1$ sont les seuls minima globaux du hamiltonien, et $\frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{\xi}^1) = \frac{1}{N} H_{N,p}^S(\underline{\xi}, -\underline{\xi}^1) = -1$.

Mais pour $p = \alpha N$, $\alpha < \alpha_c$, nous savons seulement qu'il existe un minimum local proche de chaque image originale et de chaque image combinée (voir paragraphes précédents). Peut-on déterminer la valeur de $\frac{1}{N} H_{N,p}^S(\underline{\xi}, \cdot)$ en ces points ?

Nous allons tout d'abord étudier le comportement asymptotique de $\frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{\xi}^\mu)$, $\mu \in \{1, \dots, p\}$.

$$\text{Notons que le hamiltonien peut s'écrire : } H_{N,p}^S(\underline{\xi}, \underline{x}) = -\frac{1}{N} \sum_{\mu=1}^p (\underline{\xi}^\mu, \underline{x})^2 + p,$$

où $(\underline{x}, \underline{y})$ désigne le produit scalaire de \underline{x} et \underline{y} dans \mathbf{R}^N .

$$\text{Pour } \nu \in \{1, \dots, p\}, \quad \frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{\xi}^\nu) = -1 + \frac{p}{N} - \frac{1}{N^2} \sum_{\substack{\mu=1, \dots, p \\ \mu \neq \nu}} (\underline{\xi}^\nu, \underline{\xi}^\mu)^2, \text{ d'où on déduit la}$$

borne triviale :

$$\frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{\xi}^\nu) \leq -1 + \frac{p}{N}, \text{ pour } \nu \in \{1, \dots, p\}.$$

D'autre part, nous avons :

Proposition 3.3.6 Si $\lim_{N \rightarrow \infty} \frac{p}{N} = \alpha \geq 0$, alors pour tout $\epsilon > 0$,

$$\mathbf{P} \left[\liminf_N \bigcap_{\nu=1}^p \left\{ \frac{1}{N} H_{N,p}^{\mathcal{S}}(\underline{\xi}, \underline{\xi}^\nu) \right\} > -1 - \epsilon \right] = 1 \quad (3.18)$$

Corollaire 3.3.7 Si $\lim_{N \rightarrow \infty} \frac{p}{N} = 0$, alors pour tout $\nu \in \{1, \dots, p\}$,

$$\frac{1}{N} H_{N,p}^{\mathcal{S}}(\underline{\xi}, \underline{\xi}^\nu) \xrightarrow{p.s., N \rightarrow \infty} -1 \quad (3.19)$$

Démonstration de la proposition 3.3.6 : par le lemme de Borel-Cantelli, il suffit de montrer l'existence d'un entier N_0 tel que :

$$\sum_{N > N_0} \mathbf{P} \left[\bigcup_{\nu=1}^p \left\{ \frac{1}{N} H_{N,p}^{\mathcal{S}}(\underline{\xi}, \underline{\xi}^\nu) \right\} < -1 - \epsilon \right] < \infty$$

Pour celà, nous appliquons l'inégalité exponentielle de Markov :

$$\begin{aligned} \mathbf{P} \left[\frac{1}{N} H_{N,p}^{\mathcal{S}}(\underline{\xi}, \underline{\xi}^\nu) < -1 - \epsilon \right] &\leq \inf_{t > 0} e^{-(1+\epsilon)tN} \mathbf{E} e^{-t H_{N,p}^{\mathcal{S}}(\underline{\xi}, \underline{\xi}^\nu)} \\ &\leq \inf_{t > 0} e^{-(\epsilon + \frac{p}{N})tN} \mathbf{E} \prod_{\substack{\mu=1 \\ \mu \neq \nu}}^p e^{\frac{t}{N} (\xi^\mu, \xi^\nu)^2}, \\ \text{et} \quad \mathbf{E} \prod_{\substack{\mu=1 \\ \mu \neq \nu}}^p e^{\frac{t}{N} (\xi^\mu, \xi^\nu)^2} &= \prod_{\substack{\mu=1 \\ \mu \neq \nu}}^p \mathbf{E}_{\xi^\mu} \mathbf{E}_{Y^\mu} \exp \left(\sqrt{\frac{2t}{N}} (\xi^\mu, \xi^\nu) Y^\mu \right), \end{aligned}$$

où $\underline{Y} = (Y^1, Y^2, \dots, Y^p)$ est un vecteur gaussien de v.a.i.i.d. de loi $\mathcal{N}(0, 1)$. Par le théorème de Fubini, on a alors :

$$\begin{aligned} \mathbf{E} \prod_{\substack{\mu=1 \\ \mu \neq \nu}}^p e^{\frac{t}{N} (\xi^\mu, \xi^\nu)^2} &= \prod_{\substack{\mu=1 \\ \mu \neq \nu}}^p \mathbf{E}_{Y^\mu} \left\{ \left(\cosh \left(\sqrt{\frac{2t}{N}} Y^\mu \right) \right)^N \right\} \\ &\leq \prod_{\substack{\mu=1 \\ \mu \neq \nu}}^p \mathbf{E}_{Y^\mu} \left\{ e^{t(Y^\mu)^2} \right\}, \text{ car } \forall \lambda \in \mathbf{R}, \cosh(\lambda) \leq e^{\frac{\lambda^2}{2}} \\ &= \left(\frac{1}{1-2t} \right)^{\frac{p-1}{2}}, \text{ pour } t \in]0, \frac{1}{2}[. \end{aligned}$$

Donc

$$\begin{aligned} \mathbf{P} \left[\bigcup_{\nu=1}^p \left\{ \frac{1}{N} H_{N,p}^{\mathcal{S}}(\underline{\xi}, \underline{\xi}^\nu) \right\} < -1 - \epsilon \right] &\leq p \inf_{t > 0} \exp \left\{ -(\epsilon + \frac{p}{N})tN - \frac{p-1}{2} \log(1-2t) \right\} \\ &= \exp \left\{ -N \left(\frac{\epsilon}{2} + \frac{p-1}{2N} \log \left(\frac{p-1}{p+\epsilon N} \right) - \frac{\log p}{N} - \frac{1}{2N} \right) \right\}. \end{aligned}$$

Nous devons montrer qu'il existe un entier N_0 tel que pour $N > N_0$,

$$\frac{\epsilon}{2} + \frac{p-1}{2N} \log\left(\frac{p-1}{p+\epsilon N}\right) - \frac{\log p}{N} - \frac{1}{2N} > 0. \quad (3.20)$$

-Si $\frac{p}{N} \rightarrow 0$, $\frac{p-1}{2N} \log\left(\frac{p-1}{p+\epsilon N}\right) \rightarrow_{N \rightarrow \infty} 0$, donc (3.20) est vrai.

-Si $\frac{p}{N} \rightarrow \alpha > 0$, remarquons que l'inégalité (3.20) peut aussi s'écrire

$$\frac{\epsilon N}{p-1} > \log\left(\frac{p}{p-1} + \frac{\epsilon N}{p-1}\right) + 2\frac{\log p}{p-1} + \frac{1}{p-1}.$$

Cette condition est vérifiée à partir d'un certain rang $N_0(\epsilon)$, car à la limite, on obtient :

$$\frac{\epsilon}{\alpha} > \log\left(1 + \frac{\epsilon}{\alpha}\right), \text{ qui est vrai. } \clubsuit$$

Démonstration du corollaire : pour tout $\epsilon > 0$, nous avons :

$$\begin{aligned} \mathbf{P}\left[\limsup_N \left\{ \left| \frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{\xi}^\nu) + 1 \right| > \epsilon \right\}\right] &\leq \mathbf{P}\left[\limsup_N \left\{ \frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{\xi}^\nu) > -1 + \epsilon \right\}\right] \\ &\quad + \mathbf{P}\left[\limsup_N \left\{ \frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{\xi}^\nu) + 1 < -1 - \epsilon \right\}\right] \\ &= 0, \end{aligned}$$

d'où la convergence presque sûre. \clubsuit

Concernant les images combinées, nous avons le résultat suivant :

Proposition 3.3.8 *Soit D un entier non nul fixé. Soit $\mathbf{I}(D)$ l'ensemble des vecteurs de \mathbf{R}^p , dont le nombre de composantes non nulles est D , et vérifiant (i), (ii), et (iii). Si $\lim_{N \rightarrow \infty} \frac{p}{N} = 0$, alors pour tout $\epsilon > 0$,*

$$\mathbf{P}\left[\liminf_N \bigcap_{\underline{d} \in \mathbf{I}(D)} \left\{ \left| \frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{x}(\underline{d})) + \|\underline{d}\|^2 \right| < \epsilon \right\}\right] = 1 \quad (3.21)$$

Pour montrer ce résultat, nous utilisons l'estimation de grande déviation suivante :

Lemme 3.3.9 Soient Y_1, \dots, Y_N N v.a.i.i.d. valant ± 1 , avec $\mathbf{E}Y_1 = m$, $-1 < m < 1$. Alors, pour tout $\epsilon > 0$,

$$\mathbf{P}\left[\left|\frac{1}{N}\sum_{i=1}^N Y_i - m\right| > \epsilon\right] \leq \exp\{-Nh_1(m, \epsilon)\} + \exp\{-Nh_2(m, \epsilon)\}, \quad (3.22)$$

$$\text{avec } \begin{cases} h_1(m, \epsilon) = \frac{1-m-\epsilon}{2} \log\left(1 - \frac{\epsilon}{1-m}\right) + \frac{1+m+\epsilon}{2} \log\left(1 + \frac{\epsilon}{1+m}\right) > 0 \\ h_2(m, \epsilon) = \frac{1-m+\epsilon}{2} \log\left(1 + \frac{\epsilon}{1-m}\right) + \frac{1+m-\epsilon}{2} \log\left(1 - \frac{\epsilon}{1+m}\right) > 0 \end{cases}$$

Démonstration : Soit

$$X_i = \frac{Y_i + 1}{2} = \begin{cases} 0, & \text{avec proba } q = \frac{1-m}{2} \\ 1, & \text{avec proba } p = \frac{1+m}{2} \end{cases}$$

Le lemme 3.3.5 s'applique. En effet,

$$\mathbf{P}\left[\left|\frac{1}{N}\sum_{i=1}^N Y_i - m\right| > \epsilon\right] \leq \mathbf{P}\left[\sum_{i=1}^N X_i > \frac{(m+1+\epsilon)}{2}N\right] + \mathbf{P}\left[\sum_{i=1}^N X_i < \frac{(m+1-\epsilon)}{2}N\right].$$

Il suffit alors de réécrire (3.14) pour obtenir (3.22). Une étude simple de h_1 et h_2 montre qu'elles sont strictement positives pour $-1 < m < 1$, et $\epsilon > 0$. ♣

Démonstration de la proposition : pour simplifier les notations, nous supposons désormais que les composantes non nulles de \underline{d} sont les D premières. Ecrivons alors :

$$\begin{aligned} \frac{1}{N}H_{N,p}^S(\underline{\xi}, \underline{x}(\underline{d})) &= -\frac{1}{N^2}\sum_{\mu=1}^D (\xi^\mu, \underline{x}(\underline{d}))^2 - \frac{1}{N^2}\sum_{\mu=D+1}^p (\xi^\mu, \underline{x}(\underline{d}))^2 + \frac{p}{N} \\ &\equiv -S_1(N, p, D) - S_2(N, p, D). \end{aligned}$$

-Regardons le premier terme :

$$\begin{aligned} \mathbf{P}\left[\bigcup_{\underline{d} \in \mathbf{I}(D)} \left\{ |S_1(N, p, D) - \|\underline{d}\|^2| > \epsilon \right\}\right] &\leq \mathbf{P}\left[\bigcup_{\underline{d} \in \mathbf{I}(D)} \bigcup_{\mu=1}^D \left\{ \left| \frac{1}{N^2}(\xi^\mu, \underline{x}(\underline{d}))^2 - d_\mu^2 \right| > \frac{\epsilon}{D} \right\}\right] \\ &\leq \sum_{\underline{d} \in \mathbf{I}(D)} \sum_{\mu=1}^D \mathbf{P}\left[\left| \frac{1}{N}(\xi^\mu, \underline{x}(\underline{d})) - d_\mu \right| > \frac{\epsilon}{2D} \right]. \end{aligned}$$

Par définition, $(\xi^\mu, \underline{x}(\underline{d})) = \sum_{i=1}^N \xi_i^\mu x_i(\underline{d})$. Il s'agit d'une somme de N v.a.i.i.d. valant ± 1 , d'espérance d_μ : le lemme 3.3.9 s'applique. Puisque $\text{card}(\mathbf{I}(D)) \leq (2^{D-1}p)^D$ (voir lemme 3.3.2), on obtient, pour D fini et $p/N \rightarrow 0$:

$$\mathbf{P} \left[\limsup_N \bigcup_{\underline{d} \in \mathbf{I}(D)} \left\{ |S_1(N, p, D) - \|\underline{d}\|^2| > \epsilon \right\} \right] = 0, \quad \forall \epsilon > 0. \quad (3.23)$$

-Pour le deuxième terme, $S_2(N, p, D) = \frac{1}{N^2} \sum_{\mu=D+1}^p (\underline{\xi}^\mu, \underline{x}(\underline{d}))^2 - \frac{p}{N}$, en remarquant que $\underline{x}(\underline{d})$ est indépendant de $(\underline{\xi}^\mu)_{\mu=D+1, \dots, p}$, on peut montrer, exactement comme pour la proposition 3.3.6 et son corollaire, que pour $p/N \rightarrow 0$:

$$\mathbf{P} \left[\limsup_N \bigcup_{\underline{d} \in \mathbf{I}(D)} \left\{ S_2(N, p, D) > \epsilon \right\} \right] = 0, \quad \forall \epsilon > 0. \quad (3.24)$$

De (3.23) et (3.24), on déduit (3.21). ♣

Hormis les images combinées du type (\star) , pour lesquelles nous avons montré que $\|\underline{d}\|^2 < \frac{2}{\pi}$, nous savons seulement que $\|\underline{d}\| < 1$. Dans le cas général, le corollaire 3.3.7 et la proposition 3.3.8 n'impliquent donc pas que les minima locaux proches des images combinées sont moins profonds que ceux localisés près des images originales.

La proposition 3.3.8 et le lemme 3.3.2 impliquent que, avec probabilité 1, si $p/N \rightarrow 0$, les images combinées de type (\star) ayant un grand nombre (mais indépendant de p) de composantes non nulles vérifient $\frac{1}{N} H_N^S(\underline{\xi}, \underline{x}(\underline{d})) \xrightarrow{N \rightarrow \infty} \frac{2}{\pi}$.

Nous allons montrer que sous des conditions plus restrictives pour p , ce résultat est encore vrai pour les images combinées dépendant de toutes les images originales.

Proposition 3.3.10 *Pour tout $\epsilon > 0$, il existe une constante $c(\epsilon)$ indépendante de N et p , telle que pour $p < c(\epsilon) N^{2/3}$,*

$$\mathbf{P} \left[\liminf_N \bigcap_{\underline{d} \in \{-1, +1\}^p} \left\{ \left| \frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{w}(\underline{d})) + \frac{2}{\pi} \right| < \epsilon \right\} \right] = 1 \quad (3.25)$$

Démonstration : par définition, (\cdot, \cdot) désignant le produit scalaire dans \mathbf{R}^p ,

$$\begin{aligned} H_{N,p}^S(\underline{\xi}, \underline{w}(\underline{d})) &= - \left(\sum_{i=1}^N w_i(\underline{d}) \xi_i, \sum_{j=1}^N w_j(\underline{d}) \xi_j \right) + p \\ &= - \left(\sum_{i=1}^N \operatorname{sgn} \left(\sum_{\nu=1}^p \tilde{\xi}_i^\nu \right) \tilde{\xi}_i, \sum_{j=1}^N \operatorname{sgn} \left(\sum_{\nu=1}^p \tilde{\xi}_j^\nu \right) \tilde{\xi}_j \right) + p, \quad \text{où } \tilde{\xi}_i^\nu = d_\nu \xi_i^\nu. \end{aligned}$$

Nous avons donc :

$$\mathbf{P} \left[\bigcup_{\underline{d} \in \{-1, +1\}^p} \left\{ \left| \frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{w}(\underline{d})) + \frac{2}{\pi} \right| > \epsilon \right\} \right] \leq 2^p \mathbf{P} \left[\left| \frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{s}) + \frac{2}{\pi} \right| > \epsilon \right],$$

où \underline{s} est l'image combinée correspondant à $\underline{d} = (1, 1, \dots, 1)$, i.e. $\underline{s} = \text{sgn}(\sum_{\nu=1}^p \xi^\nu)$.

Pour $i \in \{1, \dots, N\}$, nous allons construire les vecteurs aléatoires $\underline{z}_i \in \{0, 1\}^p$ de la manière suivante : soit $k_i = \sum_{\nu=1}^p \xi_i^\nu$.

- si $k_i \geq 0$, alors \underline{z}_i est le vecteur nul,

- si $k_i < 0$, alors on choisit aléatoirement $|k_i|$ indices parmi les $(p + |k_i|)/2$ indices de $\underline{\xi}_i = (\xi_i^1, \dots, \xi_i^p)$ valant -1 , et cela suivant la distribution uniforme. Soit $\{l_1, \dots, l_{|k_i|}\}$ ces indices. On pose alors :

$$\begin{cases} z_i^\nu = 1, & \text{pour } \nu \in \{l_1, \dots, l_{|k_i|}\} \\ z_i^\nu = 0, & \text{pour } \nu \in \{1, \dots, p\} \setminus \{l_1, \dots, l_{|k_i|}\} \end{cases}$$

Il est clair que $s_i \xi_i \equiv \text{sgn}(\sum_{\nu=1}^p \xi_i^\nu) \xi_i$ et $\xi_i + 2\underline{z}_i$ ont même loi. De plus, nous avons :

Lemme 3.3.11 $\mathbf{E} z_i^\mu = \frac{1}{\sqrt{2\pi p}} + R_p$, où R_p est d'ordre $\frac{1}{p}$ pour $p \rightarrow \infty$.

Démonstration : supposons p pair, et soit $m = \frac{p}{2}$.

$$\begin{aligned} \mathbf{E} z_i^\mu &= \mathbf{P} \left[z_i^\mu = 1 \right] = \sum_{k=0}^{p/2-1} \mathbf{P} \left[\sum_{\mu=1}^p \xi_i^\mu = -(p-2k) \right] \frac{p-k}{p} \frac{p-2k}{p-k} \\ &= \sum_{k=0}^{p/2-1} C_p^k \left(\frac{1}{2} \right)^p \frac{p-2k}{p} = \left(\frac{1}{2} \right)^{2m} \sum_{k=0}^{m-1} C_{2m}^k \frac{m-k}{m} \\ &= \left(\frac{1}{2} \right)^{2m} \sum_{j=1}^m C_{2m}^{m-j} \frac{j}{m} = \left(\frac{1}{2} \right)^{2m} C_{2m}^m \sum_{j=1}^m \left(\prod_{k=0}^{j-1} \frac{m-k}{m+k+1} \right) \frac{j}{m} \\ &= \frac{1}{\sqrt{2\pi p}} + R_p, \end{aligned}$$

où R_p est d'ordre $\frac{1}{p}$ pour $p \rightarrow \infty$. La dernière égalité résulte de la formule de Stirling, qui implique $C_{2m}^m \sim_{m \rightarrow \infty} 2^{2m} \left(\frac{1}{\sqrt{m\pi}} + \mathcal{O}\left(\frac{1}{p}\right) \right)$ et du lemme suivant.

Bien sûr, ce calcul peut être mené de la même manière pour p impair. Dans ce cas, on pose $p = 2m + 1$, et on obtient comme ci-dessus :

$$\mathbf{E}z_i^\mu = \left(\frac{1}{2}\right)^{2m+1} C_{2m+1}^m \left\{ \frac{1}{2m+1} + \sum_{j=1}^m \left(\prod_{k=0}^{j-1} \frac{m-k}{m+k+2} \right) \frac{2j+1}{2m+1} \right\} = \frac{1}{\sqrt{2\pi p}} + R_p,$$

pour les mêmes raisons (voir lemme ci-dessous). ♣

Lemme 3.3.12 *Pour m entier ≥ 1 ,*

$$\sum_{j=1}^m \left(\prod_{k=0}^{j-1} \frac{m-k}{m+k+1} \right) \frac{j}{m} = \frac{1}{2} \quad (3.26)$$

$$\sum_{j=1}^m \left(\prod_{k=0}^{j-1} \frac{m-k}{m+k+2} \right) \frac{2j+1}{2m+1} = \frac{m}{2m+1} \quad (3.27)$$

Démonstration : Montrons la première formule. C'est une somme de m termes. Montrons par récurrence que la somme des k derniers termes ($j = m - k + 1, \dots, m$) vaut

$$\prod_{l=0}^{m-k-1} \frac{m-l}{m+l+1} \frac{k}{2m}$$

- pour $k = 1$, il s'agit du terme correspondant à $j = m$.

- Supposons le vrai pour k , et calculons la somme des $k + 1$ derniers termes : par hypothèse de récurrence,

$$\begin{aligned} \sum_{j=m-k}^m \prod_{l=0}^{j-1} \frac{m-l}{m+l+1} \frac{j}{m} &= \prod_{l=0}^{m-k-1} \frac{m-l}{m+l+1} \frac{k}{2m} + \prod_{l=0}^{m-k-1} \frac{m-l}{m+l+1} \frac{m-k}{m} \\ &= \prod_{l=0}^{m-k-2} \frac{m-l}{m+l+1} \frac{k+1}{2m}. \end{aligned}$$

L'hypothèse est donc vraie. On peut appliquer la relation pour $k = m$, qui s'écrit

$$\sum_{j=1}^m \prod_{k=0}^{j-1} \frac{m-k}{m+k+1} \frac{j}{m} = \frac{1}{2}.$$

La deuxième relation s'obtient de la même manière en montrant par récurrence sur k que la somme des k derniers termes vaut :

$$\prod_{l=0}^{m-k} \frac{m-l}{m+l+2} \frac{2m+2-k}{2m+1}.$$

Pour $k = m$, on obtient la relation (3.27). ♣

Nous pouvons maintenant achever la démonstration de la proposition. En effet,

$$\begin{aligned} \mathbf{P} \left[\left| \frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{s}) + \frac{2}{\pi} \right| > \epsilon \right] &= \mathbf{P} \left[\left| \frac{1}{N^2} \left(\sum_{i=1}^N \xi_i + 2z_i, \sum_{j=1}^N \xi_j + 2z_j \right) - \frac{2}{\pi} + \frac{p}{N} \right| > \epsilon \right] \\ &\leq \mathbf{P} \left[\left\| \sum_{i=1}^N \xi_i \right\|^2 > \epsilon_1 N^2 \right] + \mathbf{P} \left[\left\| \sum_{i=1}^N z_i \right\|^2 - \frac{N^2}{2\pi} + \frac{pN}{4} \right| > \epsilon_2 N^2 \right] \\ &\quad + \mathbf{P} \left[\left| \left(\sum_{i=1}^N \xi_i, \sum_{j=1}^N z_j \right) \right| > \epsilon_3 N^2 \right], \end{aligned}$$

avec $\epsilon_1, \epsilon_2, \epsilon_3 > 0$ et $\epsilon_1 + 4\epsilon_2 + 4\epsilon_3 = \epsilon$.

Regardons les trois termes séparément :

a. le premier terme :

$$\begin{aligned} \mathbf{P} \left[\left\| \sum_{i=1}^N \xi_i \right\|^2 > \epsilon_1 N^2 \right] &\leq \inf_{t>0} e^{-t\epsilon_1 N} \prod_{\mu=1}^p \mathbf{E} \exp \left\{ \frac{t}{N} \left(\sum_{i=1}^N \xi_i^\mu \right)^2 \right\} \\ &\leq \inf_{t>0} e^{-t\epsilon_1 N} \left(\frac{1}{\sqrt{1-2t}} \right)^p, \text{ pour } t \in]0, \frac{1}{2}[. \end{aligned}$$

(voir démonstration de la proposition 3.3.6).

