

# Statistical Physics Methods for Community Detection

Thèse N° 7208

Présentée le 13 décembre 2019

à la Faculté informatique et communications

Laboratoire de théorie des communications

Programme doctoral en informatique et communications

pour l'obtention du grade de Docteur ès Sciences

par

**Chun Lam CHAN**

Acceptée sur proposition du jury

Prof. M. C. Gastpar, président du jury

Dr. N. Macris, directeur de thèse

Prof. H. Pfister, rapporteur

Prof. A. Coja-Oghlan, rapporteur

Prof. E. Abbé, rapporteur

2019



# Statistical Physics Methods for Community Detection

Chun Lam Chan

EPFL - Ecole Polytechnique Fédérale de Lausanne

Thesis presented to the faculty of computer and communication sciences for  
obtaining the degree of Docteur ès Sciences

Accepted by the jury:

**N. Macris**

Thesis director

**E. Abbé**

Expert

**A. Coja-Oghlen**

Expert

**H. Pfister**

Expert

**M. Gastpar**

President of the jury

Ecole Polytechnique Fédérale de Lausanne, 2019

# Abstract

---

This thesis is devoted to information-theoretic aspects of community detection. The importance of community detection is due to the massive amounts of scientific data today that describes relationships between items from a network, e.g., a social network. Items from such a network can be inherently partitioned into a known number of communities, but the partition can only be inferred from the data. To estimate the underlying partition, data scientists can apply any type of advanced statistical techniques; but the data could be very noisy, or the number of data is inadequate. A fundamental question here is about the possibility of weak recovery: does the data contain a sufficient amount of information that enables us to produce a non-trivial estimate of the partition?

For the purpose of mathematical analysis, the above problem can be formulated as Bayesian inference on generative models. These models, including the *stochastic block model* (SBM) and *censored block model* (CBM), consider a random graph generated based on a hidden partition that divides the nodes in the graph into labelled groups. In the SBM, nodes are connected with a probability depending on the labels of the endpoints. Whereas, in the CBM, hidden variables are measured through a noisy channel, and the measurement outcomes form a weighted graph. In both models, inference is the task of recovering the hidden partition from the observed graph. The criteria for weak recovery can be studied via an information-theoretic quantity called *mutual information*. Once the asymptotic mutual information is computed, phase transitions for the weak recovery can be located.

This thesis pertains to rigorous derivations of single-letter variational expressions for the asymptotic mutual information for models in community detection. These variational expressions, known as the replica predictions, come from heuristic methods of statistical physics. We present our development of new rigorous methods for confirming the replica predictions. These two methods are based on extending the recently introduced *adaptive interpolation method*.

We prove the replica prediction for the SBM in the dense-graph regime with two groups of asymmetric size. The existing proofs in the literature are indirect, as they involve mapping the model to an external problem whose mutual information is determined by a combination of methods. Here, on the

contrary, we provide a self-contained and direct proof.

Next, we extend this method to sparse models. Before this thesis, adaptive interpolation was known for providing a conceptually simple proof for replica predictions for dense graphs. Whereas, for a sparse graph, the replica prediction involves a more complicated variational expression, and rigorous confirmations are often lacking or obtained by rather complicated methods. Therefore, we focus on a simple version of CBM on sparse graphs, where hidden variables are measured through a binary erasure channel, for which we fully prove the replica prediction by the adaptive interpolation.

The key for extending the adaptive interpolation to a broader class of sparse models is a concentration result for the so-called “multi-overlaps”. This concentration forms the basis of the replica “symmetric” prediction. We prove this concentration result for a related sparse model in the context of physics. This provides inspiration for further development of the adaptive interpolation.

**Keywords:** Bayesian inference, community detection, stochastic block models, censored block models, graphical models, mutual information, statistical physics, spin glass, replica method, cavity method.

# Résumé

---

Cette thèse est dédiée à l'étude de la détection des communautés du point de vue de la théorie de l'information. L'importance de la détection des communautés peut être justifiée par l'accès, aujourd'hui, à une quantité considérable de données scientifiques qui décrivent les relations entre les composantes d'un certain réseau, tel les réseaux sociaux. Pour un réseau donné, les composantes peuvent être partitionnées en un nombre connu de communautés (le nombre étant une caractéristique du réseau), mais les partitions, en tant que telles, ne peuvent être déduites que des données. Aussi, les chercheurs sont-ils capables d'appliquer n'importe quelle technique avancée en statistique afin d'estimer les partitions sous-jacentes. Toutefois, les données peuvent être soit, corrompues par un bruit important, soit en nombre insuffisant. À partir de là, une question fondamentale se pose concernant la possibilité d'une "récupération faible": les données contiennent-elles des informations suffisantes qui nous permettent de produire une estimation, qui ne serait pas banale, de la partition?

Du point de vue de l'analyse mathématique, le problème cité ci-dessus peut être formulé en tant qu'une inférence Bayésienne sur des modèles générateurs. Ces modèles, dont le *modèle aléatoire en bloc* et le *modèle censuré en bloc*, considèrent un graph aléatoire généré à partir d'une partition cachée qui divise les noeuds du graphe en groupes étiquetés. Pour le modèle aléatoire en bloc, deux noeuds quelconques sont connectés avec une probabilité qui dépend de la classe de ces noeuds. Cependant, pour le modèle censuré en bloc, des variables cachées sont mesurées à travers un canal à bruit et les mesures réalisées forment un graph pondéré. Dans ces deux cas, l'inférence revient à déduire la partition sous-jacente et implicite du graphe observé. Le critère d'une déduction faible peut être étudié via *l'information mutuelle*, une quantité fondamentale en théorie de l'information. Intuitivement, l'information mutuelle quantifie l'information contenue dans le graphe observé à propos de la partition. Une fois l'information mutuelle asymptotique calculée, nous pouvons localiser les transitions de phases pour la "récupération faible".

Cette thèse vise à dériver, de façon rigoureuse, une caractérisation (à lettre unique) des expressions variationnelles pour l'information mutuelle asymptotique relative aux modèles utilisés pour la détection des communautés. Des méthodes heuristiques en physique statistique sont à la base de ces expres-

sions variationnelles, connues sous le nom de prédictions des répliques. Nous présentons de nouvelles méthodes pour confirmer rigoureusement les prédictions de répliques. Ces méthodes généralisent la *méthode d'interpolation adaptative* récemment développée dans la littérature.

En un premier lieu, nous calculons la prédiction des répliques pour le modèle aléatoire en bloc lorsque nous considérons un régime de graphe dense avec deux groupes de tailles asymétriques. Les preuves présentes dans la littérature sont indirectes vu qu'elles réduisent le modèle à un problème externe dont l'information mutuelle est déterminée par une combinaison de méthodes. En revanche, nous proposons une preuve directe et autonome.

En un second lieu, nous généralisons la méthode d'interpolation adaptative pour les modèles creux. L'interpolation adaptative donne une preuve simple et unifiée pour les prédictions des répliques, mais elle est restreinte aux graphes denses. Quant aux graphes creux, la prédiction des répliques implique une expression variationnelle plus compliquée, et les démonstrations rigoureuses sont souvent soit absentes soit obtenues à travers des méthodes compliquées. Pour cela, nous nous concentrons sur une version simplifiée du modèle censuré en bloc pour les graphes creux, où les variables cachées sont mesurées à travers un canal binaire à effacement. Pour cette version, nous démontrons la prédiction des répliques entièrement par l'interpolation adaptative.

La clé pour la généralisation de l'interpolation adaptative à des classes plus larges de modèles creux est un résultat de concentration pour des quantités appelées “multi-overlaps” et qui constitue la base de la prédiction des répliques. Nous démontrons ce résultat de concentration pour des modèles creux similaires à ceux de la physique. Ceci fournit une source d'inspiration pour de futurs développements sur l'interpolation adaptative.

**Mots-clés:** Inférence Bayésienne, détection de communautés, modèles aléatoires en bloc, modèles censurés en bloc, modèles graphiques, information mutuelle, physiques statistiques, verre de spin, méthode des répliques, méthode des cavités.



# Acknowledgements

---

I am very fortunate to have been a student of Nicolas Macris. I still remember his opening question in our first research meeting, “What is the Bethe free energy?” What immediately came to my mind during the meeting was all the technical derivations that start from the Gibbs’ formulation. But Nicolas showed me how to answer it concisely in one single sentence. It turns out that this conversation epitomized an important part of my PhD training, as during the last four years Nicolas took great care in teaching me to think critically and to be meticulous. Indeed, my thesis pertains to a deeper discussion of this very question. Therefore, I am greatly indebted to him.

If Nicolas was the source of free energy of this thesis, Jean Barbier would have been the catalyst (so that less free energy was required to reach the transition state). I am thankful to Jean for suggesting research topics to me and for bringing up relevant references. I also treasure going together with him to the gym regularly together after work.

I would like to thank Rüdiger Urbanke for the continual support in the lab. Being the teacher assistant of his classes was also a fun experience.

It was an honor to have Emmanuel Abbé, Amin Coja-Oghlen, Michael Gastpar and Henry Pfister on my thesis committee. Their careful reading and insightful comments improved this thesis. I also thank Amin for flying from Germany for my thesis presentation and Henry for being my thesis committee member again, after my master thesis.

My PhD studies was under the big family of Information Processing Group (IPG). I am fortunate to not have missed some inspiring lectures from Emre Teletar and Olivier Lévêque. Damir Laurenzi, France Faille, Françoise Behn and Muriel Bardet are the heroes that keep the group running smoothly. I especially thank Damir and Muriel for always helping me out with tech. / admin. / life, promptly and swiftly. Besides, I gratefully acknowledge the effort of Holly Cogliati, who helped me survive in the last stage of my PhD thesis writing. Many thanks should also go to the junior members in IPG for their invaluable friendships: Amedeo, Chien-Yi, Clément, Daria, Elie, Emanuele, Erixhen, Jingge, Karol, Kirill, Mani, Marco, Mohamad, Rafah, Raffaele, Rajai, Reka, Saied, Serj, Su and Wei. In particular, it was my pleasure being the office-mate of Daria and Mohamad, and a close neighbor of Clément. They

never let me down.

My gratitude extends to those who cheered me up regularly, beyond week-days. I am sincerely grateful to Jaggi, Mayank, Navin, Pascal and Shashank for giving me their encouraging words whenever we met at several conferences and my trips home. Thanks to them, their words always helped me to set sail again. Here is a list of things that I would definitely miss when I leave Lausanne: board games with Jian-Lin, Kun-Han and Shang-Jung, drinking with Clément, Prakhar and Praneeth, going to Henan with Hanjie and Su, going to Zayeka with Aditya, David and Naman, impromptu celebrations with Hon Ming, Kwok Yui and other Hong Kong buddies, and via ferrata with Jhih-Huang.

Last but not least, I deeply thank with love Ying Mao for her unwavering support.

# Contents

---

<b>Abstract</b>	<b>i</b>
<b>Résumé</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>Contents</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Community detection . . . . .	2
1.2 Information and coding theory . . . . .	5
1.3 Spin glass . . . . .	7
1.4 The cavity and replica methods . . . . .	11
1.5 Implication of the replica symmetric formula . . . . .	15
1.6 Related works on proving the replica prediction . . . . .	17
1.7 Contributions and organization of this thesis . . . . .	18
1.7.1 Bibliographic notes . . . . .	20
<b>2 Preliminary: Adaptive Interpolation for Spiked Wigner Models</b>	<b>21</b>
2.1 Main steps . . . . .	23
2.2 Tools . . . . .	26
2.2.1 The Nishimori identities . . . . .	26
2.2.2 Gaussian integration by parts . . . . .	27
2.2.3 Liouville formula . . . . .	27
<b>3 Dense Stochastic Block Model</b>	<b>29</b>
3.1 Introduction . . . . .	29
3.2 Setting and results: asymmetric two-group SBM . . . . .	30
3.3 Adaptive path interpolation . . . . .	34
3.3.1 The upper bound: proof of Theorem 3.1 . . . . .	37
3.3.2 The lower bound: proof of Theorem 3.2 . . . . .	38
3.4 The fundamental sum rule: proof of (3.15) . . . . .	40

3.4.1	Term $D_1$ .	41
3.4.2	Term $D_2$ .	44
3.4.3	Term $D_3$ .	44
3.4.4	Final derivations of the sum rule.	45
3.5	Concentration of overlap: proof of Lemma 3.2	46
3.5.1	Proof of the fluctuation identity: Lemma 3.8	48
3.5.2	Bound on thermal fluctuations: proof of Lemma 3.9	49
3.5.3	Bound on quenched fluctuations: proof of Lemma 3.10	49
3.6	Appendix	52
3.6.1	Approximate integration by parts: proof of lemma 3.4	52
3.6.2	Mutual information and free energy: proof of Proposition 3.1	55
3.6.3	Small error terms in the sum rule: proof of (3.30)	56
3.6.4	Concentration of free energy: proof of Lemma 3.11	58
<b>4</b>	<b>Sparse Censored Block Model</b>	<b>63</b>
4.1	Introduction	63
4.2	Setting and result	64
4.2.1	Censored block model	64
4.2.2	The replica symmetric formula for the average conditional entropy	66
4.3	Two preliminary tools	67
4.3.1	Nishimori identities	67
4.3.2	Griffiths-Kelly-Sherman inequalities for the BEC	69
4.4	The adaptive path interpolation method	69
4.4.1	The $(t, s)$ -interpolating model	70
4.4.2	Evaluating the free entropy change along the interpolation	72
4.4.3	The lower bound	73
4.4.4	The upper bound	74
4.5	Proof of the fundamental sum rule (4.28)–(4.29)	77
4.6	Concentration of overlaps I: proof of lemmas 4.3 and 4.6	79
4.6.1	Proof of lemma 4.3	80
4.6.2	Proof of lemma 4.6	80
4.7	Concentration of overlaps II: proof of lemmas 4.9, 4.10, 4.11	81
4.7.1	Useful derivative formulas	81
4.7.2	Proof of Lemma 4.9	84
4.7.3	Proof of Lemma 4.10	85
4.7.4	Proof of Lemma 4.11	86
4.8	Appendix	88
4.8.1	Direct proof of identity (4.12) for symmetric distributions	88
4.8.2	Rewriting the replica formula: proof of (4.89)	88
4.8.3	Same supremum of two free entropy functionals: proof of Lemma 4.1	89
4.8.4	Proofs of technical lemmas	93

4.8.5	Existence of the optimal interpolation path: proof of Lemma 4.8 . . . . .	96
4.8.6	Derivatives of the conditional entropy: proof of (4.60)–(4.65) . . . . .	97
4.8.7	Concentration of free entropy . . . . .	100
4.8.8	Illustration of the replica formula . . . . .	104
<b>5</b>	<b>Multi-overlaps for Ferromagnetic Spin Models on Sparse Graphs</b>	<b>107</b>
5.1	Introduction . . . . .	107
5.2	Ferromagnetic spin models and overlap concentration . . . . .	109
5.3	Proofs of concentrations for the overlaps . . . . .	113
5.3.1	Preliminary remarks . . . . .	113
5.3.2	Thermal fluctuations of overlaps: proof of Theorem 5.1 . . . . .	114
5.3.3	Disorder fluctuations of the magnetization . . . . .	114
5.3.4	Poisson fluctuations of overlaps . . . . .	116
5.3.5	Last type of fluctuations of overlaps . . . . .	118
5.3.6	Proof of Theorem 5.3 . . . . .	126
5.4	Appendix . . . . .	127
5.4.1	Proof of the approximation inequality (5.6) . . . . .	127
5.4.2	Multivariate Harris inequality . . . . .	127
5.4.3	On the concentration and existence of the free energy . . . . .	128
<b>6</b>	<b>Discussion and Conclusion</b>	<b>131</b>
	<b>Bibliography</b>	<b>135</b>



---

# 1

## Introduction

---

What are the signs that indicate we live in the age of social networks? At the time of this writing, Hong Kong has been hit by large-scale demonstrations for months. Millions of people have taken part in this movement, but no leadership behind it has been reported. It is the massive power of online social networks (Facebook, Telegram, etc.) that connects the people. Yet the overall social network has a two-community structure — the democrats and the pro-government parties share opinions in their own communities, but there are no (or very few) constructive interactions between the two communities. In general, community structures are formed in many real-world networks [1, 2]. And the task of differentiating, based on the topology of the network, the members of the communities has become an important statistical inference problem known as community detection.

In this thesis, we will focus on stochastic block models and censored block models, the two canonical random-graph models that are often used in the theoretical study of community detection. A fundamental question here is about when weak recovery of the communities is possible. The terminology of this problem resembles a classic communication problem, hence we are curious to use mutual information from information and coding theory in order to study the correlation between the underlying communities and the random graph. Indeed, weak recovery can be derived from the asymptotic mutual information. Therefore, we want to know how to compute the asymptotic mutual information. Hence, it would be interesting to look into statistical physics, as it studies macroscopic properties of complex interacting systems. In particular, the mutual information can be linked to the free energy in statistical physics. The cavity method and replica method are two heuristic statistical-physics techniques that have been used for the last four decades to provide a mean-field solution for the free energy, called the replica symmetric formula.

This formula can be used to predict phase transitions for weak recovery in community detection. A rigorous derivation of the replica symmetric formula is still needed to prove the prediction.

In this chapter, we introduce more background on the intersection of the three subjects: community detection, information and coding theory, and spin-glass systems. We will end this chapter by stating the main contributions of this thesis — in short, the development of new methods that prove the exactness of the replica symmetric formula to the asymptotic mutual information.

## 1.1 Community detection

As early as the 1930s, sociology began analyzing social network and developed the topic of “block-modeling”, an empirical procedure that effectively cluster similar objects into groups. The discussion spread to other fields of study during the last century, due to the emergence of a plethora of social networks, biological networks and computer networks. The early history of network science is well summarized in [3, 4].

Traditional algorithms for clustering are those with a min-cut approach, based on graph partitioning in computer science, where the total communication cost between computers performing parallel computing is minimized. The seminal advance in the algorithmic development was subsequently given by [1, 5]. They focused on networks with underlying community structure and proposed properties of the inter-communities edges such that removing those edges uncovered community structures. Such algorithms work well on real benchmarks. Moreover, research on community detection flourished, subsequently we have a rich literature devoted to algorithmic solutions.

Understanding the fundamental limits for community detection requires mathematical abstraction of the models. A generative model would enable us to formulate the problem as Bayesian inference, thus enable us to build theoretical understanding by using the tools of probability, hence the recent interest in generative models of community detection. Starting from [6], many more researchers studied both the fundamental limits and algorithmic solutions for the generative models. Building these models naturally appeals to the Erdős-Rényi graphs, the primitive random-graph models in graph theory in which every pair of nodes is connected with the same probability. The canonical models for community detection can be largely considered as variants of the Erdős-Rényi graphs.

Here, we introduce two canonical models called stochastic block model and censored block model. We assume that parameters in the description of the models are known. In this context, the Bayesian approach is sometimes called “Bayesian optimal”. Our interest in these models is the task of weak recovery; it will be defined at the end of this section.



### Stochastic block model

The most popular generative model is the *stochastic block model* (SBM). SBM has a long history and has attracted the attention of many disciplines. It was first introduced as a model of community detection in the network and statistics literature [7], as a problem of finding graph bisections in theoretical computer science [8], and was proposed as a model for inhomogeneous random graphs [9, 10]. A partition of nodes into labeled groups is hidden to the observer who is given only a random graph generated on the basis of the partition. The task of the observer is to recover the hidden partition from the observed graph. A simple setting that lends itself to mathematical analysis is the following. The labels of nodes are drawn i.i.d. from a prior distribution and, for the graph, the edges between pairs of nodes are placed independently according to a probability that depends only on the group labels. If the probability is slightly higher (resp. lower) when the pair of nodes have the same label, the model is called assortative (resp. disassortative). Furthermore, we assume that the parameters of the prior- and edge-probability distributions are all known. Note that the recovery task is non-trivial only when parameters are such that no information about the group label is revealed from the degrees of nodes. Much progress has been made in recent years within this simple mathematical setting, and we refer to [11] for a recent comprehensive review and references. We provide Examples 1.1 and 1.2 for readers to have a concrete example of SBM in mind.

**Example 1.1** (Symmetric two-group SBM on sparse graphs). *Let  $\sigma^0 \in \{-1, +1\}^n$  denote the hidden partition of the two communities. Labels  $\sigma_i^0$  are i.i.d. Bernoulli random variables with  $\mathbb{P}(\sigma_i^0 = 1) = 1/2$ . The observation is a graph  $\mathbf{G}$  (represented by the adjacency matrix) constructed by connecting any pair of nodes  $i$  and  $j$  with an edge with probability*

$$\mathbb{P}(G_{ij} = 1 | \sigma_i^0, \sigma_j^0) = \frac{a}{n} + \frac{b}{n} \sigma_i^0 \sigma_j^0$$

where  $a, b$  are constants. The expected degree of a node is  $a(n-1)/n = \Theta(1)$ , hence we say the graph is sparse.  $\square$

Example 1.1 is the most canonical SBM considered in the literature. In Chapter 3 of this thesis, we consider a different regime as follows:

**Example 1.2** (Asymmetric two-group SBM on dense graphs). *Let  $\sigma^0 \in \{-1, +1\}^n$  denote the hidden partition of the two communities. Labels  $\sigma_i^0$  are i.i.d. Bernoulli random variables with  $\mathbb{P}(\sigma_i^0 = 1) = r$ . For convenience we define  $X_i \equiv \phi_r(\sigma_i^0)$  with  $\phi_r(1) = \sqrt{(1-r)/r}$  and  $\phi_r(-1) = -\sqrt{r/(1-r)}$ . The observation is a graph  $\mathbf{G}$  with the transition probability*

$$\mathbb{P}(G_{ij} = 1 | \sigma_i^0, \sigma_j^0) = \bar{p}_n + \Delta_n X_i X_j.$$

where

$$n\bar{p}_n(1 - \bar{p}_n)^3 \xrightarrow{n \rightarrow \infty} \infty, \quad \frac{n\Delta_n^2}{\bar{p}_n(1 - \bar{p}_n)} \xrightarrow{n \rightarrow \infty} \lambda. \quad (1.1)$$

An example of parameters that fulfills (1.1) is  $\bar{p}_n = 1/2$  and  $\Delta_n \sim n^{-1/2}$ . The two constraints (1.1) will be further discussed in Section 3.2. Here we note that this SBM pertains to a dense graph regime, as the expected degree of a node is  $(n - 1)\bar{p}_n \xrightarrow{n \rightarrow \infty} \infty$  (the growth rate can be arbitrarily slow).  $\square$

### Censored block model

Another way to embed community structures into an Erdős-Rényi graph is the *censored block model* (CBM), studied in the context of community detection more recently in [12, 13, 14, 15]. In this model, the nodes also have a hidden partition in labeled groups. The label of each node is i.i.d. drawn from a prior distribution. Any pair of nodes is equally likely to be taken for a noisy measurement. The set of measurement outcomes can be viewed as the weighted version of an Erdős-Rényi graph. An observation is provided by a weighted edge that represents a noisy version of the product of the labels of the connected nodes. The next example is the standard CBM in the literature.

**Example 1.3** (Symmetric two-group CBM). Let  $\sigma^0 \in \{-1, +1\}^n$  denote the partition of the two communities. Labels  $\sigma_i^0$  are i.i.d. Bernoulli random variables with  $\mathbb{P}(\sigma_i^0 = 1) = 1/2$ . An instance of Erdős-Rényi graph is drawn and the weighted version  $\mathbf{G}$  (represented by the adjacency matrix of a weighted graph) is observed. Each edge has a weight  $G_{ij} \in \{-1, +1\}$  that represents the outcome of the transition probability

$$\mathbb{P}(G_{ij} | \sigma_i^0, \sigma_j^0) = (1 - q)\delta_{G_{ij}, \sigma_i^0 \sigma_j^0} + q\delta_{G_{ij}, -\sigma_i^0 \sigma_j^0}.$$

$G_{ij}$  reveals yes/no to the question of whether node  $i$  and node  $j$  are from the same group. It tells the truth with probability  $1 - q$ .  $\square$

We can generalize an Erdős-Rényi graph in the above CBM to a hypergraph and also modify the transition probability. The CBM we consider in Chapter 4 is the following.

**Example 1.4** (Symmetric two-group CBM on hypergraphs with erasures). Let  $\sigma^0$  follow the same generation as in Example 1.3. Each  $K$ -tuple  $A \equiv \{a_1, \dots, a_K\} \subset \{1, \dots, n\}$  is drawn uniformly at random. We set

$$\sigma_A^0 \equiv \sigma_{a_1}^0 \sigma_{a_2}^0 \dots \sigma_{a_K}^0.$$

The observation  $G_A$  is that with probability  $q$  we observe the true product  $\sigma_A^0$ , or an erasure otherwise. The transition probability is written as

$$\mathbb{P}(G_A | \sigma_A^0) = (1 - q)\delta_{G_A, \sigma_A^0} + q\delta_{G_A, 0}.$$

$\square$

### Weak recovery

For the symmetric two-group setting, a measure of quality of the estimate  $\sigma$  is given by the absolute value of the overlap function [16, 17, 15]:

$$|Q| = \frac{1}{n} \left| \sum_{i=1}^n \sigma_i^0 \sigma_i \right|. \quad (1.2)$$

*Exact recovery* of the communities corresponds to  $|Q| = 1$ . Whereas, as any random guess yields  $|Q| \rightarrow 0$  for large  $n$ , another interest would be to obtain an estimate better than a random guess. We say *weak recovery* is possible if there is an estimator  $\mathbf{G} \rightarrow \hat{\sigma}$  such that

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\sigma^0, \mathbf{G}, \hat{\sigma}} [|Q|] > 0. \quad (1.3)$$

(Expectation in (1.3) also carries on  $\hat{\sigma}$  because the output of the estimator can be non-deterministic.) A fundamental question is about how to characterize when (1.3) is fulfilled as a formula of the model parameters. When the size of two communities is asymmetric, the overlap function can be generalized. More details are formed in Chapter 3.

## 1.2 Information and coding theory

Mutual information and channels are concepts from information and coding theory. They were originally used to study the fundamental limits in communication. In this section, we propose that these concepts are also useful for the study of community detection. As the derivative of mutual information is intimately related to the overlap, mutual information can be used to study the phase transition for weak recovery. Whereas, channels can be used to describe the generation process of the random graphs in the block models. The description in terms of channels enables us to recognize identities known in coding theory, and also gives an intuition for comparing the difficulties in deriving the mutual information of different models.

### Mutual Information

In inference, we often want to quantify the information stored in the observations  $\mathbf{Y}$  about the ground truth  $\mathbf{X}$ . A fundamental quantity is bounded by Shannon's mutual information [18]

$$I(\mathbf{X}; \mathbf{Y}) \equiv \mathbb{E}_{\mathbf{X}, \mathbf{Y}} \ln \left( \frac{\mathbb{P}(\mathbf{X}, \mathbf{Y})}{\mathbb{P}(\mathbf{X})\mathbb{P}(\mathbf{Y})} \right) \geq 0, \quad (1.4)$$

which captures the “correlation” between the two vectors of random variables. The mutual information is zero if and only if  $\mathbf{X}$  is independent of  $\mathbf{Y}$  whereby  $\mathbb{P}(\mathbf{X}, \mathbf{Y}) = \mathbb{P}(\mathbf{X})\mathbb{P}(\mathbf{Y})$ . For the problem of community detection it is useful to

study the mutual information  $\frac{1}{n}I(\boldsymbol{\sigma}^0; \mathbf{G})$ . When the observations are pairwise (i.e., elements in  $\mathbf{G}$  are  $\{G_{ij}\}$ ), differentiating  $\frac{1}{n}I(\boldsymbol{\sigma}^0; \mathbf{G})$  with respect to some model parameters would give  $\mathbb{E}_{\boldsymbol{\sigma}^0, \mathbf{G}}\langle Q^2 \rangle$  up to a constant and rescaling, where the *Gibbs bracket*

$$\langle A(\boldsymbol{\sigma}) \rangle \equiv \sum_{\boldsymbol{\sigma}} \mathbb{P}(\boldsymbol{\sigma} | \mathbf{G}) A(\boldsymbol{\sigma})$$

is the expectation of  $A(\boldsymbol{\sigma})$  with respect to the posterior  $\mathbb{P}(\boldsymbol{\sigma} | \mathbf{G})$ . The relation of  $\mathbb{E}_{\boldsymbol{\sigma}^0, \mathbf{G}}\langle Q^2 \rangle$  to the criteria of weak recovery (1.3) is that (see [17] for details):

- $\lim_{n \rightarrow \infty} \mathbb{E}_{\boldsymbol{\sigma}^0, \mathbf{G}}\langle Q^2 \rangle = 0$  if and only if  $\lim_{n \rightarrow \infty} \mathbb{E}_{\boldsymbol{\sigma}^0, \mathbf{G}}\langle |Q| \rangle = 0$ ;
- as  $Q$  is bounded, if  $\lim_{n \rightarrow \infty} \mathbb{E}_{\boldsymbol{\sigma}^0, \mathbf{G}}\langle Q^2 \rangle > 0$ , then  $\lim_{n \rightarrow \infty} \mathbb{E}_{\boldsymbol{\sigma}^0, \mathbf{G}}\langle |Q| \rangle > 0$ ;
- if there is another estimator  $\tilde{\mathbf{G}} \rightarrow \tilde{\boldsymbol{\sigma}}$  such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i^0 \tilde{\sigma}_i \right| > 0$$

but  $\lim_{n \rightarrow \infty} \mathbb{E}_{\boldsymbol{\sigma}^0, \tilde{\mathbf{G}}}\langle |Q| \rangle = 0$ , a contradiction would occur.

In summary, the possibility of weak recovery (1.3) is determined by whether

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\boldsymbol{\sigma}^0, \mathbf{G}}\langle Q^2 \rangle > 0, \quad (1.5)$$

and (1.5) can be checked if we can compute the asymptotic mutual information

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(\boldsymbol{\sigma}^0; \mathbf{G}). \quad (1.6)$$

Whereas, when observations  $\mathbf{G}$  are interactions of  $K$  variables, the derivatives of  $\frac{1}{n}I(\boldsymbol{\sigma}^0; \mathbf{G})$  would give  $\mathbb{E}_{\boldsymbol{\sigma}^0, \mathbf{G}}\langle Q^K \rangle$  (again up to a constant and rescaling). The argument about the relation to weak recovery follows similarly to the above argument. Therefore, the asymptotic mutual information (1.6) is the central object for us to study. In Sections 1.3 and 1.4 we will review statistical physics methods to compute (1.6).

## Channel models

Another notion of coding, which can be borrowed for community detection, is the notion of “channel models”. A codeword  $\mathbf{X}$  is transmitted through a noisy channel described by a transition probability  $Q(\mathbf{Y} | \mathbf{X})$ . The codeword is often expressed as a sequence of binary digits (bits). Without loss of generality, we convert  $0 \rightarrow +1$  and  $1 \rightarrow -1$  and describe the typical channel for  $X \in \{-1, +1\}$ . Typical channels are considered to be memoryless so that the transition probability can be decomposed into an elementwise product:

$$Q(\mathbf{Y} | \mathbf{X}) = \prod_{i=1}^n Q(Y_i | X_i)$$

**Example 1.5** (Binary erasure channel – BEC( $\epsilon$ )). *The channel maps input  $X \in \{-1, +1\}$  to output  $Y \in \{-1, 0, +1\}$  where 0 denotes an erasure. The transition probability is*

$$Q(Y|X) = (1 - \epsilon)\delta_{Y,X} + \epsilon\delta_{Y,0}.$$

**Example 1.6** (Binary symmetric channel – BSC( $\epsilon$ )). *The channel maps input  $X \in \{-1, +1\}$  to output  $Y \in \{-1, +1\}$  with transition probability*

$$Q(Y|X) = (1 - \epsilon)\delta_{Y,X} + \epsilon\delta_{Y,-X}.$$

The transition probabilities of the above channels satisfy

$$\mathbb{P}(Y|X) = \mathbb{P}(-Y| -X), \quad (1.7)$$

a property known as channel symmetry. We can use the concept of channels to describe the randomness in the observations in Examples 1.1 to 1.4, as shown in Fig. 1.1. The CBM in Example 1.3 has a binary erasure channel. The CBM in Example 1.4 has a binary symmetric channel. A dense SBM with appropriate choices of parameters can be a symmetric channel. Whereas, a sparse SBM always has a very asymmetric channel. The benefit of the channel descriptions is that we can recognize some mathematical identities that are available for symmetric channels (see Section 4.2) and that we can apply them when we rigorously derive the mutual information. This also suggests that deriving the mutual information for the sparse SBM is potentially a harder problem than the others.

## 1.3 Spin glass

Statistical physics aims to understand the macroscopic properties of systems constituted of many microscopic degrees of freedom. An interesting class of systems are the disordered systems where, on top of the microscopic degrees of freedom, there exists also “random disorders” (e.g., random magnetic fields). Spin glasses are theoretical models invented to capture the statistical physics of “disordered systems”. One common phenomenon that occurs is phase transitions in the thermodynamic limit. To study the phase transition, we study the fundamental quantity called free energy.

Numerous problems in engineering and computer science can also be formulated as spin-glass models, such as LDPC codes [19], CDMA systems [20], combinatorial optimization [21], and neural networks [22]. Using statistical-physics methods, we can predict phase transitions in these problems.

In the remainder of this section, we first introduce spin-glass models, various basic definitions, and the correlation inequalities related to the models. We will then review a simple example of a spin-glass model where a phase transition can be studied from the mean-field solution of the free energy. Finally, we revisit our problems on community detection and show how the block

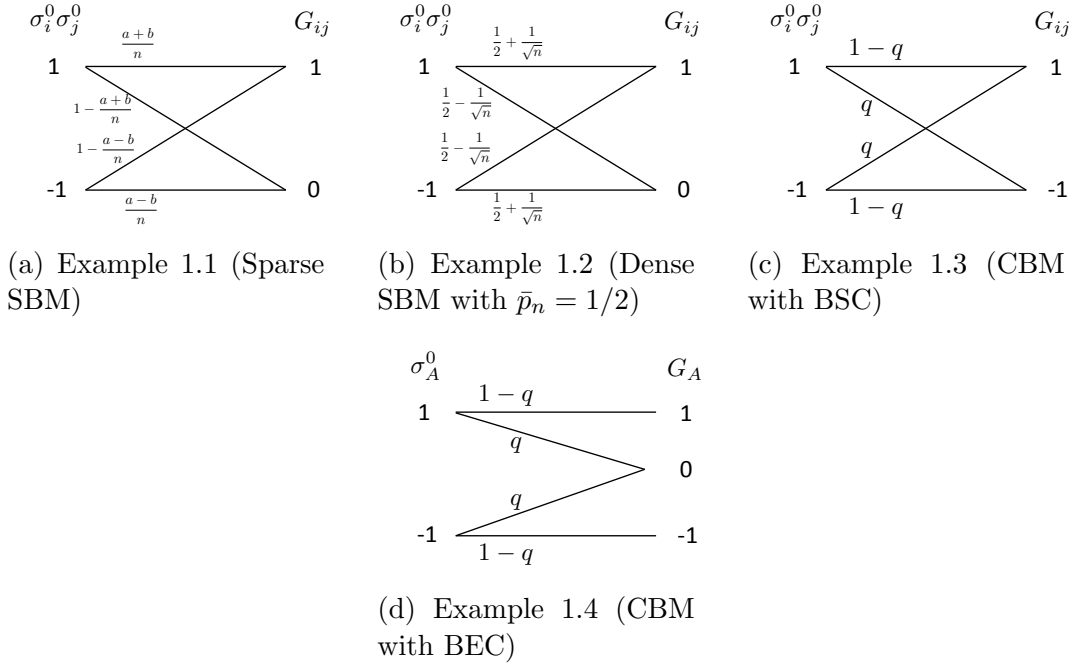


Figure 1.1 – Channels for the block model

models can be formulated as spin-glass models, as well as, show the relation between the mutual information and free energy. It is tempting to also have mean-field solutions for the free energy of the block models. We will discuss heuristic methods for deriving such solutions in Section 1.4.

## Models

A generic spin-model is a disordered system consisting of a collection of  $n$  binary spins  $\sigma_i \in \{-1, +1\}$ ,  $i = 1, \dots, n$ . For any subset  $A \subset \{1, \dots, n\}$ , we denote  $\sigma_A = \prod_{i \in A} \sigma_i$ . The model is defined by Hamiltonian

$$\mathcal{H}(\boldsymbol{\sigma}) = - \sum_{A \subset \{1, \dots, n\}} J_A \sigma_A \quad (1.8)$$

where the sum runs over all possible  $2^n$  subsets of  $\{1, \dots, n\}$ . The interaction  $J_A$  can be drawn from a probability distribution. Once we are given an instance of the Hamiltonian,  $J_A$  is fixed or frozen. Therefore, it is called a quenched random variable. The only subsets of spins that truly participate in the interactions are of course those for which  $J_A \neq 0$ . Another randomness in this model is the annealed random variables  $\boldsymbol{\sigma}$ . Any configuration  $\boldsymbol{\sigma}$  follows the Gibbs distribution

$$\mathbb{P}(\boldsymbol{\sigma} | \mathbf{J}) = \frac{1}{\mathcal{Z}_n} e^{-\beta \mathcal{H}(\boldsymbol{\sigma})},$$

where  $\beta$  is the inverse temperature and

$$\mathcal{Z}_n = \sum_{\sigma \in \{-1, +1\}^n} e^{-\beta \mathcal{H}(\sigma)}$$

is the normalization factor known as the partition function. Without loss of generality, we fix  $\beta = 1$  as this amounts to a simple global rescaling of the Hamiltonian. Macroscopic quantities of the spin model, which are often considered, include the Gibbs average of some observable  $A(\sigma)$ , conventionally denoted by

$$\langle A(\sigma) \rangle = \sum_{\sigma \in \{-1, +1\}^n} \mathbb{P}(\sigma | \mathbf{J}) A(\sigma).$$

Another fundamental object is the free energy

$$F_n \equiv -\frac{1}{n} \ln \mathcal{Z}_n, \quad f_n \equiv \mathbb{E}[F_n].$$

An interaction  $J_{ij}$  between a pair of spins is attractive when  $J_{ij} > 0$ , or repulsive when  $J_{ij} < 0$ . Extending this property, a spin model is called ferromagnetic when all  $J_A$  in (1.8) are non-negative. In such models, Griffiths-Kelly-Sherman (GKS) inequalities [23, 24, 25] are well-known correlation inequalities that state that for any subsets of variable indices  $S, T \subset \{1 \dots n\}$

$$\langle \sigma_S \rangle \geq 0, \tag{1.9}$$

$$\langle \sigma_S \sigma_T \rangle - \langle \sigma_S \rangle \langle \sigma_T \rangle \geq 0. \tag{1.10}$$

These inequalities will be used in Chapter 4 and Chapter 5.

### Phase transition in the Curie-Weiss model

The Curie-Weiss model is a model based on a complete graph where interactions are between all pairs of spins, with Hamiltonian

$$\mathcal{H}_{\text{CW}}(\sigma) \equiv -\frac{J}{n} \sum_{i < j} \sigma_i \sigma_j - h \sum_{i=1}^n \sigma_i.$$

This Hamiltonian can also be represented as a function of the magnetization (similar to the overlap function  $Q$  in (1.2))

$$m \equiv \frac{1}{n} \sum_{i=1}^n \sigma_i.$$

Following the textbook calculation such as [26], it is straightforward to see that the free energy admits a single-letter variational expression:

$$\lim_{n \rightarrow \infty} F_n = \min_{m \in [-1, +1]} f_{\text{CW}}(m), \tag{1.11}$$

where

$$f_{\text{CW}}(m) \equiv \frac{J}{2}(1 - m^2) - hm + \frac{1+m}{2} \ln\left(\frac{1+m}{2}\right) + \frac{1-m}{2} \ln\left(\frac{1-m}{2}\right).$$

The equality (1.11) is particularly useful for observing a phase transition for the magnetization. We can check that  $m^* \equiv \arg \min_{m \in [-1, +1]} f_{\text{CW}}(m)$  is attained at a stationary point satisfying

$$m = \tanh(Jm + h). \quad (1.12)$$

To evaluate the Gibbs average of the magnetization  $\langle m \rangle$ , we note that  $\langle m \rangle = -dF_n/dh$ . Differentiating both sides of (1.11) gives us

$$\langle m \rangle = m^*.$$

Let us consider the simple case  $h = 0$ . From (1.12), we can observe a phase transition for  $\langle m \rangle$ . We can deduce  $\langle m \rangle = 0$  when  $J < 1$ , and  $\langle m \rangle \neq 0$  when  $J > 1$ . This could be analogous to community detection by imagining  $m$  as the overlap and  $J$  as the signal-to-noise ratio (SNR) of observations. A non-trivial overlap could be suddenly developed when the SNR passes a threshold.

The Curie-Weiss model is simplistic in physics, but it offers a great lesson. Eq. (1.11) implies that the macroscopic behavior of a complex model with many spins and interactions can be studied through a simpler variational expression that can be explicitly computed. Intuitively, this implies that the complex model can be approximated by a decoupled model so that the difficulty in evaluating all the interactions is bypassed. Variational expressions such as (1.11) are known as mean-field solutions. Unsurprisingly, a mean-field solution is not limited to the Curie-Weiss model. Models that admit “variational solutions” are called mean-field models. In Section 1.4, we introduce the two powerful techniques in statistical physics to derive the mean-field solution.

### From mutual information to free energy

In the block model, with the assumption that  $\sigma_i^0$  are i.i.d. variables drawn from the same prior  $\mathbb{P}_0$ , the posterior distribution can be factorized as

$$\mathbb{P}(\boldsymbol{\sigma}|\mathbf{G}) = \frac{\mathbb{P}(\mathbf{G}|\boldsymbol{\sigma})\mathbb{P}_0(\boldsymbol{\sigma})}{\mathbb{P}(\mathbf{G})} = \frac{1}{\mathcal{Z}_n} \prod_{A \in \{1, \dots, n\}} \mathbb{P}(G_A | \{\sigma_i : i \in A\}) \prod_{i=1}^n \mathbb{P}_0(\sigma_i) \quad (1.13)$$

with

$$\mathcal{Z}_n = \sum_{\boldsymbol{\sigma}} \prod_{A \in \{1, \dots, n\}} \mathbb{P}(G_A | \{\sigma_i : i \in A\}) \prod_{i=1}^n \mathbb{P}_0(\sigma_i).$$

We can again define the free energy here as

$$f_n \equiv -\frac{1}{n} \mathbb{E} \ln \mathcal{Z}_n.$$



The mutual information is linked to the free energy by the formula,

$$\frac{1}{n}I(\boldsymbol{\sigma}^0; \mathbf{G}) = \frac{1}{n}\mathbb{E} \ln \left( \frac{\mathbb{P}(\boldsymbol{\sigma}|\mathbf{G})}{\mathbb{P}(\boldsymbol{\sigma})} \right) = f_n + \frac{1}{n}\mathbb{E} \left[ \sum_{A \in \{1, \dots, n\}} \ln \mathbb{P}(G_A | \{\sigma_i : i \in A\}) \right].$$

The second term can be reduced to a simple expression once an explicit model is chosen. Therefore, computing the mutual information is essentially reduced to the problem of computing the free energy.

## 1.4 The cavity and replica methods

The replica method and the cavity method are two heuristic methods for computing the free energy [27, 28]. The two methods involve mathematically unjustified steps or need to assume some structure on the model, and eventually enable us to derive a variational expression called replica symmetric formula. It is predicted that this formula is exactly equal to the limit of free energy.

### The replica method

The replica method is based on the formula

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\ln Z_n] = \lim_{n \rightarrow \infty} \lim_{N \rightarrow 0} \frac{1}{n} \frac{\mathbb{E}[Z_n^N] - 1}{N} \quad (1.14)$$

or the equivalent formula

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[\ln Z_n] = \lim_{n \rightarrow \infty} \lim_{N \rightarrow 0} \frac{1}{n} \frac{d}{dN} \mathbb{E}[Z_n^N].$$

To yield the replica symmetric formula from the L.H.S. of these formulae, we assume the validity of swapping the two limits. Then the moments  $\mathbb{E}[Z_n^N]$  are computed as if  $N$  would be an integer, despite that the limit  $N \rightarrow 0$  is taken in the final step.

### Factor graphs

The replica method is compact but elusive in meaning. The cavity method is another method, in principle equivalent to the replica method, but it has a clearer interpretation. It is best described in the language of factor graph.

A factor graph  $\mathcal{G} = (\mathcal{V}, \mathcal{F}, \mathcal{E})$  for  $n$  variables  $\mathbf{x} = (x_1, \dots, x_n)$  is a bipartite graph containing a set of variable nodes  $\mathcal{V}$ , a set of factor nodes  $\mathcal{F}$ , and a set of edges  $\mathcal{E}$  connecting variable nodes to factor nodes. We associate each variable node  $i \in \mathcal{V}$  with a variable  $x_i \in \mathcal{X}_i$ , and associate each factor node  $a \in \mathcal{F}$  with a function  $\psi_a$ . We denote  $\partial i \equiv \{i \in \mathcal{V} : (i, a) \in \mathcal{E}\}$  the neighbors of node  $i$  and

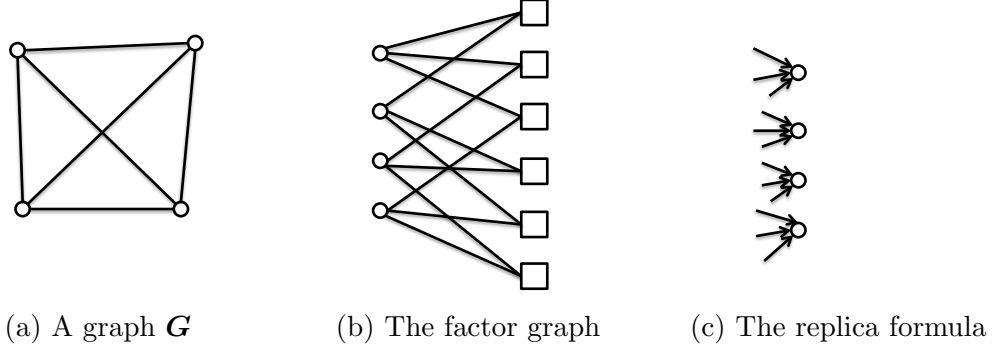


Figure 1.2 – A graphical illustration from a graph  $\mathbf{G}$  to the replica formula

$\partial a \equiv \{a \in \mathcal{F} : (i, a) \in \mathcal{E}\}$  the neighbors of node  $a$ . The argument of function  $\psi_a$  is  $\mathbf{x}_{\partial a} \equiv \{x_i, i \in \partial a\}$ . The factor graph represents the probability

$$p(\mathbf{x}) = \frac{1}{\mathcal{Z}_n} \prod_{a \in \mathcal{F}} \psi_a(\mathbf{x}_{\partial a}), \quad (1.15)$$

where  $\mathcal{Z}_n \equiv \sum_{\mathbf{x}} \prod_{a \in \mathcal{F}} \psi_a(\mathbf{x}_{\partial a})$ .

An example is in Fig. 1.2a and Fig. 1.2b. Let's assume we want to infer  $\boldsymbol{\sigma}$  from the graph  $\mathbf{G}$  in (a). Each edge represents some weight  $G_{ij}$ . The posterior  $\mathbb{P}(\boldsymbol{\sigma}|\mathbf{G})$  follows the expression in (1.13) where  $A = (i, j)$ . The associated factor graph is constructed in Fig. 1.2b with factor nodes represented in boxes. Each factor node represents a factor  $\mathbb{P}(G_{ij}|\sigma_i, \sigma_j)$ .

### The cavity method

The cavity method starts with a variational approach for approximating  $F_n \equiv n^{-1} \ln \mathcal{Z}_n$ . For the probability  $p(\mathbf{x})$  in (1.15) that admits a factor graph  $\mathcal{G}$ , and for any trial probability  $b(\mathbf{x})$  on  $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ , we define the Gibbs free energy  $F_{\text{Gibbs}}(b)$  to be

$$F_{\text{Gibbs}}(b) = -\frac{1}{n} \sum_{\mathbf{x}} b(\mathbf{x}) \ln \left( \prod_{a \in \mathcal{F}} \psi_a(\mathbf{x}_{\partial a}) \right) + \frac{1}{n} \sum_{\mathbf{x}} b(\mathbf{x}) \ln b(\mathbf{x}). \quad (1.16)$$

This Gibbs free energy satisfies

$$F_{\text{Gibbs}}(b) = F_{\text{Gibbs}}(p) + D(b||p),$$

where  $D(b||p) \equiv \sum_{\mathbf{x}} b(\mathbf{x}) \ln (b(\mathbf{x})/p(\mathbf{x}))$  is the Kullback-Leibler divergence. We can think  $b$  as an estimate of  $p$ . Since  $D(b||p) \geq 0$  and the equality is attained if and only if  $p(\mathbf{x}) = b(\mathbf{x})$  for all  $\mathbf{x} \in \mathcal{X}$ , computing the free energy  $F_n \equiv F_{\text{Gibbs}}(p)$  can be formulated as the Gibbs variational problem:

$$\inf_b F_{\text{Gibbs}}(b) \quad (1.17)$$

$$\text{st. } 0 \leq b(\mathbf{x}) \leq 1, \quad \mathbf{x} \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n, \quad (1.18)$$

$$\sum_{\mathbf{x}} b(\mathbf{x}) = 1. \quad (1.19)$$

Finding the optimal distribution for (1.17) is not easier than directly computing  $F_{\text{Gibbs}}(p)$ . We can attempt the special case where the underlying factor graph does not have a cycle. In this case,  $p(\mathbf{x})$  can be factorized as (see [26])

$$p(\mathbf{x}) = \prod_{a \in \mathcal{F}} p_a(\mathbf{x}_{\partial a}) \prod_{i \in \mathcal{V}} p_i(x_i)^{1-|\partial i|}, \quad (1.20)$$

where  $p_i(x_i)$  (resp.  $p_a(\mathbf{x}_{\partial a})$ ) is the marginal probability at the variable node  $i$  (resp. at the factor node  $a$ ). Eq. (1.20) implies that in the Gibbs variational problem it suffices to assume

$$b(\mathbf{x}) = \prod_{a \in \mathcal{F}} b_a(\mathbf{x}_{\partial a}) \prod_{i \in \mathcal{V}} b_i(x_i)^{1-|\partial i|}, \quad (1.21)$$

where  $\{b_i, b_a : i \in \mathcal{V}, a \in \mathcal{F}\}$  are estimates of  $\{p_i, p_a : i \in \mathcal{V}, a \in \mathcal{F}\}$ . The resulting formula after substituting (1.21) into (1.16) is called the Bethe free energy, given by

$$\begin{aligned} F_{\text{Bethe}}(\{b_i, b_a\}) &= -\frac{1}{n} \sum_{a \in \mathcal{F}} \sum_{\mathbf{x}_a} b_a(\mathbf{x}_{\partial a}) \ln(\psi_a(\mathbf{x}_{\partial a})) + \frac{1}{n} \sum_{a \in \mathcal{F}} \sum_{\mathbf{x}_{\partial a}} b_a(\mathbf{x}_{\partial a}) \ln b_a(\mathbf{x}_{\partial a}) \\ &\quad + \frac{1}{n} \sum_{i \in \mathcal{V}} (1 - |\partial i|) \sum_{x_i} b_i(x_i) \ln b_i(x_i). \end{aligned}$$

Also, the Gibbs variational problem (1.17)–(1.19) leads to

$$\inf_{\{b_i, b_a\}} F_{\text{Bethe}}(\{b_i, b_a\}) \quad (1.22)$$

$$\text{st. } \sum_{x_i} b_i(x_i) = 1, \quad \forall i \in \mathcal{V}, \quad (1.23)$$

$$\sum_{\mathbf{x}_{\partial a}} b_a(\mathbf{x}_{\partial a}) = 1, \quad \forall a \in \mathcal{F}, \quad (1.24)$$

$$\sum_{\mathbf{x}_{\partial a} \setminus x_i} b_a(\mathbf{x}_{\partial a}) = b_i(x_i), \quad \forall (i, a) \in \mathcal{E}, x_i \in \mathcal{X}_i, \quad (1.25)$$

$$0 \leq b_i(x_i) \leq 1, \quad \forall i \in \mathcal{V}, x_i \in \mathcal{X}_i, \quad (1.26)$$

$$0 \leq b_a(\mathbf{x}_{\partial a}) \leq 1, \quad \forall a \in \mathcal{F}, \mathbf{x}_{\partial a} \in \prod_{i \in \partial a} \mathcal{X}_i. \quad (1.27)$$

For general factor graphs, possibly with cycles, the infimum obtained from (1.22)–(1.27) can be used as an approximation of  $F_n$ . This is known as the Bethe approximation.

Another way to approximate  $F_n$  is belief propagation. It is an iterative message-passing algorithm that involves two kinds of messages: message  $m_{i \rightarrow a}(\mathbf{x}_a)$  from variable node  $i$  to factor node  $a$ , and message  $\tilde{m}_{a \rightarrow i}(x_i)$  from factor node  $a$  to variable node  $i$ . All messages are initialized to 1 and are

updated according to the update rules:

$$\begin{aligned} m_{i \rightarrow a}(\mathbf{x}_{\partial a}) &\propto \prod_{c \in \partial i \setminus a} \tilde{m}_{c \rightarrow i}(x_i) \\ \tilde{m}_{a \rightarrow i}(x_i) &\propto \sum_{\mathbf{x}_{\partial a} \setminus x_i} \psi_a(\mathbf{x}_{\partial a}) \prod_{j \in \partial a \setminus i} m_{j \rightarrow a}(x_j). \end{aligned} \quad (1.28)$$

The updated messages are always normalized. When the messages converge,  $F_n$  can be approximated by

$$\begin{aligned} F_{\text{BP}}(\{m, \tilde{m}\}) &\equiv -\frac{1}{n} \sum_{i \in V} Z_i(\{\tilde{m}_{a \rightarrow i}\}_{a \in \partial i}) - \frac{1}{n} \sum_{a \in \mathcal{F}} \ln Z_a(\{m_{i \rightarrow a}\}_{i \in \partial a}) \\ &\quad + \frac{1}{n} \sum_{i \in \mathcal{V}} \sum_{a \in \partial i} \ln Z_{i,a}(\tilde{m}_{a \rightarrow i}, m_{i \rightarrow a}), \end{aligned}$$

where

$$\begin{aligned} Z_i(\{\tilde{m}_{a \rightarrow i}\}_{a \in \partial i}) &\equiv \sum_{x_i} \prod_{a \in \partial i} \tilde{m}_{a \rightarrow i}(x_i), \\ Z_a(\{m_{i \rightarrow a}\}_{i \in \partial a}) &\equiv \sum_{\mathbf{x}_a} \psi_a(\mathbf{x}_a) \prod_{i \in \partial a} m_{i \rightarrow a}(x_i), \\ Z_{i,a}(\tilde{m}_{a \rightarrow i}, m_{i \rightarrow a}) &\equiv \sum_{x_i} \tilde{m}_{a \rightarrow i}(x_i) m_{i \rightarrow a}(x_i). \end{aligned}$$

The authors of [29] show that any stationary point  $\{b_i, b_a\}$  of the Bethe free energy  $F_{\text{Bethe}}(\{b_i, b_a\})$  has a one-to-one mapping with a fixed point  $\{m, \tilde{m}\}$  of belief propagation. This implies that for the pair of stationary point and fixed point

$$F_{\text{Bethe}}(\{b_i, b_a\}) = F_{\text{BP}}(\{m, \tilde{m}\}).$$

One of the fixed-point equations is setting (1.28) to be an equality. From this fixed-point equation, we can express  $\tilde{m}$  as a function of  $m$  and simplify  $F_{\text{BP}}(\{m, \tilde{m}\})$  to  $F_{\text{BP}}(\{m\})$ .

The cavity method then assumes that for every  $(i, a) \in \mathcal{E}$ ,  $\{m_{i \rightarrow a}(\mathbf{x}_{\partial a}) : \mathbf{x}_{\partial a} \in \prod_{i \in \partial a} \mathcal{X}_i\}$  is an i.i.d. random vector that is normalized to 1, and the vector is drawn from a trial distribution  $\mathbf{m}$ . The *replica symmetric formula* is defined to be

$$f_{\text{RS}}(\mathbf{m}) \equiv \mathbb{E}_{\mathcal{G}, \mathbf{m}}[F_{\text{BP}}(\{m\})].$$

The conjecture is that <sup>1</sup>

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{G}}[F_n] = \inf_{\mathbf{m} \in \mathfrak{M}} f_{\text{RS}}(\mathbf{m}). \quad (1.29)$$

<sup>1</sup>This is conjectured at least for inference problems. In general, the choice of the extremum over  $\{m, \tilde{m}\}$  of  $F_{\text{BP}}(\{m, \tilde{m}\})$  might be different.

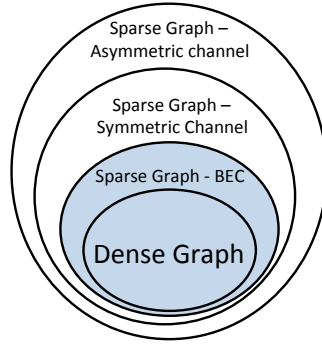


Figure 1.3 – Intuitive difficulty level to prove (1.29): the outermost circle is the most difficult

The support of the infimum is a certain set of distribution  $\mathfrak{M}$ ; this set can be tightened depending on the known properties of a given model. Fig. 1.2c pictures the associated graph with the replica symmetric formula. The factor nodes are replaced by i.i.d. messages fed into the variable nodes. This corresponds to a new decoupled problem where the variables from the messages are inferred. The difficulty of proving (1.29) has a hierarchy (as depicted in Fig. 1.3) based on the channels and the graph regime. Statements on channel universality [30] or applications of Lindeberg's Theorem [31] can often be used to show that the dense graph model can be mapped to another model with Gaussian channels. Hence, we do not differentiate the channel models for dense graphs. For models on dense factor graphs or models with binary erasure channels,  $m$  can be transformed to a trial scalar parameter  $m$  in a bounded interval, and this also simplifies the replica symmetric formula. Therefore, we have the two innermost circles in Fig. 1.3. Also, if we recall channels in Sec. 1.2, we understand that symmetric channels are more structured and more tools are available to analyze symmetric channels. Therefore, we obtain the order of the two other circles in Fig. 1.3. This view implies that the difficulty in analyzing community detection problem is in the order: Example 1.1 (Sparse SBM)  $\geq$  Example 1.3 (Sparse CBM with BSC)  $\geq$  Example 1.4 (Sparse CBM with BEC)  $\geq$  Example 1.2 (Dense SBM).

## 1.5 Implication of the replica symmetric formula

Given a community detection problem, we can identify the free energy associated with the mutual information. We can then derive the replica symmetric formula  $f_{\text{RS}}$  in order to predict the asymptotic free energy or the asymptotic mutual information. In Fig. 1.4, we illustrate  $f_{\text{RS}}$  as a function of the trial parameter  $m$  and the signal-to-noise ratio (SNR) in the community detection problem. The trial parameter  $m$  can be interpreted as the SNR of some decoupled observations. We always have the trivial stationary point  $m = 0$ ; it implies that the decoupled observations are uninformative.

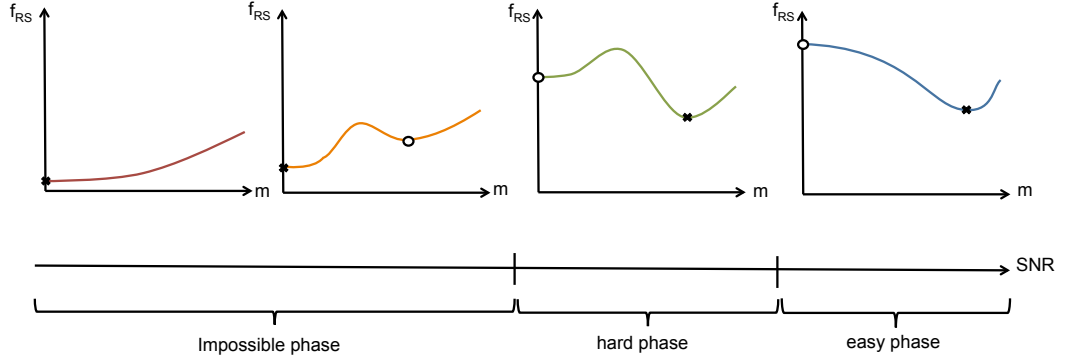


Figure 1.4 – The replica symmetric formula  $f_{\text{RS}}$  as a function of the trial parameter  $m$  and the SNR of the model. A black point denotes a global minima, and a white circle denotes a local minimum.

As the derivation of the replica symmetric formula  $f_{\text{RS}}$  is intimately related to message-passing algorithms (which is belief propagation when the graph is sparse or, is the approximated version of belief propagation when the graph is dense), only the replica symmetric formula provides details about the performance of the message-passing algorithm. It is known that the stationary point of the replica symmetric formula corresponds to the fixed-point solution of the density-evolution equation [32] or the state-evolution equation [33] that tracks the performance of the message-passing algorithm. As the message-passing algorithm iterates, we can understand its performance as a kind of “gradient descent” starting from the uninformative point  $m = 0$  and eventually converging to a local minimum. Message-passing algorithms therefore achieve weak recovery if the “gradient descent” can reach a local minimum not equal to zero.

If we recall the discussion in Section 1.2, we understand that the derivative of the mutual information with respect to the SNR is linked to the overlap function. Now, assuming the asymptotic mutual information is given by evaluating the replica symmetric formula at the global minimum, we can inspect the overlap function from the derivative of  $\min_m f_{\text{RS}}(m)$ . The replica symmetric formula predicts that a phase transition (say, as a function of SNR) for weak recovery would occur at the non-analytical point of  $\min_m f_{\text{RS}}(m)$ , at which the global minima of the replica symmetric formula jumps (say, as a function of SNR).

Therefore, the replica symmetric formula enables us to identify three phases, as illustrated in Fig. 1.4. At the “easy phase” (the blue curve), any negligible amount of side information would initiate the “gradient descent”, and the performance of message-passing algorithms is given by the global minimum, which is strictly positive. Message-passing algorithms can be used to achieve weak recovery. As the SNR decreases, we will enter the “hard phase” (the green curve). The “gradient descent” cannot overcome the barrier and is stuck at the uninformative point  $m = 0$ . Message-passing algorithms cannot achieve weak recovery. However, weak recovery is possible because the global minimum is

attained at  $m > 0$ . As the SNR further decreases, we can expect the “impossible phase” (the yellow and red curves). The SNR is low and weak recovery is impossible. This is illustrated by the curves where the global minimum is attained at  $m = 0$ .

## 1.6 Related works on proving the replica prediction

The most notable replica symmetric formula in statistical physics is the one for the Sherrington-Kirkpatrick model. Although it was proposed by Parisi [34] in 1980, it was rigorously proved relatively recently by Talagrand in [35]. The underlying techniques of the proof — the Guerra-Toninelli interpolation method [36] and the Aizenman-Sims-Starr scheme [37] — remains beneficial for analyzing many other problems. During the last two decades, the utility of replica symmetric formula has attracted the attention in high-dimensional Bayesian inference, and much progress has been achieved. Examples where *full* proofs have been achieved are random linear estimation and compressed sensing [38, 39], learning for single-layer networks [40], generalized estimation in multi-layer settings [41, 42], and low-rank matrix and tensor estimation [43, 17, 44, 45]. The Guerra-Toninelli interpolation method is used to obtain a one-sided bound in most of the existing proofs. The converse bound is arguably more difficult. The Aizenman-Sims-Starr scheme [17, 44, 46], or other state-of-the-art techniques such as spatial coupling [43, 38], are used to address this converse bound. More recently, the authors in [47, 48] have discovered a unified proof via an elaborated version of the Guerra-Toninelli interpolation method, called *adaptive interpolation*. The method is generic to problems with a dense underlying factor graph. The new method is quite generic, once the correct decoupled problem has been identified and is directly applicable when the concentration of the overlap can be proved. The successes of the adaptive interpolation method have so far been limited to inference models with a dense underlying factor graph, and it has not been applied to any block models before this thesis. It is desirable to see to what extent the method can be developed for these open cases.

Let us very briefly summarize the interpolation methods. Given the free energy  $f_n$  and the replica symmetric formula  $f_{\text{RS}}$ , we can identify the Hamiltonian  $\mathcal{H}_{\text{base}}$  associated with  $f_n$  and another Hamiltonian  $\mathcal{H}_{\text{decoupled}}$  associated with a decoupled model. The Guerra-Toninelli interpolation method defines an interpolated Hamiltonian  $\mathcal{H}_t$  with time  $t \in [0, 1]$  such that  $\mathcal{H}_0 = \mathcal{H}_{\text{base}}$  and  $\mathcal{H}_1 = \mathcal{H}_{\text{decoupled}}$ . Let  $f_t$  be the interpolated free energy associated with  $\mathcal{H}_t$ . Interpolation is based on using a formula due to the fundamental theorem of calculus:

$$f_{t=0} = f_{t=1} - \int_0^1 dt \frac{df_t}{dt}.$$

It turns out that  $f_{t=1}$  is closely related to  $f_{\text{RS}}$  (although not equal) and thus lead us to the *sum rule*

$$f_n = f_{\text{RS}} + \text{remainder}.$$

The remainder often is either positive or negative. Thus, a one-sided bound can be obtained immediately.

The Guerra-Toninelli method is based on a *fixed* interpolation path. The idea of *adaptive interpolation* is to adopt a larger class of paths for the interpolated Hamiltonian so that we have more flexibility to control the remainder in the sum rule. In particular, the remainder can often be expressed as a function of overlap. The other bound is obtained by suitably controlling the overlap and “adapting” the interpolation path according to a differential equation. More details for a simple example are found in Chapter 2.

## 1.7 Contributions and organization of this thesis

In this thesis, we develop new methods to rigorously derive the replica symmetric formula for community detection on both dense graphs and sparse graphs. The methods are based on extending the adaptive interpolation method in [47, 48].

In Chapter 2, we review the adaptive interpolation for proving the replica symmetric formula. The material is extracted from [47, 48]. We expose to the readers what was known about the adaptive interpolation before this thesis, in order to make a good link to the contribution of this thesis. To this end, we outline the essential steps in the method without going into all the calculations.

Chapters 3 through 5 constitute the original contributions of this thesis. Let us give a brief summary here.

In **Chapter 3**, we develop the adaptive interpolation for community detection on dense graphs. We rigorously derive the replica symmetric formula for the mutual information of the asymmetric two-group SBM in the dense-graph regime (see **Theorems 3.1 and 3.2**). A brief description of this model is in Example 1.2 in Section 1.1. The replica symmetric formula for this model is not new. The formula was derived in [49] by using heuristic methods in statistical physics. Rigorous proofs are given in [16, 17] but with two weaknesses: (1) the proofs are indirect, as they involve first mapping the problem to spiked Wigner models, and (2) the proofs have not fully covered a certain regime of “fully dense” graphs. Our contribution is a *direct* proof that include *all* regimes of dense graphs.

For analyzing for the symmetric case, where the two groups are of equal size, we can rely on the fact that the information-theoretic phase transition is continuous and of the second-order type. This enables a proof [16] using a message-passing argument. The asymmetric case is more challenging, as it can involve a first-order (discontinuous) phase transition. In this case, the authors



of [17] tackle the problem by combining two methods (Guerra-Toninelli interpolation and the Aizenman-Sims-Starr scheme). Whereas, our proof addresses the asymmetric case by using only a single method, namely, the adaptive interpolation. Thus, our proof is also conceptually simpler.

In **Chapter 4**, we develop the adaptive interpolation for community detection on sparse graphs. We rigorously derive the replica symmetric formula for the mutual information of a symmetric two-group CBM on the binary erasure channel in the sparse-graph regime (**Theorem 4.1**). A brief description of this model is in Example 1.4 in Section 1.1. Our proof demonstrates the first example of using the adaptive interpolation for a model on a sparse graph.

The replica symmetric formula for the mutual information in a sparse model is known to be more complicated. We need to take both the randomness in the observations and in the sparse graph into account; in addition, the replica symmetric formula is a functional over a set of probability distributions (instead of scalars as in the dense-graph case). Existing rigorous derivations of the formulas for sparse models require a combination of methods. To this end, the adaptive interpolation we have developed in Chapter 4 serves as a first step towards an analysis of more complicated models via a unified approach.

When we apply the adaptive interpolation to the sparse CBM, the sum rule contains a set of “multi-overlaps”  $\{Q_p : p \geq 1\}$ , where

$$Q_p \equiv \frac{1}{n} \sum_{i=1}^n \sigma_i^{(1)} \cdots \sigma_i^{(p)}$$

is the overlap of  $p$  independent replicas  $\boldsymbol{\sigma}^{(1)}, \dots, \boldsymbol{\sigma}^{(p)}$ . (This is in contrast to the dense-graph case where only the first or second overlaps matter.) The adaptive interpolation requires controlling the total fluctuation of all multi-overlaps in the form

$$\mathbb{E} \langle (Q_p - \mathbb{E} \langle Q_p \rangle)^2 \rangle. \quad (1.30)$$

In the simple situation where the measurement channel is the binary erasure channel, we can prove the concentration of all multi-overlaps, and therefore fully develop the adaptive interpolation method. If we take for granted the concentration of multi-overlaps, the adaptive interpolation method developed in Chapter 4 can be directly extended to a larger class of channels. Even though we were unable to resolve the proof of concentration in general cases, in Chapter 5 we did find a solution for similar models in the context of physics.

In **Chapter 5**, we depart from community detection and study ferromagnetic “mean-field” spin models on sparse random graphs. We prove the concentration of the total fluctuation of all multi-overlaps (**Theorem 5.3**). To the best of our knowledge, this concentration result is unprecedented for multi-overlaps in any spin model. This concentration result is also significant in the context of community detection, or Bayesian inference in general, because it sheds light on the potential of extending the adaptive interpolation for sparse graphs (developed in Chapter 4).

The concentration result involves recognizing the total fluctuation in (1.30) as the addition of two types:

$$\mathbb{E}\langle(Q_p - \langle Q_p \rangle)^2\rangle \quad \text{and} \quad \mathbb{E}[(\langle Q_p \rangle - \mathbb{E}\langle Q_p \rangle)^2].$$

The main novelties are involved in showing the concentration of the latter type. The approach in our proof suggests that, for general “mean-field” models, the concentration of multi-overlaps could be obtained from the concentration of the first overlap.

In Chapter 6, we provide our closing remarks and discuss some open challenges in extending the adaptive interpolation method to other models.

### 1.7.1 Bibliographic notes

A summary of Chapter 3 was presented in [50] and the extended version is found in [51]. A summary of Chapter 4 was presented in [52] and the extended version is found in [53]. The research yielding Chapter 5 is found in [54].

---

# Preliminary: Adaptive Interpolation for Spiked Wigner Models

---

# 2

This chapter gives a quick review of adaptive interpolation for the spiked Wigner model. The reader can refer to [55, 56] for the detailed calculation or follow the subsequent chapters in this thesis for a complete proof on related models.

The spiked Wigner model is the first model that has a full proof by the adaptive interpolation. This model consists of i.i.d. random variable  $X_i, i = 1, \dots, n$ . Each  $X_i$  is drawn from a prior distribution  $\mathbb{P}_0$  with support on a bounded interval  $[-S, S]$ . We observe the matrix  $\mathbf{Y} = (Y_{ij})_{i,j=1}^n$  where each matrix element is generated by the process:

$$Y_{ij} = \sqrt{\frac{\lambda}{n}} X_i X_j + Z_{ij}, \quad 1 \leq i \leq j \leq n, \quad (2.1)$$

where  $\lambda > 0$ ,  $Z_{ij} \sim \mathcal{N}(0, 1)$  are i.i.d. Gaussian random variables for  $i \leq j$  and symmetric  $Z_{ij} = Z_{ji}$ . We can always rescale  $\lambda$  so that we can assume  $\mathbb{E}[X_1^2] = 1$ . The total signal-to-noise ratio per parameter is  $\text{\#observations} \cdot \text{SNR}_{\text{obs}} / \text{\#parameters}$  to infer, where  $\text{SNR}_{\text{obs}}$  is the SNR per observation. The  $\text{SNR}_{\text{obs}}$  for the diagonal element  $Y_{ii}$  is  $\lambda \mathbb{E}[X_1^4]/n = \mathcal{O}(1/n)$ , and the  $\text{SNR}_{\text{obs}}$  for the off-diagonal element  $Y_{ij}$  is  $(\lambda/n) \mathbb{E}[X_1^2]^2 = \lambda/n$ . Therefore, the total signal-to-noise ratio per parameter

$$\frac{(n(n-1)/2) \cdot (\lambda/n) + n(\lambda \mathbb{E}[X_1^4]/n)}{n} = \frac{\lambda}{2} + \mathcal{O}\left(\frac{1}{n}\right) = \Theta(1)$$

defines a non-trivial inference problem.

The transition probability in this model is

$$\mathbb{P}(\mathbf{Y}|\mathbf{X}) = \frac{1}{(2\pi)^{n(n+1)/2}} \exp\left(-\frac{1}{2} \sum_{i \leq j} (Y_{ij} - \sqrt{\frac{\lambda}{n}} X_i X_j)^2\right).$$

Using Bayes' rule, we obtain the posterior distribution

$$\mathbb{P}(\mathbf{x}|\mathbf{Y}) = \frac{\mathbb{P}(\mathbf{Y}|\mathbf{x})\mathbb{P}(\mathbf{x})}{\mathbb{P}(\mathbf{Y})} \propto \exp\left\{-\sum_{i \leq j} \left(\frac{\lambda}{2n}x_i^2x_j^2 - 2\sqrt{\frac{\lambda}{n}}Y_{ij}\right)\right\} \prod_{i=1}^n \mathbb{P}_0(x_i).$$

For the convenience of using the tools from statistical physics, we write the posterior distribution following the conventions of Gibbs distributions:

$$\mathbb{P}(\mathbf{x}|\mathbf{Y}) = \frac{1}{\mathcal{Z}(\mathbf{Y})} e^{-\mathcal{H}_{\text{SW}}(\mathbf{x}, \mathbf{Y}(\mathbf{X}, \mathbf{Z}))} \prod_{i=1}^n \mathbb{P}_0(x_i),$$

where the Hamiltonian  $\mathcal{H}_{\text{SW}}$  and the partition function  $\mathcal{Z}(\mathbf{Y})$  are defined to be

$$\begin{aligned} \mathcal{H}_{\text{SW}}(\mathbf{x}, \mathbf{Y}(\mathbf{X}, \mathbf{Z})) &\equiv -\sum_{i \leq j} \left( \frac{\lambda x_i x_j X_i X_j}{n} + \frac{\sqrt{\lambda}}{n} x_i x_j Z_{ij} - \frac{x_i^2 x_j^2}{2n} \right), \\ \mathcal{Z}(\mathbf{Y}) &\equiv \int d\mathbf{x} e^{-\mathcal{H}_{\text{SW}}(\mathbf{x}, \mathbf{Y}(\mathbf{X}, \mathbf{Z}))} \prod_{i=1}^n \mathbb{P}_0(x_i). \end{aligned}$$

Furthermore, we define the free energy of this model  $f_{\text{SW}}$  to be

$$f_{\text{SW}} \equiv -\frac{1}{n} \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{Y}|\mathbf{X}} \ln \mathcal{Z}(\mathbf{Y}).$$

A straightforward calculation shows that

$$\frac{1}{n} I(\mathbf{X}; \mathbf{Y}) = \frac{1}{n} \mathbb{E} \ln \left( \frac{\mathbb{P}(\mathbf{X}|\mathbf{Y})}{\mathbb{P}(\mathbf{X})} \right) = f_{\text{SW}} + \frac{\lambda}{4} + \mathcal{O}(1/n).$$

When we ask if  $\lim_{n \rightarrow \infty} \frac{1}{n} I(\mathbf{X}; \mathbf{Y})$  exists, and what the value is if it exists, the non-trivial part to answer these questions is about how to evaluate  $f_{\text{SW}}$ . The heuristic from the replica or cavity method predicts that

$$\lim_{n \rightarrow \infty} f_{\text{SW}} = \min_{q \in [0, \lambda]} f_{\text{RS}}(q), \quad (2.2)$$

where the so-called replica-symmetric formula  $f_{\text{RS}}$  is a single-letter variational expression given by

$$f_{\text{RS}}(q) \equiv \frac{q^2}{4\lambda} - \mathbb{E} \ln \int dx \mathbb{P}_0(x) e^{-(\frac{q}{2}x^2 - qxX - \sqrt{q}xZ)}. \quad (2.3)$$

Recalling the derivation of the cavity method in Sec. 1.4, the replica-symmetric formula is always associated with an inference problem with a decoupled factor graph. Here,  $f_{\text{RS}}(q)$  is associated with inferring  $X \sim \mathbb{P}_0$  from the observation

$$Y = \sqrt{q}X + Z, \quad (2.4)$$

where  $Z \sim \mathcal{N}(0, 1)$  is a Gaussian random variable.

Eq. (2.2) follows from the combination of the two matching bounds:

$$\limsup_{n \rightarrow \infty} f_{\text{SW}} \leq \min_{q \in [0, \lambda]} f_{\text{RS}}(q), \quad (2.5)$$

$$\liminf_{n \rightarrow \infty} f_{\text{SW}} \geq \min_{q \in [0, \lambda]} f_{\text{RS}}(q). \quad (2.6)$$

Adaptive interpolation is a method to prove these two bounds in a unified way.

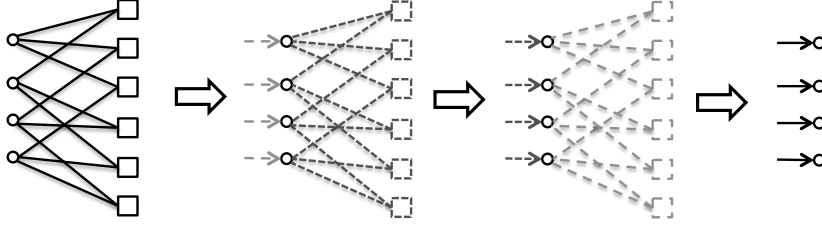


Figure 2.1 – Interpolation for dense graphs

## 2.1 Main steps

### Step 1: Set up an interpolating inference model

Since each of the two expressions,  $f_{\text{SW}}$  and  $f_{\text{RS}}$ , is associated with a respective inference problem defined in (2.1) and (2.4), the idea of interpolation method is based on constructing an interpolating inference model parameterized by time  $t \in [0, 1]$  such that the associated free energies at  $t = 0$  and  $t = 1$  recover  $f_{\text{SW}}$  and  $f_{\text{RS}}$ . A simple interpolating model at time  $t$  joins the two kinds of observations, and rescales the SNR of each kind by  $(1-t)$  and  $t$ . The particular feature of adaptive interpolation is replacing this linear rescaling by a function

$$R(t, \epsilon) \equiv \epsilon + \int_0^t ds q(s, \epsilon), \quad (2.7)$$

where  $\epsilon \in [s_n, 2s_n]$  with  $s_n = n^{-\theta}$  for some  $\theta \in (0, 1)$  such that  $s_n$  tends to  $0_+$ . The adaptive interpolating model thus involves these two kinds of observations:

$$\begin{aligned} Y_{ij}(t) &= \frac{(1-t)\lambda}{n} X_i X_j + Z_{ij}, & 1 \leq i \leq j \leq n, \\ \tilde{Y}_i(t) &= \sqrt{R(t, \epsilon)} X_i + \tilde{Z}_i, & 1 \leq i \leq n, \end{aligned}$$

with  $\tilde{Z}_i \sim \mathcal{N}(0, 1)$  i.i.d. Gaussian random variables. The perturbation  $\epsilon$  can be viewed as a negligible amount of hints for one to start to infer. Fig. 2.1 illustrates the factor graph of the interpolating model evolved with time  $t$ . The change of color and the intensity of the dash indicates the change of SNR of the observations.

We then set up the notations to link to the free energy. The posterior distribution of this interpolating model expressed in the convention of Gibbs distribution is

$$\mathbb{P}_t(\mathbf{x} | \mathbf{Y}, \tilde{\mathbf{Y}}) = \frac{e^{-\mathcal{H}_{t, \epsilon}(\mathbf{x}; \mathbf{Y}, \tilde{\mathbf{Y}})} \prod_{i=1}^n \mathbb{P}_0(x_i)}{\mathcal{Z}_{t, \epsilon}}$$

with the Hamiltonian

$$\begin{aligned} \mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{Y}(\mathbf{X}, \mathbf{Z}), \tilde{\mathbf{Y}}(\mathbf{X}, \tilde{\mathbf{Z}})) \\ \equiv - \sum_{i \leq j} \left( \frac{(1-t)\lambda x_i x_j X_i X_j}{n} + \frac{\sqrt{(1-t)\lambda}}{n} x_i x_j Z_{ij} - \frac{x_i^2 x_j^2}{2n} \right) \\ - \sum_{i=1}^n \left( R(t, \epsilon) X_i x_i + R(t, \epsilon) Z_i x_i - R(t, \epsilon) \frac{x_i^2}{2} \right) \end{aligned}$$

and the partition function  $\mathcal{Z}_{t,\epsilon} \equiv \int d\mathbf{x} e^{-\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{Y}, \tilde{\mathbf{Y}})} \prod_{i=1}^n \mathbb{P}_0(x_i)$ . We also define the Gibbs-bracket and the free energy for the interpolating model in the standard manner:

$$\begin{aligned} \langle A(\mathbf{x}) \rangle_{t,\epsilon} &\equiv \int d\mathbf{x} A(\mathbf{x}) \mathbb{P}_t(\mathbf{x} | \mathbf{Y}, \tilde{\mathbf{Y}}), \\ f_{t,\epsilon} &\equiv -\frac{1}{n} \mathbb{E} \ln \mathcal{Z}_{t,s}. \end{aligned}$$

We check that  $f_{0,0} = f_{\text{SW}}$  and  $f_{1,0} = f_{\text{RS}}(R(1,0)) - \frac{R(1,0)^2}{4\lambda}$ . Using the mean value theorem, it is also true that

$$\begin{cases} f_{0,\epsilon} &= f_{\text{SW}} + \mathcal{O}(s_n), \\ f_{1,\epsilon} &= f_{\text{RS}}(\int_0^1 q(s, \epsilon) ds) - \frac{(\int_0^1 q(s, \epsilon) ds)^2}{4\lambda} + \mathcal{O}(s_n). \end{cases} \quad (2.8)$$

## Step 2: Write down the sum rule

The difference of the free energies can then be derived by the formula

$$f_{0,\epsilon} = f_{1,\epsilon} - \int_0^1 dt \frac{df_{t,\epsilon}}{dt}. \quad (2.9)$$

The derivative  $\frac{df_{t,\epsilon}}{dt}$  can be obtained by the Nishimori identity<sup>1</sup> and Gaussian integration-by-parts formula (see Sec. 2.2.1 and 2.2.2) with the expression

$$\frac{df_{t,\epsilon}}{dt} = \frac{\lambda}{4} \mathbb{E} \langle Q^2 \rangle_{t,\epsilon} - \frac{1}{2} q(t, \epsilon) \mathbb{E} \langle Q \rangle_{t,\epsilon} + \mathcal{O}\left(\frac{1}{n}\right), \quad (2.10)$$

where  $Q \equiv n^{-1} \sum_{i=1}^n X_i x_i$  is called the overlap.

Substituting (2.8) and (2.10) into (2.9), we obtain

$$\begin{aligned} f_{\text{SW}} &= f_{\text{RS}} \left( \int_0^1 q(s, \epsilon) ds \right) - \frac{(\int_0^1 q(s, \epsilon) ds)^2}{4\lambda} - \frac{1}{4\lambda} \int_0^1 dt \mathbb{E} \langle (\lambda Q)^2 - 2q(t, \epsilon) Q \rangle_{t,\epsilon} \\ &\quad + \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}(s_n) \\ &= f_{\text{RS}} \left( \int_0^1 q(s, \epsilon) ds \right) + \mathcal{R}_1 - \frac{1}{4\lambda_n} \int_0^1 dt \mathcal{R}_2(t) + \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}(s_n) \end{aligned} \quad (2.11)$$

<sup>1</sup>The original Nishimori identity [57, 58] is obtained by “gauge invariance” of the posterior and the properties derived from channel symmetry (1.7). In Sec. 2.2.1 we show that a generalized result can be obtained from Bayes’ rule and transforms of dummy variables. Nevertheless, we call the generalized result Nishimori identities.

where

$$\begin{aligned}\mathcal{R}_1 &= \frac{1}{4\lambda} \left( \int_0^1 q(t, \epsilon)^2 dt - \left( \int_0^1 q(t, \epsilon) dt \right)^2 \right) \geq 0, \\ \mathcal{R}_2 &= \mathbb{E} \langle (\lambda Q - q(t, \epsilon))^2 \rangle_{t, \epsilon} \geq 0.\end{aligned}$$

Eq. (2.11) is the fundamental sum rule. We group the remainders  $\mathcal{R}_1$  and  $\mathcal{R}_2$  in such a way so that proving (2.5) or (2.6) amounts to choosing suitable  $R(t, \epsilon)$  to remove either of the remainders in (2.11).

### Step 3: Upper Bound (2.5)

The approach to obtaining (2.5) is to recover the simple version of the interpolation method. We set  $\epsilon = 0$  and  $q(t, \epsilon) = q$  a non-negative constant so that  $\mathcal{R}_1 = 0$ . With  $\mathcal{R}_2 \geq 0$ , (2.11) implies

$$f_{\text{SW}} \leq f_{\text{RS}}(q) + \mathcal{O}\left(\frac{1}{n}\right).$$

Optimizing over  $q \in [0, \lambda]$  and passing to the limit  $\limsup_{n \rightarrow \infty}$  yields (2.5).

### Step 4: Lower Bound (2.6)

For the other bound, it is tempting to cancel  $\mathcal{R}_2(t)$ . However, exact cancellation of  $\mathcal{R}_2(t)$  is impossible because  $Q$  is a random variable and  $q(t, \epsilon)$  is a parameter to be fixed. We can decompose  $\mathcal{R}_2(t)$  in the way analogous to the bias-variance decomposition:

$$\mathcal{R}_2(t) = (\lambda \mathbb{E} \langle Q \rangle_{t, \epsilon} - q(t, \epsilon))^2 + \lambda^2 \mathbb{E} \langle (Q - \mathbb{E} \langle Q \rangle)^2 \rangle_{t, \epsilon}, \quad (2.12)$$

Now  $\mathbb{E} \langle Q \rangle_{t, \epsilon}$  is no longer a random variable and one expects that at fixed  $\epsilon$  there is a choice

$$q(t, \epsilon) = \lambda \mathbb{E} \langle Q \rangle_{t, \epsilon}, \quad 0 \leq t \leq 1 \quad (2.13)$$

to remove the first term in (2.12). A crucial observation is recognizing  $q(t, \epsilon) = \frac{dR}{dt}(t, \epsilon)$  (recalling  $R(t, \epsilon) = \epsilon + \int_0^1 q(s, \epsilon) ds$ ) and that  $\lambda \mathbb{E} \langle Q \rangle_{t, \epsilon}$  is a bounded function  $G(t, R(t, \epsilon))$  in  $[0, \lambda]$ . Eq. (2.13) can thus be recast as a first-order differential equation

$$\frac{dR}{dt}(t, \epsilon) = G(t, R(t, \epsilon)) \quad \text{with} \quad R(0, \epsilon) = \epsilon. \quad (2.14)$$

The Cauchy-Lipschitz theorem (see for example [59, Chapter 5]) implies that (2.14) admits a unique global solution  $R^*(t, \epsilon)$  over  $t \in [0, 1]$ . Moreover, we can use (2.18) and (2.20) to see that  $\frac{dG}{dR}(t, R(t, \epsilon)) \geq 0$ . Through Liouville's formula (see Sec. 2.2.3)

$$\frac{dR^*}{d\epsilon}(t, \epsilon) = \exp \int_0^t dt' \frac{dG}{dR}(t', R^*(t', \epsilon)), \quad (2.15)$$

the non-negativity of  $dG/dR$  implies  $dR^*/d\epsilon \geq 1$ . When the second term in (2.12) is also evaluated at  $R = R^*$ , the property  $dR^*/d\epsilon \geq 1$  implies

$$\frac{1}{s_n} \int_{s_n}^{2s_n} d\epsilon \int_0^1 dt \mathbb{E} \langle (Q - \mathbb{E} \langle Q \rangle_{t,\epsilon})^2 \rangle_{t,\epsilon} \leq \frac{C(S)}{(s_n^4 n)^{1/3}} \quad (2.16)$$

for a positive constant  $C(S)$  depending on the support  $S$  of the prior  $\mathbb{P}_0$ .

Now we revisit (2.11) at  $R = R^*$  and average the equation over  $\epsilon \in [s_n, 2s_n]$ . Using  $\mathcal{R}_1 \geq 0$  and the above discussion on  $\mathcal{R}_2$ , we obtain

$$\begin{aligned} f_{\text{SW}} &\geq \frac{1}{s_n} \int_{s_n}^{2s_n} d\epsilon f_{\text{RS}}(R^*(1, \epsilon)) + \frac{C(S)}{(s_n^4 n)^{1/3}} + \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}(s_n) \\ &\geq \min_{q \in [0, \lambda]} f_{\text{RS}}(q) + \frac{C(S)}{(s_n^4 n)^{1/3}} + \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}(s_n). \end{aligned} \quad (2.17)$$

Setting  $s_n = n^{-\theta}$  with  $\theta \in (0, 1/4)$  ensures the extra terms on the r.h.s. of (2.17). Furthermore, taking  $\liminf_{n \rightarrow \infty}$  on both sides of (2.17) yields (2.5).

Before the end of this section, we remark that the main calculations we have hidden are (2.10) and (2.16). This is the question when one applies the method to other problems.

## 2.2 Tools

The tools presented in this section are generic and will be used again in the subsequent chapters.

### 2.2.1 The Nishimori identities

Let  $(\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}})$  be a couple of random variables with joint distribution

$$\mathbb{P}(\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}}) = \mathbb{P}(\mathbf{X}) \mathbb{P}(\mathbf{Y} | \mathbf{X}) \mathbb{P}(\tilde{\mathbf{Y}} | \mathbf{X})$$

and conditional distribution  $P(\cdot | \mathbf{Y}, \tilde{\mathbf{Y}})$ . Let  $k \geq 1$  and let  $x^{(1)}, \dots, x^{(k)}$  be i.i.d. copies from the conditional distribution. Let us denote  $\langle - \rangle$  the expectation w.r.t. the product distribution  $P(\cdot | \mathbf{Y}, \tilde{\mathbf{Y}})^{\otimes \infty}$  over copies and  $\mathbb{E}$  the expectation w.r.t. the joint distribution. Then, for all continuous bounded functions  $g$  we have

$$\mathbb{E} \langle g(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}) \rangle = \mathbb{E} \langle g(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{X}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(k)}) \rangle. \quad (2.18)$$

The expectation  $\mathbb{E}$  is over  $(\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}})$ .

*Proof.* This is a simple consequence of Bayes formula. We have

$$\begin{aligned} &\mathbb{E} \langle g(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}) \rangle \\ &= \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{Y}, \tilde{\mathbf{Y}} | \mathbf{X}} \mathbb{E}_{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)} | \mathbf{Y}, \tilde{\mathbf{Y}}} [g(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)})] \\ &= \mathbb{E}_{\mathbf{X}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}, \mathbf{Y}, \tilde{\mathbf{Y}}} [g(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)})] \\ &= \mathbb{E}_{\mathbf{x}^{(1)}} \mathbb{E}_{\mathbf{Y}, \tilde{\mathbf{Y}} | \mathbf{x}^{(1)}} \mathbb{E}_{\mathbf{X}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(k)} | \mathbf{Y}, \tilde{\mathbf{Y}}} [g(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)})]. \end{aligned} \quad (2.19)$$



Since  $\mathbf{X}$  and  $\mathbf{x}^{(1)}$  are dummy, we can do the transform  $(\mathbf{X}, \mathbf{x}^{(1)}) \rightarrow (\mathbf{x}^{(1)}, \mathbf{X})$ . We can continue (2.19) with

$$\begin{aligned} & \mathbb{E}\langle g(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}) \rangle \\ &= \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{Y}, \tilde{\mathbf{Y}} | \mathbf{X}} \mathbb{E}_{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)} | \mathbf{Y}, \tilde{\mathbf{Y}}} [g(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{X}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(k)})] \\ &= \mathbb{E}\langle g(\mathbf{Y}, \tilde{\mathbf{Y}}, \mathbf{X}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(k)}) \rangle. \end{aligned}$$

□

### 2.2.2 Gaussian integration by parts

Integration by parts implies that for any bounded and differentiable function  $g$  of  $Z \sim \mathcal{N}(0, 1)$ , we have

$$\mathbb{E}[Zg(Z)] = \mathbb{E}[g'(Z)]. \quad (2.20)$$

### 2.2.3 Liouville formula

Consider the differential equation (2.14) with  $G(t, R(t, \epsilon)) = \lambda \mathbb{E}\langle Q \rangle_{t, \epsilon}$ . Differentiating w.r.t  $\epsilon$  and using the chain rule gives

$$\frac{d}{dt} \frac{dR}{d\epsilon}(t, \epsilon) = \frac{dR}{d\epsilon}(t, \epsilon) \frac{dG}{dR}(t, R(t, \epsilon)).$$

Therefore we have

$$\frac{d}{dt} \ln \left\{ \frac{dR}{d\epsilon}(t, \epsilon) \right\} = \frac{dG}{dR}(t, R(t, \epsilon)). \quad (2.21)$$

Integrating (2.21) over  $t \in [0, t']$ , we have

$$\ln \left\{ \frac{dR}{d\epsilon}(t', \epsilon) \right\} - \ln \left\{ \frac{dR}{d\epsilon}(0, \epsilon) \right\} = \int_0^{t'} dt \frac{dG}{dR}(t, R(t, \epsilon)). \quad (2.22)$$

Using  $R(0, \epsilon) = \epsilon$ , (2.22) implies

$$\frac{dR}{d\epsilon}(t', \epsilon) = \exp \left\{ \int_0^{t'} dt \frac{dG}{dR}(t, R(t, \epsilon)) \right\}. \quad (2.23)$$

This is known as Liouville's formula for one-dimensional ordinary differential equations.



---

# Dense Stochastic Block Model

---

# 3

## 3.1 Introduction

In this chapter, we focus on the mutual information of the *two-group* SBM with possibly *asymmetric* group sizes, in *dense regimes* where the expected degree of the nodes diverges with the total number of nodes (and is independent of the group label). We rigorously determine a *single-letter variational expression* for the asymptotic mutual information by means of the adaptive interpolation method.

Single-letter variational expressions for the mutual information of the SBM are not new. They were first analytically derived in heuristic ways by methods of statistical physics and in this context are often called replica or cavity formulas [49]. Rigorous proofs then appeared in [16, 17]. These approaches are indirect in the sense that the SBM is first mapped on a spiked Wigner model, and then the spiked Wigner model is solved. In [16], the particular case of two equal-size communities is considered and the analysis relies on the fact that in this case the information-theoretic phase transition is of the second-order type (i.e., continuous), which allows to use message-passing arguments. The asymmetric case is more challenging because first-order (discontinuous) phase transitions appears for large enough asymmetry. In [17], this case is tackled through a Guerra-Toninelli interpolation combined with a rigorous version of the cavity method or Aizenman-Sims-Starr scheme [60]. Strictly speaking, the analysis [17] does not cover the widest possible regime of dense graphs (see section 3.2 for details). We note that the mutual information of the spiked Wigner model had also been determined earlier in [61] for the symmetric case and more recently for the general case in [43] using a spatial coupling method.

The proof presented here covers the asymmetric two-group SBM and has the virtue of being completely unified. It uses a single method, namely, the

adaptive interpolation, which is conceptually simpler and is direct, as it does not make any detour through another model. The method is a powerful evolution of the classic Guerra-Toninelli interpolation [36] and allows to derive tight upper *and* lower bounds for the mutual information, whereas the classic interpolation only yields a one-sided inequality. It has been successfully applied to a range of Bayesian inference problems, e.g., [45, 62]. Here, besides various new technical aspects, the main novelty is that we do not use Gaussian integration by parts, as is generally the case in interpolation methods. Instead, we develop a general *approximate integration-by-parts* formula and apply it to the Bernoulli random elements of the adjacency matrix of the graph. We note that related approximate integration-by-parts formulas have already been used by [63, 64] in the context of the Hopfield and Sherrington-Kirkpatrick models.

This chapter is organized as follows. In Section 3.2, we give a precise formulation of the model and state the main result of this chapter (Theorem 3.4). In Section 3.3, we formulate the adaptive interpolation method for the dense SBM. The derivation of the sum rule is provided in Section 3.4. Overlap concentration is proved in Section 3.5. A technical lemma important for the sum rule is proved in Section 3.6.1. The rest of the technical results are found in the appendices.

## 3.2 Setting and results: asymmetric two-group SBM

We first formulate the SBM for two communities that may be of different sizes. Suppose we have  $n$  nodes belonging to two communities where the partition is denoted by a vector  $\boldsymbol{\sigma}^0 \in \{-1, 1\}^n$ . Labels  $X_i^0$  are i.i.d. Bernoulli random variables with  $\mathbb{P}(\sigma_i^0 = 1) = r \in (0, 1/2]$ . The size of each community is  $nr$  and  $n(1-r)$  up to fluctuations of  $\mathcal{O}(\sqrt{n})$ . The labels  $\boldsymbol{\sigma}^0$  are hidden and instead one is given a random undirected graph  $\mathbf{G}$  constructed as follows (equivalently one is given an adjacency matrix). An edge between node  $i$  and  $j$  is present with probability  $\mathbb{P}(G_{ij} = 1 | \sigma_i^0, \sigma_j^0)$  and absent with the complementary probability. To specify  $\mathbb{P}(G_{ij} = 1 | \sigma_i^0, \sigma_j^0)$ , first we define  $d_n$  such that

$$\mathbb{E}[\deg(i) | \sigma_i^0 = 1] \equiv \frac{(n-1)d_n}{n} \approx d_n, \quad (3.1)$$

$$\mathbb{E}[\deg(i) | \sigma_i^0 = -1] \equiv \frac{(n-1)d_n}{n} \approx d_n. \quad (3.2)$$

We require these two constraints for the inference problem to be non-trivial, in the sense that no information about the labels stems from the nodes' degrees. The two constraints imply

$$\mathbb{E}[\deg(i)] = r \mathbb{E}[\deg(i) | \sigma_i^0 = 1] + (1-r) \mathbb{E}[\deg(i) | \sigma_i^0 = -1] = \frac{(n-1)d_n}{n} \approx d_n$$

so that we can interpret  $d_n$  as the average degree of a node. Then we define  $\mathbb{P}(G_{ij} = 1 | \sigma_i^0, \sigma_j^0) = M_{\sigma_i^0, \sigma_j^0}$  where  $M_{\sigma_i^0, \sigma_j^0}$  are the four possible matrix elements of

$$M = \frac{d_n}{n} \begin{bmatrix} a_n & b_n \\ b_n & c_n \end{bmatrix}.$$

Because of (3.1) and (3.2), we have the equations

$$\begin{aligned} \mathbb{E}[\deg(i) | \sigma_i^0 = 1] &= \frac{(n-1)d_n}{n} (ra_n + (1-r)b_n) = \frac{(n-1)d_n}{n}, \\ \mathbb{E}[\deg(i) | \sigma_i^0 = -1] &= \frac{(n-1)d_n}{n} (rb_n + (1-r)c_n) = \frac{(n-1)d_n}{n}. \end{aligned}$$

Solving this system imposes  $a_n = 1 - (1 - 1/r)(1 - b_n)$  and  $c_n = 1 - (1 - b_n)/(1 - 1/r)$ . Therefore there are three independent parameters, namely  $d_n$ ,  $b_n$  and  $r$ . A more convenient re-parametrization is often used [16] instead of  $b_n, d_n$ :

$$\bar{p}_n \equiv \frac{d_n}{n}, \quad \text{and} \quad \Delta_n \equiv \frac{d_n(1 - b_n)}{n}.$$

Here  $\bar{p}_n \in (0, 1)$  is the average probability for the presence of an edge. We will look at the *dense* asymmetric SBM (the symmetric model corresponding to  $r = 1/2$ ) regimes where  $d_n = n\bar{p}_n \rightarrow +\infty$ . In our analysis the growth of  $d_n$  spans the whole spectrum from arbitrarily slow, at the verge of a sparse graph, to linear  $d_n = vn$ ,  $v \in (0, 1)$ , for fully dense graphs.

In this chapter we rigorously determine the asymptotic mutual information for this problem  $\lim_{n \rightarrow \infty} \frac{1}{n} I(\boldsymbol{\sigma}^0; \mathbf{G})$  in the *dense graph regime* wherein  $\bar{p}_n$  and  $\Delta_n$  satisfy:

- (h1) (Dense SBM)  $n\bar{p}_n(1 - \bar{p}_n)^3 \xrightarrow{n \rightarrow \infty} \infty$ .
- (h2) (Appropriate scaling of signal-to-noise ratio)  $\lambda_n \equiv n\Delta_n^2 / (\bar{p}_n(1 - \bar{p}_n)) = d_n(1 - b_n)^2 / (1 - d_n/n) \xrightarrow{n \rightarrow \infty} \lambda$  finite.

The first condition ensures that the graph is dense in the sense that  $d_n \rightarrow +\infty$ , still maintaining  $\bar{p}_n \in (0, 1)$ . The second ensures the mutual information has a well defined non-trivial limit when  $n \rightarrow +\infty$ . Note that the second condition requires  $\Delta_n \ll \bar{p}_n(1 - \bar{p}_n)^2$  as  $\Delta_n / (\bar{p}_n(1 - \bar{p}_n)^2) = \sqrt{\lambda_n / (n\bar{p}_n(1 - \bar{p}_n)^3)} \rightarrow 0$  as  $n \rightarrow \infty$ , hence  $\Delta_n \ll \bar{p}_n$  and  $\Delta_n \ll (1 - \bar{p}_n)^2$ . The reader may wish to keep in mind two simple typical examples. The first example is a dense graph with  $d_n = vn$ ,  $v \in [0, 1]$  so  $\bar{p}_n = v$  and  $\Delta_n \approx \sqrt{\lambda v(1 - v)/n}$ . The second example is  $d_n = vn^{1-\theta}$  with  $\theta \in (0, 1)$ , so  $\bar{p}_n = vn^{-\theta}$  and  $\Delta_n \approx \sqrt{\lambda vn^{-1-\theta}}$ . These are easily translated back to the matrix  $M$ .

We note that in the sparse graph version of the model one would have a finite limit for  $d_n$  but the second condition would be the same. The analysis of the sparse case is however more difficult and is not addressed in this chapter.

Instead of working with the spin  $\pm 1$  variables it is convenient to change the alphabet. We define  $X_i \equiv \phi_r(\sigma_i^0)$  with  $\phi_r(1) = \sqrt{(1 - r)/r}$  and  $\phi_r(-1) = -\sqrt{r/(1 - r)}$ . The hidden labels of the nodes now belong to the alphabet

$\mathcal{X} \equiv \{\mathcal{X}_1 = \sqrt{(1-r)/r}, \mathcal{X}_2 = -\sqrt{r/(1-r)}\}$  and  $\mathbf{X} \in \mathcal{X}^n$ . An edge is then present with conditional probability

$$\mathbb{P}(G_{ij} = 1 | X_i X_j) = \bar{p}_n + \Delta_n X_i X_j. \quad (3.3)$$

This can be viewed as an asymmetric binary-input binary-output channel  $\mathbf{X} \rightarrow \mathbf{G}$  and the inference problem is to recover the input  $\mathbf{X}$  (or  $\boldsymbol{\sigma}^0$ ) from the channel output  $\mathbf{G}$ . Henceforth we adopt the notation

$$\mathbb{P}_r \equiv r\delta_{\mathcal{X}_1} + (1-r)\delta_{\mathcal{X}_2}$$

for the probability distribution of the hidden labels  $\mathcal{X} \in \mathcal{X}$ . Note that  $\mathbb{E}[X^2] = 1$ .

We now formulate our results which provide a single-letter variational formula for the asymptotic mutual information. Let  $Z \sim \mathcal{N}(0, 1)$  and  $X \sim \mathbb{P}_r$  independently, and set for  $q > 0$ :

$$i_{\text{RS}}(q, \lambda, r) \equiv \frac{\lambda}{4} + \frac{q^2}{4\lambda} - \mathbb{E} \ln \sum_{x \in \mathcal{X}} \mathbb{P}_r(x) e^{\sqrt{q} Z x + q X x - \frac{q}{2} x^2}.$$

The so-called replica formula conjectures the identity

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(\boldsymbol{\sigma}^0; \mathbf{G}) = \min_{q \in [0, \lambda]} i_{\text{RS}}(q, \lambda, r). \quad (3.4)$$

We prove that (3.4) is correct, namely:

**Theorem 3.1** (Upper bound). *For the SBM under concern in the regime (h1), (h2),*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} I(\boldsymbol{\sigma}^0; \mathbf{G}) \leq \min_{q \in [0, \lambda]} i_{\text{RS}}(q, \lambda, r).$$

**Theorem 3.2** (Lower bound). *For the SBM under concern in the regime (h1), (h2),*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} I(\boldsymbol{\sigma}^0; \mathbf{G}) \geq \min_{q \in [0, \lambda]} i_{\text{RS}}(q, \lambda, r).$$

*Remark 1:* Of course we have  $I(\boldsymbol{\sigma}^0; \mathbf{G}) = I(\mathbf{X}; \mathbf{G})$  and in the following we will work with  $I(\mathbf{X}; \mathbf{G})$  where

$$\mathbf{X} \in \mathcal{X} = \left\{ \mathcal{X}_1 = \sqrt{\frac{1-r}{r}}, \mathcal{X}_2 = -\sqrt{\frac{r}{1-r}} \right\}.$$

*Remark 2:* Elementary analysis shows that the minimum over  $q \geq 0$  of  $i_{\text{RS}}(q, \lambda, r)$  is attained for  $q \in [0, \lambda]$ .

*Remark 3:* From (3.4) one can derive the information theoretic phase transition thresholds. Let  $r_* \equiv (1 - 1/\sqrt{3})/2$ . For "small" asymmetry between group sizes  $r \in [r_*, 1/2]$  there is a continuous phase transition at  $\lambda_c = 1$  while

for “large” asymmetry  $r \in (0, r_*)$  the phase transition becomes discontinuous. An information theoretic-to-algorithmic gap occurs in the second situation as discussed in detail in [17].

Let us explain the relation of these theorems with previous works. In [16] they were obtained for the symmetric case  $r = 1/2$  by a mapping of the model on a rank-one matrix estimation problem via an application of Lindeberg’s theorem. The regime treated is essentially the same than ours except that in place of  $(h1)$  [16] has  $n\bar{p}_n(1-\bar{p}_n) \rightarrow +\infty$ . Note that the difference only matters if  $p_n \rightarrow 1$  which is the complete graph limit. Still using the same mapping to matrix factorization, [17] treats the asymmetric case, however in a limit where  $n \rightarrow +\infty$  first and  $d_n \rightarrow +\infty$  after (in fact, this analysis can accomodate any growth slower than  $d_n \approx n^{1/2}$ ). It is unclear whether this is possible for denser regimes. Our analysis covers this gap and the whole spectrum of growth for  $d_n$  up to linear growth is allowed. Besides, we propose a self-contained and direct method using the adaptive interpolation method [55]. A technical limitation of interpolation methods has often been the need to use Gaussian integration by parts. We by-pass this limitation using an (approximate) integration-by-parts formula for the edge binary variables  $G_{ij} \in \{0, 1\}$ .

Before we formulate the adaptive interpolation, let us set up more explicitly the quantities that we compute. The distribution of  $G$  given the hidden partition  $X$  is the inhomogeneous Erdős-Rényi graph measure:

$$\mathbb{P}(\mathbf{G}|\mathbf{X}) = \prod_{i < j} (\bar{p}_n + \Delta_n X_i X_j)^{G_{ij}} (1 - \bar{p}_n - \Delta_n X_i X_j)^{1-G_{ij}}.$$

Using this measure and Bayes rule, we find the posterior distribution of the SBM

$$\begin{aligned} \mathbb{P}(\mathbf{X} = \mathbf{x}|\mathbf{G}) &= \mathbb{P}(\mathbf{x}|\mathbf{G}) = \frac{\mathbb{P}(\mathbf{G}|\mathbf{x})\mathbb{P}(\mathbf{x})}{\mathbb{P}(\mathbf{G})} \propto \mathbb{P}(\mathbf{G}|\mathbf{x})\mathbb{P}(\mathbf{x}) \\ &= \exp \left\{ \sum_{i < j} \left( G_{ij} \ln(\bar{p}_n + \Delta_n x_i x_j) + (1 - G_{ij}) \ln(1 - \bar{p}_n - \Delta_n x_i x_j) \right) \right\} \prod_{i=1}^n \mathbb{P}_r(x_i) \\ &= \exp \left\{ \sum_{i < j} \left( G_{ij} \ln\left(1 + \frac{\Delta_n}{\bar{p}_n} x_i x_j\right) + (1 - G_{ij}) \ln\left(1 - \frac{\Delta_n}{1 - \bar{p}_n} x_i x_j\right) \right) \right. \\ &\quad \left. + D_n(\bar{p}_n, \mathbf{G}) \right\} \prod_{i=1}^n \mathbb{P}_r(x_i) \end{aligned}$$

where  $D_n(\bar{p}_n, \mathbf{G}) \equiv \sum_{i < j} G_{ij} \ln \bar{p}_n + (1 - G_{ij}) \ln(1 - \bar{p}_n)$ . Therefore, the posterior distribution becomes

$$\begin{aligned} \mathbb{P}(\mathbf{x}|\mathbf{G}) &= \frac{1}{\mathcal{Z}(\mathbf{G})} e^{-\mathcal{H}_{\text{SBM}}(\mathbf{x}; \mathbf{G})} \prod_{i=1}^n \mathbb{P}_r(x_i), \\ \mathcal{H}_{\text{SBM}}(\mathbf{x}; \mathbf{G}) &\equiv - \sum_{i < j} \left\{ G_{ij} \ln\left(1 + x_i x_j \frac{\Delta_n}{\bar{p}_n}\right) + (1 - G_{ij}) \ln\left(1 - x_i x_j \frac{\Delta_n}{1 - \bar{p}_n}\right) \right\}. \end{aligned}$$

We use the statistical mechanics terminology and therefore call this posterior distribution the Gibbs distribution. The normalizing factor

$$\mathcal{Z}(\mathbf{G}) \equiv \sum_{\mathbf{x} \in \mathcal{X}^n} e^{-\mathcal{H}_{\text{SBM}}(\mathbf{x}; \mathbf{G})} \prod_{i=1}^n \mathbb{P}_r(x_i)$$

is the partition function, and  $\mathcal{H}_{\text{SBM}}$  is the Hamiltonian. A straightforward computation, using the scaling regime (h1) and (h2), gives the following formula (see the proof in Appendix 3.6.2):

**Proposition 3.1** (Linking the mutual information and log-partition function). *For the SBM under concern we have*

$$\frac{1}{n} I(\mathbf{X}; \mathbf{G}) = -\frac{1}{n} \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \ln \mathcal{Z}(\mathbf{G}) + \frac{\lambda_n}{4} + o_n(1) \quad (3.5)$$

where  $\lim_{n \rightarrow \infty} o_n(1) = 0$ .

The problem thus boils down to compute minus the expected log-partition function, or the expected free energy, in the limit  $n \rightarrow +\infty$ . This will be achieved via an interpolation towards the log-partition function of  $n$  independent scalar Gaussian channels where the observations about the hidden labels are of the form

$$Y_i = \sqrt{q} X_i + Z_i, \quad 1 \leq i \leq n, \quad (3.6)$$

with  $Z_i \sim \mathcal{N}(0, 1)$  i.i.d. Gaussian random variables and  $q > 0$  the signal-to-noise ratio (SNR). An important feature of our technique is the freedom to adapt a suitable interpolation path to the problem at hand. This is explained in the next section.

### 3.3 Adaptive path interpolation

We design an *interpolating model* parametrized by  $t \in [0, 1]$  and  $\epsilon \geq 0$  s.t. at  $t = \epsilon = 0$  we recover the original SBM, while at  $t = 1$  we have a decoupled channel similar to (3.6). For  $t \in (0, 1)$  the model is a mixture of the SBM with parameters  $(\bar{p}_n, \sqrt{1-t} \Delta_n)$  and the extra *decoupled* Gaussian observations (3.6) with SNR replaced by

$$q \rightarrow R(t, \epsilon) \equiv \epsilon + \int_0^t ds q(s, \epsilon)$$



with  $q(s, \epsilon) \geq 0$ . The transition kernels for the channels  $\mathbf{X} \rightarrow \mathbf{G}$  and  $\mathbf{X} \rightarrow \mathbf{Y}$  at time  $t \in [0, 1]$  are

$$\begin{aligned} \mathbb{P}_t(\mathbf{G}|\mathbf{X}) &= \prod_{i < j} (\bar{p}_n + \sqrt{1-t}\Delta_n X_i X_j)^{G_{ij}} (1 - \bar{p} - \sqrt{1-t}\Delta_n X_i X_j)^{1-G_{ij}} \\ &= \exp \sum_{i < j} \left( G_{ij} \ln(\bar{p} + \sqrt{1-t}\Delta_n X_i X_j) \right. \\ &\quad \left. + (1 - G_{ij}) \ln(1 - \bar{p}_n - \sqrt{1-t}\Delta_n X_i X_j) \right), \end{aligned} \quad (3.7)$$

$$\mathbb{P}_t(\mathbf{Y}|\mathbf{X}) = \frac{1}{(2\pi)^{n/2}} \exp \left( -\frac{1}{2} \sum_{i=1}^n (Y_i - \sqrt{R(t, \epsilon)} X_i)^2 \right). \quad (3.8)$$

We constrain  $\epsilon \in [s_n, 2s_n]$  where  $s_n \rightarrow 0_+$  as  $n \rightarrow +\infty$  at an appropriate rate to be fixed later on. The interpolating Hamiltonian is then defined to be

$$\mathcal{H}_{t, \epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{Y}) \equiv \mathcal{H}_{\text{SBM}; t}(\mathbf{x}; \mathbf{G}) + \mathcal{H}_{\text{dec}; t, \epsilon}(\mathbf{x}; \mathbf{Y})$$

where

$$\begin{aligned} \mathcal{H}_{\text{SBM}; t}(\mathbf{x}; \mathbf{G}) &\equiv - \sum_{i < j} \left( G_{ij} \ln(1 + x_i x_j \sqrt{1-t} \frac{\Delta_n}{\bar{p}_n}) \right. \\ &\quad \left. + (1 - G_{ij}) \ln(1 - x_i x_j \sqrt{1-t} \frac{\Delta_n}{1 - \bar{p}_n}) \right), \quad (3.9) \\ \mathcal{H}_{\text{dec}; t, \epsilon}(\mathbf{x}; \mathbf{Y}(\mathbf{X}, \mathbf{Z})) &\equiv - \sum_{i=1}^n \left( \sqrt{R(t, \epsilon)} Y_i x_i - R(t, \epsilon) \frac{x_i^2}{2} \right) \\ &= - \sum_{i=1}^n \left( R(t, \epsilon) X_i x_i + \sqrt{R(t, \epsilon)} Z_i x_i - R(t, \epsilon) \frac{x_i^2}{2} \right). \end{aligned} \quad (3.10)$$

The posterior distribution expressed with the Hamiltonian  $\mathcal{H}_{t, \epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{Y})$  then reads

$$\mathbb{P}_t(\mathbf{x}|\mathbf{G}, \mathbf{Y}) = \frac{\prod_{i=1}^n \mathbb{P}_r(x_i) \exp(-\mathcal{H}_{t, \epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{Y}))}{\sum_{\mathbf{x} \in \mathcal{X}^n} \prod_{i=1}^n \mathbb{P}_r(x_i) \exp(-\mathcal{H}_{t, \epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{Y}))}.$$

Therefore, the *Gibbs-bracket* (i.e., the expectation operator w.r.t. the posterior distribution) for the interpolating model is

$$\langle A \rangle_{t, \epsilon} \equiv \sum_{\mathbf{x} \in \mathcal{X}^n} A(\mathbf{x}) \mathbb{P}_t(\mathbf{x}|\mathbf{G}, \mathbf{Y}) = \frac{1}{\mathcal{Z}_{t, \epsilon}(\mathbf{G}, \mathbf{Y})} \sum_{\mathbf{x} \in \mathcal{X}^n} A(\mathbf{x}) e^{-\mathcal{H}_{t, \epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{Y})} \prod_{i=1}^n \mathbb{P}_r(x_i)$$

with the partition function  $\mathcal{Z}_{t, \epsilon}(\mathbf{G}, \mathbf{Y}) \equiv \sum_{\mathbf{x} \in \mathcal{X}^n} e^{-\mathcal{H}_{t, \epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{Y})} \prod_{i=1}^n \mathbb{P}_r(x_i)$ . The reader should keep in mind that Gibbs-brackets are therefore functions of the *quenched* random variables  $(\mathbf{Y}(\mathbf{X}, \mathbf{Z}), \mathbf{G}(\mathbf{X}))$ . The free energy for a given

graph  $\mathbf{G} = \mathbf{G}(\mathbf{X})$  (that depends on the ground truth partition) and decoupled observation  $\mathbf{Y}(\mathbf{X}, \mathbf{Z})$  is

$$F_{t,\epsilon}(\mathbf{G}, \mathbf{Y}) = F_{t,\epsilon} \equiv -\frac{1}{n} \ln \mathcal{Z}_{t,\epsilon}(\mathbf{G}, \mathbf{Y}), \quad (3.11)$$

and its expectation

$$f_{t,\epsilon} \equiv \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Y}|\mathbf{X}} F_{t,\epsilon} = \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}. \quad (3.12)$$

By construction,

$$\begin{aligned} f_{t=0,\epsilon} &= -\frac{1}{n} \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} \ln \left( \sum_{\mathbf{x} \in \mathcal{X}^n} \exp \left\{ \sum_{i < j} \left( G_{ij} \ln \left( 1 + \frac{\Delta_n}{\bar{p}_n} x_i x_j \right) \right. \right. \right. \\ &\quad \left. \left. \left. + (1 - G_{ij}) \ln \left( 1 - \frac{\Delta_n}{1 - \bar{p}_n} x_i x_j \right) \right) \right. \right. \\ &\quad \left. \left. + \sum_{i=1}^n \left( \sqrt{\epsilon} Z_i x_i + \epsilon X_i x_i - \frac{\epsilon}{2} x_i^2 \right) \right\} \prod_{i=1}^n \mathbb{P}_r(x_i) \right), \\ f_{t=1,\epsilon} &= -\frac{1}{n} \mathbb{E}_{\mathbf{Z}} \ln \left( \sum_{\mathbf{x} \in \mathcal{X}^n} \exp \left\{ \sum_{i=1}^n \left( \sqrt{R(1, \epsilon)} Z_i x_i \right. \right. \right. \\ &\quad \left. \left. \left. + R(1, \epsilon) X_i x_i - \frac{R(1, \epsilon)}{2} x_i^2 \right) \right\} \prod_{i=1}^n \mathbb{P}_r(x_i) \right) \\ &= i_{\text{RS}}(R(1, \epsilon), \lambda_n, r) - \frac{\lambda_n}{4} - \frac{R(1, \epsilon)^2}{4\lambda_n}. \end{aligned}$$

In particular, when  $t = \epsilon = 0$  we have

$$f_{0,0} = \frac{1}{n} I(\mathbf{X}; \mathbf{G}) - \frac{\lambda_n}{4} + o_n(1).$$

Therefore

$$\begin{aligned} \frac{1}{n} I(\mathbf{X}; \mathbf{G}) &= f_{0,0} + \frac{\lambda_n}{4} + o_n(1) \\ &= i_{\text{RS}}(R(1, \epsilon), \lambda_n, r) - \frac{R(1, \epsilon)^2}{4\lambda_n} - f_{1,\epsilon} + f_{0,0} + o_n(1) \quad (3.13) \\ &= i_{\text{RS}}(R(1, \epsilon), \lambda_n, r) - \frac{R(1, \epsilon)^2}{4\lambda_n} - \int_0^1 dt \frac{df_{t,\epsilon}}{dt} + (f_{0,0} - f_{0,\epsilon}) + o_n(1) \quad (3.14) \end{aligned}$$

where  $o_n(1)$  collects all contributions that tend to zero uniformly in  $\epsilon$  when  $n \rightarrow \infty$ . Eventually, we reach the following fundamental sum rule (see section 3.4 for the derivation):

$$\frac{1}{n} I(\mathbf{X}; \mathbf{G}) = i_{\text{RS}}(R(1, \epsilon), \lambda_n, r) + \mathcal{R}_1 - \frac{1}{4\lambda_n} \int_0^1 dt \mathcal{R}_2(t) - \mathcal{R}_3 \quad (3.15)$$

where

$$\begin{aligned}\mathcal{R}_1 &\equiv \frac{1}{4\lambda_n} \left( \int_0^1 q(t, \epsilon)^2 dt - \left( \int_0^1 q(t, \epsilon) dt \right)^2 \right) \geq 0, \\ \mathcal{R}_2(t) &\equiv \mathbb{E} \langle (\lambda_n Q - q(t, \epsilon))^2 \rangle_{t, \epsilon} \geq 0, \\ \mathcal{R}_3 &\equiv \frac{\epsilon}{4\lambda_n} \left( \epsilon + 2 \int_0^1 q(t, \epsilon) dt \right) - \frac{1}{2} \int_0^\epsilon d\epsilon' \mathbb{E} \langle Q \rangle_{0, \epsilon'} + o_n(1),\end{aligned}$$

and the *overlap* is

$$Q(\mathbf{X}, \mathbf{x}) = Q \equiv \frac{1}{n} \sum_{i=1}^n X_i x_i.$$

Two generic tools that we will widely use in our proof are (2.18) and (2.20). To adapt (2.18) to the present case, we map  $(\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}}) \rightarrow (\mathbf{X}, \mathbf{G}, \mathbf{Y})$  with joint law

$$\mathbb{P}_t(\mathbf{X} | \mathbf{G}, \mathbf{Y}) \prod_{i=1}^n \mathbb{P}_r(X_i).$$

Let us take  $k$  i.i.d. copies  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}$  drawn from the posterior distribution  $\mathbb{P}_t(\cdot | \mathbf{G}, \mathbf{Y})$ . Then for any continuous bounded function  $g$

$$\mathbb{E} \langle g(\mathbf{G}, \mathbf{Y}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)}, \mathbf{X}) \rangle_{t, \epsilon} = \mathbb{E} \langle g(\mathbf{G}, \mathbf{Y}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)}, \mathbf{x}^{(k)}) \rangle_{t, \epsilon}. \quad (3.16)$$

where  $\mathbb{E}$  is over  $(\mathbf{G}, \mathbf{Y})$ . More precisely,  $\mathbb{E} = \mathbb{E}_{\prod_{i=1}^n \mathbb{P}_r(X_i)} \mathbb{E}_{\mathbb{P}_t(\mathbf{G} | \mathbf{X})} \mathbb{E}_{\mathbb{P}_t(\mathbf{Y} | \mathbf{X})}$ . Note that, by a slight abuse of notation, we continue to use the Gibbs-bracket notation for expressions depending on multiple i.i.d. copies from the posterior, so that  $\langle - \rangle_{t, \epsilon}$  corresponds to the expectation w.r.t. the product measure  $\mathbb{P}_t(\cdot | \mathbf{G}, \mathbf{Y})^{\otimes \infty}$ .