Donc

$$\mathbf{P} \left[\left\| \sum_{i=1}^N \xi_i \right\|^2 > \epsilon_1 N^2 \right] \leq \exp \left\{ -t\epsilon_1 N \left(1 - \frac{2p}{\epsilon_1 N} \right) \right\},$$

pour $t \in]0, \frac{1}{4}[$, car $-\frac{1}{2} \log(1-u) < u$, pour $u < \frac{1}{2}$.

b. le deuxième terme :

$$\begin{aligned}
\mathbf{P}\left[\left|\left\|\sum_{i=1}^N \underline{z}_i\right\|^2 - \frac{N^2}{2\pi} + \frac{pN}{4}\right| > \epsilon_2 N^2\right] &\leq \mathbf{P}\left[\bigcup_{\mu=1}^p \left\{\left|\left(\sum_{i=1}^N z_i^\mu\right)^2 - \frac{1}{2\pi} \frac{N^2}{p} + \frac{N}{4}\right| > \epsilon_2 \frac{N^2}{p}\right\}\right] \\
&\leq p \mathbf{P}\left[\left|\left(\sum_{i=1}^N z_i^1\right)^2 - \frac{1}{2\pi} \frac{N^2}{p} + \frac{N}{4}\right| > \epsilon_2 \frac{N^2}{p}\right] \\
&\leq p \mathbf{P}\left[\sum_{i=1}^N z_i^1 > \frac{N}{\sqrt{p}} \sqrt{\frac{1}{2\pi} + \epsilon_2 - \frac{p}{4N}}\right] \\
&\quad + \mathbf{P}\left[\sum_{i=1}^N z_i^1 < \frac{N}{\sqrt{p}} \sqrt{\frac{1}{2\pi} - \epsilon_2 \frac{p}{4N}}\right].
\end{aligned}$$

Le lemme de Chernoff (3.3.5) s'applique, et on obtient :

$$\mathbf{P}\left[\left|\left\|\sum_{i=1}^N \underline{z}_i\right\|^2 - \frac{N^2}{2\pi} + \frac{pN}{4}\right| > \epsilon_2 N^2\right] \leq p \exp\left\{-N\left(\frac{c(\epsilon_2)}{\sqrt{p}} + R_p\right)\right\},$$

où $c(\epsilon_2)$ est une constante positive indépendante de N et p , et R_p est d'ordre $\frac{1}{p}$, pour $p \rightarrow \infty$.

- le troisième terme :

$$\begin{aligned}
\mathbf{P}\left[\left|\left(\sum_{i=1}^N \underline{\xi}_i, \sum_{j=1}^N \underline{z}_j\right)\right| > \epsilon_3\right] &\leq \mathbf{P}\left[\left\|\sum_{i=1}^N \underline{\xi}_i\right\| \left\|\sum_{j=1}^N \underline{z}_j\right\| > \epsilon_2 N^2\right] \\
&= \mathbf{P}\left[\left\{\left\|\sum_{i=1}^N \underline{\xi}_i\right\| \left\|\sum_{j=1}^N \underline{z}_j\right\| > \epsilon_2 N^2\right\} \cap \left\{\left\|\sum_{i=1}^N \underline{\xi}_i\right\| > \epsilon_4 N\right\}\right] \\
&\quad + \mathbf{P}\left[\left\{\left\|\sum_{i=1}^N \underline{\xi}_i\right\| \left\|\sum_{j=1}^N \underline{z}_j\right\| > \epsilon_2 N^2\right\} \cap \left\{\left\|\sum_{i=1}^N \underline{\xi}_i\right\| \leq \epsilon_4 N\right\}\right] \\
&\leq \mathbf{P}\left[\left\|\sum_{i=1}^N \underline{\xi}_i\right\| > \epsilon_4 N\right] + \mathbf{P}\left[\left\|\sum_{j=1}^N \underline{z}_j\right\| > \frac{\epsilon_2}{\epsilon_4} N\right]
\end{aligned}$$

Pour $\epsilon_4 > 0$ suffisamment petit, tel que $\frac{\epsilon_2}{\epsilon_4} > \frac{1}{\sqrt{2\pi}}$ on peut appliquer *b.* pour majorer le dernier terme. Et le premier terme est analogue à *a.* Finalement, on obtient la majoration globale suivante : il existe une constante $c_0(\epsilon) > 0$, telle que pour N et p suffisamment grands,

$$\mathbf{P}\left[\bigcup_{\underline{d} \in \{-1, +1\}^p} \left\{\left|\frac{1}{N} H_{N,p}^S(\underline{\xi}, \underline{X}(\underline{d})) + \frac{2}{\pi}\right| > \epsilon\right\}\right] \leq \exp\left\{-\frac{N}{\sqrt{p}}(c_0(\epsilon) - \frac{p^{\frac{3}{2}}}{N}(\log 2 + o(1)))\right\}.$$

On a donc (3.25) par le lemme de Borel-Cantelli.♣

Chapitre 4

Étude de la dynamique parallèle déterministe de Little

4.1 Définitions

Contrairement à la dynamique séquentielle, la dynamique parallèle autorise la modification de plusieurs spins x_i , à chaque instant entier t . Le modèle de Little tire son origine des travaux de W.Little et G.Shaw (voir les articles originaux [29],[30],[31]). La transformation $\mathcal{U}_{\mathcal{P}}$ associée est définie par :

$$\begin{aligned} \mathcal{U}_{\mathcal{P}}(\underline{\xi}, \cdot) : X_N &\longrightarrow X_N \\ \underline{x} &\longmapsto \underline{x}', \text{ avec } x'_i = \text{sgn}\left(\sum_{j=1}^N J_{ij}x_j\right), \quad i \in \{1, \dots, N\}. \end{aligned}$$

Les coefficients J_{ij} sont toujours définis par :

$$\begin{cases} J_{ij} = \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu, \text{ pour } i \neq j, \\ J_{ii} = 0. \end{cases}$$

À chaque condition initiale $\underline{x}(0) \in X_N$ correspond alors un ensemble de trajectoires $\left(\underline{x}(t) = \mathcal{U}_{\mathcal{P}}(\underline{\xi}(\omega), \underline{x}(t-1)), t \geq 1\right)_{\omega \in \Omega}$.

Le hamiltonien défini au chapitre précédent ne convient plus. Cependant, dans ce cas également, il est possible de construire une fonction énergie de type Lyapounov, i.e. bornée inférieurement, et décroissant sur chaque trajectoire induite par la dynamique (voir [41]).

Soit

$$\begin{aligned}
 H_{N,p}^{\mathcal{P}} : \Omega \times X_N &\longrightarrow \mathbf{R} \\
 (\omega, \underline{x}) &\longmapsto H_{N,p}^{\mathcal{P}}(\underline{\xi}(\omega), \underline{x}) = -\frac{1}{N} \sum_{i=1}^N \left| \sum_{j=1}^N J_{ij} x_j \right| \\
 &= -\frac{1}{N} \sum_{i=1}^N \left| \sum_{\substack{j=1 \\ j \neq i}}^N \sum_{\mu=1}^p \xi_i^\mu(\omega) \xi_j^\mu(\omega) x_j \right|
 \end{aligned}$$

Remarquons que $H_{N,p}^{\mathcal{P}}(\underline{\xi}, \underline{x}(t))$ peut s'écrire comme une fonction dépendant des deux états successifs $\underline{x}(t)$, $\underline{x}(t+1) \equiv \mathcal{U}_{\mathcal{P}}(\underline{\xi}, \underline{x}(t))$:

$$H_{N,p}^{\mathcal{P}}(\underline{\xi}, \underline{x}(t)) = -\frac{1}{N} \sum_{i,j=1}^N J_{ij} x_j(t) x_i(t+1).$$

Proposition 4.1.1 *Partant d'un état initial quelconque de X_N , la dynamique parallèle déterministe conduit :*

- soit à un point fixe de $\mathcal{U}_{\mathcal{P}}(\underline{\xi}, \cdot)$,
- soit à un 2-cycle de $\mathcal{U}_{\mathcal{P}}(\underline{\xi}, \cdot)$, i.e. un couple $(\underline{x}, \underline{y})$ de $X_N \times X_N$ vérifiant

$$\mathcal{U}_{\mathcal{P}}(\underline{\xi}, \underline{x}) = \underline{y}, \quad \mathcal{U}_{\mathcal{P}}(\underline{\xi}, \underline{y}) = \underline{x}.$$

Démonstration : pour $t \geq 1$,

$$\begin{aligned}
 &H_{N,p}^{\mathcal{P}}(\underline{\xi}, \underline{x}(t+1)) - H_{N,p}^{\mathcal{P}}(\underline{\xi}, \underline{x}(t-1)) \\
 &= -\frac{1}{N} \sum_{i,j=1}^N J_{ij} \left[x_j(t+1) x_i(t) - x_j(t) x_i(t-1) \right] \\
 &= -\frac{1}{N} \sum_{i,j=1}^N J_{ij} x_j(t) \left[x_i(t+1) - x_i(t-1) \right], \text{ car } J_{ij} = J_{ji}, \\
 &= -\frac{1}{N} \sum_{i=1}^N \left(x_i(t+1) - x_i(t-1) \right) \sum_{j=1}^N J_{ij} x_j(t).
 \end{aligned}$$

Quand $x_i(t+1) \neq x_i(t-1)$, le signe de leur différence est le signe de $x_i(t+1)$. Et c'est aussi le signe du second facteur, par hypothèse. Donc $H_{N,p}^{\mathcal{P}}(\underline{\xi}, \cdot)$ est une fonction de Lyapounov. Tant que le système évolue, c'est-à-dire qu'il existe un i tel que $x_i(t+1) \neq x_i(t-1)$, la fonction énergie décroît. Toutefois, il peut exister un ou plusieurs indices i tels que $\sum_{j=1}^N J_{ij} x_j(t) = 0$, $x_i(t+1) \neq x_i(t-1)$, et $H_{N,p}^{\mathcal{P}}(\underline{\xi}, \underline{x}(t+1)) - H_{N,p}^{\mathcal{P}}(\underline{\xi}, \underline{x}(t-1)) = 0$.

Mais cela signifie : $x_i(t-1) = -1$, $x_i(t+1) = 1$. Après un nombre fini (au plus N) de tels changements, on a : $x_i(t+1) = x_i(t-1)$, pour tout $i \in \{1, \dots, N\}$, d'où la proposition.♣

4.2 Étude asymptotique de la stabilité et de la convergence directe

Proposition 4.2.1 *les résultats du paragraphe III.2 restent vrais pour la dynamique parallèle déterministe, i.e. on peut remplacer \mathcal{U}_S par \mathcal{U}_P dans les théorèmes III.2.1, III.2.2, et III.2.4.*

Démonstration :

- Pour la stabilité des images originales, on peut reproduire mot pour mot la démonstration du théorème III.2.1.

- Pour la convergence directe, il suffit de noter que

$$\begin{aligned} \left\{ \omega \in \Omega : \mathcal{U}_P(\underline{\xi}(\omega), \underline{y}) \neq \underline{\xi}^\nu(\omega) \right\} &= \bigcup_{i=1}^N \left\{ \omega \in \Omega : \pi_i \circ \mathcal{U}_P(\underline{\xi}(\omega), \underline{y}) \neq \xi_i^\nu(\omega) \right\} \\ &= \bigcup_{i=1}^N \left\{ \omega \in \Omega : \operatorname{sgn} \left(\sum_{j=1}^N J_{ij} y_j \right) \neq \xi_i^\nu(\omega) \right\}. \end{aligned}$$

On peut donc utiliser, comme pour la dynamique séquentielle, les lemmes III.2.3 et III.2.5, pour obtenir l'analogie des propositions citées, dans le cas de la dynamique parallèle.♣

Nous allons maintenant montrer un autre résultat de convergence directe. Dans [40], A.Patrick et V.Zagrebnov étudient l'évolution temporelle des paramètres de recouvrement $m_\mu(t) = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \underline{x}(t)$, $\mu \in \{1, \dots, p\}$, quand la configuration initiale est le vecteur $\{x_i(0)\}_{i=1, \dots, N}$ formé de N variables aléatoires indépendantes corrélées avec une seule image originale $\underline{\xi}^\nu$, i.e. $\mathbf{P}[x_i(0)\xi_i^\nu = 1] = \frac{1}{2}(1 + \delta_{\mu\nu}m_0)$, avec $m_0 \in]0, 1]$. Ceci est un problème délicat, qui n'est résolu que pour des versions modifiées du modèle de Hopfield : réseaux multicouches [34], réseaux extrêmement dilués à liaisons d'interaction J_{ij} asymétriques [14]. L'étude de la convergence de telles configurations vers l'image $\underline{\xi}^\nu$ est donc un problème intéressant pour le modèle de Hopfield classique.

Theorème 4.2.2 *Soient $m_0 \in]0, 1]$, $\epsilon > 0$, $p \leq (m_0 - \epsilon)\sqrt{\frac{N}{2 \log N}}$, et $\underline{x}^\nu(0)$ le vecteur formé de N v.a. indépendantes telles que*

$$\begin{aligned} \mathbf{P}[x_i(0)^\nu \xi_i^\nu = 1] &= \frac{1}{2}(1 + \delta_{\mu\nu}m_0), \\ \mathbf{P}[x_i(0)^\nu \xi_i^\nu = -1] &= \frac{1}{2}(1 - \delta_{\mu\nu}m_0). \end{aligned}$$

Alors :

$$\lim_{N \rightarrow \infty} \mathbf{P} \left[\mathcal{U}_{\mathcal{P}}(\underline{\xi}, \underline{x}^{\nu}(0)) = \underline{\xi}^{\nu} \right] = 1. \quad (4.1)$$

Démonstration : par définition, pour $i = 1, \dots, N$,

$$\begin{aligned} x_i(t=1) &= \operatorname{sgn} \left(\sum_{\mu=1}^p \xi_i^{\mu} \sum_{\substack{j=1 \\ j \neq i}}^N \xi_j^{\mu} x_j(t=0) \right) \\ &= \xi_i^{\nu} \operatorname{sgn} \left(K_{N,1} + K_{N,2} \right), \end{aligned}$$

$$\text{avec } \begin{cases} K_{N,1} = \sum_{j=1, j \neq i}^N X_j^{\nu} \\ K_{N,2} = \xi_i^{\nu} \sum_{\mu=1, \mu \neq \nu}^p \xi_i^{\mu} \sum_{j=1, j \neq i}^N X_j^{\nu} \\ X_j^{\mu} = x_j^{\nu}(0) \xi_j^{\mu}. \end{cases}$$

Nous devons majorer $\mathbf{P} [K_{N,1} + K_{N,2} < 0]$. Pour tout $\epsilon > 0$, nous avons :

$$\mathbf{P} [K_{N,1} + K_{N,2} < 0] \leq \mathbf{P} [K_{N,1} < (m_0 - \epsilon)N] + \mathbf{P} [K_{N,2} < (-m_0 + \epsilon)N].$$

• $K_{N,1}$ est une somme de v.a. indépendantes valant ± 1 , d'espérance m_0 . Par le lemme III.3.9, il existe donc une constante strictement positive $h(m_0, \epsilon)$, ne dépendant pas de N telle que

$$\mathbf{P} [K_{N,1} < (m_0 - \epsilon)N] \leq \exp(-Nh(m_0, \epsilon)).$$

• $K_{N,2}$ n'est pas une somme de v.a. indépendantes, car les v.a. $(X_j^{\mu})_{\mu=1, \dots, p}$ ne sont pas nécessairement indépendantes. Cependant, nous pouvons écrire :

$$\begin{aligned}
 \mathbf{P}[K_{N,2} < (-m_0 + \epsilon)N] &= \mathbf{P}[-K_{N,2} > (m_0 - \epsilon)N] \\
 &\leq \sum_{\substack{\mu=1 \\ \mu \neq \nu}}^p \mathbf{P}\left[\sum_{\substack{j=1 \\ j \neq i}}^N \xi_i^\nu \xi_i^\mu X_j^\mu > (m_0 - \epsilon)\frac{N}{p}\right] \\
 &\leq \inf_{t>0} \sum_{\substack{\mu=1 \\ \mu \neq \nu \\ p}}^p e^{-t(m_0 - \epsilon)\frac{N}{p}} \mathbf{E} e^{t \sum_{\substack{j=1 \\ j \neq i}}^N \xi_i^\nu \xi_i^\mu X_j^\mu} \\
 &= \inf_{t>0} \sum_{\substack{\mu=1 \\ \mu \neq \nu \\ p}}^p e^{-t(m_0 - \epsilon)\frac{N}{p}} (\cosh t)^{N-1} \\
 &\leq (p-1) e^{-t(m_0 - \epsilon)\frac{N}{p} + N\frac{t^2}{2}} \\
 &\leq (p-1) e^{-N\frac{(m_0 - \epsilon)^2}{2p^2}}.
 \end{aligned}$$

Il est clair que :

$$\mathbf{P}[\mathcal{U}_p(\underline{\xi}, \underline{x}^\nu(0)) \neq \underline{\xi}^\nu] \leq \sum_{i=1}^N \mathbf{P}[x_i(t=1) \neq \xi_i^\nu].$$

La relation (4.1) est donc vérifiée si :

$$-\frac{N(m_0 - \epsilon)^2}{p^2} + \log N + \log p \xrightarrow{N \rightarrow \infty} -\infty.$$

Puisque $-\frac{N(m_0 - \epsilon)^2}{p^2} + \log N + \log p = -N\left(\frac{(m_0 - \epsilon)^2}{2p^2} - \frac{\log N}{N} - \frac{\log p}{N}\right)$, cette relation est vérifiée si $p < (m_0 - \epsilon)\sqrt{\frac{N}{2 \log N}}$. ♣

Malheureusement, nous n'avons pu montrer un résultat analogue quand le nombre d'images est proportionnel à $\frac{N}{\log N}$.

Chapitre 5

Étude de la dynamique stochastique

5.1 Description à volume fini

Aux chapitres III et IV, nous avons défini les dynamiques déterministes contrôlant l'évolution temporelle de la configuration du réseau. Nous avons notamment montré que partant d'un état initial \underline{x}_0 arbitraire, cette dynamique conduit nécessairement à une configuration minimisant au moins localement le hamiltonien du système. Quand un tel état est atteint, la dynamique déterministe interdit toute transition vers un autre état (la dynamique parallèle est un peu particulière puisqu'elle autorise une oscillation entre deux états terminaux).

Si \underline{x}_0 est suffisamment proche d'une image originale $\underline{\xi}^\nu$, la dynamique conduit à une configuration proche de $\underline{\xi}^\nu$, minimisant localement la fonction énergie, d'où la propriété de *mémoire associative*. Cependant, comme le révèlent le paragraphe III.3.2, relatif aux images combinées, il existe d'autres minima locaux, points fixes potentiels pour la dynamique déterministe. Si $\frac{p}{N} \rightarrow 0$, nous savons que ces minima ne sont pas plus profonds que ceux localisés près des images originales. Et pour p proportionnel à N , des arguments heuristiques et des simulations (voir [4]) laissent présager que ces derniers sont en fait les uniques minima globaux du hamiltonien. Il est donc naturel de s'intéresser à une évolution temporelle conduisant aux attracteurs de plus basse énergie.

Une méthode pour éviter que la configuration du réseau ne reste piégée en un minimum local inintéressant consiste à adopter une dynamique stochastique, *la dynamique de Glauber*, qui autorise, avec une probabilité non nulle, une transition d'un état \underline{x}_1 vers un état \underline{x}_2 , même si $H_{N,p}(\underline{\xi}^\nu, \underline{x}_2) > H_{N,p}(\underline{\xi}^\nu, \underline{x}_1)$.

Il convient maintenant de distinguer le cas séquentiel du cas parallèle.

5.1.1 La dynamique séquentielle de Glauber

Rappelons qu'à un instant t donné, un seul spin, déterminé par l'application de balayage, peut être modifié. La *dynamique d'évolution de Glauber* est alors définie par une chaîne de Markov (sur l'espace des configurations), dont la matrice de transition est donnée par :

$$P = \prod_{i=0}^{N-1} P_{N-i} = P_N P_{N-1} \dots P_1, \quad (5.1)$$

où P_i est la matrice $2^N \times 2^N$ suivante (on suppose les éléments de X_N numérotés de 1 à 2^N) :

$$P_i(\underline{x}, \underline{y}) = \begin{cases} \frac{\exp(\beta h_i(\underline{x}) y_i)}{\exp(\beta h_i(\underline{x})) + \exp(-\beta h_i(\underline{x}))}, & \text{si } y_j = x_j, \text{ pour } j \neq i, \\ 0, & \text{sinon,} \end{cases}$$

et

$$h_i(\underline{x}) = \sum_{j=1}^N J_{ij} x_j = \sum_{\substack{j=1 \\ j \neq i}}^N \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu x_j.$$

β est un paramètre strictement positif fixé; $T = \beta^{-1}$ est appelé *température* en mécanique statistique. D'une certaine manière, il mesure le degré d'agitation du système.

Remarques : - pour $\beta \nearrow \infty$, i.e. $T \searrow 0$, on retrouve la dynamique déterministe, puisque

$$P_i(\underline{x}, \underline{y}) \xrightarrow{\beta \nearrow \infty} \begin{cases} 1, & \text{si } y_i h_i(\underline{x}) > 0, \\ 0, & \text{sinon,} \end{cases}$$

c'est-à-dire

$$y_i = \text{sgn}(h_i(\underline{x})).$$

- A l'opposé, pour $\beta \searrow 0$, i.e. $T \nearrow \infty$, l'agitation thermique désordonne le système :

$$P_i(\underline{x}, \underline{y}) \xrightarrow{\beta \searrow 0} \begin{cases} \frac{1}{2}, & \text{si } \underline{y} = (x_1, \dots, x_{i-1}, \pm 1, x_{i+1}, \dots, x_N), \\ 0, & \text{sinon.} \end{cases}$$

Toute l'information est perdue, et le système évolue de manière totalement erratique.

Le produit de matrices de l'équation (5.1) signifie simplement que pour une condition initiale $\underline{x}(0)$ donnée, l'état du réseau à un instant t entier est déterminé par la probabilité conditionnelle :

$$\begin{aligned}
& \mathbf{P} \left[\underline{x}(t) = \underline{y} \mid \underline{x}(t-1) = \underline{x} \right] \\
= & \sum_{\underline{z}_1, \dots, \underline{z}_{N-1} \in X_N} \mathbf{P} \left[\underline{x}(t) = \underline{y} \mid \underline{x}(t - \frac{1}{N}) = \underline{z}_{N-1} \right] \mathbf{P} \left[\underline{x}(t - \frac{1}{N}) = \underline{z}_{N-1} \mid \underline{x}(t - \frac{2}{N}) = \underline{z}_{N-2} \right] \dots \\
& \dots \mathbf{P} \left[\underline{x}(t - 1 + \frac{1}{N}) = \underline{z}_1 \mid \underline{x}(t) = \underline{x} \right] \\
= & \sum_{\underline{z}_1, \dots, \underline{z}_{N-1} \in X_N} P_N(\underline{z}_{N-1}, \underline{y}) P_{N-1}(\underline{z}_{N-2}, \underline{z}_{N-1}) \dots P_1(\underline{x}, \underline{z}_1) \\
= & P_N \dots P_1(\underline{x}, \underline{y}).
\end{aligned}$$

Le balayage étant séquentiel, nous avons supposé implicitement que chaque spin est susceptible d'être modifié aux instants $\{t + \frac{i}{N}\}_{t \in \mathbf{N}}$, pour $i = 1, \dots, N$, et cela suivant la probabilité conditionnelle

$$\mathbf{P} \left[\underline{x}(t + \frac{i}{N}) = \underline{y} \mid \underline{x}(t + \frac{i-1}{N}) = \underline{x} \right] = P_i(\underline{x}, \underline{y}).$$

Nous allons maintenant rappeler et appliquer un résultat classique d'ergodicité concernant les chaînes de Markov (voir [18]). Considérons une chaîne de Markov sur l'espace des états $\{x_1, \dots, x_r\}$, $r \geq 2$, définie par la *matrice stochastique de transition*

$$(p_{ij} = p(x_i, x_j))_{i,j=1,\dots,r}, \text{ i.e. } 0 \leq p_{ij} \leq 1, \text{ et } \sum_{j=1}^r p_{ij} = 1, \text{ pour chaque } i \in \{1, \dots, r\}.$$

Définition 5.1.1 (i) La matrice p est dite *irréductible* si pour chaque paire (x_i, x_j) , $i \neq j$,

ou bien $p_{ij} > 0$,

ou bien il existe une suite $x_{i_1} = x_i, x_{i_2}, \dots, x_{i_k} = x_j$, avec $k \geq 2$, telle que $p_{i_l i_{l+1}} > 0$, pour chaque $l \in \{1, \dots, k-1\}$.

(ii) La matrice p est dite *apériodique* si pour chaque $i \in \{1, \dots, r\}$, l'ensemble des entiers positifs n tels que $(p^n)_{ii} > 0$ est non vide et a pour p.g.c.d. 1.

Theorème 5.1.2 (i) Une matrice stochastique p est irréductible et apériodique si et seulement si il existe un entier $M \geq 1$ tel que pour chaque (i, j) , $(p^M)_{ij} > 0$.

(ii) Si p est une matrice $r \times r$ irréductible et apériodique, alors la limite $\lim_{n \rightarrow \infty} (p^n)_{ij} = \nu_j$ existe, et est indépendante de i . Chaque ν_j est positive, et $\{\nu_j\}_{j=1,\dots,r}$ est l'unique solution des équations

$$\begin{cases} \sum_{i=1}^r \nu_i p_{ij} = \nu_j \\ \sum_{i=1}^r \nu_i = 1. \end{cases}$$

Nous appelons $\nu = \sum_{i=1}^r \nu_i \delta_{x_i}$ la mesure invariante correspondant à p .

Nous allons maintenant appliquer ces résultats à la chaîne de Markov définie précédemment.

Proposition 5.1.3 *la mesure de Gibbs à volume fini sur l'espace des configurations, définie par*

$$\mathcal{G}_{N,p,\beta}(\underline{\xi}, \cdot) = \frac{1}{Z_{N,p,\beta}(\underline{\xi})} \sum_{\underline{x} \in X_N} e^{-\beta H_{N,p}^S(\underline{\xi}, \underline{x})} \delta_{\underline{x}}(\cdot),$$

(où $Z_{N,p,\beta}(\underline{\xi})$ est la constante de renormalisation, appelée fonction de partition à volume fini), est l'unique mesure invariante associée à P .

Remarque : en fait, il faudrait écrire $\left\{ \mathcal{G}_{N,p,\beta}(\underline{\xi}(\omega), \cdot) \right\}_{\omega \in \Omega}$: la mesure de Gibbs associée au hamiltonien aléatoire est une mesure de probabilité *aléatoire* relativement à l'espace $(\Omega, \mathcal{F}, \mathbf{P})$.

Démonstration : il est clair que pour chaque $\omega \in \Omega$ fixé, P est irréductible et apériodique. Il suffit donc de vérifier que $\mathcal{G}_{N,p,\beta}(\underline{\xi}, \cdot)$ est vecteur propre gauche de P , i.e. $\forall \underline{y} \in X_N$,

$$\sum_{\underline{x} \in X_N} \mathcal{G}_{N,p,\beta}(\underline{\xi}, \underline{x}) P(\underline{x}, \underline{y}) = \mathcal{G}_{N,p,\beta}(\underline{\xi}, \underline{y}).$$

En fait, par définition de P , il suffit de montrer que pour chaque $i \in \{1, \dots, N\}$,

$$\sum_{\underline{x} \in X_N} \mathcal{G}_{N,p,\beta}(\underline{\xi}, \underline{x}) P_i(\underline{x}, \underline{y}) = \mathcal{G}_{N,p,\beta}(\underline{\xi}, \underline{y}).$$

Posons $\Lambda_i(\underline{y}) = \{\underline{x} \in X_N : x_j = y_j, \text{ pour } j \neq i\}$. Nous avons, par définition,

$$\begin{aligned}
\sum_{\underline{x} \in X_N} \mathcal{G}_{N,p,\beta}(\underline{\xi}, \underline{x}) P_i(\underline{x}, \underline{y}) &= \frac{1}{Z_{N,p,\beta}(\underline{\xi})} \sum_{\underline{x} \in \Lambda_i(\underline{y})} \frac{\exp\left(-\beta H_{N,p}^S + \beta \sum_{j=1}^N J_{ij} x_j y_i\right)}{\exp\left(\beta \sum_{j=1}^N J_{ij} x_j\right) + \exp\left(-\beta \sum_{j=1}^N J_{ij} x_j\right)} \\
&= \frac{1}{Z_{N,p,\beta}(\underline{\xi})} \sum_{\underline{x} \in \Lambda_i(\underline{y})} \frac{\exp\left(\beta \sum_{j,k=1}^N J_{kj} x_k x_j + \sum_{j=1}^N J_{ij} x_j y_i\right)}{\exp\left(\beta \sum_{j=1}^N J_{ij} x_j\right) + \exp\left(-\beta \sum_{j=1}^N J_{ij} x_j\right)} \\
&= \frac{1}{Z_{N,p,\beta}(\underline{\xi})} \sum_{\underline{x} \in \Lambda_i(\underline{y})} \frac{\exp\left(\beta \sum_{j=1}^N J_{ij} x_i x_j + \sum_{\substack{j,k=1 \\ k \neq i}}^N J_{kj} x_k x_j + \sum_{j=1}^N J_{ij} x_j y_i\right)}{\exp\left(\beta \sum_{j=1}^N J_{ij} x_j\right) + \exp\left(-\beta \sum_{j=1}^N J_{ij} x_j\right)} \\
&= \frac{1}{Z_{N,p,\beta}(\underline{\xi})} \sum_{x_i = \pm 1} \frac{\exp\left(\beta \sum_{j=1}^N J_{ij} x_i y_j\right)}{\exp\left(\beta \sum_{j=1}^N J_{ij} x_j\right) + \exp\left(-\beta \sum_{j=1}^N J_{ij} x_j\right)} \\
&\quad \times \exp\left(\sum_{\substack{j,k=1 \\ k \neq i}}^N J_{kj} y_k y_j + \sum_{j=1}^N J_{ij} y_j y_i\right) \\
&= \frac{1}{Z_{N,p,\beta}(\underline{\xi})} \exp\left(\sum_{j,k=1}^N J_{kj} y_k y_j\right) \\
&= \mathcal{G}_{N,p,\beta}(\underline{\xi}, \underline{y}). \spadesuit
\end{aligned}$$

Remarques : - Contrairement à la dynamique déterministe, la situation à l'équilibre n'est pas statique. Il s'agit d'un "équilibre dynamique", au sens suivant : la probabilité d'atteindre l'état \underline{y} , étant dans l'état \underline{x} , est identique à la probabilité d'atteindre l'état \underline{x} , étant dans l'état \underline{y} . Ainsi, la situation à l'équilibre est décrite par une mesure de probabilité sur l'espace des configurations, la mesure de Gibbs à volume fini.

- Le résultat précédent implique que la distribution au temps t du réseau converge, quand $t \rightarrow \infty$, vers la mesure $\mathcal{G}_{N,p,\beta}(\underline{\xi}, \cdot)$, et ceci indépendamment de l'état initial. Ceci exclut donc toute possibilité de mémoire associative, puisque la condition initiale est oubliée. La seule alternative est de considérer la limite thermodynamique (i.e. $N \rightarrow \infty$) de ce système, pour laquelle la situation suivante peut se produire : quand la taille de la matrice P augmente, la différence entre les deux plus grandes valeurs propres diminue. Ainsi, pour une matrice infiniment grande (i.e. un système thermodynamique), apparaît une dégénérescence asymptotique : le temps nécessaire pour atteindre la probabilité d'équilibre, en fait le temps nécessaire pour que la plus grande valeur propre domine la dynamique et efface le souvenir de la condition initiale, devient de plus en plus grand. L'étude de la limite thermodynamique du modèle de Hopfield est l'objet du paragraphe V.2.

5.1.2 La dynamique parallèle de Glauber

L'étude de la dynamique stochastique parallèle peut être menée de manière analogue à celle de la dynamique séquentielle. Maintenant, à chaque instant entier t , tous les spins sont susceptibles de changer de valeur.

La dynamique parallèle à température $T = \frac{1}{\beta} > 0$ fixée est définie par les probabilités conditionnelles suivantes :

$$\mathbf{P} \left[\underline{x}(t) = \underline{y} \mid \underline{x}(t-1) = \underline{x} \right] = \prod_{i=1}^N \mathbf{P} \left[x_i(t) = y_i \mid \underline{x}(t-1) = \underline{x} \right],$$

avec, pour $i = 1, \dots, N$,

$$\begin{aligned} \mathbf{P} \left[x_i(t) = y_i \mid \underline{x}(t-1) = \underline{x} \right] &= \frac{\exp(\beta h_i(\underline{x}) y_i)}{\exp(\beta h_i(\underline{x})) + \exp(-\beta h_i(\underline{x}))} \\ &= \frac{1}{2} \left[1 + \tanh(\beta h_i(\underline{x}) y_i) \right]. \end{aligned}$$

Remarque : ici encore, pour $\beta \nearrow \infty$, on retrouve la dynamique déterministe (de Little).

Supposons les vecteurs de X_N numérotés arbitrairement de 1 à 2^N . La matrice $\left(Q(\underline{x}, \underline{y}) \right)_{\underline{x}, \underline{y} \in X_N}$ telle que

$$Q(\underline{x}, \underline{y}) = \mathbf{P} \left[\underline{x}(1) = \underline{y} \mid \underline{x}(0) = \underline{x} \right]$$

définit une chaîne de Markov homogène, irréductible et apériodique sur l'espace des configurations.

Proposition 5.1.4 *Soit $\beta > 0$, et*

$$\begin{aligned} E_{N,p,\beta}^{\mathcal{P}} : \Omega \times X_N &\longrightarrow \mathbf{R} \\ (\omega, \underline{x}) &\longmapsto E_{N,p,\beta}^{\mathcal{P}}(\underline{\xi}(\omega), \underline{x}) = -\frac{1}{\beta} \sum_{i=1}^N \log 2 \cosh \left(\beta \sum_{j=1}^N J_{ij} x_j \right). \end{aligned}$$

La mesure (aléatoire) de Gibbs à volume fini sur l'espace des configurations, définie par :

$$\mathcal{G}_{N,p,\beta}^{\mathcal{P}}(\underline{\xi}, \cdot) = \frac{1}{Z_{N,p,\beta}^{\mathcal{P}}(\underline{\xi})} \sum_{\underline{x} \in X_N} e^{-\beta E_{N,p,\beta}^{\mathcal{P}}(\underline{\xi}, \underline{x})} \delta_{\underline{x}}(\cdot),$$

(où $Z_{N,p,\beta}^P(\underline{\xi})$ est la constante de renormalisation, appelée fonction de partition à volume fini), est l'unique mesure invariante associée à Q .

Démonstration : il suffit de vérifier que pour tous $\underline{x}, \underline{y} \in X_N$,

$$\sum_{\underline{x} \in X_N} \mathcal{G}_{N,p,\beta}^P(\underline{\xi}, \underline{x}) Q(\underline{x}, \underline{y}) = \mathcal{G}_{N,p,\beta}^P(\underline{\xi}, \underline{y}).$$

Par définition, nous avons

$$\begin{aligned} \sum_{\underline{x} \in X_N} \mathcal{G}_{N,p,\beta}^P(\underline{\xi}, \underline{x}) Q(\underline{x}, \underline{y}) &= \frac{1}{Z_{N,p,\beta}^P(\underline{\xi})} \sum_{\underline{x} \in X_N} \prod_{i=1}^N \left\{ 2 \cosh\left(\beta \sum_{j=1}^N J_{ij} x_j\right) \right\} \frac{\exp\left(\sum_{i=1}^N \beta h_i(\underline{x}) y_i\right)}{\prod_{i=1}^N 2 \cosh(\beta h_i(\underline{x}))} \\ &= \frac{1}{Z_{N,p,\beta}^P(\underline{\xi})} \sum_{\underline{x} \in X_N} \exp\left(\beta \sum_{i,j=1}^N J_{ij} x_j y_i\right) \\ &= \frac{1}{Z_{N,p,\beta}^P(\underline{\xi})} \prod_{j=1}^N \sum_{x_j = \pm 1} \exp\left(\beta x_j \sum_{i=1}^N J_{ij} y_i\right) \\ &= \frac{1}{Z_{N,p,\beta}^P(\underline{\xi})} \prod_{j=1}^N 2 \cosh\left(\beta \sum_{i=1}^N J_{ij} y_i\right) \\ &= \frac{1}{Z_{N,p,\beta}^P(\underline{\xi})} \prod_{j=1}^N 2 \cosh\left(\beta \sum_{j=1}^N J_{ij} y_j\right), \text{ car } J_{ij} = J_{ji}, \\ &= \mathcal{G}_{N,p,\beta}^P(\underline{\xi}, \underline{y}). \clubsuit \end{aligned}$$

5.2 Description à volume infini

Nous avons montré précédemment que pour obtenir une propriété de mémoire associative pour la dynamique stochastique ($T > 0$), il est nécessaire de considérer la limite thermodynamique de ce système. Bien que n'ayant pas travaillé sur ce problème, nous allons rappeler quelques définitions et propriétés générales des mesures définies sur l'ensemble des configurations à volume infini, puis nous citerons les résultats obtenus récemment par A.Bovier, V.Gaynard et P.Picco, concernant l'existence d'états de Gibbs pour $\frac{p}{N} \rightarrow 0$ [10].

•Formalisme de la mécanique statistique

Pour un système fini, nous avons montré l'unicité de la probabilité invariante, qui se traduit en mécanique statistique par l'absence de transition de phase. Dans un cadre plus général, l'intérêt de la limite thermodynamique réside dans le fait qu'il existe plusieurs mesures limites, appelées *états de Gibbs*, correspondant chacun à une phase. L'étude des

transitions de phases pour toute température est en général un problème difficile (consulter par exemple le livre [21]). La description des états de Gibbs pour les modèles de mécanique statistique peut être envisagée de plusieurs manières : pour les “bonnes” interactions (par exemple les interactions à courte portée), il s’agit des solutions des équations DLR, dues à Dobrushin, Landford et Ruelle (voir [42],[21]). Ces solutions coïncident avec l’enveloppe fermée convexe de toutes les limites faibles des mesures à volume fini, notion introduite par Minlos et Ruelle [35],[43]. Par contre, pour les interactions de type champ moyen, comme le modèle de Hopfield, où les interactions dépendent de tout le volume et ne sont pas définies pour le système infini, on ne peut définir les états de Gibbs via les équations DLR. Il faut alors exploiter la notion de limites faibles, que nous allons maintenant expliciter.

• **L’espace X des configurations à volume infini, et les mesures de probabilités sur X** ([16],[15],[11],[3])

Soit $X_0 = \{-1, +1\}$ l’espace des états. Nous définissons l’espace produit $X = \prod_{i \in \mathbf{N}} X_i$, où les X_i sont des copies de X_0 , et \mathbf{N} est l’ensemble des entiers naturels.

Pour $\Lambda \subset \mathbf{N}$, de cardinalité k finie, nous définissons $X_\Lambda = X_{i_1} \times \dots \times X_{i_k}$, si $\Lambda = \{i_1, \dots, i_k\}$, et la projection

$$\begin{aligned} \pi_\Lambda : X &\longrightarrow X_\Lambda \\ \underline{x} &\longmapsto \underline{x}^\Lambda = \{x_i\}_{i \in \Lambda} \end{aligned}$$

Soit $I_n = \{i_1, \dots, i_n\} \subset \mathbf{N}$. Alors $C_{I_n}(B_n) = \{\underline{x} \in X : \{x_i\}_{i \in I_n} \in B_n \subset X_{I_n}\}$ est un cylindre dans X de base B_n et de support I_n . Par définition, $C_{I_{n+1}}(B_n \times X_{i_{n+1}}) = C_{I_n}(B_n)$, et la réunion et l’intersection de cylindres est encore un cylindre. Si nous notons Σ la σ -algèbre engendrée par l’ensemble des cylindres $C(X)$, alors (X, Σ) est un espace mesurable.

Soit X_0 muni de la topologie discrète τ_0 . Alors X peut être muni de la topologie produit $\tau = \prod_{i \in \mathbf{N}} \tau_i$, i.e. la topologie la plus faible rendant continues les projections $\pi_i : X \longrightarrow X_i$, $i \in \mathbf{N}$. La base de cette topologie est formée des ensembles

$$\left\{ \bigcap_{i \in I} (\pi_i^{-1}[U_i]) : U_i \in \tau_i \right\}_I,$$

où I décrit les ensembles finis de \mathbf{N} . Notons que la σ -algèbre $\mathcal{B}(X)$ engendrée par les ouverts de la topologie produit τ (i.e. la tribu borélienne de X) coïncide avec Σ .

Par le théorème de Tychonoff, X est un espace compact pour la topologie produit τ , car X_0 est compact pour τ_0 .

D’autre part, on peut définir une distance ρ sur X , par

$$\rho(\underline{x}, \underline{x}') = \sum_{i \in \mathbf{N}} \frac{1}{2^i} |x_i - x'_i|,$$

et la topologie définie sur X par ρ coïncide avec la topologie produit τ , faisant de (X, ρ) un *espace métrique compact*, et donc un espace *complet et séparable*.

Rappelons les résultats suivants :

Théorème de Stone-Weierstrass : Soit E un espace métrique compact, et $\mathcal{C}(E)$ l'espace de Banach des fonctions continues sur E , à valeurs réelles. Si une sous-algèbre A de $\mathcal{C}(E)$ contient les fonctions constantes et sépare les points de E , alors A est dense dans l'espace $\mathcal{C}(E)$.

Théorème de représentation de Riesz-Markov : Soit E un espace métrique compact. Alors toute forme linéaire positive $f \mapsto \lambda(f)$ définie sur $\mathcal{C}(E)$, a une représentation unique sous la forme $\lambda(f) = \int f d\mu$, où μ est une mesure bornée sur E .

Nous pouvons appliquer ces résultats : soit $\mathcal{C}(X)$ l'ensemble des fonctions continues, bornées sur X , à valeurs réelles, de norme $\|f\|_X = \sup_{\underline{x} \in X} |f(\underline{x})| < \infty$. L'ensemble des fonctions cylindriques $\mathcal{C}_c(X)$ formé des fonctions constantes et des combinaisons linéaires finies de fonctions caractéristiques $\sum_j \lambda_j \chi_{C_j}(\underline{x})$, où $C_j \in \mathcal{C}(X), \forall j$, est une sous algèbre de $\mathcal{C}(X)$, séparant les points de X . Donc $\mathcal{C}_c(X)$ est dense dans $\mathcal{C}(X)$ (théorème de Stone-Weierstrass).

Nous pouvons définir les mesures de probabilité sur X par l'intermédiaire du théorème de représentation de Riesz-Markov. Les mesures boréliennes de probabilité $\mathcal{M}(X)$ sont en correspondance bijective avec les formes linéaires \mathcal{L} positives sur $\mathcal{C}(X)$, de norme égale à 1 : pour tout $\lambda \in \mathcal{L}$, il existe une unique mesure μ sur $\mathcal{B}(X)$ telle que

$$\lambda(f) = \int_X f d\mu, \quad \forall f \in \mathcal{C}(X).$$

Pour pouvoir définir les états de Gibbs sur $(X, \mathcal{B}(X))$, nous considérons alors la topologie sur $\mathcal{M}(X)$, induite par la *topologie vague* (ou *topologie faible* *) sur le dual $\mathcal{C}^*(X)$: la topologie la plus faible rendant continues toutes les fonctions $\mathcal{F}_f : \lambda \mapsto \lambda(f), f \in \mathcal{C}(X)$. (notez que $\{\mathcal{F}_f\} \subset \mathcal{C}^{**}(X)$).

Un système fondamental de voisinages d'un point $Q \in \mathcal{M}(X)$ est donné par les ensembles de la forme

$$N_Q(f_1, \dots, f_k; \epsilon) = \left\{ P \in \mathcal{M}(X) : \left| \int_X f_i dP - \int_X f_i dQ \right| < \epsilon, f_i \in \mathcal{C}(X), i = 1, \dots, k \right\}$$

Ceci correspond en fait à la topologie de la convergence faible sur $\mathcal{M}(X)$, telle que

$$P_n \implies P \text{ si } \lim_{n \rightarrow \infty} \int f dP_n \xrightarrow{n \rightarrow \infty} \int f dP, \forall f \in \mathcal{C}(X).$$

Et nous avons le résultat suivant :

Proposition 5.2.1 $\mathcal{M}(X)$ est compact pour la topologie de la convergence faible, et $P_n \implies P$ dans $\mathcal{M}(X)$ si et seulement si $P_n(C) \rightarrow P(C)$ pour tout cylindre $C \in \mathcal{C}(X)$.

Remarque : la première partie de cette proposition résulte du théorème de Banach-Alaoglu-Bourbaki concernant la compacité de la boule unité de $\mathcal{C}^*(X)$ pour la topologie faible * ([11]), et la seconde partie du théorème de Stone-Weierstrass ([3]).

Nous pouvons maintenant décrire les états de Gibbs sur X , tels qu'ils furent définis par Minlos et Ruelle [35],[43].

$$\text{Pour } \bar{x} \in X, \text{ soit } X_{\bar{\Lambda}}(\bar{x}) = \left\{ x \in X : x_j = \bar{x}_j, j \in \bar{\Lambda} = \mathbf{N} \setminus \Lambda \right\}.$$

Pour tout $\Lambda \subset X$, $\text{card}(\Lambda) < \infty$, on se donne des interactions $I_{\Lambda} : X_{\Lambda} \rightarrow \mathbf{R}$. L'énergie $H_{\Lambda}(\underline{y} | \bar{x})$ d'une configuration \underline{y} sur une partie finie Λ , avec condition au bord $\bar{x}_{\bar{\Lambda}} = \pi_{\bar{\Lambda}}(\bar{x})$, est alors définie par :

$$H_{\Lambda}(\underline{y} | \bar{x}_{\bar{\Lambda}}) = \sum_{\Lambda' \cap \Lambda \neq \emptyset} I_{\Lambda'}(\underline{y} | \bar{x}_{\bar{\Lambda}}),$$

où $(\underline{y} | \bar{x}_{\bar{\Lambda}})$ est la configuration telle que

$$(\underline{y} | \bar{x}_{\bar{\Lambda}})_i = \begin{cases} y_i, & \text{si } i \in \Lambda, \\ (\bar{x}_{\bar{\Lambda}})_i, & \text{si } i \in \bar{\Lambda} = \mathbf{N} \setminus \Lambda. \end{cases}$$

On peut remarquer que \bar{x} intervient dans la définition de $H_{\Lambda}(\cdot | \bar{x}_{\bar{\Lambda}})$ uniquement pour les Λ' tels que $\Lambda' \not\subset \Lambda$.