We are now ready to provide the proofs of the bounds on the mutual information.

### 3.3.1 The upper bound: proof of Theorem 3.1

Set  $\epsilon = 0$  and  $q(t, \epsilon) = q$  a non-negative constant. Then we have  $\mathcal{R}_1 = 0$ ,  $\mathcal{R}_3 = o_n(1)$ . Since  $\mathcal{R}_2 \geq 0$ , (3.15) implies

$$\frac{1}{n} I(\mathbf{X}; \mathbf{G}) \leq i_{\text{RS}}(q, \lambda_n, r) + o_n(1).$$

Since  $i_{\text{RS}}$  is continuous w.r.t its second argument  $\limsup_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{X}; \mathbf{G}) \leq i_{\text{RS}}(q, \lambda, r)$ . Optimizing over  $q \in [0, \lambda]$  yields the bound (optimization over  $q \in [0, +\infty)$  does not yield a sharper bound, see remark 2).

### 3.3.2 The lower bound: proof of Theorem 3.2

The basic idea is to “remove”  $\mathcal{R}_2$  from (3.15) by *adapting*  $q(t, \epsilon)$ . Then taking the limit  $n \rightarrow \infty$  and  $\epsilon \rightarrow 0_+$  will provide the desired bound since  $\mathcal{R}_1 \geq 0$  and  $\mathcal{R}_3 \rightarrow 0$  will disappear. To implement this idea we first decompose  $\mathcal{R}_2$  into

$$\mathcal{R}_2 = (\lambda_n \mathbb{E}\langle Q \rangle_{t,\epsilon} - q(t, \epsilon))^2 + \lambda_n^2 \mathbb{E}\langle (Q - \mathbb{E}\langle Q \rangle_{t,\epsilon})^2 \rangle_{t,\epsilon} \quad (3.17)$$

and address each part with the following two lemmas. The proof of Lemma 3.2 can be found in section 3.5.

**Lemma 3.1.** *For every  $\epsilon \in [0, 1]$  and  $t \in [0, 1]$  there exists a (unique) bounded solution  $R_n^*(t, \epsilon) = \epsilon + \int_0^t ds q_n^*(s, \epsilon)$  to the first order differential equation*

$$\frac{dR}{dt}(t, \epsilon) = \lambda_n \mathbb{E}\langle Q \rangle_{t,\epsilon} \quad \text{with} \quad R(0, \epsilon) = \epsilon. \quad (3.18)$$

Furthermore

$$q_n^*(t, \epsilon) = \lambda_n \mathbb{E}\langle Q \rangle_{t,\epsilon} \in [0, \lambda_n], \quad \text{and} \quad \frac{dR_n^*}{d\epsilon}(t, \epsilon) \geq 1.$$

*Proof.* Let  $G_n(t, R(t, \epsilon)) \equiv \lambda_n \mathbb{E}\langle Q \rangle_{t,\epsilon}$ . Equation (3.18) is thus a first-order differential equation. Also note that, letting  $dG_n/dR$  be the derivative w.r.t. the second argument,

$$\begin{aligned} & \frac{dG_n}{dR}(t, R(t, \epsilon)) \\ &= \frac{\lambda_n}{n} \sum_{i=1}^n \mathbb{E} \left[ X_i \sum_{\mathbf{x} \in \mathcal{X}^n} x_i \mathbb{P}_r(\mathbf{x}) \frac{d}{dR} \frac{e^{-\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{Y})}}{\mathcal{Z}_{t,\epsilon}(\mathbf{G}, \mathbf{Y})} \right] \\ &= \frac{\lambda_n}{n} \sum_{i=1}^n \mathbb{E} \left[ X_i \sum_{\mathbf{x} \in \mathcal{X}^n} x_i \mathbb{P}_r(\mathbf{x}) \right. \\ & \quad \times \left( - \frac{e^{-\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{Y})}}{\mathcal{Z}_{t,\epsilon}(\mathbf{G}, \mathbf{Y})} \frac{d\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{Y})}{dR} - \frac{e^{-\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{Y})}}{\mathcal{Z}_{t,\epsilon}(\mathbf{G}, \mathbf{Y})} \frac{\frac{d}{dR} \mathcal{Z}_{t,\epsilon}(\mathbf{G}, \mathbf{Y})}{\mathcal{Z}_{t,\epsilon}(\mathbf{G}, \mathbf{Y})} \right) \Big] \\ &= \frac{\lambda_n}{n} \sum_{i,j=1}^n \mathbb{E} \left[ X_i \left\langle x_i (x_j X_j + \frac{x_j Z_j}{2\sqrt{R(t, \epsilon)}} - \frac{x_j^2}{2}) \right\rangle_{t,\epsilon} \right. \\ & \quad \left. - X_i \langle x_i \rangle_{t,\epsilon} \left\langle x_j X_j + \frac{x_j Z_j}{2\sqrt{R(t, \epsilon)}} - \frac{x_j^2}{2} \right\rangle_{t,\epsilon} \right] \\ &= \frac{\lambda_n}{2n} \sum_{i,j=1}^n \mathbb{E} \left[ 2X_i X_j \langle x_i x_j \rangle_{t,\epsilon} - X_i \langle x_i x_j \rangle_{t,\epsilon} \langle x_j \rangle_{t,\epsilon} \right. \\ & \quad \left. - 2X_i X_j \langle x_i \rangle_{t,\epsilon} \langle x_j \rangle_{t,\epsilon} + 2X_i \langle x_i \rangle_{t,\epsilon} \langle x_j \rangle_{t,\epsilon}^2 - X_i \langle x_i x_j \rangle_{t,\epsilon} \langle x_j \rangle_{t,\epsilon} \right] \end{aligned} \quad (3.20)$$

To get the last identity, we used Gaussian integration by parts, which reads when applied to Gibbs brackets,

$$\mathbb{E}[Z_j \langle f \rangle_{t,\epsilon}] = \sqrt{R(t, \epsilon)} \mathbb{E}[\langle f x_j \rangle_{t,\epsilon} - \langle f \rangle_{t,\epsilon} \langle x_j \rangle_{t,\epsilon}].$$

Indeed, one must be careful that in the definition of the Gibbs bracket both the Hamiltonian *and* partition function are functions of the quenched variable  $\mathbf{Z}$ , thus the appearance of two terms when we differentiate w.r.t  $Z$ . Now, using the Nishimori identity to replace the hidden partition  $\mathbf{X}$  by a new independent sample from the posterior in (3.20) (which yields, e.g.,  $\mathbb{E}[X_i X_j \langle x_i x_j \rangle_{t,\epsilon}] = \mathbb{E}[\langle x_i x_j \rangle_{t,\epsilon}^2]$  or  $\mathbb{E}[X_i \langle x_i x_j \rangle_{t,\epsilon} \langle x_j \rangle_{t,\epsilon}] = \mathbb{E}[\langle x_i \rangle_{t,\epsilon} \langle x_i x_j \rangle_{t,\epsilon} \langle x_j \rangle_{t,\epsilon}]$ ) we reach

$$\frac{dG_n}{dR}(t, R(t, \epsilon)) = \frac{\lambda_n}{n} \sum_{i,j=1}^n \mathbb{E}[(\langle x_i x_j \rangle_{t,\epsilon} - \langle x_i \rangle_{t,\epsilon} \langle x_j \rangle_{t,\epsilon})^2]. \quad (3.21)$$

The function  $G_n$  is bounded and takes values in  $[0, \lambda_n]$ . Indeed,  $\mathbb{E}\langle Q \rangle_{t,\epsilon} = \mathbb{E}[X_1 \langle x_1 \rangle_{t,\epsilon}] = \mathbb{E}[\langle x_1 \rangle_{t,\epsilon}^2]$  by the Nishimori identity, thus  $\mathbb{E}\langle Q \rangle_{t,\epsilon} \leq \mathbb{E}\langle x_1^2 \rangle_{t,\epsilon} = \mathbb{E}[X_1^2]$  again by the Nishimori identity, and finally  $\mathbb{E}[X_1^2] = 1$ . In addition of being bounded,  $G_n$  is differentiable w.r.t. its second argument, with bounded derivative as seen from (3.21). The Cauchy-Lipschitz theorem then implies that (3.18) admits a unique global solution over  $t \in [0, 1]$ . Finally, Liouville's formula (see section 2.2.3) gives

$$\frac{dR_n^*}{d\epsilon}(t, \epsilon) = \exp \int_0^t dt' \frac{dG_n}{dR}(t', R_n^*(t', \epsilon)). \quad (3.22)$$

The non-negativity of  $dG_n/dR$  then implies  $dR_n^*/d\epsilon \geq 1$ .  $\square$

We now state a crucial concentration result for the overlap. Its validity is a consequence of the fact that the problem is analyzed in the so-called Bayesian optimal setting. This means that all hyper-parameters in the problem, namely  $(\mathbb{P}_r, r, \bar{p}_n, \Delta_n)$ , are assumed to be known, so that the posterior of the model can be written exactly. It implies the validity of the Nishimori identity which in turn allows to prove the following result (see section 3.5):

**Lemma 3.2** (Overlap concentration). *Let  $R$  be the solution  $R_n^*$  in Lemma 3.1. Then for any bounded positive sequence  $s_n$  there exists a sequence  $C_n(r, \lambda_n) > 0$  converging to a constant and such that*

$$\frac{1}{s_n} \int_{s_n}^{2s_n} d\epsilon \mathbb{E}[\langle (Q - \mathbb{E}\langle Q \rangle_{t,\epsilon})^2 \rangle_{t,\epsilon}] \leq \frac{C_n(r, \lambda_n)}{(s_n^4 n)^{1/3}}.$$

Now we average (3.15) over a small interval  $\epsilon \in [s_n, 2s_n]$  (note that  $I(\mathbf{X}; \mathbf{G})$  is independent of  $\epsilon$ ) and set  $R$  to the solution  $R_n^*$  of (3.18) in Lemma 3.1; therefore,  $q_n^*(t, \epsilon) = \lambda_n \mathbb{E}\langle Q \rangle_{t,\epsilon}$ . This choice cancels the first term of  $\mathcal{R}_2$  in the

decomposition (3.17). The second term in (3.17) is then upper bounded using Lemma 3.2. Finally  $\mathcal{R}_1 \geq 0$ . Combining all these observations we obtain

$$\frac{1}{n}I(\mathbf{X}; \mathbf{G}) \geq \frac{1}{s_n} \int_{s_n}^{2s_n} d\epsilon [i_{\text{RS}}(R_n^*(1, \epsilon), \lambda_n, r) - \mathcal{R}_3] - \frac{C_n(r, \lambda_n)\lambda_n}{4(s_n^4 n)^{1/3}} \quad (3.23)$$

where we used Fubini's theorem to switch the  $t$  and  $\epsilon$  integrals when using Lemma 3.2. Using  $q_n^* \in [0, \lambda_n]$  and  $\epsilon \in [s_n, 2s_n]$ , we see that  $\mathcal{R}_3$  is bounded uniformly in  $\epsilon$ :

$$|\mathcal{R}_3| \leq \frac{2s_n}{4\lambda_n}(2s_n + 2\lambda_n) + o_n(1) = \frac{s_n^2}{\lambda_n} + s_n + o_n(1).$$

Therefore, the average of  $\mathcal{R}_3$  over  $\epsilon$  has the same upper bound. Now, since

$$\frac{d}{d\lambda} i_{\text{RS}}(R_n^*(1, \epsilon), \lambda, r) = \frac{1}{4} - \frac{R_n^*(1, \epsilon)^2}{4\lambda}$$

and  $R_n^*(1, \epsilon) \in [s_n, 2s_n + \lambda_n]$  we have  $-\frac{1}{4} \leq \frac{d}{d\lambda} i_{\text{RS}}(R_n^*(1, \epsilon), \lambda) \leq \frac{1}{4}$  (we use  $n$  large enough for the l.h.s inequality). Therefore, by remark 2 and the mean value theorem

$$\begin{aligned} \frac{1}{s_n} \int_{s_n}^{2s_n} d\epsilon i_{\text{RS}}(R_n^*(1, \epsilon), \lambda_n, r) &= \frac{1}{s_n} \int_{s_n}^{2s_n} d\epsilon i_{\text{RS}}(R_n^*(1, \epsilon), \lambda, r) \\ &\quad + \frac{1}{s_n} \int_{s_n}^{2s_n} d\epsilon (i_{\text{RS}}(R_n^*(1, \epsilon), \lambda_n, r) - i_{\text{RS}}(R_n^*(1, \epsilon), \lambda, r)) \\ &\geq \min_{q \in [0, \lambda]} i_{\text{RS}}(q, \lambda, r) - \frac{1}{4} |\lambda_n - \lambda| \end{aligned}$$

These remarks imply a relaxation of (3.23):

$$\frac{1}{n}I(\mathbf{X}, \mathbf{G}) \geq \min_{q \in [0, \lambda]} i_{\text{RS}}(q, \lambda_n, r) - \frac{1}{4} |\lambda_n - \lambda| - \frac{C_n(r, \lambda_n)\lambda_n}{4(s_n^4 n)^{1/3}} - \frac{s_n^2}{\lambda_n} - s_n - o_n(1). \quad (3.24)$$

Finally, setting  $s_n = n^{-\theta}$  with  $\theta \in (0, 1/4)$  ensures the extra terms on the r.h.s. of (3.23) vanish as  $n \rightarrow +\infty$ . Then taking the  $\liminf_{n \rightarrow +\infty}$  and using  $\lambda_n \rightarrow \lambda$  we finally reach the desired bound.

### 3.4 The fundamental sum rule: proof of (3.15)

In this section we use the notation  $F_{t, \epsilon}$  for (3.11) without explicitly indicating the dependence in its arguments. When  $G_{ij}$  is set to zero for a specific pair  $(i, j)$  all other  $G_{k, l}$ ,  $(k, l) \neq (i, j)$  being fixed we write  $F_{t, \epsilon}(G_{ij} = 0)$ . Expectation with respect to the set of all  $G_{k, l}$ ,  $(k, l) \neq (i, j)$  is denoted by  $\mathbb{E}_{\sim G_{ij}}$ .

The derivative of the averaged free energy can be decomposed into three terms:

$$\frac{df_{t,\epsilon}}{dt} = D_1 + D_2 + D_3 \quad (3.25)$$

where

$$\begin{aligned} D_1 &\equiv \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{Y}|\mathbf{X}} \sum_{\mathbf{G}} F_{t,\epsilon} \frac{d}{dt} \mathbb{P}_t(\mathbf{G}|\mathbf{X}), \\ D_2 &\equiv \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \int d\mathbf{Y} F_{t,\epsilon} \frac{d}{dt} \mathbb{P}_t(\mathbf{Y}|\mathbf{X}), \\ D_3 &\equiv \frac{1}{n} \mathbb{E} \left\langle \frac{d}{dt} \mathcal{H}_{\text{dec};t,\epsilon} \right\rangle_{t,\epsilon} + \frac{1}{n} \mathbb{E} \left\langle \frac{d}{dt} \mathcal{H}_{\text{SBM};t} \right\rangle_{t,\epsilon}. \end{aligned}$$

### 3.4.1 Term $D_1$ .

**Lemma 3.3.** *We have*

$$D_1 = \frac{\lambda_n}{4} \mathbb{E} \langle Q^2 \rangle_{t,\epsilon} + \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}\left(\frac{\lambda_n^{3/2}}{\sqrt{n\bar{p}_n(1-\bar{p}_n)^3}}\right).$$

*Proof.* Note that by (3.7) we have

$$\begin{aligned} \frac{d}{dt} \mathbb{P}_t(\mathbf{G}|\mathbf{X}) &= \mathbb{P}_t(\mathbf{G}|\mathbf{X}) \sum_{i < j} \frac{1}{2} \frac{\Delta_n}{\sqrt{1-t}} X_i X_j \left( - \frac{G_{ij}}{\bar{p}_n + \sqrt{1-t} \Delta_n X_i X_j} \right. \\ &\quad \left. + \frac{1 - G_{ij}}{1 - \bar{p}_n - \sqrt{1-t} \Delta_n X_i X_j} \right). \end{aligned}$$

This gives

$$\begin{aligned} D_1 &= \frac{\Delta_n}{2\sqrt{1-t}} \sum_{i < j} \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{Y}|\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \left[ X_i X_j \left( \frac{(1 - G_{ij}) F_{t,\epsilon}}{1 - \bar{p}_n - \sqrt{1-t} \Delta_n X_i X_j} \right. \right. \\ &\quad \left. \left. - \frac{G_{ij} F_{t,\epsilon}}{\bar{p}_n + \sqrt{1-t} \Delta_n X_i X_j} \right) \right] \\ &= \frac{\Delta_n}{2\sqrt{1-t}} (D_1^{(a)} + D_1^{(b)}) \end{aligned} \quad (3.26)$$

with the definitions

$$\begin{aligned} D_1^{(a)} &\equiv \sum_{i < j} \mathbb{E}_{\sim G_{ij}} \left[ X_i X_j \frac{\mathbb{E}_{G_{ij}|\mathbf{X}} F_{t,\epsilon} - \mathbb{E}_{G_{ij}|\mathbf{X}} [G_{ij} F_{t,\epsilon}]}{1 - \mathbb{E}_{G_{ij}|X_i, X_j} G_{ij}} \right], \\ D_1^{(b)} &\equiv - \sum_{i < j} \mathbb{E}_{\sim G_{ij}} \left[ X_i X_j \frac{\mathbb{E}_{G_{ij}|\mathbf{X}} [G_{ij} F_{t,\epsilon}]}{\mathbb{E}_{G_{ij}|X_i, X_j} G_{ij}} \right], \end{aligned}$$

where  $\mathbb{E}_{\sim G_{ij}} \equiv \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{Y}|\mathbf{X}} \mathbb{E}_{\mathbf{G} \setminus G_{ij}|\mathbf{X}}$ , and recalling

$$\mathbb{E}_{G_{ij}|X_i, X_j} G_{ij} = \bar{p}_n + \sqrt{1-t} \Delta_n X_i X_j.$$

Both  $D_1^{(a)}$  and  $D_1^{(b)}$  involve the term  $\mathbb{E}_{G_{ij}|\mathbf{X}}[G_{ij} F_{t,\epsilon}]$ . In Section 3.6.1 we derive an *approximate integration-by-parts formula* that, when applied in the present case, yields

**Lemma 3.4.** *Fix  $i, j \in \{1, \dots, n\}^2$  and recall that  $G_{ij} \in \{0, 1\}$  with conditional mean  $\mathbb{E}_{G_{ij}|X_i, X_j}[G_{ij}] = \bar{p}_n + \sqrt{1-t} \Delta_n X_i X_j$ . Let  $F_{t,\epsilon}^{(1)}(G_{ij})$  be the first partial derivative of  $F_{t,\epsilon}$  with respect to  $G_{ij}$ . We have the approximate integration by parts formula*

$$\begin{aligned} \mathbb{E}_{G_{ij}|X_i, X_j}[G_{ij} F_{t,\epsilon}(G_{ij})] &= \mathbb{E}_{G_{ij}|X_i, X_j}[F_{t,\epsilon}^{(1)}(G_{ij})] \mathbb{E}_{G_{ij}|X_i, X_j}[G_{ij}] \\ &\quad + F_{t,\epsilon}(G_{ij} = 0) \mathbb{E}_{G_{ij}|X_i, X_j}[G_{ij}] \\ &\quad + \mathcal{O}\left(\frac{\sqrt{1-t} \lambda_n}{n^2(1-\bar{p}_n)}\right). \end{aligned} \quad (3.27)$$

where

$$F_{t,\epsilon}^{(1)}(G_{ij}) = -\frac{1}{n \bar{p}_n(1-\bar{p}_n)} \Delta_n \sqrt{1-t} \langle x_i x_j \rangle_{t,\epsilon} + \mathcal{O}\left(\frac{1}{n} \left(\frac{\Delta_n}{\bar{p}_n(1-\bar{p}_n)}\right)^2 (1-t)\right)$$

and  $F_{t,\epsilon}(G_{ij} = 0)$  is the evaluation of  $F_{t,\epsilon}$  at  $G_{ij} = 0$  all other variables  $G_{kl}$ ,  $(k, l) \neq (i, j)$  being fixed.

The approximate integration by part formula (3.27) implies that the term  $D_1^{(b)}$  of (3.26) can be written as (recall  $\bar{p}_n(1-\bar{p}_n) \gg \Delta_n$ )

$$\begin{aligned} &\frac{\Delta_n}{2\sqrt{1-t}} D_1^{(b)} \\ &= -\frac{\Delta_n}{2\sqrt{1-t}} \sum_{i < j} \mathbb{E}_{\sim G_{ij}} \left[ X_i X_j (F_{t,\epsilon}(G_{ij} = 0) - \frac{\sqrt{1-t} \Delta_n}{n \bar{p}_n(1-\bar{p}_n)} \mathbb{E}_{G_{ij}|X_i, X_j} \langle x_i x_j \rangle_{t,\epsilon}) \right] \\ &\quad + \mathcal{O}\left(\frac{\lambda_n \Delta_n}{\bar{p}_n(1-\bar{p}_n)}\right) \\ &= \frac{\Delta_n^2}{2n \bar{p}_n(1-\bar{p}_n)} \sum_{i < j} \mathbb{E}[X_i X_j \langle x_i x_j \rangle_{t,\epsilon}] - \frac{\Delta_n}{2\sqrt{1-t}} \sum_{i < j} \mathbb{E}_{\sim G_{ij}} [X_i X_j F_{t,\epsilon}(G_{ij} = 0)] \\ &\quad + \mathcal{O}\left(\frac{\lambda_n \Delta_n}{\bar{p}_n(1-\bar{p}_n)}\right). \end{aligned} \quad (3.28)$$



Applying again the approximate integration by parts formula (3.27) the term  $D_1^{(a)}$  of (3.26) can be written as (recall  $(1 - \bar{p}_n)^2 \gg \Delta_n$ )

$$\begin{aligned}
& \frac{\Delta_n}{2\sqrt{1-t}} D_1^{(a)} \\
&= -\frac{\Delta_n}{2\sqrt{1-t}} \sum_{i < j} \mathbb{E}_{\sim G_{ij}} \left[ X_i X_j \frac{\mathbb{E}_{G_{ij}|X_i, X_j} G_{ij}}{1 - \mathbb{E}_{G_{ij}|X_i, X_j} G_{ij}} (F_{t,\epsilon}(G_{ij} = 0) \right. \\
&\quad \left. - \frac{\sqrt{1-t}\Delta_n}{n\bar{p}_n(1-\bar{p}_n)} \mathbb{E}_{G_{ij}|X_i, X_j} \langle x_i x_j \rangle_{t,\epsilon} \right] \\
&\quad + \frac{\Delta_n}{2\sqrt{1-t}} \sum_{i < j} \mathbb{E}_{\sim G_{ij}} \left[ X_i X_j \frac{\mathbb{E}_{G_{ij}|X_i, X_j} F_{t,\epsilon}}{1 - \mathbb{E}_{G_{ij}|X_i, X_j} G_{ij}} \right] + \mathcal{O}\left(\frac{\lambda_n \Delta_n}{(1 - \bar{p}_n)^2}\right) \\
&= E_1 + E_2 + \frac{\Delta_n}{2\sqrt{1-t}} \sum_{i < j} \mathbb{E}_{\sim G_{ij}} [X_i X_j F_{t,\epsilon}(G_{ij} = 0)] + \mathcal{O}\left(\frac{\lambda_n \Delta_n}{(1 - \bar{p}_n)^2}\right) \quad (3.29)
\end{aligned}$$

where we define

$$\begin{aligned}
E_1 &\equiv \frac{\Delta_n}{2\sqrt{1-t}} \sum_{i < j} \mathbb{E}_{\sim G_{ij}} \left[ X_i X_j \frac{\mathbb{E}_{G_{ij}|X_i, X_j} F_{t,\epsilon} - F_{t,\epsilon}(G_{ij} = 0)}{1 - \mathbb{E}_{G_{ij}|X_i, X_j} G_{ij}} \right], \\
E_2 &\equiv \frac{\Delta_n^2}{2n\bar{p}_n(1-\bar{p}_n)} \sum_{i < j} \mathbb{E} \left[ \frac{\mathbb{E}_{G_{ij}|X_i, X_j} G_{ij}}{1 - \mathbb{E}_{G_{ij}|X_i, X_j} G_{ij}} X_i X_j \langle x_i x_j \rangle_{t,\epsilon} \right].
\end{aligned}$$

We show in Appendix 3.6.3 that in (3.29) the terms  $E_1$  and  $E_2$  approximately cancel so that

$$\frac{\Delta_n}{2\sqrt{1-t}} D_1^{(a)} = \frac{\Delta_n}{2\sqrt{1-t}} \sum_{i < j} \mathbb{E}_{\sim G_{ij}} [X_i X_j F_{t,\epsilon}(G_{ij} = 0)] + \mathcal{O}\left(\frac{\lambda_n \Delta_n}{(1 - \bar{p}_n)^2}\right). \quad (3.30)$$

Finally, substituting (3.28) and (3.30) into (3.26) gives

$$\begin{aligned}
& \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{Y}|\mathbf{X}} \sum_{\mathbf{G}} F_{t,\epsilon} \frac{d}{dt} \mathbb{P}_t(\mathbf{G}|\mathbf{X}) \\
&= \frac{\Delta_n^2}{2n\bar{p}_n(1-\bar{p}_n)} \sum_{i < j} \mathbb{E}[X_i X_j \langle x_i x_j \rangle_{t,\epsilon}] + \mathcal{O}\left(\frac{\lambda_n \Delta_n}{\bar{p}_n(1-\bar{p}_n)}\right) + \mathcal{O}\left(\frac{\lambda_n \Delta_n}{(1-\bar{p}_n)^2}\right) \\
&= \frac{\lambda_n}{4} \mathbb{E}\langle Q^2 \rangle_{t,\epsilon} + \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}\left(\frac{\lambda_n \Delta_n}{\bar{p}_n(1-\bar{p}_n)}\right) + \mathcal{O}\left(\frac{\lambda_n \Delta_n}{(1-\bar{p}_n)^2}\right) \\
&= \frac{\lambda_n}{4} \mathbb{E}\langle Q^2 \rangle_{t,\epsilon} + \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}\left(\frac{\lambda_n^{3/2}}{\sqrt{n\bar{p}_n(1-\bar{p}_n)^3}}\right),
\end{aligned}$$

where, in the last two equalities, we used  $\lambda_n = n\Delta_n^2/(\bar{p}_n(1-\bar{p}_n))$  and  $Q = \frac{1}{n} \sum_{i=1}^n X_i x_i$ . With (h1) and (h2), all the error terms represented by the big-O notations tend to zero.  $\square$

### 3.4.2 Term $D_2$ .

**Lemma 3.5.** *We have*

$$D_2 = -\frac{1}{2}q(t, \epsilon)\mathbb{E}\langle Q \rangle_{t, \epsilon}.$$

*Proof.* Recall (3.8). Using Gaussian integration by parts (2.20) we obtain

$$\begin{aligned} D_2 &\equiv \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \int d\mathbf{Y} F_{t, \epsilon} \frac{d}{dt} \mathbb{P}_t(\mathbf{Y}|\mathbf{X}) \\ &= \sum_{i=1}^n \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Y}|\mathbf{X}} \left[ (Y_i - \sqrt{R(t, \epsilon)} X_i) \frac{q(t, \epsilon) X_i}{2\sqrt{R(t, \epsilon)}} F_{t, \epsilon} \right] \\ &= \frac{q(t, \epsilon)}{2\sqrt{R(t, \epsilon)}} \sum_{i=1}^n \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} [Z_i X_i F_{t, \epsilon}] \\ &= -\frac{q(t, \epsilon)}{2n\sqrt{R(t, \epsilon)}} \sum_{i=1}^n \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} [X_i \langle \sqrt{R(t, \epsilon)} x_i \rangle_{t, \epsilon}] \\ &= -\frac{1}{2}q(t, \epsilon)\mathbb{E}\langle Q \rangle_{t, \epsilon}, \end{aligned}$$

where we used that  $\frac{dF_{t, \epsilon}}{dZ} = -\frac{1}{n}\langle \sqrt{R(t, \epsilon)} x_i \rangle_{t, \epsilon}$ , and then the definition of the overlap.  $\square$

### 3.4.3 Term $D_3$ .

**Lemma 3.6.** *We have  $D_3 = 0$ .*

*Proof.* Using the Nishimori identity (3.16) we obtain

$$\begin{aligned} \mathbb{E} \left\langle \frac{d}{dt} \mathcal{H}_{\text{dec}; t, \epsilon} \right\rangle_{t, \epsilon} &= -q(t, \epsilon) \sum_{i=1}^n \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Y}|\mathbf{X}} \left\langle \frac{Y_i x_i}{2\sqrt{R(t, \epsilon)}} - \frac{x_i^2}{2} \right\rangle_{t, \epsilon} \\ &= -q(t, \epsilon) \sum_{i=1}^n \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Y}|\mathbf{X}} \left[ \frac{Y_i X_i}{2\sqrt{R(t, \epsilon)}} - \frac{X_i^2}{2} \right] \\ &= -q(t, \epsilon) \sum_{i=1}^n \mathbb{E}_{X_i} \mathbb{E}_{Z_i} \frac{Z_i X_i}{2\sqrt{R(t, \epsilon)}} \\ &= 0 \end{aligned}$$

by independence of the centered noise  $\mathbf{Z}$  and the hidden partition  $\mathbf{X}$ .

Again the Nishimori identity (3.16) is used to obtain

$$\begin{aligned}
& \mathbb{E} \left\langle \frac{d}{dt} \mathcal{H}_{\text{SBM},t} \right\rangle_{t,\epsilon} \\
&= \frac{1}{2\sqrt{1-t}} \sum_{i < j} \mathbb{E} \left\langle \Delta_n x_i x_j \left( \frac{G_{ij}}{\bar{p}_n + \sqrt{1-t}\Delta_n x_i x_j} - \frac{1 - G_{ij}}{1 - \bar{p}_n - \sqrt{1-t}\Delta_n x_i x_j} \right) \right\rangle_{t,\epsilon} \\
&= \frac{1}{2\sqrt{1-t}} \sum_{i < j} \mathbb{E} \left[ \Delta_n X_i X_j \left( \frac{G_{ij}}{\bar{p}_n + \sqrt{1-t}\Delta_n X_i X_j} - \frac{1 - G_{ij}}{1 - \bar{p}_n - \sqrt{1-t}\Delta_n X_i X_j} \right) \right] \\
&= \frac{1}{2\sqrt{1-t}} \sum_{i < j} \mathbb{E}_{X_i, X_j} \left[ \Delta_n X_i X_j \left( \frac{\mathbb{E}_{G_{ij}|X_i, X_j} G_{ij}}{\bar{p}_n + \sqrt{1-t}\Delta_n X_i X_j} \right. \right. \\
&\quad \left. \left. - \frac{1 - \mathbb{E}_{G_{ij}|X_i, X_j} G_{ij}}{1 - \bar{p}_n - \sqrt{1-t}\Delta_n X_i X_j} \right) \right] \\
&= 0,
\end{aligned}$$

where the last line follows from  $\mathbb{E}_{G_{ij}|X_i, X_j} G_{ij} = \bar{p}_n + \sqrt{1-t}\Delta_n X_i X_j$ .  $\square$

### 3.4.4 Final derivations of the sum rule.

The last missing term in order to simplify the sum rule (3.14) is:

**Lemma 3.7.** *We have*

$$f_{0,0} - f_{0,\epsilon} = \frac{1}{2} \int_0^\epsilon d\epsilon' \mathbb{E} \langle Q \rangle_{0,\epsilon'}.$$

*Proof.* Using Gaussian integration by parts (2.20) and from (3.16) the specific Nishimori identity  $\mathbb{E}[\langle x_i \rangle_{0,\epsilon'}^2] = \mathbb{E}[X_i \langle x_i \rangle_{0,\epsilon'}]$  we have (recall also that  $R(0, \epsilon') = \epsilon'$ )

$$\begin{aligned}
f_{0,0} - f_{0,\epsilon} &= - \int_0^\epsilon d\epsilon' \frac{df_{0,\epsilon'}}{d\epsilon'} = - \int_0^\epsilon d\epsilon' \left\langle \frac{d}{d\epsilon'} \mathcal{H}_{\text{dec};t,\epsilon'} \right\rangle_{0,\epsilon'} \\
&= \int_0^\epsilon d\epsilon' \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left\langle X_i x_i - \frac{x_i^2}{2} + \frac{1}{2\sqrt{\epsilon'}} Z_i x_i \right\rangle_{0,\epsilon'} \\
&= \int_0^\epsilon d\epsilon' \frac{1}{n} \sum_{i=1}^n \left( \mathbb{E} \langle X_i x_i \rangle_{0,\epsilon'} - \frac{1}{2} \mathbb{E}[\langle x_i \rangle_{0,\epsilon'}^2] \right) \\
&= \frac{1}{2} \int_0^\epsilon d\epsilon' \mathbb{E} \langle Q \rangle_{0,\epsilon'}.
\end{aligned}$$

$\square$

Recall  $R(1, \epsilon) = \epsilon + \int_0^1 q(t, \epsilon) dt$ . Substituting (3.25), and Lemmas 3.3, 3.5 and 3.6 as well as 3.7 into (3.14) yields

$$\begin{aligned}
\frac{1}{n} I(\mathbf{X}; \mathbf{G}) &= i_{\text{RS}}(R(1, \epsilon), \lambda_n, r) - \frac{(\epsilon + \int_0^1 q(t, \epsilon) dt)^2}{4\lambda_n} + \frac{1}{2} \int_0^\epsilon d\epsilon' \mathbb{E}\langle Q \rangle_{0, \epsilon'} \\
&\quad - \int_0^1 dt \left( \frac{\lambda_n}{4} \mathbb{E}\langle Q^2 \rangle_{t, \epsilon} - \frac{1}{2} q(t, \epsilon) \mathbb{E}\langle Q \rangle_{t, \epsilon} \right) + o_n(1) \\
&= i_{\text{RS}}(R(1, \epsilon), \lambda_n, r) + \frac{1}{4\lambda_n} \left( \int_0^1 q(t, \epsilon)^2 dt - \left( \int_0^1 q(t, \epsilon) dt \right)^2 \right) \\
&\quad - \frac{1}{4\lambda_n} \int_0^1 dt \mathbb{E}\langle (\lambda_n Q - q(t, \epsilon))^2 \rangle_{t, \epsilon} \\
&\quad - \frac{\epsilon}{4\lambda_n} \left( \epsilon + 2 \int_0^1 q(t, \epsilon) dt \right) + \frac{1}{2} \int_0^\epsilon d\epsilon' \mathbb{E}\langle Q \rangle_{0, \epsilon'} + o_n(1)
\end{aligned}$$

which is the sum rule (3.15).

### 3.5 Concentration of overlap: proof of Lemma 3.2

Concentration of overlap has been shown for various Bayesian inference problems, see, e.g., [62, 55, 56]. These proofs can be adapted to the present case. The idea is to bound the fluctuations of the overlap by those of another, easier to control, object  $\mathcal{L}$  defined below. This object is more natural to work with as it is directly related to derivatives of the free energy. Let us present the main steps of the proof, and then provide the proof details afterwards.

Let

$$\mathcal{L} \equiv \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i^2}{2} - x_i X_i - \frac{x_i Z_i}{2\sqrt{R(t, \epsilon)}} \right). \quad (3.31)$$

As said previously, we can relate the fluctuations of the overlap to those of  $\mathcal{L}$ :

**Lemma 3.8** (A fluctuation identity). *We have*

$$\mathbb{E}\langle (Q - \mathbb{E}\langle Q \rangle_{t, \epsilon})^2 \rangle_{t, \epsilon} \leq 4 \mathbb{E}\langle (\mathcal{L} - \mathbb{E}\langle \mathcal{L} \rangle_{t, \epsilon})^2 \rangle_{t, \epsilon}.$$

Therefore, it remains to show the concentration of  $\mathcal{L}$ . We divide the task into two parts:

$$\mathbb{E}\langle (\mathcal{L} - \mathbb{E}\langle \mathcal{L} \rangle_{t, \epsilon})^2 \rangle_{t, \epsilon} = \mathbb{E}\langle (\mathcal{L} - \langle \mathcal{L} \rangle_{t, \epsilon})^2 \rangle_{t, \epsilon} + \mathbb{E}[(\langle \mathcal{L} \rangle_{t, \epsilon} - \mathbb{E}\langle \mathcal{L} \rangle_{t, \epsilon})^2]. \quad (3.32)$$

These two terms are controlled by the following lemmas:

**Lemma 3.9** (Thermal fluctuations). *Let  $R(t, \epsilon) = \epsilon + \int_0^t ds q(s, \epsilon) \geq \epsilon$  be such that  $dR/d\epsilon \geq 1$ . We then have*

$$\int_{s_n}^{2s_n} d\epsilon \mathbb{E} \langle (\mathcal{L} - \langle \mathcal{L} \rangle_{t,\epsilon})^2 \rangle_{t,\epsilon} \leq \frac{1}{n}.$$

**Lemma 3.10** (Quenched fluctuations). *Let  $R(t, \epsilon) = \epsilon + \int_0^t ds q(s, \epsilon)$ , with  $\epsilon \in [s_n, 2s_n]$  and  $q$  taking values in  $[0, \lambda_n]$ , be such that  $dR/d\epsilon \geq 1$ . There exists a sequence  $C_n(r, \lambda_n) > 0$  converging to a constant such that*

$$\int_{s_n}^{2s_n} d\epsilon \mathbb{E} [(\langle \mathcal{L} \rangle_{t,\epsilon} - \mathbb{E} \langle \mathcal{L} \rangle_{t,\epsilon})^2] \leq \frac{C_n(r, \lambda_n)}{(s_n n)^{1/3}}. \quad (3.33)$$

The proof of Lemma 3.9 and Lemma 3.10 employ some useful identities for the derivatives of the free energy (recall  $F_{t,\epsilon} \equiv -\frac{1}{n} \ln \mathcal{Z}_{t,\epsilon}(\mathbf{G}, \mathbf{Y})$ ):

$$\frac{dF_{t,\epsilon}}{dR} = \langle \mathcal{L} \rangle_{t,\epsilon}, \quad (3.34)$$

$$\frac{1}{n} \frac{d^2 F_{t,\epsilon}}{dR^2} = -(\langle \mathcal{L}^2 \rangle_{t,\epsilon} - \langle \mathcal{L} \rangle_{t,\epsilon}^2) + \frac{1}{4n^2 R^{3/2}} \sum_{i=1}^n \langle x_i \rangle_{t,\epsilon} Z_i, \quad (3.35)$$

where we simply denote, when no confusion can arise,  $R = R(t, \epsilon)$ . Taking expectation on both sides of (3.34) and (3.35) we have

$$\frac{df_{t,\epsilon}}{dR} = \mathbb{E} \langle \mathcal{L} \rangle_{t,\epsilon} = -\frac{1}{2n} \sum_{i=1}^n \mathbb{E} [\langle x_i \rangle_{t,\epsilon}^2], \quad (3.36)$$

$$\frac{1}{n} \frac{d^2 f_{t,\epsilon}}{dR^2} = -\mathbb{E} [\langle \mathcal{L}^2 \rangle_{t,\epsilon} - \langle \mathcal{L} \rangle_{t,\epsilon}^2] + \frac{1}{4n^2 R} \sum_{i=1}^n \mathbb{E} [\langle x_i^2 \rangle_{t,\epsilon} - \langle x_i \rangle_{t,\epsilon}^2] \quad (3.37)$$

$$= -\frac{1}{2n^2} \sum_{i,j=1}^n \mathbb{E} [(\langle x_i x_j \rangle_{t,\epsilon} - \langle x_i \rangle_{t,\epsilon} \langle x_j \rangle_{t,\epsilon})^2]. \quad (3.38)$$

The proof of Lemma 3.2 is ended by applying Lemmas 3.8, 3.9 and 3.10 in conjunction with (3.32):

$$\frac{1}{s_n} \int_{s_n}^{2s_n} d\epsilon \mathbb{E} \langle (Q - \mathbb{E} \langle Q \rangle_{t,\epsilon})^2 \rangle_{t,\epsilon} \leq \frac{4}{s_n n} + \frac{4C_n(r, \lambda_n)}{(s_n^4 n)^{1/3}}.$$

We now provide the proofs of Lemmas 3.8 to 3.11. For the sake of readability, we simply denote  $\langle - \rangle \equiv \langle - \rangle_{t,\epsilon}$  for the rest of this section.

### 3.5.1 Proof of the fluctuation identity: Lemma 3.8

We start by proving

$$-2 \mathbb{E} \langle Q(\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle = \mathbb{E} \langle (Q - \mathbb{E} \langle Q \rangle)^2 \rangle + \mathbb{E} \langle (Q - \langle Q \rangle)^2 \rangle. \quad (3.39)$$

Using the definitions  $Q \equiv \frac{1}{n} \sum_{i=1}^n x_i X_i$  and (3.31) gives

$$\begin{aligned} 2 \mathbb{E} \langle Q(\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle &= \frac{1}{n^2} \sum_{i,j=1}^n \left\{ \mathbb{E} \left[ X_i \langle x_i x_j^2 \rangle - 2X_i X_j \langle x_i x_j \rangle - \frac{Z_j}{\sqrt{R}} X_i \langle x_i x_j \rangle \right] \right. \\ &\quad \left. - \mathbb{E}[X_i \langle x_i \rangle] \mathbb{E} \left[ \langle x_j^2 \rangle - 2X_j \langle x_j \rangle - \frac{Z_j}{\sqrt{R}} \langle x_j \rangle \right] \right\}. \end{aligned} \quad (3.40)$$

Gaussian integration by parts then yields

$$\begin{aligned} \mathbb{E} \left[ \frac{Z_j}{\sqrt{R}} X_i \langle x_i x_j \rangle \right] &= \mathbb{E} [X_i \langle x_i x_j^2 \rangle - X_i \langle x_i x_j \rangle \langle x_j \rangle], \quad \text{and} \\ \mathbb{E} \left[ \frac{Z_j}{\sqrt{R}} \langle x_j \rangle \right] &= \mathbb{E} [\langle x_j^2 \rangle - \langle x_j \rangle^2]. \end{aligned}$$

These two formulas simplify (3.40) to

$$\begin{aligned} 2 \mathbb{E} \langle Q(\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle &= \frac{1}{n^2} \sum_{i,j=1}^n \left\{ \mathbb{E} [X_i \langle x_j \rangle \langle x_i x_j \rangle - 2X_i X_j \langle x_i x_j \rangle] \right. \\ &\quad \left. - \mathbb{E}[X_i \langle x_i \rangle] \mathbb{E} [\langle x_j \rangle^2 - 2X_j \langle x_j \rangle] \right\}. \end{aligned} \quad (3.41)$$

The Nishimori identity implies

$$\begin{aligned} \mathbb{E} [\langle x_j \rangle^2] &= \mathbb{E} [X_j \langle x_j \rangle], \quad \text{and} \\ \mathbb{E} [X_i \langle x_j \rangle \langle x_i x_j \rangle] &= \mathbb{E} [\langle x_i \rangle \langle x_j \rangle \langle x_i x_j \rangle] = \mathbb{E} [\langle x_i \rangle \langle x_j \rangle X_i X_j]. \end{aligned}$$

These formulas further simplify (3.41) to

$$\begin{aligned} 2 \mathbb{E} \langle Q(\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle &= \frac{1}{n^2} \sum_{i,j=1}^n \left\{ \mathbb{E} [\langle x_i \rangle \langle x_j \rangle X_i X_j - 2X_i X_j \langle x_i x_j \rangle] + \mathbb{E} [X_i \langle x_i \rangle] \mathbb{E} [X_j \langle x_j \rangle] \right\} \\ &= \mathbb{E} [\langle Q \rangle^2] - 2 \mathbb{E} \langle Q^2 \rangle + \mathbb{E} [\langle Q \rangle]^2 \\ &= -(\mathbb{E} \langle Q^2 \rangle - \mathbb{E} [\langle Q \rangle]^2) - (\mathbb{E} \langle Q^2 \rangle - \mathbb{E} [\langle Q \rangle]^2) \end{aligned}$$

which is (3.39).

Identity (3.39) implies

$$2 |\mathbb{E} \langle Q(\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle| = 2 |\mathbb{E} \langle (Q - \mathbb{E} \langle Q \rangle)(\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle) \rangle| \geq \mathbb{E} \langle (Q - \mathbb{E} \langle Q \rangle)^2 \rangle$$

and application of the Cauchy-Schwarz inequality then gives

$$2 \{ \mathbb{E} \langle (Q - \mathbb{E} \langle Q \rangle)^2 \rangle \mathbb{E} \langle (\mathcal{L} - \mathbb{E} \langle \mathcal{L} \rangle)^2 \rangle \}^{1/2} \geq \mathbb{E} \langle (Q - \mathbb{E} \langle Q \rangle)^2 \rangle.$$

This ends the proof of Lemma 3.8.

### 3.5.2 Bound on thermal fluctuations: proof of Lemma 3.9

First note that  $\frac{d^2 f_{t,\epsilon}}{dR^2} \leq 0$ . Then, using (3.37),  $dR/d\epsilon \geq 1$ ,  $R(t, \epsilon) \geq \epsilon$ , and the Nishimori identity  $\mathbb{E}\langle x_i^2 \rangle = \mathbb{E}[X_i^2] = 1$ ,

$$\begin{aligned} \mathbb{E}(\langle \mathcal{L} - \langle \mathcal{L} \rangle)^2) &= -\frac{1}{n} \frac{d^2 f_{t,\epsilon}}{dR^2} + \frac{1}{4n^2 R} \sum_{i=1}^n \mathbb{E}[\langle x_i^2 \rangle - \langle x_i \rangle^2] \\ &\leq -\frac{1}{n} \frac{dR}{d\epsilon} \frac{d^2 f_{t,\epsilon}}{dR^2} + \frac{1}{4n\epsilon} = -\frac{1}{n} \frac{d}{d\epsilon} \left( \frac{df_{t,\epsilon}}{dR} \right) + \frac{1}{4n\epsilon}, \end{aligned}$$

From (3.36)  $df_{t,\epsilon}/dR \in [-1/2, 0]$ , therefore  $[df_{t,\epsilon}/dR]_{\epsilon=s_n}^{\epsilon=2s_n} \geq -1/2$ . Integrating over  $\epsilon$  then gives

$$\begin{aligned} \int_{s_n}^{2s_n} d\epsilon \mathbb{E}(\langle \mathcal{L} - \langle \mathcal{L} \rangle)^2) &\leq \int_{s_n}^{2s_n} d\epsilon \left\{ -\frac{1}{n} \frac{d}{d\epsilon} \left( \frac{df_{t,\epsilon}}{dR} \right) + \frac{1}{4n\epsilon} \right\} \\ &= -\frac{1}{n} \left[ \frac{df_{t,\epsilon}}{dR} \right]_{\epsilon=s_n}^{\epsilon=2s_n} + \frac{\ln 2}{4n} \\ &\leq \frac{2 + (\ln 2)}{4n} \leq \frac{1}{n}. \end{aligned}$$

### 3.5.3 Bound on quenched fluctuations: proof of Lemma 3.10

Lemma 3.10 is based on the concentration of the free energy, a very general fact in “well-behaved” statistical mechanics models. The proof of the following lemma uses more or less standard methods and can be found in Appendix 3.6.4.

**Lemma 3.11** (Free energy fluctuations). *There exists a sequence  $C_n(r, \lambda_n) > 0$  converging to a constant when  $n \rightarrow +\infty$ , such that*

$$\text{Var}(F_{t,\epsilon}) = \mathbb{E}[(F_{t,\epsilon} - f_{t,\epsilon})^2] \leq \frac{C_n(r, \lambda_n)}{n}. \quad (3.42)$$

Recall  $R = R(t, \epsilon)$ . Let

$$\tilde{F}_{t,\epsilon}(R) \equiv F_{t,\epsilon} + \sqrt{R \frac{1-r}{r}} \frac{1}{n} \sum_{i=1}^n |Z_i|, \quad \tilde{f}_{t,\epsilon}(R) \equiv f_{t,\epsilon} + \sqrt{R \frac{1-r}{r}} \frac{1}{n} \sum_{i=1}^n \mathbb{E}|Z_i|. \quad (3.43)$$

From (3.38) we see that  $\tilde{f}_{t,\epsilon}(R)$  is concave in  $R$ . Furthermore, from (3.35) and  $|x_i| \leq \sqrt{\frac{1-r}{r}}$  for  $0 \leq r \leq 1/2$ , we see that  $\tilde{F}_{t,\epsilon}(R)$  is also concave in  $R$ . Hence, we can employ the following lemma (see the end of this section for a proof):

**Lemma 3.12** (A bound on the difference of derivatives due to concavity). *Let  $G(x)$  and  $g(x)$  be concave functions. Let  $\delta > 0$  and define  $C_\delta^+(x) \equiv g'(x) - g'(x + \delta) \geq 0$  and  $C_\delta^-(x) \equiv g'(x - \delta) - g'(x) \geq 0$ . Then*

$$|G'(x) - g'(x)| \leq \delta^{-1} \sum_{u \in \{x-\delta, x, x+\delta\}} |G(u) - g(u)| + C_\delta^+(x) + C_\delta^-(x).$$

From (3.43) we have

$$\tilde{F}_{t,\epsilon} - \tilde{f}_{t,\epsilon} = F_{t,\epsilon} - f_{t,\epsilon} + \sqrt{R \frac{1-r}{r}} A_n, \quad A_n \equiv \frac{1}{n} \sum_{i=1}^n (|Z_i| - \mathbb{E}|Z_i|),$$

and from (3.34) and (3.36) we have

$$\frac{d\tilde{F}_{t,\epsilon}}{dR} - \frac{d\tilde{f}_{t,\epsilon}}{dR} = \langle \mathcal{L} \rangle - \mathbb{E}\langle \mathcal{L} \rangle + \frac{1}{2} \sqrt{\frac{1-r}{Rr}} A_n.$$

Using Lemma 3.12, we then get

$$\begin{aligned} |\langle \mathcal{L} \rangle - \mathbb{E}\langle \mathcal{L} \rangle| &\leq \delta^{-1} \sum_{u \in \{R-\delta, R, R+\delta\}} (|F_{t,\epsilon}(R=u) - f_{t,\epsilon}(R=u)| + \sqrt{u \frac{1-r}{r}} |A_n|) \\ &\quad + C_\delta^+(R) + C_\delta^-(R) + \frac{1}{2} \sqrt{\frac{1-r}{Rr}} A_n, \end{aligned}$$

where  $C_\delta^+(R) \equiv \tilde{f}'_{t,\epsilon}(R) - \tilde{f}'_{t,\epsilon}(R + \delta) \geq 0$  and  $C_\delta^-(R) \equiv \tilde{f}'_{t,\epsilon}(R - \delta) - \tilde{f}'_{t,\epsilon}(R) \geq 0$ . Then squaring this inequality, using  $(\sum_{i=1}^p v_i)^2 \leq p \sum_{i=1}^p v_i^2$ , taking the expectation, and recalling that  $R = R(t, \epsilon) \geq \epsilon$  we reach

$$\begin{aligned} \frac{1}{9} \mathbb{E}[(\langle \mathcal{L} \rangle - \mathbb{E}\langle \mathcal{L} \rangle)^2] &\leq \delta^{-2} \sum_{u \in \{R-\delta, R, R+\delta\}} \left\{ \mathbb{E}[(F_{t,\epsilon}(u) - f_{t,\epsilon}(u))^2] + u \frac{1-r}{r} \mathbb{E}[A_n^2] \right\} \\ &\quad + C_\delta^+(R)^2 + C_\delta^-(R)^2 + \frac{1-r}{4\epsilon r} \mathbb{E}[A_n^2]. \end{aligned} \quad (3.44)$$

Note that  $\mathbb{E}[A_n^2] = a/n$  with  $a = 1 - 2/\pi$ . Recall  $q^*(t, \epsilon) \in [0, \lambda_n]$  from Lemma 3.1. We can upper bound  $u$  by  $\lambda_n + 2s_n + \delta$ . These remarks with Lemma 3.11 simplify (3.44) to

$$\begin{aligned} \frac{1}{9} \mathbb{E}[(\langle \mathcal{L} \rangle - \mathbb{E}\langle \mathcal{L} \rangle)^2] &\leq \frac{3}{n\delta^2} \left( C_n(r, \lambda_n) + a(\lambda_n + 2s_n + \delta) \frac{1-r}{r} \right) \\ &\quad + C_\delta^+(R)^2 + C_\delta^-(R)^2 + \frac{1}{4\epsilon} \frac{1-r}{r} \frac{a}{n}. \end{aligned} \quad (3.45)$$

Recall (3.36) and that  $\mathbb{E}[\langle x_i \rangle^2] \leq \mathbb{E}[x_i^2] = \mathbb{E}[X_i^2] = 1$ . We have

$$|\tilde{f}'_{t,\epsilon}(R)| \leq \frac{1}{2} \left( 1 + \sqrt{\frac{1-r}{rR}} \right)$$



and therefore  $0 \leq C_\delta^\pm(R) \leq 1 + \sqrt{\frac{1-r}{r(R-\delta)}}$ . Using  $dR/d\epsilon \geq 1$  and  $R \geq s_n$ , we then have

$$\begin{aligned}
& \int_{s_n}^{2s_n} d\epsilon \{C_\delta^+(R)^2 + C_\delta^-(R)^2\} \\
& \leq 2 \left(1 + \sqrt{\frac{1-r}{r(s_n-\delta)}}\right) \int_{s_n}^{2s_n} d\epsilon \{C_\delta^+(R) + C_\delta^-(R)\} \\
& = 2 \left(1 + \sqrt{\frac{1-r}{r(s_n-\delta)}}\right) \int_{s_n}^{2s_n} d\epsilon \left( \frac{d\tilde{f}_{t,\epsilon}(R-\delta)}{dR} - \frac{d\tilde{f}_{t,\epsilon}(R+\delta)}{dR} \right) \\
& \leq 2 \left(1 + \sqrt{\frac{1-r}{r(s_n-\delta)}}\right) \int_{s_n}^{2s_n} d\epsilon \frac{dR}{d\epsilon} \left( \frac{d\tilde{f}_{t,\epsilon}(R-\delta)}{dR} - \frac{d\tilde{f}_{t,\epsilon}(R+\delta)}{dR} \right) \\
& = 2 \left(1 + \sqrt{\frac{1-r}{r(s_n-\delta)}}\right) \int_{s_n}^{2s_n} d\epsilon \left( \frac{d\tilde{f}_{t,\epsilon}(R(t,\epsilon)-\delta)}{d\epsilon} - \frac{d\tilde{f}_{t,\epsilon}(R(t,\epsilon)+\delta)}{d\epsilon} \right) \\
& = 2 \left(1 + \sqrt{\frac{1-r}{r(s_n-\delta)}}\right) \{ (\tilde{f}_{t,2s_n}(R(t,2s_n)-\delta) - \tilde{f}_{t,2s_n}(R(t,2s_n)+\delta)) \\
& \quad + (\tilde{f}_{t,s_n}(R(t,s_n)+\delta) - \tilde{f}_{t,s_n}(R(t,s_n)-\delta)) \} \\
& \leq 4\delta \left(1 + \sqrt{\frac{1-r}{r(s_n-\delta)}}\right)^2
\end{aligned}$$

using the mean value theorem for the last step. Therefore, upon integrating (3.45) over  $\epsilon \in (s_n, 2s_n)$  we have

$$\begin{aligned}
\frac{1}{9} \int_{s_n}^{2s_n} d\epsilon \mathbb{E}[(\langle \mathcal{L} \rangle - \mathbb{E}(\langle \mathcal{L} \rangle))^2] & \leq \frac{3s_n}{n\delta^2} (C_n(r, \lambda_n) + a(\lambda_n + 2s_n + \delta) \frac{1-r}{r}) \\
& \quad + 4\delta \left(1 + \sqrt{\frac{1-r}{r(s_n-\delta)}}\right)^2 + \frac{a(1-r) \ln 2}{4rn}. \quad (3.46)
\end{aligned}$$

The bound is optimized by choosing  $\delta = (s_n^2/n)^{1/3}$ . This ends the proof.

*Proof of Lemma 3.12.* Concavity implies that for any  $\delta > 0$  we have

$$\begin{aligned}
G'(x) - g'(x) & \geq \frac{G(x+\delta) - G(x)}{\delta} - g'(x) \\
& \geq \frac{G(x+\delta) - G(x)}{\delta} - g'(x) + g'(x+\delta) - \frac{g(x+\delta) - g(x)}{\delta} \\
& = \frac{G(x+\delta) - g(x+\delta)}{\delta} - \frac{G(x) - g(x)}{\delta} - C_\delta^+(x), \\
G'(x) - g'(x) & \leq \frac{G(x) - G(x-\delta)}{\delta} - g'(x) + g'(x-\delta) - \frac{g(x) - g(x-\delta)}{\delta} \\
& = \frac{G(x) - g(x)}{\delta} - \frac{G(x-\delta) - g(x-\delta)}{\delta} + C_\delta^-(x).
\end{aligned}$$

Combining these two inequalities ends the proof.  $\square$

## 3.6 Appendix

### 3.6.1 Approximate integration by parts: proof of lemma 3.4

The following general formula follows from Taylor expansion with Lagrange remainder. When the r.h.s is small in specific applications, the formula can be seen as an approximate integration-by-parts formula that generalizes Gaussian integration by parts.

**Lemma 3.13.** *Let  $g(U)$  be a  $\mathcal{C}^4$  function of a random variable  $U$  such that for  $k = 1, 2, 3, 4$  we have  $\sup_U |g^{(k)}(U)| \leq C_k$  for some constants  $C_k \geq 0$  and  $g^{(k)}(U) \equiv d^k g(U)/dU^k$ . Suppose that the first four moments of  $U$  are finite. Then*

$$\begin{aligned} & \left| \mathbb{E}[Ug(U)] - \mathbb{E}[g'(U)]\mathbb{E}[U^2] - g(0)\mathbb{E}U \right| \\ & \leq C_2 \left( \frac{|\mathbb{E}[U^3]|}{2} + \mathbb{E}[U^2]\mathbb{E}U \right) + C_3 \left( \frac{\mathbb{E}[U^4]}{24} + \frac{\mathbb{E}[U^2]^2}{2} \right) + \frac{C_4}{6} |\mathbb{E}[U^3]| \mathbb{E}[U^2]. \end{aligned} \quad (3.47)$$

*Proof.* By Taylor's theorem, any  $\mathcal{C}^4$  function  $h(U)$  can be written as

$$h(U) = h(0) + h^{(1)}(0)U + \frac{1}{2}h^{(2)}(0)U^2 + \frac{1}{2} \int_0^U h^{(3)}(s)(U-s)^2 ds.$$

Taking the expectation on both sides:

$$\mathbb{E}h(U) = h(0) + h^{(1)}(0)\mathbb{E}U + \frac{1}{2}h^{(2)}(0)\mathbb{E}[U^2] + \frac{1}{2}\mathbb{E} \int_0^U h^{(3)}(s)(U-s)^2 ds. \quad (3.48)$$

When (3.48) is applied to  $h(U) = g^{(1)}(U)$ , we have

$$\mathbb{E}g^{(1)}(U) = g^{(1)}(0) + g^{(2)}(0)\mathbb{E}U + \frac{1}{2}g^{(3)}(0)\mathbb{E}[U^2] + \frac{1}{2}\mathbb{E} \int_0^U g^{(4)}(s)(U-s)^2 ds. \quad (3.49)$$

Whereas, when (3.48) is applied to  $h(U) = Ug(U)$ , using  $(Ug(U))^{(k)} = Ug^{(k)}(U) + kg^{(k-1)}(U)$  we have

$$\mathbb{E}[Ug(U)] - g(0)\mathbb{E}U = g^{(1)}(0)\mathbb{E}[U^2] + \frac{1}{2}\mathbb{E} \int_0^U (sg^{(3)}(s) + 3g^{(2)}(s))(U-s)^2 ds. \quad (3.50)$$

Subtracting (3.49) and (3.50), we have the bound

$$\begin{aligned}
& \left| \mathbb{E}[Ug(U)] - \mathbb{E}g^{(1)}(U)\mathbb{E}[U^2] - g(0)\mathbb{E}U \right| \\
&= \left| \frac{1}{2}\mathbb{E} \int_0^U (sg^{(3)}(s) + 3g^{(2)}(s))(U-s)^2 ds - g^{(2)}(0)\mathbb{E}[U^2]\mathbb{E}U - \frac{1}{2}g^{(3)}(0)\mathbb{E}[U^2]^2 \right. \\
&\quad \left. - \frac{1}{2}\mathbb{E}[U^2]\mathbb{E} \int_0^U g^{(4)}(s)(U-s)^2 ds \right| \\
&\leq \frac{C_3}{2} \left| \mathbb{E} \int_0^U s(U-s)^2 ds \right| + \frac{3C_2}{2} \left| \mathbb{E} \int_0^U (U-s)^2 ds \right| + C_2\mathbb{E}[U^2]\mathbb{E}U \\
&\quad + \frac{C_3}{2}\mathbb{E}[U^2]^2 + \frac{C_4}{2}\mathbb{E}[U^2] \left| \mathbb{E} \int_0^U (U-s)^2 ds \right| \\
&= \frac{C_3}{24}\mathbb{E}[U^4] + \frac{C_2}{2}|\mathbb{E}[U^3]| + C_2\mathbb{E}U\mathbb{E}[U^2] + \frac{C_3}{2}\mathbb{E}[U^2]^2 + \frac{C_4}{6}|\mathbb{E}[U^3]|\mathbb{E}[U^2],
\end{aligned} \tag{3.51}$$

which is the right hand side of (3.47) after factorization.  $\square$

We now apply Lemma 3.13 to our specific problem in order to derive the approximate integration-by-parts formula (3.27).