Considérons une suite $(\Lambda_n)_n$, $\Lambda_n \nearrow \mathbf{N}$, et la suite P_n des mesures sur $\mathbf{B}(X)$, définies par

$$\mathbf{P}_n(A) = \mathbf{P}_{\Lambda_n, \bar{x}}(\pi_{\Lambda}(A \cap X_{\bar{\Lambda}}(\bar{x}))), \text{ pour } A \in \mathbf{B}(X),$$

où $\mathbf{P}_{\Lambda, \bar{x}}$ est la mesure de Gibbs (à volume fini Λ , et température β^{-1}), associée au hamiltonien H_{Λ} , avec condition au bord \bar{x} , c'est-à-dire

$$\mathbf{P}_{\Lambda, \bar{x}}(x_{\Lambda}) = \frac{\exp\{-\beta H_{\Lambda}(x_{\Lambda} | \bar{x}_{\bar{\Lambda}})\}}{\sum_{y \in X_{\Lambda}} \exp\{-\beta H_{\Lambda}(y | \bar{x}_{\bar{\Lambda}})\}}.$$

Définition 5.2.2 Nous disons qu'une mesure \mathcal{G} sur $\mathcal{B}(X)$ est un état de Gibbs (à volume infini) pour la suite des mesures de Gibbs à volume fini précédente, si \mathcal{G} appartient à l'enveloppe convexe fermée \mathcal{G}_β de l'ensemble des points d'accumulation (pour la topologie faible) de la suite $\{\mathbf{P}_{\Lambda, \bar{x}}(\cdot)\}_{\Lambda \subset \mathbf{N}}$ pour $\Lambda \nearrow \mathbf{N}$.

Remarque : pour les modèles de type champ moyen, on retrouve la même difficulté que pour les équations DLR, puisque le hamiltonien $H_\Lambda(x_\Lambda \mid \bar{x}_\Lambda)$ n'est pas défini si $\bar{x} \in X$. Dans ce cas, il faut imposer la condition extérieure identiquement nulle, à cause des interactions de portée infinie (voir [3] pour la définition des états de Gibbs correspondant au modèle de Curie-Weiss avec champ externe aléatoire).

• Résultats connus pour le modèle de Hopfield

L'existence des états de Gibbs pour p proportionnel à N est un problème ouvert. Cependant, des résultats intéressants, dus à A.Bovier, V.Gayrard, et P.Picco [10], existent pour $\frac{p}{N} \rightarrow 0$.

Soit \mathcal{M}_N l'application aléatoire

$$\begin{aligned} \mathcal{M}_N(\underline{\xi}, \cdot) : X_N &\longrightarrow \mathbf{R}^p \\ \underline{x} &\longmapsto \mathcal{M}_N(\underline{\xi}, \underline{x}) = (m_N^1(\underline{\xi}, \underline{x}), \dots, m_N^p(\underline{\xi}, \underline{x})), \end{aligned}$$

où m_n^μ est le paramètre (aléatoire) de recouvrement, défini par

$$m_n^\mu(\underline{\xi}, \underline{x}) = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu x_i, \quad \mu = 1, \dots, p.$$

Pour $\eta \in \mathbf{N}$, soit $\mathcal{G}_{N,p,\beta,h}^\eta(\underline{\xi}, \cdot)$ la mesure de probabilité aléatoire sur $(X_N, \mathbf{B}(X_N))$, telle que pour $\underline{x} \in X_N$,

$$\mathcal{G}_{N,p,\beta,h}^\eta(\underline{\xi}, \underline{x}) = \frac{1}{Z_{N,p,\beta,h}^\eta} e^{-\beta H_{N,p}^S(\underline{\xi}, \underline{x}) - \beta h \sum_{i=1}^N \xi_i^\eta x_i},$$

où $Z_{N,p,\beta,h}^\eta$ est la constante de renormalisation, et h est appelé *champ magnétique* associé à l'image $\underline{\xi}^\eta$.

Cette mesure ne dépend de \underline{x} que par l'intermédiaire des paramètres de recouvrement, puisque le hamiltonien peut s'écrire

$$H_{N,p}^S(\underline{\xi}, \underline{x}) = -N \sum_{\mu=1}^p (m_N^\mu(\underline{\xi}, \underline{x}))^2 + \frac{p}{N}.$$

Il est donc naturel de considérer les mesures $\mathcal{Q}_{N,p,\beta,h}^\eta(\underline{\xi}, \cdot)$ sur $(\mathbf{R}^p, \mathbf{B}(\mathbf{R}^p))$ induites par $\mathcal{G}_{N,p,\beta,h}^\eta(\underline{\xi}, \cdot)$ par la transformation \mathcal{M}_N , i.e.

$$\mathcal{Q}_{N,p,\beta,h}^\eta(\underline{\xi}, \cdot) = \mathcal{G}_{N,p,\beta,h}^\eta(\underline{\xi}, \mathcal{M}_N(\underline{\xi}, \cdot)^{-1}).$$

Avec ces notations, le résultat obtenu est le suivant :

Theorème 5.2.3 (Bovier, Gaynard, Picco) *soit p non décroissant et tel que $\frac{p}{N} \rightarrow 0$. Soit a^\pm la plus grande (resp. plus petite) solution de l'équation $a = \tanh(\beta a)$. Alors, pour tout $\beta > 0$,*

$$\lim_{h \rightarrow 0^\pm} \lim_{N \rightarrow \infty} \mathcal{Q}_{N,p,\beta,h}^\eta(\underline{\xi}, \cdot) = \delta_{a^\pm(\beta)\underline{e}^\eta}, \mathbf{P} - p.s. \quad (5.2)$$

où il s'agit de la convergence au sens faible; $\delta_{a^\pm(\beta)\underline{e}^\eta}$ désigne la masse de Dirac au point $a^\pm(\beta)\underline{e}^\eta$, et \underline{e}^η est le η -ème vecteur unité dans \mathbf{R}^N .

Remarques :- pour $\beta \leq 1$, $a^+(\beta) = a^-(\beta) = 0$, donc il existe une unique mesure limite à haute température.

- Pour $\beta > 1$, le champ magnétique associé à $\underline{\xi}^\eta$ a permis d'exhiber la mesure extrémale correspondant à cette image mémorisée.

Si maintenant on s'intéresse au cas $\frac{p}{N} \rightarrow \alpha > 0$, le résultat obtenu est plus faible. Pour $\delta > 0$, soit $a(\delta, \beta)$ la plus grande solution de l'équation $\delta a = \tanh(\beta a)$.

Posons

$$B_\rho^{(\nu,s)} = \left\{ \underline{x} \in \mathbf{R}^N : \|\underline{x} - sa(1 - 2\sqrt{\alpha}, \beta)\underline{e}^\eta\| < \rho \right\},$$

où $\|\cdot\|$ est la norme l^2 sur \mathbf{R}^N , et

$$B_\rho = \bigcup_{(\nu,s) \in \mathbf{N} \times \{-1,+1\}} B_\rho^{(\nu,s)}.$$

Avec ces notations, nous avons :

Theorème 5.2.4 (Bovier, Gaynard, Picco) *Il existe un $\alpha_0 > 0$, tel que si $\frac{p}{N} \rightarrow \alpha$, avec $\alpha \leq \alpha_0$, alors pour $\beta > 1 + 3\sqrt{\alpha}$, si $\rho^2 > C(a(1 - 2\sqrt{\alpha}, \beta))^{\frac{3}{2}} \alpha^{\frac{1}{8}} |\log \alpha|^{\frac{1}{4}}$, pour presque tout ω ,*

$$\lim_{N \rightarrow \infty} \mathcal{Q}_{N,p,\beta,h=0}(\underline{\xi}(\omega), B_\rho) = 1 \quad (5.3)$$

Ce théorème suggère l'existence de mesures de Gibbs extrémales ayant chacune pour support une des boules $B_\rho^{(s,\nu)}$.

Chapitre 6

Étude de l'énergie libre associée au modèle de Hopfield

6.1 Définition de l'énergie libre

Pour $\underline{\xi} = (\underline{\xi}^\mu)_{\mu=1,\dots,p}$, ensemble de p images mémorisées, en fait pN v.a.i.i.d. $(\xi_i^\mu)_{i=1,\dots,N}^{\mu=1,\dots,p}$, avec $\mathbf{P}[\xi_i^\mu = 1] = \mathbf{P}[\xi_i^\mu = -1] = \frac{1}{2}$, nous définissons l'énergie libre comme étant la variable aléatoire suivante :

$$F_{N,p,\beta} : \Omega \longrightarrow \mathbf{R} \\ \omega \longmapsto F_{N,p,\beta}(\underline{\xi}(\omega)) = -\frac{1}{\beta} \log Z_{N,p,\beta}(\underline{\xi}(\omega)),$$

où β est un paramètre strictement positif,

$$Z_{N,p,\beta}(\underline{\xi}(\omega)) = \sum_{\underline{x} \in X_N} e^{-\beta H_{N,p}(\underline{\xi}(\omega), \underline{x})}$$

est la *fonction de partition* du système, et

$$H_{N,p}(\underline{\xi}(\omega), \underline{x}) = -\frac{1}{2N} \sum_{\substack{i,j=1,\dots,N \\ i \neq j}} \sum_{\mu=1}^p \xi_i^\mu(\omega) \xi_j^\mu(\omega) x_i x_j$$

est le *hamiltonien* (aléatoire) associé à la *dynamique séquentielle*. (Le facteur $\frac{1}{2}$ ne figure pas dans le hamiltonien défini au chapitre III. Il ne change en rien l'étude menée ici, mais il est préférable de le faire figurer pour rester cohérent avec les articles cités au cours de ce chapitre).

L'énergie libre joue un rôle important en mécanique statistique : lorsqu'elle dépend d'autres paramètres que la température $T = \frac{1}{\beta}$, l'état d'équilibre correspond à un minimum de $F_{N,p,\beta}$. Et plusieurs quantités intéressantes sont déterminées à partir de l'énergie libre :

-Posons $T = \frac{1}{\beta}$; alors $S = \frac{\partial F_{N,p,\beta}}{\partial T}$ est l'entropie du système.

-L'espérance du hamiltonien par rapport à la mesure de Gibbs à volume fini :

$$\sum_{\underline{x} \in X_N} H_{N,p}(\underline{\xi}(\omega), \underline{x}) \mathcal{G}_{N,p,\beta}(\underline{\xi}(\omega), \underline{x}), \text{ notée } \left\langle H_{N,p}(\underline{\xi}(\omega), \cdot) \right\rangle_{\mathcal{G}_{N,p,\beta}},$$

peut s'exprimer de la manière suivante :

$$\left\langle H_{N,p}(\underline{\xi}(\omega), \cdot) \right\rangle_{\mathcal{G}_{N,p,\beta}} = - \frac{\partial F_{N,p,\beta}(\underline{\xi}(\omega))}{\partial \beta}.$$

-En ajoutant le terme auxiliaire $\sum_{i=1}^N h_i x_i$ au hamiltonien, où $\underline{h} = (h_1, \dots, h_N)$ est un vecteur de \mathbf{R}^N , on peut calculer la valeur moyenne (par rapport à la mesure de Gibbs à volume fini) de l'activité du neurone i :

$$\left\langle x_i \right\rangle_{\mathcal{G}_{N,p,\beta}} \stackrel{\text{déf}}{=} \sum_{\underline{x} \in X_N} x_i \mathcal{G}_{N,p,\beta}(\underline{\xi}, \underline{x}) = - \lim_{\underline{h} \rightarrow 0} \frac{\partial F_{N,p,\beta}(\underline{\xi})}{\partial h_i},$$

et la covariance entre les deux spins x_i et x_j :

$$\left\langle x_i x_j \right\rangle_{\mathcal{G}_{N,p,\beta}} - \left\langle x_i \right\rangle_{\mathcal{G}_{N,p,\beta}} \left\langle x_j \right\rangle_{\mathcal{G}_{N,p,\beta}} = \lim_{\underline{h} \rightarrow 0} \frac{\partial^2 F_{N,p,\beta}}{\partial h_i \partial h_j}.$$

Remarquons que ces quantités sont des v.a. pour l'espace $(\Omega, \mathcal{F}, \mathbf{P})$. La valeur moyenne d'autres quantités encore peut être calculée, pourvu qu'on sache les faire apparaître dans la fonction de partition.

Nous définissons naturellement l'énergie libre par site :

$$f_{N,p,\beta}(\underline{\xi}) = \frac{1}{N} F_{N,p,\beta}(\underline{\xi}).$$

Nous allons maintenant étudier les propriétés de cette variable aléatoire et en particulier nous pencher sur la question de l'existence de sa *limite thermodynamique* (i.e. $N \rightarrow \infty$). $f_{N,p,\beta}$ s'exprime notamment en fonction des paramètres p et β . Le comportement asymptotique de $f_{N,p,\beta}$ dépend fortement de la valeur de ces paramètres.

L'inégalité de Jensen, rappelée ci dessous, permet d'obtenir deux inégalités simples concernant l'énergie libre.

Proposition 6.1.1 (*Inégalité de Jensen*) Soit Y un vecteur aléatoire intégrable, défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$, à valeurs dans \mathcal{A} convexe de \mathbf{R}^k . Soit Φ une fonction convexe de \mathcal{A} dans \mathcal{R} . Alors la v.a. $\Phi(Y)$ est quasi-intégrable, i.e. $\mathbf{E}Y^+$, ou $\mathbf{E}Y^-$ est finie, et

$$\mathbf{E}\Phi(Y) \geq \Phi(\mathbf{E}Y).$$

(Si Φ est strictement convexe, l'inégalité est stricte dès que Y n'est pas p.s. constante).

Démonstration : voir [13]♣

Soit K la mesure équadistribuée sur $\{-1, +1\}$, et K_N la mesure produit associée sur $\{-1, +1\}^N$. Alors la fonction de partition s'écrit :

$$Z_{N,p,\beta}(\underline{\xi}) = 2^N \int_{X_N} e^{-\beta H_{N,p}(\underline{\xi}, \underline{x})} K_N(d\underline{x}),$$

donc par convexité de la fonction exponentielle,

$$Z_{N,p,\beta}(\underline{\xi}) \geq 2^N \exp\left\{-\beta \int_{X_N} H_{N,p}(\underline{\xi}, \underline{x}) K_N(d\underline{x})\right\} = 2^N,$$

et l'énergie libre par site est donc majorée, uniformément en ω, p , et N , par une constante :

$$f_{N,p,\beta}(\underline{\xi}(\omega)) \leq -\frac{1}{\beta} \log 2 \tag{6.1}$$

De même, puisque $(-\log)$ est une fonction convexe, on a :

$$\mathbf{E}f_{N,p,\beta}(\underline{\xi}) \geq \bar{f}_{N,p,\beta} \stackrel{\text{déf}}{=} -\frac{1}{\beta N} \log \mathbf{E}Z_{N,p,\beta}(\underline{\xi}) \tag{6.2}$$

$\bar{f}_{N,p,\beta}$, qui n'est pas une variable aléatoire, est appelée énergie libre "annealed" (*recuite*), alors que $f_{N,p,\beta}$ est parfois appelée énergie libre "quenched" (*tremppée*).

6.2 Étude de l'énergie libre "annealed"

L'étude de l'énergie libre "annealed" est plus simple que celle de l'énergie libre "quenched". Toutefois, son intérêt se limite à la région haute température, dès que p , le nombre d'images, est une fonction croissante en N .

Proposition 6.2.1 (i) Si $\beta < 1$, $\lim_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow \alpha \geq 0}} \bar{f}_{N,p,\beta} = -\frac{1}{\beta} \log 2 + \frac{\alpha}{2\beta} \log(1 - \beta) + \frac{\alpha}{2}$

(ii) Si $\beta > 1$, $\lim_{p,N \rightarrow \infty} \bar{f}_{N,p,\beta} = -\infty$
 $\lim_{\substack{N \rightarrow \infty \\ p \text{ fini}}} \bar{f}_{N,p,\beta} = -\frac{1}{\beta} \log 2 + \frac{p}{\beta} \inf_{x \in \mathbf{R}} \left\{ \frac{x^2}{2} - \log \cosh(\sqrt{\beta}x) \right\}.$

Démonstration : pour montrer cette proposition, nous écrivons la fonction de partition sous la forme d'une intégrale gaussienne p -dimensionnelle. Cette écriture sera également utile pour étudier l'énergie libre "quenched".

Pour k entier, nous notons $(\cdot, \cdot)_k$ le produit scalaire de \mathbf{R}^k .

La fonction de partition s'écrit :

$$Z_{N,p,\beta}(\underline{\xi}) = e^{-\frac{\beta p}{2}} \sum_{\underline{x} \in X_N} \prod_{\mu=1}^p \exp \left\{ \frac{\beta}{2N} (\underline{\xi}^\mu, \underline{x})_N^2 \right\}.$$

En utilisant la formule :

$$e^{\frac{x^2}{2a}} = \sqrt{\frac{a}{2\pi}} \int_{\mathbf{R}} e^{-\frac{a}{2}t^2 + tx} dt,$$

il vient alors :

$$Z_{N,p,\beta}(\underline{\xi}) = e^{-\frac{\beta p}{2}} \left(\frac{N}{2\pi} \right)^{\frac{p}{2}} \int_{\mathbf{R}^p} dm_1 \dots dm_p \sum_{\underline{x} \in X_N} \exp \left\{ -\frac{N}{2} \sum_{\mu=1}^p m_\mu^2 + \sqrt{\beta} \sum_{\mu=1}^p m_\mu \sum_{i=1}^N x_i \xi_i^\mu \right\}$$

On peut sommer en \underline{x} :

$$Z_{N,p,\beta}(\underline{\xi}) = e^{-\frac{\beta p}{2}} \left(\frac{N}{2\pi} \right)^{\frac{p}{2}} \int_{\mathbf{R}^p} d\underline{m} \exp \left\{ -\frac{N}{2} \|\underline{m}\|^2 + \sum_{i=1}^N \log 2 \cosh(\sqrt{\beta}(\underline{m}, \underline{\xi}_i)_p) \right\}$$

Le théorème de Fubini s'applique, et donne :

$$\mathbf{E} Z_{N,p,\beta}(\underline{\xi}) = 2^N e^{-\frac{\beta p}{2}} \left(\frac{N}{2\pi} \right)^{\frac{p}{2}} \int_{\mathbf{R}^p} d\underline{m} \exp \left\{ -\frac{N}{2} \|\underline{m}\|^2 \right\} \mathbf{E} \exp \left\{ \sum_{i=1}^N \log \cosh(\sqrt{\beta}(\underline{m}, \underline{\xi}_i)_p) \right\}$$

Par indépendance des v.a. ξ_i^μ , pour $i = 1, \dots, N$, $\mu = 1, \dots, p$,

$$\begin{aligned} \mathbf{E} \exp \left\{ \sum_{i=1}^N \log \cosh(\sqrt{\beta}(\underline{m}, \underline{\xi}_i)_p) \right\} &= \prod_{i=1}^N \mathbf{E} \cosh(\sqrt{\beta} \sum_{\mu=1}^p m_\mu \xi_i^\mu) \\ &= \prod_{i=1}^N \prod_{\mu=1}^p \cosh \sqrt{\beta} m_\mu, \end{aligned}$$

car

$$\begin{aligned} \cosh \left(\sqrt{\beta} \sum_{\mu=1}^k m_\mu \xi_i^\mu \right) &= \cosh \left(\sqrt{\beta} \sum_{\mu=1}^{k-1} m_\mu \xi_i^\mu \right) \cosh \left(\sqrt{\beta} m_k \xi_i^k \right) \\ &\quad + \sinh \left(\sqrt{\beta} \sum_{\mu=1}^{k-1} m_\mu \xi_i^\mu \right) \sinh \left(\sqrt{\beta} m_k \xi_i^k \right) \end{aligned}$$

et

$$\mathbf{E} \sinh \left(\sqrt{\beta} m_k \xi_i^k \right) = 0, \quad \forall k = 1, \dots, p.$$

On a donc

$$\begin{aligned} \mathbf{E} Z_{N,p,\beta}(\underline{\xi}) &= 2^N e^{-\frac{\beta p}{2}} \left(\frac{N}{2\pi} \right)^{\frac{p}{2}} \int_{\mathbf{R}^p} d\underline{m} \exp \left\{ -\frac{N}{2} \|\underline{m}\|^2 + N \sum_{\mu=1}^p \log \cosh(\sqrt{\beta} m_\mu) \right\} \\ &= 2^N e^{-\frac{\beta p}{2}} \left[\left(\frac{N}{2\pi} \right)^{\frac{1}{2}} \int_{\mathbf{R}} dm \exp \left\{ -\frac{N}{2} m^2 + N \log \cosh(\sqrt{\beta} m) \right\} \right]^p \end{aligned}$$

L'énergie libre "annealed" s'écrit alors :

$$\begin{aligned} \bar{f}_{N,p,\beta} &= -\frac{1}{\beta} \log 2 + \frac{p}{2N} - \frac{p}{2\beta N} - \frac{p}{2\beta N} \log \frac{N}{2\pi} \\ &\quad - \frac{p}{\beta N} \log \int_{\mathbf{R}} \exp \left\{ -N \left(\frac{m^2}{2} - \log \cosh(\sqrt{\beta} m) \right) \right\}, \end{aligned}$$

et pour N suffisamment grand, la méthode de Laplace (voir [24]) implique :

$$\frac{1}{N} \log \int_{\mathbf{R}} \exp \left\{ -N \left(\frac{m^2}{2} - \log \cosh(\sqrt{\beta} m) \right) \right\} = \inf_{x \in \mathbf{R}} \left\{ \frac{x^2}{2} - \log \cosh(\sqrt{\beta} x) \right\} + R_N,$$

avec $\lim_{N \rightarrow \infty} R_N = 0$.

- Il est facile de voir que pour $\beta > 1$,

$$\inf_{x \in \mathbf{R}} \left\{ \frac{x^2}{2} - \log \cosh(\sqrt{\beta}x) \right\} = c_\beta < 0.$$

Il existe donc un entier $N_0(\beta)$ tel que pour $N > N_0(\beta)$,

$$-\frac{1}{N} \log \int_{\mathbf{R}} \exp \left\{ -N \left(\frac{m^2}{2} - \log \cosh(\sqrt{\beta}m) \right) \right\} = c_\beta - R_N < 0,$$

et $\bar{f}_{N,p,\beta} = -\frac{1}{\beta} \log 2 + \frac{p}{2N} - \frac{p}{2\beta N} - \frac{p}{2\beta N} \log \frac{N}{2\pi} + \frac{p}{\beta} (c_\beta - R_N)$ diverge vers $-\infty$, dès que $p \xrightarrow{N \rightarrow \infty} \infty$.

- Pour $\beta < 1$, $\inf_{x \in \mathbf{R}} \left\{ \frac{x^2}{2} - \log \cosh(\sqrt{\beta}x) \right\} = 0$.

Il faut donc connaître la valeur de R_N pour pouvoir conclure. Nous allons pour cela utiliser une méthode plus précise. Puisque $\beta < 1$, il est possible d'écrire :

$$\begin{aligned} & \left(\frac{N}{2\pi} \right)^{\frac{p}{2}} \left[\int_{\mathbf{R}} dm \exp \left\{ -\frac{N}{2} m^2 + N \log \cosh(\sqrt{\beta}m) \right\} \right]^p \\ &= \left(\frac{N}{2\pi} \right)^{\frac{p}{2}} \left[\int_{\mathbf{R}} dm \exp \left\{ -\frac{N}{2} m^2 (1 - \beta) - \frac{N}{2} m^2 \beta + N \log \cosh(\sqrt{\beta}m) \right\} \right]^p \\ &= \left(\frac{1}{2\pi(1 - \beta)} \right)^{\frac{p}{2}} \left[\int_{\mathbf{R}} dm \exp \left\{ -\frac{m^2}{2} - \frac{\beta}{1 - \beta} \frac{m^2}{2} + N \log \cosh \left(\sqrt{\frac{\beta}{1 - \beta}} \frac{m}{\sqrt{N}} \right) \right\} \right]^p \end{aligned}$$

Donc

$$\begin{aligned} \bar{f}_{N,p,\beta} &= -\frac{1}{\beta} \log 2 + \frac{p}{2N} + \frac{p}{2\beta N} \log(1 - \beta) \\ &\quad - \frac{p}{\beta N} \log \left[\frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} dm \exp \left\{ -\frac{m^2}{2} - \frac{\beta}{1 - \beta} \frac{m^2}{2} + N \log \cosh \left(\sqrt{\frac{\beta}{1 - \beta}} \frac{m}{\sqrt{N}} \right) \right\} \right] \end{aligned}$$

Le terme entre crochets peut s'écrire : $\mathbf{E}_Z[Y_N]$, avec Z v.a. gaussienne de loi $\mathcal{N}(0, 1)$,

$$\text{et } Y_N(\omega) = \exp \left\{ -\frac{\beta}{1 - \beta} \frac{Z(\omega)^2}{2} + N \log \cosh \left(\sqrt{\frac{\beta}{1 - \beta}} \frac{Z(\omega)}{\sqrt{N}} \right) \right\}.$$

Pour tout ω , $0 \leq Y_N(\omega) \leq 1$, car $\log \cosh x \leq \frac{x^2}{2}$, et par la formule de Taylor

$$Y_N(\omega) = \exp \left\{ \frac{1}{N} \left(\frac{\beta}{1 - \beta} \right)^2 \int_0^{Z(\omega)} \frac{(Z(\omega) - t)^3}{6!} \left(\frac{4}{\cosh^2 t} - \frac{6}{\cosh^4 t} \right) dt \right\},$$

6.3. CONVERGENCE PRESQUE SÛRE DE L'ÉNERGIE LIBRE ET DE SON ESPÉRANCE, POUR

donc Y_N converge presque sûrement vers 1.

Par le théorème de convergence dominée, $\mathbf{E}_Z Y_N \xrightarrow{N \rightarrow \infty} 1$. On peut donc conclure :

$$\lim_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow \alpha \geq 0}} \bar{f}_{N,p,\beta} = -\frac{1}{\beta} \log 2 + \frac{\alpha}{2} + \frac{\alpha}{2\beta} \log(1 - \beta). \clubsuit$$

Remarque : en corollaire de la proposition 6.2.1, et des inégalités (6.1) et (6.2), il est facile de montrer la convergence en probabilité de l'énergie libre, dans la région haute température ($\beta < 1$), vers la constante $-\frac{1}{\beta} \log 2$, si $\frac{p}{N} \rightarrow 0 : \forall \epsilon > 0$,

$$\begin{aligned} \mathbf{P} \left[\left| f_{N,p,\beta} + \frac{1}{\beta} \log 2 \right| > \epsilon \right] &= \mathbf{P} \left[-f_{N,p,\beta} - \frac{1}{\beta} \log 2 > \epsilon \right], \text{ car } f_{N,p,\beta} \leq -\frac{1}{\beta} \log 2, \\ &\leq \frac{1}{\epsilon} \left(-\mathbf{E} f_{N,p,\beta} - \frac{1}{\beta} \log 2 \right) \\ &\leq \frac{1}{\epsilon} \left(-\bar{f}_{N,p,\beta} - \frac{1}{\beta} \log 2 \right) \xrightarrow[\frac{p}{N} \rightarrow 0]{N \rightarrow \infty} 0, \text{ pour } \beta < 1. \end{aligned}$$

L'objectif du paragraphe suivant est d'améliorer ce résultat.

6.3 Convergence presque sûre de l'énergie libre et de son espérance, pour $\frac{p}{N} \rightarrow 0$

Theorème 6.3.1 Soient $\alpha \in [0, 1[$, $\delta \in]0, 1[$, tels que $\delta - 4\sqrt{\alpha}(1 - \delta) > 0$.

$$\begin{aligned} \text{Soient } \Phi_\delta :]0, \infty[\times \mathbf{R} &\longrightarrow \mathbf{R} \\ (\beta, h) &\longmapsto \Phi_\delta(\beta, h) = -\frac{1}{\beta} \log \cosh(\beta h) + \frac{1-\delta}{2} h^2, \end{aligned}$$

$$\Psi^1(\beta, \alpha, \delta) = \min_{h \in \mathbf{R}} \Phi_\delta(\beta, h) - \frac{1}{\beta} \log 2 + \frac{\alpha}{2} + \frac{\alpha}{\beta} \log \left(\delta - 4\sqrt{\alpha}(1 - \delta) \right),$$

et

$$\Psi^2(\beta, \alpha) = \min_{h \in \mathbf{R}} \Phi_0(\beta, h) - \frac{1}{\beta} \log 2 + \frac{\alpha}{2}.$$

Alors, pour tout $\beta > 0$,

(i) Pour \mathbf{P} -presque tout $\omega \in \Omega$,

$$\Psi^1(\beta, \alpha, \delta) \leq \liminf_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow \alpha}} f_{N,p,\beta}(\underline{\xi}(\omega)) \leq \limsup_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow \alpha}} f_{N,p,\beta}(\underline{\xi}(\omega)) \leq \Psi^2(\beta, \alpha) \quad (6.3)$$

(ii)

$$\Psi^1(\beta, \alpha, \delta) \leq \liminf_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow \alpha}} \mathbf{E} f_{N,p,\beta}(\underline{\xi}) \leq \limsup_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow \alpha}} \mathbf{E} f_{N,p,\beta}(\underline{\xi}) \leq \Psi^2(\beta, \alpha) \quad (6.4)$$

En particulier, pour $\alpha = 0$, on obtient par continuité, en prenant $\delta \searrow 0$:

Corollaire 6.3.2 *Si $\alpha = \lim_{N \rightarrow \infty} \frac{p}{N} = 0$, alors pour tout $\beta > 0$, l'énergie libre converge presque sûrement vers une constante :*

$$f_{N,p,\beta}(\underline{\xi}) \xrightarrow{p.s., N \rightarrow \infty} \min_{h \in \mathbf{R}} \Phi_0(\beta, h) - \frac{1}{\beta} \log 2 \quad (6.5)$$

Et

$$\lim_{N \rightarrow \infty} \mathbf{E} f_{N,p,\beta}(\underline{\xi}) = \min_{h \in \mathbf{R}} \Phi_0(\beta, h) - \frac{1}{\beta} \log 2 \quad (6.6)$$

Remarques : 1. Cette constante est l'énergie libre du modèle de Curie Weiss, et il est facile de voir que :

$$\begin{aligned} &\text{pour } 0 < \beta \leq 1, \min_{h \in \mathbf{R}} \Phi_0(\beta, h) = \Phi_0(\beta, 0) = 0, \\ &\text{pour } \beta > 1, \min_{h \in \mathbf{R}} \Phi_0(\beta, h) = c(\beta) < 0. \end{aligned}$$

2. Au paragraphe V.4, nous montrons que la variance de l'énergie libre converge vers 0 quand $N \rightarrow \infty$, pour tout β . L'équation (6.5) et ce résultat, que nous admettons provisoirement, impliquent que l'énergie libre converge aussi dans $\mathbf{L}^2(\Omega, \mathcal{F}, \mathbf{P})$ vers la constante $\min_{h \in \mathbf{R}} \Phi_0(\beta, h) - \frac{1}{\beta} \log 2$.

3. Pour $\beta < 1$, les énergies libres “annealed” et “quenched” coïncident asymptotiquement (au sens de la convergence p.s. et dans \mathbf{L}^2), et la minoration (6.2) devient une égalité :

$$\lim_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow 0}} \frac{1}{N} \mathbf{E} \log Z_{N,p,\beta}(\underline{\xi}) = \lim_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow 0}} \frac{1}{N} \log \mathbf{E} Z_{N,p,\beta}(\underline{\xi}).$$

L'idée de la démonstration du théorème 6.3.1 est due à Koch [27]. Dans l'article original, seuls les encadrements (6.4) sont formulés. Or, il s'avère que les estimations obtenues impliquent sans plus de difficulté des estimations presque sûres. Nous allons maintenant démontrer en détail ce théorème, en précisant notamment quelques points non clairement justifiés par Koch.

• La démonstration de la **minoration** se fait en plusieurs étapes :

1. Écriture de l'énergie libre sous forme d'intégrale gaussienne p -dimensionnelle :

$$\text{Posons } S_N(\underline{\xi}, \underline{m}) = \frac{\|\underline{m}\|^2}{2} - \frac{1}{N} \sum_{i=1}^N \log \cosh\left(\sqrt{\beta}(\underline{\xi}, \underline{m})\right).$$

En reprenant la méthode utilisée au paragraphe précédent, nous pouvons écrire $Z_{N,p,\beta}$ et $f_{N,p,\beta}(\underline{\xi})$ sous la forme :

$$Z_{N,p,\beta}(\underline{\xi}) = e^{-\frac{\beta p}{2}} 2^N \left(\frac{N}{2\pi}\right)^{\frac{p}{2}} \int_{\mathbf{R}^p} d\underline{m} \exp\left\{-NS_N(\underline{\xi}, \underline{m})\right\} \quad (6.7)$$

$$f_{N,p,\beta} = -\frac{1}{\beta} \log 2 + \frac{p}{2N} - \frac{1}{\beta N} \log \left[\left(\frac{N}{2\pi}\right)^{\frac{p}{2}} \int_{\mathbf{R}^p} d\underline{m} \exp\left\{-NS_N(\underline{\xi}, \underline{m})\right\} \right] \quad (6.8)$$

2. Restriction du domaine d'intégration de l'intégrale p -dimensionnelle :

Nous utilisons ici des résultats dus à S.Albeverio, B.Tirozzi, et B.Zegarlinski [2].

Pour $0 < \epsilon < 1$, $\beta > 0$, soit $R_\epsilon = \frac{2\sqrt{\beta p}}{\epsilon}$,

et $C_N = \frac{1}{N} \exp\left\{-\beta \frac{1-\epsilon}{\epsilon^2} p(N-2) + p \log \frac{\epsilon}{(1-\epsilon)\sqrt{2\beta}}\right\}$, alors

$$\left| \frac{1}{N} \log \int_{\mathbf{R}^p} d\underline{m} \exp\left\{-NS_N(\underline{\xi}, \underline{m})\right\} - \frac{1}{N} \log \int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| \leq R_\epsilon]} \exp\left\{-NS_N(\underline{\xi}, \underline{m})\right\} \right| \leq C_N \quad (6.9)$$

Notation : $\chi_{\mathcal{A}}$ désigne la fonction indicatrice de l'ensemble \mathcal{A} .

De (6.9), on déduit trivialement :

$$f_{N,p,\beta}(\underline{\xi}) \geq -\frac{1}{\beta} \log 2 + \frac{p}{2N} + f_{N,p,\beta}^\epsilon(\underline{\xi}) - \frac{C_N}{\beta}, \quad (6.10)$$

où $f_{N,p,\beta}^\epsilon = -\frac{1}{\beta N} \log \left[\left(\frac{N}{2\pi}\right)^{\frac{p}{2}} \int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| \leq R_\epsilon]} \exp\left\{-NS_N(\underline{\xi}, \underline{m})\right\} \right]$

Pour obtenir la relation (6.10), il suffit d'effectuer une partition de l'unité, qui donne :

$$\begin{aligned} \frac{1}{N} \log \int_{\mathbf{R}^p} d\underline{m} \exp\left\{-NS_N(\underline{\xi}, \underline{m})\right\} &= \frac{1}{N} \log \int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| \leq R_\epsilon]} \exp\left\{-NS_N(\underline{\xi}, \underline{m})\right\} \\ &\quad + \frac{1}{N} \log \left[1 + \frac{\int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| > R_\epsilon]} \exp\left\{-NS_N(\underline{\xi}, \underline{m})\right\}}{\int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| \leq R_\epsilon]} \exp\left\{-NS_N(\underline{\xi}, \underline{m})\right\}} \right], \end{aligned}$$

et de montrer que
$$\frac{\int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| > R_\epsilon]} \exp\{-N S_N(\underline{\xi}, \underline{m})\}}{\int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| \leq R_\epsilon]} \exp\{-N S_N(\underline{\xi}, \underline{m})\}} \leq N C_N.$$

Cette dernière inégalité résulte des estimations suivantes, dont la démonstration est donnée en 5.

Lemme 6.3.3 *Pour tout $\epsilon \in]0, 1[$, et $R_\epsilon = \frac{2\sqrt{\beta p}}{\epsilon}$,*

$$(i) \int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| > R_\epsilon]} \exp\{-N S_N(\underline{\xi}, \underline{m})\} \leq \left(\frac{2\pi}{1-\epsilon}\right)^{\frac{p}{2}} \exp\left\{-\frac{1-\epsilon}{2} R_\epsilon^2 (N-1)\right\}$$

$$(ii) \int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| \leq R_\epsilon]} \exp\{-N S_N(\underline{\xi}, \underline{m})\} \geq \left(2\pi\beta p(1-\epsilon)\right)^{\frac{p}{2}} \exp\left\{-\frac{N}{2} \frac{1-\epsilon}{2} R_\epsilon^2\right\}$$

3. Etude et minoration de “l'énergie libre tronquée” $f_{N,p,\beta}^\epsilon$:

Considérons la matrice de recouvrement A , dont les coefficients sont donnés par

$$A_{\mu\nu} = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \xi_i^\nu - \delta_{\mu\nu}, \text{ pour } \mu, \nu = 1, \dots, p \text{ (} \delta_{\mu\nu} \text{ désigne le symbole de Kronecker).}$$

Par le changement de variable $\underline{m}' = \frac{1}{\sqrt{\beta}} \underline{m}$, dans (6.10), on a :

$$\begin{aligned} f_{N,p,\beta}^\epsilon(\underline{\xi}) &= -\frac{1}{\beta N} \log \left[\left(\frac{N\beta}{2\pi}\right)^{\frac{p}{2}} \int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| \leq \frac{R_\epsilon}{\sqrt{\beta}}]} \exp\left\{-N\beta \|\underline{m}\|^2 + \sum_{i=1}^N \log \cosh(\beta(\underline{\xi}_i, \underline{m})_p)\right\} \right] \\ &= -\frac{1}{\beta N} \log \left[\left(\frac{N\beta}{2\pi}\right)^{\frac{p}{2}} \int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| \leq \frac{R_\epsilon}{\sqrt{\beta}}]} \exp\left\{\frac{-N\beta}{2} \left(\underline{m}, (\delta I - (1-\delta)A)\underline{m}\right)_p - \beta \sum_{i=1}^N \Phi_\delta(\beta, (\underline{\xi}_i, \underline{m})_p)\right\} \right], \end{aligned}$$

d'où la minoration évidente :

$$f_{N,p,\beta}^\epsilon(\underline{\xi}) \geq \min_{h \in \mathbf{R}} \Phi_\delta(\beta, h) - \frac{1}{\beta N} \log \left[\left(\frac{N\beta}{2\pi}\right)^{\frac{p}{2}} \int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| \leq \frac{R_\epsilon}{\sqrt{\beta}}]} \exp\left\{\frac{-N\beta}{2} \left(\underline{m}, (\delta I - (1-\delta)A)\underline{m}\right)_p\right\} \right] \quad (6.11)$$

$$\begin{aligned} \text{Posons } U_N(\underline{\xi}) &\equiv U_N(\underline{\xi}, \delta, p, \beta, \epsilon) \\ &= \frac{1}{\beta N} \log \left[\left(\frac{N\beta}{2\pi}\right)^{\frac{p}{2}} \int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| \leq \frac{R_\epsilon}{\sqrt{\beta}}]} \exp\left\{\frac{-N\beta}{2} \left(\underline{m}, (\delta I - (1-\delta)A)\underline{m}\right)_p\right\} \right] \end{aligned}$$

Il s'agit maintenant de majorer $U_N(\underline{\xi})$, et $\mathbf{E}U_N(\underline{\xi})$.

Notation : pour A matrice $p \times p$, $\|A\| = \sup_{\substack{\underline{m} \in \mathbf{R}^p \\ \underline{m} \neq 0}} \frac{\|A\underline{m}\|}{\|\underline{m}\|}$.

Soit $\lambda > 0$. On suppose que δ est choisi tel que $\delta - \lambda(1 - \delta) > 0$. On écrit alors :

$$U_N(\underline{\xi}) = U_N(\underline{\xi})\chi_{[\omega: \|A(\underline{\xi}(\omega))\| > \lambda]} + U_N(\underline{\xi})\chi_{[\omega: \|A(\underline{\xi}(\omega))\| \leq \lambda]},$$

et on étudie les deux termes séparément.

- Si $\|A(\underline{\xi}(\omega))\| \leq \lambda$, $\left(\underline{m}, A\underline{m}\right)_p \leq \lambda \left(\underline{m}, \underline{m}\right)_p$, donc

$$\frac{-N\beta}{2} \left(\underline{m}, (\delta I - (1 - \delta)A)\underline{m}\right)_p \leq \frac{-N\beta}{2} (\delta - (1 - \delta)\lambda) (\underline{m}, \underline{m})_p.$$

En tenant compte de cette majoration, on peut alors calculer l'intégrale gaussienne, en omettant la restriction $[\|\underline{m}\| \leq \frac{R_\epsilon}{\sqrt{\beta}}]$, et on obtient :

$$U_N(\underline{\xi})\chi_{[\omega: \|A(\underline{\xi}(\omega))\| \leq \lambda]} \leq \frac{-p}{2\beta N} \log(\delta - (1 - \delta)\lambda) \quad (6.12)$$

- Pour $\|A(\underline{\xi}(\omega))\| > \lambda$, nous utilisons le lemme suivant :

Lemme 6.3.4 *Soit $\lambda > 0$, et K un entier strictement positif. Soit $\tilde{\lambda} = \lambda e^{-\frac{1}{K}}$. Si A est une matrice symétrique telle que $\|A\| > \lambda$, alors pour $n \geq K$,*

$$\|A - \tilde{\lambda}I\| \leq \tilde{\lambda}^{-2n+1} \text{tr } A^{2n} \quad (6.13)$$

Soit $\tilde{\lambda} = \lambda e^{-\frac{1}{K}}$, où K est un entier strictement positif.

$\tilde{\lambda} < \lambda$, donc $\delta - \tilde{\lambda}(1 - \delta) > \delta - \lambda(1 - \delta) > 0$.

$$\begin{aligned} \left(\underline{m}, (\delta I - (1 - \delta)A)\underline{m}\right)_p &= -\left(\underline{m}, (\delta - (1 - \delta)\tilde{\lambda})\underline{m}\right)_p + (1 - \delta)\left(\underline{m}, (A - \tilde{\lambda}I)\underline{m}\right)_p \\ &\leq -(\delta - \lambda(1 - \delta))\|\underline{m}\|^2 + (1 - \delta)\|A - \tilde{\lambda}I\|\|\underline{m}\|^2, \end{aligned}$$

d'où

$$\begin{aligned} U_N(\underline{\xi})\chi_{[\|A\| > \lambda]} &\leq \chi_{[\|A\| > \lambda]} \left\{ \frac{R_\epsilon^2}{2\beta} (1 - \delta)\|A - \tilde{\lambda}I\| \right. \\ &\quad \left. + \frac{1}{\beta N} \log \left\{ \left(\frac{N}{2\pi}\right)^{\frac{p}{2}} \int_{\mathbf{R}^p} d\underline{m} \exp\left(-\frac{N\beta}{2} (\delta - (1 - \delta)\lambda)\|\underline{m}\|^2\right) \right\} \right\}, \end{aligned}$$

c'est-à-dire, en utilisant (6.13), et en calculant l'intégrale :

$$U_N(\underline{\xi})\chi_{[\|A\|>\lambda]} \leq \frac{2(1-\delta)}{\epsilon^2} \tilde{\lambda}^{-2n+1} p \operatorname{tr} A^{2n} \chi_{[\|A\|>\lambda]} - \frac{p}{2\beta N} \log(\delta - \lambda(1-\delta)), \quad (6.14)$$

pour tout $n \geq K$.

De (6.12) et (6.14), on déduit :

$$U_N(\underline{\xi}) \leq -\frac{p}{\beta N} \log(\delta - \lambda(1-\delta)) + \frac{2(1-\delta)}{\epsilon^2} \tilde{\lambda}^{-2n+1} p \operatorname{tr} A^{2n} \chi_{[\|A\|>\lambda]} \quad (6.15)$$

4. Conclusion :

Pour pouvoir conclure, nous avons besoin d'une majoration de $\mathbf{E} \operatorname{tr} A^n$. La proposition suivante, intéressante en soi, nous servira également au paragraphe suivant.

Proposition 6.3.5 *Soit A la matrice de recouvrement, dont les coefficients sont définis par*

$$A_{\mu\nu} = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \xi_i^\nu - \delta_{\mu\nu}, \text{ pour } \mu, \nu = 1, \dots, p.$$

Alors, pour $n \geq 4$ entier,

$$\mathbf{E} \operatorname{tr} A^n \leq 4^n N^{-n} \max_{\substack{u, v \geq 1 \\ u+v \leq n+1, v \leq \frac{n}{2}}} (p+n^2)^u N^v \quad (6.16)$$

En particulier, si $p+n^2 \leq N$,

$$\mathbf{E} \operatorname{tr} A^n \leq 4^n \left(\frac{p+n^2}{N} \right)^{\frac{n}{2}} (p+n^2) \quad (6.17)$$

-Montrons la minoration de l'expression (6.4). De (6.15) et (6.17), on déduit :

$$\mathbf{E} U_N(\underline{\xi}) \leq -\frac{p}{\beta N} \log(\delta - \lambda(1-\delta)) + \frac{2(1-\delta)}{\epsilon^2} \tilde{\lambda}^{-2n+1} 4^{2n} p (p+4n^2) \left(\frac{p+4n^2}{N} \right)^n,$$

pour n tel que $p+4n^2 \leq N$.

$$\text{Posons } \begin{cases} n = K = [N^{\frac{1}{4}}] \text{ (partie entière de } N^{\frac{1}{4}}) \\ \tilde{\lambda} = 4(1 + n^{-\frac{1}{2}}) \left(\frac{p+4n^2}{N} \right)^{\frac{1}{2}}, \lambda = \tilde{\lambda} e^{\frac{1}{n}} \end{cases}$$

Alors :

$$\begin{aligned} \tilde{\lambda}^{-2n} 4^{2n} (p + 4n^2) \left(\frac{p+4n^2}{N} \right)^n &= (p + 4n^2) (1 + n^{-\frac{1}{2}})^{-2n} \\ &= (p + 4n^2) \exp \left\{ -2n \log \left(1 + \frac{1}{\sqrt{n}} \right) \right\} \\ &\leq (p + 4n^2) \exp \left\{ -2\sqrt{n} + 1 \right\}, \end{aligned}$$

car $\log(1 + u) \geq u - \frac{u^2}{2}, \forall u > 0$.