*Proof of lemma 3.4.* In order to apply lemma 3.13 to the SBM, consider  $U = G_{ij}$  and  $g(U) = F_{t,\epsilon}(G_{ij})$  the free energy (3.11) seen as a function of  $G_{ij}$  (all other variables being fixed). For the expectation we take  $\mathbb{E} = \mathbb{E}_{G_{ij}|X_i, X_j}$ . At time  $t$  and for any integer  $k$

$$\mathbb{E}_{G_{ij}|X_i, X_j}[G_{ij}^k] = \mathbb{E}_{G_{ij}|X_i, X_j}G_{ij} = \bar{p}_n + \sqrt{1-t}\Delta_n X_i X_j = \mathcal{O}(\bar{p}_n),$$

because  $G_{ij} \in \{0, 1\}$ . For the derivatives we note that using the Taylor expansion of the logarithm, one obtains for any  $v_n \in \mathbb{R}$  and  $v_n \rightarrow 0$ ,  $\ln(1+v_n) - v_n = \mathcal{O}(|v_n|^2)$ , which also implies  $\ln(1+v_n) = \mathcal{O}(|v_n|)$ . (The reader should keep this fact in mind, as it is used again in the appendices whenever we need to expand the logarithm.) Now, this fact implies

$$\begin{aligned}
-F_{t,\epsilon}^{(1)}(G_{ij}) &= \frac{1}{n} \left\langle \ln\left(1 + \frac{\Delta_n}{\bar{p}_n} \sqrt{1-t} x_i x_j\right) - \ln\left(1 - \frac{\Delta_n}{1-\bar{p}_n} \sqrt{1-t} x_i x_j\right) \right\rangle_{t,\epsilon} \\
&= \frac{1}{n} \left( \frac{\Delta_n}{\bar{p}_n} + \frac{\Delta_n}{1-\bar{p}_n} \right) \sqrt{1-t} \langle x_i x_j \rangle_{t,\epsilon} + \mathcal{O}\left(\frac{1}{n} \left( \frac{\Delta_n}{1-\bar{p}_n} \right)^2 (1-t)\right) \\
&\quad + \mathcal{O}\left(\frac{1}{n} \left( \frac{\Delta_n}{(1-\bar{p}_n)} \right)^2 (1-t)\right) \\
&= \frac{1}{n} \frac{\Delta_n}{\bar{p}_n(1-\bar{p}_n)} \sqrt{1-t} \langle x_i x_j \rangle_{t,\epsilon} + \mathcal{O}\left(\frac{1}{n} \left( \frac{\Delta_n}{\bar{p}_n(1-\bar{p}_n)} \right)^2 (1-t)\right),
\end{aligned}$$

$$\begin{aligned}
-F_{t,\epsilon}^{(2)}(G_{ij}) &= \frac{1}{n} \left\langle \left( \ln(1 + \frac{\Delta_n}{\bar{p}_n} \sqrt{1-t} x_i x_j) - \ln(1 - \frac{\Delta_n}{1-\bar{p}_n} \sqrt{1-t} x_i x_j) \right)^2 \right\rangle_{t,\epsilon} \\
&\quad - \frac{1}{n} \left\langle \ln(1 + \frac{\Delta_n}{\bar{p}_n} \sqrt{1-t} x_i x_j) - \ln(1 - \frac{\Delta_n}{1-\bar{p}_n} \sqrt{1-t} x_i x_j) \right\rangle_{t,\epsilon}^2 \\
&= \mathcal{O}\left(\frac{1}{n} \left(\frac{\Delta_n}{\bar{p}_n(1-\bar{p}_n)}\right)^2 (1-t)\right).
\end{aligned}$$

To obtain these identities the reader has again to be careful in performing the derivatives: both the exponential of the Hamiltonian and the partition function appearing in the definition of the Gibbs-bracket depend on  $(G_{ij})$  (see the derivation of (3.20) for similar computations). In general,

$$|F_{t,\epsilon}^{(k)}(G_{ij})| = \mathcal{O}\left(\frac{1}{n} \left(\frac{\Delta_n}{\bar{p}_n(1-\bar{p}_n)}\right)^k (1-t)^{k/2}\right).$$

Using Lemma 3.13, we have

$$\begin{aligned}
A_n &\equiv \left| \mathbb{E}_{G_{ij}|X_i,X_j}[G_{ij}F_{t,\epsilon}] + \mathbb{E}_{G_{ij}|X_i,X_j}[G_{ij}] \right. \\
&\quad \times \left\{ \frac{1}{n} \frac{\Delta_n}{\bar{p}_n(1-\bar{p}_n)} \sqrt{1-t} \mathbb{E}_{G_{ij}|X_i,X_j}[\langle x_i x_j \rangle_{t,\epsilon}] \right. \\
&\quad \left. + \mathcal{O}\left(\frac{1-t}{n} \left(\frac{\Delta_n}{\bar{p}_n(1-\bar{p}_n)}\right)^2\right) - F_{t,\epsilon}(G_{ij}=0) \right\} \Big| \\
&= \mathcal{O}\left(\frac{\sqrt{1-t}}{n} \left( \left(\frac{\Delta_n}{\bar{p}_n(1-\bar{p}_n)}\right)^2 (\bar{p}_n + \bar{p}_n^2) \right. \right. \\
&\quad \left. \left. + \left(\frac{\Delta_n}{\bar{p}_n(1-\bar{p}_n)}\right)^3 (\bar{p}_n + \bar{p}_n^2) + \left(\frac{\Delta_n}{\bar{p}_n(1-\bar{p}_n)}\right)^4 \bar{p}_n^2 \right) \right) \\
&= \mathcal{O}\left(\frac{\sqrt{1-t}}{n} \frac{\Delta_n^2}{\bar{p}_n(1-\bar{p}_n)^2}\right).
\end{aligned}$$

Then, by the triangle inequality we extract

$$\begin{aligned}
&\left| \mathbb{E}_{G_{ij}|X_i,X_j}[G_{ij}F_{t,\epsilon}] \right. \\
&\quad \left. + \mathbb{E}_{G_{ij}|X_i,X_j}[G_{ij}] \left\{ \frac{1}{n} \frac{\Delta_n}{\bar{p}_n} \sqrt{1-t} \mathbb{E}_{G_{ij}|X_i,X_j}[\langle x_i x_j \rangle_{t,\epsilon}] - F_{t,\epsilon}(G_{ij}=0) \right\} \right| \\
&\leq A_n + (\bar{p}_n + \sqrt{1-t} \Delta_n X_i X_j) \mathcal{O}\left(\frac{1-t}{n} \left(\frac{\Delta_n}{\bar{p}_n(1-\bar{p}_n)}\right)^2\right) \\
&= \mathcal{O}\left(\frac{\sqrt{1-t}}{n} \frac{\Delta_n^2}{\bar{p}_n(1-\bar{p}_n)^2}\right) \\
&= \mathcal{O}\left(\frac{\sqrt{1-t} \lambda_n}{n^2(1-\bar{p}_n)}\right).
\end{aligned}$$

and recognize formula (3.27).

### 3.6.2 Mutual information and free energy: proof of Proposition 3.1

Using (3.3), we have the expression

$$\begin{aligned} I(\mathbf{X}; \mathbf{G}) &\equiv \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \ln \left\{ \frac{\mathbb{P}(\mathbf{G}|\mathbf{X})}{\mathbb{P}(\mathbf{G})} \right\} = \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \ln \left\{ \frac{\mathbb{P}(\mathbf{G}|\mathbf{X})}{\sum_{\mathbf{x} \in \mathcal{X}^n} \mathbb{P}_r(\mathbf{x}) \mathbb{P}(\mathbf{G}|\mathbf{x})} \right\} \\ &= \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \ln \left\{ \frac{\prod_{i < j} (\bar{p}_n + \Delta_n X_i X_j)^{G_{ij}} (1 - \bar{p}_n - \Delta_n X_i X_j)^{1-G_{ij}}}{\sum_{\mathbf{x} \in \mathcal{X}^n} \mathbb{P}_r(\mathbf{x}) \prod_{i < j} (\bar{p}_n + \Delta_n x_i x_j)^{G_{ij}} (1 - \bar{p}_n - \Delta_n x_i x_j)^{1-G_{ij}}} \right\}. \end{aligned}$$

We divide both the numerator and denominator by the same factor, and then rewrite the denominator in exponential form:

$$\begin{aligned} I(\mathbf{X}; \mathbf{G}) &= \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \ln \left\{ \frac{\prod_{i < j} (1 + \frac{\Delta_n}{\bar{p}_n} X_i X_j)^{G_{ij}} (1 - \frac{\Delta_n}{1-\bar{p}_n} X_i X_j)^{1-G_{ij}}}{\sum_{\mathbf{x} \in \mathcal{X}^n} \mathbb{P}_r(\mathbf{x}) \prod_{i < j} (1 + \frac{\Delta_n}{\bar{p}_n} x_i x_j)^{G_{ij}} (1 - \frac{\Delta_n}{1-\bar{p}_n} x_i x_j)^{1-G_{ij}}} \right\} \\ &= \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \ln \left\{ \prod_{i < j} \left( 1 + \frac{\Delta_n}{\bar{p}_n} X_i X_j \right)^{G_{ij}} \left( 1 - \frac{\Delta_n}{1-\bar{p}_n} X_i X_j \right)^{1-G_{ij}} \right\} \\ &\quad - \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \ln \mathcal{Z}(\mathbf{G}). \end{aligned} \tag{3.52}$$

Recall  $\mathbb{E}_{G_{ij}|X_i, X_j} G_{ij} = \bar{p}_n + \Delta_n X_i X_j$ . The first term in (3.52) equals

$$\begin{aligned} &\sum_{i < j} \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \left\{ G_{ij} \ln \left( 1 + \frac{\Delta_n}{\bar{p}_n} X_i X_j \right) + (1 - G_{ij}) \ln \left( 1 - \frac{\Delta_n}{1-\bar{p}_n} X_i X_j \right) \right\} \\ &= \sum_{i < j} \mathbb{E}_{\mathbf{X}} \left\{ (\bar{p}_n + \Delta_n X_i X_j) \ln \left( 1 + \frac{\Delta_n}{\bar{p}_n} X_i X_j \right) \right. \\ &\quad \left. + (1 - \bar{p}_n - \Delta_n X_i X_j) \ln \left( 1 - \frac{\Delta_n}{1-\bar{p}_n} X_i X_j \right) \right\}. \end{aligned} \tag{3.53}$$

Let  $X \sim \mathbb{P}_r$ . We can further write explicitly the expectation in (3.53) that leads us to conclude

$$\begin{aligned} \frac{1}{n} I(\mathbf{X}; \mathbf{G}) &= \frac{n-1}{2} \left\{ r^2 (\bar{p}_n + \Delta_n \frac{1-r}{r}) \ln \left( 1 + \frac{\Delta_n}{\bar{p}_n} \frac{1-r}{r} \right) \right. \\ &\quad + r^2 (1 - \bar{p}_n - \Delta_n \frac{1-r}{r}) \ln \left( 1 - \frac{\Delta_n}{1-\bar{p}_n} \frac{1-r}{r} \right) \\ &\quad + (1-r)^2 (\bar{p}_n + \Delta_n \frac{r}{1-r}) \ln \left( 1 + \frac{\Delta_n}{\bar{p}_n} \frac{r}{1-r} \right) \\ &\quad + (1-r)^2 (1 - \bar{p}_n - \Delta_n \frac{r}{1-r}) \ln \left( 1 - \frac{\Delta_n}{1-\bar{p}_n} \frac{r}{1-r} \right) \\ &\quad + 2r(1-r)(\bar{p}_n - \Delta_n) \ln \left( 1 - \frac{\Delta_n}{\bar{p}_n} \right) + 2r(1-r)(1 - \bar{p}_n + \Delta_n) \ln \left( 1 + \frac{\Delta_n}{1-\bar{p}_n} \right) \left. \right\} \\ &\quad - \frac{1}{n} \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \ln \mathcal{Z}(\mathbf{G}). \end{aligned} \tag{3.54}$$

Using the Taylor expansion of the logarithm, (3.54) becomes

$$\begin{aligned} \frac{1}{n}I(\mathbf{X}; \mathbf{G}) &= \frac{\lambda_n(n-1)}{4n} - \frac{1}{n}\mathbb{E}_{\mathbf{X}}\mathbb{E}_{\mathbf{G}|\mathbf{X}} \ln \mathcal{Z}(\mathbf{G}) \\ &\quad + \frac{n-1}{2} \sum_{k=3}^{\infty} \frac{\Delta_n^k}{k(k-1)} \left( \frac{1}{\bar{p}_n^{k-1}} + \frac{(-1)^k}{(1-\bar{p}_n)^{k-1}} \right) \mathbb{E}[X^k]^2, \end{aligned}$$

where  $\mathbb{E}[X^k]^2 = r^2(\frac{1-r}{r})^k + (1-r)^2(\frac{r}{1-r})^k + (-1)^k 2r(1-r)$ . This becomes the expression in (3.5) by noting that the last term is  $\mathcal{O}\left(n\Delta_n^3/(\bar{p}_n(1-\bar{p}_n))^2\right) = \mathcal{O}(\lambda_n^{3/2}/\sqrt{n\bar{p}_n(1-\bar{p}_n)})$ .

### 3.6.3 Small error terms in the sum rule: proof of (3.30)

Recalling the definitions (3.9) and (3.10), let

$$\begin{aligned} \mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y}) &\equiv \mathcal{H}_{\text{SBM};t}(\mathbf{x}; \mathbf{G} \setminus G_{ij}) + \mathcal{H}_{\text{dec};t,\epsilon}(\mathbf{x}; \mathbf{Y}), \quad (3.55) \\ \mathcal{H}_{\text{SBM};t}(\mathbf{x}; \mathbf{G} \setminus G_{ij}) &\equiv - \sum_{k < l: (k,l) \notin \{(i,j), (j,i)\}} \left\{ G_{kl} \ln\left(1 + \frac{\Delta_n}{\bar{p}_n} \sqrt{1-t} x_k x_l\right) \right. \\ &\quad \left. + (1 - G_{kl}) \ln\left(1 - \frac{\Delta_n}{1-\bar{p}_n} \sqrt{1-t} x_k x_l\right) \right\}. \end{aligned}$$

Also let  $F_{t,\epsilon;\sim G_{ij}} \equiv n^{-1} \ln \sum_{\mathbf{x} \in \mathcal{X}^n} e^{-\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y})} \mathbb{P}_r(\mathbf{x})$ , and  $\langle - \rangle_{t,\epsilon;\sim G_{ij}}$  be the Gibbs-bracket associated to the measure proportional to  $\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y})$ . The difference of the free energies when changing one  $G_{ij}$  can be written in terms of this Gibbs-bracket:

$$\begin{aligned} &\mathbb{E}_{G_{ij}|X_i, X_j} F_{t,\epsilon} - F_{t,\epsilon}(G_{ij} = 0) \\ &= \mathbb{P}_t(G_{ij} = 1 | X_i, X_j) (F_{t,\epsilon}(G_{ij} = 1) - F_{t,\epsilon}(G_{ij} = 0)) \quad (3.56) \\ &= \mathbb{P}_t(G_{ij} = 1 | X_i, X_j) \{ (F_{t,\epsilon}(G_{ij} = 1) - F_{t,\epsilon;\sim G_{ij}}) - (F_{t,\epsilon}(G_{ij} = 0) - F_{t,\epsilon;\sim G_{ij}}) \} \\ &= -(\bar{p}_n + \sqrt{1-t}\Delta_n X_i X_j) \frac{1}{n} \\ &\quad \left\{ \ln \frac{\sum_{\mathbf{x} \in \mathcal{X}^n} e^{-\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y}) + (\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y}) - \mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}, G_{ij}=1, \mathbf{Y}))} \mathbb{P}_r(\mathbf{x})}{\sum_{\mathbf{x} \in \mathcal{X}^n} e^{-\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y})} \mathbb{P}_r(\mathbf{x})} \right. \\ &\quad \left. - \ln \frac{\sum_{\mathbf{x} \in \mathcal{X}^n} e^{-\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y}) + (\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y}) - \mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}, G_{ij}=0, \mathbf{Y}))} \mathbb{P}_r(\mathbf{x})}{\sum_{\mathbf{x} \in \mathcal{X}^n} e^{-\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y})} \mathbb{P}_r(\mathbf{x})} \right\} \\ &= -(\bar{p}_n + \sqrt{1-t}\Delta_n X_i X_j) \frac{1}{n} \left\{ \ln \left\langle e^{\mathcal{H}_{\text{SBM};t}(\mathbf{x}; \mathbf{G} \setminus G_{ij}) - \mathcal{H}_{\text{SBM};t}(\mathbf{x}; \mathbf{G}, G_{ij}=1)} \right\rangle_{t,\epsilon;\sim G_{ij}} \right. \\ &\quad \left. - \ln \left\langle e^{\mathcal{H}_{\text{SBM};t}(\mathbf{x}; \mathbf{G} \setminus G_{ij}) - \mathcal{H}_{\text{SBM};t}(\mathbf{x}; \mathbf{G}, G_{ij}=0)} \right\rangle_{t,\epsilon;\sim G_{ij}} \right\} \\ &= -(\bar{p}_n + \sqrt{1-t}\Delta_n X_i X_j) \frac{1}{n} \left\{ \ln \left\langle 1 + \frac{\Delta_n}{\bar{p}_n} \sqrt{1-t} x_i x_j \right\rangle_{t,\epsilon;\sim G_{ij}} \right. \\ &\quad \left. - \ln \left\langle 1 - \frac{\Delta_n}{1-\bar{p}_n} \sqrt{1-t} x_i x_j \right\rangle_{t,\epsilon;\sim G_{ij}} \right\}. \quad (3.57) \end{aligned}$$

Using the Taylor expansion of the logarithms in (3.57), we have

$$\begin{aligned}
& \mathbb{E}_{G_{ij}|X_i, X_j} F_{t,\epsilon} - F_{t,\epsilon}(G_{ij} = 0) \\
&= -(\bar{p}_n + \sqrt{1-t}\Delta_n X_i X_j) \frac{1}{n} \left\{ \frac{\Delta_n \sqrt{1-t}}{\bar{p}_n(1-\bar{p}_n)} \langle x_i x_j \rangle_{t,\epsilon;\sim G_{ij}} \right. \\
&\quad \left. + \sum_{k=2}^{\infty} \frac{\Delta_n^k}{k} \left( \frac{(-1)^k}{\bar{p}_n^k} - \frac{1}{(1-\bar{p}_n)^k} \right) (1-t)^{k/2} \langle x_i x_j \rangle_{t,\epsilon;\sim G_{ij}}^k \right\} \\
&= -(\bar{p}_n + \sqrt{1-t}\Delta_n X_i X_j) \frac{\Delta_n \sqrt{1-t}}{n\bar{p}_n(1-\bar{p}_n)} \langle x_i x_j \rangle_{t,\epsilon;\sim G_{ij}} + \mathcal{O}\left(\frac{\Delta_n^2(1-t)}{n\bar{p}_n(1-\bar{p}_n)^2}\right).
\end{aligned}$$

Therefore, replacing in the expression of  $E_1$ , we find

$$E_1 = E_1^{(a)} + E_1^{(b)}$$

where

$$\begin{aligned}
E_1^{(a)} &= \frac{\Delta_n^2}{2n\bar{p}_n(1-\bar{p}_n)} \sum_{i<j} \mathbb{E}_{\sim G_{ij}} \left[ \frac{\bar{p}_n + \sqrt{1-t}\Delta_n X_i X_j}{1-\bar{p}_n - \sqrt{1-t}\Delta_n X_i X_j} X_i X_j \langle x_i x_j \rangle_{t,\epsilon;\sim G_{ij}} \right], \\
E_1^{(b)} &= \mathcal{O}\left(\frac{\Delta_n n^2}{\sqrt{1-t}(1-\bar{p}_n)} \cdot \frac{\Delta_n^2(1-t)}{n\bar{p}_n(1-\bar{p}_n)^2}\right) \\
&= \mathcal{O}\left(\frac{n\Delta_n^3}{\bar{p}_n(1-\bar{p}_n)^3}\right) = \mathcal{O}\left(\frac{\lambda_n \Delta_n}{(1-\bar{p}_n)^2}\right). \tag{3.58}
\end{aligned}$$

We then observe that

$$\begin{aligned}
E_1^{(a)} + E_2 &= \frac{\Delta_n^2}{2n\bar{p}_n(1-\bar{p}_n)} \sum_{i<j} \mathbb{E}_{\sim G_{ij}} \left[ \frac{\bar{p}_n + \sqrt{1-t}\Delta_n X_i X_j}{1-\bar{p}_n - \sqrt{1-t}\Delta_n X_i X_j} X_i X_j \right. \\
&\quad \left. (\mathbb{E}_{G_{ij}|X_i, X_j} [\langle x_i x_j \rangle_{t,\epsilon}] - \langle x_i x_j \rangle_{t,\epsilon;\sim G_{ij}}) \right]. \tag{3.59}
\end{aligned}$$

The difference between the Gibbs-brackets in (3.59) can be expanded as

$$\begin{aligned}
& \mathbb{E}_{G_{ij}|X_i, X_j} [\langle x_i x_j \rangle_{t,\epsilon}] - \langle x_i x_j \rangle_{t,\epsilon;\sim G_{ij}} \\
&= \mathbb{P}_t(G_{ij} = 1|X_i, X_j) (\langle x_i x_j \rangle_{t,\epsilon;G_{ij}=1} - \langle x_i x_j \rangle_{t,\epsilon;\sim G_{ij}}) \\
&\quad + \mathbb{P}_t(G_{ij} = 0|X_i, X_j) (\langle x_i x_j \rangle_{t,\epsilon;G_{ij}=0} - \langle x_i x_j \rangle_{t,\epsilon;\sim G_{ij}}), \tag{3.60}
\end{aligned}$$

and we can evaluate  $\langle x_i x_j \rangle_{t, \epsilon; G_{ij}=1} - \langle x_i x_j \rangle_{t, \epsilon; \sim G_{ij}}$  by an interpolation:

$$\begin{aligned}
& \langle x_i x_j \rangle_{t, \epsilon; G_{ij}=1} - \langle x_i x_j \rangle_{t, \epsilon; \sim G_{ij}} \\
&= \int_0^1 ds \frac{d}{ds} \left\{ \left( \sum_{\mathbf{x} \in \mathcal{X}^n} x_i x_j \exp \left\{ -\mathcal{H}_{t, \epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y}) \right. \right. \right. \\
&\quad \left. \left. \left. + s \ln \left( 1 + x_i x_j \sqrt{1 - t \frac{\Delta_n}{\bar{p}_n}} \right) \right\} \mathbb{P}_r(\mathbf{x}) \right) \right. \\
&\quad \left. \left/ \left( \sum_{\mathbf{x} \in \mathcal{X}^n} \exp \left\{ -\mathcal{H}_{t, \epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y}) \right. \right. \right. \right. \\
&\quad \left. \left. \left. + s \ln \left( 1 + x_i x_j \sqrt{1 - t \frac{\Delta_n}{\bar{p}_n}} \right) \right\} \mathbb{P}_r(\mathbf{x}) \right) \right\} \\
&= \int_0^1 ds \left\{ \langle x_i x_j \ln \left( 1 + x_i x_j \sqrt{1 - t \frac{\Delta_n}{\bar{p}_n}} \right) \rangle_{t, \epsilon; s} \right. \\
&\quad \left. - \langle x_i x_j \rangle_{t, \epsilon; s} \langle \ln \left( 1 + x_i x_j \sqrt{1 - t \frac{\Delta_n}{\bar{p}_n}} \right) \rangle_{t, \epsilon; s} \right\}, \quad (3.61)
\end{aligned}$$

where  $\langle - \rangle_{t, \epsilon; s}$  is the Gibbs-bracket associated to the measure proportional to

$$\exp \left\{ -\mathcal{H}_{t, \epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y}) + s \ln \left( 1 + x_i x_j \sqrt{1 - t \frac{\Delta_n}{\bar{p}_n}} \right) \right\} \mathbb{P}_r(\mathbf{x})$$

with  $\mathcal{H}_{t, \epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y})$  defined in (3.55). By the Taylor expansion of the logarithms in (3.61) and using  $\mathbb{P}_t(G_{ij} = 1 | X_i, X_j) = \mathcal{O}(\bar{p}_n)$ , we see that the first term of (3.60) is  $\mathcal{O}(\Delta_n)$ . The same kind of calculation is used to see that the second term of (3.60) is also  $\mathcal{O}(\Delta_n)$ . This implies for (3.59)

$$E_1^{(a)} + E_2 = \mathcal{O}\left(\frac{n\Delta_n^3}{(1 - \bar{p}_n)^2}\right) = \mathcal{O}\left(\frac{\lambda_n \Delta_n \bar{p}_n}{(1 - \bar{p}_n)^2}\right), \quad (3.62)$$

which tends to zero. Now we conclude by noting that  $E_1 + E_2 = E_1^{(a)} + E_1^{(b)} + E_2$  and using (3.58) and (3.62) to obtain (3.30).

### 3.6.4 Concentration of free energy: proof of Lemma 3.11

The generation of quenched variables can be divided into two stages: firstly  $\mathbf{X}$ , then  $\mathbf{G}$  given  $\mathbf{X}$ , and independently the Gaussian noise  $\mathbf{Z}$ . We expand the variance of free energy according to the two stages (recall  $f_{t, \epsilon} = \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t, \epsilon}$ ):

$$\mathbb{E}[(F_{t, \epsilon} - f_{t, \epsilon})^2] = \mathbb{E}[(F_{t, \epsilon} - \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t, \epsilon})^2] + \mathbb{E}[(\mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t, \epsilon} - f_{t, \epsilon})^2]. \quad (3.63)$$

In each stage the variables are all independently generated. This enables us to use Efron-Stein inequality to show the concentration of free energy.

Let  $\mathbf{Z}^{(i)}$  be a vector such that  $\mathbf{Z}^{(i)}$  differs from  $\mathbf{Z}$  only at the  $i$ -th which becomes  $Z'_i$  drawn independently from the same distribution as the one of



$Z_i \sim \mathcal{N}(0, 1)$ . We define  $\mathbf{G}^{(ij)}$  and  $\mathbf{X}^{(i)}$  in the similar manner with respect to  $\mathbf{G}$  and  $\mathbf{X}$ . Efron-Stein's inequality tells us that

$$\begin{aligned} & \mathbb{E}[(F_{t,\epsilon} - \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon})^2] \\ & \leq \frac{1}{2} \sum_{i=1}^n \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} \mathbb{E}_{Z'_i} [(F_{t,\epsilon}(\mathbf{Z}) - F_{t,\epsilon}(\mathbf{Z}^{(i)}))^2] \\ & \quad + \frac{1}{2} \sum_{i < j} \mathbb{E}_{\mathbf{X}} \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{G}'_{ij}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} [(F_{t,\epsilon}(\mathbf{G}) - F_{t,\epsilon}(\mathbf{G}^{(ij)}))^2], \end{aligned} \quad (3.64)$$

as well as

$$\begin{aligned} & \mathbb{E}[(\mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon} - f_{t,\epsilon})^2] \\ & \leq \frac{1}{2} \sum_{i=1}^n \mathbb{E}_{\mathbf{X}} \mathbb{E}_{X'_i} [(\mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}) - \mathbb{E}_{\mathbf{G}|\mathbf{X}^{(i)}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}^{(i)}))^2]. \end{aligned} \quad (3.65)$$

By (3.63) it suffices to show that both (3.64) and (3.65) are upper bounded by  $C_n(r, \lambda_n)/n$  for some large enough sequence  $C_n(r, \lambda_n)$  that converges to a constant.

### Bound on (3.64)

The bound obtained from Efron-Stein's inequality is a sum of local variances of the free energy. The bound on the difference due to a local change can be estimated by interpolation. For the first one we have

$$\begin{aligned} & |F_{t,\epsilon}(\mathbf{Z}) - F_{t,\epsilon}(\mathbf{Z}^{(i)})| \\ & = \frac{1}{n} \left| \int_0^1 ds \frac{d}{ds} \ln \sum_{\mathbf{x} \in \mathcal{X}^n} \exp \{ -s \mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{X}, \mathbf{Z}) \right. \\ & \quad \left. - (1-s) \mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{X}, \mathbf{Z}^{(i)}) \} \mathbb{P}_r(\mathbf{x}) \right| \\ & = \frac{1}{n} \left| \int_0^1 ds \langle \mathcal{H}_{\text{dec};t,\epsilon}(\mathbf{x}; \mathbf{X}, \mathbf{Z}^{(i)}) - \mathcal{H}_{\text{dec};t,\epsilon}(\mathbf{x}; \mathbf{X}, \mathbf{Z}) \rangle_s \right| \\ & = \frac{1}{n} \left| \int_0^1 ds \sqrt{R(t, \epsilon)} \langle x_i \rangle_s (Z'_i - Z_i) \right| \\ & \leq \frac{1}{n} \sqrt{(2s_n + \lambda_n) \frac{1-r}{r}} |Z'_i - Z_i| \end{aligned}$$

where the Gibbs-bracket  $\langle - \rangle_s$  is associated to the measure proportional to  $\exp\{-s \mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{X}, \mathbf{Z}) - (1-s) \mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{X}, \mathbf{Z}^{(i)})\}$ . This implies an upper bound on the first sum in (3.64):

$$\begin{aligned} & \frac{1}{2} \sum_{i=1}^n \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} \mathbb{E}_{Z'_i} [(F_{t,\epsilon}(\mathbf{Z}) - F_{t,\epsilon}(\mathbf{Z}^{(i)}))^2] \\ & \leq \frac{1}{2n^2} (2s_n + \lambda_n) \frac{1-r}{r} \sum_{i=1}^n \mathbb{E}[(Z'_i - Z_i)^2] \leq \frac{C_n(r, \lambda_n)}{n}. \end{aligned}$$

Another interpolation gives

$$\begin{aligned}
& |F_{t,\epsilon}(\mathbf{G}) - F_{t,\epsilon}(\mathbf{G}^{(ij)})| \\
&= \frac{1}{n} \left| \int_0^1 ds \frac{d}{ds} \ln \sum_{\mathbf{x} \in \mathcal{X}^n} \exp \left\{ -s \mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{X}, \mathbf{Z}) \right. \right. \\
&\quad \left. \left. - (1-s) \mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}^{(ij)}, \mathbf{X}, \mathbf{Z}) \right\} \mathbb{P}_r(\mathbf{x}) \right| \\
&= \frac{1}{n} \left| (G'_{ij} - G_{ij}) \left\langle \ln \left( 1 + \frac{\Delta_n}{\bar{p}_n} \sqrt{1 - t x_i x_j} \right) - \ln \left( 1 - \frac{\Delta_n}{1 - \bar{p}_n} \sqrt{1 - t x_i x_j} \right) \right\rangle_s \right| \\
&\leq \frac{C(r) \Delta_n}{2n \bar{p}_n (1 - \bar{p}_n)} |G'_{ij} - G_{ij}|
\end{aligned}$$

for some constant  $C(r)$ , and where  $\langle - \rangle_s$  is associated to the measure proportional to  $\exp \{ -s \mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}, \mathbf{X}, \mathbf{Z}) - (1-s) \mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G}^{(ij)}, \mathbf{X}, \mathbf{Z}) \}$ . This bounds the second sum in (3.64) as

$$\begin{aligned}
& \frac{C(r) \Delta_n^2}{2n^2 \bar{p}_n^2 (1 - \bar{p}_n)^2} \sum_{i < j} \mathbb{E}_{G_{ij}|X_i, X_j} \mathbb{E}_{G'_{ij}|X_i, X_j} [(G'_{ij} - G_{ij})^2] \\
&= \frac{C(r) \Delta_n^2}{n^2 \bar{p}_n^2 (1 - \bar{p}_n)^2} \sum_{i < j} \text{Var}_{G_{ij}|X_i, X_j}(G_{ij}) \leq \frac{C_n(r, \lambda_n)}{n},
\end{aligned}$$

using that  $(G_{ij})$  are 0, 1 Bernoulli variables, and the variance

$$\text{Var}_{G_{ij}|X_i, X_j}(G_{ij}) = (\bar{p}_n + \Delta_n \sqrt{1 - t X_i X_j})(1 - \bar{p}_n + \Delta_n \sqrt{1 - t X_i X_j})$$

as well as  $(\Delta_n / (p_n(1 - \bar{p}_n)))^2 = \lambda_n / (n \bar{p}_n (1 - \bar{p}_n))$  in the last inequality.

### Bound on (3.65)

We relax (3.65) with inequality  $((a - c) + (c - b))^2 \leq 2(a - c)^2 + 2(c - b)^2$  so that

$$\begin{aligned}
& \mathbb{E}[(\mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon} - f_{t,\epsilon})^2] \\
&\leq \sum_{i=1}^n \mathbb{E}_{\mathbf{X}} \mathbb{E}_{X'_i} [(\mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}) - \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}^{(i)}))^2] \\
&\quad + \sum_{i=1}^n \mathbb{E}_{\mathbf{X}} \mathbb{E}_{X'_i} [(\mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}^{(i)}) - \mathbb{E}_{\mathbf{G}|\mathbf{X}^{(i)}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}^{(i)}))^2]. \quad (3.66)
\end{aligned}$$

The difference in the first sum is given by

$$\begin{aligned}
& |F_{t,\epsilon}(\mathbf{X}) - F_{t,\epsilon}(\mathbf{X}^{(i)})| \\
&= \frac{1}{n} \left| \int_0^1 ds \frac{d}{ds} \ln \sum_{\mathbf{x} \in \mathcal{X}^n} \exp \left\{ -s \mathcal{H}_{t,\epsilon}(\mathbf{G}, \mathbf{X}, \mathbf{Z}, \mathbf{x}) \right. \right. \\
&\quad \left. \left. - (1-s) \mathcal{H}_{t,\epsilon}(\mathbf{G}, \mathbf{X}^{(i)}, \mathbf{Z}, \mathbf{x}) \right\} \mathbb{P}_r(\mathbf{x}) \right| \\
&= \frac{1}{n} \left| \int_0^1 ds R(t, \epsilon) \langle x_i \rangle_s (X'_i - X_i) \right|
\end{aligned}$$

where  $\langle - \rangle_s$  is associated to the measure proportional to

$$\exp\{-s\mathcal{H}_{t,\epsilon}(\mathbf{G}, \mathbf{X}, \mathbf{Z}, \mathbf{x}) - (1-s)\mathcal{H}_{t,\epsilon}(\mathbf{G}, \mathbf{X}^{(i)}, \mathbf{Z}, \mathbf{x})\}.$$

Therefore the sum of square is bounded by  $C_n(r, \lambda_n)/n$  using  $R(t, \epsilon) \in [0, \lambda_n]$ .

For the second sum we use another interpolation:

$$\begin{aligned} & \mathbb{E}_{\mathbf{G}|\mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}^{(i)}) - \mathbb{E}_{\mathbf{G}|\mathbf{X}^{(i)}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}^{(i)}) \\ &= \int_0^1 ds \sum_{\mathbf{G}} \frac{d}{ds} \mathbb{P}_{t,s}(\mathbf{G}|\mathbf{X}, X'_i) \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}^{(i)}), \end{aligned} \quad (3.67)$$

where

$$\begin{aligned} \mathbb{P}_{t,s}(\mathbf{G}|\mathbf{X}, X'_i) &\equiv \prod_{j:j \neq i}^n (\bar{p}_n + \sqrt{1-t}\Delta_n((1-s)X_i + sX'_i)X_j)^{G_{ij}} \\ &\quad (1 - \bar{p}_n - \sqrt{1-t}\Delta_n((1-s)X_i + sX'_i)X_j)^{1-G_{ij}} \\ &\quad \times \prod_{\substack{k < l: \\ k, l \neq i}} (\bar{p}_n + \sqrt{1-t}\Delta_n X_k X_l)^{G_{kl}} (1 - \bar{p}_n - \sqrt{1-t}\Delta_n X_k X_l)^{1-G_{kl}}. \end{aligned}$$

As  $G_{ij} \in \{0, 1\}$ , we have various ways to write  $\mathbb{P}_{t,s}(\mathbf{G}|\mathbf{X}, X'_i)$ . A convenient way is using

$$\begin{aligned} P_{ij} &\equiv (\bar{p}_n + \sqrt{1-t}\Delta_n((1-s)X_i + sX'_i)X_j)^{G_{ij}} \\ &\quad (1 - \bar{p}_n - \sqrt{1-t}\Delta_n((1-s)X_i + sX'_i)X_j)^{1-G_{ij}} \\ &= G_{ij} \{ \bar{p}_n + \sqrt{1-t}\Delta_n((1-s)X_i + sX'_i)X_j \} \\ &\quad + (1 - G_{ij}) \{ 1 - \bar{p}_n - \sqrt{1-t}\Delta_n((1-s)X_i + sX'_i)X_j \}. \end{aligned}$$

A compact formula for  $dP_{ij}/ds$  can then be derived:

$$\begin{aligned} \frac{dP_{ij}}{ds} &= (2G_{ij} - 1) \sqrt{1-t} \Delta_n(X'_i - X_i) X_j \\ &= (-1)^{1+G_{ij}} \sqrt{1-t} \Delta_n(X'_i - X_i) X_j. \end{aligned} \quad (3.68)$$

Let  $\mathbf{G}_{\sim(i,j)} \equiv \mathbf{G} \setminus G_{ij}$  and  $\mathbb{P}_{t,s}(\mathbf{G}_{\sim(i,j)}|\mathbf{X}, X'_i) \equiv \sum_{G_{ij} \in \{0,1\}} \mathbb{P}_{t,s}(\mathbf{G}|\mathbf{X}, X'_i)$  be the marginal of this sub-graph. Using (3.68) we obtain

$$\begin{aligned} \frac{d}{ds} \mathbb{P}_{t,s}(\mathbf{G}|\mathbf{X}, X'_i) &= \sum_{j:j \neq i}^n \frac{dP_{ij}}{ds} \mathbb{P}_{t,s}(\mathbf{G}_{\sim(i,j)}|\mathbf{X}, X'_i) \\ &= \sum_{j:j \neq i}^n \sqrt{1-t} \Delta_n(X'_i - X_i) X_j (-1)^{1+G_{ij}} \mathbb{P}_{t,s}(\mathbf{G}_{\sim(i,j)}|\mathbf{X}, X'_i). \end{aligned} \quad (3.69)$$

Substituting (3.69) into (3.67) gives

$$\begin{aligned}
& \int_0^1 ds \sum_{\mathbf{G}} \sum_{j:j \neq i}^n \sqrt{1-t} \Delta_n(X'_i - X_i) X_j (-1)^{1+G_{ij}} \mathbb{P}_{t,s}(\mathbf{G}_{\sim(i,j)} | \mathbf{X}, X'_i) \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}^{(i)}) \\
&= \int_0^1 ds \sum_{j:j \neq i}^n \sqrt{1-t} \Delta_n(X'_i - X_i) X_j \sum_{G_{ij} \in \{0,1\}} (-1)^{1+G_{ij}} \mathbb{E}_{\mathbf{G}_{\sim(i,j)} | \mathbf{X}, X'_i} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}^{(i)}) \\
&= \int_0^1 ds \sum_{j:j \neq i}^n \sqrt{1-t} \Delta_n(X'_i - X_i) X_j \\
&\quad \mathbb{E}_{\mathbf{G}_{\sim(i,j)} | \mathbf{X}, X'_i} \mathbb{E}_{\mathbf{Z}} [F_{t,\epsilon}(\mathbf{X}^{(i)}, G_{ij} = 1) - F_{t,\epsilon}(\mathbf{X}^{(i)}, G_{ij} = 0)], \quad (3.70)
\end{aligned}$$

where  $\mathbb{E}_{\mathbf{G}_{\sim(i,j)} | \mathbf{X}, X'_i}$  corresponds to the expectation with respect to the distribution  $\mathbb{P}_{t,s}(\mathbf{G}_{\sim(i,j)} | \mathbf{X}, X'_i)$ . To evaluate the difference of free energy in (3.70), first we define  $\mathbf{Y}^{(i)} = \sqrt{R(t, \epsilon)} \mathbf{X}^{(i)} + \mathbf{Z}$ , and  $\langle - \rangle_{t,\epsilon; \mathbf{X}^{(i)}, \sim G_{ij}}$  is associated to  $\exp\{-\mathcal{H}_{t,\epsilon}(\mathbf{x}; \mathbf{G} \setminus G_{ij}, \mathbf{Y}^{(i)})\}$  defined in (3.55). The same calculation as in (3.56) – (3.57) gives

$$\begin{aligned}
& F_{t,\epsilon}(\mathbf{X}^{(i)}, G_{ij} = 1) - F_{t,\epsilon}(\mathbf{X}^{(i)}, G_{ij} = 0) \\
&= -\frac{1}{n} \left\{ \ln \left\langle 1 + \frac{\Delta_n}{\bar{p}_n} \sqrt{1-t} x_i x_j \right\rangle_{t,\epsilon; \mathbf{X}^{(i)}, \sim G_{ij}} - \ln \left\langle 1 - \frac{\Delta_n}{1 - \bar{p}_n} \sqrt{1-t} x_i x_j \right\rangle \right\}. \quad (3.71)
\end{aligned}$$

Expanding the logarithms we can see (3.71) is  $\mathcal{O}(\Delta_n / (n \bar{p}_n (1 - \bar{p}_n)))$ . Using this fact and that all other terms inside the sum of (3.70) are upper bounded by constants, we see that (3.70) is  $\mathcal{O}(\Delta_n^2 / (\bar{p}_n (1 - \bar{p}_n))) = \mathcal{O}(\lambda_n / n)$ . We can then upper bound the second term of (3.66):

$$\sum_{i=1}^n \mathbb{E}_{\mathbf{X}} \mathbb{E}_{X'_i} [(\mathbb{E}_{\mathbf{G} | \mathbf{X}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}^{(i)}) - \mathbb{E}_{\mathbf{G} | \mathbf{X}^{(i)}} \mathbb{E}_{\mathbf{Z}} F_{t,\epsilon}(\mathbf{X}^{(i)}))^2] \leq \frac{C_n(r, \lambda_n)}{n}.$$

---

# Sparse Censored Block Model

---

# 4

## 4.1 Introduction

In this chapter, we move to the community detection models on sparse graphs. Other than community detection, sparse models can also be found in communication such as the low-density parity-check codes and low-density generator-matrix codes. It is fair to say that the replica symmetric formulas for the mutual information is much more complicated in such models. Indeed, besides the measurements (or channel outputs): (i) the sparse graph is also random; (ii) the single-letter variational problem involves a functional over a set of probability distributions (instead of scalars as in the dense-graph case). Existing rigorous derivations of the replica symmetric formulas for sparse models have so far been achieved in [46] by using a combination of the interpolation method (first developed by [65] for sparse models) and the rigorous version of the cavity method [66], and achieved in [67] by using spatial coupling.

In this work, we consider a simple version of the censored block model [12, 13, 14, 15], for which we fully develop the adaptive interpolation method. We believe that this constitutes the first step towards an analysis of more complicated models via this method.

In the censored block model, we have a set of  $n$  hidden binary variables. We observe  $\alpha n$  products of random  $K$ -tuples, where  $\alpha > 0$  is called the fraction of measurements, through a noisy channel. The goal is to reconstruct an estimate of the hidden variables from the noisy observations. There are other interpretations of this model. For example, it can be interpreted as a low-density generator-matrix code ensemble, with design communication rate  $1/\alpha$ , on a factor graph with degree- $K$  factor nodes and variable nodes with degrees following the Poisson distribution (when  $n \rightarrow +\infty$ ). Another possible interpretation is, as a model of statistical mechanics, an Ising model on a

sparse random graph with  $K$ -spin interactions. The censored block model has been discussed when the measurement channel is a binary symmetric channel in [68], and the replica formula is proved in this case [46]. Here, we consider a simpler situation where the measurement channel is the binary erasure channel (BEC), for which the adaptive interpolation method can be completely developed. As we will see, this method requires concentration results for a whole set of suitable “overlaps” in the case of sparse graphs. Here, we solve this issue for the BEC, and it is currently the only aspect of the method that is missing for extending our analysis to other channels.

This chapter is organized as follows. In Section 4.2, we give a precise formulation of the model and state the main result of this chapter (Theorem 4.1). In Section 4.3, we review two important tools used throughout our analysis, namely, the Nishimori identities and the Griffiths-Kelly-Sherman inequalities. The adaptive interpolation method for the sparse graph models is formulated in Section 4.4 and the core of the proof of Theorem 4.1 is also developed. This section contains the main new technical ideas of this chapter. Overlap concentration is proved in Section 4.6 and 4.7. A series of more technical results are found in Section 4.8.4 and in the appendices.

## 4.2 Setting and result

### 4.2.1 Censored block model

We shall denote binary variables by  $\sigma_i \in \{-1, +1\}$ ,  $i = 1, \dots, n$  and vectors of such variables by  $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n) \in \{-1, +1\}^n$ . Subsets  $S \subset \{1, \dots, n\}$  with at least two elements are always denoted by capital letters. For the product of binary variables in a subset  $S$  we use the shorthand notation  $\sigma_S \equiv \prod_{i \in S} \sigma_i$ . If there is a possible confusion between small and capital letter subscripts we occasionally use more specific notations. Below, the integer  $K \geq 2$  and the fraction  $\alpha \in \mathbb{R}_+$  are fixed independent of  $n$ .

In the censored block model considered in this chapter,  $n$  *hidden* binary variables  $\boldsymbol{\sigma}^0 = (\sigma_1^0, \dots, \sigma_n^0)$  are i.i.d. uniform, i.e. drawn independently according to a  $\text{Ber}(1/2)$  prior  $\mathbb{P}_0(\sigma_i^0) = \frac{1}{2}\delta_{\sigma_i^0, +1} + \frac{1}{2}\delta_{\sigma_i^0, -1}$ . A *noiseless measurement* consists in a product  $\sigma_{a_1}^0 \sigma_{a_2}^0 \dots \sigma_{a_K}^0$  of a  $K$ -tuple of variables drawn uniformly at random. The  $K$ -tuple is identified with a subset  $A \equiv \{a_1, \dots, a_K\} \subset \{1, \dots, n\}$  and we set  $\sigma_A^0 \equiv \sigma_{a_1}^0 \sigma_{a_2}^0 \dots \sigma_{a_K}^0$ . Of course  $\sigma_A^0 = \pm 1$ . The *true observations*  $G_A \in \mathbb{R}$  are noisy versions of these products obtained through a binary input memoryless channel described by some transition probability  $Q(G_A | \sigma_A^0)$ . For large  $n$  the total number of observations  $m$  asymptotically follows a Poisson distribution with mean  $\alpha n$ , i.e.,  $m \sim \text{Poi}(\alpha n)$ . We shall also index the observations as  $A = 1, \dots, m$ .

Let us now describe the Bayesian setting used here to determine the information theoretic limits for reconstructing the hidden variables. From the

Bayes rule we have that the posterior given the observations is

$$\mathbb{P}(\boldsymbol{\sigma}|\mathbf{G}) = \frac{\prod_{i=1}^n \mathbb{P}_0(\sigma_i) \prod_{A=1}^m Q(G_A|\sigma_A)}{\sum_{\boldsymbol{\sigma} \in \{-1, +1\}^n} \prod_{i=1}^n \mathbb{P}_0(\sigma_i) \prod_{A=1}^m Q(G_A|\sigma_A)}.$$

Dividing both the numerator and denominator by  $\prod_{A=1}^m Q(G_A|\sigma_A = +1)$ , the posterior  $\mathbb{P}(\boldsymbol{\sigma}|\mathbf{G})$  can be rewritten as

$$\mathbb{P}(\boldsymbol{\sigma}|\mathbf{J}) = \frac{1}{\mathcal{Z}} \exp \left( \sum_{A=1}^m J_A (\sigma_A - 1) \right), \quad (4.1)$$

where

$$J_A \equiv \frac{1}{2} \ln \frac{Q(G_A|+1)}{Q(G_A|-1)},$$

$$\mathcal{Z} \equiv \sum_{\boldsymbol{\sigma} \in \{-1, +1\}^n} \exp \left( \sum_{A=1}^m J_A (\sigma_A - 1) \right).$$

We will use the language and notations of statistical mechanics. The normalization  $\mathcal{Z}$  shall be called the partition function. The bipartite factor graph  $\mathcal{G}$  underlying (4.1) contains variable nodes  $i = 1, \dots, n$  and constraint (or factor) nodes  $A = 1, \dots, m$ . Each variable node  $i$  “carries” the binary variable  $\sigma_i$  and each constraint node  $A$  “carries” the half-log-likelihood ratio  $J_A$  and uniformly connects to  $K$  variable nodes  $a_1, \dots, a_K$ . As said before, we identify  $A \equiv \{a_1, \dots, a_K\}$ . Distribution (4.1) can be interpreted as the Gibbs distribution of a *random* spin system (or spin glass). The expectation of a quantity  $A(\boldsymbol{\sigma})$  with respect to the posterior (4.1) will be denoted by a Gibbs bracket

$$\langle A(\boldsymbol{\sigma}) \rangle \equiv \sum_{\boldsymbol{\sigma} \in \{-1, +1\}^n} A(\boldsymbol{\sigma}) \mathbb{P}(\boldsymbol{\sigma}|\mathbf{J}).$$

The posterior distribution as well as the expectations  $\langle A(\boldsymbol{\sigma}) \rangle$  are random because of the randomness in: (i) the factor graph  $\mathcal{G}$  ensemble; (ii) the observations  $\mathbf{G}$  given the hidden vector  $\boldsymbol{\sigma}^0$ ; and (iii) the hidden vector  $\boldsymbol{\sigma}^0$ .

It is equivalent to work in terms of observations  $\mathbf{G}$  or associated half-log-likelihood ratios  $\mathbf{J}$ . The latter are (formally) distributed according to

$$\prod_{i=1}^n \mathbb{P}_0(\sigma_i^0) \prod_{A=1}^m c(J_A|\sigma_A^0) dJ_A \equiv \prod_{i=1}^n \mathbb{P}_0(\sigma_i^0) \prod_{A=1}^m Q(G_A|\sigma_A^0) dG_A. \quad (4.2)$$

Most of the time it will be more convenient for us to refer directly to half-log-likelihood ratios. The graph, the observations and the hidden vector are called *quenched* random variables (r.v.) because *given instance* of the problem their realization is *fixed*. In contrast, the r.v.  $\boldsymbol{\sigma}$  is sampled from the posterior 4.1, and hence is often called an annealed variable. Expectations with respect to the quenched variables are denoted  $\mathbb{E}_{\mathcal{G}}$  and  $\mathbb{E}_{\boldsymbol{\sigma}^0} \mathbb{E}_{\mathbf{J}|\boldsymbol{\sigma}^0}$ . To alleviate notations

we shall often simply use  $\mathbb{E}$  when the expectation is taken with respect to *all* quenched r.v in the ensuing expression. The bracket  $\langle - \rangle$  is reserved for expectations with respect to the posterior (4.1).

Let  $H(\boldsymbol{\sigma}|\mathbf{J}) \equiv -\sum_{\boldsymbol{\sigma} \in \{-1,+1\}^n} \mathbb{P}(\boldsymbol{\sigma}|\mathbf{J}) \ln \mathbb{P}(\boldsymbol{\sigma}|\mathbf{J})$  be the conditional entropy of the hidden variables given fixed observations. It is easy to see that the average conditional entropy (per variable) is given by the average *free entropy* (the r.h.s of the formula)

$$\frac{1}{n} \mathbb{E}_{\mathcal{G}} \mathbb{E}_{\boldsymbol{\sigma}^0} \mathbb{E}_{\mathbf{J}|\boldsymbol{\sigma}^0} H(\boldsymbol{\sigma}|\mathbf{J}) = \frac{1}{n} \mathbb{E}_{\mathcal{G}} \mathbb{E}_{\boldsymbol{\sigma}^0} \mathbb{E}_{\mathbf{J}|\boldsymbol{\sigma}^0} \ln \mathcal{Z}. \quad (4.3)$$

We refer readers to [69] for details. One can easily convert between the average mutual information and the average conditional entropy by the relation

$$\frac{1}{n} I(\boldsymbol{\sigma}; \mathbf{J}) = \ln 2 - \frac{1}{n} \mathbb{E}_{\mathcal{G}} \mathbb{E}_{\boldsymbol{\sigma}^0} \mathbb{E}_{\mathbf{J}|\boldsymbol{\sigma}^0} \ln \mathcal{Z}.$$

The singularities, as a function of the measurement fraction  $\alpha$ , of this limiting quantity when  $n \rightarrow +\infty$  give us the information-theoretic thresholds, or the location of *static phase transitions* in physics language.

#### 4.2.2 The replica symmetric formula for the average conditional entropy

The cavity method [27] predicts that the asymptotic average conditional entropy per variable is accessible from the following “replica symmetric” functional. This functional is an “average form” of the Bethe free entropy expression.

**Definition 4.1** (The replica symmetric free entropy functional). *Let  $V$  be a r.v. with distribution  $\mathbf{x}$ , and  $V_i$ ,  $i = 1, \dots, K$  i.i.d. copies of  $V$ . Let*

$$U = \tanh^{-1} \left( \tanh J \prod_{i=1}^{K-1} \tanh V_i \right), \quad (4.4)$$

*and  $U_B$ ,  $B = 1, \dots, l$  i.i.d. copies of  $U$  where  $l \sim \text{Poi}(\alpha K)$  is a Poisson distributed integer. Let  $\sigma^0 \sim \mathbb{P}_0$  and  $\prod_{a=1}^K \sigma_a$  be the product of  $K$  independent copies. Let  $J \sim \mathbf{c}(J | \prod_{a=1}^K \sigma_a)$  (see equation (4.2)). The replica symmetric free entropy functional is defined to be*

$$\begin{aligned} h_{\text{RS}}(\mathbf{x}) \equiv & \mathbb{E}_l \mathbb{E}_{\sigma_1, \dots, \sigma_K} \mathbb{E}_{J | \prod_{a=1}^K \sigma_a} \mathbb{E}_U \mathbb{E}_V \left[ \ln \left( \prod_{B=1}^l (1 + \tanh U_B) + \prod_{B=1}^l (1 - \tanh U_B) \right) \right. \\ & \left. - \alpha(K-1) \ln \left( 1 + \tanh J \prod_{i=1}^K \tanh V_i \right) - \alpha \ln(1 + \tanh J) \right]. \end{aligned} \quad (4.5)$$

**Remark 4.1.** *For uniform  $\mathbb{P}_0$  we can replace the product  $\prod_{a=1}^K \sigma_a$  by a single binary variable  $\sigma_0 \sim \mathbb{P}_0$ .*



While a substantial part of our analysis holds for general (symmetric) memoryless channels, our main result is fully proved for the BEC. This channel has transition probability

$$Q(G_A|\sigma_A^0) = (1 - q)\delta_{G_A, \sigma_A^0} + q\delta_{G_A, 0},$$

and from (4.2) we get in this case

$$c(J_A|\sigma_A^0) = (1 - q)\delta_{\sigma_A^0 J_A, +\infty} + q\delta_{J_A, 0}.$$

The set of distributions with point masses at  $\{0, +\infty\}$  plays a special role and will be called  $\mathfrak{B}$ . We adopt the notation (from coding theory)  $\Delta_0$  and  $\Delta_\infty$  for the two point masses at 0 and  $+\infty$ . Any distribution  $\mathbf{x} \in \mathfrak{B}$  is of the form  $\mathbf{x} = x\Delta_0 + (1 - x)\Delta_\infty$ , with  $x \in [0, 1]$ . In this case the replica symmetric free entropy functional becomes (4.5) becomes a function of  $x \in [0, 1]$ . A numerical illustration is found in Appendix 4.8.8.

Our main result is the proof, through the use of the adaptive interpolation method for sparse graphs, of the following theorem:

**Theorem 4.1** (The replica symmetric formula is exact for the BEC channel). *For a censored block model with observations obtained through a binary erasure channel as described above, we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{\mathcal{G}} \mathbb{E}_{\sigma^0} \mathbb{E}_{\mathbf{J}|\sigma^0} H(\sigma|\mathbf{J}) = \sup_{\mathbf{x} \in \mathfrak{B}} h_{\text{RS}}(\mathbf{x}). \quad (4.6)$$

This theorem is a direct consequence of two main Propositions 4.1 and 4.2 proved in Sec. 4.4.

## 4.3 Two preliminary tools

In this section we review standard material which is needed in our analysis. For more details the reader can consult [24, 25, 70].

### 4.3.1 Nishimori identities

#### Nishimori identities for symmetric channels

Consider the quantity  $\prod_{S \in \mathcal{C}} \sigma_S^0 \prod_{S \in \mathcal{C}} \langle \sigma_S \rangle$  for a given graph and any collection  $\mathcal{C}$  of subsets  $S \subset \{1, \dots, n\}$ . The same subset can occur many times in a collection. Using the map  $(\mathbf{X}, \mathbf{Y}, \tilde{\mathbf{Y}}) \rightarrow (\sigma^0, \mathbf{J}, \emptyset)$  on (2.18), we obtain

$$\mathbb{E}_{\sigma^0} \mathbb{E}_{\mathbf{J}|\sigma^0} \left[ \prod_{S \in \mathcal{C}} \sigma_S^0 \prod_{S \in \mathcal{C}} \langle \sigma_S \rangle \right] = \mathbb{E}_{\sigma^0} \mathbb{E}_{\mathbf{J}|\sigma^0} \left[ \left\langle \prod_{S \in \mathcal{C}} \sigma_S \right\rangle \prod_{S \in \mathcal{C}} \langle \sigma_S \rangle \right]. \quad (4.7)$$

For symmetric channels, this identity can be further specialized. This is specially important for us since the BEC is a symmetric channel. By definition, symmetric channels are those satisfying  $Q(G_A|\sigma_A^0) = Q(-G_A|-\sigma_A^0)$  or equivalently  $\mathbf{c}(J_A|\sigma_A^0) = \mathbf{c}(-J_A|-\sigma_A^0)$ .

Given  $\sigma^0$  the Gibbs distribution (4.1) is *invariant* under the *gauge transformation*  $\sigma_i \rightarrow \sigma_i^0 \sigma_i$ ,  $J_A \rightarrow \sigma_A^0 J_A$ . Let us denote by  $\sigma^0 \star J$  the “component-wise” product  $(\sigma_A^0 J_A)_{A=1}^m$ . Now, we perform a gauge transformation on both sides of (4.7). For the left hand side we have

$$\mathbb{E}_{J|\sigma^0} \left[ \prod_{S \in \mathcal{C}} \sigma_S^0 \prod_{S \in \mathcal{C}} \langle \sigma_S \rangle \right] = \mathbb{E}_{\sigma^0 \star J|\sigma^0} \left[ \prod_{S \in \mathcal{C}} \langle \sigma_S \rangle \right]. \quad (4.8)$$

Moreover, from  $\mathbf{c}(J_A|\sigma_A^0) = \mathbf{c}(-J_A|-\sigma_A^0)$  one can see that for a *symmetric channel*  $\mathbf{c}(\sigma_A^0 J_A|\sigma_A^0) = \mathbf{c}(J_A|1)$ , and therefore in (4.8) we can replace  $\mathbb{E}_{\sigma^0 \star J|\sigma^0}$  by  $\mathbb{E}_{J|1}$ . We get

$$\mathbb{E}_{J|\sigma^0} \left[ \prod_{S \in \mathcal{C}} \sigma_S^0 \prod_{S \in \mathcal{C}} \langle \sigma_S \rangle \right] = \mathbb{E}_{J|1} \left[ \prod_{S \in \mathcal{C}} \langle \sigma_S \rangle \right]. \quad (4.9)$$

The same steps show that the right hand side of (4.7) also satisfies

$$\mathbb{E}_{J|\sigma^0} \left[ \left\langle \prod_{S \in \mathcal{C}} \sigma_S \right\rangle \prod_{S \in \mathcal{C}} \langle \sigma_S \rangle \right] = \mathbb{E}_{J|1} \left[ \left\langle \prod_{S \in \mathcal{C}} \sigma_S \right\rangle \prod_{S \in \mathcal{C}} \langle \sigma_S \rangle \right]. \quad (4.10)$$

From (4.9), (4.10), (4.7) we get the final Nishimori identity

$$\mathbb{E}_{J|1} \left[ \prod_{S \in \mathcal{C}} \langle \sigma_S \rangle \right] = \mathbb{E}_{J|1} \left[ \left\langle \prod_{S \in \mathcal{C}} \sigma_S \right\rangle \prod_{S \in \mathcal{C}} \langle \sigma_S \rangle \right]. \quad (4.11)$$

### A special Nishimori identity for symmetric distributions

An important role is played by the space  $\mathfrak{X}$  of *symmetric distributions* which we define as follows. Take a transition probability (a “channel”) satisfying  $q(G|\sigma^0) = q(-G|-\sigma^0)$ ,  $\sigma^0 \in \{-1, +1\}$ ,  $G \in \mathbb{R}$ . The associated half-log-likelihood variable is  $h = \frac{1}{2} \ln \frac{q(G|+1)}{q(G|-1)}$ . The space  $\mathfrak{X}$  is the space of symmetric distributions over the half-log-likelihood variable is formally defined by  $\mathbf{x}(dh) = q(G|+1)dJ$ . It is easy to deduce from  $q(G|\sigma^0) = q(-G|-\sigma^0)$  that a symmetric distribution  $\mathbf{x} \in \mathfrak{X}$  satisfies  $\mathbf{x}(-dh) = e^{-2h}\mathbf{x}(dh)$ . We note that  $\mathfrak{B} \subset \mathfrak{X}$  (recall  $\mathfrak{B}$  is the set of convex combinations of point masses at 0 and  $+\infty$ ).