Donc

$$\mathbf{E}U_N(\underline{\xi}) \leq -\frac{p}{\beta N} \log(\delta - \lambda(1 - \delta)) + \frac{2(1 - \delta)}{\epsilon^2} \tilde{\lambda} p (p + 4n^2) \exp \left\{ -2\sqrt{n} + 1 \right\} \quad (6.18)$$

$$\text{où } \begin{cases} n = [N^{\frac{1}{4}}] \\ \tilde{\lambda} = 4(1 + N^{-\frac{1}{8}}) \left(\frac{p}{N} + \frac{4}{\sqrt{N}} \right)^{\frac{1}{2}} = 4\sqrt{\alpha} + \mathcal{O}(\sqrt{\frac{\alpha}{N}}). \end{cases}$$

Des relations (6.10),(6.11),(6.18), on déduit immédiatement le résultat cherché :

$$\liminf_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow \alpha}} \mathbf{E}f_{N,p,\beta}(\underline{\xi}) \geq \Psi^1(\beta, \alpha, \delta).$$

-Nous pouvons également, à partir de (6.15) et (6.17), déduire une bonne majoration de $U_N(\underline{\xi}(\omega))$, pour ω appartenant à un sous-ensemble $\tilde{\Omega}$ de Ω , de mesure 1. Il suffit pour celà de reprendre les mêmes hypothèses que ci-dessus,

$$\text{i.e. } \begin{cases} n = [N^{\frac{1}{4}}] \\ \tilde{\lambda} \equiv \tilde{\lambda}_N = 4(1 + n^{-\frac{1}{2}}) \left(\frac{p+4n^2}{N} \right)^{\frac{1}{2}} \\ \lambda \equiv \lambda_N = \tilde{\lambda}_N e^{\frac{1}{n}}. \end{cases}$$

Alors

$$U_N(\underline{\xi}) \leq -\frac{p}{\beta N} \log(\delta - \lambda_N(1 - \delta)) + \frac{2(1 - \delta)}{\epsilon^2} \tilde{\lambda}_N^{-2n+1} p \text{ tr } A^{2n}. \quad (6.19)$$

Posons

$$\tilde{\Omega}_N = \left\{ \omega : \frac{2(1 - \delta)}{\epsilon^2} \tilde{\lambda}_N^{-2n+1} p \text{ tr } A(\underline{\xi}(\omega))^{2n} \leq \frac{1}{N} \right\},$$

et

$$\tilde{\Omega} = \bigcup_{N_0 \geq 1} \bigcap_{N \geq N_0} \tilde{\Omega}_N \equiv \liminf_N \tilde{\Omega}_N.$$

Par le lemme de Borel-Cantelli, $\mathbf{P}[\tilde{\Omega}] = 1$, car

$$\mathbf{P}\left[(\tilde{\Omega}_N)^C\right] \leq V_N \stackrel{\text{déf}}{=} \frac{2(1-\delta)}{\epsilon^2} \tilde{\lambda}_N pN(p+4n^2) \exp\left\{-2\sqrt{n}+1\right\},$$

et (V_N) est sommable. Des relations (6.10),(6.11),(6.19), on déduit que pour $\omega \in \tilde{\Omega}$,

$$\liminf_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow \alpha}} f_{N,p,\beta}(\underline{\xi}(\omega)) \geq \Psi^1(\beta, \alpha, \delta).$$

5. Démonstration des résultats intermédiaires :

Démonstration du lemme 6.3.3 :

$$(i) \text{ Pour } \epsilon \in]0, 1[, S_N(\underline{\xi}, \underline{m}) = \frac{1-\epsilon}{2} \frac{\|\underline{m}\|^2}{2} + \frac{\epsilon}{2} \frac{\|\underline{m}\|^2}{2} - \frac{1}{N} \sum_{i=1}^N \log \cosh\left(\sqrt{\beta}(\underline{\xi}, \underline{m})\right),$$

$$\begin{aligned} \log \cosh\left(\sqrt{\beta}(\underline{\xi}, \underline{m})\right) &\leq \sqrt{\beta} |(\underline{\xi}, \underline{m})|, \text{ car } \log \cosh x \leq |x|, \\ &\leq \sqrt{\beta p} \|\underline{m}\|, \end{aligned}$$

donc

$$S_N(\underline{\xi}, \underline{m}) \geq \frac{1-\epsilon}{2} \|\underline{m}\|^2, \text{ pour } \underline{m} \text{ tel que } \|\underline{m}\| > R_\epsilon,$$

et

$$\begin{aligned} \int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| > R_\epsilon]} \exp\left\{-NS_N(\underline{\xi}, \underline{m})\right\} &\leq \int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| > R_\epsilon]} \exp\left\{-N \frac{1-\epsilon}{2} \|\underline{m}\|^2\right\} \\ &\leq \exp\left(-\frac{1-\epsilon}{2} R_\epsilon^2\right) \int_{\mathbf{R}^p} d\underline{m} \exp\left\{-\frac{1-\epsilon}{2} \|\underline{m}\|^2\right\} \\ &= \left(\frac{2\pi}{1-\epsilon}\right)^{\frac{p}{2}} \exp\left(-\frac{1-\epsilon}{2} R_\epsilon^2 (N-1)\right). \end{aligned}$$

(ii) Pour tout γ positif, tel que $\gamma < R_\epsilon$,

$$\int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| \leq R_\epsilon]} \exp\left\{-NS_N(\underline{\xi}, \underline{m})\right\} \geq \int_{\mathbf{R}^p} d\underline{m} \chi_{[\|\underline{m}\| \leq \gamma]} \exp\left\{-N \max_{\|\underline{m}\| \leq \gamma} S_N(\underline{\xi}, \underline{m})\right\}.$$

Puisque $S_N(\underline{\xi}, \underline{m}) \leq \frac{\|\underline{m}\|^2}{2}$, on a donc

$$\int_{\mathbf{R}^p} d\mathbf{m}\chi_{[\|\mathbf{m}\| \leq R_\epsilon]} \exp\left\{-NS_N(\underline{\xi}, \mathbf{m})\right\} \geq \exp\left(-\frac{N}{2}\gamma^2\right) \int_{\mathbf{R}^p} d\mathbf{m}\chi_{[\|\mathbf{m}\| \leq \gamma]}$$

On sait calculer la dernière intégrale (volume de la boule p -dimensionnelle de rayon γ) : (voir [24], p.61)

$$\int_{\mathbf{R}^p} d\mathbf{m}\chi_{[\|\mathbf{m}\| \leq \gamma]} = \gamma^p \frac{\pi^{\frac{p}{2}}}{\Gamma(1 + \frac{p}{2})} \leq \gamma^p \frac{\pi^{\frac{p}{2}}}{[\frac{p}{2}]!}.$$

En posant $\gamma = \left(\frac{1-\epsilon}{2}\right)^{\frac{1}{2}} R_\epsilon$, on obtient (ii).♣

Démonstration du lemme 6.3.4 : Soient $\lambda_1, \dots, \lambda_p$ les valeurs propres de A , et $\rho(A) = \max_{i=1, \dots, p} |\lambda_i|$ le rayon spectral de A .

A est une matrice symétrique, donc $\|A\| = \rho(A) > \lambda$, par hypothèse.

$$\|A - \tilde{\lambda}I\| \leq \|A\| + \tilde{\lambda} = \tilde{\lambda} \left(\frac{\rho(A)}{\tilde{\lambda}} + 1 \right), \text{ avec } \frac{\rho(A)}{\tilde{\lambda}} > \frac{\rho(A)}{\lambda} > 1,$$

donc $\forall n \geq 1$,

$$\|A - \tilde{\lambda}I\|^n \leq \tilde{\lambda}^n \left(\left(\frac{\rho(A)}{\tilde{\lambda}} \right)^n + 1 \right).$$

$\frac{\rho(A)}{\tilde{\lambda}} > 1$, donc il existe un entier $n_0(\frac{\rho(A)}{\tilde{\lambda}})$, tel que pour $n \geq n_0(\frac{\rho(A)}{\tilde{\lambda}})$, on ait

$$\left(\frac{\rho(A)}{\tilde{\lambda}} \right)^n + 1 \leq \left(\frac{\rho(A)}{\tilde{\lambda}} \right)^{2n} \leq \frac{\text{tr } A^{2n}}{\tilde{\lambda}^{2n}},$$

ce qui donne $\|A - \tilde{\lambda}I\|^n \leq \tilde{\lambda} \frac{\text{tr } A^{2n}}{\tilde{\lambda}^{2n}}$, pour $n \geq n_0(\frac{\rho(A)}{\tilde{\lambda}})$.

Pour $x > 1$, $n_0(x)$ doit vérifier la relation

$$\begin{aligned} n_0(x) &\geq \inf \left\{ n \in \mathbf{N} : x^n + 1 \leq x^{2n} \right\} \\ &= \inf \left\{ n \in \mathbf{N} : n \geq \log \left(\frac{1+\sqrt{5}}{2} \right) \frac{1}{\log x} \right\} \end{aligned}$$

Par hypothèse, $\log \frac{\rho(A)}{\tilde{\lambda}} > \log \frac{\lambda}{\tilde{\lambda}} = \frac{1}{K}$, donc $n_0(\frac{\rho(A)}{\tilde{\lambda}}) = K$ convient. ♣

Démonstration de la proposition 6.3.5 : par définition,

$$\operatorname{tr} A^n = \frac{1}{N^n} \sum_{\substack{\mu_1, \dots, \mu_n=1 \\ \mu_1 \neq \mu_2, \mu_2 \neq \mu_3, \dots, \mu_n \neq \mu_1}}^p \sum_{i_1, \dots, i_n=1}^N \xi_{i_1}^{\mu_1} \xi_{i_1}^{\mu_2} \xi_{i_2}^{\mu_2} \cdots \xi_{i_{n-1}}^{\mu_n} \xi_{i_n}^{\mu_n} \xi_{i_n}^{\mu_1},$$

donc $N^n \mathbf{E} \operatorname{tr} A^n$ est égal au nombre de produits ordonnés

$$\Pi_0 \stackrel{\text{déf}}{=} \xi_{i_1}^{\mu_1} \xi_{i_1}^{\mu_2} \xi_{i_2}^{\mu_2} \cdots \xi_{i_{n-1}}^{\mu_n} \xi_{i_n}^{\mu_n} \xi_{i_n}^{\mu_1}, \quad (6.20)$$

où pour $k = 1, \dots, n$, $\mu_k \in \{1, \dots, p\}$, $i_k \in \{1, \dots, N\}$,
 $\mu_k \neq \mu_{k+1}$ (avec la convention $\mu_{n+1} \equiv \mu_1$),

et où chaque variable aléatoire ξ_i^μ apparaît un nombre pair de fois. De tels produits sont appelés *contribuants*, et vérifient évidemment $\mathbf{E} \Pi_0 = \Pi_0 = 1$, puisque $\xi_i^\mu = \pm 1$. Les autres produits sont d'espérance nulle, et ne contribuent pas à $\mathbf{E} \operatorname{tr} A^n$. Si n est impair, $\mathbf{E} \operatorname{tr} A^n$ est nulle. Nous supposons désormais n pair.

La démonstration de l'inégalité (6.16) se fait en trois étapes.

1) Pour un produit Π_0 donné, posons

$$\begin{aligned} u(\Pi_0) &= \operatorname{card} \left\{ \mu \in \{1, \dots, p\}, \exists k \in \{1, \dots, n\}, \mu_k = \mu \right\} \\ v(\Pi_0) &= \operatorname{card} \left\{ i \in \{1, \dots, N\}, \exists k \in \{1, \dots, n\}, i_k = i \right\} \end{aligned}$$

Nous allons montrer la propriété suivante :

$$\Pi_0 \text{ contribuant} \Rightarrow \begin{cases} u(\Pi_0) + v(\Pi_0) \leq n + 1 \\ 2v(\Pi_0) \leq n \end{cases} \quad (6.21)$$

a. $2v(\Pi_0) \leq n$:

Nous dirons que deux variables ξ_i^μ et ξ_j^ν sont *appariées* si $i = j$ et $\mu = \nu$.

Du fait que $\mu_l \neq \mu_{l+1}$, deux termes consécutifs ayant même indice i_k ne peuvent être appariés, donc pour chaque $k \in \{1, \dots, n\}$, il existe un $j \in \{1, \dots, n\}$, $j \neq k$, tel que $i_j = i_k$. D'où la majoration $v(\Pi_0) \leq \frac{n}{2}$.

b. $u(\Pi_0) + v(\Pi_0) \leq n + 1$:

Soit Π_0 un produit contribuant tel que $v(\Pi_0) = k \in \{1, \dots, \frac{n}{2}\}$. Alors nous devons montrer que $u(\Pi_0) \leq n - k + 1$.

Si $v(\Pi_0) = k$, il existe k indices inférieurs distincts i_1, \dots, i_k , donc, quite à renuméroter les indices supérieurs, Π_0 peut s'écrire sous la forme :

$$\Pi_0 = \dots \xi_{i_1}^{\mu_1} \xi_{i_2}^{\mu_1} \dots \xi_{i_2}^{\mu_2} \xi_{i_3}^{\mu_2} \dots \xi_{i_{k-1}}^{\mu_{k-1}} \xi_{i_k}^{\mu_{k-1}} \dots,$$

i.e. il existe au moins $(k - 1)$ couples de variables adjacentes non appariées.

Puisque toutes les variables ξ_i^μ sont appariées (Π_0 est contribuant),

- $\xi_{i_2}^{\mu_1}$ est appariée, donc il existe un indice $\mu^{(1)}$ tel que $\mu^{(1)} \equiv \mu_1$.

- $\xi_{i_3}^{\mu_2}$ est appariée, donc l'indice μ_2 est identifié avec un autre indice $\mu^{(2)}$. En effet, même si $\mu_1 = \mu_2$ (par la première identification), puisque $i_2 \neq i_3$, $\xi_{i_2}^{\mu_1}$ et $\xi_{i_3}^{\mu_2}$ ne sont pas appariés.

- ...

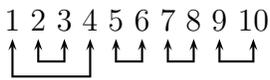
- $\xi_{i_k}^{\mu_{k-1}}$ est appariée, donc l'indice μ_{k-1} est identifié avec un autre indice $\mu^{(k-1)}$. (Même si μ_{k-1} est un des $\mu^{(j)}$, $j \leq k - 2$, $\xi_{i_k}^{\mu_{k-1}}$ ne peut être apparié avec $\xi_{i_{j+1}}^{\mu^{(j)}}$, car les indices inférieurs sont distincts).

On a donc prouvé que au moins $(k - 1)$ identifications doivent être effectuées sur les indices supérieurs, pour avoir un produit contribuant, i.e. $u(\Pi_0) \leq n - (k - 1)$.

2) Cas des produits simples :

Un produit contribuant sera dit *simple* si ses $2n$ facteurs peuvent être groupés en n paires de variables aléatoires identiques, de telle manière que si chaque paire est délimité par des crochets (par exemple $\xi_k^\mu \left\{ \xi_i^\mu \xi_i^\nu \xi_j^\nu \xi_j^\mu \xi_i^\mu \right\} \xi_i^\lambda$), alors les n paires de crochets ont une structure emboîtée, et/ou disjointe. On peut remarquer que la donnée des n crochets gauches suffit à reconstruire les appariements, dans le cas d'un produit simple.

Prenons un exemple : pour $n = 5$, si on sait qu'il y a un crochet gauche aux positions 1, 2, 5, 7, et 9, on peut compléter et obtenir tous les appariements :



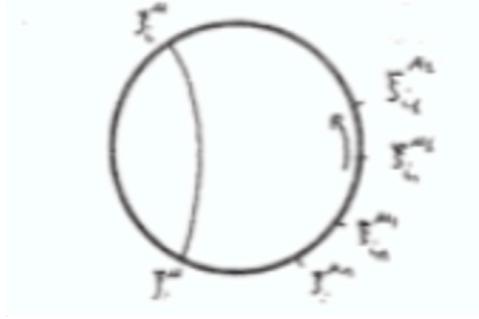
Ainsi, le nombre de produits simples est majoré par :

$$M_n(p) = C_{2n}^n \max_{\substack{u+v \leq n+1 \\ v \leq \frac{n}{2}}} p^u N^v \quad (6.22)$$

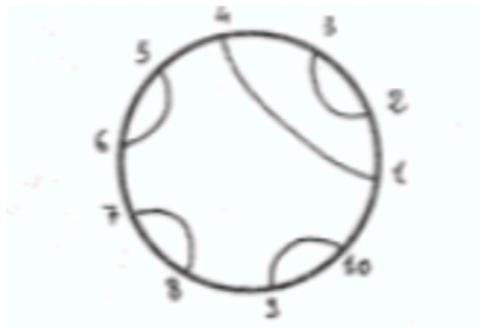
3) Cas des produits non simples :

Pour dénombrer les produits non simples, nous construisons un algorithme permettant de transformer tout graphe non simple, avec $1 \leq i_k \leq N$, $1 \leq \mu_k \leq p$, $\forall k = 1, \dots, n$, en un graphe simple, où $1 \leq i_k \leq N$, $1 \leq \mu_k \leq \tilde{p}$, $\forall k = 1, \dots, n$, $\tilde{p} > p$, de manière bijective. On pourra alors utiliser le **2)** pour majorer le nombre de produits non simples par $M_n(\tilde{p})$.

Nous allons définir de manière plus précise la notion de graphe simple, et de graphe non simple : un produit $\xi_{i_1}^{\mu_1} \xi_{i_1}^{\mu_2} \xi_{i_2}^{\mu_2} \dots \xi_{i_{n-1}}^{\mu_{n-1}} \xi_{i_n}^{\mu_n} \xi_{i_n}^{\mu_1}$ peut être représenté par un cercle orienté, indexé par les $2n$ points ξ_i^μ , la position de la première variable étant fixée.



Si on relie les variables appariées, il est facile de voir que pour un produit simple, il est possible d'obtenir un appariement complet, tel que les liens d'appariement ne se croisent pas : on dit que le graphe est simple. Pour l'exemple précédent, on obtient :



Définition : Dans un produit $\Pi_0 = \xi_{i_1}^{\mu_1} \xi_{i_1}^{\mu_2} \xi_{i_2}^{\mu_2} \dots \xi_{i_{n-1}}^{\mu_{n-1}} \xi_{i_n}^{\mu_n} \xi_{i_n}^{\mu_1}$, deux variables $\xi_{i_{k-1}}^{\mu_k}, \xi_{i_k}^{\mu_k}$ ($k = 2, \dots, n$), ainsi que les deux variables $\xi_{i_1}^{\mu_1}, \xi_{i_n}^{\mu_1}$ sont dites *adjacentes*.

Dans un produit non simple, il existe nécessairement au moins deux couples de variables adjacentes non appariées. Nous allons donner un algorithme tel que chaque étape crée au moins un nouveau couple de variables adjacentes appariées. En au plus n étapes, on obtient donc un produit simple.

Soit Π_{m-1} produit non simple obtenu après $(m - 1)$ étapes. Il existe au moins deux couples de variables adjacentes non appariées. Soit ξ_i^μ la variable en deuxième position dans le premier de ces couples. Alors ξ_i^μ est appariée avec une autre variable, non adjacente :

$$\xi_j^\mu \xi_i^\mu \dots \xi_{i'}^{\mu'}, \text{ où } \mu = \mu', i = i'.$$

deux cas se présentent alors :

a. $\xi_{i'}^{\mu'}$ est en deuxième position de son couple :

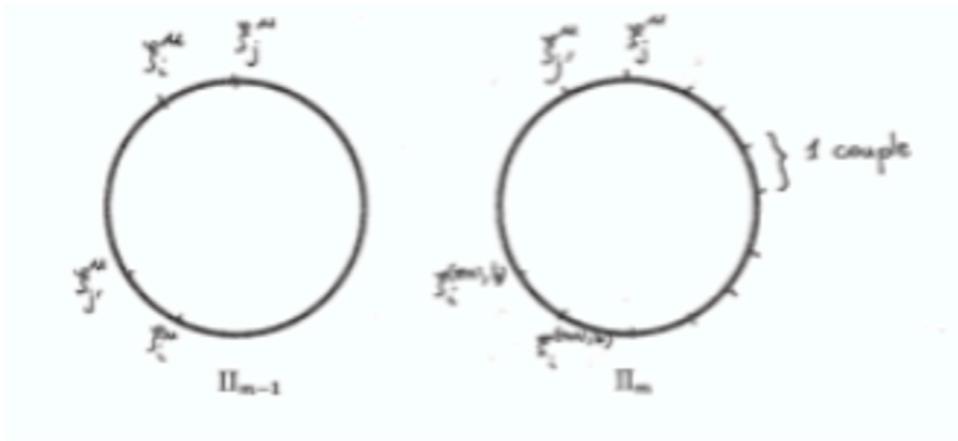
$$\exists j, \text{ tel que } \Pi_{m-1} = \dots \xi_j^\mu \xi_i^\mu \xi_{i'}^{\mu'} \dots \xi_{j'}^{\mu'} \xi_{j'}^\mu \xi_i^\mu \dots$$

Dans ce cas, on effectue la transformation suivante : on inverse la séquence $\xi_i^\mu \dots \xi_{j'}^\mu$, ce qui donne : $\dots \xi_j^\mu \xi_{j'}^\mu \xi_{j'}^{\nu'} \dots \xi_j^\nu \xi_i^\mu \xi_i^\mu \dots$

Cette transformation crée une nouvelle paire adjacente appariée. Pour pouvoir retrouver la configuration initiale, on remplace l'indice μ par un double indice (a, b) , $a \in \{1, \dots, n\}$, $b \in \{0, \dots, n-1\}$,

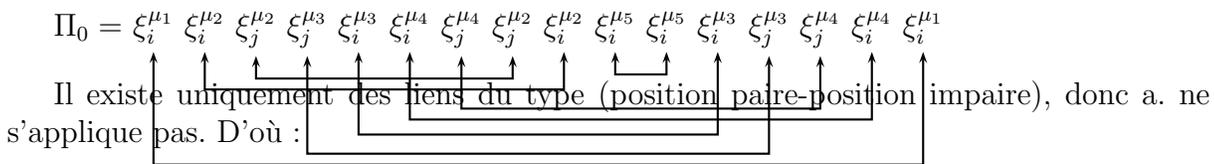
$$\text{où } \begin{cases} a = m, & \text{s'il s'agit de la } m\text{ième transformation,} \\ b & \text{est la "distance" du couple } \xi_i^{(a,b)} \xi_i^{(a,b)} \\ & \text{au couple } \xi_j^\mu \xi_{j'}^\mu, \text{ sur le cercle orienté.} \end{cases}$$

La transformation $\Pi_{m-1} \rightarrow \Pi_m$ peut être schématisé de la manière suivante :



b est le nombre de couples séparant les deux couples considérés. (Sur l'exemple, $b = 4$).

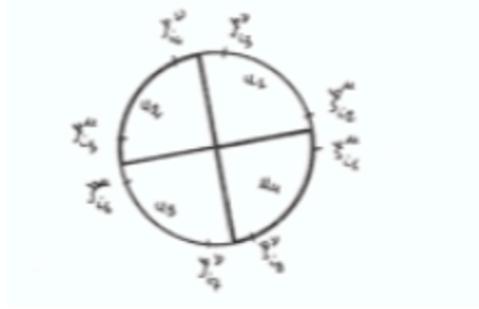
Contrairement à ce qui est affirmé dans la démonstration de Koch, il peut se produire un autre cas : si toute variable en deuxième position de son couple, non appariée avec sa variable adjacente, est appariée avec une variable en première position de son couple, on ne peut pas appliquer ce qui précède. Voici un exemple, avec $n = 8$, illustrant cette possibilité :



b. Passage de Π_{m-1} à Π_m , quand Π_m est un graphe non simple, tel que tous les liens sont du type (position paire-position impaire), i.e. :

$$\xi_i^\mu \xi_i^\mu, \text{ ou bien } \xi_i^\mu \xi_j^\mu \dots \xi_{j'}^\mu \xi_i^\mu, \text{ ou bien } \xi_j^\mu \xi_i^\mu \dots \xi_i^\mu \xi_{j'}^\mu.$$

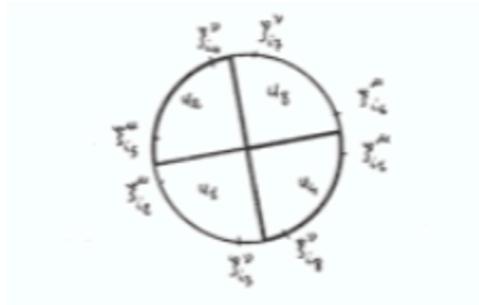
Tout graphe non simple contient des liens d'appariement entremêlés, par définition, et peut donc être représenté de la manière suivante :



$$\text{avec : } \begin{cases} i_1 \neq i_2, i_3 \neq i_4 \\ i_5 \neq i_6, i_7 \neq i_8 \end{cases} \quad \text{et} \quad \begin{cases} i_1 = i_6, \text{ et/ou } i_2 = i_5 \\ i_3 = i_8, \text{ et/ou } i_4 = i_7 \end{cases}$$

$$\text{Soient } \begin{cases} u_1 \text{ le produit } \xi_{i_2}^\mu \dots \xi_{i_3}^\nu \\ u_2 \text{ le produit } \xi_{i_4}^\nu \dots \xi_{i_5}^\mu \\ u_3 \text{ le produit } \xi_{i_6}^\mu \dots \xi_{i_7}^\nu \\ u_4 \text{ le produit } \xi_{i_8}^\nu \dots \xi_{i_1}^\mu \end{cases}$$

La transformation consiste à inverser l'ordre des produits, en respectant l'ordre des termes de chaque produit : $(u_1.u_2.u_3.u_4) \mapsto (u_4.u_3.u_2.u_1)$. On suppose aussi que $\xi_{i_1}^\mu$ (et donc tous les termes de u_4) restent à la même position sur le cercle.