There is an important special case of the Nishimori identity (4.11). Namely the one satisfied by the system constituted by a single uniform hidden variable  $\sigma^0 \sim P_0$ , observed through a noisy “channel”  $q(G|\sigma^0)$ . The Gibbs distribution is simply in this case  $e^{h\sigma}/(2 \cosh h)$  where  $h$  is the half-log-likelihood of the “channel”. Since  $\langle \sigma \rangle = \tanh h$ , an application of (4.11) (where the singleton set is taken  $k$  times) yields

$$\int (\tanh h)^{2k-1} \mathbf{x}(dh) = \int (\tanh h)^{2k} \mathbf{x}(dh), \quad k \in \mathbb{N}^*. \quad (4.12)$$

In Appendix 4.8.1 we show an independent and direct way that any  $\mathbf{x} \in \mathfrak{X}$  satisfies (4.12).

### Conditional entropy for symmetric channels

Since the Gibbs distribution is invariant under a gauge transformation, the partition function  $\mathcal{Z}$  also is, and therefore, for a given graph  $\mathcal{G}$  and hidden vector  $\sigma^0$ , we have

$$\mathbb{E}_{\mathbf{J}|\sigma^0} \ln \mathcal{Z} = \mathbb{E}_{\sigma^0 * \mathbf{J}|\sigma^0} \ln \mathcal{Z}. \quad (4.13)$$

For symmetric channels the r.h.s. equals  $\mathbb{E}_{\mathbf{J}|1} \ln \mathcal{Z}$  and thus the average conditional entropy 4.3 becomes

$$\frac{1}{n} \mathbb{E}_{\mathcal{G}} \mathbb{E}_{\sigma^0} \mathbb{E}_{\mathbf{J}|\sigma^0} H(\sigma|\mathbf{J}) = \frac{1}{n} \mathbb{E}_{\mathcal{G}} \mathbb{E}_{\mathbf{J}|1} \ln \mathcal{Z}. \quad (4.14)$$

### Summary

When one is dealing with symmetric measurement channels, in order to compute the average conditional entropy, or *certain* averages, one may assume that  $\sigma_i^0 = 1$ ,  $i = 1, \dots, n$  and that the quenched variables  $\mathbf{J}$  have distribution  $\mathbf{c}(J_A|1)$ ,  $A = 1, \dots, m$ . From now on this is understood unless explicitly specified otherwise.

### 4.3.2 Griffiths-Kelly-Sherman inequalities for the BEC

The BEC is a symmetric channel so as shown before, without loss of generality for analysis purposes, we assume  $\sigma_i^0 = 1$ ,  $i = 1, \dots, n$  and that  $\mathbf{J}$  have distribution  $\mathbf{c}(J_A|1)$ ,  $A = 1, \dots, m$ . Since  $\mathbf{c}(J_A|1) = (1-q)\Delta_\infty + q\Delta_0$  the Gibbs distribution (4.1) has non-negative coupling constants  $J_A$ ,  $A = 1, \dots, m$ . Therefore the Gibbs distribution satisfies the GKS inequalities in (1.9) and (1.10). These two inequalities play an important role in the proof of Theorem 4.1.

## 4.4 The adaptive path interpolation method

For  $t = 1, \dots, T$  let  $V_i^{(t)}$  be i.i.d. r.v. distributed according to  $\mathbf{x}^{(t)} \in \mathfrak{X}$ . Consider the r.v.

$$U^{(t)} = \tanh^{-1} \left( \tanh J \prod_{i=1}^{K-1} V_i^{(t)} \right) \quad (4.15)$$

and independent copies denoted  $U_B^{(t)}$  where  $B$  is a subscript which runs over  $l^{(t)} \sim \text{Poi}(\frac{K}{RT})$  of these copies. Later on, we call  $\tilde{\mathbf{x}}^{(t)}$  the distribution of  $U^{(t)}$  (induced by  $\mathbf{x}^{(t)}$  and  $\mathbf{c}$ ).

Let also define two extra random variables,  $H$  with distribution  $\epsilon\Delta_\infty + (1-\epsilon)\Delta_0 \in \mathfrak{B}$ , and  $\tilde{H}$  with distribution  $\delta n^{-\theta}\Delta_\infty + (1-\delta n^{-\theta})\Delta_0 \in \mathfrak{B}$ , where  $\epsilon, \delta \in (0, 1)$  and  $\theta \in (0, 1)$  (eventually we will have to take  $\theta \in (0, 1/5]$  in the final estimates).

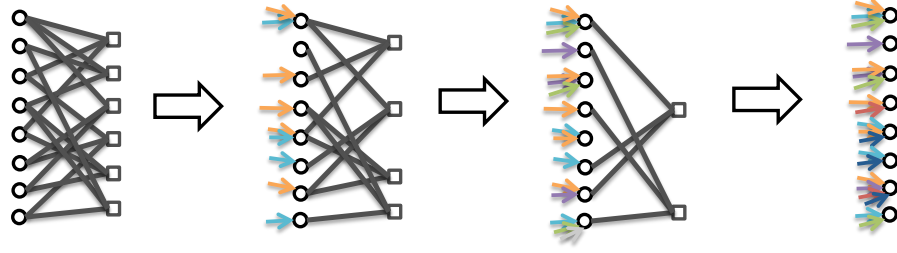


Figure 4.1 – Interpolation for sparse graphs

Let us set  $\mathbf{x} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)})$ . We define the *generalized free entropy functional*:

$$\begin{aligned} \tilde{h}_{\epsilon, \delta}(\mathbf{x}) \equiv & \mathbb{E} \left[ \ln \left( \prod_{t=1}^T \prod_{B=1}^{l^{(t)}} (1 + \tanh U_B^{(t)}) + e^{-2(H+\tilde{H})} \prod_{t=1}^T \prod_{B=1}^{l^{(t)}} (1 - \tanh U_B^{(t)}) \right) \right. \\ & \left. - \frac{\alpha(K-1)}{T} \sum_{t=1}^T \ln \left( 1 + \tanh J \prod_{i=1}^K \tanh V_i^{(t)} \right) - \alpha \ln(1 + \tanh J) \right]. \end{aligned} \quad (4.16)$$

One can easily check that if  $\mathbf{x}^{(t)} = \mathbf{x}$  for all  $t$ , then  $\tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x}) = h_{\text{RS}}(\mathbf{x})$ . More is true as the following lemma shows:

**Lemma 4.1.** *Let  $\mathcal{X}^T = \mathfrak{X} \times \mathfrak{X} \times \dots \times \mathfrak{X}$ . We have for  $\mathbf{x} \in \mathfrak{X}^T$*

$$\sup_{\mathbf{x} \in \mathcal{X}^T} \tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x}) = \sup_{\mathbf{x} \in \mathcal{X}} h_{\text{RS}}(\mathbf{x}). \quad (4.17)$$

**Remark 4.2.** *We prove this lemma in Sec. 4.8.3. For distributions in  $\mathbf{x} \in \mathfrak{B}^T$  the supremum carries over  $(x^{(1)}, \dots, x^{(T)}) \in [0, 1]^T$  and the proof only requires real analysis.*

#### 4.4.1 The $(t, s)$ –interpolating model

Consider the construction of an interpolating factor graph ensemble  $\mathcal{G}_{t,s}$  involving discrete and a continuous interpolation parameters,  $t \in \{1, 2, \dots, T\}$  and  $s \in [0, 1]$ . This is the sparse graph counterpart of the interpolating ensemble initially developed for dense graphs in [47] (and the simplified in [48]).

Fig. 4.1 illustrates the interpolating factor graph  $\mathcal{G}_{t,s}$ . As time  $t$  proceeds by a unit, a set of check nodes of random size is removed, and a new set of messages also of random size is added with its  $t$ -dependence labeled by color. The interpolating graph is designed such that  $\mathcal{G}_{t,1}$  is statistically equivalent to  $\mathcal{G}_{t+1,0}$ ; in addition,  $\mathcal{G}_{t,s}$  maintains the degree distribution of variable nodes invariant: For any  $(t, s)$  the degree of each variable node is an independent

**Algorithm 1** Construction of  $\mathcal{G}_{t,s}$ 


---

```

for  $i = 1, \dots, n$  do
  for  $t' = 1, \dots, t - 1$  do
    _draw a random number  $e_i^{(t')} \sim \text{Poi}(\frac{\alpha K}{T})$ 
    for  $B = 1, \dots, e_i^{(t')}$  do
      _connect variable node  $i$  with a half edge and assign a weight
       $U_{B \rightarrow i}^{(t')} \sim \tilde{x}^{(t')}$  to this half-edge
    _draw a random number  $e_{i,s}^{(t)} \sim \text{Poi}(\frac{\alpha K s}{T})$ 
    for  $C = 1, \dots, e_{i,s}^{(t)}$  do
      _connect variable node  $i$  with a half edge and assign a weight  $U_{C \rightarrow i}^{(t)} \sim$ 
       $\tilde{x}^{(t)}$  to this half-edge

  _draw a random number  $m_s^{(t)} \sim \text{Poi}(\frac{\alpha n(T-t+1-s)}{T})$ 
  for  $A = 1, \dots, m_s^{(t)}$  do
    _assign to factor node  $A$  a r.v  $J_A \sim c$ 
    _uniformly and randomly connect factor node  $A$  to  $K$  variable nodes (this
    subset of variable nodes is also denoted  $A$  by a slight abuse of notation)

```

---

$\text{Poi}(K/R)$  random variable. The Hamiltonian associated with  $\mathcal{G}_{t,s}$  is

$$\begin{aligned}
 \mathcal{H}_{t,s}(\boldsymbol{\sigma}, \mathbf{J}, \mathbf{U}, \mathbf{m}, \mathbf{e}) = & - \sum_{i=1}^n \left\{ \sum_{t'=1}^{t-1} \sum_{B=1}^{e_i^{(t')}} U_{B \rightarrow i}^{(t')} + \sum_{C=1}^{e_{i,s}^{(t)}} U_{C \rightarrow i}^{(t)} \right\} (\sigma_i - 1) \\
 & - \sum_{A=1}^{m_s^{(t)}} J_A (\sigma_A - 1). \tag{4.18}
 \end{aligned}$$

We further consider a generalized version of (4.18) by adding two kinds of perturbations that can be interpreted as small additional observations from side-channels for each node  $i = 1, \dots, n$ . These perturbations are then removed at the end of the analysis. let  $H_i$  and  $\tilde{H}_i$  be half-log-likelihood variables, where  $H_i$  and  $\tilde{H}_i$  have the same distribution as  $H$  and  $\tilde{H}$  defined at the beginning of this section. Our final *interpolating Hamiltonian* is

$$\mathcal{H}_{t,s;\epsilon,\delta}(\boldsymbol{\sigma}, \mathbf{J}, \mathbf{U}, \mathbf{m}, \mathbf{e}, \mathbf{H}, \tilde{\mathbf{H}}) \equiv \mathcal{H}_{t,s}(\boldsymbol{\sigma}, \mathbf{J}, \mathbf{U}, \mathbf{m}, \mathbf{e}) - \sum_{i=1}^n (H_i + \tilde{H}_i)(\sigma_i - 1). \tag{4.19}$$

The associated interpolating partition function, Gibbs expectation and free

entropy are:

$$\mathcal{Z}_{t,s;\epsilon,\delta} \equiv \sum_{\sigma \in \{-1,+1\}^n} e^{-\mathcal{H}_{t,s;\epsilon,\delta}(\sigma, \mathbf{J}, \mathbf{U}, \mathbf{m}, \mathbf{e}, \mathbf{H}, \tilde{\mathbf{H}})}, \quad (4.20)$$

$$\langle A(\sigma) \rangle_{t,s;\epsilon,\delta} \equiv \frac{1}{\mathcal{Z}_{t,s;\epsilon,\delta}} \sum_{\sigma \in \{-1,+1\}^n} A(\sigma) e^{-\mathcal{H}_{t,s;\epsilon,\delta}(\sigma, \mathbf{J}, \mathbf{U}, \mathbf{m}, \mathbf{e}, \mathbf{H}, \tilde{\mathbf{H}})}, \quad (4.21)$$

$$h_{t,s;\epsilon,\delta} \equiv \frac{1}{n} \mathbb{E} \ln \mathcal{Z}_{t,s;\epsilon,\delta}, \quad (4.22)$$

$$H_{t,s;\epsilon,\delta} \equiv \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}} \ln \mathcal{Z}_{t,s;\epsilon,\delta}. \quad (4.23)$$

Recall our notation: The expectation  $\mathbb{E}$  here carries over *all* quenched variables entering in the interpolating system, thus  $\mathbf{J}, \mathbf{U}, \mathbf{m}, \mathbf{e}, \mathbf{H}$  and  $\tilde{\mathbf{H}}$ . Note that Nishimori's identity (4.11) and GKS inequalities (1.9), (1.10) still apply to the Gibbs expectation  $\langle - \rangle_{t,s;\epsilon,\delta}$ .

One may check that, at the initial point of the interpolating path ( $t = 1, s = 0$ ) the free entropy  $h_{1,0;\epsilon=0,\delta=0}$  is equal to the averaged conditional entropy of the original model (see formula (4.27) below), and at the end-point ( $t = T, s = 1$ ) the free entropy  $h_{T,1;\epsilon,\delta}$  is given by a part of the generalized entropy functional (4.16) (see formula (4.45)).

The connection between the unperturbed and perturbed free entropies is given by (see Sec. 4.8.4)

**Lemma 4.2.** *Let  $\mathbf{c} \in \mathfrak{B}$  and  $\mathbf{x} \in \mathfrak{B}^T$ . We have*

$$|h_{t,s;\epsilon,\delta} - h_{t,s;\epsilon=0,\delta=0}| \leq (\epsilon + \frac{\delta}{n^\theta}) \ln 2, \quad (4.24)$$

$$|\tilde{h}_{\epsilon,\delta}(\mathbf{x}) - \tilde{h}_{\epsilon=0,\delta=0}(\mathbf{x})| \leq (\epsilon + \frac{\delta}{n^\theta}) \ln 2. \quad (4.25)$$

#### 4.4.2 Evaluating the free entropy change along the interpolation

By interpolating  $h_{t,s;\epsilon,\delta}$  from the initial state ( $t = 1, s = 0$ ) to the final one ( $t = T, s = 1$ ), we have

$$h_{1,0;\epsilon,\delta} = h_{T,1;\epsilon,\delta} + \sum_{t=1}^T (h_{t,0;\epsilon,\delta} - h_{t,1;\epsilon,\delta}) = h_{T,1;\epsilon,\delta} - \sum_{t=1}^T \int_0^1 ds \frac{dh_{t,s;\epsilon,\delta}}{ds}. \quad (4.26)$$

We have  $m_{s=0}^{(t=1)} = m \sim \text{Poi}(\alpha n)$  and thus

$$\mathcal{H}_{1,0;\epsilon,\delta} = - \sum_{A=1}^m J_A (\sigma_A - 1) - \sum_{i=1}^n (H_i + \tilde{H}_i) (\sigma_i - 1).$$

Therefore, the initial interpolating free entropy without perturbation equals the average conditional entropy per variable:

$$h_{1,0;\epsilon=0,\delta=0} = \frac{1}{n} \mathbb{E} H(\sigma | \mathbf{J}). \quad (4.27)$$

On the other hand  $h_{T,1;\epsilon,\delta}$  corresponds to a part of the generalized free entropy functional (4.16). A subsequent computation (see Sec. 4.5) on (4.26) leads to the *fundamental sum rule*

$$h_{1,0;\epsilon,\delta} = \tilde{h}_{\epsilon,\delta}(\mathbf{x}) + \frac{\alpha}{T} \sum_{t=1}^T \int_0^1 ds \mathcal{R}_{t,s;\epsilon,\delta} \quad (4.28)$$

where

$$\mathcal{R}_{t,s;\epsilon,\delta} = \sum_{p=1}^{\infty} \frac{\mathbb{E}[(\tanh J)^{2p}]}{2p(2p-1)} \mathbb{E} \langle Q_{2p}^K - K(q_{2p}^{(t)})^{K-1}(Q_{2p} - q_{2p}^{(t)}) - (q_{2p}^{(t)})^K \rangle_{t,s;\epsilon,\delta} \quad (4.29)$$

with

$$Q_p \equiv \frac{1}{n} \sum_{i=1}^n \sigma_i^{(1)} \cdots \sigma_i^{(p)}$$

the *overlap* of  $p$  independent *replicas*  $\boldsymbol{\sigma}^{(1)}, \dots, \boldsymbol{\sigma}^{(p)}$  and

$$q_p^{(t)} \equiv \mathbb{E}[(\tanh V^{(t)})^p].$$

In (4.29) the Gibbs average  $\langle - \rangle_{t,s;\epsilon,\delta}$  over a polynomial of  $Q_p$  must be understood as an average over the product measure

$$\prod_{\alpha=1}^p \frac{1}{\mathcal{Z}_{t,s;\epsilon,\delta}} e^{-\mathcal{H}_{t,s;\epsilon,\delta}(\boldsymbol{\sigma}^{(\alpha)}, \mathbf{J}, \mathbf{U}, \mathbf{m}, \mathbf{e}, \mathbf{H}, \tilde{\mathbf{H}})}$$

where the quenched variables have the same realization for all replicas. We still denote this Gibbs average by  $\langle - \rangle_{t,s;\epsilon,\delta}$  for simplicity.

### 4.4.3 The lower bound

In order to show the lower bound we need the following important concentration lemma (proven in Sec. 4.6), which is at the core of the “replica symmetric” behavior of the model:

**Lemma 4.3** (Concentration of  $Q_p^K$  on  $\langle Q_p \rangle_{t,s;\epsilon,\delta}^K$ ). *For any  $\mathbf{c} \in \mathfrak{B}$ ,  $\mathbf{x} \in \mathfrak{B}^T$  we have*

$$\int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \mathbb{E} \langle |Q_p^K - \langle Q_p \rangle_{t,s;\epsilon,\delta}^K| \rangle_{t,s;\epsilon,\delta} \leq K \left( \frac{3p(\varepsilon_1 - \varepsilon_0)}{n} \right)^{1/2}. \quad (4.30)$$

**Proposition 4.1** (Lower bound). *For  $\mathbf{c} \in \mathfrak{B}$  we have*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} H(\boldsymbol{\sigma} | \mathbf{J}) \geq \sup_{\mathbf{x} \in \mathfrak{B}} h_{\text{RS}}(\mathbf{x}). \quad (4.31)$$

**Remark 4.3.** *The methods of this chapter can be extended to show this proposition for  $\mathbf{c}, \mathbf{x} \in \mathfrak{X}$ .*

*Proof.* Eq. (4.28) implies

$$\begin{aligned} h_{1,0;\epsilon=0,\delta=0} &= \tilde{h}_{\epsilon=0,\delta=0}(\mathbf{x}) + \frac{\alpha}{T} \sum_{t=1}^T \int_0^1 ds \mathcal{R}_{t,s;\epsilon,\delta}(\mathbf{x}) \\ &\quad + (\tilde{h}_{\epsilon,\delta}(\mathbf{x}) - \tilde{h}_{\epsilon=0,\delta=0}(\mathbf{x})) - (h_{1,0;\epsilon,\delta} - h_{1,0;\epsilon=0,\delta=0}). \end{aligned} \quad (4.32)$$

We fix  $\delta = 0$ . From (4.29) we have  $\mathcal{R}_{t,s;\epsilon,\delta=0}$  equal to

$$\begin{aligned} \sum_{p=1}^{\infty} \frac{\mathbb{E}[(\tanh J)^{2p}]}{2p(2p-1)} &\left( \mathbb{E} \left[ \langle Q_{2p} \rangle_{t,s;\epsilon,0}^K - K(q_{2p}^{(t)})^{K-1} (\langle Q_{2p} \rangle_{t,s;\epsilon,0} - q_{2p}^{(t)}) - (q_{2p}^{(t)})^K \right] \right. \\ &\quad \left. - \mathbb{E} \langle (Q_p^K - \langle Q_p \rangle_{t,s;\epsilon,0}^K) \rangle_{t,s;\epsilon,0} \right). \end{aligned}$$

Note that the convexity  $x \mapsto x^K$  for  $x \in \mathbb{R}_+$  implies  $x^K - y^K \geq Ky^{K-1}(x-y) \geq 0$  for any  $x, y \in \mathbb{R}_+$ . As  $\langle Q_{2p} \rangle_{t,s;\epsilon,0} = \frac{1}{n} \sum_{i=1}^n \langle \sigma_i \rangle_{t,s;\epsilon,0}^{2p} \geq 0$  and  $q_{2p}^{(t)} \geq 0$ ,

$$\langle Q_{2p} \rangle_{t,s;\epsilon,0}^K - K(q_{2p}^{(t)})^{K-1} (\langle Q_{2p} \rangle_{t,s;\epsilon,0} - q_{2p}^{(t)}) - (q_{2p}^{(t)})^K \geq 0.$$

Thus with Lemma 4.3 we obtain

$$\frac{1}{\epsilon_n} \int_{\epsilon_n}^{2\epsilon_n} d\epsilon \mathcal{R}_{t,s;\epsilon,0} \geq -(\ln 2) K \left( \frac{3p}{\epsilon_n n} \right)^{1/2}. \quad (4.33)$$

Now we average both side of (4.32) over  $\epsilon \in [\epsilon_n, 2\epsilon_n]$  for some sequence  $\epsilon_n$  specified at the end. Using (4.33) and Lemma 4.2

$$h_{1,0;\epsilon=0,\delta=0} \geq \tilde{h}_{\epsilon=0,\delta=0}(\mathbf{x}) + \mathcal{O}\left(\frac{1}{\sqrt{\epsilon_n n}}\right) + \mathcal{O}(\epsilon_n).$$

Choosing  $\epsilon_n = n^{-\gamma}$  with  $0 < \gamma < 1$ , we conclude

$$\liminf_{n \rightarrow \infty} h_{1,0;\epsilon=0,\delta=0} \geq \tilde{h}_{\epsilon=0,\delta=0}(\mathbf{x}).$$

Finally one can take the supremum of the right hand side and use (4.17) as well as (4.27) to obtain (4.31).  $\square$

#### 4.4.4 The upper bound

In this paragraph we crucially use the specialties of the BEC. We take interpolating paths  $\mathbf{x} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)}) \in \mathfrak{B}^T$ , where  $\mathbf{x}^{(t)} = x^{(t)} \Delta_0 + (1 - x^{(t)}) \Delta_\infty$ . In particular we use the following lemma (proven in Sec. 4.8.4):

**Lemma 4.4.** *For any  $\mathbf{c} \in \mathfrak{B}$ , and any interpolation path  $\mathbf{x}(\epsilon) \in \mathfrak{B}^T$  depending on  $\epsilon$ , and any  $A \subseteq \{1, \dots, n\}$  we have  $\langle \sigma_A \rangle_{t,s;\epsilon,\delta} \in \{0, 1\}$ .*



Notice that  $\langle Q_p^k \rangle_{t,s;\epsilon,\delta} = \frac{1}{n^k} \sum_{i_1, \dots, i_k=1}^n \langle \sigma_{i_1} \cdots \sigma_{i_k} \rangle_{t,s;\epsilon,\delta}^p$  for any  $k \in \mathbb{N}$ . Lemma 4.4 then implies

$$\langle Q_p^k \rangle_{t,s;\epsilon,\delta} = \langle Q_1^k \rangle_{t,s;\epsilon,\delta}$$

for all  $p \in \mathbb{N}^*$ . We also have  $\tanh V^{(t)} \in \{0, 1\}$  because  $\mathbf{x}^{(t)} \in \mathfrak{B}$  and thus

$$q_p^{(t)} = q_1^{(t)} \quad \forall p \in \mathbb{N}^*.$$

Finally recall that  $\mathbf{c}(J|1) = (1 - q)\Delta_\infty + q\Delta_0$ , therefore  $\mathbb{E}[(\tanh J)^{2p}] = 1 - q$ . These facts reduce (4.29) to

$$\mathcal{R}_{t,s;\epsilon,\delta} = (1 - q)(\ln 2)\mathbb{E}\left[\langle Q_1^K \rangle_{t,s;\epsilon,\delta} - K(q_1^{(t)})^{K-1}(\langle Q_1 \rangle_{t,s;\epsilon,\delta} - q_1^{(t)}) - (q_1^{(t)})^K\right]. \quad (4.34)$$

We then split the remainder as follows:

$$\begin{aligned} \mathcal{R}_{t,s;\epsilon,\delta} &= (\mathcal{R}_{t,s;\epsilon,\delta} - \mathcal{R}_{t,0;\epsilon,\delta}) + \mathcal{R}_{t,0;\epsilon,\delta} \\ &= (\mathcal{R}_{t,s;\epsilon,\delta} - \mathcal{R}_{t,0;\epsilon,\delta}) + (1 - q)(\ln 2)\left(\mathbb{E}[\langle Q_1 \rangle_{t,0;\epsilon,0}]^K \right. \\ &\quad \left. - K(q_1^{(t)})^{K-1}(\mathbb{E}\langle Q_1 \rangle_{t,0;\epsilon,0} - q_1^{(t)}) - (q_1^{(t)})^K\right) \\ &\quad + (1 - q)(\ln 2)(\mathbb{E}\langle Q_1^K \rangle_{t,0;\epsilon,\delta} - \mathbb{E}[\langle Q_1 \rangle_{t,0;\epsilon,\delta}]^K) \\ &\quad + (1 - q)(\ln 2)(\mathbb{E}[\langle Q_1 \rangle_{t,0;\epsilon,\delta}]^K - \mathbb{E}[\langle Q_1 \rangle_{t,0;\epsilon,0}]^K \\ &\quad \left. - K(q_1^{(t)})^{K-1}(\mathbb{E}[\langle Q_1 \rangle_{t,0;\epsilon,\delta}] - \mathbb{E}[\langle Q_1 \rangle_{t,0;\epsilon,0}])\right), \end{aligned} \quad (4.35)$$

and treat each part thanks to the three following lemmas. Lemma 4.5 is proven in Sec. 4.8.4, Lemma 4.6 in Sec. 4.6, and Lemma 4.7 in Sec. 4.8.4).

**Lemma 4.5** (Weak  $s$ -dependence at fixed  $t$ ). *For any  $k \in \mathbb{N}$  and  $s \in [0, 1]$  we have*

$$|\mathbb{E}\langle Q_1^k \rangle_{t,s;\epsilon,\delta} - \mathbb{E}\langle Q_1^k \rangle_{t,0;\epsilon,\delta}| \leq \frac{2\alpha(K+1)n}{T}. \quad (4.36)$$

**Lemma 4.6** (Concentration of  $\langle Q_1 \rangle_{t,s;\epsilon,\delta}^K$  on  $\mathbb{E}[\langle Q_1 \rangle_{t,s;\epsilon,\delta}]^K$ ). *For any  $\mathbf{c} \in \mathfrak{B}$  and  $\mathbf{x}(\epsilon) \in \mathfrak{B}^T$  such that every component satisfies  $dx^{(t)}/d\epsilon \geq 0$  we have for  $\theta \in (0, 1/5]$*

$$\begin{aligned} &\int_{\delta_0}^{\delta_1} d\delta \int_{\epsilon_0}^{\epsilon_1} d\epsilon \mathbb{E}[|\langle Q_1 \rangle_{t,s;\epsilon,\delta}^K - \mathbb{E}[\langle Q_1 \rangle_{t,s;\epsilon,\delta}]^K|] \\ &= \mathcal{O}\left(\left((\delta_1 - \delta_0)(\epsilon_1 - \epsilon_0) \frac{\delta_1 - \delta_0 + \epsilon_1 - \epsilon_0}{n^\theta}\right)^{1/2}\right). \end{aligned} \quad (4.37)$$

**Lemma 4.7.** *For any  $\mathbf{c} \in \mathfrak{B}$  and  $\mathbf{x}(\epsilon) \in \mathfrak{B}^T$  such that every component satisfies  $dx^{(t)}/d\epsilon \geq 0$  we have*

$$\begin{aligned} &\left| \int_{\delta_0}^{\delta_1} d\delta \int_{\epsilon_0}^{\epsilon_1} d\epsilon \left\{ \mathbb{E}[\langle Q_1 \rangle_{t,0;\epsilon,\delta}]^K - \mathbb{E}[\langle Q_1 \rangle_{t,0;\epsilon,0}]^K \right. \right. \\ &\quad \left. \left. - K(q_1^{(t)})^{K-1}(\mathbb{E}[\langle Q_1 \rangle_{t,0;\epsilon,\delta}] - \mathbb{E}[\langle Q_1 \rangle_{t,0;\epsilon,0}]) \right\} \right| \leq \frac{3K(\delta_1^2 - \delta_0^2)}{n^\theta(1 - \delta_1/n^\theta)}. \end{aligned} \quad (4.38)$$

where  $\theta \in (0, 1/5]$ .

Now we look into each term of (4.35). Lemma 4.5 and (4.34) imply

$$|\mathcal{R}_{t,s;\epsilon,\delta} - \mathcal{R}_{t,0;\epsilon,\delta}| \leq \frac{2(\ln 2)\alpha(K+1)^2n}{T} = \mathcal{O}\left(\frac{n}{T}\right).$$

Since  $T$  is a free parameter (controlling the mean of  $e_{i,s}^{(t)}$  and  $m_s^{(t)}$ ) we can set it significantly larger than  $n$ . The first term of (4.35) thus can be neglected and it is sufficient to work with  $\mathcal{R}_{t,0;\epsilon,\delta}$ . This separation is important because we use that  $\mathbb{E}\langle Q_1 \rangle_{t,0;\epsilon,0}$  (in the second term of (4.35)) is independent of  $\{\mathbf{x}^{(t')}\}_{t' \geq t}$ . Also recall  $q_1^{(t)} \equiv \mathbb{E} \tanh V^{(t)}$ . This allows us to sequentially choose a distribution  $\hat{\mathbf{x}}_n^{(t)}$  for  $V^{(t)}$  along our interpolation from  $t = 1$  to  $T$  such that the following equation is satisfied:

$$q_1^{(t)} = \mathbb{E}\langle Q_1 \rangle_{t,0;\epsilon,0}. \quad (4.39)$$

In other words, the interpolation path is *adapted* so that (4.39) holds, which then cancels the second term in (4.35). This path is also independent of  $\delta$  because we have set  $\delta = 0$  in the Gibbs expectation (4.39) as well as in the second term of (4.35). We must still check that equation (4.39) possesses a (unique) solution, see Sec. 4.8.5 for the proof:

**Lemma 4.8** (Existence of the optimal interpolation path). *Eq. (4.39) has a unique solution  $\hat{\mathbf{x}}_n(\epsilon) = \{\hat{\mathbf{x}}_n^{(t)}\}_{t=1}^T \in \mathfrak{B}^T$ . The solution  $\hat{\mathbf{x}}_n \equiv \hat{x}_n^{(t)}\Delta_\infty + (1 - \hat{x}_n^{(t)})\Delta_0$  satisfies  $d\hat{x}_n^{(t)}/d\epsilon \geq 0$ .*

Fixing  $\mathbf{x} = \hat{\mathbf{x}}_n(\epsilon)$ , lemmas 4.6 and 4.7 are used to upper bound the last two terms of (4.35) upon integrating over  $\delta, \epsilon$ . The solution of (4.39), that eliminates  $\mathcal{R}_{t,s;\epsilon,\delta}$ , therefore can be considered as the “optimal interpolation path”. In summary, using lemmas 4.5 to 4.8 on (4.35) we have

$$\left| \int_0^1 d\delta \int_{\epsilon_n}^{2\epsilon_n} d\epsilon \mathcal{R}_{t,s;\epsilon,\delta}(\hat{\mathbf{x}}_n(\epsilon)) \right| = \mathcal{O}\left(\frac{n\epsilon_n}{T}\right) + \mathcal{O}\left(\frac{\sqrt{\epsilon_n}}{n^{\theta/2}}\right) + \mathcal{O}\left(\frac{1}{n^\theta}\right) \quad (4.40)$$

for any sequence  $\epsilon_n$ . We are now ready to prove the upper bound.

**Proposition 4.2** (Upper bound). *For any  $\mathbf{c} \in \mathfrak{B}$  we have*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} H(\boldsymbol{\sigma} | \mathbf{J}) \leq \sup_{\mathbf{x} \in \mathfrak{B}} h_{\text{RS}}(\mathbf{x}). \quad (4.41)$$

*Proof.* We evaluate (4.32) at  $\mathbf{x} = \hat{\mathbf{x}}_n(\epsilon)$  and average the equation over  $\delta \in [0, 1]$ ,  $\epsilon \in [\epsilon_n, 2\epsilon_n]$ . Using (4.40) and Lemma 4.2

$$h_{1,0;\epsilon=0,\delta=0} = \tilde{h}_{\epsilon=0,\delta=0}(\hat{\mathbf{x}}_n(\epsilon)) + \mathcal{O}\left(\frac{n}{T}\right) + \mathcal{O}\left(\frac{1}{\sqrt{\epsilon_n}n^\theta}\right) + \mathcal{O}\left(\frac{1}{\epsilon_n n^\theta}\right) + \mathcal{O}\left(\epsilon_n + \frac{1}{n^\theta}\right). \quad (4.42)$$

A trivial upper bound together with Lemma 4.17 gives

$$\tilde{h}_{\epsilon=0,\delta=0}(\hat{\mathbf{x}}_n(\epsilon)) \leq \sup_{\mathbf{x} \in \mathfrak{B}^T} \tilde{h}_{\epsilon=0,\delta=0}(\mathbf{x}) = \sup_{\mathbf{x} \in \mathfrak{B}} h_{\text{RS}}(\mathbf{x}), \quad (4.43)$$

where the r.h.s. is independent of  $n$  and  $T$ . We substitute (4.43) into (4.42). Then we choose  $\epsilon_n = n^{-\gamma}$  with  $0 < \gamma < 1$  and pass both sides of (4.42) to the  $\limsup \limsup_{n \rightarrow \infty} \limsup_{T \rightarrow \infty}$ . Note that  $h_{1,0;\epsilon=0,\delta=0}$  is independent of  $T$ . This implies that the left hand side of (4.42) at the limit becomes

$$\limsup_{n \rightarrow \infty} \limsup_{T \rightarrow \infty} \tilde{h}_{\epsilon=0,\delta=0}(\hat{\mathbf{x}}_n(\epsilon)) = \limsup_{n \rightarrow \infty} h_{1,0;\epsilon=0,\delta=0}. \quad (4.44)$$

The proof is ended by also using (4.27) on (4.44).  $\square$

## 4.5 Proof of the fundamental sum rule (4.28)–(4.29)

Similar computations go back to [65] and were applied in Nishimori symmetric situations in [71, 72, 69], so we will be relatively brief. We compute  $h_{T,1;\epsilon,\delta}$  and  $\frac{dh_{t,s;\epsilon,\delta}}{ds}$  in (4.26). From the definitions (4.19), (4.20), (4.22), and the identity  $e^{\sigma x} = (1 + \sigma \tanh x) \cosh x$  for  $\sigma \in \{-1, +1\}$ , we can expand  $h_{T,1;\epsilon,\delta}$  as

$$\begin{aligned} h_{T,1;\epsilon,\delta} = \mathbb{E} \Big[ \ln \Big( \prod_{t'=1}^T \prod_{B=1}^{e_i^{(t')}} (1 + \tanh U_B^{(t')}) + e^{-2(H+\tilde{H})} \prod_{t'=1}^T \prod_{B=1}^{e_i^{(t')}} (1 - \tanh U_B^{(t')}) \Big) \\ - \frac{\alpha K}{T} \sum_{t'=1}^T \ln(1 + \tanh U^{(t')}) \Big]. \end{aligned} \quad (4.45)$$

Note that the first term is part of (4.16). For  $\frac{dh_{t,s;\epsilon,\delta}}{ds}$  we use the following property of the Poisson distribution: any function  $g : \mathbb{N} \rightarrow \mathbb{R}$  of a random variable  $X \sim \text{Poi}(\nu)$  with Poisson distribution and mean  $\nu$ , and such that both  $\mathbb{E} g(X)$  and  $\mathbb{E} g(X+1)$  exist, satisfies

$$\begin{aligned} \frac{d \mathbb{E} g(X)}{d\nu} &= \sum_{k=0}^{\infty} \frac{d}{d\nu} \left\{ \frac{\nu^k e^{-\nu}}{k!} \right\} g(k) = \sum_{k=1}^{\infty} \frac{\nu^{k-1} e^{-\nu}}{(k-1)!} g(k) - \sum_{k=0}^{\infty} \frac{\nu^k e^{-\nu}}{k!} g(k) \\ &= \sum_{k=0}^{\infty} \frac{\nu^k e^{-\nu}}{k!} g(k+1) - \sum_{k=0}^{\infty} \frac{\nu^k e^{-\nu}}{k!} g(k) \\ &= \mathbb{E} g(X+1) - \mathbb{E} g(X). \end{aligned} \quad (4.46)$$

This allows us to write

$$\begin{aligned} \frac{dh_{t,s;\epsilon,\delta}}{ds} &= -\frac{\alpha}{T} \mathbb{E}_{t,s;\epsilon,\delta} \mathbb{E}_{B,\tilde{J}_B} \ln \langle e^{J_B(\sigma_B-1)} \rangle_{t,s;\epsilon,\delta} \\ &\quad + \frac{\alpha K}{nT} \sum_{i=1}^n \mathbb{E}_{t,s;\epsilon,\delta} \mathbb{E}_{U_i^{(t)}} \ln \langle e^{U_i^{(t)}(\sigma_i-1)} \rangle_{t,s;\epsilon,\delta} \end{aligned} \quad (4.47)$$

where we distinguish the expectation  $\mathbb{E}_{t,s;\epsilon,\delta}$  with respect to the original interpolating model with Hamiltonian (4.19) and the expectation with respect to an “extra measurement” and its neighborhood  $\mathbb{E}_{B,J_B}$  and an “extra field”  $\mathbb{E}_{U_i^{(t)}}$ . Standard algebra, using again the identity  $e^{\pm x} = (1 \pm \tanh x) \cosh x$ , leads to

$$\begin{aligned} & \mathbb{E} \ln \langle e^{J_B(\sigma_B - 1)} \rangle_{t,s;\epsilon,\delta} \\ &= \mathbb{E}_{t,s;\epsilon,\delta} \mathbb{E}_{B,J_B} \ln (1 + \langle \sigma_B \rangle_{t,s;\epsilon,\delta} \tanh J_B) - \mathbb{E}_{J_B} \ln(1 + \tanh J_B) \\ &= \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p} \frac{1}{n^K} \sum_{i_1, \dots, i_K} \mathbb{E}[\langle \sigma_{i_1} \cdots \sigma_{i_K} \rangle_{t,s;\epsilon,\delta}^p] \mathbb{E}[(\tanh J)^p] - \mathbb{E} \ln(1 + \tanh J) \end{aligned}$$

and similarly, using (4.15),

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \mathbb{E} \ln \langle e^{U_i^{(t)}(\sigma_i - 1)} \rangle_{t,s;\epsilon,\delta} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{t,s;\epsilon,\delta} \mathbb{E}_{U_i^{(t)}} \ln (1 + \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \tanh U_i^{(t)}) - \frac{1}{n} \sum_{i=1}^n \mathbb{E} \ln(1 + \tanh U_i^{(t)}) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{t,s;\epsilon,\delta} \mathbb{E}_{J, \mathbf{V}^{(t)}} \ln (1 + \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \tanh J \prod_{j=1}^{K-1} \tanh V_j^{(t)}) - \mathbb{E} \ln(1 + \tanh U^{(t)}) \\ &= \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p} \mathbb{E}[(\tanh J)^p] \mathbb{E}[(\tanh V^{(t)})^p]^{K-1} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\langle \sigma_i \rangle_{t,s;\epsilon,\delta}^p] \\ & \quad - \mathbb{E} \ln(1 + \tanh U^{(t)}). \end{aligned}$$

Recall  $Q_p \equiv \frac{1}{n} \sum_i \sigma_i^{(1)} \cdots \sigma_i^{(p)}$  and thus

$$\langle Q_p \rangle_{t,s;\epsilon,\delta} = \frac{1}{n} \sum_i \langle \sigma_i \rangle_{t,s;\epsilon,\delta}^p, \quad \langle Q_p^K \rangle_{t,s;\epsilon,\delta} = \frac{1}{n^K} \sum_{i_1, \dots, i_K} \langle \sigma_{i_1} \cdots \sigma_{i_K} \rangle_{t,s;\epsilon,\delta}^p.$$

Recall also  $q_p^{(t)} \equiv \mathbb{E}[(\tanh V^{(t)})^p]$ . Then (4.47) becomes

$$\begin{aligned} & \frac{dh_{t,s;\epsilon,\delta}}{ds} \\ &= -\frac{\alpha}{T} \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p} \mathbb{E}[(\tanh J)^p] \mathbb{E}[\langle Q_p^K \rangle_{t,s;\epsilon,\delta} - K(q_p^{(t)})^{K-1} \langle Q_p \rangle_{t,s;\epsilon,\delta}] \\ & \quad + \frac{\alpha}{T} \mathbb{E} \ln(1 + \tanh J) - \frac{\alpha K}{T} \mathbb{E} \ln(1 + \tanh U^{(t)}) \\ &= -\frac{\alpha}{T} \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p} \mathbb{E}[(\tanh J)^p] \mathbb{E} \langle Q_p^K - K(q_p^{(t)})^{K-1} (Q_p - q_p^{(t)}) - (q_p^{(t)})^K \rangle_{t,s;\epsilon,\delta} \\ & \quad + \frac{\alpha(K-1)}{T} \mathbb{E} \ln (1 + \tanh J \prod_{j=1}^K \tanh V_j^{(t)}) \\ & \quad + \frac{\alpha}{T} \mathbb{E} \ln(1 + \tanh J) - \frac{\alpha K}{T} \mathbb{E} \ln(1 + \tanh U^{(t)}). \end{aligned} \tag{4.48}$$

Substituting (4.45) and (4.48) into (4.26) gives (4.28), where

$$\mathcal{R}_{t,s;\epsilon,\delta} = \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p} \mathbb{E}[(\tanh J)^p] \mathbb{E} \langle Q_p^K - K(q_p^{(t)})^{K-1}(Q_p - q_p^{(t)}) - (q_p^{(t)})^K \rangle_{t,s;\epsilon,\delta}. \quad (4.49)$$

An application of (4.11) yields

$$\mathbb{E} \langle Q_{2p-1}^m \rangle_{t,s;\epsilon,\delta} = \mathbb{E} \langle Q_{2p}^m \rangle_{t,s;\epsilon,\delta}$$

for all  $m \in \mathbb{N}$  and  $p \geq 1$ . Similarly an application of (4.12) yields

$$q_{2p-1}^{(t)} = q_{2p}^{(t)} \quad \text{as well as} \quad \mathbb{E}[(\tanh J)^{2p-1}] = \mathbb{E}[(\tanh J)^{2p}]$$

for  $p \geq 1$ . Therefore combining the odd and even terms of (4.49) we obtain the form in (4.29).

## 4.6 Concentration of overlaps I: proof of lemmas 4.3 and 4.6

In this section we prove lemmas 4.3 and 4.6. We need the following lemmas proved in the next section 4.7. For lemma 4.3 it suffices to take an interpolation path  $\mathbf{x} \in \mathfrak{B}^T$  independent of  $\epsilon$  and  $\delta$ . However, for 4.6, we need to take  $\mathbf{x}(\epsilon) \in \mathfrak{B}^T$  dependent on  $\epsilon$  (and independent of  $\delta$ ). We therefore formulate the lemmas below for an  $\epsilon$ -dependent interpolation path.

**Lemma 4.9** (Concentration of  $Q_p$  on  $\langle Q_p \rangle_{t,s;\epsilon,\delta}$ ). *For any  $\mathbf{c} \in \mathfrak{B}$  and any choice of interpolating path  $\mathbf{x}(\epsilon) \in \mathfrak{B}^T$  such that every component satisfies  $dx^{(t)}/d\epsilon \geq 0$ , we have*

$$\int_{\epsilon_0}^{\epsilon_1} d\epsilon \mathbb{E} \langle (Q_p - \langle Q_p \rangle_{t,s;\epsilon,\delta})^2 \rangle_{t,s;\epsilon,\delta} \leq \frac{3p}{n} \quad (4.50)$$

uniformly in  $t, s, \delta$ .

**Lemma 4.10** (Concentration of  $\langle Q_1 \rangle$  on  $\mathbb{E}_{\tilde{\mathbf{H}}} \langle Q_1 \rangle_{t,s;\epsilon,\delta}$ ). *For any  $\mathbf{c} \in \mathfrak{B}$  and  $\mathbf{x}(\epsilon) \in \mathfrak{B}^T$  and any choice of interpolating path such that every component satisfies  $dx^{(t)}/d\epsilon \geq 0$ , we have*

$$\int_{\epsilon_0}^{\epsilon_1} d\epsilon \mathbb{E} [(\langle Q_1 \rangle_{t,s;\epsilon,\delta} - \mathbb{E}_{\tilde{\mathbf{H}}} \langle Q_1 \rangle_{t,s;\epsilon,\delta})^2] \leq \frac{3\delta}{n^\theta} \quad (4.51)$$

for any  $\theta \in (0, 1]$ , uniformly in  $t, s, \delta$ .

**Lemma 4.11** (Concentration of  $\langle Q_1 \rangle_{t,s;\epsilon,\delta}$  on  $\mathbb{E}\langle Q_1 \rangle_{t,s;\epsilon,\delta}$ ). *For  $\mathbf{c} \in \mathfrak{B}$  and any choice of interpolating path  $\mathbf{x}(\epsilon) \in \mathfrak{B}^T$ , we have*

$$\int_{\delta_0}^{\delta_1} d\delta \mathbb{E}[(\mathbb{E}_{\tilde{\mathbf{H}}} \langle Q_1 \rangle_{t,s;\epsilon,\delta} - \mathbb{E}\langle Q_1 \rangle_{t,s;\epsilon,\delta})^2] \leq \left( \frac{15C(\delta_1 - \delta_0)}{(\ln 2)^2} + 4 \right) n^{-(1-2\theta)/3} \quad (4.52)$$

for any  $\theta \in (0, 1/2)$ , uniformly in  $t, s, \epsilon$ , with  $C > 0$  a constant (this constant is obtained from Lemma 4.12).

**Remark 4.4.** We have already seen that Lemma 4.4 implies for the BEC  $\langle Q_1 \rangle_{t,s;\epsilon,\delta} = \langle Q_p \rangle_{t,s;\epsilon,\delta}$ , and therefore the last two concentration lemmas are valid for all overlaps.

### 4.6.1 Proof of lemma 4.3

In lemma 4.3 we take  $\mathbf{x}$  independent of  $\epsilon$ , thus  $dx^{(t)}/d\epsilon = 0$ . We have

$$\begin{aligned} \mathbb{E}\langle |Q_p^K - \langle Q_p \rangle_{t,s;\epsilon,\delta}^K| \rangle_{t,s;\epsilon,\delta} &= \mathbb{E}\left\langle \left| (Q_p - \langle Q_p \rangle_{t,s;\epsilon,\delta}) \sum_{k=0}^{K-1} Q_p^{K-k-1} \langle Q_p \rangle_{t,s;\epsilon,\delta}^k \right| \right\rangle_{t,s;\epsilon,\delta} \\ &\leq K \mathbb{E}\langle |Q_p - \langle Q_p \rangle_{t,s;\epsilon,\delta}| \rangle_{t,s;\epsilon,\delta}. \end{aligned} \quad (4.53)$$

We can apply the Cauchy-Schwarz inequality to get

$$\begin{aligned} \int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \mathbb{E}\langle |Q_p^K - \langle Q_p \rangle_{t,s;\epsilon,\delta}^K| \rangle_{t,s;\epsilon,\delta} \\ \leq K \left\{ (\varepsilon_1 - \varepsilon_0) \int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \mathbb{E}\langle (Q_p - \langle Q_p \rangle_{t,s;\epsilon,\delta})^2 \rangle_{t,s;\epsilon,\delta} \right\}^{1/2}. \end{aligned} \quad (4.54)$$

Thanks to (4.50) we obtain

$$\int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \mathbb{E}\langle |Q_p^K - \langle Q_p \rangle_{t,s;\epsilon,\delta}^K| \rangle_{t,s;\epsilon,\delta} \leq K \left( \frac{3p(\varepsilon_1 - \varepsilon_0)}{n} \right)^{1/2}.$$

This proves Lemma 4.3.

### 4.6.2 Proof of lemma 4.6

Similar to (4.53) and (4.54), it is easy to show

$$\begin{aligned} \int_{\delta_0}^{\delta_1} d\delta \int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \mathbb{E}[\langle |Q_1 \rangle_{t,s;\epsilon,\delta}^K - \mathbb{E}[\langle Q_1 \rangle_{t,s;\epsilon,\delta}]^K] \\ \leq K \left\{ (\delta_1 - \delta_0)(\varepsilon_1 - \varepsilon_0) \int_{\delta_0}^{\delta_1} d\delta \int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \mathbb{E}[(\langle Q_1 \rangle_{t,s;\epsilon,\delta} - \mathbb{E}\langle Q_1 \rangle_{t,s;\epsilon,\delta})^2] \right\}^{1/2}. \end{aligned} \quad (4.55)$$

We decompose  $\mathbb{E}[(\langle Q_1 \rangle_{t,s;\epsilon,\delta} - \mathbb{E}\langle Q_1 \rangle_{t,s;\epsilon,\delta})^2]$  in three parts:

$$\begin{aligned} & \mathbb{E}\langle (Q_1 - \langle Q_1 \rangle_{t,s;\epsilon,\delta})^2 \rangle_{t,s;\epsilon,\delta} + \mathbb{E}[(\langle Q_p \rangle_{t,s;\epsilon,\delta} - \mathbb{E}_{\tilde{H}}\langle Q_p \rangle_{t,s;\epsilon,\delta})^2] \\ & + \mathbb{E}[(\mathbb{E}_{\tilde{H}}\langle Q_1 \rangle_{t,s;\epsilon,\delta} - \mathbb{E}\langle Q_1 \rangle_{t,s;\epsilon,\delta})^2]. \end{aligned}$$

With Fubini's theorem we are free to switch the  $\delta$  and  $\epsilon$  integrals. Lemmas 4.9, 4.10 and 4.11 then imply

$$\begin{aligned} & \int_{\delta_0}^{\delta_1} d\delta \int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \mathbb{E}[(\langle Q_1 \rangle_{t,s;\epsilon,\delta} - \mathbb{E}\langle Q_1 \rangle_{t,s;\epsilon,\delta})^2] \\ & \leq \frac{3(\delta_1 - \delta_0)}{n} + \frac{3(\delta_1^2 - \delta_0^2)}{n^\theta} + \left( \frac{15C(\delta_1 - \delta_0)}{(\ln 2)^2} + 4 \right) \frac{\varepsilon_1 - \varepsilon_0}{n^{(1-2\theta)/3}} \\ & \leq \frac{3(\delta_1 - \delta_0)(1 + \delta_0 + \delta_1)}{n^\theta} + \left( \frac{15C(\delta_1 - \delta_0)}{(\ln 2)^2} + 4 \right) \frac{\varepsilon_1 - \varepsilon_0}{n^{(1-2\theta)/3}}. \end{aligned} \quad (4.56)$$

The bound (4.56) is optimal for  $\theta = (1 - 2\theta)/3$ , i.e.,  $\theta = 1/5$ . But any  $\theta \in (0, 1/2)$  will do. Lemma 4.6 is then obtained by substituting (4.56) into (4.55):

$$\begin{aligned} & \int_{\delta_0}^{\delta_1} d\delta \int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \mathbb{E}[|\langle Q_1 \rangle_{t,s;\epsilon,\delta}^K - \mathbb{E}[\langle Q_1 \rangle_{t,s;\epsilon,\delta}^K]|] \\ & \leq \frac{K}{n^{\theta/2}} \sqrt{(\delta_1 - \delta_0)(\varepsilon_1 - \varepsilon_0) \left( 3(\delta_1 - \delta_0)(1 + \delta_0 + \delta_1) + (\varepsilon_1 - \varepsilon_0) \left( \frac{15C(\delta_1 - \delta_0)}{(\ln 2)^2} + 4 \right) \right)} \\ & = \mathcal{O}\left( ((\delta_1 - \delta_0)(\varepsilon_1 - \varepsilon_0))^{\frac{\delta_1 - \delta_0 + \varepsilon_1 - \varepsilon_0}{n^\theta}} \right)^{1/2}. \end{aligned}$$

## 4.7 Concentration of overlaps II: proof of lemmas 4.9, 4.10, 4.11

We start with useful preliminary results on the derivatives of the free entropy of the interpolated system, and then prove the three concentration lemmas.

### 4.7.1 Useful derivative formulas

We first remark that, according to (4.15), the distribution  $\tilde{\mathbf{x}}^{(t)}$  is a function of  $\mathbf{x}^{(t)}$  and  $\mathbf{c}$ . Therefore  $\tilde{\mathbf{x}}^{(t)} \in \mathfrak{B}$  when  $\mathbf{x}^{(t)}, \mathbf{c} \in \mathfrak{B}$ . Let  $\mathbf{x}^{(t)} = x^{(t)}\Delta_\infty + (1 - x^{(t)})\Delta_0$  and  $\tilde{\mathbf{x}}^{(t)} = \tilde{x}^{(t)}\Delta_\infty + (1 - \tilde{x}^{(t)})\Delta_0$ <sup>1</sup>. From (4.15) we have the relation

$$\tilde{x}^{(t)} = (1 - q)x^{(t)K-1}. \quad (4.57)$$

<sup>1</sup>We prefer to write  $1 - x^{(t)}$  and  $1 - \tilde{x}^{(t)}$  as the erasure probability in this interpolation to align with the way we define the distribution of  $H$  and  $\tilde{H}$ .

We now provide another view of the interpolating Hamiltonian (4.18). Consider an “effective half-edge” fed into node  $i$  with random half-log-likelihood variable

$$\bar{H}_i^{(t,s)} \equiv \sum_{t'=1}^{t-1} \sum_{B=1}^{e_i^{(t')}} U_{B \rightarrow i}^{(t')} + \sum_{C=1}^{e_{i,s}^{(t)}} U_{C \rightarrow i}^{(t)} + H_i + \tilde{H}_i,$$

equal to  $\infty$  with probability

$$\bar{\epsilon}_i^{(t,s)} \equiv 1 - (1 - \epsilon)(1 - \frac{\delta}{n^\theta})(1 - \tilde{x}^{(t)})^{e_i^{(t,s)}} \prod_{t'=1}^{t-1} (1 - \tilde{x}^{(t')})^{e_i^{(t')}}.$$

and equal to 0 complementary probability. Set  $\bar{\mathcal{E}} = (\bar{\epsilon}_1^{(t,s)}, \dots, \bar{\epsilon}_n^{(t,s)})$ . The Hamiltonian (4.18) is equal in distribution to

$$\bar{\mathcal{H}}_{t,s;\bar{\mathcal{E}}}(\boldsymbol{\sigma}, \mathbf{J}, \bar{\mathbf{H}}) \equiv - \sum_{i=1}^n \bar{H}_i^{(t,s)} (\sigma_i - 1) - \sum_{A=1}^{m_s^{(t)}} J_A (\sigma_A - 1). \quad (4.58)$$

Let  $n^{-1} \mathbb{E} \ln \bar{\mathcal{Z}}_{t,s;\bar{\mathcal{E}}}$  the associated averaged free entropy. Clearly this is a function of  $(\mathbb{E}[\bar{\epsilon}_1^{(t,s)}], \dots, \mathbb{E}[\bar{\epsilon}_n^{(t,s)}])$  where for all  $i = 1, \dots, n$

$$\begin{aligned} \mathbb{E}[\bar{\epsilon}_i^{(t,s)}] &\equiv \mathbb{E}_{e_i^{(1)}, \dots, e_i^{(t-1)}, e_i^{(t,s)}} [\bar{\epsilon}_i^{(t,s)}] \\ &= 1 - (1 - \epsilon)(1 - \frac{\delta}{n^\theta}) e^{-\frac{K}{RT}(s\tilde{x}^{(t)} + \sum_{t'=1}^{t-1} \tilde{x}^{(t')})}. \end{aligned} \quad (4.59)$$

Moreover, it is clear that  $h_{t,s;\epsilon,\delta} = n^{-1} \mathbb{E} \ln \bar{\mathcal{Z}}_{t,s;\bar{\mathcal{E}}}$ . Therefore we see that the dependence in  $\epsilon$  and  $\delta$  effectively comes through the combination (4.59). Since this is independent of  $i$  we denote it by  $\mathbb{E}[\bar{\epsilon}^{(t,s)}]$ . The reader should keep in mind that in this combination there is always an explicit  $(\epsilon, \delta)$ , and that there may also be an implicit one through the choice of the interpolating path  $(x^{(t)}, \tilde{x}^{(t)})$ .

We are now ready to state derivative formulas playing an important role.



Their detailed derivation is provided in Appendix 4.8.6:

$$\begin{aligned} \frac{d}{d\mathbb{E}[\bar{\epsilon}(t,s)]} h_{t,s;\epsilon,\delta} &= -\frac{\ln 2}{n} \sum_{i=1}^n (1 - \mathbb{E}\langle \sigma_i \rangle_{t,s;\epsilon,\delta;\sim \bar{H}_i^{(t,s)}}) \\ &= -\frac{\ln 2}{n(1 - \mathbb{E}[\bar{\epsilon}(t,s)])} \sum_{i=1}^n (1 - \mathbb{E}\langle \sigma_i \rangle_{t,s;\epsilon,\delta}), \end{aligned} \quad (4.60)$$

$$\frac{d^2}{d\mathbb{E}[\bar{\epsilon}(t,s)]^2} h_{t,s;\epsilon,\delta} = \frac{\ln 2}{n(1 - \mathbb{E}[\bar{\epsilon}(t,s)])^2} \sum_{i \neq j} \mathbb{E}[\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta} - \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \langle \sigma_j \rangle_{t,s;\epsilon,\delta}], \quad (4.61)$$

$$\begin{aligned} \frac{d}{d\delta} H_{t,s;\epsilon,\delta} &= -\frac{\ln 2}{n^{1+\theta}} \sum_{i=1}^n (1 - \mathbb{E}_{\tilde{\mathbf{H}}} \langle \sigma_i \rangle_{t,s;\epsilon,\delta;\sim \tilde{H}_i}) \\ &= -\frac{\ln 2}{n^{1+\theta}(1 - \delta/n^\theta)} \sum_{i=1}^n (1 - \mathbb{E}_{\tilde{\mathbf{H}}} \langle \sigma_i \rangle_{t,s;\epsilon,\delta}), \end{aligned} \quad (4.62)$$

$$\begin{aligned} \frac{d^2}{d\delta^2} H_{t,s;\epsilon,\delta} &= \frac{1}{n^{2\theta}(1 - \delta/n^\theta)^2} \\ &\quad \sum_{i \neq j} \mathbb{E}_{\tilde{\mathbf{H}}} \ln \left\{ \frac{1 + \langle \sigma_i \rangle_{\sim \tilde{H}_i, \tilde{H}_j} + \langle \sigma_j \rangle_{\sim \tilde{H}_i, \tilde{H}_j} + \langle \sigma_i \sigma_j \rangle_{\sim \tilde{H}_i, \tilde{H}_j}}{1 + \langle \sigma_i \rangle_{\sim \tilde{H}_i, \tilde{H}_j} + \langle \sigma_j \rangle_{\sim \tilde{H}_i, \tilde{H}_j} + \langle \sigma_i \rangle_{\sim \tilde{H}_i, \tilde{H}_j} \langle \sigma_j \rangle_{\sim \tilde{H}_i, \tilde{H}_j}} \right\}, \end{aligned} \quad (4.63)$$

where  $\langle \sigma_i \rangle_{t,s;\epsilon,\delta;\sim \bar{H}_i^{(t,s)}}$  is the Gibbs expectation with fixed  $\bar{H}_i^{(t,s)} = 0$  and  $\langle \sigma_i \rangle_{\sim \tilde{H}_i, \tilde{H}_j} \equiv \langle \sigma_i \rangle_{t,s;\epsilon,\delta;\sim \tilde{H}_i, \tilde{H}_j}$  is the Gibbs expectation with fixed  $\tilde{H}_i = \tilde{H}_j = 0$ . If we choose  $\mathbf{x} = \mathbf{x}(\epsilon)$  independent of  $\delta$  we have furthermore

$$\begin{aligned} \frac{d}{d\delta} h_{t,s;\epsilon,\delta} &= -\frac{\ln 2}{n^{1+\theta}} \sum_{i=1}^n (1 - \mathbb{E}\langle \sigma_i \rangle_{t,s;\epsilon,\delta;\sim \tilde{H}_i}) \\ &= -\frac{\ln 2}{n^{1+\theta}(1 - \delta/n^\theta)} \sum_{i=1}^n (1 - \mathbb{E}\langle \sigma_i \rangle_{t,s;\epsilon,\delta}), \end{aligned} \quad (4.64)$$

$$\frac{d^2}{d\delta^2} h_{t,s;\epsilon,\delta} = \frac{\ln 2}{n^{1+2\theta}(1 - \delta/n^\theta)^2} \sum_{i \neq j} \mathbb{E}[\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta} - \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \langle \sigma_j \rangle_{t,s;\epsilon,\delta}], \quad (4.65)$$

$$\frac{d}{d\delta} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E}\langle \sigma_i \rangle_{t,s;\epsilon,\delta} \right) = \frac{1}{n^{1+\theta}(1 - \frac{\delta}{n^\theta})} \sum_{i,j=1}^n \mathbb{E}[\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta} - \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \langle \sigma_j \rangle_{t,s;\epsilon,\delta}]. \quad (4.66)$$

The first equalities of (4.60), (4.62), (4.64), together with (1.9), tell us that

$$\left| \frac{d}{d\mathbb{E}[\bar{\epsilon}(t,s)]} h_{t,s;\epsilon,\delta} \right| \leq \ln 2, \quad \left| \frac{d}{d\delta} H_{t,s;\epsilon,\delta} \right| \leq \frac{\ln 2}{n^\theta}, \quad \left| \frac{d}{d\delta} h_{t,s;\epsilon,\delta} \right| \leq \frac{\ln 2}{n^\theta}. \quad (4.67)$$

Moreover from (4.63), (4.65) and the second GKS inequality (1.10), we see that  $h_{t,s;\epsilon,\delta}$  and  $H_{t,s;\epsilon,\delta}$  are convex in  $\delta$ .