Puisque $i_5 = i_2$ et/ou $i_1 = i_6$, il se crée un ou deux nouveaux couples adjacents appariés ayant comme indice supérieur μ . Pour un de ces couples, noté $\xi_i^\mu \xi_i^\mu$, on remplace l'indice μ par le couple (a, b) , où $a \in \{1, \dots, n\}$, $b \in \{0, \dots, n-1, \dots\}$, et

$$\begin{cases} a = m, & \text{s'il s'agit de la } m^{\text{ème}} \text{ transformation,} \\ b & \text{est le nombre de couples entre } \xi_i^\mu \xi_i^\mu \text{ et l'autre couple} \\ & \text{ayant comme indice supérieur } \mu. \end{cases}$$

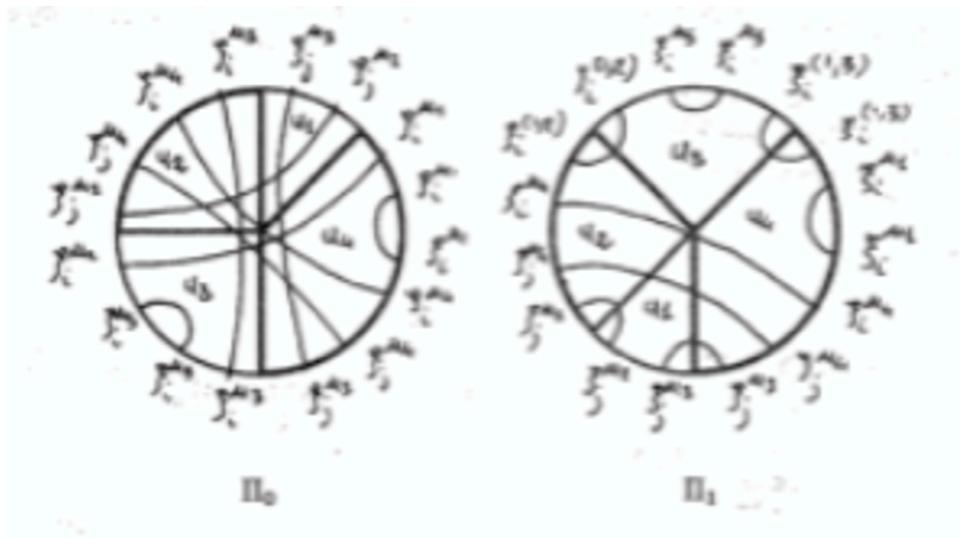
$$\text{De même, } \xi_j^\nu \xi_j^\nu \mapsto \xi_j^{(m,b')} \xi_j^{(m,b')}.$$

La donnée de b et b' permet d'effectuer la transformation inverse, car on connaît alors les produits u_4, u_3, u_2 , et u_1 . Il suffit d'inverser l'ordre des suites, et de remplacer les indices supérieurs initiaux.

Remarque : quand on remplace l'indice μ (respectivement ν) par (m, b) (respectivement (m, b')), on ne perd pas la valeur de cet indice, car b donne la position du couple ayant

même indice supérieur. En quelque sorte, on remplace l'information redondante par une information utile.

Exemple : appliquons cet algorithme à l'exemple précédant b. Π_0 peut être représenté par :



La donnée de Π_1 (qui est un graphe simple) équivaut à celle de Π_0 .

c. Finalement, à partir d'un produit simple Π_m obtenu après m transformations ($m \leq n$), on peut retrouver la configuration Π_0 , en appliquant les transformations inverses, étape par étape. On commence par considérer le (ou les) couple(s) dont l'indice supérieur est du type (m, \cdot) .

- si un seul couple a pour indice supérieur (m, \cdot) , appliquer a.
- si deux couples ont pour indices supérieurs (m, \cdot) , appliquer b.

$\text{card}\{(a, b) \in \{1, \dots, n\} \times \{0, \dots, n-1\}\} = n^2$, donc le nombre de produits contribuant est majoré par le nombre de produits simples, mais avec la condition suivante : les indices supérieurs peuvent prendre $(p + n^2)$ valeurs distinctes. Par **2**), on a alors

$$\mathbf{E} \text{tr} A^n \leq N^{-n} M_n(p + n^2) = N^{-n} C_{2n}^n \max_{\substack{u+v \leq n+1 \\ v \leq \frac{n}{2}}} (p + n^2)^u N^v$$

$$\mathbf{E} \text{tr} A^n \leq 4^n (p + n^2) \left(\frac{p + n^2}{N}\right)^{\frac{n}{2}}, \text{ si } p + n^2 \leq N.$$

Cela achève la démonstration de la proposition 6.3.5.♣

• Nous allons maintenant prouver les **majorations** des encadrements (6.3) et (6.4). Pour cela, nous montrons d'abord que l'énergie libre du modèle de Curie-Weiss (qui n'est pas

une v.a.) est, à une constante près, une borne supérieure triviale de l'énergie libre du modèle de Hopfield. Nous calculons ensuite, par la méthode de Laplace, la limite thermodynamique de l'énergie libre du modèle de Curie-Weiss.

Nous avons :

$$\begin{aligned} Z_{N,p,\beta}(\underline{\xi}) &= \sum_{\underline{x} \in X_N} \exp \left\{ \frac{\beta}{2N} \sum_{\mu=1}^p \left(\underline{\xi}^\mu, \underline{x} \right)_N^2 - \beta \frac{p}{2} \right\} \\ &= \sum_{\underline{x} \in X_N} \exp \left\{ \frac{\beta}{2N} \left(\underline{\xi}^1, \underline{x} \right)_N^2 - \frac{\beta}{2} \right\} \exp \left\{ \frac{\beta}{2N} \sum_{\mu=2}^p \left(\underline{\xi}^\mu, \underline{x} \right)_N^2 - \beta \frac{p-1}{2} \right\} \\ &\geq e^{-\beta \frac{p-1}{2}} Z_{N,1,\beta}(\underline{\xi}) = e^{-\beta \frac{p-1}{2}} Z_{N,\beta}^{CW}, \end{aligned}$$

où $Z_{N,\beta}^{CW} = \sum_{\underline{x} \in X_N} \exp \left\{ \frac{\beta}{2N} \sum_{\substack{i,j=1 \\ i \neq j}}^N x_i x_j \right\}$ est la fonction de partition du modèle de Curie-Weiss. Nous avons donc, pour tout $\omega \in \Omega$,

$$f_{N,p,\beta}(\underline{\xi}(\omega)) \leq f_{N,\beta}^{CW} + \frac{p-1}{2N} \quad (6.23)$$

En reprenant la méthode utilisée au paragraphe 2, il est facile d'obtenir :

$$\begin{aligned} \frac{1}{N} \log Z_{N,\beta}^{CW} &= -\frac{\beta}{2N} + \log 2 + \frac{1}{2N} \log \frac{N}{2\pi} \\ &\quad + \frac{1}{N} \log \int_{\mathbf{R}} dm \exp \left\{ -N \left(\frac{m^2}{2} - \log \cosh \sqrt{\beta} m \right) \right\}, \end{aligned}$$

et la méthode de Laplace permet d'estimer le dernier terme, ce qui donne :

$$f_{N,\beta}^{CW} = -\frac{1}{\beta} \log 2 + \min_{m \in \mathbf{R}} \left\{ \frac{m^2}{2} - \frac{1}{\beta} \log \cosh \beta m \right\} + R_N, \text{ où } R_N \xrightarrow{N \rightarrow \infty} 0 \quad (6.24)$$

De (6.23) et (6.24), on déduit que pour tout $\omega \in \Omega$,

$$\limsup_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow \alpha \geq 0}} f_{N,p,\beta}(\underline{\xi}(\omega)) \leq -\frac{1}{\beta} \log 2 + \min_{h \in \mathbf{R}} \Phi_0(\beta, h) + \frac{\alpha}{2}.$$

Ceci achève la démonstration du théorème 6.3.1. ♣

6.4 Propriété de self-averaging de l'énergie libre pour toute température et $\frac{p}{N} \rightarrow \alpha \geq 0$

Dans la pratique, les réseaux neuronaux ne sont intéressants que s'ils peuvent mémoriser un nombre d'images proportionnel à la taille du réseau, i.e $p = \alpha N$, où α est une constante positive. Et il s'avère que des simulations, effectuées pour des réseaux formés de quelques centaines de neurones, confirment cette hypothèse (voir par exemple [4]). Or les résultats de convergence obtenues au paragraphe précédent ne sont valables que sous la contrainte $\frac{p}{N} \rightarrow 0$, qu'il semble d'ailleurs difficile d'affaiblir. Il est donc judicieux de chercher quelles propriétés vérifie l'énergie libre dans le cas où p peut croître proportionnellement à N . Dans cette optique, nous allons montrer que l'énergie libre vérifie la propriété de *self-averaging* (*auto-moyennation*) pour toute température.

Définition 6.4.1 Soit $(Y_n)_{n \geq 1}$ une suite de variables aléatoires sur l'espace $(\Omega, \mathcal{F}, \mathbf{P})$. (Y_n) vérifie la propriété de self-averaging (s.a.) si

$$\lim_{n \rightarrow \infty} \mathbf{E} \left(Y_n - \mathbf{E} Y_n \right)^2 = 0,$$

autrement dit la variance de Y_n est asymptotiquement nulle.

Théorème 6.4.2 Soient $\beta > 0$, $\alpha \in [0, 1]$, et $p \leq \alpha N$. Alors :

$$\mathbf{E} \left(f_{N,p,\beta}(\underline{\xi}) - \mathbf{E} f_{N,p,\beta}(\underline{\xi}) \right)^2 \leq \frac{1}{N} (20 + \alpha + \mathcal{O}(N^{-\frac{1}{8}})) \quad (6.25)$$

En particulier,

$$\lim_{\substack{N \rightarrow \infty \\ \frac{p}{N} \rightarrow \alpha}} \mathbf{E} \left(f_{N,p,\beta}(\underline{\xi}) - \mathbf{E} f_{N,p,\beta}(\underline{\xi}) \right)^2 = 0.$$

La méthode utilisée pour montrer ce théorème, originalement due à V. Girko [22], a été reprise par L.Pastur et M.Shcherbina [39] pour montrer la propriété de s.a. de l'énergie libre pour le modèle de Sherrington-Kirkpatrick. Il est possible d'adapter cette technique de calcul pour le modèle de Hopfield, et des majorations grossières impliquent la propriété s.a. pour $p \leq \gamma(N) \sqrt{N}$, où $\gamma(N) \rightarrow_{N \rightarrow \infty} 0$ [47]. Récemment, M.Shcherbina et B.Tirozzi [45] ont étendu ce résultat au cas $\frac{p}{N} \rightarrow \alpha \geq 0$: une estimation de $\mathbf{E} \operatorname{tr} A^n$ (voir proposition 6.3.5), obtenue de manière différente de celle présentée ici, leur permet d'améliorer les majorations citées précédemment. Nous allons donc suivre les articles [47]

et [45]. Toutefois, nous utilisons l'estimation (6.16), un peu différente de celle proposée dans [45].

Notation : $(\mathcal{F}_k)_{k=0,\dots,N}$ est la famille croissante de tribus telle que $\mathcal{F}_0 = \{\emptyset, \Omega\}$, et pour $k = 1, \dots, N$, \mathcal{F}_k est la tribu engendrée par les v.a. $\underline{\xi}_1, \dots, \underline{\xi}_k$.

Le théorème 6.4.2 repose sur les deux lemmes suivants :

Lemme 6.4.3 Soit G_N une fonction dépendant des N v.a. $\underline{\xi}_1, \dots, \underline{\xi}_N$, et

$$\Psi_k = \mathbf{E}\left[G_N \mid \mathcal{F}_k\right] - \mathbf{E}\left[G_N \mid \mathcal{F}_{k-1}\right], \quad k \in \{1, \dots, N\}.$$

Si $\mathbf{E}\left[(\Psi_k)^2\right]$ est borné uniformément en k par une fonction C , dépendant de N , alors on a la majoration suivante :

$$\mathbf{E}\left[\left(\frac{G_N}{N} - \mathbf{E}\frac{G_N}{N}\right)^2\right] \leq \frac{C(N)}{N}.$$

Démonstration : ce lemme se montre facilement en remarquant que

$$G_N - \mathbf{E}G_N = \sum_{k=1}^N \Psi_k, \quad \text{et } \mathbf{E}\Psi_i\Psi_j = 0, \quad \text{pour } i \neq j,$$

d'où

$$\mathbf{E}\left[\left(\frac{G_N}{N} - \mathbf{E}\frac{G_N}{N}\right)^2\right] = \frac{1}{N^2} \sum_{k=1}^N \mathbf{E}\Psi_k^2 \leq \frac{C(N)}{N},$$

par hypothèse.♣

L'étape suivante consiste à appliquer le lemme 6.4.3 dans le cas qui nous intéresse. Posons

$$\Psi_k = \mathbf{E}\left[F_{N,p,\beta}(\underline{\xi}) \mid \mathcal{F}_k\right] - \mathbf{E}\left[F_{N,p,\beta}(\underline{\xi}) \mid \mathcal{F}_{k-1}\right].$$

(Rappelons que $F_{N,p,\beta} = Nf_{N,p,\beta} = -\frac{1}{\beta} \log Z_{N,p,\beta}$).

Lemme 6.4.4 Soit $\alpha \in [0, 1]$, $p \leq \alpha N$. Alors $\mathbf{E}\left[(\Psi_k)^2\right]$ est borné uniformément en $k \in \{1, \dots, N\}$, et

$$\mathbf{E}\left[(\Psi_k)^2\right] \leq 20 + \alpha + \mathcal{O}(N^{-\frac{1}{8}}) \quad (6.26)$$

Démonstration : Soit

$$H_k = -\frac{1}{2N} \sum_{\mu=1}^p \sum_{\substack{i,j=1,\dots,N \\ i \neq j; i,j \neq k}} \xi_i^\mu \xi_j^\mu x_i x_j,$$

$$R_k = -\frac{1}{N} \sum_{\mu=1}^p \sum_{\substack{l=1,\dots,N \\ l \neq k}} \xi_k^\mu \xi_l^\mu x_k x_l,$$

et

$$\tilde{H}_k(t) = H_k + tR_k.$$

Remarquons que $\tilde{H}_k(1) = H_{N,p,\beta}(\underline{\xi})$.

Posons $\tilde{f}_k(t) = -\frac{1}{\beta} \left\{ \log \tilde{Z}_k(t) - \log \tilde{Z}_k(0) \right\}$,

où $\tilde{Z}_k(t)$ est la fonction de partition associée au hamiltonien $\tilde{H}_k(t)$:

$$\tilde{Z}_k(t) = \sum_{\underline{x} \in X_N} e^{-\beta \tilde{H}_k(t)}.$$

Nous avons :

$$\Psi_k = \mathbf{E} \left[\tilde{f}_k(1) \mid \mathcal{F}_k \right] - \mathbf{E} \left[\tilde{f}_k(1) \mid \mathcal{F}_{k-1} \right], \quad (6.27)$$

car $\mathbf{E} \left[\log \tilde{Z}_k(0) \mid \mathcal{F}_k \right] = \mathbf{E} \left[\log \tilde{Z}_k(0) \mid \mathcal{F}_{k-1} \right]$.

D'autre part, toute fonction $g \in C^1([0, 1])$ vérifie :

$$g(1) = g(0) + \int_0^1 \frac{dg}{dt}(t) dt.$$

Cette relation, appliquée à $\tilde{f}_k(t)$, implique :

$$\Psi_k = \frac{1}{N} \int_0^1 dt \left\{ \sum_{\mu=1}^p \sum_{\substack{i=1 \\ i \neq k}}^N \mathbf{E} \left[\xi_i^\mu \xi_k^\mu \langle x_i x_k \rangle_{t,k} \mid \mathcal{F}_{k-1} \right] - \xi_k^\mu \mathbf{E} \left[\xi_i^\mu \langle x_i x_k \rangle_{t,k} \mid \mathcal{F}_k \right] \right\},$$

où $\langle \cdot \rangle_{t,k}$ désigne l'intégration par rapport à la mesure de Gibbs associée au hamiltonien $\tilde{H}_k(t)$:

$$\langle g(\cdot) \rangle_{t,k} = \frac{1}{\tilde{Z}_k(t)} \sum_{\underline{x} \in X_N} g(\underline{x}) e^{-\beta \tilde{H}_k(t, \underline{x})}.$$

Dans [47], nous concluons en remarquant simplement que $|\langle x_i x_k \rangle_{t,k}| \leq 1$, d'où

$$\mathbf{E}(\Psi_k)^2 \leq 4p^2 \quad (6.28)$$

Cela implique la propriété de s.a. uniquement pour $p \leq \gamma(N)\sqrt{N}$, où γ est une fonction convergeant arbitrairement lentement vers 0 en $+\infty$.

Nous allons suivre la méthode proposée par M.Shcherbina et B.Tirozzi pour améliorer la majoration (6.28). Reprenant la relation (6.27), nous écrivons :

$$\begin{aligned} (\Psi_k)^2 &= \left(\mathbf{E}[\tilde{f}_k(1) | \mathcal{F}_k] \right)^2 - 2\mathbf{E}[\tilde{f}_k(1) | \mathcal{F}_{k-1}] \mathbf{E}[\tilde{f}_k(1) | \mathcal{F}_k] \\ &\quad + \left(\mathbf{E}[\tilde{f}_k(1) | \mathcal{F}_{k-1}] \right)^2 \\ &= \mathbf{E}[\tilde{f}_k(1) \mathbf{E}[\tilde{f}_k(1) | \mathcal{F}_k] | \mathcal{F}_k] - 2\mathbf{E}[\tilde{f}_k(1) \mathbf{E}[\tilde{f}_k(1) | \mathcal{F}_{k-1}] | \mathcal{F}_k] \\ &\quad + \mathbf{E}[\tilde{f}_k(1) \mathbf{E}[\tilde{f}_k(1) | \mathcal{F}_{k-1}] | \mathcal{F}_{k-1}] \end{aligned}$$

d'où

$$\mathbf{E}[(\Psi_k)^2] = \mathbf{E}[\tilde{f}_k(1) \Psi_k] \leq \sqrt{\mathbf{E}[\tilde{f}_k(1)^2]} \sqrt{\mathbf{E}[(\Psi_k)^2]}$$

et

$$\mathbf{E}[(\Psi_k)^2] \leq \mathbf{E}[\tilde{f}_k(1)^2] \quad (6.29)$$

D'autre part,

$$\frac{d}{dt} \tilde{f}_k(t) = -\frac{1}{\beta} \frac{1}{\tilde{Z}_k(t)} \frac{d}{dt} (\tilde{Z}_k(t)) = \langle R_k \rangle_{k,t}$$

$$\frac{d^2}{dt^2} \tilde{f}_k(t) = \beta \left(\langle R_k \rangle_{k,t}^2 - \langle R_k^2 \rangle_{k,t} \right) \leq 0,$$

donc $\tilde{f}'_k(t) \leq \tilde{f}'_k(0)$, pour $t \in [0, 1]$. (On note $f'(\cdot) \equiv \frac{d}{dt}f(\cdot)$).

Puisque $\tilde{f}_k(1) = \tilde{f}_k(0) + \int_0^1 \tilde{f}'_k(t)dt$, et $\tilde{f}_k(0) = 0$, il vient : $\tilde{f}_k(1) \leq \tilde{f}'_k(0)$.

De même, $\tilde{f}_k(1) \geq \tilde{f}'_k(1)$. On en déduit

$$\tilde{f}'_k(1) \leq \tilde{f}_k(1) \leq \tilde{f}'_k(0),$$

et

$$\mathbf{E} \left[(\tilde{f}_k(1))^2 \right] \leq \mathbf{E} \left[(\tilde{f}'_k(1))^2 \right] + \mathbf{E} \left[(\tilde{f}'_k(0))^2 \right] \quad (6.30)$$

• Estimation du premier terme du second membre de (6.30) : soit J la matrice dont les coefficients sont définis par

$$\begin{cases} J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu, & \text{si } i \neq j, \\ J_{ii} = 0. \end{cases}$$

(Dans les chapitres précédents, J_{ij} était défini sans facteur $\frac{1}{N}$). Notons $\langle \cdot \rangle_{H_{N,p,\beta}}$ l'intégration par rapport à la mesure de Gibbs associée au hamiltonien $H_{N,p,\beta}$.

$$\begin{aligned} \mathbf{E} \left[(\tilde{f}'_k(1))^2 \right] &= \mathbf{E} \left\langle R_k \right\rangle_{k,t=1}^2 \equiv \mathbf{E} \left\langle R_k \right\rangle_{H_{N,p,\beta}}^2 \\ &\leq \mathbf{E} \left\langle \sum_{l_1, l_2=1}^N J_{kl_1} J_{kl_2} x_{l_1} x_{l_2} \right\rangle_{H_{N,p,\beta}} \\ &= \frac{1}{N} \mathbf{E} \left\langle \sum_{k=1}^N \sum_{l_1, l_2=1}^N J_{kl_1} J_{kl_2} x_{l_1} x_{l_2} \right\rangle_{H_{N,p,\beta}}, \text{ par symétrie de } H_{N,p,\beta} \text{ en ses indices,} \\ &= \mathbf{E} \left\langle \|J\vec{x}\|^2 \right\rangle_{H_{N,p,\beta}}, \text{ avec } \vec{x} = \left(\frac{x_1}{\sqrt{N}}, \dots, \frac{x_N}{\sqrt{N}} \right). \end{aligned}$$

Le vecteur \vec{x} est de norme 1, et par définition, $\|J\|^2 = \sup_{\vec{s} \in \mathbf{R}^N: \|\vec{s}\|=1} \|J\vec{s}\|^2$, donc

$$\mathbf{E} \left[(\tilde{f}'_k(1))^2 \right] \leq \mathbf{E} \left[\|J\|^2 \right] \quad (6.31)$$

• Estimation du deuxième terme :

$$\mathbf{E} \left[(\tilde{f}'_k(0))^2 \right] = \frac{1}{N^2} \mathbf{E} \left[\sum_{l_1, l_2 \neq k} \sum_{\mu_1, \mu_2=1}^p \xi_k^{\mu_1} \xi_{l_1}^{\mu_1} \xi_k^{\mu_2} \xi_{l_2}^{\mu_2} \left\langle x_{l_1} x_{l_2} \right\rangle_{k,t=0} \right].$$

Le terme $\langle x_{l_1} x_{l_2} \rangle_{k,t=0}$ est indépendant de ξ_k , donc

$$\begin{aligned}
\mathbf{E} \left[(f'_k(0))^2 \right] &= \frac{1}{N^2} \mathbf{E} \left[\sum_{l_1, l_2 \neq k} \sum_{\mu=1}^p \xi_{l_1}^\mu \xi_{l_2}^\mu \langle x_{l_1} x_{l_2} \rangle_{k,t=0} \right] \\
&= \frac{N-1}{N^2} p + \frac{1}{N^2} \mathbf{E} \left[\sum_{\substack{l_1, l_2=1 \\ l_1 \neq l_2}}^N \sum_{\mu=1}^p \xi_{l_1}^\mu \xi_{l_2}^\mu \langle x_{l_1} x_{l_2} \rangle_{k,t=0} \right] \\
&\leq \alpha + \mathbf{E} \left[\left\langle \sum_{l_1, l_2=1}^N J_{l_1 l_2} \frac{x_{l_1}}{\sqrt{N}} \frac{x_{l_2}}{\sqrt{N}} \right\rangle_{k,t=0} \right] \\
&= \alpha + \mathbf{E} \left[\left\langle (J\vec{x}, \vec{x}) \right\rangle_{k,t=0} \right].
\end{aligned}$$

La matrice J est symétrique donc $\|J\| = \sup_{\vec{s} \in \mathbf{R}^N: \|\vec{s}\|=1} (J\vec{s}, \vec{s})$. On peut donc écrire :

$$\mathbf{E} \left[(f'_k(0))^2 \right] \leq \alpha + \mathbf{E} \left[\|J\| \right] \quad (6.32)$$

• Pour majorer $\mathbf{E} \left[\|J\| \right]$ et $\mathbf{E} \left[\|J\|^2 \right]$, nous pouvons utiliser les estimations de la proposition 6.3.5. Posons $\tilde{N} = p, \tilde{p} = N$, et pour $\mu, \nu \in \{1, \dots, \tilde{p}\}$,

$$\tilde{A}_{\mu\nu} = \frac{1}{\tilde{N}} \sum_{i=1}^{\tilde{N}} \xi_i^\mu \xi_i^\nu - \delta_{\mu\nu}$$

Nous avons : $\mathbf{E} \left[\|J\|^k \right] = \alpha^k \mathbf{E} \left[\|\tilde{A}\|^k \right], \forall k \geq 1$.

$$\|\tilde{A}\|^k \leq \max_{\mu=1, \dots, \tilde{p}} \sum_{\nu=1}^{\tilde{p}} |\tilde{A}_{\mu\nu}| \leq \tilde{p}.$$

$$\begin{aligned}
\text{Soit } \lambda > 0, \mathbf{E} \left[\|\tilde{A}\|^k \right] &= \mathbf{E} \left[\|\tilde{A}\|^k \chi_{\left[\|\tilde{A}\| > \lambda \right]} \right] + \mathbf{E} \left[\|\tilde{A}\|^k \chi_{\left[\|\tilde{A}\| \leq \lambda \right]} \right] \\
&\leq \tilde{p}^k \mathbf{P} \left[\|\tilde{A}\| > \lambda \right] + \lambda^k \\
&\leq \tilde{p}^k \lambda^{-n} \mathbf{E} \left[\|\tilde{A}\|^n \right] + \lambda^k, \text{ pour } n \geq 1.
\end{aligned}$$

\tilde{A} est une matrice symétrique, donc $\|\tilde{A}\| = \rho(\tilde{A}) = \max_{i=1, \dots, \tilde{N}} |\lambda_i|$, où $\lambda_1, \dots, \lambda_N$ sont les valeurs propres de \tilde{A} .

Nous supposons que n est pair, alors

$$\|\tilde{A}\|^n \leq \max_{i=1,\dots,\tilde{N}} \lambda_i^n \leq \text{tr } \tilde{A}^n,$$

et

$$\begin{aligned} \mathbf{E} \left[\|\tilde{A}\|^k \right] &\leq \tilde{p}^k \lambda^{-n} \mathbf{E} \text{tr } \tilde{A}^N + \lambda^k, \text{ pour } n \text{ pair.} \\ &\leq \tilde{p}^k 4^n \tilde{N}^{-n} \lambda^{-n} \max_{\substack{u,v \geq 1 \\ u+v \leq n+1; v \leq \frac{n}{2}}} (\tilde{p} + n^2)^u \tilde{N}^v + \lambda^k, \text{ par (6.16).} \end{aligned}$$

$\tilde{p} = N, \tilde{N} = p$, et par hypothèse $p \leq N$, donc

$$\mathbf{E} \left[\|\tilde{J}\|^k \right] \leq \alpha^k \lambda^{-n} 4^n N^k (N + n^2)^n p^{1-n} + \lambda^k \alpha^k.$$

$$\text{Posons } \begin{cases} \lambda = 4 \left(\frac{N + n^2}{p} \right) \left(1 + \frac{1}{\sqrt{n}} \right), \\ n = 2 \lfloor N^{\frac{1}{4}} \rfloor \end{cases}$$

$$\begin{aligned} \mathbf{E} \left[\|J\|^k \right] &\leq \alpha^k N^k p \left(1 + \frac{1}{\sqrt{n}} \right)^{-n} + \lambda^k \alpha^k \\ &\leq \alpha^k N^k p \exp \left(-\sqrt{n} + \frac{1}{2} \right) + \lambda^k \alpha^k \\ &\leq \alpha^k N^k p \exp \left(-\sqrt{2} N^{\frac{1}{8}} + \frac{1}{2} \right) + 4^k \left(\frac{N + n^2}{p} \right)^k \left(1 + \frac{1}{\sqrt{n}} \right)^k \alpha^k. \end{aligned}$$

Pour $N \rightarrow \infty$, on obtient

$$\mathbf{E} \left[\|J\|^k \right] \leq 4^k + \mathcal{O}(N^{-\frac{1}{8}}), \text{ pour } k \geq 1 \quad (6.33)$$

Des relations (6.29),(6.30),(6.31), (6.32) et (6.33), on déduit

$$\mathbf{E}(\Psi_k)^2 \leq 20 + \alpha + \mathcal{O}(N^{-\frac{1}{8}}),$$

ce qui est la relation cherchée.♣

Remarques : 1. Nous avons supposé que $p \leq N$. En fait, l'hypothèse minimale pour avoir la propriété de s.a. est $\frac{p}{N^2} \rightarrow 0$ (d'après la relation (6.25), qui est toujours valable).

2. Il est possible de définir l'énergie libre associée au modèle de Little, de la même manière que pour le modèle de Hopfield, en remplaçant $H_{N,p}$ par $E_{N,p,\beta}^P$ (voir proposition VI.4). Dans l'article [5], Amit, Gutfreund, et Sompolinsky obtiennent par des approximations non rigoureuses que pour toute température et p fini, la limite thermodynamique

de l'énergie libre du modèle de Little est exactement 2 fois celle du modèle de Hopfield. Cependant, l'étude rigoureuse des propriétés de l'énergie libre du modèle de Little est un problème difficile ; étant donnée la forme du hamiltonien $E_{N,p,\beta}^{\mathcal{P}}$, il n'est plus possible d'écrire la fonction de partition sous forme d'intégrale gaussienne, et les résultats obtenus sont plus faibles. Ainsi, la convergence presque sûre de cette variable aléatoire et de son espérance (analogue du corollaire VI.3.2) restent un problème ouvert.

En utilisant la même méthode que pour le modèle de Hopfield, nous pouvons montrer que l'énergie libre associée au modèle de Little vérifie la propriété de *self-averaging*, si $p < \gamma(N)\sqrt{N}$, où $\gamma(N) \xrightarrow{N \rightarrow \infty} 0$.

En effet, posons pour $k = 1, \dots, N$, et $t \in [0, 1]$,

$$\tilde{E}_{N,p,\beta}(t, k) = -\frac{1}{\beta} \sum_{\substack{i=1 \\ i \neq k}}^N \log 2 \cosh \left(\beta \sum_{j=1}^N J_{ij} x_j + t \beta J_{ik} x_k \right) - \frac{t}{\beta} \log 2 \cosh \left(\beta \sum_{j=1}^N J_{kj} x_j \right),$$

(on a $\tilde{E}_{N,p,\beta}(t = 1, k) = E_{N,p,\beta}^{\mathcal{P}}$),

$$F_{N,p,\beta}^{\mathcal{P}}(\underline{\xi}; k, t) = -\frac{1}{\beta} \log \sum_{\underline{x} \in X_N} \exp \left(-\beta \tilde{E}_{N,p,\beta}(t, k) \right),$$

et

$$\Psi_k^{\mathcal{P}} = \mathbf{E}[F_{N,p,\beta}^{\mathcal{P}}(\underline{\xi}; k, t = 1) \mid \mathcal{F}_k] - \mathbf{E}[F_{N,p,\beta}^{\mathcal{P}}(\underline{\xi}; k, t = 1) \mid \mathcal{F}_{k-1}].$$

Pour le modèle de Hopfield, nous avons majoré $\mathbf{E}[(\Psi_k)^2]$ par une constante (lemme VI.4.4). Malheureusement, nous obtenons ici une estimation plus faible, car les techniques développées par Shcherbina et Tirozzi ne fonctionnent plus. Nous avons seulement :

$$E[(\Psi_k^{\mathcal{P}})^2] \leq C p^2. \quad (6.34)$$

On en déduit que la propriété de *self-averaging* est vérifiée si $\frac{p^2}{N} \rightarrow 0$ (par le lemme VI.4.3).

Montrons la majoration (6.34) : la fonction $\Psi_k^{\mathcal{P}}$ peut s'écrire

$$\begin{aligned} \Psi_k^{\mathcal{P}} &= \int_0^1 dt \left\{ \mathbf{E} \left[\frac{d}{dt} F_{N,p,\beta}^{\mathcal{P}}(\underline{\xi}; k, t) \mid \mathcal{F}_k \right] - \mathbf{E} \left[\frac{d}{dt} F_{N,p,\beta}^{\mathcal{P}}(\underline{\xi}; k, t) \mid \mathcal{F}_{k-1} \right] \right\} \\ &= \int_0^1 dt \left\{ \mathbf{E} \left[\left\langle \frac{d}{dt} (-\beta \tilde{E}_{N,p,\beta}(t, k)) \right\rangle_{t,k} \mid \mathcal{F}_k \right] - \mathbf{E} \left[\left\langle \frac{d}{dt} (-\beta \tilde{E}_{N,p,\beta}(t, k)) \right\rangle_{t,k} \mid \mathcal{F}_{k-1} \right] \right\}, \end{aligned}$$

et

$$\frac{d}{dt}(-\beta\tilde{E}_{N,p,\beta}(t, k)) = \sum_{i=1}^N \beta J_{ik} x_k \tanh\left(\beta \sum_{j=1}^N J_{ij} x_j + t\beta J_{ik} x_k\right) + \log \cosh\left(\beta \sum_{j=1}^N J_{kj} x_j\right) + \log 2,$$

d'où

$$\begin{aligned} \left| \frac{d}{dt}(-\beta\tilde{E}_{N,p,\beta}(t, k)) \right| &\leq 2\beta \sum_{i=1}^N |J_{ik}| + \log 2 \\ &\leq 2\beta p + \log 2 \end{aligned}$$

et

$$(\Psi_k)^2 \leq 16\beta^2 p^2 + \mathcal{O}(p).$$

Remarquons qu'il s'agit d'estimations grossières que l'on peut espérer améliorer.

Chapitre 7

Conclusion

Nous allons résumer l'ensemble de cette thèse, et discuter des problèmes ouverts.

Nous avons d'abord considéré le modèle déterministe de Hopfield : après avoir vu que la région de mémorisation sans erreurs est limitée à $\frac{p}{c \log N}$, nous avons étudié la convergence directe des images originales bruitées, et des mosaïques de fragments d'images. Étant donnée la complexité des itérées successives de l'application associée à la dynamique, nous nous sommes limités à une convergence en une étape. L'étude de la convergence en plusieurs étapes reste un problème très riche. Les états stables atteints par la dynamique étant des minima au moins locaux du hamiltonien, nous nous sommes intéressés à la structure de ces minima. Nous avons obtenu la convergence presque sûre du hamiltonien normalisé vers des constantes bien déterminées, pour les images originales et des combinaisons de ces images ; ceci suggère que les images originales sont des minima globaux de la fonction énergie. Cependant, ce problème n'est pas résolu, et on souhaiterait également montrer que ce sont les seuls minima globaux (résultat obtenu par simulation). D'autre part, il serait intéressant de confirmer la stabilité d'un nombre exponentiel (c^p , $c < 2$) d'états parasites, annoncée par J.Komlos et R.Paturi.

Les propriétés de stabilité et de convergence directe pour la dynamique parallèle déterministe (modèle de Little) se déduisent aisément de ceux obtenus dans le cas séquentiel. De plus, nous montrons la convergence directe de configurations corrélées avec une image originale, pour $p < \sqrt{\frac{N}{\log N}}$. L'étude des fluctuations de la fonction énergie associée, plus difficile que pour le modèle de Hopfield, reste à faire.

Nous avons ensuite exposé les résultats connus concernant la dynamique stochastique, ce qui conduit naturellement au formalisme des états de Gibbs pour un réseau de taille infini, notion introduite dans les années 60 – 70 pour aborder les problèmes de transition de phases en Mécanique Statistique. Comme nous l'avons déjà souligné, une des questions ouvertes les plus intéressantes pour le modèle de Hopfield est de prouver l'existence des mesures de Gibbs extrémales correspondant aux différentes images mémorisées, pour p proportionnel à N .

Enfin, nous avons étudié les propriétés asymptotiques de l'énergie libre : cette variable aléatoire converge presque sûrement et dans $\mathbf{L}^2(\Omega, \mathcal{F}, \mathbf{P})$ vers une constante dépendant de $\beta = T^{-1}$, pour toute température $T > 0$, et sous la condition $\frac{p}{N} \rightarrow 0$. De plus, elle vérifie la propriété de self-averaging, i.e. sa variance converge vers 0, pour toute température, et sous des conditions plus faibles pour p ($p \leq N$ convient). Ces propriétés impliquent en particulier que pour $\beta < 1$, si $\frac{p}{N} \rightarrow 0$, les énergies libres “annealed” et “quenched” coïncident asymptotiquement (au sens de la convergence p.s. et dans \mathbf{L}^2). Il est naturel de se demander si cette propriété est encore vraie pour $p = \alpha N$ ($\alpha > 0$). Ce problème a été étudié par E.Scacciatelli et B.Tirozzi [44] : ils obtiennent la convergence en probabilité et dans \mathbf{L}^r ($r < \infty$) de l'énergie libre vers la limite de l'énergie libre “annealed”, dans la région ergodique $T > 1 + \sqrt{\alpha}$. Cependant, la méthode utilisée est calquée sur celle conçue par M.Aizenman, J.L.Lebowitz, et D.Ruelle pour le modèle de Sherrington-Kirkpatrick [1] : il s'agit de montrer par un calcul de moments la convergence en loi de la fonction de partition (divisée par son espérance) vers une variable de loi log-normale. La méthode, déjà pénible pour le modèle SK, pose des problèmes combinatoires très difficiles de développement en graphes. Après avoir longuement travaillé sur l'article de Scacciatelli et Tirozzi, nous sommes convaincus que certaines estimations données dans l'annexe de cet article ne sont pas complètement prouvées. Ceci reste donc, à mon avis, un problème intéressant : il serait agréable de trouver une démonstration plus “élégante”. F.Comets et J.Neuveu ont récemment mis à profit la théorie du Calcul Stochastique pour le modèle SK, généralisant ainsi les résultats de Aizenman, Lebowitz et Ruelle. Il est légitime d'espérer que cette théorie puisse fonctionner également dans le cas du modèle de Hopfield, bien que les difficultés techniques soient plus grandes.

Nous allons terminer cette thèse en évoquant des extensions possibles du modèle de Hopfield : des résultats existent pour le modèle de Potts (neurones à q états possibles, avec $q \geq 2$)[20][19], et le modèle de Hopfield dilué (les interactions synaptiques sont détruites aléatoirement avec une probabilité $1 - \tau(N)$, $\tau(N)$ décroissant vers 0 plus lentement que $\mathcal{O}(\frac{1}{N})$)[9]. Dans l'article [19], les auteurs reprennent les calculs de C.Newman [38], et les résultats coïncident : les deux modèles sont fondamentalement très proches (le modèle de Hopfield correspond à $q = 2$), et ils obtiennent naturellement comme capacité de stockage $p \leq \alpha_c(q)N$. Dans l'article [9], il est prouvé que si $\frac{p}{\tau(N)N} \rightarrow 0$, alors l'énergie libre associé au réseau dilué a les mêmes propriétés que celle du modèle de Hopfield classique. Cependant, les images originales sont toujours considérées indépendantes ; or dans les applications des modèles de réseaux neuronaux, les données à manipuler sont très souvent corrélées, et pour modéliser des phénomènes cognitifs biologiques, il est indispensable également de le supposer. Cela ouvre un vaste champ de recherche, car actuellement, les seuls résultats connus ont été obtenus par des simulations et des techniques non rigoureuses (méthode des répliques). Il existe deux types de corrélations possibles : les corrélations “sémantiques”, entre différentes images, et les corrélations “spatiales”, entre différents sites d'un réseau (voir [46]). Cet article soulève des questions intéressantes, puisque les auteurs concluent que les corrélations spatiales augmentent la capacité de stockage du modèle de Hopfield.

En conclusion, le modèle de Hopfield, de par sa complexité et la richesse de ses comportements, intéresse Informaticiens, Physiciens, Mathématiciens et Neurophysiologistes. Les

outils des Probabilités et de la Mécanique Statistique ont déjà permis d'obtenir de nombreux résultats. Il mérite cependant d'être plus connu des Probabilistes, car les problèmes qu'il soulève seront peut-être résolus grâce à des théories encore inexploitées, comme le Calcul Stochastique, par exemple.

Bibliographie

- [1] Aizenman, M., Lebowitz, J. L., Ruelle, D. : Some Rigorous Results on the Sherrington-Kirkpatrick Spin Glass Model. *Commun. Math. Phys.* **112**, 3-20 (1987).
- [2] Albeverio, S., Tirozzi, B., Zegarlini, B. : Rigorous Results for the Free Energy in the Hopfield Model. Preprint 121 Institut für Mathematik Ruhr-Universität Bochum.
- [3] Amaro De Matos, J., Patrick, A., Zagrebnov, V. : Random Infinite-Volume Gibbs States for the Curie-Weiss Random Field. *Journal of Stat.Physics* **66**, 1/2, 139-164 (1992)
- [4] Amit, D., *Modeling Brain Function, the World of Attractor Neural Networks.* Cambridge : Cambridge University Press (1989)
- [5] Amit, D., Gutfreund, G., Sompolinsky, H. : Spin Glass Models of Neural Networks. *Physical Review A*, **32**, 1007-1018 (1985).
- [6] Amit, D., Gutfreund, G., Sompolinsky, H. : Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks. *Physical Review Letters*, **55**, 1530-1533 (1985)
- [7] Amit, D., Gutfreund, G., Sompolinsky, H. : Statistical Mechanisms of Neural Networks near Saturation. *Annals of Physics*, **173**, 30-67 (1987)
- [8] Aubin, J.P. : *Mathematical Methods for Neural Networks.* Polycopié du cours dispensé au Centre de Physique des Houches en Mars 1992.
- [9] Bovier, A., Gayraud, V. : Rigorous Results on the Thermodynamics of the Dilute Hopfield Model. *Journal Stat.Physics*, **72** (1993)
- [10] Bovier, A., Gayraud, V., Picco, P. : Gibbs States of the Hopfield Model in the Regime of Perfect Memory. Prépublication du Centre de Physique Théorique de Marseille (1993)
- [11] Brezis, H. : *Analyse Fonctionnelle : Théorie et Applications.* Paris : Masson (1983)
- [12] Comets, F., Neveu, J. : *The Sherrington-Kirkpatrick Model of Spin Glasses and Stochastic Calculus : the High Temperature Case.* Prépublication (1993)
- [13] Dacunha-Castelle, D., Dufflo, M. : *Probabilités et Statistiques.* Paris : Masson (1982)
- [14] Derrida, B., Gardner, E., Zippelius, A. : An Exactly Solvable Asymptotic Neural Network Model. *Europhysics Letter*, **4**, 167-173 (1987)
- [15] Dieudonné, J. : *Éléments d'Analyse I.* Paris : Gauthier Villars (1968)

- [16] Dunford, N., Schwartz, J. : Linear Operators I. New York : Intersciences publishers (1958)
- [17] Erdos, P., Spencer, J. : Probabilistic Methods in Combinatorics. New York : Academic Press (1974)
- [18] Feller, W. : An Introduction to Probability Theory and its Applications. New York : John Wiley and Sons (1950)
- [19] Ferrari, P., Martinez, S., Picco, P. : A Lower Bound for the Memory Capacity in the Potts-Hopfield Model *Journal Stat.Physics*, **66**, 5/6, 1643-1652 (1992)
- [20] Gayrard, V. : The Thermodynamic Limit of the q-State Potts-Hopfield Model with Infinitely many Patterns. *Journal Stat.Physics*, **68**, 977-1011 (1992)
- [21] Georgii, H. : Gibbs Measures and Phase Transitions. Berlin : Walter de Gruyter (1988)
- [22] Girko, V. : Spectral Theory of Random Matrices. Moscow, Nauka (1988)
- [23] Hebb, D. : The Organization of Behavior : a Neurophysiological Theory. New York : Wiley (1949)
- [24] Henrici, P. : Applied and Computational Complex Analysis, volume II. New York : John Wiley and Sons (1977)
- [25] Hertz, J., Krogh, A., Palmer, R. : Introduction to the Theory of Neural Computation. Redwood City : Addison-Wesley (1991)
- [26] Hopfield, J. : Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proc.Natl.Acad.Sci.USA*, **79**, 2554-2558 (1982)
- [27] Koch, H. : A Free Energy Bound for the Hopfield Model. *Journal of Physics A : Math.Gen.*, **26**, L353-355 (1993)
- [28] Komlos, J., Paturi, R. : Convergence Results in a Autoassociative Memory Model. *Neural Networks* **1** 239-250 (1988)
- [29] Little, W. : The existence of Persistent States in The Brain. *Mathematical Biosciences*, **19**, 101-120 (1974)
- [30] Little, W., Shaw, G. : A Statistical Theory of Short and Long Term Memory . *Behavioral Biology*, **14** 115-133 (1975)
- [31] Little, W., Shaw, G. : Analytic Study of the Memory Storage Capacity of a Neural Network. *Mathematical Biosciences*, **39**, 281-290 (1978)
- [32] MacCulloch, W., Pitts, W. : A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bull.Math.Biophys.*, **5**, 115-133 (1943)
- [33] MacEliece, R., Posner, E., Rodemich, E., Venkatesh, S. : The Capacity of the Hopfield Associative Memory. *IEEE Transactions on Information Theory*, **33**, 4, 461-482 (1987)
- [34] Maes, C., Zagrebnev, V. : On the Parallel Dynamics of a Multi-layered Perceptron. Preprint de l'Institut de Physique Théorique de Louvain (1991)
- [35] Minlos, R. : Gibbs'Limit Distribution. *Funct.Anal.Appl.* **2**, 60-73 ; **3**, 40-53 (1967)

- [36] Montgomery, B., Vijaya Kumar, B. : Evaluation of the Use of the Hopfield Neural Network Model as a Nearest-Neighbour Algorithm. *Applied Optics*, **25**, 3759-3766 (1986)
- [37] Muller, B., Reinhardt, J. : *Neural Networks, An Introduction*. Berlin : Springer Verlag (1991)
- [38] Newman, C. : Memory Capacity in Neural Networks Model : Rigorous Lower Bounds. *Neural Networks*, **1**, 223-238 (1988)
- [39] Pastur, L., Shcherbina, M. : Absence of Self-Averaging in the Order Parameter in the Sherrington-Kirkpatrick Model. *Journal of Stat.Physics*, **62**, 1-19 (1991)
- [40] Patrick, A, Zagrebnoy, V. : A Probabilistic Approach to Parallel dynamics for the Little-Hopfield Model. *Journal of Physics A : Math.Gen.*, **24**, 3413-3426 (1991)
- [41] Peretto, P. : Collective Properties of Neural Networks : a Statistical Physics Approach. *Biological Cybernetics*, **50**, 51-62 (1984)
- [42] Prum, B. : *Processus sur un réseau et mesures de Gibbs*. Paris : Masson (1986)
- [43] Ruelle, D. : Correlation Functions of Classical Gases. *Ann. of Phys. (N.Y.)*, **25**, 109-120 (1963)
- [44] Scacciatelli, E., Tirozzi, B. : Fluctuation of the Free Energy in the Hopfield Model. *Journal of Stat.Physics*, **67**, 5/6, 981-1008 (1992)
- [45] Shcherbina, M., Tirozzi, B. : Exact Mean Field Theory for some Hopfield Model. Preprint (1992)
- [46] Tarkowski, W., Lewenstein, M. : Storage of Sets of Correlated Data in Neural Network Memories. *Journal of Physics A : Math.Gen.*, **26**, 2453-2469 (1993)
- [47] Vermet, F. : Convergence de la Variance de l'Énergie Libre pour le Modèle de Hopfield. *C.R.A.S.* **315**, Série I, 1001-1004 (1992)