### 4.7.2 Proof of Lemma 4.9

From the definition of  $Q_p$  we have

$$\begin{aligned} \mathbb{E}\langle (Q_p - \langle Q_p \rangle_{t,s;\epsilon,\delta})^2 \rangle_{t,s;\epsilon,\delta} &= \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}[\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta}^p - \langle \sigma_i \rangle_{t,s;\epsilon,\delta}^p \langle \sigma_j \rangle_{t,s;\epsilon,\delta}^p] \\ &= \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}\left[\left(\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta} - \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \langle \sigma_j \rangle_{t,s;\epsilon,\delta}\right) \sum_{l=0}^{p-1} \langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta}^{p-1-l} \langle \sigma_i \rangle_{t,s;\epsilon,\delta}^l \langle \sigma_j \rangle_{t,s;\epsilon,\delta}^l\right]. \end{aligned} \quad (4.68)$$

By (1.10), we have  $0 \leq \langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta} - \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \langle \sigma_j \rangle_{t,s;\epsilon,\delta}$ . This permits us to upper bound (4.68) as

$$\begin{aligned} &\mathbb{E}\langle (Q_p - \langle Q_p \rangle_{t,s;\epsilon,\delta})^2 \rangle_{t,s;\epsilon,\delta} \\ &\leq \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}\left[\left(\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta} - \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \langle \sigma_j \rangle_{t,s;\epsilon,\delta}\right) \sum_{l=0}^{p-1} \left|\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta}^{p-1-l} \langle \sigma_i \rangle_{t,s;\epsilon,\delta}^l \langle \sigma_j \rangle_{t,s;\epsilon,\delta}^l\right|\right] \\ &\leq \frac{p}{n^2} \sum_{i,j=1}^n \mathbb{E}\left[\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta} - \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \langle \sigma_j \rangle_{t,s;\epsilon,\delta}\right]. \end{aligned} \quad (4.69)$$

Hence, integrating (4.69) over  $\epsilon \in [\varepsilon_0, \varepsilon_1]$  and recalling the formula (4.61), we obtain

$$\begin{aligned} &\int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \mathbb{E}\langle (Q_p - \langle Q_p \rangle_{t,s;\epsilon,\delta})^2 \rangle_{t,s;\epsilon,\delta} \\ &\leq p \int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}[\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta} - \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \langle \sigma_j \rangle_{t,s;\epsilon,\delta}] \\ &\leq \frac{p}{n} + \frac{p}{n \ln 2} \int_{\varepsilon_0}^{\varepsilon_1} d\epsilon (1 - \mathbb{E}[\bar{\epsilon}^{(t,s)}]) \frac{d^2}{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]^2} h_{t,s;\epsilon,\delta}. \end{aligned} \quad (4.70)$$

Recall (4.59) for the expression of  $\mathbb{E}[\bar{\epsilon}^{(t,s)}]$ . Under the hypothesis  $dx^{(t)}/d\epsilon \geq 0$  for all  $t \in \{1, \dots, T\}$  and using (4.57), we have  $d\tilde{x}^{(t)}/d\epsilon \geq 0$ . This gives

$$\frac{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]}{d\epsilon} = \frac{1 - \mathbb{E}[\bar{\epsilon}^{(t,s)}]}{1 - \epsilon} + (1 - \mathbb{E}[\bar{\epsilon}^{(t,s)}]) \frac{\alpha K}{T} \left( s \frac{d\tilde{x}^{(t)}}{d\epsilon} + \sum_{t'=1}^{t-1} \frac{d\tilde{x}^{(t')}}{d\epsilon} \right) \geq 1 - \mathbb{E}[\bar{\epsilon}^{(t,s)}] \quad (4.71)$$

and allows us to relax the second term of (4.70):

$$\begin{aligned}
 \int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \mathbb{E} \langle (Q_p - \langle Q_p \rangle_{t,s;\epsilon,\delta})^2 \rangle_{t,s;\epsilon,\delta} &\leq \frac{p}{n} + \frac{p}{n \ln 2} \int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \frac{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]}{d\epsilon} \frac{d^2}{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]^2} h_{t,s;\epsilon,\delta} \\
 &= \frac{p}{n} + \frac{p}{n \ln 2} \int_{\varepsilon_0}^{\varepsilon_1} d\epsilon \frac{d}{d\epsilon} \left( \frac{d}{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]} h_{t,s;\epsilon,\delta} \right) \\
 &= \frac{p}{n} + \frac{p}{n \ln 2} \left[ \frac{d}{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]} h_{t,s;\epsilon,\delta} \right]_{\epsilon=\varepsilon_0}^{\epsilon=\varepsilon_1} \\
 &\leq \frac{3p}{n},
 \end{aligned}$$

using the first bound in (4.67) for the last inequality.

### 4.7.3 Proof of Lemma 4.10

Let  $\tilde{H}'_j$  be an i.i.d. copy of  $\tilde{H}_j$ . Let also  $\tilde{\mathbf{H}}^j$  be a vector same as  $\tilde{\mathbf{H}}$  except the  $j$ -th component is replaced by  $\tilde{H}'_j$ . By the Efron-Stein inequality we have

$$\begin{aligned}
 &\mathbb{E}_{\tilde{\mathbf{H}}} [(\langle Q_1 \rangle_{\tilde{\mathbf{H}}} - \mathbb{E}_{\tilde{\mathbf{H}}} \langle Q_1 \rangle_{\tilde{\mathbf{H}}})^2] \\
 &\leq \frac{1}{2} \sum_{j=1}^n \mathbb{E}_{\tilde{\mathbf{H}}} \mathbb{E}_{\tilde{H}'_j} [(\langle Q_1 \rangle_{\tilde{\mathbf{H}}^j} - \langle Q_1 \rangle_{\tilde{\mathbf{H}}})^2] \\
 &= \frac{1}{2} \sum_{j=1}^n \mathbb{E}_{\tilde{\mathbf{H}}} \mathbb{E}_{\tilde{H}'_j} [(\langle Q_1 \rangle_{\tilde{\mathbf{H}}^j} - \langle Q_1 \rangle_{\tilde{\mathbf{H}}})^2 \mathbb{I}(\tilde{H}'_j = \infty) \mathbb{I}(\tilde{H}_j = 0)] \\
 &\quad + \frac{1}{2} \sum_{j=1}^n \mathbb{E}_{\tilde{\mathbf{H}}} \mathbb{E}_{\tilde{H}'_j} [(\langle Q_1 \rangle_{\tilde{\mathbf{H}}^j} - \langle Q_1 \rangle_{\tilde{\mathbf{H}}})^2 \mathbb{I}(\tilde{H}'_j = 0) \mathbb{I}(\tilde{H}_j = \infty)] \\
 &= \sum_{j=1}^n \mathbb{E}_{\tilde{\mathbf{H}}} \mathbb{E}_{\tilde{H}'_j} [(\langle Q_1 \rangle_{\tilde{\mathbf{H}}^j} - \langle Q_1 \rangle_{\tilde{\mathbf{H}}})^2 \mathbb{I}(\tilde{H}'_j = \infty) \mathbb{I}(\tilde{H}_j = 0)]. \quad (4.72)
 \end{aligned}$$

To see the last equality we can exchange  $\tilde{H}_j$  and  $\tilde{H}'_j$  in the r.h.s of the second equality to see that the two terms are equal. This symmetry allows us to simplify the expression to (4.72). The GKS inequalities (1.9), (1.10) imply  $0 \leq \langle Q_1 \rangle_{\tilde{\mathbf{H}}^j} - \langle Q_1 \rangle_{\tilde{\mathbf{H}}} \leq 1$ . This allows us to relax (4.72) to

$$\begin{aligned}
 &\mathbb{E}_{\tilde{\mathbf{H}}} [(\langle Q_1 \rangle_{\tilde{\mathbf{H}}} - \mathbb{E}_{\tilde{\mathbf{H}}} \langle Q_1 \rangle_{\tilde{\mathbf{H}}})^2] \\
 &\leq \sum_{j=1}^n \mathbb{E}_{\tilde{\mathbf{H}}} \mathbb{E}_{\tilde{H}'_j} [(\langle Q_1 \rangle_{\tilde{\mathbf{H}}^j} - \langle Q_1 \rangle_{\tilde{\mathbf{H}}}) \mathbb{I}(\tilde{H}'_j = \infty) \mathbb{I}(\tilde{H}_j = 0)]. \quad (4.73)
 \end{aligned}$$

When  $\tilde{H}'_j = \infty$  and  $\tilde{H}_j = 0$ , we have

$$\begin{aligned}
& \langle Q_1 \rangle_{\tilde{H}^j} - \langle Q_1 \rangle_{\tilde{H}} \\
&= \frac{1}{n} \sum_{i=1}^n (\langle \sigma_i \rangle_{\tilde{H}^j} - \langle \sigma_i \rangle_{\tilde{H}}) = \frac{1}{n} \sum_{i=1}^n \left( \frac{\langle \sigma_i e^{\tilde{H}'_j(\sigma_j-1)} \rangle_{\tilde{H}}}{\langle e^{\tilde{H}'_j(\sigma_j-1)} \rangle_{\tilde{H}}} - \langle \sigma_i \rangle_{\tilde{H}} \right) \\
&= \frac{1}{n} \sum_{i=1}^n \left( \frac{\langle \sigma_i(1+\sigma_j) \rangle_{\tilde{H}}}{1 + \langle \sigma_j \rangle_{\tilde{H}}} - \langle \sigma_i \rangle_{\tilde{H}} \right) = \frac{1}{n} \sum_{i=1}^n \int_0^1 d\tau \frac{d}{d\tau} \frac{\langle \sigma_i \rangle_{\tilde{H}} + \tau \langle \sigma_i \sigma_j \rangle_{\tilde{H}}}{1 + \tau \langle \sigma_j \rangle_{\tilde{H}}} \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^1 d\tau \frac{\langle \sigma_i \sigma_j \rangle_{\tilde{H}} - \langle \sigma_i \rangle_{\tilde{H}} \langle \sigma_j \rangle_{\tilde{H}}}{(1 + \tau \langle \sigma_j \rangle_{\tilde{H}})^2} \leq \frac{1}{n} \sum_{i=1}^n (\langle \sigma_i \sigma_j \rangle_{\tilde{H}} - \langle \sigma_i \rangle_{\tilde{H}} \langle \sigma_j \rangle_{\tilde{H}}) \\
\end{aligned} \tag{4.74}$$

$$= \frac{1}{n} \sum_{i=1}^n \int_0^1 d\tau \frac{\langle \sigma_i \sigma_j \rangle_{\tilde{H}} - \langle \sigma_i \rangle_{\tilde{H}} \langle \sigma_j \rangle_{\tilde{H}}}{(1 + \tau \langle \sigma_j \rangle_{\tilde{H}})^2} \leq \frac{1}{n} \sum_{i=1}^n (\langle \sigma_i \sigma_j \rangle_{\tilde{H}} - \langle \sigma_i \rangle_{\tilde{H}} \langle \sigma_j \rangle_{\tilde{H}}) \tag{4.75}$$

where the first equality of (4.74) follows from the identity  $e^{\tilde{H}'_j \sigma_j} \equiv \cosh \tilde{H}'_j (1 + \sigma_j \tanh \tilde{H}'_j)$ , and the last bound in (4.75) uses the first GKS inequality (1.9). Substituting (4.75) into (4.73), we get

$$\begin{aligned}
& \int_{\varepsilon_0}^{\varepsilon_1} d\varepsilon \mathbb{E}[(\langle Q_1 \rangle_{\tilde{H}} - \mathbb{E}_{\tilde{H}} \langle Q_1 \rangle_{\tilde{H}})^2] \\
& \leq \int_{\varepsilon_0}^{\varepsilon_1} d\varepsilon \frac{1}{n} \sum_{i,j=1}^n \mathbb{E}[(\langle \sigma_i \sigma_j \rangle_{\tilde{H}} - \langle \sigma_i \rangle_{\tilde{H}} \langle \sigma_j \rangle_{\tilde{H}}) \mathbb{I}(\tilde{H}'_j = \infty) \mathbb{I}(\tilde{H}_j = 0)] \\
& = \mathbb{P}(\tilde{H}'_j = \infty) \int_{\varepsilon_0}^{\varepsilon_1} d\varepsilon \frac{1}{n} \sum_{i,j=1}^n \mathbb{E}[(\langle \sigma_i \sigma_j \rangle_{\tilde{H}} - \langle \sigma_i \rangle_{\tilde{H}} \langle \sigma_j \rangle_{\tilde{H}}) \mathbb{I}(\tilde{H}_j = 0)] \\
& \leq \mathbb{P}(\tilde{H}'_j = \infty) \int_{\varepsilon_0}^{\varepsilon_1} d\varepsilon \frac{1}{n} \sum_{i,j=1}^n \mathbb{E}[(\langle \sigma_i \sigma_j \rangle_{\tilde{H}} - \langle \sigma_i \rangle_{\tilde{H}} \langle \sigma_j \rangle_{\tilde{H}})] \\
& = \delta n^{1-\theta} \int_{\varepsilon_0}^{\varepsilon_1} d\varepsilon \mathbb{E}[(Q_1 - \langle Q_1 \rangle_{\tilde{H}})^2]_{\tilde{H}} \\
& \leq \frac{3\delta}{n^\theta},
\end{aligned}$$

using Lemma 4.9 to yield the last inequality.

#### 4.7.4 Proof of Lemma 4.11

We write  $H_{t,s;\varepsilon}(\delta) \equiv H_{t,s;\varepsilon,\delta}$  and  $h_{t,s;\varepsilon}(\delta) \equiv h_{t,s;\varepsilon,\delta}$  to emphasize  $\delta$  in this proof. For both quantities we have taken the expectation over  $\tilde{\mathbf{H}}$  and therefore their

derivatives w.r.t.  $\delta$  are well-defined.<sup>2</sup> From (4.62) and (4.64) we have

$$|\mathbb{E}_{\tilde{\mathbf{H}}} \langle Q_1 \rangle_{t,s;\epsilon,\delta} - \mathbb{E} \langle Q_1 \rangle_{t,s;\epsilon,\delta}| = \frac{n^\theta(1 - \frac{\delta}{n^\theta})}{\ln 2} \left| \frac{d}{d\delta} H_{t,s;\epsilon}(\delta) - \frac{d}{d\delta} h_{t,s;\epsilon}(\delta) \right| \quad (4.76)$$

$$\leq \frac{n^\theta}{\ln 2} \left| \frac{d}{d\delta} H_{t,s;\epsilon}(\delta) - \frac{d}{d\delta} h_{t,s;\epsilon}(\delta) \right|. \quad (4.77)$$

Recall that  $H_{t,s;\epsilon}(\delta)$  and  $h_{t,s;\epsilon}(\delta)$  are convex in  $\delta$ . Lemma 3.12 then implies that for any  $\xi > 0$  we have

$$\begin{aligned} & \left| \frac{d}{d\delta} H_{t,s;\epsilon}(\delta) - \frac{d}{d\delta} h_{t,s;\epsilon}(\delta) \right| \\ & \leq \xi^{-1} \sum_{u \in \{\delta - \xi, \xi, \delta + \xi\}} |H_{t,s;\epsilon}(u) - h_{t,s;\epsilon}(u)| + C_\xi^+(\delta) + C_\xi^-(\delta) \end{aligned} \quad (4.78)$$

where

$$\begin{cases} C_\xi^+(\delta) \equiv \frac{d}{d\delta} h_{t,s;\epsilon}(\delta + \xi) - \frac{d}{d\delta} h_{t,s;\epsilon}(\delta) \geq 0, \\ C_\xi^-(\delta) \equiv \frac{d}{d\delta} h_{t,s;\epsilon}(\delta) - \frac{d}{d\delta} h_{t,s;\epsilon}(\delta - \xi) \geq 0. \end{cases} \quad (4.79)$$

We substitute (4.78) into (4.77), then square both sides and apply  $(\sum_{r=1}^k u_r)^2 \leq k \sum_{r=1}^k u_r^2$ . The resulting inequality upon full expectation is written as

$$\begin{aligned} \mathbb{E}[(\mathbb{E}_{\tilde{\mathbf{H}}} \langle Q_1 \rangle_{t,s;\epsilon,\delta} - \mathbb{E} \langle Q_1 \rangle_{t,s;\epsilon,\delta})^2] & \leq \frac{5n^{2\theta}}{(\xi \ln 2)^2} \sum_{u \in \{\delta - \xi, \xi, \delta + \xi\}} \mathbb{E}[(H_{t,s;\epsilon}(u) - h_{t,s;\epsilon}(u))^2] \\ & + \frac{5n^{2\theta}}{(\ln 2)^2} ((C_\xi^+(\delta))^2 + (C_\xi^-(\delta))^2). \end{aligned} \quad (4.80)$$

We now make use of a concentration result for the interpolated free entropy. In Appendix 4.8.7 we prove:

**Lemma 4.12** (Free entropy concentration). *For any  $s$  in  $[0, 1]$  and  $t = 1, \dots, T$  there is a constant  $C > 0$  such that*

$$\mathbb{E}[(H_{t,s;\epsilon,\delta} - h_{t,s;\epsilon,\delta})^2] \leq \frac{C}{n}. \quad (4.81)$$

Using Lemma 4.12, the first term on the r.h.s is found to be smaller than  $15C/((\ln 2)^2 n^{1-2\theta} \xi^2)$ . Next, using  $\frac{\ln 2}{n^\theta} \leq \frac{dh_{t,s;\epsilon}(\delta)}{d\delta} \leq 0$  allows us to assert from

---

<sup>2</sup>The proof here differs from the standard strategy in [40, 47] in the way that an extra parameter  $\delta$  is required and  $x(\epsilon)$  has to be independent of  $\delta$ . This is because we need a well-defined derivative of free entropy such that we can obtain a controllable upper bound like in (4.77).

(4.79) that  $|C_\xi^\pm(\delta)| \leq \frac{\ln 2}{n^\theta}$ . Then using  $C_\xi^\pm(\delta) \geq 0$

$$\int_{\delta_0}^{\delta_1} d\delta (C_\xi^+(\delta)^2 + C_\xi^-(\delta)^2) \leq \frac{\ln 2}{n^\theta} \int_{\delta_0}^{\delta_1} d\delta (C_\xi^+(\delta) + C_\xi^-(\delta)) \quad (4.82)$$

$$= \frac{\ln 2}{n^\theta} [(h_{t,s;\epsilon}(\delta_1 + \xi) - h_{t,s;\epsilon}(\delta_1 - \xi)) + (h_{t,s;\epsilon}(\delta_0 - \xi) - h_{t,s;\epsilon}(\delta_0 + \xi))] \quad (4.83)$$

$$\leq \frac{4(\ln 2)^2 \xi}{n^{2\theta}} \quad (4.84)$$

where the mean value theorem has been used to get the last inequality. Thus when (4.80) is integrated over  $\delta$  we obtain

$$\int_{\delta_0}^{\delta_1} d\delta \mathbb{E}[(\mathbb{E}_{\tilde{H}}\langle Q_1 \rangle_{t,s;\epsilon,\delta} - \mathbb{E}\langle Q_1 \rangle_{t,s;\epsilon,\delta})^2] \leq \frac{15C(\delta_1 - \delta_0)}{(\ln 2)^2 n^{1-2\theta} \xi^2} + 4\xi. \quad (4.85)$$

The proof is ended by choosing  $\xi$  such that  $1/(n^{1-2\theta} \xi^2) = \xi$ , i.e.,  $\xi = n^{-(1-2\theta)/3}$  and  $\theta \in (0, 1/2)$ .

## 4.8 Appendix

### 4.8.1 Direct proof of identity (4.12) for symmetric distributions

If  $x(-dh) = e^{-2h}x(dh)$  holds, then we have

$$\begin{aligned} \int_{-\infty}^{\infty} (\tanh h)^{2k-1} x(dh) &= \int_0^{\infty} (\tanh h)^{2k-1} x(dh) - \int_0^{\infty} (\tanh h)^{2k-1} x(-dh) \\ &= \int_0^{\infty} (\tanh h)^{2k-1} (1 - e^{-2h}) x(dh) = \int_0^{\infty} (\tanh h)^{2k} (1 + e^{-2h}) x(dh) \\ &= \int_0^{\infty} (\tanh h)^{2k} x(dh) + \int_0^{\infty} (\tanh h)^{2k} x(-dh) = \int_{-\infty}^{\infty} (\tanh h)^{2k} x(dh). \end{aligned}$$

### 4.8.2 Rewriting the replica formula: proof of (4.89)

We copy again

$$\begin{aligned} \tilde{h}_{\epsilon,\delta}(\mathbf{x}) &= \mathbb{E} \left[ \ln \left( \prod_{t=1}^T \prod_{b=1}^l (1 + \tanh U_b^{(t)}) + e^{-2(H+\tilde{H})} \prod_{t=1}^T \prod_{b=1}^l (1 - \tanh U_b^{(t)}) \right) \right. \\ &\quad \left. - \frac{\alpha(K-1)}{T} \sum_{t=1}^T \ln \left( 1 + \tanh \tilde{J} \prod_{i=1}^K \tanh V_i^{(t)} \right) - \alpha \ln(1 + \tanh \tilde{J}) \right]. \end{aligned}$$

The first term can be rewritten as

$$\begin{aligned}
& \mathbb{E} \ln \left( \prod_{t=1}^T \prod_{b=1}^l (1 + \tanh U_b^{(t)}) + e^{-2(H+\tilde{H})} \prod_{t=1}^T \prod_{b=1}^l (1 - \tanh U_b^{(t)}) \right) \\
&= \mathbb{E} \ln \left( \prod_{t=1}^T \prod_{b=1}^l (1 + \tanh U_b^{(t)}) \right) + \mathbb{E} \ln \left( 1 + e^{-2(H+\tilde{H})} \prod_{t=1}^T \prod_{b=1}^l \frac{1 - \tanh U_b^{(t)}}{1 + \tanh U_b^{(t)}} \right) \\
&= \mathbb{E} \ln \left( \prod_{t=1}^T \prod_{b=1}^l (1 + \tanh U_b^{(t)}) \right) + \mathbb{E} \ln \left( 1 + e^{-2(\sum_{t=1}^T \sum_{b=1}^l U_b^{(t)} + H + \tilde{H})} \right) \\
&= -\alpha K H \left( \frac{1}{T} \sum_{t=1}^T \mathbf{c} \boxtimes (\mathbf{x}^{(t)})^{\boxtimes(K-1)} \right) + \alpha K \ln 2 \\
&\quad + H \left( \mathbf{h} \boxtimes \Lambda^{\boxtimes} \left( \frac{1}{T} \sum_{t=1}^T \mathbf{c} \boxtimes (\mathbf{x}^{(t)})^{\boxtimes(K-1)} \right) \right).
\end{aligned}$$

The second term can be easily seen to be equal to

$$\begin{aligned}
& -\frac{\alpha(K-1)}{T} \sum_{t=1}^T \ln(1 + \tanh J \prod_{i=1}^K \tanh V_i^{(t)}) \\
&= \frac{\alpha(K-1)}{T} \sum_{t=1}^T H(\mathbf{c} \boxtimes (\mathbf{x}^{(t)})^{\boxtimes K}) - \alpha(K-1) \ln 2.
\end{aligned}$$

The remaining term is

$$-\alpha \mathbb{E} \ln(1 + \tanh J) = \alpha(H(\mathbf{c}) - \ln 2).$$

### 4.8.3 Same supremum of two free entropy functionals: proof of Lemma 4.1

We first note that the generalized entropy functionals can easily be shown to be upper bounded and are defined on a closed convex set of probability measures. Hence we can replace the supremum in the lemma by a maximum. For the BEC  $\mathbf{x} \in \mathfrak{B}^T$  and  $\tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x})$  becomes a function of  $\mathbf{x} \in [0, 1]^T$ , therefore the proof of the lemma can be carried out directly by elementary real analysis calculations.

Here we give an analysis that applies more generally to functionals over  $\mathbf{x} \in \mathfrak{X}^T$  in the general case of symmetric channels. Let us outline the strategy of the proof: (i) We first show that the stationarity condition for  $\tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x})$  implies that all  $\mathbf{x}^{(t)}$  are equal for  $t = 1, \dots, T$ ; (ii) We then show that there exists a sequence of distributions that converges to a stationary point; (iii) Finally, we show that a maximum of  $\tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x})$  is necessarily a stationary point.

Before carrying out point (i) it is convenient to express  $\tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x})$  more explicitly in terms of the distribution  $\mathbf{x}$  thanks to a formalism from coding theory (see e.g. [32, 73]). We define an entropy functional<sup>3</sup>  $H : \mathfrak{X} \rightarrow \mathbb{R}$  as

$$H(\mathbf{x}) \equiv \int \ln(1 + e^{-2a})\mathbf{x}(da) = \ln 2 - \int \ln(1 + \tanh a)\mathbf{x}(da). \quad (4.86)$$

The argument  $a$  is to be interpreted as a half-log-likelihood ratio. Two convolution operators  $\circledast, \boxtimes : \mathfrak{X} \times \mathfrak{X} \rightarrow \mathfrak{X}$  are defined for  $a_1 \sim \mathbf{x}_1, a_2 \sim \mathbf{x}_2$  such that  $\mathbf{x}_1 \circledast \mathbf{x}_2$  is the distribution of  $a_1 + a_2$  and  $\mathbf{x}_1 \boxtimes \mathbf{x}_2$  is the distribution of  $\tanh^{-1}(\tanh a_1 \tanh a_2)$ . Therefore, the entropies of convolutions are

$$H(\circledast_{i=1}^k \mathbf{x}_i) = \int \ln(1 + e^{-2\sum_{i=1}^k a_i}) \prod_{i=1}^k \mathbf{x}_i(da_i), \quad (4.87)$$

$$H(\boxtimes_{i=1}^k \mathbf{x}_i) = \ln 2 - \int \ln \left( 1 + \prod_{i=1}^k \tanh a_i \right) \prod_{i=1}^k \mathbf{x}_i(da_i). \quad (4.88)$$

We define  $\mathbf{x}^{\circledast 0} \equiv \Delta_0$ , where  $\Delta_0$  is the identity of  $\circledast$  and it is a distribution with solely a point mass at 0. We also define  $\Lambda^{\circledast}(\mathbf{x}) \equiv \sum_{l=0}^{\infty} \Lambda_l \mathbf{x}^{\circledast l}$ , where  $\Lambda_l = \frac{(\alpha K)^l}{l!} e^{-\alpha K}$  denotes the probability that a variable node has degree  $l$ , and  $\lambda^{\circledast}(\mathbf{x}) \equiv \sum_{l=1}^{\infty} \lambda_l \mathbf{x}^{\circledast(l-1)}$ , where  $\lambda_l = \frac{l\Lambda_l}{\Lambda'(1)} = \frac{(\alpha K)^{l-1}}{(l-1)!} e^{-\alpha K}$  denotes the probability that an edge is connected to a variable node of degree  $l \geq 1$ . One can check that (see Appendix 4.8.2)

$$\begin{aligned} \tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x}) &= \alpha H(\mathbf{c}) + \frac{\alpha(K-1)}{T} \sum_{t=1}^T H(\mathbf{c} \boxtimes (\mathbf{x}^{(t)})^{\boxtimes K}) \\ &\quad - \alpha K H\left(\frac{1}{T} \sum_{t=1}^T \mathbf{c} \boxtimes (\mathbf{x}^{(t)})^{\boxtimes(K-1)}\right) + H\left(\Lambda^{\circledast}\left(\frac{1}{T} \sum_{t=1}^T \mathbf{c} \boxtimes (\mathbf{x}^{(t)})^{\boxtimes(K-1)}\right)\right). \end{aligned} \quad (4.89)$$

We will need differentiation rules for functionals. The directional (or Gateaux) derivative of a functional<sup>4</sup>  $F : \mathbf{x} \in \mathfrak{X} \rightarrow \mathbb{R}$  at point  $\mathbf{x} \in \mathfrak{X}$  in the direction  $\eta = \mathbf{x}_2 - \mathbf{x}_1$  where  $\mathbf{x}_1, \mathbf{x}_2 \in \mathfrak{X}$  is by definition the following linear functional of  $\eta$ :

$$dF(\mathbf{x})[\eta] \equiv \lim_{\gamma \rightarrow 0} \frac{F(\mathbf{x} + \gamma\eta) - F(\mathbf{x})}{\gamma}.$$

We employ the following computational rules that are easily proved for linear functionals  $F$ :

<sup>3</sup>The notation  $H$  for the entropy should not be confused with the notation  $H$  for the perturbation field in the model.

<sup>4</sup>To have well defined directional derivatives it is understood that we extend the space  $\mathfrak{X}$  to the Banach space of signed probability measures over  $\bar{\mathbb{R}}$ .



**Lemma 4.13** ([32, Propositions 14 and 15]). *Let  $F : \mathfrak{X} \rightarrow \mathbb{R}$  be a linear functional, and  $*$  be either  $\otimes$  or  $\boxtimes$ . Then for  $k \geq 1$  integer,  $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in \mathfrak{X}$ , and setting  $\eta = \mathbf{x}_2 - \mathbf{x}_1$ , we have*

$$dF(\mathbf{x}^{*k})[\eta] = kF(\mathbf{x}^{*(k-1)} * \eta).$$

For any polynomials  $p, q$ , we have

$$dF(p^{\otimes}(q^{\boxtimes}(\mathbf{x})))[\eta] = F(p'^{\otimes}(q^{\boxtimes}(\mathbf{x}) \otimes (q'^{\boxtimes}(\mathbf{x}) \boxtimes \eta))).$$

where  $p'$  and  $q'$  are the derivatives of the polynomials.

**Lemma 4.14** ([73, Theorem 4.41]). *For any  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathfrak{X}$ , we have*

$$H((\mathbf{x}_1 - \mathbf{x}_2) \otimes \mathbf{x}_3) + H((\mathbf{x}_1 - \mathbf{x}_2) \boxtimes \mathbf{x}_3) = H(\mathbf{x}_1 - \mathbf{x}_2).$$

We can now proceed to prove (i). Fix  $t \in \{1, \dots, T\}$ . Consider the functional  $\tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x})$  as functional w.r.t its  $t$ -th component only. We denote  $d_t \tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x})[\eta^{(t)}]$ , the directional derivative of this functional at the point  $\mathbf{x}$  with respect to its  $t$ -th component (only) in the direction  $\eta^{(t)}$ . This a linear functional of  $\eta^{(t)}$  and corresponds to a “partial” Gateaux derivative as indicated by the notation  $d_t$ . Let

$$\mathbb{T}(\mathbf{c}, \mathbf{x}) \equiv \lambda^{\otimes} \left( \frac{1}{T} \sum_{t=1}^T \mathbf{c} \boxtimes \mathbf{x}^{(t) \boxtimes (K-1)} \right).$$

Using Lemma 4.13,  $d_t \tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x})[\eta^{(t)}]$  is the sum of the following three terms:

$$\frac{\alpha(K-1)}{T} \sum_{t=1}^T d_t H(\mathbf{c} \boxtimes (\mathbf{x}^{(t) \boxtimes K}))[\eta^{(t)}] = \frac{\alpha K(K-1)}{T} H(\mathbf{c} \boxtimes \mathbf{x}^{(t) \boxtimes (K-1)} \boxtimes \eta^{(t)}). \quad (4.90)$$

$$- \alpha K d_t H\left(\frac{1}{T} \sum_{t=1}^T \mathbf{c} \boxtimes \mathbf{x}^{(t) \boxtimes (K-1)}\right)[\eta^{(t)}] = -\frac{\alpha K(K-1)}{T} H(\mathbf{c} \boxtimes \mathbf{x}^{(t) \boxtimes (K-2)} \boxtimes \eta^{(t)}), \quad (4.91)$$

$$d_t H\left(\Lambda^{\otimes} \left( \frac{1}{T} \sum_{t=1}^T \mathbf{c} \boxtimes \mathbf{x}^{(t) \boxtimes (K-1)} \right)\right)[\eta^{(t)}] = \frac{\alpha K(K-1)}{T} H(\mathbb{T}(\mathbf{c}, \mathbf{x}) \otimes (\mathbf{c} \boxtimes \mathbf{x}^{(t) \boxtimes (K-2)} \boxtimes \eta^{(t)})), \quad (4.92)$$

In addition, we use Lemma 4.14 to rewrite (4.92) as

$$\frac{\alpha K(K-1)}{T} \left\{ H(\mathbf{c} \boxtimes \mathbf{x}^{(t) \boxtimes (K-2)} \boxtimes \eta^{(t)}) - H(\mathbb{T}(\mathbf{c}, \mathbf{x}) \boxtimes (\mathbf{c} \boxtimes \mathbf{x}^{(t) \boxtimes (K-2)} \boxtimes \eta^{(t)})) \right\}. \quad (4.93)$$

Putting (4.91), (4.90) and (4.93) together, we have

$$d_t \tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x})[\eta^{(t)}] = \frac{\alpha K(K-1)}{T} H((\mathbf{x}^{(t)} - \mathbf{T}(\mathbf{c}, \mathbf{x})) \boxtimes (\mathbf{c} \boxtimes (\mathbf{x}^{(t) \boxtimes (K-2)} \boxtimes \eta^{(t)})), \quad (4.94)$$

which implies that  $\mathbf{x}$  is a stationary point of  $\tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x})[\eta^{(t)}]$  if and only if it satisfies the equation

$$\mathbf{x}^{(t)} = \mathbf{T}(\mathbf{c}, \mathbf{x}), \quad t = 1, \dots, T. \quad (4.95)$$

In particular, we have  $\mathbf{x}^{(1)} = \dots = \mathbf{x}^{(T)}$  as claimed.

Now, we prove (ii). We use an iterative method that outputs a sequence of distributions ordered by “degradation” defined as follows. For  $\mathbf{x} \in \mathfrak{X}$  and  $f : [0, 1] \rightarrow \mathbb{R}$ , let

$$I_f(\mathbf{x}) \equiv \int f(|\tanh(a)|) \mathbf{x}(da).$$

For  $\mathbf{x}_1, \mathbf{x}_2 \in \mathfrak{X}$ ,  $\mathbf{x}_2$  is said to be degraded with respect to  $\mathbf{x}_1$  (denoted  $\mathbf{x}_1 \preceq \mathbf{x}_2$ ) if  $I_f(\mathbf{x}_1) \leq I_f(\mathbf{x}_2)$  for all concave non-increasing  $f$ . For  $\mathbf{x}_1 = (\mathbf{x}_1^{(1)}, \dots, \mathbf{x}_1^{(T)})$ ,  $\mathbf{x}_2 = (\mathbf{x}_2^{(1)}, \dots, \mathbf{x}_2^{(T)}) \in \mathfrak{X}^T$ , we use  $\mathbf{x}_1 \preceq \mathbf{x}_2$  to denote  $\mathbf{x}_1^{(t)} \preceq \mathbf{x}_2^{(t)}$  for all  $t = 1, \dots, T$ . In the iteration method, we consider the update equation  $\mathbf{T}(\mathbf{c}, \mathbf{x})$ . Let  $\mathbf{T}^{(l)}(\mathbf{c}, \mathbf{x})$  be the distributions obtained by  $l$  iterations of  $\mathbf{T}(\mathbf{c}, \mathbf{x})$  on  $\mathbf{x}$  while keeping  $\mathbf{c}$  fixed. This iteration has the properties:

**Lemma 4.15** ([73, Section 4.6], [32, Lemma 34]). *The operator  $\mathbf{T}(\mathbf{c}, \mathbf{x}) : \mathfrak{X} \times \mathfrak{X}^T \rightarrow \mathfrak{X}^T$  satisfies the following for all  $1 \leq l < \infty$ : If  $\mathbf{T}(\mathbf{c}, \mathbf{x}) \preceq \mathbf{x}$ , then*

$$\mathbf{T}^{(l+1)}(\mathbf{c}, \mathbf{x}) \preceq \mathbf{T}^{(l)}(\mathbf{c}, \mathbf{x}).$$

Also, the limit  $\mathbf{T}^{(\infty)}(\mathbf{c}, \mathbf{x})$  exists and satisfies

$$\mathbf{T}^{(\infty)}(\mathbf{c}, \mathbf{x}) \preceq \mathbf{T}^{(l)}(\mathbf{c}, \mathbf{x}), \quad \mathbf{T}(\mathbf{c}, \mathbf{T}^{(\infty)}(\mathbf{c}, \mathbf{x})) = \mathbf{T}^{(\infty)}(\mathbf{c}, \mathbf{x}).$$

Using Lemma 4.15, we see that  $\mathbf{T}^{(\infty)}(\mathbf{c}, \mathbf{x})$  with  $\mathbf{x} = (\Delta_0, \dots, \Delta_0)$  is a vector of distributions in  $\mathfrak{X}$  that satisfies (4.95). Hence, this fixed point converges to a stationary point of  $\tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x})[\eta^{(t)}]$ .

Finally, we prove (iii). We proceed by contradiction and show that: if  $\mathbf{x}$  is not a stationary point then it cannot be a maximum. From the Taylor expansion of the logarithm and (4.12) we find for any  $\mathbf{x} \in \mathfrak{X}$

$$\begin{aligned} H(\mathbf{x}) &= \ln 2 - \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p} \int \mathbf{x}(da) (\tanh a)^p \\ &= \ln 2 - \sum_{p=1}^{\infty} \frac{1}{2p(2p-1)} \int \mathbf{x}(da) (\tanh a)^{2p}. \end{aligned} \quad (4.96)$$

Let  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \in \mathfrak{X}$ . From (4.88) and (4.96)

$$\begin{aligned} & H((\mathbf{x}_1 - \mathbf{x}_2) \boxtimes (\mathbf{x}_3 - \mathbf{x}_4)) \\ &= - \sum_{p=1}^{\infty} \frac{1}{2p(2p-1)} \left\{ \int (\mathbf{x}_1 - \mathbf{x}_2)(da)(\tanh a)^{2p} \right\} \left\{ \int (\mathbf{x}_3 - \mathbf{x}_4)(da)(\tanh a)^{2p} \right\} \end{aligned} \quad (4.97)$$

which implies that (4.94) can be written as

$$\begin{aligned} & d_t \tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x})[\eta^{(t)}] \\ &= - \frac{\alpha K(K-1)}{T} \sum_{p=1}^{\infty} \frac{1}{2p(2p-1)} \left\{ \int (\mathbf{x}^{(t)} - \mathbb{T}(\mathbf{c}, \mathbf{x}))(da)(\tanh a)^{2p} \right\} \\ & \quad \left\{ \int (\mathbf{c} \boxtimes \mathbf{x}^{(t) \boxtimes (K-2)})(da)(\tanh a)^{2p} \right\} \left\{ \int \eta^{(t)}(da)(\tanh a)^{2p} \right\}. \end{aligned} \quad (4.98)$$

Now, take an  $\mathbf{x}$  that is not a stationary point. Then there must exist an  $t^*$  such that  $\mathbf{x}^{(t^*)} \neq \mathbb{T}(\mathbf{c}, \mathbf{x})$ . Hence we can look at the directional derivative in the non-trivial direction  $\eta^{(t^*)} = \mathbf{x}^{(t^*)} - \mathbb{T}(\mathbf{c}, \mathbf{x})$ . From (4.98) we see that

$$\begin{aligned} & d_{t^*} \tilde{h}_{\epsilon=0, \delta=0}(\mathbf{x})[\eta^{(t^*)}] \\ &= \frac{\alpha K(K-1)}{T} \sum_{p=1}^{\infty} \frac{1}{2p(2p-1)} \left\{ \int (\mathbf{c} \boxtimes \mathbf{x}^{(t) \boxtimes (K-2)})(da)(\tanh a)^{2p} \right\} \\ & \quad \times \left\{ \int (\mathbf{x}^{(t^*)} - \mathbb{T}(\mathbf{h}, \mathbf{c}, \mathbf{x}))(da)(\tanh a)^{2p} \right\}^2 \end{aligned}$$

so the directional derivative is strictly positive. Hence  $\mathbf{x}$  cannot be a maximum since there exists one direction in which the functional increases.

#### 4.8.4 Proofs of technical lemmas

##### Proof of Lemma 4.2

Here we consider the effect of removing the perturbation, so we shall assume  $\frac{dx^{(t)}}{d\epsilon} = 0$ . From formula (4.60) proved in section 4.7.1 we have

$$\begin{aligned} \frac{d}{d\epsilon} h_{t,s;\epsilon,\delta} &= \frac{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]}{d\epsilon} \frac{d}{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]} h_{t,s;\epsilon,\delta} = - \frac{\ln 2}{n(1-\epsilon)} \sum_{i=1}^n (1 - \mathbb{E}\langle \sigma_i \rangle_{t,s;\epsilon,\delta}) \\ &= - \frac{\ln 2}{n} \sum_{i=1}^n (1 - \mathbb{E}\langle \sigma_i \rangle_{t,s;\epsilon,\delta; \sim H_i}). \end{aligned} \quad (4.99)$$

where  $\langle \sigma_i \rangle_{t,s;\epsilon,\delta; \sim H_i}$  is the Gibbs expectation with fixed  $H_i = 0$ . Thus  $|\frac{d}{d\epsilon} h_{t,s;\epsilon,\delta}| \leq \ln 2$ . We also remarked below equation (4.64) that  $|\frac{d}{d\delta} h_{t,s;\epsilon,\delta}| \leq \ln(2)/n^\theta$ . Thus

by the mean value theorem

$$|h_{t,s;\epsilon,\delta} - h_{t,s;\epsilon=0,\delta}| \leq \epsilon \ln 2, \quad (4.100)$$

$$|h_{t,s;\epsilon,\delta} - h_{t,s;\epsilon,\delta=0}| \leq \frac{\ln 2}{n^\theta}. \quad (4.101)$$

By triangle inequality we get (4.24). Note that  $\tilde{h}_{\epsilon,\delta}(\mathbf{x})$  is in the form  $h_{T,1;\epsilon,\delta}(\mathbf{x}) + g(\mathbf{x})$ , therefore

$$\tilde{h}_{\epsilon,\delta}(\mathbf{x}) - \tilde{h}_{\epsilon=0,\delta=0}(\mathbf{x}) = h_{T,1;\epsilon,\delta}(\mathbf{x}) - h_{T,1;\epsilon=0,\delta=0}(\mathbf{x}).$$

Consequently (4.25) follows immediately from (4.24).

#### Proof of Lemma 4.4

The first GKS inequality (1.9) implies

$$\langle \sigma_A \rangle_{t,s;\epsilon,\delta} (1 - \langle \sigma_A \rangle_{t,s;\epsilon,\delta}) \geq 0. \quad (4.102)$$

Moreover, Nishimori's identity (4.11) implies

$$\mathbb{E} \langle \sigma_A \rangle_{t,s;\epsilon,\delta} = \mathbb{E} [\langle \sigma_A \rangle_{t,s;\epsilon,\delta}^2] \quad (4.103)$$

which can be written as

$$\mathbb{E} [\langle \sigma_A \rangle_{t,s;\epsilon,\delta} (1 - \langle \sigma_A \rangle_{t,s;\epsilon,\delta})] = 0. \quad (4.104)$$

As a result of (4.102) and (4.104), we have  $\langle \sigma_A \rangle_{t,s;\epsilon,\delta}$  equal to either 0 or 1.

#### Proof of Lemma 4.5

Using the fundamental theorem of calculus, the desired difference has an integral form

$$\begin{aligned} \mathbb{E} \langle Q_1^k \rangle_{t,s';\epsilon,\delta} - \mathbb{E} \langle Q_1^k \rangle_{t,0;\epsilon,\delta} &= \frac{1}{n^k} \int_0^{s'} ds \sum_{i_1, \dots, i_k=1}^n \frac{d}{ds} \mathbb{E} \langle \sigma_{i_1} \cdots \sigma_{i_k} \rangle_{t,s;\epsilon_n} \\ &= \frac{1}{n^k} \int_0^{s'} ds \sum_{i_1, \dots, i_k=1}^n \left\{ \frac{\alpha K}{T} \sum_{j=1}^n (\mathbb{E} \langle \sigma_{i_1} \cdots \sigma_{i_k} \rangle_{e_{j,s}^{(t)}+1} - \mathbb{E} \langle \sigma_{i_1} \cdots \sigma_{i_k} \rangle_{e_{j,s}^{(t)}}) \right. \\ &\quad \left. - \frac{\alpha n}{T} (\mathbb{E} \langle \sigma_{i_1} \cdots \sigma_{i_k} \rangle_{m_s^{(t)}+1} - \mathbb{E} \langle \sigma_{i_1} \cdots \sigma_{i_k} \rangle_{m_s^{(t)}}) \right\} \end{aligned} \quad (4.105)$$

where (4.105) follows from the Poisson property (4.46). Since  $|\langle \sigma_{i_1} \cdots \sigma_{i_k} \rangle| \leq 1$  we see that the absolute value of (4.105) is bounded by  $2\alpha(K+1)n/T$ .

**Proof of Lemma 4.7**

As indices  $t, s, \epsilon$  are fixed in this proof, we omit them for concision. Recall that (4.66) and (1.10) imply  $\mathbb{E}\langle Q_1 \rangle_\delta$  is increasing in  $\delta$  and therefore  $\mathbb{E}[\langle Q_1 \rangle_\delta] \geq \mathbb{E}[\langle Q_1 \rangle_0]$ . Using also  $|\langle Q_1 \rangle| \leq 1$  we obtain the inequality

$$\begin{aligned} \mathbb{E}[\langle Q_1 \rangle_\delta]^K - \mathbb{E}[\langle Q_1 \rangle_0]^K &= (\mathbb{E}[\langle Q_1 \rangle_\delta] - \mathbb{E}[\langle Q_1 \rangle_0]) \sum_{k=0}^{K-1} \mathbb{E}[\langle Q_1 \rangle_\delta]^{K-k} \mathbb{E}[\langle Q_1 \rangle_\delta]^k \\ &\leq K(\mathbb{E}[\langle Q_1 \rangle_\delta] - \mathbb{E}[\langle Q_1 \rangle_0]). \end{aligned}$$

This inequality, together with  $q_1^{(t)} \equiv \mathbb{E} \tanh V^{(t)} \in [0, 1]$  and  $\mathbb{E}[\langle Q_1 \rangle_\delta] \geq \mathbb{E}[\langle Q_1 \rangle_0]$ , gives

$$\begin{aligned} &\int_{\delta_0}^{\delta_1} d\delta \int_{\epsilon_0}^{\epsilon_1} d\epsilon \left| \mathbb{E}[\langle Q_1 \rangle_\delta]^K - \mathbb{E}[\langle Q_1 \rangle_0]^K - K(q_1^{(t)})^{K-1}(\mathbb{E}[\langle Q_1 \rangle_\delta] - \mathbb{E}[\langle Q_1 \rangle_0]) \right| \\ &\leq 2K \int_{\delta_0}^{\delta_1} d\delta \int_{\epsilon_0}^{\epsilon_1} d\epsilon (\mathbb{E}[\langle Q_1 \rangle_\delta] - \mathbb{E}[\langle Q_1 \rangle_0]) \\ &= 2K \int_{\delta_0}^{\delta_1} d\delta \left( \int_{\epsilon_0}^{\epsilon_1} d\epsilon \mathbb{E}[\langle Q_1 \rangle_\delta] - \int_{\epsilon_0}^{\epsilon_1} d\epsilon \mathbb{E}[\langle Q_1 \rangle_0] \right). \end{aligned} \quad (4.106)$$

We use the mean value theorem to upper bound (4.106) as

$$\begin{aligned} &2K \int_{\delta_0}^{\delta_1} d\delta \delta \max_{\delta' \in [0, \delta]} \left( \frac{d}{d\delta} \int_{\epsilon_0}^{\epsilon_1} d\epsilon \mathbb{E}[\langle Q_1 \rangle_\delta] \right) \Big|_{\delta=\delta'} \\ &= 2K \int_{\delta_0}^{\delta_1} d\delta \delta \max_{\delta' \in [0, \delta]} \left( \int_{\epsilon_0}^{\epsilon_1} d\epsilon \frac{d}{d\delta} \mathbb{E}[\langle Q_1 \rangle_\delta] \right) \Big|_{\delta=\delta'} \end{aligned}$$

where the equality follows from the fact that  $\mathbf{x}(\epsilon)$  is independent of  $\delta$  and therefore we can exchange the order of derivative and integral. Using (4.66) the last equation equals

$$\begin{aligned} &2K \int_{\delta_0}^{\delta_1} d\delta \delta \max_{\delta' \in [0, \delta]} \left( \int_{\epsilon_0}^{\epsilon_1} d\epsilon \frac{1}{n^{1+\theta}(1-\delta/n^\theta)} \sum_{i,j=1}^n \mathbb{E}[\langle \sigma_i \sigma_j \rangle_\delta - \langle \sigma_i \rangle_\delta \langle \sigma_j \rangle_\delta] \right) \Big|_{\delta=\delta'} \\ &= 2Kn^{1-\theta} \int_{\delta_0}^{\delta_1} d\delta \frac{\delta}{1-\delta/n^\theta} \max_{\delta' \in [0, \delta]} \left( \int_{\epsilon_0}^{\epsilon_1} d\epsilon \mathbb{E}[\langle (Q_1 - \langle Q_1 \rangle_\delta)^2 \rangle_\delta] \right) \Big|_{\delta=\delta'} \\ &\leq \frac{6K}{n^\theta} \int_{\delta_0}^{\delta_1} d\delta \frac{\delta}{1-\delta/n^\theta} \\ &\leq \frac{3K(\delta_1^2 - \delta_0^2)}{n^\theta(1-\delta_1/n^\theta)} \end{aligned} \quad (4.107)$$

where (4.107) follows from Lemma 4.9.

### 4.8.5 Existence of the optimal interpolation path: proof of Lemma 4.8

For each  $n$ , we seek distributions  $\mathbf{x}^{(t)} \in \mathcal{B}$  for  $V^{(t)}$ ,  $t = 1, \dots, T$  which solve equation (4.39). By symmetry between vertices  $\mathbb{E}\langle Q_1 \rangle_{t,0;\epsilon,0} = \mathbb{E}\langle \sigma_1 \rangle_{t,0;\epsilon,0}$  so the equation becomes

$$\mathbb{E} \tanh V^{(t)} = \mathbb{E}\langle \sigma_1 \rangle_{t,0;\epsilon,0}. \quad (4.108)$$

Recall that in our interpolation scheme the right hand side depends only on  $\{\mathbf{x}^{(t')}\}_{t' < t}$  and is thus independent of  $\mathbf{x}^{(t)}$ . Thus it suffices to choose  $V^{(t)}$  for  $t = 1, \dots, T$  as follows:  $V^{(t)} = +\infty$  with probability  $\mathbb{E}\langle \sigma_1 \rangle_{t,0;\epsilon,0}$  and  $V^{(t)} = 0$  with probability  $1 - \mathbb{E}\langle \sigma_1 \rangle_{t,0;\epsilon,0}$ . These are the distributions  $\hat{\mathbf{x}}_n^{(t)} \in \mathcal{B}$  of the Lemma. It is clear that this solution is unique and  $\hat{\mathbf{x}}_n^{(t)} = \mathbb{E} \tanh V^{(t)}$ .

Finally, we verify  $d\hat{\mathbf{x}}_n^{(t)}/d\epsilon \geq 0$ . To simplify the notation we use  $\langle - \rangle \equiv \langle - \rangle_{t,0;\epsilon,0}$  and  $\langle - \rangle_{\sim \bar{H}_j^{(t,0)}}$  denotes  $\langle - \rangle_{t,0;\epsilon,0}$  with  $\bar{H}_j^{(t,0)}$  set to 0. Recall the definition of  $\mathbb{E}[\bar{\epsilon}^{(t,s)}]$  in section 4.7.1. By the chain rule we have

$$\frac{d\hat{\mathbf{x}}^{(t)}}{d\epsilon} = \frac{d\mathbb{E}[\bar{\epsilon}^{(t,0)}]}{d\epsilon} \frac{d\mathbb{E}\langle \sigma_1 \rangle_{t,0;\epsilon,0}}{d\mathbb{E}[\bar{\epsilon}^{(t,0)}]}. \quad (4.109)$$

To compute the last derivative, first we use the identity  $e^{\pm x} = (1 \pm \tanh x) \cosh x$  to write

$$\begin{aligned} \mathbb{E}\langle \sigma_1 \rangle_{t,0;\epsilon,0} &= \mathbb{E} \left[ \frac{\langle e^{\bar{H}_j^{(t,0)}(\sigma_j-1)} \sigma_1 \rangle_{\sim \bar{H}_j^{(t,0)}}}{\langle e^{\bar{H}_j^{(t,0)}(\sigma_j-1)} \rangle_{\sim \bar{H}_j^{(t,0)}}} \right] = \mathbb{E} \left[ \frac{\langle \sigma_1 \rangle_{\sim \bar{H}_j^{(t,0)}} + \langle \sigma_1 \sigma_j \rangle_{\sim \bar{H}_j^{(t,0)}} \tanh \bar{H}_j^{(t,0)}}{1 + \langle \sigma_j \rangle_{\sim \bar{H}_j^{(t,0)}} \tanh \bar{H}_j^{(t,0)}} \right] \\ &= \mathbb{E}[\bar{\epsilon}_j^{(t,0)}] \cdot \mathbb{E} \left[ \frac{\langle \sigma_1 \rangle_{\sim \bar{H}_j^{(t,0)}} + \langle \sigma_1 \sigma_j \rangle_{\sim \bar{H}_j^{(t,0)}}}{1 + \langle \sigma_j \rangle_{\sim \bar{H}_j^{(t,0)}}} \right] + (1 - \mathbb{E}[\bar{\epsilon}_j^{(t,0)}]) \mathbb{E}\langle \sigma_1 \rangle_{\sim \bar{H}_j^{(t,0)}}. \end{aligned}$$

Then it is straightforward to compute

$$\frac{d\mathbb{E}\langle \sigma_1 \rangle_{t,0;\epsilon,0}}{d\mathbb{E}[\bar{\epsilon}^{(t,0)}]} = \sum_{j=1}^n \frac{d\mathbb{E}\langle \sigma_1 \rangle_{t,0;\epsilon,0}}{d\mathbb{E}[\bar{\epsilon}_j^{(t,0)}]} = \sum_{j=1}^n \mathbb{E} \left[ \frac{\langle \sigma_1 \sigma_j \rangle_{\sim \bar{H}_j^{(t,0)}} - \langle \sigma_1 \rangle_{\sim \bar{H}_j^{(t,0)}} \langle \sigma_j \rangle_{\sim \bar{H}_j^{(t,0)}}}{1 + \langle \sigma_j \rangle_{\sim \bar{H}_j^{(t,0)}}} \right]. \quad (4.110)$$

The second GKS inequality (1.10) ensures (4.110) non-negative, leaving the sign of  $d\mathbf{x}^{(t)}/d\epsilon$  determined by  $d\mathbb{E}[\bar{\epsilon}^{(t,0)}]/d\epsilon$ . Using (4.57) and (4.59),

$$\begin{aligned} \frac{\mathbb{E}[\bar{\epsilon}^{(1,0)}]}{d\epsilon} &= 1, \\ \frac{\mathbb{E}[\bar{\epsilon}^{(t,0)}]}{d\epsilon} &= \frac{1 - \mathbb{E}[\bar{\epsilon}^{(t,0)}]}{1 - \epsilon} + (1 - \mathbb{E}[\bar{\epsilon}^{(t,0)}]) \frac{\alpha K(K-1)(1-q)}{T} \sum_{t'=1}^{t-1} \hat{\mathbf{x}}^{(t')K-2} \frac{d\hat{\mathbf{x}}^{(t')}}{d\epsilon}. \end{aligned}$$

This equation implies that the claim  $d\hat{\mathbf{x}}^{(t)}/d\epsilon \geq 0$  is true for  $t = 1$  by direct calculation. Then we also get the claim for  $t \geq 2$  by induction.

### 4.8.6 Derivatives of the conditional entropy: proof of (4.60)–(4.65)

A large part of this appendix is an adaptation of [25, 69]. We recall that  $h_{t,s;\epsilon,\delta} = n^{-1} \mathbb{E} \ln \bar{\mathcal{Z}}_{t,s;\bar{\epsilon}}$  where  $\bar{\mathcal{Z}}_{t,s;\bar{\epsilon}}$  is the partition function associated to the hamiltonian (4.58). Therefore, as explained in section (4.7), the free entropy only depends on  $(\epsilon, \delta)$  through the combination (4.59), with an explicit dependence as well as (possibly) an implicit one through the choice of  $\mathbf{x}$ . To alleviate the notations in this appendix we drop the subscripts  $t, s; \bar{\epsilon}$  in the Gibbs brackets.

#### Proof of (4.60)

Let  $\mathcal{H}_{t,s;\bar{\epsilon}}^{\sim i}(\boldsymbol{\sigma}, \mathbf{J}, \bar{\mathbf{H}})$  be the Hamiltonian  $\mathcal{H}_{t,s;\bar{\epsilon}}(\boldsymbol{\sigma}, \mathbf{J}, \bar{\mathbf{H}})$  with  $\bar{H}_i^{(t,s)} = 0$ . Let  $\mathcal{Z}_{t,s;\bar{\epsilon}}^{\sim i}$  and  $\langle - \rangle_{\sim i}$  be the partition function and the Gibbs expectation associated with  $\mathcal{H}_{t,s;\bar{\epsilon}}^{\sim i}(\boldsymbol{\sigma}, \mathbf{J}, \bar{\mathbf{H}})$ . The identities

$$\begin{aligned} \ln \left\{ \frac{\mathcal{Z}_{t,s;\bar{\epsilon}}}{\mathcal{Z}_{t,s;\bar{\epsilon}}^{\sim i}} \right\} &= \ln \langle e^{\bar{H}_i^{(t,s)}(\sigma_i - 1)} \rangle_{\sim i}, \\ e^{\bar{H}_i^{(t,s)}(\sigma_i - 1)} &= \frac{1 + \sigma_i \tanh \bar{H}_i^{(t,s)}}{1 + \tanh \bar{H}_i^{(t,s)}}, \end{aligned} \quad (4.111)$$

imply

$$h_{t,s;\epsilon,\delta} = \frac{1}{n} \mathbb{E} \ln \mathcal{Z}_{t,s;\bar{\epsilon}}^{\sim i} + \frac{1}{n} \mathbb{E} \ln \left\{ \frac{1 + \langle \sigma_i \rangle_{\sim i} \tanh \bar{H}_i^{(t,s)}}{1 + \tanh \bar{H}_i^{(t,s)}} \right\}. \quad (4.112)$$

As  $\tanh \bar{H}_i^{(t,s)}$  and  $\langle \sigma_i \rangle_{\sim i}$  equal either 0 or 1, (4.112) simplifies to

$$h_{t,s;\epsilon,\delta} = \frac{1}{n} \mathbb{E} \ln \mathcal{Z}_{t,s;\bar{\epsilon}}^{\sim i} - \frac{1}{n} \mathbb{E} [\bar{\epsilon}_i^{(t,s)}] \ln 2 (1 - \mathbb{E} \langle \sigma_i \rangle_{\sim i}). \quad (4.113)$$

Therefore, we have

$$\begin{aligned} \frac{d}{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]} h_{t,s;\epsilon,\delta} &= \sum_{i=1}^n \frac{d}{d\mathbb{E}[\bar{\epsilon}_i^{(t,s)}]} h_{t,s;\bar{\epsilon}} \Big|_{\mathbb{E}[\bar{\epsilon}_1^{(t,s)}] = \dots = \mathbb{E}[\bar{\epsilon}_n^{(t,s)}] = \mathbb{E}[\bar{\epsilon}^{(t,s)}]} \\ &= -\frac{\ln 2}{n} \sum_{i=1}^n (1 - \mathbb{E} \langle \sigma_i \rangle_{\sim i}), \end{aligned} \quad (4.114)$$

which is the first equality in (4.60).

To obtain the second equality, simply notice that as  $1 - \langle \sigma_i \rangle_{t,s;\epsilon,\delta} = 0$  when  $\bar{H}_i^{(t,s)} = +\infty$  (which happens with probability  $\bar{\epsilon}_i^{(t,s)}$ ). Performing the expectation over  $\bar{H}_i^{(t,s)}$  in the following expression we get

$$\begin{aligned} 1 - \mathbb{E} \langle \sigma_i \rangle_{t,s;\epsilon,\delta} &= \mathbb{E} [1 - \langle \sigma_i \rangle_{t,s;\epsilon,\delta}] \\ &= \mathbb{E} [(1 - \bar{\epsilon}_i^{(t,s)})(1 - \mathbb{E} \langle \sigma_i \rangle_{\sim i}) + \bar{\epsilon}_i^{(t,s)}(1 - \mathbb{E} \langle \sigma_i \rangle_{\bar{H}_i^{(t,s)} = \infty})] \\ &= (1 - \mathbb{E}[\bar{\epsilon}^{(t,s)}])(1 - \mathbb{E} \langle \sigma_i \rangle_{\sim i}). \end{aligned} \quad (4.115)$$

Replacing in (4.114) yields the second equality in (4.60).

**Proof of (4.61)**

Let  $\mathcal{H}_{t,s;\bar{\epsilon}}^{\sim i,j}(\boldsymbol{\sigma}, \mathbf{J}, \bar{\mathbf{H}})$  be the Hamiltonian  $\mathcal{H}_{t,s;\bar{\epsilon}}(\boldsymbol{\sigma}, \mathbf{J}, \bar{\mathbf{H}})$  with  $\bar{H}_i^{(t,s)} = \bar{H}_j^{(t,s)} = 0$ . Let  $\mathcal{Z}_{t,s;\bar{\epsilon}}^{\sim i,j}$  and  $\langle - \rangle_{\sim i,j}$  be the partition function and the Gibbs expectation associated with  $\mathcal{H}_{t,s;\bar{\epsilon}}^{\sim i,j}(\boldsymbol{\sigma}, \mathbf{J}, \bar{\mathbf{H}})$ . Also let  $t_i^{(t,s)} \equiv \tanh \bar{H}_i^{(t,s)}$ . Using again (4.111) on the identity

$$\ln \left\{ \frac{\mathcal{Z}_{t,s;\epsilon,\delta}}{\mathcal{Z}_{t,s;\bar{\epsilon}}^{\sim i,j}} \right\} = \ln \langle e^{\bar{H}_i^{(t,s)}(\sigma_i-1) + \bar{H}_j^{(t,s)}(\sigma_j-1)} \rangle_{\sim i,j},$$

we have

$$\begin{aligned} h_{t,s;\bar{\epsilon}} &= \frac{1}{n} \mathbb{E} \ln \mathcal{Z}_{t,s;\bar{\epsilon}}^{\sim i,j} + \frac{1}{n} \mathbb{E} \ln \left\{ \frac{1 + \langle \sigma_i \rangle_{\sim i,j} t_i^{(t,s)} + \langle \sigma_j \rangle_{\sim i,j} t_j^{(t,s)} + \langle \sigma_i \sigma_j \rangle_{\sim i,j} t_i^{(t,s)} t_j^{(t,s)}}{1 + t_i^{(t,s)} + t_j^{(t,s)} + t_i^{(t,s)} t_j^{(t,s)}} \right\} \\ &= \frac{1}{n} \mathbb{E} \ln \mathcal{Z}_{t,s;\bar{\epsilon}}^{\sim i,j} + \frac{\mathbb{E}[\bar{\epsilon}_i^{(t,s)}] \mathbb{E}[\bar{\epsilon}_j^{(t,s)}]}{n} \mathbb{E} \ln \left\{ \frac{1 + \langle \sigma_i \rangle_{\sim i,j} + \langle \sigma_j \rangle_{\sim i,j} + \langle \sigma_i \sigma_j \rangle_{\sim i,j}}{4} \right\} \\ &\quad + \frac{\mathbb{E}[\bar{\epsilon}_i^{(t,s)}] (1 - \mathbb{E}[\bar{\epsilon}_j^{(t,s)}])}{n} \mathbb{E} \ln \left\{ \frac{1 + \langle \sigma_i \rangle_{\sim i,j}}{2} \right\} \\ &\quad + \frac{(1 - \mathbb{E}[\bar{\epsilon}_i^{(t,s)}]) \mathbb{E}[\bar{\epsilon}_j^{(t,s)}]}{n} \mathbb{E} \ln \left\{ \frac{1 + \langle \sigma_j \rangle_{\sim i,j}}{2} \right\}, \end{aligned} \quad (4.116)$$

where (4.116) follows from taking the expectation over  $\bar{H}_i^{(t,s)}$  and  $\bar{H}_j^{(t,s)}$ . From (4.113) one can deduce that  $\frac{d^2}{d\mathbb{E}[\bar{\epsilon}_i^{(t,s)}]^2} h_{t,s;\epsilon,\delta} = 0$ . Therefore

$$\begin{aligned} \frac{d^2}{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]^2} h_{t,s;\epsilon,\delta} &= \sum_{i,j=1}^n \frac{d^2}{d\mathbb{E}[\bar{\epsilon}_j^{(t,s)}] d\mathbb{E}[\bar{\epsilon}_i^{(t,s)}]} h_{t,s;\bar{\epsilon}} \Big|_{\epsilon_1=\dots=\epsilon_n=\epsilon} \\ &= \sum_{i \neq j} \frac{d^2}{d\mathbb{E}[\bar{\epsilon}_j^{(t,s)}] d\mathbb{E}[\bar{\epsilon}_i^{(t,s)}]} h_{t,s;\bar{\epsilon}} \Big|_{\mathbb{E}[\bar{\epsilon}_i^{(t,s)}]=\dots=\mathbb{E}[\bar{\epsilon}_n^{(t,s)}]=\mathbb{E}[\bar{\epsilon}^{(t,s)}]}. \end{aligned}$$

The derivatives  $\frac{d^2}{d\mathbb{E}[\bar{\epsilon}_j^{(t,s)}] d\mathbb{E}[\bar{\epsilon}_i^{(t,s)}]} h_{t,s;\epsilon,\delta}$  can be readily obtained from (4.116). This provides

$$\frac{d^2}{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]^2} h_{t,s;\epsilon,\delta} = \frac{1}{n} \sum_{i \neq j} \mathbb{E} \ln \left\{ \frac{1 + \langle \sigma_i \rangle_{\sim i,j} + \langle \sigma_j \rangle_{\sim i,j} + \langle \sigma_i \sigma_j \rangle_{\sim i,j}}{1 + \langle \sigma_i \rangle_{\sim i,j} + \langle \sigma_j \rangle_{\sim i,j} + \langle \sigma_i \rangle_{\sim i,j} \langle \sigma_j \rangle_{\sim i,j}} \right\}. \quad (4.117)$$

We now simplify each term in the sum (4.117). Given that  $\langle \sigma_S \rangle_{\sim i,j}$  equals either 0 or 1 for any subsets  $S \subset \{1 \dots n\}$ , one can verify that the numerator



and denominator of (4.117) can be written as

$$\begin{aligned} \ln(1 + \langle \sigma_i \rangle_{\sim i,j} + \langle \sigma_j \rangle_{\sim i,j} + \langle \sigma_i \sigma_j \rangle_{\sim i,j}) &= (\langle \sigma_i \rangle_{\sim i,j} + \langle \sigma_j \rangle_{\sim i,j} + \langle \sigma_i \sigma_j \rangle_{\sim i,j}) \ln 2 \\ &+ (\langle \sigma_i \rangle_{\sim i,j} \langle \sigma_j \rangle_{\sim i,j} + \langle \sigma_i \rangle_{\sim i,j} \langle \sigma_i \sigma_j \rangle_{\sim i,j} + \langle \sigma_j \rangle_{\sim i,j} \langle \sigma_i \sigma_j \rangle_{\sim i,j}) (\ln 3 - 2 \ln 2) \\ &+ \langle \sigma_i \rangle_{\sim i,j} \langle \sigma_j \rangle_{\sim i,j} \langle \sigma_i \sigma_j \rangle_{\sim i,j} (5 \ln 2 - 3 \ln 3), \end{aligned} \quad (4.118)$$

and

$$\ln(1 + \langle \sigma_i \rangle_{\sim i,j} + \langle \sigma_j \rangle_{\sim i,j} + \langle \sigma_i \rangle_{\sim i,j} \langle \sigma_j \rangle_{\sim i,j}) = (\langle \sigma_i \rangle_{\sim i,j} + \langle \sigma_j \rangle_{\sim i,j}) \ln 2.$$

Special cases of the Nishimori identities (4.11),

$$\begin{aligned} \mathbb{E}[\langle \sigma_i \rangle_{\sim i,j} \langle \sigma_j \rangle_{\sim i,j}] &= \mathbb{E}[\langle \sigma_i \rangle_{\sim i,j} \langle \sigma_j \rangle_{\sim i,j} \langle \sigma_i \sigma_j \rangle_{\sim i,j}], \\ \mathbb{E}[\langle \sigma_i \rangle_{\sim i,j} \langle \sigma_i \sigma_j \rangle_{\sim i,j}] &= \mathbb{E}[\langle \sigma_i \rangle_{\sim i,j} \langle \sigma_j \rangle_{\sim i,j} \langle \sigma_i \sigma_j \rangle_{\sim i,j}], \\ \mathbb{E}[\langle \sigma_j \rangle_{\sim i,j} \langle \sigma_i \sigma_j \rangle_{\sim i,j}] &= \mathbb{E}[\langle \sigma_i \rangle_{\sim i,j} \langle \sigma_j \rangle_{\sim i,j} \langle \sigma_i \sigma_j \rangle_{\sim i,j}], \end{aligned}$$

can now be used to simplify (4.118) so that each term in the sum (4.117) becomes

$$\ln(2) \mathbb{E}[\langle \sigma_i \sigma_j \rangle_{\sim i,j} - \langle \sigma_i \rangle_{\sim i,j} \langle \sigma_j \rangle_{\sim i,j}]. \quad (4.119)$$

Moreover, as  $\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle = 0$  when  $\bar{H}_i^{(t,s)}$  and/or  $\bar{H}_j^{(t,s)}$  equal  $+\infty$ , we obtain

$$\mathbb{E}[\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle] = (1 - \mathbb{E}[\bar{\epsilon}_i^{(t,s)}])(1 - \mathbb{E}[\bar{\epsilon}_j^{(t,s)}]) \mathbb{E}[\langle \sigma_i \sigma_j \rangle_{\sim i,j} - \langle \sigma_i \rangle_{\sim i,j} \langle \sigma_j \rangle_{\sim i,j}]. \quad (4.120)$$

Finally, from (4.117), (4.119) and (4.120) we obtain (4.61).

### Derivation of (4.62) and (4.63)

The derivation of (4.62) is the same as Sec. 4.8.6 except that the steps should be done on  $\tilde{H}_i$  instead of  $\bar{H}_i^{(t,s)}$ . The derivation of (4.62) is the same as Sec. 4.8.6 except that the steps should be done on  $\tilde{H}_i, \tilde{H}_j$  instead of  $\bar{H}_i^{(t,s)}, \bar{H}_j^{(t,s)}$ .

### Proof of (4.64) and (4.65)

For  $\mathbf{x}(\epsilon)$  independent of  $\delta$ , from (4.59) we have

$$\frac{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]}{d\delta} = \frac{1}{n^\theta} (1 - \epsilon) e^{-\frac{K}{RT}(s\bar{x}^t + \sum_{t'=1}^{t-1} \bar{x}^{t'})} = \frac{1 - \mathbb{E}[\bar{\epsilon}^{(t,s)}]}{n^\theta - \delta} \quad \text{and} \quad \frac{d^2\mathbb{E}[\bar{\epsilon}^{(t,s)}]}{d\delta^2} = 0.$$

Together with (4.60) and (4.61) we can immediately derive

$$\begin{aligned} \frac{d}{d\delta} h_{t,s;\epsilon,\delta} &= \frac{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]}{d\delta} \frac{dh_{t,s;\epsilon,\delta}}{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]} = -\frac{\ln 2}{n^{1+\theta}(1 - \delta/n^\theta)} \sum_{i=1}^n (1 - \mathbb{E}[\langle \sigma_i \rangle_{t,s;\epsilon,\delta}]) \\ \frac{d^2}{d\delta^2} h_{t,s;\epsilon} &= \frac{d}{d\delta} \left( \frac{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]}{d\delta} \frac{dh_{t,s;\epsilon,\delta}}{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]} \right) = \left( \frac{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]}{d\delta} \right)^2 \frac{d^2 h_{t,s;\epsilon,\delta}}{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]^2} + \frac{d^2\mathbb{E}[\bar{\epsilon}^{(t,s)}]}{d\delta^2} \frac{dh_{t,s;\epsilon,\delta}}{d\mathbb{E}[\bar{\epsilon}^{(t,s)}]} \\ &= \frac{\ln 2}{n^{1+2\theta}(1 - \delta/n^\theta)^2} \sum_{i \neq j} \mathbb{E}[\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta} - \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \langle \sigma_j \rangle_{t,s;\epsilon,\delta}]. \end{aligned}$$

The first equality of (4.64) follows from applying the same argument in (4.115) to  $\tilde{H}_i$ .

**Proof of (4.66)**

We rearrange (4.64) to obtain

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} \langle \sigma_i \rangle_{t,s;\epsilon,\delta} = \frac{n^\theta(1 - \delta/n^\theta)}{\ln 2} \frac{d}{d\delta} h_{t,s;\epsilon,\delta}. \quad (4.121)$$

Then using (4.64) and (4.65) we have

$$\begin{aligned} \frac{d}{d\delta} \left( \frac{1}{n} \sum_{i=1}^n \mathbb{E} \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \right) &= \frac{n^\theta(1 - \delta/n^\theta)}{\ln 2} \frac{d^2}{d\delta^2} h_{t,s;\epsilon,\delta} - \frac{1}{\ln 2} \frac{d}{d\delta} h_{t,s;\epsilon,\delta} \\ &= \frac{1}{n^{1+\theta}(1 - \delta/n^\theta)} \left( \sum_{i \neq j} \mathbb{E} [\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta} - \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \langle \sigma_j \rangle_{t,s;\epsilon,\delta}] - \sum_{i=1}^n (1 - \mathbb{E} \langle \sigma_i \rangle_{t,s;\epsilon,\delta}) \right) \\ &= \frac{1}{n^{1+\theta}(1 - \delta/n^\theta)} \sum_{i,j=1}^n \mathbb{E} [\langle \sigma_i \sigma_j \rangle_{t,s;\epsilon,\delta} - \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \langle \sigma_j \rangle_{t,s;\epsilon,\delta}] \end{aligned}$$

where the last equality uses one of the Nishimori identities (4.11), namely  $\mathbb{E} \langle \sigma_i \rangle = \mathbb{E} [\langle \sigma_i \rangle^2]$ .

### 4.8.7 Concentration of free entropy

Let  $\mathcal{J}$  collect both the realization of  $\mathbf{J}$  and the graph realization of all the factor nodes carrying elements in  $\mathbf{J}$ . Let  $\mathcal{U}$  collect both the realization of  $\mathbf{U}$  and the graph realization of all the half edges carrying elements in  $\mathbf{U}$ . The proof of Lemma 4.12 can be decomposed into the following three lemmas. We stress that the three Lemmas 4.16, 4.17 and 4.18 are valid under the condition that  $\mathbf{J}, \mathbf{U}$  are non-negative such that we can make use of the consequence  $\langle \sigma_S \rangle_{t,s;\epsilon,\delta} \geq 0$  where  $S$  is any subset of  $\{1, \dots, n\}$ . Finally recall definitions (4.22) and (4.23).

**Lemma 4.16** (Concentration w.r.t.  $\mathbf{H}$ ). *For any  $s, \epsilon, \delta$  all in  $[0, 1]$ ,  $t = 1, \dots, T$ ,  $\nu > 0$  and any realization  $\mathbf{H}$  we have*

$$\mathbb{P}(|H_{t,s;\epsilon,\delta} - \mathbb{E}_{\mathbf{H}} H_{t,s;\epsilon,\delta}| \geq \nu/3) \leq 2 \exp \left( - \frac{2n\nu^2}{(3 \ln 2)^2} \right). \quad (4.122)$$

**Lemma 4.17** (Concentration w.r.t.  $\mathcal{J}$ ). *For any  $s, \epsilon, \delta$  all in  $[0, 1]$ ,  $t = 1, \dots, T$ ,  $\nu > 0$  and any realization  $\mathcal{J}$  there exists a constant  $C_1 > 0$  such that*

$$\mathbb{P}(|\mathbb{E}_{\mathbf{H}} H_{t,s;\epsilon,\delta} - \mathbb{E}_{\mathbf{H}, \mathcal{J}} H_{t,s;\epsilon,\delta}| \geq \nu/3) \leq 3 \exp(-n\nu^2 C_1). \quad (4.123)$$

**Lemma 4.18** (Concentration w.r.t.  $\mathcal{U}$ ). *For any  $s, \epsilon, \delta$  all in  $[0, 1]$ ,  $t = 1, \dots, T$ ,  $\nu > 0$  and any realization  $\mathcal{U}$  there exists a constant  $C_2 > 0$  such that*

$$\mathbb{P}(|\mathbb{E}_{\mathbf{H}, \mathcal{J}} H_{t,s;\epsilon,\delta} - h_{t,s;\epsilon,\delta}| \geq \nu/3) \leq 3 \exp(-n\nu^2 C_2). \quad (4.124)$$

Lemmas 4.16 to 4.18 are consequences of McDiarmid's inequality, which states that if  $X_1, \dots, X_N$  are independent variables and  $g$  is a function satisfying the bounded difference property

$$|g(x_1, \dots, x_i, \dots, x_N) - g(x_1, \dots, x'_i, \dots, x_N)| \leq d_i \quad \forall i = 1, \dots, N$$

then for any  $\nu > 0$  we have

$$\mathbb{P}(|g(\mathbf{X}) - \mathbb{E}_{\mathbf{X}} g(\mathbf{X})| \geq \nu) \leq 2 \exp\left(-\frac{2\nu^2}{\sum_{i=1}^N d_i^2}\right).$$

We provide the proof of those three lemmas at the end of this section.

From the triangle inequality and the union bound we have

$$\begin{aligned} \mathbb{P}(|H_{t,s;\epsilon,\delta} - h_{t,s;\epsilon,\delta}| \geq \nu) &\leq \mathbb{P}(|H_{t,s;\epsilon,\delta} - \mathbb{E}_{\mathbf{H}} H_{t,s;\epsilon,\delta}| \geq \nu/3) \\ &\quad + \mathbb{P}(|\mathbb{E}_{\mathbf{H}} H_{t,s;\epsilon,\delta} - \mathbb{E}_{\mathbf{H}, \mathcal{J}} H_{t,s;\epsilon,\delta}| \geq \nu/3) \\ &\quad + \mathbb{P}(|\mathbb{E}_{\mathbf{H}, \mathcal{J}} H_{t,s;\epsilon,\delta} - h_{t,s;\epsilon,\delta}| \geq \nu/3). \end{aligned} \quad (4.125)$$

From (4.122), (4.123), (4.124)

$$\mathbb{P}(|H_{t,s;\epsilon,\delta} - h_{t,s;\epsilon,\delta}| \geq \nu) \leq 8 \exp(-n\nu^2 C_0). \quad (4.126)$$

where  $C_0 \equiv \min\{\frac{2}{(3 \ln 2)^2}, C_1, C_2\}$ . Let  $D \equiv |H_{t,s;\epsilon,\delta} - h_{t,s;\epsilon,\delta}|$ . We have

$$\begin{aligned} \int_0^\infty d\nu \nu \mathbb{P}(D \geq \nu) &= \int_0^\infty d\nu \nu \mathbb{E}_D \mathbb{I}(D \geq \nu) = \mathbb{E}_D \int_0^\infty d\nu \nu \mathbb{I}(D \geq \nu) \\ &= \mathbb{E}_D \int_0^D d\nu \nu = \frac{1}{2} \mathbb{E}_D [D^2]. \end{aligned} \quad (4.127)$$

Substituting (4.126) into (4.127), we have the required bound for Lemma 4.12 with  $C = 8/C_0$ :

$$\begin{aligned} \mathbb{E}[(H_{t,s;\epsilon,\delta} - h_{t,s;\epsilon,\delta})^2] &= 2 \int_0^\infty d\nu \nu P(|H_{t,s;\epsilon,\delta} - h_{t,s;\epsilon,\delta}| \geq \nu) \\ &\leq 16 \int_0^\infty d\nu \nu e^{-n\nu^2 C_0} = \frac{C}{n}. \end{aligned}$$

**Proof of Lemma 4.16**

Consider  $g(H_1, \dots, H_n) \equiv H_{t,s;\epsilon,\delta}$  with  $H_i \in \{0, \infty\}$  (note that  $H_{t,s;\epsilon,\delta}$  given by (4.23) is already averaged over  $\tilde{\mathbf{H}}$ , but not over  $\mathbf{H}$ ). As for all  $i = 1, \dots, n$  the function  $g$  satisfies

$$\begin{aligned} |g(H_1, \dots, H_i, \dots, H_n) - g(H_1, \dots, H'_i, \dots, H_n)| &= \left| \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}} \ln \langle e^{H_i(\sigma_i-1)} \rangle_{t,s;\epsilon,\delta} \right| \\ &= \left| \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}} \ln(1 + \langle \sigma_i \rangle_{t,s;\epsilon,\delta} \tanh H_i) - \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}} \ln(1 + \tanh H_i) \right| \\ &\leq \frac{\ln 2}{n}. \end{aligned}$$

McDiarmid's inequality immediately gives the lemma.

**Proof of Lemma 4.17**

Let  $|\mathbf{J}|$  be the number of components of the vector  $\mathbf{J}$ . From the construction of  $\mathcal{G}_{t,s}$  in Sec. 4.4.1, we have  $\mathbb{E}[|\mathbf{J}|] = \frac{\alpha n}{T}(T-t+1-s) \leq \alpha n$ . Set  $m_{\max} = (1+\gamma)\alpha n$  for  $\gamma > 0$ . The probability of the event  $|\mathbf{J}| > m_{\max}$  can be bounded by a relaxed form of the Chernoff bound as follows.

**Lemma 4.19** (Chernoff bound, [74, Theorem 4.4]). *Let  $X = \sum_{i=1}^N X_i$  where  $X_i = 1$  with probability  $p_i$  and  $X_i = 0$  with probability  $1 - p_i$ , and all  $X_i$  are independent. Let  $\mu = \mathbb{E}[X] = \sum_{i=1}^N p_i$ . Then for all  $\gamma > 0$*

$$\mathbb{P}(X > (1 + \gamma)\mu) \leq \exp\left(-\frac{\mu}{3} \min\{\gamma, \gamma^2\}\right).$$

By the Chernoff bound we have

$$\mathbb{P}(|\mathbf{J}| > m_{\max}) \leq \exp\left(-\frac{\alpha n}{3} \min\{\gamma, \gamma^2\}\right). \quad (4.128)$$

Conditioned on  $|\mathbf{J}| \leq m_{\max}$ , we can have the representation  $\mathcal{J} = (c_1, \dots, c_{m_{\max}})$  where for  $a = 1, \dots, m_{\max}$  the profile  $c_a \equiv (A_a, J_a)$  encodes that a factor node with weight  $J_a$  is connected to a  $K$ -tuple identified by  $A_a$ . For  $m < a \leq m_{\max}$  we denote  $c_a = (\emptyset, 0)$ .

Now consider  $g(c_1, \dots, c_{m_{\max}}) \equiv \mathbb{E}_{\mathbf{H}} H_{t,s;\epsilon,\delta}$  and pick a  $c_a$  for a given  $a$ . Let  $c'_a \equiv (A'_a, J'_a)$  be a new profile with either  $A_a \neq A'_a$  or  $J_a \neq J'_a$ . Also let  $c''_a \equiv (A_a, 0)$  and  $c'''_a \equiv (A'_a, 0)$ . Note that

$$g(c_1, \dots, c''_a, \dots, c_{m_{\max}}) = g(c_1, \dots, c'''_a, \dots, c_{m_{\max}}).$$

We then have

$$\begin{aligned}
& |g(c_1, \dots, c_a, \dots, c_{m_{\max}}) - g(c_1, \dots, c'_a, \dots, c_{m_{\max}})| \\
&= |g(c_1, \dots, c_a, \dots, c_{m_{\max}}) - g(c_1, \dots, c''_a, \dots, c_{m_{\max}}) \\
&\quad + g(c_1, \dots, c''_a, \dots, c_{m_{\max}}) - g(c_1, \dots, c'_a, \dots, c_{m_{\max}})| \\
&\leq |g(c_1, \dots, c_a, \dots, c_{m_{\max}}) - g(c_1, \dots, c''_a, \dots, c_{m_{\max}})| \\
&\quad + |g(c_1, \dots, c''_a, \dots, c_{m_{\max}}) - g(c_1, \dots, c'_a, \dots, c_{m_{\max}})| \\
&= \left| \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{H}} \ln \langle e^{J_a(\sigma_{A_a}-1)} \rangle_{t,s;\epsilon,\delta} \right| + \left| \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{H}} \ln \langle e^{J'_a(\sigma_{A'_a}-1)} \rangle_{t,s;\epsilon,\delta} \right| \\
&= \left| \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{H}} \ln(1 + \langle \sigma_{A_a} \rangle_{t,s;\epsilon,\delta} \tanh J_a) - \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{H}} \ln(1 + \tanh J_a) \right| \\
&\quad + \left| \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{H}} \ln(1 + \langle \sigma_{A'_a} \rangle_{t,s;\epsilon,\delta} \tanh J'_a) - \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{H}} \ln(1 + \tanh J'_a) \right| \\
&\leq \frac{2 \ln 2}{n}
\end{aligned}$$

This allows the use of McDiarmid's inequality to obtain

$$\mathbb{P}(|\mathbb{E}_{\mathbf{H}} H_{t,s;\epsilon,\delta} - \mathbb{E}_{\mathbf{H}, \mathcal{J}} H_{t,s;\epsilon,\delta}| \geq \nu/3 \mid |\mathbf{J}| \leq m_{\max}) \leq 2 \exp \left( - \frac{n\nu^2}{18\alpha(\ln 2)^2} \right). \quad (4.129)$$

Finally, we take the union bound based on (4.128) and (4.129):

$$\begin{aligned}
& \mathbb{P}(|\mathbb{E}_{\mathbf{H}} H_{t,s;\epsilon,\delta} - \mathbb{E}_{\mathbf{H}, \mathcal{J}} H_{t,s;\epsilon,\delta}| \geq \nu/3) \\
& \leq 2 \exp \left( - \frac{n\nu^2}{18\alpha(\ln 2)^2} \right) + \exp \left( - \frac{\alpha n}{3} \min\{\gamma, \gamma^2\} \right).
\end{aligned}$$

Choosing  $\nu^2 = \min\{\gamma, \gamma^2\}$  and  $C_1 = \min\{\frac{1}{18\alpha(\ln 2)^2}, \frac{\alpha}{3}\}$ , we obtain the lemma.

### Proof of Lemma 4.18

This proof can adopt the same presentation as in the proof of Lemma 4.17 by noting that in the construction of  $\mathcal{G}_{t,s}$  the Poisson process of adding half edges with weight  $U_{a \rightarrow i}^{(t')}$  can be rephrased as follows:

1. (*Create all the messages without specifying their location*): We draw the random numbers  $e_i^{(t')}$ ,  $e_{i,s}^{(t)}$  and create the associated number of copies of  $U^{(t')}$  for  $t' = 1, \dots, t$ . We collect all  $U^{(t')}$  to form a set  $\{U_k\}_{k=1}^w$ , where  $w$  follows a Poisson distribution with mean  $\frac{n\alpha K}{T}(t-1+s) \leq n\alpha K$ .
2. (*Specify the location of the messages*): Given the number  $w$  and the set  $\{U_k\}$ , we attach each  $U_k$  to variable node  $i$  chosen randomly and uniformly.

Let  $w_{\max} = (1 + \gamma)n\alpha K$ . The Chernoff bound (Lemma 4.19) provides that

$$\mathbb{P}(w > w_{\max}) \leq \exp\left(-\frac{n\alpha K}{3} \min\{\gamma, \gamma^2\}\right). \quad (4.130)$$

Conditioned on  $w \leq w_{\max}$ , we have the representation  $\mathcal{U} = (u_1, \dots, u_{w_{\max}})$  where for  $k = 1, \dots, w_{\max}$  the profile  $u_k = (i_k, U_k)$  represents that a half edge with weight  $U_k$  is connected to variable node  $i_k$ . For  $w < k \leq w_{\max}$  we denote  $u_k = (\emptyset, 0)$ .

Now consider  $g(u_1, \dots, u_{w_{\max}}) \equiv \mathbb{E}_{\mathbf{H}, \mathcal{J}} h_{t, s; \epsilon, \delta}$  and pick any  $u_k$ . Let  $u'_k \equiv (i'_k, U_k)$  be a new profile with either  $i_k \neq i'_k$  or  $U_k \neq U'_k$ . Also let  $u''_k = (i_k, 0)$  and  $u'''_k = (i'_k, 0)$ . Note that  $g(u_1, \dots, u''_k, \dots, u_{w_{\max}}) = g(u_1, \dots, u'''_k, \dots, u_{w_{\max}})$ . We then have

$$\begin{aligned} & |g(u_1, \dots, u_k, \dots, u_{w_{\max}}) - g(u_1, \dots, u'_k, \dots, u_{w_{\max}})| \\ &= |g(u_1, \dots, u_k, \dots, u_{w_{\max}}) - g(u_1, \dots, u''_k, \dots, u_{w_{\max}}) \\ &\quad + g(u_1, \dots, u''_k, \dots, u_{w_{\max}}) - g(u_1, \dots, u'_k, \dots, u_{w_{\max}})| \\ &\leq |g(u_1, \dots, u_k, \dots, u_{w_{\max}}) - g(u_1, \dots, u''_k, \dots, u_{w_{\max}})| \\ &\quad + |g(u_1, \dots, u''_k, \dots, u_{w_{\max}}) - g(u_1, \dots, u'_k, \dots, u_{w_{\max}})| \\ &= \left| \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{H}, \mathcal{J}} \ln \langle e^{U_k(\sigma_{i_k} - 1)} \rangle_{t, s; \epsilon, \delta} \right| + \left| \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{H}, \mathcal{J}} \ln \langle e^{U'_k(\sigma_{i'_k} - 1)} \rangle_{t, s; \epsilon, \delta} \right| \\ &= \left| \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{H}, \mathcal{J}} \ln(1 + \langle \sigma_{i_k} \rangle_{t, s; \epsilon, \delta} \tanh U_k) - \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{H}, \mathcal{J}} \ln(1 + \tanh U_k) \right| \\ &\quad + \left| \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{H}, \mathcal{J}} \ln(1 + \langle \sigma_{i'_k} \rangle_{t, s; \epsilon, \delta} \tanh U'_k) - \frac{1}{n} \mathbb{E}_{\tilde{\mathbf{H}}, \mathbf{H}, \mathcal{J}} \ln(1 + \tanh U'_k) \right| \\ &\leq \frac{2 \ln 2}{n}. \end{aligned}$$

McDiarmid's inequality is then used to obtain

$$\mathbb{P}(|\mathbb{E}_{\mathbf{H}, \mathcal{J}} H_{t, s; \epsilon, \delta} - h_{t, s; \epsilon, \delta}| \geq \nu/3 \mid w \leq w_{\max}) \leq 2 \exp\left(-\frac{n\nu^2}{18(\ln 2)^2 \alpha K}\right). \quad (4.131)$$

Finally, we take the union bound based on (4.130) and (4.131):

$$\begin{aligned} & \mathbb{P}(|\mathbb{E}_{\mathbf{H}, \mathcal{J}} H_{t, s; \epsilon, \delta} - h_{t, s; \epsilon, \delta}| \geq \nu/3) \\ & \leq 2 \exp\left(-\frac{n\nu^2}{18(\ln 2)^2 \alpha K}\right) + \exp\left(-\frac{n\alpha K}{3} \min\{\gamma, \gamma^2\}\right). \end{aligned}$$

Choosing  $\nu^2 = \min\{\gamma, \gamma^2\}$  and  $C_2 = \min\{\frac{R}{18(\ln 2)^2 K}, \frac{K}{3R}\}$ , we obtain the lemma.

### 4.8.8 Illustration of the replica formula

Recall that the distribution of  $V$  is denoted by  $\mathbf{x} = x\Delta_0 + (1-x)\Delta_\infty$ ,  $x \in [0, 1]$ . From (4.4) the distribution of  $\mathbf{U}$  is  $\tilde{\mathbf{x}} = \tilde{x}\Delta_0 + (1-\tilde{x})\Delta_\infty$  where  $\tilde{x} = 1 - (1 -$

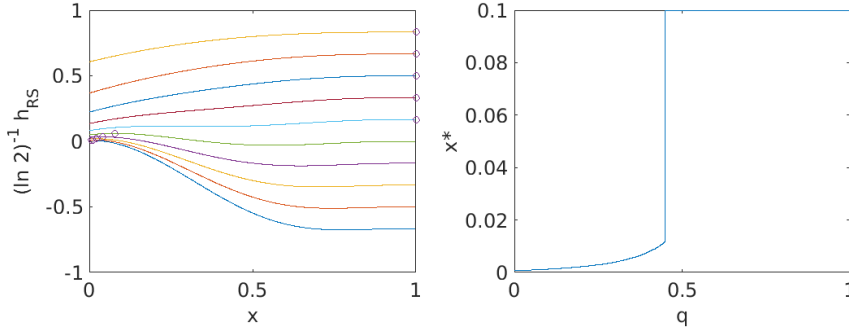


Figure 4.2 – Illustration of  $h_{\text{RS}}(x)$  with  $K = 3$  and  $\alpha = 1/5$ . (Left)  $h_{\text{RS}}(x)$  as a function of  $x$  for  $q = 0, 0.1, 0.2, \dots, 0.9$ .  $h_{\text{RS}}(x)$  increases with  $q$  when  $x$  is fixed. Circles locate the maximum of  $h_{\text{RS}}(x)$  for every  $q$ . (Right) The first order phase transition for  $x^*(q) = \arg \max_x h_{\text{RS}}(x)$  as a function of  $q$

$q)(1-x)^{K-1} \in [0, 1]$ . The first term of (4.5) can be simplified as

$$\begin{aligned}
 & \mathbb{E} \ln \left( \prod_{B=1}^l (1 + \tanh U_B) + \prod_{B=1}^l (1 - \tanh U_B) \right) \\
 &= \mathbb{E}_l \mathbb{E}_U \left[ l \ln(1 + \tanh U) \right] + \mathbb{E}_l \mathbb{E}_U \ln \left( 1 + \prod_{B=1}^l \frac{1 - \tanh U_B}{1 + \tanh U_B} \right) \\
 &= \mathbb{E}_l [l] (1 - \tilde{x}) \ln 2 + \mathbb{E}_l [\tilde{x}^l] \ln 2 \\
 &= \alpha K (1 - q) (1 - x)^{K-1} \ln 2 + e^{-\alpha K (1-q)(1-x)^{K-1}} \ln 2
 \end{aligned}$$

The remaining terms of (4.5) can also be simplified straightforwardly. Eventually, for BEC we can write (4.5) with a scalar expression:

$$\begin{aligned}
 h_{\text{RS}}(x) &= (\ln 2) \left[ e^{-\alpha K (1-q)(1-x)^{K-1}} + \alpha K (1-q) (1-x)^{K-1} \right. \\
 &\quad \left. - \alpha (K-1) (1-q) (1-x)^K - \alpha (1-q) \right]
 \end{aligned}$$

We have illustrated  $h_{\text{RS}}(x)$  with  $K = 3$  and  $\alpha = 1/5$  in Fig. 4.2.





---

# Multi-overlaps for Ferromagnetic Spin Models on Sparse Graphs

---

# 5

## 5.1 Introduction

The adaptive interpolation for sparse graphs (developed in Chapter 4) shows that the sum rule involves a sequence of overlap parameters  $\{Q_p : p \geq 1\}$ . This is in contrast to the dense graphs, where only the lowest order overlaps are involved. Concentration of total fluctuation in the form  $\mathbb{E}((Q_p - \mathbb{E}\langle Q_p \rangle)^2)$  is critically required for the adaptive interpolation to fully validate the replica symmetric formula. The control for  $p \geq 2$  sets a new challenge. Thanks to Lemma 4.4, the binary erasure channel is a special case where the overlaps  $Q_p$  for all  $p \geq 1$  are the same and thus only  $Q_1$  matters. Therefore, Lemma 4.6 (full concentration of  $Q_1$ ) suffices to complete the proof in Sec. 4.4.4. How to generalize Lemma 4.6 for  $p \geq 1$  and for other channels in an inference setting remains an open problem. Therefore, we shall return to the origin in physics and consider even simpler models for inspiration and future progress.

Spin models in physics have the same notion of overlaps. For a system with binary spins  $\sigma_i \in \{-1, +1\}$ ,  $i = 1, \dots, n$ , the overlap parameters are generally defined as  $Q_p \equiv \frac{1}{n} \sum_{i=1}^n \sigma_i^{(1)} \sigma_i^{(2)} \dots \sigma_i^{(p)}$  where  $p \geq 1$  is an integer and  $(\sigma_i^{(\alpha)})_{i=1, \dots, n}^{\alpha=1, \dots, p}$  are distributed according to the replicated Gibbs distribution, in other words the product of  $p$  copies of the Gibbs distribution. It is known folklore that for “mean-field” spin models the concentration of the overlaps is an important ingredient for the validity of a replica symmetric expression for the free energy if it exists. However, to our knowledge, there is no direct logical implication that has been mathematically settled in a clear way. In [75], the authors show (for models on complete graphs) that if the free energy is not given by a replica symmetric expression then the overlaps cannot concentrate. Also, the Guerra-Toninelli interpolation method [76, 77, 78] makes it clear that if the free energy is not replica symmetric then the overlaps of an

“interpolated model” cannot concentrate. The Ghirlanda-Guerra identities in [79] provide a way to show the concentration of  $Q_1$  or  $Q_2$  for any spin model under a suitable perturbation. This is further explained in [80, 81, 82]. In [83, 84, 85, 86], non-trivial constraints analogous to the Aizenman-Contucci identities and Ghirlanda-Guerra identities [87] are derived for *all* overlaps of sparse models. But one cannot deduce their concentration from these constraints alone (nevertheless some of the techniques used in the present chapter are inspired from these works).

In this chapter, our main interest is the study of fluctuations and concentration properties of *all* overlaps  $\{Q_p, p \geq 1\}$  for the ferromagnetic spin models on sparse random hypergraphs (typically of Erdős-Rényi type). We distinguish two types of fluctuations, namely, the thermal ones and those with respect to the disorder, informally measured by the two quantities

$$\mathbb{E}\langle(Q_p - \langle Q_p \rangle)^2\rangle \quad \text{and} \quad \mathbb{E}[(\langle Q_p \rangle - \mathbb{E}\langle Q_p \rangle)^2].$$

Adding these two fluctuations one finds the total fluctuations

$$\mathbb{E}\langle(Q_p - \mathbb{E}\langle Q_p \rangle)^2\rangle.$$

Our main result states that both types of fluctuations vanish in the thermodynamic limit for all temperatures. For this result to hold at all temperatures, we must add suitable “infinitesimal” one-body perturbations to the Hamiltonian<sup>1</sup>. Indeed, concentration may hold only within a “pure state”<sup>2</sup>, and it is well known that one should add suitable perturbations in order to select pure states (that may coexist at low temperatures).

We would like to stress that, for disordered systems, the nature of the perturbation that one should add is not always clear. For example, all multi-spin infinitesimal interactions are sometimes added to the two-body Sherrington-Kirkpatrick Hamiltonian, and it is perhaps not so clear what the physical interpretation of such perturbations is [88]. Here, we limit ourselves to *simple one-body perturbations* that can physically be interpreted as infinitesimal external magnetic fields.

To the best of our knowledge, this is the first time a concentration result is established for *all* overlaps  $\{Q_p, p \geq 1\}$  in a dilute disordered spin model for all temperatures. Examples of models that are covered by our results are the pure and mixed  $K$ -spin ferromagnets on random sparse Erdős-Rényi hypergraphs. With minor adjustments in the formulation of the models, we can also cover ferromagnets on dense graphs. The coupling constants are ferromagnetic and this allows the use of the Griffith-Kelly-Sherman (GKS) inequality which plays an important role in our analysis.

<sup>1</sup>By infinitesimal perturbations we mean perturbations that do not change the thermodynamic limit of the free energy when we take the limit of zero perturbation after the thermodynamic limit.

<sup>2</sup>Recall the Curie-Weiss model introduced in Sec. 1.3. Without the external field  $h$ , the magnetization can be in different “pure states” at low temperatures.

In Section 5.2, we formulate the models and state our main theorems. The proofs are found in Section 5.3. The appendices contain technical intermediate results.

## 5.2 Ferromagnetic spin models and overlap concentration

Consider a collection of  $n$  binary spins  $\sigma_i \in \{-1, 1\}$ ,  $i = 1, \dots, n$ . For any subset  $A \subset \{1, \dots, n\}$  we denote  $\sigma_A = \prod_{i \in A} \sigma_i$ . A generic ferromagnetic spin system has Hamiltonian

$$\mathcal{H}_0(\boldsymbol{\sigma}) \equiv - \sum_{A \subset \{1, \dots, n\}} J_A \sigma_A \quad (5.1)$$

where  $J_A \geq 0$  and the sum runs over all possible  $2^n$  subsets of  $\{1, \dots, n\}$ . The only subsets of spins that truly participate in the interactions are of course those for which  $J_A > 0$ . The random models that we consider have independently distributed coupling constants  $J_A$ ,  $A \subset \{1, \dots, n\}$ , with distribution supported on  $\mathbb{R}_{\geq 0}$ . As said in the introduction our main interest is in sparse systems, a typical example of which is given below.

The thermodynamic potential of interest is the free energy

$$F_n \equiv -\frac{1}{n} \ln \mathcal{Z} = -\frac{1}{n} \ln \sum_{\boldsymbol{\sigma} \in \{\pm 1\}^n} \exp(-\mathcal{H}_0(\boldsymbol{\sigma}))$$

where  $\mathcal{Z}$  is the partition function of the model. The average free energy is defined as  $f_n \equiv \mathbb{E} F_n$  where  $\mathbb{E}$  is the expectation over all the coupling constants. For models of physical interest one expects that  $F_n$  concentrates over  $f_n$ . Our theorems on overlap concentration stated below are formulated in a generic setting and hold as long as the concentration of the free energy holds:

$$\mathbb{E}[(F_n - f_n)^2] \leq \frac{C_F}{n} \quad (5.2)$$

for  $C_F > 0$  a constant independent of  $n$ . In Appendix 5.4.3 we verify by standard arguments that the following simple condition implies (5.2) (we do not need  $J_A \geq 0$  for this implication, see Proposition 5.1):

**Condition 5.1.** *We assume that  $J_A$ ,  $A \subset \{1, \dots, n\}$  are independent random variables with finite second moment and such that  $\sum_{A \subset \{1, \dots, n\}} \text{Var}(J_A) \leq C_F n$  for a numerical constant  $C_F > 0$  independent of  $n$ .*

Models of physical interest also have a well-defined thermodynamic limit for  $f_n$ . This requires a little bit more structure on the distribution of the couplings  $J_A$  and will not be used (see [89, 90] for proofs of the existence of such limits). For completeness we give a simple hypothesis and standard argument

in Appendix 5.4.3 that guarantees the existence of the thermodynamic limit in the case of ferromagnetic models (see Proposition 5.2).

Let us give a canonical example of ferromagnetic spin system on a sparse random graph where Condition 5.1 is satisfied as well as the existence of the thermodynamic limit of the free energy. Note that our results also cover dense graph systems as long as  $J_A$  are suitably rescaled with  $n$  so that Condition 5.1 is met.

**Example 5.1** ( $K$ -spin models on the Erdős-Rényi hypergraph). *A sparse ferromagnetic  $K$ -spin model with coupling strength  $J > 0$  and magnetic field  $H > 0$  can be constructed as follows. For all subsets  $A \subset \{1, \dots, n\}$  with cardinalities different from 1 and  $K$  set  $J_A = 0$ . For all  $A$  such that  $|A| = 1$  set  $J_A = H$ . In other words the Hamiltonian contains the one-body term  $-H \sum_{i=1}^n \sigma_i$ . For all subsets with  $|A| = K$  take for  $J_A$  Bernoulli random variables with  $\mathbb{P}(J_A = J) = \gamma n / \binom{n}{K}$  and  $\mathbb{P}(J_A = 0) = 1 - \gamma n / \binom{n}{K}$  where  $\gamma > 0$ . The Hamiltonian thus contains on average of the order of  $\gamma n$  interaction terms of the form  $-J \sigma_{a_1} \sigma_{a_2} \dots \sigma_{a_K}$  where  $a_i, i = 1, \dots, K$  are chosen uniformly at random in  $\{1, \dots, n\}$  without repetition.*

*This model can be generalized to mixed  $K$ -spin models as follows: Fix  $H > 0, J_2, \dots, J_{K^*} > 0, \gamma_2, \dots, \gamma_{K^*} > 0$ . Let  $k = 2, \dots, K^*$ . We then draw Bernoulli random variables for the couplings of subsets with cardinality  $|A| = k$  such that  $\mathbb{P}(J_A = J_k) = \gamma_k n / \binom{n}{k}$  and  $\mathbb{P}(J_A = 0) = 1 - \gamma_k n / \binom{n}{k}$ . And again of course  $J_A = H$  for  $A$  such that  $|A| = 1$  and  $J_A = 0$  if instead  $|A| \neq 1, 2, \dots, K^*$ . These models are generalizations of the Ising two-body ferromagnet on a standard Erdős-Rényi random graph. In Appendix 5.4.3 we verify that Condition 5.1 is satisfied so that  $F_n$  concentrates on  $f_n$ , and also that the thermodynamic limit of  $f_n$  exists.*

Any observable is a linear combination over subsets  $T \subset \{1, \dots, n\}$  of  $\sigma_T \equiv \prod_{i \in T} \sigma_i$  and their Gibbs expectation is denoted by

$$\langle \sigma_T \rangle \equiv \frac{1}{Z} \sum_{\sigma \in \{\pm 1\}^n} \sigma_T \exp(-\mathcal{H}_0(\sigma)).$$

The crucial property of ferromagnetic models that will be instrumental in our analysis are the Griffiths-Kelly-Sherman (GKS) correlation inequalities (1.9) and (1.10).

The *multi-overlaps* (simply called overlaps) are defined for any integer  $p \geq 1$  as

$$Q_p \equiv \frac{1}{n} \sum_{i=1}^n \sigma_i^{(1)} \dots \sigma_i^{(p)} \quad (5.3)$$

where  $\{\sigma^{(\alpha)}, \alpha = 1, \dots, p\}$  is a set of  $p$  replicas of the spin configurations drawn according to the  $p$ -fold tensor product of the Gibbs measure. We emphasize that in this work the replicas are always uncoupled and i.i.d.. The Gibbs

average w.r.t. the tensor product Gibbs measure is still indicated as  $\langle - \rangle$ . Note that

$$\langle Q_p \rangle = \frac{1}{n} \sum_{i=1}^n \langle \sigma_i^{(1)} \cdots \sigma_i^{(p)} \rangle = \frac{1}{n} \sum_{i=1}^n \langle \sigma_i \rangle^p, \quad p \geq 1.$$

It is well known that concentration results for overlaps generally require the addition of small perturbation terms whose role is to select “pure states”. With suitable such perturbations, and under a suitable concentration hypothesis for the free energy (of the perturbed model) we show that for large  $n$ : *i) For any instance of the random model (i.e. of the quenched disorder)  $Q_p$  concentrates over  $\langle Q_p \rangle$ ; and ii)  $\langle Q_p \rangle$  concentrates over  $\mathbb{E} \langle Q_p \rangle$ .* These two concentration properties imply that overall  $Q_p$  concentrates on  $\mathbb{E} \langle Q_p \rangle$ . One then expects that the free energy is given by the replica symmetric formula but this is still an open problem.

We consider one-body perturbation terms  $\mathcal{H}_{\text{pert}}(\boldsymbol{\sigma})$  added to the generic Hamiltonian (5.1):

$$\mathcal{H}_{\text{pert}}(\boldsymbol{\sigma}) \equiv -h_0 \sum_{i=1}^n \sigma_i - h_1 \sum_{i=1}^n \tau_i \sigma_i \quad (5.4)$$

with  $h_0 \in [0, 1]$ ,  $h_1 \in [0, 1]$ ,  $\tau_i \sim \text{Poi}(\alpha n^{\theta-1})$  i.i.d. for  $i = 1, \dots, n$  and  $\alpha \in [0, 1]$ ,  $\theta \in (1/2, 7/8]$ . The first part of the perturbation, proportional to  $h_0$ , is called *homogeneous perturbation* while the second part proportional to  $h_1$  is called *Poisson perturbation*. Both are purely ferromagnetic such that the GKS inequalities remain valid. Note that in distribution  $h_1 \sum_{i=1}^n \tau_i \sigma_i \stackrel{d}{=} h_1 \sum_{v=1}^{\Gamma} \sigma_{i_v}$  where  $\Gamma \sim \text{Poi}(\alpha n^{\theta})$  and  $i_v$  is randomly and uniformly chosen from  $\{1, \dots, n\}$ . While this second expression might seem more natural, the first equivalent expression allows a more compact notation in our analysis. We associate to the total Hamiltonian (5.1) + (5.4), i.e.

$$\mathcal{H}(\boldsymbol{\sigma}) \equiv \mathcal{H}_0(\boldsymbol{\sigma}) + \mathcal{H}_{\text{pert}}(\boldsymbol{\sigma}), \quad (5.5)$$

its partition function  $\mathcal{Z}_{h_0, h_1, \alpha}$ , Gibbs expectation  $\langle - \rangle_{h_0, h_1, \alpha}$ , free energy

$$F_n(h_0, h_1, \alpha) \equiv -\frac{1}{n} \ln \mathcal{Z}_{h_0, h_1, \alpha}$$

and average free energy  $f_n(h_0, h_1, \alpha) \equiv \mathbb{E} F_n(h_0, h_1, \alpha)$  defined similarly as before with  $\mathcal{H}_0$  replaced by  $\mathcal{H}$ . Here  $\mathbb{E}$  is the expectation over all quenched variables, i.e.,  $J_A$ ,  $A \subset \{1, \dots, n\}$  and  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)$ . An elementary argument shows that the perturbation does not modify the thermodynamic properties as long as  $h_0 \rightarrow 0_+$  ( $\alpha$  can be taken fixed). More precisely in Appendix 5.4.1 we show

$$|f_n(h_0, h_1, \alpha) - f_n(0, 0, 0)| \leq h_0 + \frac{\alpha h_1}{n^{1-\theta}}. \quad (5.6)$$

In particular,  $\lim_{h_0 \rightarrow 0_+} \lim_{n \rightarrow +\infty} |f_n(h_0, h_1, \alpha) - f_n| = 0$ . Note also that the pressure associated with the perturbed Hamiltonian satisfies the concentration requirement (5.2) as soon as the unperturbed one does.

We can now state the main concentration results. From now on, in the rest of the paper it is understood that  $n$  is always large enough.

**Theorem 5.1** (Thermal concentration of the overlaps). *Assume the Hamiltonian  $\mathcal{H}$  given by (5.5) satisfies  $\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \geq 0$  for all  $i, j = 1, \dots, n$ . Then for any  $[\underline{h}, \bar{h}] \subset (0, 1)$ ,  $h_1 \in [0, 1]$ ,  $\alpha \in [0, 1]$ , we have for any instance of the random Hamiltonian*

$$\int_{\underline{h}}^{\bar{h}} dh_0 \langle (Q_p - \langle Q_p \rangle_{h_0, h_1, \alpha})^2 \rangle_{h_0, h_1, \alpha} \leq \frac{2p}{n}.$$

The next two results use the concentration of the free energy (5.2) for the total Hamiltonian (5.5). This holds under Condition 5.1 and the constant  $C_F$  can easily be made independent of  $h_0, h_1, \alpha$ .

**Theorem 5.2** (Total concentration of the magnetization/first overlap). *Assume the Hamiltonian  $\mathcal{H}$  given by (5.5) satisfies  $\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \geq 0$  for all  $i, j = 1, \dots, n$  and assume also that Condition 5.1 holds. Then for any  $[\underline{h}, \bar{h}] \subset (0, 1)$ ,  $h_1 \in [0, 1]$ ,  $\alpha \in [0, 1]$  we have*

$$\int_{\underline{h}}^{\bar{h}} dh_0 \mathbb{E} \langle (Q_1 - \mathbb{E} \langle Q_1 \rangle_{h_0, h_1, \alpha})^2 \rangle_{h_0, h_1, \alpha} \leq \frac{15C_F + 40}{n^{1/3}}.$$

*Remark:* Let us make a few remarks about these two theorems. First of all, the Poisson perturbation is not needed and we can set  $h_1 = \alpha = 0$ . Second, when both GKS inequalities hold  $2p/n$  can be replaced by  $p/n$  in Theorem 5.1 as it will become clear from the proof. More interestingly: Assuming only  $\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \geq 0$  is weaker than assuming both GKS inequalities and even weaker than assuming only the second one. This assumption is satisfied (for example) for all Hamiltonians satisfying the Fortuin-Kasteleyn-Ginibre (FKG) inequality. An example is  $J_A \geq 0$  except for one-body terms (magnetic fields) that may have arbitrary sign. So in particular, Theorems 5.1 and 5.2 cover the random field Ising model (RFIM). Another example is strong enough ferromagnetic two-body terms and any sign for magnetic fields and higher order interactions.

The next theorem assumes both GKS inequalities and its extension to systems satisfying only FKG is an open problem. It would be of interest to extend this theorem to the RFIM. Moreover, both the homogeneous and Poisson perturbations play an important role in the proof.

**Theorem 5.3** (Total concentration of the overlaps). *Assume the Hamiltonian  $\mathcal{H}$  given by (5.5) satisfies both GKS inequalities (in other words assume all*

$J_A \geq 0$ ) and also that Condition 5.1 holds. Then for any  $[\underline{h}, \bar{h}] \subset (0, 1)$ ,  $[\underline{\alpha}, \bar{\alpha}] \subset (0, 1)$ ,  $h_1 \in (0, 1]$  and  $\theta \in (1/2, 7/8]$  we have

$$\int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \mathbb{E} \langle (Q_p - \mathbb{E} \langle Q_p \rangle_{h_0, h_1, \alpha})^2 \rangle_{h_0, h_1, \alpha} \leq \frac{4p^2}{(\tanh h_1)^p} \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3}}.$$

The bound yields a decay  $\mathcal{O}(n^{-1/8})$  for  $\theta = 7/8$ . Also note that the prefactor on the r.h.s. grows exponentially fast with  $k$ . The details of the proof show that a slowly growing  $h_1$  can be accommodated and we can take  $h_1 = \mathcal{O}(\ln n)$  to mitigate this growth.

## 5.3 Proofs of concentrations for the overlaps

The main aim of this section is to prove Theorem 5.3. The proof is generic and essentially requires only two ingredients: *i*) That the Hamiltonian  $\mathcal{H}$  given by (5.5) is purely ferromagnetic so that the two GKS inequalities, (1.9) and (1.10) are verified; *ii*) the free energy of the perturbed model concentrates in the sense of (5.2). In the process we also obtain the proofs of Theorems 5.1 and 5.2.

To ease the notations in this section we do not indicate explicitly the arguments  $h_0, h_1, \alpha$  in the Gibbs brackets and free energy.

### 5.3.1 Preliminary remarks

Theorem 5.3 will be a consequence of the individual control of three types of overlap fluctuations. One can verify by expanding the squares that the total overlaps fluctuations can be decomposed as

$$\begin{aligned} \mathbb{E} \langle (Q_p - \mathbb{E} \langle Q_p \rangle)^2 \rangle &= \mathbb{E} \langle (Q_p - \langle Q_p \rangle)^2 \rangle + \mathbb{E} [(\langle Q_p \rangle - \mathbb{E}_{\boldsymbol{\tau}} \langle Q_p \rangle)^2] \\ &\quad + \mathbb{E} [(\mathbb{E}_{\boldsymbol{\tau}} \langle Q_p \rangle - \mathbb{E} \langle Q_p \rangle)^2]. \end{aligned} \quad (5.7)$$

The first type of fluctuations are purely *thermal fluctuations* and are controlled in Theorem 5.1 thanks to the homogeneous part of the perturbation in (5.4); for the analysis of these fluctuations the Poisson perturbation can be dropped. The last two terms are the *disorder fluctuations* due to the quenched variables. Their control requires the Poisson perturbation<sup>3</sup>. The second term are fluctuations directly related to the Poisson perturbation itself, called *Poisson fluctuations*, and is controlled by Lemma 5.2. Here,  $\mathbb{E}_{\boldsymbol{\tau}}$  is the expectation with respect to the Poisson random variables  $\boldsymbol{\tau}$ , with  $\tau_i \sim \text{Poi}(\alpha n^{\theta-1})$  i.i.d. for  $i = 1, \dots, n$ . The third term are the fluctuations due to all other quenched couplings in the unperturbed Hamiltonian and is controlled by Lemma 5.6. In Section 5.3.6 we show how to combine all these concentration results in order to obtain Theorem 5.3.

<sup>3</sup>It is an open problem to assess if these can be dropped and the fluctuations controlled only thanks to the homogeneous perturbation.

### 5.3.2 Thermal fluctuations of overlaps: proof of Theorem 5.1

We start by considering the thermal fluctuations for a fixed realization of quenched variables. Note that

$$\frac{dF_n}{dh_0} = -\frac{1}{n} \sum_{i=1}^n \langle \sigma_i \rangle = -\langle Q_1 \rangle, \quad (5.8)$$

$$\frac{1}{n} \frac{d^2 F_n}{dh_0^2} = -\frac{1}{n^2} \sum_{i,j=1}^n (\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle) = -\langle (Q_1 - \langle Q_1 \rangle)^2 \rangle. \quad (5.9)$$

The second identity shows that the free energy  $F_n$ , as well as its expectation  $f_n$ , are concave in  $h_0$  (a generic fact in statistical mechanics models).

By the definition (5.3) of  $Q_p$  we have

$$\begin{aligned} \langle (Q_p - \langle Q_p \rangle)^2 \rangle &= \langle Q_p^2 \rangle - \langle Q_p \rangle^2 \\ &= \frac{1}{n^2} \sum_{i,j=1}^n (\langle \sigma_i \sigma_j \rangle^p - \langle \sigma_i \rangle^p \langle \sigma_j \rangle^p) \\ &= \frac{1}{n^2} \sum_{i,j=1}^n (\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle) \sum_{l=0}^{p-1} \langle \sigma_i \sigma_j \rangle^{p-1-l} \langle \sigma_i \rangle^l \langle \sigma_j \rangle^l. \end{aligned} \quad (5.10)$$

Using  $\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \geq 0$  for all  $i, j = 1, \dots, n$  and the triangle inequality, (5.10) is upper bounded as

$$\langle (Q_p - \langle Q_p \rangle)^2 \rangle \leq \frac{p}{n^2} \sum_{i,j=1}^n (\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle).$$

Hence, integrating this inequality over  $h_0 \in [\underline{h}, \bar{h}]$  and using (5.9) we have

$$\int_{\underline{h}}^{\bar{h}} dh_0 \langle (Q_p - \langle Q_p \rangle)^2 \rangle \leq -p \int_{\underline{h}}^{\bar{h}} dh_0 \frac{1}{n} \frac{d^2 F_n}{dh_0^2} = \frac{p}{n} [\langle Q_1 \rangle]_{h_0=\underline{h}}^{h_0=\bar{h}} \leq \frac{2p}{n}.$$

Note that if the first GKS inequality also holds then  $\langle Q_1 \rangle \geq 0$ . Therefore,

$$0 \leq [\langle Q_1 \rangle]_{h_0=\underline{h}}^{h_0=\bar{h}} \leq 1,$$

and  $2p/n$  becomes  $p/n$ . □

### 5.3.3 Disorder fluctuations of the magnetization

Before considering the concentration of general overlaps we need to control the quenched fluctuations of the first overlap  $Q_1$ , that is the magnetization. Indeed, our proof of the concentration of the overlaps w.r.t. the quenched variables in Section 5.3.5 is based on an induction argument where the induction is on the order  $p$  of the overlaps  $Q_p$ , and the following lemma will serve as the base case for the induction. In order to control these fluctuations the homogeneous perturbation alone is again sufficient.



**Lemma 5.1** (Concentration of the magnetization w.r.t. the quenched disorder). *Assume that Condition 5.1 holds. Then for any  $[\underline{h}, \bar{h}] \subset (0, 1)$  we have*

$$\int_{\underline{h}}^{\bar{h}} dh_0 \mathbb{E}[(\langle Q_1 \rangle - \mathbb{E}\langle Q_1 \rangle)^2] \leq \frac{15C_F + 40}{n^{1/3}}.$$

*Remark:* There is no need to assume  $J_A \geq 0$  here. So this lemma holds generally even if GKS or FKG inequalities do not hold.

*Remark:* Combining Theorem 5.1 and Lemma 5.1 yields Theorem 5.2.

*Proof.* Below it will be convenient to indicate explicitly the  $h_0$  dependence in the free energy. Recall  $f_n(h_0) = \mathbb{E} f_n(h_0)$ . From (5.9) we have

$$\langle Q_1 \rangle - \mathbb{E}\langle Q_1 \rangle = \frac{df_n(h_0)}{dh_0} - \frac{dF_n(h_0)}{dh_0}.$$

Since  $F_n$  and  $f_n$  are concave in  $h_0$  as seen from (5.9), we can use Lemma 3.12 to obtain

$$|\langle Q_1 \rangle - \mathbb{E}\langle Q_1 \rangle| \leq \delta^{-1} \sum_{u \in \{h_0 - \delta, h_0 + \delta\}} |F_n(u) - f_n(u)| + C_\delta^+(h_0) + C_\delta^-(h_0)$$

where

$$\begin{cases} C_\delta^+(h_0) \equiv \frac{df_n(h_0)}{dh_0} - \frac{df_n(h_0 + \delta)}{dh_0} \geq 0, \\ C_\delta^-(h_0) \equiv \frac{df_n(h_0 - \delta)}{dh_0} - \frac{df_n(h_0)}{dh_0} \geq 0. \end{cases} \quad (5.11)$$

Squaring both sides, applying  $(\sum_{r=1}^k u_r)^2 \leq k \sum_{r=1}^k u_r^2$ , and then taking an expectation we have

$$\begin{aligned} \mathbb{E}[(\langle Q_1 \rangle - \mathbb{E}\langle Q_1 \rangle)^2] &\leq 5\delta^{-2} \sum_{u \in \{h_0 - \delta, h_0 + \delta\}} \mathbb{E}[(F_n(u) - f_n(u))^2] \\ &\quad + 5C_\delta^+(h_0)^2 + 5C_\delta^-(h_0)^2. \end{aligned} \quad (5.12)$$

Under the assumption (5.2) about concentration of the free energy, the first term is smaller than  $15C_F/(n\delta^2)$ . Next, using  $|\frac{df_n}{dh_0}| = |\mathbb{E}\langle Q_1 \rangle| \leq 1$  allows to assert from (5.11) the crude bound  $C_\delta^\pm(h_0) \leq 2$ . Then using  $C_\delta^\pm(h_0) \geq 0$ ,

$$\begin{aligned} &\int_{\underline{h}}^{\bar{h}} dh_0 (C_\delta^+(h_0)^2 + C_\delta^-(h_0)^2) \\ &\leq 2 \int_{\underline{h}}^{\bar{h}} dh_0 (C_\delta^+(h_0) + C_\delta^-(h_0)) \\ &= 2[(f_n(\bar{h} - \delta) - f_n(\bar{h} + \delta)) + (f_n(\underline{h} + \delta) - f_n(\underline{h} - \delta))] \\ &\leq 8\delta, \end{aligned}$$

where the mean value theorem has been used to get the last inequality. When (5.12) is integrated over  $h_0$  we reach

$$\int_{\underline{h}}^{\bar{h}} dh_0 \mathbb{E}[(\langle Q_1 \rangle - \mathbb{E}\langle Q_1 \rangle)^2] \leq \frac{15C_F}{n\delta^2} + 40\delta.$$

The proof is ended by optimizing the bound by choosing  $\delta = n^{-1/3}$  for  $n$  large enough.  $\square$

### 5.3.4 Poisson fluctuations of overlaps

**Lemma 5.2** (Concentration of overlaps w.r.t. the Poisson perturbation). *Assume the Hamiltonian  $\mathcal{H}$  given by (5.5) is fully ferromagnetic so that both GKS inequalities hold. Then for any  $[\underline{h}, \bar{h}] \subset (0, 1)$  we have*

$$\int_{\underline{h}}^{\bar{h}} dh_0 \mathbb{E}_{\tau}[(\langle Q_p \rangle - \mathbb{E}_{\tau}\langle Q_p \rangle)^2] \leq \frac{\alpha p h_1}{n^{1-\theta}}.$$

*Proof.* Recall  $\tau$  is a random vector with i.i.d. components  $\tau_j \sim \text{Poi}(\alpha n^{\theta-1})$  for  $j \in \{1, \dots, n\}$ . Let  $\tau^j$  be the random vector that differs from  $\tau$  only at the  $j$ -th component, which is replaced by a new  $\tau'_j \sim \text{Poi}(\alpha n^{\theta-1})$  drawn independently from everything else. For this proof we explicitly keep track of the  $\tau$  dependence in Gibbs expectations  $\langle \cdot \rangle_{\tau}$ . The Efron–Stein inequality states ( $\mathbb{1}(\cdot)$  is the indicator function)

$$\begin{aligned} \mathbb{E}_{\tau}[(\langle Q_p \rangle_{\tau} - \mathbb{E}_{\tau}\langle Q_p \rangle_{\tau})^2] &\leq \frac{1}{2} \sum_{j=1}^n \mathbb{E}_{\tau \setminus \tau_j} \mathbb{E}_{\tau_j} \mathbb{E}_{\tau'_j}[(\langle Q_p \rangle_{\tau^j} - \langle Q_p \rangle_{\tau})^2] \\ &= \frac{1}{2} \sum_{j=1}^n \mathbb{E}_{\tau \setminus \tau_j} \mathbb{E}_{\tau_j} \mathbb{E}_{\tau'_j}[\mathbb{1}(\tau'_j > \tau_j)(\langle Q_p \rangle_{\tau^j} - \langle Q_p \rangle_{\tau})^2] \\ &\quad + \frac{1}{2} \sum_{j=1}^n \mathbb{E}_{\tau \setminus \tau_j} \mathbb{E}_{\tau_j} \mathbb{E}_{\tau'_j}[\mathbb{1}(\tau'_j = \tau_j)(\langle Q_p \rangle_{\tau^j} - \langle Q_p \rangle_{\tau})^2] \\ &\quad + \frac{1}{2} \sum_{j=1}^n \mathbb{E}_{\tau \setminus \tau_j} \mathbb{E}_{\tau_j} \mathbb{E}_{\tau'_j}[\mathbb{1}(\tau'_j < \tau_j)(\langle Q_p \rangle_{\tau^j} - \langle Q_p \rangle_{\tau})^2] \\ &= \sum_{j=1}^n \mathbb{E}_{\tau \setminus \tau_j} \mathbb{E}_{\tau_j} \mathbb{E}_{\tau'_j}[\mathbb{1}(\tau'_j > \tau_j)(\langle Q_p \rangle_{\tau^j} - \langle Q_p \rangle_{\tau})^2]. \end{aligned} \tag{5.13}$$

To get the second equality we used that the term with  $\tau'_j = \tau_j$  vanishes and that the terms with  $\tau'_j > \tau_j$  and  $\tau'_j < \tau_j$  are equal by symmetry (under exchange of  $\tau'_j$  and  $\tau_j$ ). The two GKS inequalities imply

$$\frac{d\langle Q_p \rangle_{\tau}}{d\tau_j} = \frac{ph_1}{n} \sum_{i=1}^n \langle \sigma_i \rangle_{\tau}^{p-1} (\langle \sigma_i \sigma_j \rangle_{\tau} - \langle \sigma_i \rangle_{\tau} \langle \sigma_j \rangle_{\tau}) \geq 0$$

and therefore  $\langle Q_p \rangle_{\tau^j} - \langle Q_p \rangle_{\tau} \geq 0$  when  $\tau'_j > \tau_j$  (here note that  $\tau_j$  is an integer but we formally consider it real when computing a derivative and then restrict the obtained monotonicity result to the integer case). This together with  $0 \leq \langle Q_p \rangle_{\tau} \leq 1$  (by GKS) implies  $(\langle Q_p \rangle_{\tau^j} - \langle Q_p \rangle_{\tau}) \in [0, 1]$ . Then (5.13) implies

$$\mathbb{E}_{\tau}[(\langle Q_p \rangle_{\tau} - \mathbb{E}_{\tau} \langle Q_p \rangle_{\tau})^2] \leq \sum_{j=1}^n \mathbb{E}_{\tau \setminus \tau_j} \mathbb{E}_{\tau_j} \mathbb{E}_{\tau'_j} [\mathbb{1}(\tau'_j > \tau_j) (\langle Q_p \rangle_{\tau^j} - \langle Q_p \rangle_{\tau})]. \quad (5.14)$$

Let  $\Delta_j \equiv \tau'_j - \tau_j$  and  $\mathbf{u}^j = (0, \dots, 0, 1, 0, \dots, 0)$  with  $u_j = 1$ . This allows us to rewrite  $\tau^j = \tau + \Delta_j \mathbf{u}^j$ . An interpolation gives

$$\begin{aligned} & \langle Q_p \rangle_{\tau + \Delta_j \mathbf{u}^j} - \langle Q_p \rangle_{\tau} \\ &= \int_0^1 ds \frac{d}{ds} \langle Q_p \rangle_{\tau + s \Delta_j \mathbf{u}^j} \\ &= \frac{ph_1 \Delta_j}{n} \int_0^1 ds \sum_{i=1}^n \langle \sigma_i \rangle_{\tau + s \Delta_j \mathbf{u}^j}^{p-1} (\langle \sigma_i \sigma_j \rangle_{\tau + s \Delta_j \mathbf{u}^j} - \langle \sigma_i \rangle_{\tau + s \Delta_j \mathbf{u}^j} \langle \sigma_j \rangle_{\tau + s \Delta_j \mathbf{u}^j}). \end{aligned} \quad (5.15)$$

Under the condition  $\tau'_j > \tau_j$ , the integrand in (5.15) is non-negative by the two GKS inequalities. Also note that, again by the two GKS inequalities,

$$\frac{d}{d\tau_j} (\langle \sigma_i \sigma_j \rangle_{\tau} - \langle \sigma_i \rangle_{\tau} \langle \sigma_j \rangle_{\tau}) = -2h_1 \langle \sigma_j \rangle_{\tau} (\langle \sigma_i \sigma_j \rangle_{\tau} - \langle \sigma_i \rangle_{\tau} \langle \sigma_j \rangle_{\tau}) \leq 0$$

so that, as  $\Delta_j > 0$ ,

$$\langle \sigma_i \sigma_j \rangle_{\tau + s \Delta_j \mathbf{u}^j} - \langle \sigma_i \rangle_{\tau + s \Delta_j \mathbf{u}^j} \langle \sigma_j \rangle_{\tau + s \Delta_j \mathbf{u}^j} \leq \langle \sigma_i \sigma_j \rangle_{\tau} - \langle \sigma_i \rangle_{\tau} \langle \sigma_j \rangle_{\tau}. \quad (5.16)$$

Therefore, substituting (5.15), (5.16) into (5.14) and simply upper bounding  $\langle \sigma_i \rangle_{\tau + s \Delta_j \mathbf{u}^j}^{k-1}$  by 1, we obtain

$$\begin{aligned} & \mathbb{E}_{\tau}[(\langle Q_p \rangle_{\tau} - \mathbb{E}_{\tau} \langle Q_p \rangle_{\tau})^2] \\ & \leq \frac{ph_1}{n} \sum_{i,j=1}^n \mathbb{E}_{\tau} \mathbb{E}_{\tau'_j} [\mathbb{1}(\tau'_j > \tau_j) (\tau'_j - \tau_j) (\langle \sigma_i \sigma_j \rangle_{\tau} - \langle \sigma_i \rangle_{\tau} \langle \sigma_j \rangle_{\tau})]. \end{aligned} \quad (5.17)$$

For given  $\tau_j$  fixed, the part containing  $\tau'_j$  has an upper bound independent of  $j$ :

$$\mathbb{E}_{\tau'_j} [\mathbb{1}(\tau'_j > \tau_j) (\tau'_j - \tau_j)] \leq \mathbb{E}_{\tau'_j} [\tau'_j] = \frac{\alpha}{n^{1-\theta}}$$

because  $\tau_j \geq 0$  and  $\tau'_j \sim \text{Poi}(\alpha n^{\theta-1})$ . This further relaxes (5.17) to

$$\begin{aligned} & \mathbb{E}_{\tau}[(\langle Q_p \rangle_{\tau} - \mathbb{E}_{\tau} \langle Q_p \rangle_{\tau})^2] \\ & \leq \frac{\alpha ph_1}{n^{2-\theta}} \sum_{i,j=1}^n \mathbb{E}_{\tau} [\langle \sigma_i \sigma_j \rangle_{\tau} - \langle \sigma_i \rangle_{\tau} \langle \sigma_j \rangle_{\tau}] = \alpha ph_1 n^{\theta} \mathbb{E}_{\tau} \langle (Q_1 - \langle Q_1 \rangle_{\tau})^2 \rangle_{\tau} \end{aligned} \quad (5.18)$$

(recall (5.9) for the last equality). Finally, integrating (5.18) over  $h_0 \in [\underline{h}, \bar{h}]$  and using Theorem 5.1 with  $p = 1$  (the factor 2 can be removed here because we assume both GKS inequalities) ends the proof.  $\square$

### 5.3.5 Last type of fluctuations of overlaps

In this section we tackle the last kind of fluctuations in the decomposition (5.7). Before proceeding let us say a few words about the strategy. The proof is decomposed in three steps (where the first two follow the ideas used in proving the Ghirlanda-Guerra identities for spin glasses [79]). The first step shows that the  $\alpha$ -derivative of the free energy concentrates which will lead to Lemma 5.3 (recall  $\alpha$  controls the mean of the Poisson quenched variables  $\tau_i$ ). In the second step we derive an identity which links a “generating series” containing overlap covariances to the product of an overlap and the free energy  $\alpha$ -derivative fluctuations. Using the concentration result of step one we can show that this generating series concentrates, leading to Lemma 5.4. In the third step, from the concentration of this generating series we extract the concentration of *each* overlap covariance, leading to Lemma 5.6. In particular, this will imply the control of the third kind of fluctuations in (5.7). The third step is non-trivial as the generating series has alternating signs. Nevertheless, we overcome this problem using an induction argument over  $p$  (the order of the overlap) thanks to the GKS inequalities and to Lemma 5.1 for the base case  $p = 1$ .

#### Step 1: Concentration of the free energy $\alpha$ -derivative

Let  $\hat{F}_n(\alpha) \equiv \mathbb{E}_{\tau} F_n$ . Note that the free energy  $f_n(\alpha) = \mathbb{E} \hat{F}_n$  is obtained by taking an expectation over the rest of the quenched variables. We start with a few preliminaries about these functions. We emphasize the  $\alpha$  dependence in this section. As before let  $\mathbf{u}^j = (0, \dots, 0, 1, 0, \dots, 0)$  with  $u_j = 1$ . Recall  $\tau_i \sim \text{Poi}(\alpha n^{\theta-1})$ . Then a straightforward algebra using the Poisson property (4.46) yields the following identities:

$$\begin{aligned} \frac{d\hat{F}_n(\alpha)}{d\alpha} &= -\frac{1}{n^{2-\theta}} \sum_{i=1}^n \mathbb{E}_{\tau} \ln \langle e^{h_1 \sigma_i} \rangle_{\tau} \\ &= -\frac{1}{n^{2-\theta}} \sum_{i=1}^n \mathbb{E}_{\tau} \ln(1 + \langle \sigma_i \rangle_{\tau} \tanh h_1) - \frac{1}{n^{1-\theta}} \ln \cosh h_1, \end{aligned} \quad (5.19)$$

$$\begin{aligned} \frac{d^2 \hat{F}_n(\alpha)}{d\alpha^2} &= -\frac{1}{n^{3-2\theta}} \sum_{i,j=1}^n \left( \mathbb{E}_{\tau} \ln \langle e^{h_1 \sigma_i} \rangle_{\tau + \mathbf{u}^j} - \mathbb{E}_{\tau} \ln \langle e^{h_1 \sigma_i} \rangle_{\tau} \right) \\ &= -\frac{1}{n^{3-2\theta}} \sum_{i,j=1}^n \left( \mathbb{E}_{\tau} \ln(1 + \langle \sigma_i \rangle_{\tau + \mathbf{u}^j} \tanh h_1) \right. \\ &\quad \left. - \mathbb{E}_{\tau} \ln(1 + \langle \sigma_i \rangle_{\tau} \tanh h_1) \right), \end{aligned} \quad (5.20)$$

where we used  $e^{\sigma x} = \cosh x(1 + \sigma \tanh x)$  for  $\sigma = \pm 1$ . The derivatives for  $f_n(\alpha)$  can directly be obtained by taking an expectation over the rest of the quenched variables:

$$\frac{df_n(\alpha)}{d\alpha} = -\frac{1}{n^{2-\theta}} \sum_{i=1}^n \mathbb{E} \ln(1 + \langle \sigma_i \rangle_{\tau} \tanh h_1) - \frac{1}{n^{1-\theta}} \ln \cosh h_1, \quad (5.21)$$

$$\begin{aligned} \frac{d^2 f_n(\alpha)}{d\alpha^2} = & -\frac{1}{n^{3-2\theta}} \sum_{i,j=1}^n \left( \mathbb{E} \ln(1 + \langle \sigma_i \rangle_{\tau+\mathbf{u}^j} \tanh h_1) \right. \\ & \left. - \mathbb{E} \ln(1 + \langle \sigma_i \rangle_{\tau} \tanh h_1) \right). \end{aligned} \quad (5.22)$$

The second GKS inequality (1.10) implies that  $d\langle \sigma_i \rangle_{\tau}/d\tau_j = h_1(\langle \sigma_i \sigma_j \rangle_{\tau} - \langle \sigma_i \rangle_{\tau} \langle \sigma_j \rangle_{\tau})$  is non-negative, and therefore  $\langle \sigma_i \rangle_{\tau} \leq \langle \sigma_i \rangle_{\tau+\mathbf{u}^j}$ . Thus the identities (5.20) and (5.22) imply (using also  $1 + \langle \sigma_i \rangle \tanh h_1 \geq 0$ ) that  $d^2 \hat{F}_n(\alpha)/d\alpha^2 \leq 0$  and  $d^2 f_n(\alpha)/d\alpha^2 \leq 0$ , which means that  $\hat{F}_n(\alpha)$  and  $f_n(\alpha)$  are concave in  $\alpha$  (note that in order to obtain this concavity we used only the second GKS inequality, without the first one here). One can also see that

$$\left| \frac{df_n(\alpha)}{d\alpha} \right| = \left| \frac{1}{n^{2-\theta}} \sum_{i=1}^n \mathbb{E} \ln \langle e^{h_1 \sigma_i} \rangle_{\tau} \right| \leq \frac{h_1}{n^{1-\theta}}, \quad (5.23)$$

noting that  $|\ln \langle e^{h_1 \sigma_i} \rangle_{\tau}| \leq h_1$  because  $h_1 > 0$  and  $\sigma_i \in \{-1, +1\}$ .

We can now show a concentration result for the  $\alpha$ -derivative of the free energy based on Lemma 3.12.

**Lemma 5.3** (Concentration of the free energy  $\alpha$ -derivative). *Assume the Hamiltonian  $\mathcal{H}$  given by (5.5) satisfies  $\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \geq 0$  for all  $i, j = 1, \dots, n$  and assume also that Condition 5.1 holds. Then for any  $[\underline{\alpha}, \bar{\alpha}] \subset (0, 1)$ ,  $h_0 \in (0, 1)$ ,  $h_1 \in (0, 1]$  and  $\theta \in (1/2, 1)$  we have*

$$\int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \mathbb{E} \left[ \left( \frac{d\hat{P}_n(\alpha)}{d\alpha} - \frac{df_n(\alpha)}{d\alpha} \right)^2 \right] \leq \frac{15C_F + 40h_1^2}{n^{(5-4\theta)/3}}.$$

*Proof.* The proof is similar to the one of Lemma 5.1. Using Lemma 3.12, the concavity of  $\hat{F}_n$  and  $f_n$  in  $\alpha$  implies that for any  $\delta > 0$  we have

$$\begin{aligned} \left| \frac{d\hat{F}_n(\alpha)}{d\alpha} - \frac{df_n(\alpha)}{d\alpha} \right| & \leq \delta^{-1} \sum_{u \in \{\alpha-\delta, \alpha, \alpha+\delta\}} |\hat{F}_n(u) - f_n(u)| \\ & \quad + C_{\delta}^{+}(\alpha) + C_{\delta}^{-}(\alpha) \end{aligned}$$

(note we can take  $\delta$  small enough so that  $\alpha - \delta > 0$ ) where

$$C_{\delta}^{+}(\alpha) \equiv \frac{df_n(\alpha)}{d\alpha} - \frac{df_n(\alpha + \delta)}{d\alpha} \geq 0, \quad C_{\delta}^{-}(\alpha) \equiv \frac{df_n(\alpha - \delta)}{d\alpha} - \frac{df_n(\alpha)}{d\alpha} \geq 0.$$

Squaring both sides, applying  $(\sum_{r=1}^k u_r)^2 \leq k \sum_{r=1}^k u_r^2$  and averaging we get

$$\mathbb{E} \left[ \left( \frac{d\hat{F}_n(\alpha)}{d\alpha} - \frac{df_n(\alpha)}{d\alpha} \right)^2 \right] \leq 5\delta^{-2} \sum_{u \in \{\alpha-\delta, \alpha, \alpha+\delta\}} \mathbb{E}[(\hat{F}_n(u) - f_n(u))^2] + 5C_\delta^+(\alpha)^2 + 5C_\delta^-(\alpha)^2. \quad (5.24)$$

It is easy to check that

$$\begin{aligned} \mathbb{E}[(\hat{F}_n(\alpha) - f_n(\alpha))^2] &= \mathbb{E}[(F_n(\alpha) - f_n(\alpha))^2] - \mathbb{E}[(F_n(\alpha) - \hat{F}_n(\alpha))^2] \\ &\leq \mathbb{E}[(F_n(\alpha) - f_n(\alpha))^2]. \end{aligned}$$

Thus under the concentration assumption (5.2) for the free energy, the first term in the r.h.s. of (5.24) is smaller than  $15C_F/(n\delta^2)$ . Next, we recall (5.23) which implies the crude bound  $C_\delta^\pm(h_0) \leq 2h_1/n^{1-\theta}$ , so using  $C_\delta^\pm(h_0) \geq 0$ ,

$$\begin{aligned} &\int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha (C_\delta^+(\alpha)^2 + C_\delta^-(\alpha)^2) \\ &\leq \frac{2h_1}{n^{1-\theta}} \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha (C_\delta^+(\alpha) + C_\delta^-(\alpha)) \\ &= \frac{2h_1}{n^{1-\theta}} [(f_n(\bar{\alpha} - \delta) - f_n(\bar{\alpha} + \delta)) + (f_n(\underline{\alpha} + \delta) - f_n(\underline{\alpha} - \delta))] \\ &\leq \frac{8\delta h_1^2}{n^{2-2\theta}} \end{aligned}$$

where we used the mean value theorem for the last inequality. Thus when (5.24) is integrated over  $\alpha$  we obtain

$$\int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \mathbb{E} \left[ \left( \frac{d\hat{F}_n(\alpha)}{d\alpha} - \frac{df_n(\alpha)}{d\alpha} \right)^2 \right] \leq \frac{15C_F}{n\delta^2} + \frac{40\delta h_1^2}{n^{2-2\theta}}. \quad (5.25)$$

The proof is ended by choosing  $\delta$  such that  $n^{-1}\delta^{-2} = \delta n^{-2+2\theta}$ , in other words  $\delta = n^{(1-2\theta)/3}$ , which is possible for  $\theta > 1/2$  (because we must have  $\delta$  small enough in (5.25)). With this choice the upper bound in (5.25) becomes  $(15C_F + 40h_1^2)/n^{(5-4\theta)/3}$ . Note that  $5 - 4\theta > 0$  because  $\theta < 1$  anyway.  $\square$

## Step 2: Linking the fluctuations of the free energy $\alpha$ -derivative to a series of overlap covariances

In this step  $P \geq 1$  is an integer fixed throughout. Define the set of multi-overlap covariances (w.r.t. the quenched variables except the Poisson ones  $\tau$ ) as

$$\text{Cov}_{P,p} \equiv \mathbb{E}[\mathbb{E}_\tau \langle Q_P \rangle \mathbb{E}_\tau \langle Q_p \rangle] - \mathbb{E} \langle Q_P \rangle \mathbb{E} \langle Q_p \rangle, \quad k \geq 1. \quad (5.26)$$

The task is to bound the variance of  $\mathbb{E}_\tau \langle Q_p \rangle$  using Lemma 5.3. However, here is a case where constructing a bound for the covariances is more flexible

and feasible. Roughly speaking, we will show in this step that a *generating series* for the set  $\{\text{Cov}_{P,p}, p \geq 1\}$  is small. From this knowledge, and despite this series has alternating signs, we will in step 3 deduce that all individual covariances  $\text{Cov}_{P,p}$  are also small. In particular, this will hold for the variance term  $p = P$ .

**Lemma 5.4** (Concentration of a generating series). *Assume the Hamiltonian  $\mathcal{H}$  given by (5.5) satisfies  $\langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \geq 0$  for all  $i, j = 1, \dots, n$  and assume also that Condition 5.1 holds. Then for any  $[\underline{\alpha}, \bar{\alpha}] \subset (0, 1)$ ,  $h_0 \in (0, 1)$ ,  $h_1 \in (0, 1]$ ,  $\theta \in (1/2, 1)$  and any fixed integer  $P \geq 1$  we have*

$$\int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \sum_{k=1}^{\infty} \frac{(-1)^{p+1}}{p} (\tanh h_1)^p \text{Cov}_{P,p} \right| \leq \frac{\sqrt{15C_F + 40h_1^2}}{n^{(\theta-1/2)/3}}. \quad (5.27)$$

*Proof.* By the Cauchy-Schwarz inequality and Lemma 5.3 we have

$$\begin{aligned} & \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \mathbb{E} \left[ \mathbb{E}_{\tau} \langle Q_P \rangle \left( \frac{d\hat{F}_n(\alpha)}{d\alpha} - \frac{df_n(\alpha)}{d\alpha} \right) \right] \right| \\ & \leq \left\{ \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \mathbb{E} \left[ (\mathbb{E}_{\tau} \langle Q_P \rangle)^2 \right] \right\}^{1/2} \left\{ \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \mathbb{E} \left[ \left( \frac{d\hat{F}_n(\alpha)}{d\alpha} - \frac{df_n(\alpha)}{d\alpha} \right)^2 \right] \right\}^{1/2} \\ & \leq \frac{\sqrt{15C_F + 40h_1^2}}{n^{(5-4\theta)/6}}. \end{aligned} \quad (5.28)$$

The next step is to expand the  $\alpha$ -derivatives of the free energy. For that we recall the formulas (5.19) and (5.21). Taylor expanding the logarithms in (5.19) and recalling  $n^{-1} \sum_{i=1}^n \langle \sigma_i \rangle^p = \langle Q_p \rangle$  gives

$$\frac{d\hat{F}_n(\alpha)}{d\alpha} = -\frac{1}{n^{1-\theta}} \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p} (\tanh h)^p \mathbb{E}_{\tau} \langle Q_p \rangle - \frac{1}{n^{1-\theta}} \ln \cosh h_1.$$

The series expansion of  $\frac{df_n}{d\alpha}$  is obtained similarly based on (5.21), and is thus the same with  $\mathbb{E}_{\tau}$  replaced by the full expectation  $\mathbb{E}$ . Substituting these expansions in the left-most hand side of (5.28) yields

$$\begin{aligned} & \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \sum_{k=1}^{\infty} \frac{(-1)^{p+1}}{p} (\tanh h_1)^p \{ \mathbb{E}[\mathbb{E}_{\tau} \langle Q_P \rangle \mathbb{E}_{\tau} \langle Q_p \rangle] - \mathbb{E} \langle Q_P \rangle \mathbb{E} \langle Q_p \rangle \} \right| \\ & \leq \frac{\sqrt{15C_F + 40h_1^2}}{n^{(5-4\theta)/6}} n^{1-\theta}. \end{aligned}$$

Recognizing (5.26) then ends the proof.  $\square$

**Step 3: An induction argument over the overlap covariances**

We start with a useful monotonicity lemma that will allow us to control the alternating signs of the generating series in Lemma 5.4.

**Lemma 5.5** (A property on monotonicity). *Assume the Hamiltonian  $\mathcal{H}$  given by (5.5) is fully ferromagnetic and thus satisfies both GKS inequalities. Then we have*

$$\frac{1}{p} \text{Cov}_{P,p} - \frac{\tanh h_1}{p+1} \text{Cov}_{P,p+1} \geq 0.$$

*Proof.* Let  $\mathbf{J} \equiv (J_A, A \subset \{1, \dots, n\})$ . Define  $g_P(\mathbf{J}) \equiv \mathbb{E}_{\tau} \langle Q_P \rangle$  and  $\tilde{g}_p(\mathbf{J}) \equiv \frac{1}{p} \mathbb{E}_{\tau} \langle Q_p \rangle - \frac{\tanh h_1}{p+1} \mathbb{E}_{\tau} \langle Q_{p+1} \rangle$ . One can then recognize

$$\frac{1}{k} \text{Cov}_{P,p} - \frac{\tanh h_1}{p+1} \text{Cov}_{P,p+1} = \mathbb{E}[g_P(\mathbf{J}) \tilde{g}_p(\mathbf{J})] - \mathbb{E} g_P(\mathbf{J}) \mathbb{E} \tilde{g}_p(\mathbf{J}) \quad (5.29)$$

so it is enough to verify that  $g_P(\mathbf{J})$  and  $\tilde{g}_p(\mathbf{J})$  are positively correlated. Note that the expectations in (5.29) only carry over the set of i.i.d. random coupling constants  $J_A, A \subset \{1, \dots, n\}$ . By the GKS inequalities the following derivatives are non-negative:

$$\begin{aligned} \frac{d}{dJ_A} g_P(\mathbf{J}) &= \frac{P}{n} \sum_{i=1}^n \mathbb{E}_{\tau} [\langle \sigma_i \rangle^{P-1} (\langle \sigma_i \sigma_A \rangle - \langle \sigma_i \rangle \langle \sigma_A \rangle)] \geq 0, \\ \frac{d}{dJ_A} \tilde{g}_p(\mathbf{J}) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\tau} [\langle \sigma_i \rangle^{p-1} (1 - \langle \sigma_i \rangle \tanh h_1) (\langle \sigma_i \sigma_A \rangle - \langle \sigma_i \rangle \langle \sigma_A \rangle)] \geq 0. \end{aligned}$$

In other words,  $g_P(\mathbf{J})$  and  $\tilde{g}_p(\mathbf{J})$  have same monotonicity w.r.t. each  $J_A$  for all  $A \subset \{1, \dots, n\}$ . We can then apply the Harris inequality (reproduced in Lemma 5.7, Appendix 5.4.2) to finish the proof.  $\square$

Now we have all the necessary ingredients in order to inductively extract the concentration of each individual overlap from Lemma 5.4.

**Lemma 5.6** (Concentration of the overlaps w.r.t. the quenched variables). *Assume the Hamiltonian  $\mathcal{H}$  given by (5.5) is fully ferromagnetic so that it satisfies both GKS inequalities, and also that Condition 5.1 holds. Then for any  $[\underline{h}, \bar{h}] \subset (0, 1)$ ,  $[\underline{\alpha}, \bar{\alpha}] \subset (0, 1)$ ,  $\theta \in (1/2, 1)$ ,  $h_1 \in (0, 1]$  and any  $p, P \geq 1$  we have*

$$\int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,p}| \leq \frac{pM_p}{(\tanh h_1)^p} \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3}},$$

where  $M_p$  is defined by  $M_1 = 1$ ,  $M_{2p} = M_{2p-1} + 1$ ,  $M_{2p+1} = M_{2p} + 2$  (so  $M_p < 3p/2$ ). In particular for  $p = P$ ,

$$\int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \mathbb{E}[(\mathbb{E}_{\tau} \langle Q_p \rangle - \mathbb{E} \langle Q_p \rangle)^2] \leq \frac{3p^2}{2(\tanh h_1)^p} \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3}}.$$



*Proof.* We start the induction with the base case  $k = 1$ . From (5.26) we note that

$$\begin{aligned} \text{Cov}_{P,1} &= \mathbb{E}[\mathbb{E}_{\tau}\langle Q_P \rangle \mathbb{E}_{\tau}\langle Q_1 \rangle] - \mathbb{E}\langle Q_P \rangle \mathbb{E}\langle Q_1 \rangle \\ &= \mathbb{E}[\mathbb{E}_{\tau}\langle Q_P \rangle \langle Q_1 \rangle] - \mathbb{E}\langle Q_P \rangle \mathbb{E}\langle Q_1 \rangle = \mathbb{E}[\mathbb{E}_{\tau}\langle Q_P \rangle (\langle Q_1 \rangle - \mathbb{E}\langle Q_1 \rangle)]. \end{aligned}$$

Then, using successively Fubini's theorem, the Cauchy-Schwarz inequality and Lemma 5.1, we have

$$\begin{aligned} \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,1}| &= \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \int_{\underline{h}}^{\bar{h}} dh_0 |\mathbb{E}[\mathbb{E}_{\tau}\langle Q_P \rangle (\langle Q_1 \rangle - \mathbb{E}\langle Q_1 \rangle)]| \\ &\leq \left\{ \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \int_{\underline{h}}^{\bar{h}} dh_0 \mathbb{E}[(\mathbb{E}_{\tau}\langle Q_P \rangle)^2] \right\}^{1/2} \left\{ \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \int_{\underline{h}}^{\bar{h}} dh_0 \mathbb{E}[(\langle Q_1 \rangle - \mathbb{E}\langle Q_1 \rangle)^2] \right\}^{1/2} \\ &\leq \frac{\sqrt{15C_F + 40}}{n^{1/6}} \leq \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3} \tanh h_1}. \end{aligned} \quad (5.30)$$

Note that the last inequality is valid because  $0 < \tanh h_1 \leq 1$  and  $\theta < 1$ . For  $p \geq 2$  we adopt an induction in two steps: From  $2p - 1$  to  $2p$  and then from  $2p$  to  $2p + 1$ .

We start with the induction step from  $2p - 1$  to  $2p$ . Suppose

$$\int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,2p-1}| \leq \frac{(2p-1)M_{2p-1}}{(\tanh h_1)^{2p-1}} \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3}}. \quad (5.31)$$

The left hand side of (5.27) is

$$\begin{aligned} \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \sum_{p'=1}^{\infty} \frac{(-1)^{p'+1}}{p'} (\tanh h_1)^{p'} \text{Cov}_{P,p'} \right| \\ &= \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \sum_{p'=1}^{\infty} (\tanh h_1)^{2p'-1} \left( \frac{1}{2p'-1} \text{Cov}_{P,2p'-1} - \frac{\tanh h_1}{2p'} \text{Cov}_{P,2p'} \right) \right| \\ &= \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \sum_{p'=1}^{\infty} (\tanh h_1)^{2p'-1} \left| \frac{1}{2p'-1} \text{Cov}_{P,2p'-1} - \frac{\tanh h_1}{2p'} \text{Cov}_{P,2p'} \right| \end{aligned} \quad (5.32)$$

$$\geq \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \frac{(\tanh h_1)^{2p-1}}{2p-1} \text{Cov}_{P,2p-1} - \frac{(\tanh h_1)^{2p}}{2p} \text{Cov}_{P,2p} \right| \quad (5.33)$$

where (5.32) follows from  $h_1 \geq 0$  and Lemma 5.5. By the triangle inequality

we have

$$\begin{aligned}
& \frac{(\tanh h_1)^{2p}}{2p} \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,2p}| \\
&= \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \left( \frac{(\tanh h_1)^{2p-1}}{2p-1} \text{Cov}_{P,2p-1} - \frac{(\tanh h_1)^{2p}}{2p} \text{Cov}_{P,2p} \right) \right. \\
&\quad \left. - \frac{(\tanh h_1)^{2p-1}}{2p-1} \text{Cov}_{P,2p-1} \right| \\
&\leq \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \frac{(\tanh h_1)^{2p-1}}{2p-1} \text{Cov}_{P,2p-1} - \frac{(\tanh h_1)^{2p}}{2p} \text{Cov}_{P,2p} \right| \\
&\quad + \frac{(\tanh h_1)^{2p-1}}{2p-1} \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,2p-1}| \\
&\leq \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \sum_{p'=1}^{\infty} \frac{(-1)^{p'+1}}{p'} (\tanh h_1)^{p'} \text{Cov}_{P,p'} \right| \\
&\quad + \frac{(\tanh h_1)^{2p-1}}{2p-1} \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,2p-1}| \tag{5.34}
\end{aligned}$$

$$\leq \frac{\sqrt{15C_F + 40h_1^2}}{n^{(\theta-1/2)/3}} + M_{2p-1} \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3}} \tag{5.35}$$

$$\leq (M_{2p-1} + 1) \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3}} \tag{5.36}$$

where (5.34) follows from (5.33), then (5.35) follows from Lemma 5.4 and the hypothesis (5.31), and finally (5.36) uses  $h_1 \in (0, 1]$ . Summarizing, we have shown

$$\int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,2p}| \leq \frac{2pM_{2p}}{(\tanh h_1)^{2p}} \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3}} \tag{5.37}$$

with  $M_{2p} = M_{2p-1} + 1$ .

Now we proceed similarly for the induction from  $2p$  to  $2p+1$ . This time we start with

$$\begin{aligned}
& \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \sum_{p'=2}^{\infty} \frac{(-1)^{p'+1}}{p'} (\tanh h_1)^{p'} \text{Cov}_{P,p'} \right| \\
&= \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \sum_{p'=1}^{\infty} (\tanh h_1)^{2p'} \left( \frac{1}{2p'} \text{Cov}_{P,2p'} - \frac{\tanh h_1}{2p'+1} \text{Cov}_{P,2p'+1} \right) \right| \\
&= \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \sum_{p'=1}^{\infty} (\tanh h_1)^{2p'} \left| \frac{1}{2p'} \text{Cov}_{P,2p'} - \frac{\tanh h_1}{2p'+1} \text{Cov}_{P,2p'+1} \right| \tag{5.38}
\end{aligned}$$

$$\geq \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \frac{(\tanh h_1)^{2p}}{2p} \text{Cov}_{P,2p} - \frac{(\tanh h_1)^{2p+1}}{2p+1} \text{Cov}_{P,2p+1} \right| \tag{5.39}$$

where (5.38) follows from Lemma 5.5 and  $h_1 \geq 0$ . Also we have

$$\begin{aligned}
& \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \sum_{p'=2}^{\infty} \frac{(-1)^{p'+1}}{p'} (\tanh h_1)^{p'} \text{Cov}_{P,p'} \right| \\
&= \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \sum_{p'=1}^{\infty} \frac{(-1)^{p'+1}}{p'} (\tanh h_1)^{p'} \text{Cov}_{P,p'} - \tanh h_1 \text{Cov}_{P,1} \right| \\
&\leq \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \sum_{p'=1}^{\infty} \frac{(-1)^{p'+1}}{p'} (\tanh h_1)^{p'} \text{Cov}_{P,p'} \right| + \tanh h_1 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,1}|. \quad (5.40)
\end{aligned}$$

Then we proceed as

$$\begin{aligned}
& \frac{(\tanh h_1)^{2p+1}}{2p+1} \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,2p+1}| \\
&= \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \left( \frac{(\tanh h_1)^{2p}}{2p} \text{Cov}_{P,2p} - \frac{(\tanh h_1)^{2p+1}}{2p+1} \text{Cov}_{P,2p+1} \right) \right. \\
&\quad \left. - \frac{(\tanh h_1)^{2p}}{2p} \text{Cov}_{P,2p} \right| \\
&\leq \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \frac{(\tanh h_1)^{2p}}{2p} \text{Cov}_{P,2p} - \frac{(\tanh h_1)^{2p+1}}{2p+1} \text{Cov}_{P,2p+1} \right| \\
&\quad + \frac{(\tanh h_1)^{2p}}{2p} \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,2p}| \\
&\leq \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \sum_{p'=2}^{\infty} \frac{(-1)^{p'+1}}{p'} (\tanh h_1)^{p'} \text{Cov}_{P,p'} \right| \\
&\quad + \frac{(\tanh h_1)^{2p}}{2p} \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,2p}| \quad (5.41)
\end{aligned}$$

$$\begin{aligned}
&\leq \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \left| \sum_{p'=1}^{\infty} \frac{(-1)^{p'+1}}{p'} (\tanh h_1)^{p'} \text{Cov}_{P,p'} \right| \\
&\quad + \tanh h_1 \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,1}| \\
&\quad + \frac{(\tanh h_1)^{2p}}{2p} \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,2p}| \quad (5.42)
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{\sqrt{15C_F + 40h_1^2}}{n^{(\theta-1/2)/3}} + \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3}} + M_{2p} \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3}} \\
&\leq (M_{2p} + 2) \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3}} \quad (5.43)
\end{aligned}$$

where (5.41) follows from (5.39), then (5.42) follows from (5.40), and finally

(5.43) follows from Lemma 5.4, (5.30) and (5.37). Summarizing,

$$\int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha |\text{Cov}_{P,2p+1}| \leq \frac{(2p+1)M_{2p+1} \sqrt{15C_F + 40}}{(\tanh h_1)^{2p+1} n^{(\theta-1/2)/3}}$$

with  $M_{2p+1} = M_{2p} + 2$ , which ends the induction argument.  $\square$

### 5.3.6 Proof of Theorem 5.3

We finally show how to combine all the concentration results we obtained in order to prove the following theorem. This theorem is a mild variant of Theorem 5.3. Inequality (5.44) below is exactly Theorem 5.3.

**Theorem 5.4** (Overlap concentration). *Assume the Hamiltonian  $\mathcal{H}$  given by (5.5) satisfies both GKS inequalities and also that Condition 5.1 holds. Then for any moment  $k \geq 2$ ,  $[\underline{h}, \bar{h}] \subset (0, 1)$ ,  $[\underline{\alpha}, \bar{\alpha}] \subset (0, 1)$ ,  $h_1 \in (0, 1]$ ,  $\theta \in (1/2, 7/8]$ ,*

$$\int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \mathbb{E} \langle |Q_p^k - \mathbb{E}[\langle Q_p \rangle]^k| \rangle \leq \frac{2pk}{(\tanh h_1)^{p/2}} \frac{(15C_F + 40)^{1/4}}{n^{(\theta-1/2)/6}}.$$

*Proof.* We integrate both sides of (5.7) over  $h_0$  and  $\alpha$ . As all the square terms are bounded by 1, by Fubini's theorem we are free to exchange the order of the integrals. Theorem 5.1, and Lemmas 5.2 and 5.6 are applied accordingly and lead to the estimate (for any  $p \geq 1$ )

$$\begin{aligned} & \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \mathbb{E} \langle (Q_p - \mathbb{E} \langle Q_p \rangle)^2 \rangle \\ & \leq \frac{2p}{n} + \frac{(\bar{\alpha}^2 - \underline{\alpha}^2)ph_1}{2n^{1-\theta}} + \frac{pM_p}{(\tanh h_1)^p} \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3}} \\ & \leq \frac{4p^2}{(\tanh h_1)^p} \frac{\sqrt{15C_F + 40}}{n^{(\theta-1/2)/3}} \end{aligned} \quad (5.44)$$

using  $\theta \in (1/2, 7/8]$  (the  $7/8$  is enforced by  $(\theta - 1/2)/3 \leq 1 - \theta$ ),  $[\underline{\alpha}, \bar{\alpha}] \subset (0, 1)$ ,  $h_1 \in (0, 1]$  and  $M_p < 3p/2$ . Finally, observe that

$$\begin{aligned} \mathbb{E} \langle |Q_p^k - \mathbb{E}[\langle Q_p \rangle]^k| \rangle &= \mathbb{E} \left\langle \left| (Q_p - \mathbb{E} \langle Q_p \rangle) \sum_{l=0}^{k-1} Q_p^{k-1-l} \mathbb{E}[\langle Q_p \rangle]^l \right| \right\rangle \\ &\leq k \mathbb{E} \langle |Q_p - \mathbb{E} \langle Q_p \rangle| \rangle. \end{aligned}$$

By the Cauchy-Schwarz inequality we then have

$$\begin{aligned} & \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \mathbb{E} \langle |Q_p^k - \mathbb{E}[\langle Q_p \rangle]^k| \rangle \\ & \leq k \left\{ \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \right\}^{1/2} \left\{ \int_{\underline{h}}^{\bar{h}} dh_0 \int_{\underline{\alpha}}^{\bar{\alpha}} d\alpha \mathbb{E} \langle (Q_p - \mathbb{E} \langle Q_p \rangle)^2 \rangle \right\}^{1/2} \end{aligned}$$

which ends the proof once combined with (5.44).  $\square$

## 5.4 Appendix

### 5.4.1 Proof of the approximation inequality (5.6)

Note that

$$\begin{aligned} & |f_n(h_0, h_1, \alpha) - f_n(0, 0, 0)| \\ &= |f_n(h_0, h_1, \alpha) - f_n(0, h_1, 0)| \\ &\leq |f_n(h_0, h_1, \alpha) - f_n(0, h_1, \alpha)| + |f_n(0, h_1, \alpha) - f_n(0, h_1, 0)|. \end{aligned}$$

We have  $|\frac{df_n(h_0, h_1, \alpha)}{d\alpha}| = |\mathbb{E}\langle Q_1 \rangle| \leq 1$  and from (5.23) we also have  $|\frac{df_n(0, h_1, \alpha)}{d\alpha}| \leq h_1 n^{-(1-\theta)}$ . Thus by the mean value theorem we obtain (5.6), i.e.  $|f_n(h_0, h_1, \alpha) - f_n(0, 0, 0)| \leq h_0 + \alpha h_1 n^{-(1-\theta)}$ .

### 5.4.2 Multivariate Harris inequality

For completeness we provide here a simple proof of the multivariate version of the Harris inequality. We refer to [91] for more information.

**Lemma 5.7** (Multivariate version of the Harris inequality). *Let  $g, \tilde{g} : \mathbb{R}^n \mapsto \mathbb{R}$  be two functions of the random vector  $\mathbf{x} = (x_1, \dots, x_n)$  where all components are independent random variables. If for all  $i \in \{1, \dots, n\}$   $g$  and  $\tilde{g}$  are both monotone w.r.t.  $x_i$  with same monotonicity, i.e.  $\partial_{x_i} g(\mathbf{x}) \partial_{x_i} \tilde{g}(\mathbf{x}) \geq 0 \forall i$ , then  $\mathbb{E}[g(\mathbf{x}) \tilde{g}(\mathbf{x})] - \mathbb{E} g(\mathbf{x}) \mathbb{E} \tilde{g}(\mathbf{x}) \geq 0$ .*

*Proof.* Let  $\mathbf{x}_i^j \equiv (x_i, x_{i+1}, \dots, x_j)$ . The monotonicity w.r.t.  $x_1$  implies

$$\mathbb{E}_{x_1} \mathbb{E}_{x_1'} [(g(x_1, \mathbf{x}_2^n) - g(x_1', \mathbf{x}_2^n)) (\tilde{g}(x_1, \mathbf{x}_2^n) - \tilde{g}(x_1', \mathbf{x}_2^n))] \geq 0$$

which by expanding the product can be simplified to

$$\mathbb{E}_{x_1} [g(\mathbf{x}) \tilde{g}(\mathbf{x})] - \mathbb{E}_{x_1} g(\mathbf{x}) \mathbb{E}_{x_1} \tilde{g}(\mathbf{x}) \geq 0.$$

The proof then proceeds by induction. Suppose

$$\mathbb{E}_{\mathbf{x}_1^{i-1}} [g(\mathbf{x}) \tilde{g}(\mathbf{x})] - \mathbb{E}_{\mathbf{x}_1^{i-1}} g(\mathbf{x}) \mathbb{E}_{\mathbf{x}_1^{i-1}} \tilde{g}(\mathbf{x}) \geq 0. \quad (5.45)$$

Again, the monotonicity w.r.t.  $x_i$  implies

$$\begin{aligned} & \mathbb{E}_{x_i} \mathbb{E}_{x_i'} [(\mathbb{E}_{\mathbf{x}_1^{i-1}} g(\mathbf{x}_1^{i-1}, x_i, \mathbf{x}_{i+1}^n) - \mathbb{E}_{\mathbf{x}_1^{i-1}} g(\mathbf{x}_1^{i-1}, x_i', \mathbf{x}_{i+1}^n)) \\ & \quad \cdot (\mathbb{E}_{\mathbf{x}_1^{i-1}} \tilde{g}(\mathbf{x}_1^{i-1}, x_i, \mathbf{x}_{i+1}^n) - \mathbb{E}_{\mathbf{x}_1^{i-1}} \tilde{g}(\mathbf{x}_1^{i-1}, x_i', \mathbf{x}_{i+1}^n))] \geq 0 \end{aligned}$$

which can be simplified to

$$\mathbb{E}_{x_i} [\mathbb{E}_{\mathbf{x}_1^{i-1}} g(\mathbf{x}) \mathbb{E}_{\mathbf{x}_1^{i-1}} \tilde{g}(\mathbf{x})] - \mathbb{E}_{x_i} g(\mathbf{x}) \mathbb{E}_{x_i} \tilde{g}(\mathbf{x}) \geq 0. \quad (5.46)$$

The induction is ended by noting that with the hypothesis (5.45) the identity (5.46) can further be relaxed to

$$\mathbb{E}_{\mathbf{x}_1^i} [g(\mathbf{x}) \tilde{g}(\mathbf{x})] - \mathbb{E}_{\mathbf{x}_1^i} g(\mathbf{x}) \mathbb{E}_{\mathbf{x}_1^i} \tilde{g}(\mathbf{x}) \geq 0.$$

This ends the induction argument and the proof.  $\square$

### 5.4.3 On the concentration and existence of the free energy

We consider Hamiltonian (5.1) with independent random couplings  $J_A$ ,  $A \subset \{1, \dots, n\}$  and prove the following generic result used in (5.2). We then discuss a simple argument and condition that guarantees the existence of the thermodynamic limit using the first GKS inequality. We verify that these results applied to Example 5.1.

**Proposition 5.1** (Concentration of the free energy). *Let  $J_A$ ,  $A \subset \{1, \dots, n\}$  be independent random variables such that  $\sum_{A \subset \{1, \dots, n\}} \text{Var}(J_A) \leq C_F n$  for some numerical constant  $C_F > 0$ . Then we have  $\mathbb{E}[(F_n - f_n)^2] \leq C_F/n$ .*

*Proof.* The proof is a simple application of the Efron-Stein inequality. Set  $\mathbf{J} \equiv (J_A, A \subset \{1, \dots, n\})$ . Let  $\mathbf{J}^{(A)}$  be a vector such that  $\mathbf{J}^{(A)}$  differs from  $\mathbf{J}$  only at the  $A$ -th component which becomes  $J'_A$  drawn independently from the same distribution as the one of  $J_A$  (note that the random variables  $J_A$  for different  $A$  do not necessarily have the same distribution). Efron Stein's inequality tells us that

$$\mathbb{E}[(F_n - \mathbb{E} F_n)^2] \leq \frac{1}{2} \sum_{A \subset \{1, \dots, n\}} \mathbb{E}_{\mathbf{J} \setminus J_A} \mathbb{E}_{J_A} \mathbb{E}_{J'_A} [(F_n(\mathbf{J}) - F_n(\mathbf{J}^{(A)}))^2]. \quad (5.47)$$

An elementary interpolation gives

$$\begin{aligned} & |F_n(\mathbf{J}) - F_n(\mathbf{J}^{(A)})| \\ &= \frac{1}{n} \left| \int_0^1 ds \frac{d}{ds} \ln \sum_{\sigma \in \{\pm 1\}^n} \exp \{ -\mathcal{H}_0(\sigma, \mathbf{J}^{(A)}) + s(J_A - J'_A)\sigma_A \} \right| \\ &= \frac{1}{n} \left| \int_0^1 ds (J_A - J'_A) \langle \sigma_A \rangle_s \right| \leq \frac{1}{n} |J_A - J'_A|. \end{aligned}$$

Replacing in (5.47) (and recalling  $f_n \equiv \mathbb{E} F_n$ ) gives

$$\mathbb{E}[(F_n - f_n)^2] \leq \frac{1}{2n^2} \sum_{X \subset \{1, \dots, n\}} \mathbb{E}_{J_A} \mathbb{E}_{J'_A} [(J_A - J'_A)^2] = \frac{1}{n^2} \sum_{X \subset \{1, \dots, n\}} \text{Var}(J_A).$$

With the hypothesis on  $\text{Var}(J_A)$ , the proof is complete.  $\square$

An easy and more or less standard superadditivity argument proves that the thermodynamic limit exists for the ferromagnetic model (5.1). We give the argument for completeness. For simplicity we consider that there exists a maximal size  $x_{\max}$  independent of  $n$  such that  $|A| \leq x_{\max}$ . We suppose furthermore that all  $J_A$  are independent with a distribution that depends only on the cardinalities  $|A|$  (in other words, given a cardinality they are i.i.d.) and also

$$\frac{1}{n} \sum_{A \subset \{1, \dots, n\}} \mathbb{E} J_A = \frac{1}{n} \sum_{|A|=1}^{x_{\max}} \binom{n}{|A|} m(|A|) \leq C \quad (5.48)$$

where  $m(|A|) \equiv \mathbb{E} J_A$  and  $C$  a positive constant independent of  $n$ .

**Proposition 5.2** (Existence of the thermodynamic limit of the free energy). *Let  $J_A, A \subset \{1, \dots, n\}$  be independent random variables with a probability distribution supported on  $\mathbb{R}_{\geq 0}$  depending only on  $|A|$ . Moreover, assume  $J_A = 0$  for  $|A| > a_{\max}$  independent of  $n$ . Let (5.48) be satisfied. Then  $\lim_{n \rightarrow +\infty} f_n$  exists and is finite.*

*Proof.* Fix non-zero integers  $n_1, n_2$  such that both are greater than  $x_{\max}$  and  $n \equiv n_1 + n_2$ . Consider a set of realizations  $S \equiv \{J_A, A \subset \{1, \dots, n\}\}$ . This set can be split in three disjoint sets  $S = S_1 \cup S_2 \cup S_{12}$  with  $S_1 \equiv \{J_A, A \subset \{1, \dots, n_1\}\}$ ,  $S_2 \equiv \{J_A, A \subset \{n_1 + 1, \dots, n\}\}$  and  $S_{12} \equiv \{J_A, A \cap \{1, \dots, n_1\} \neq \emptyset, A \cap \{n_1 + 1, \dots, n\} \neq \emptyset\}$ . Let  $-n^{-1} \ln \mathcal{Z}(S)$  be the free energy corresponding to the Hamiltonian with couplings in  $S$ , and  $-n_1^{-1} \ln \mathcal{Z}(S_1)$  and  $-n_2^{-1} \ln \mathcal{Z}(S_2)$  be the free energies corresponding to the Hamiltonians with couplings from  $S_1$  and  $S_2$  only. One can show, using the first GKS inequality, that

$$\ln \mathcal{Z}(S) \geq \ln \mathcal{Z}(S_1) + \ln \mathcal{Z}(S_2).$$

Then averaging over all coupling constants in  $S$ , using that they are independent with distributions depending only on the cardinality  $|A|$  and that all cardinalities are contained in  $S, S_1$  and  $S_2$ , we obtain

$$\mathbb{E}_S \ln \mathcal{Z}(S) \geq \mathbb{E}_{S_1} \ln \mathcal{Z}(S_1) + \mathbb{E}_{S_2} \ln \mathcal{Z}(S_2)$$

which is equivalent to  $nf_n \leq n_1 f_{n_1} + n_2 f_{n_2}$  (for  $n_1, n_2$  greater than  $x_{\max}$ ). This means that the function  $n \mapsto np_n$  is a subadditive sequence and therefore by Fekete's lemma the limit  $\lim_{n \rightarrow +\infty} f_n$  equals  $\inf_n f_n$ . To show that  $\inf_n f_n$  is finite note that

$$f_n \leq \frac{1}{n} \mathbb{E} \min_{\sigma} \mathcal{H}(\sigma) - \ln 2 = -\frac{1}{n} \sum_{A \subset \{1, \dots, n\}} \mathbb{E} J_A - \ln 2 \leq C - \ln 2$$

using  $J_A \geq 0$  and condition (5.48). This ends the proof.  $\square$

Consider now Example 5.1 for  $n$  large and  $K$  fixed. We have  $J_A = 0$  for all subsets with cardinalities  $|A|$  different from 1 and  $K$ . For  $|A| = 1$  the coupling constants  $J_A = H$  are deterministic so obviously  $\text{Var}(J_A) = 0$ . For  $|A| = K$  the couplings  $J_A$  are independent Bernoulli variables taking value  $J$  with probability  $\gamma n \binom{n}{K}^{-1}$  and 0 with complementary probability, so  $\text{Var}(J_A) = J^2 \gamma n \binom{n}{K}^{-1} (1 - \gamma n \binom{n}{K}^{-1})$ . Thus

$$\sum_{A \subset \{1, \dots, n\}} \text{Var}(J_A) = \binom{n}{K} J^2 \gamma n \binom{n}{K}^{-1} \left(1 - \gamma n \binom{n}{K}^{-1}\right) < J^2 \gamma n.$$

Therefore, Proposition 5.1 applies. Similarly, the condition for the existence of the thermodynamic limit of the free energy is also met because the left hand side of (5.48) equals

$$\frac{1}{n} \sum_{A \subset \{1, \dots, n\}} \mathbb{E} J_A = H + J\gamma.$$

The mixed  $K$ -spin models can be treated similarly.



---

# Discussion and Conclusion

---

# 6

In this thesis, we have developed adaptive interpolation methods for computing exactly the asymptotic mutual information for the dense SBM in Chapter 3 and the sparse CBM in Chapter 4. The proofs are direct (without first mapping to external models), conceptually simple and self-contained. They provide new techniques for adaptive interpolation methods and make the methods robust to a broader class of models. If we recall Figure 1.3, we can see that in this thesis we have addressed models belonging to the blue regions in Figure 1.3, namely, dense-graph models and the sparse-graph models with binary erasure channels. It is desirable to extend our adaptive interpolation methods to the other sparse-graph models in Figure 1.3, such as Example 1.1 (sparse SBM) and Example 1.3 (sparse CBM with binary symmetric channels). One major bottleneck of our method for general sparse graphs is proving the concentration of multi-overlaps. In Chapter 5, we have proved the required concentration for related spin models on sparse graphs. The concentration result suggests that the issue about concentration could be addressed for more general models, and this would enable further development of the adaptive interpolation method. Finally, we point out some interesting open questions about the adaptive interpolation method for further development of this thesis.

## **Q1: Does continuous interpolation for sparse graphs exist?**

In Chapter 2 and 3, we have presented continuous adaptive interpolation for dense factor graphs. In fact, the primitive version of interpolation in [55] started with a discrete one. When applying the discrete version to the model in Chapter 2, time is divided into integer steps  $t = 0, 1, \dots, T$ . At each step  $t$ ,

a new decoupled observation

$$\tilde{Y}_i^{(t)} = \frac{q(t/T, \epsilon)}{T} X_i + \frac{\tilde{Z}_i^{(t)}}{\sqrt{T}}, \quad i = 1, \dots, n$$

is observed where  $\tilde{Z}_i^{(t)} \sim \mathcal{N}(0, 1)$ . At the end  $t = T$ , the sum of all these observations is

$$\tilde{Y}_i = \sum_{t=0}^T \left\{ \frac{q(t/T, \epsilon)}{T} X_i + \frac{\tilde{Z}_i^{(t)}}{\sqrt{T}} \right\} \stackrel{d}{=} \frac{1}{T} \sum_{t=0}^T q(t/T, \epsilon) X_i + \tilde{Z}_i, \quad \tilde{Z}_i \sim \mathcal{N}(0, 1).$$

We consider that  $T$  tends to infinity (faster than  $n$ ). This reduces the sum to Riemann integral

$$\frac{1}{T} \sum_{t=0}^T q(t/T, \epsilon) \xrightarrow{T \rightarrow \infty} \int_0^1 ds q(s, \epsilon),$$

and recovers the integral part of  $R(1, \epsilon)$  in (2.7).

On the contrary, in Chapter 4, we have presented a discrete adaptive interpolation for sparse graphs. We might wonder if it could be simplified to a continuous version. This is however not obvious. The interpolation in Chapter 4 considers time  $t = 1, \dots, T$ . Recall (4.18). At time  $t = T$ , ignoring the perturbation by  $H_i$  and  $\tilde{H}_i$ , the decoupled observations has the sum of log-likelihood ratio

$$U_i \equiv \sum_{t=1}^T \sum_{B=1}^{e_i^{(t)}} U_{B \rightarrow i}^{(t)}, \quad i = 1, \dots, n, \quad (6.1)$$

where  $e_i^{(t)} \sim \text{Poi}(\alpha K/T)$  follows a Poisson distribution. A continuous interpolation amounts to finding an integral expression for the distribution of  $U_i$  when  $T$  tends to infinity. For BEC, where  $U_{B \rightarrow i}^{(t)}$  either equals to  $\infty$  with probability  $\tilde{x}^{(t)}$  or equals to 0 otherwise, the distribution of  $U_i$  is characterized by the probability

$$\mathbb{P}(U_i = \infty) = 1 - \prod_{t=1}^T (1 - \tilde{x}^{(t)})^{e_i^{(t)}}.$$

It is not clear how to write this probability with an integral expression similar to  $R(1, \epsilon)$  for the dense graphs.

## Q2: Can we extend Theorem 4.1 to an asymmetric prior?

We might want to understand how to extend Theorem 4.1 if we generalize the prior to  $\mathbb{P}_0(\sigma_i^0) = r\delta_{\sigma_i^0, +1} + (1-r)\delta_{\sigma_i^0, -1}$  with  $r \in (0, 1/2]$ . In this case the posterior (4.1) would be replaced by

$$\mathbb{P}(\boldsymbol{\sigma} | \mathbf{J}) = \frac{1}{\mathcal{Z}} \exp \left\{ \sum_{A=1}^m J_A (\sigma_A - 1) + \sum_{i=1}^n \hat{H}_i \sigma_i \right\} \quad (6.2)$$

where

$$\hat{H}_i \equiv \frac{1}{2} \ln \frac{r}{1-r},$$

$$\mathcal{Z} \equiv \sum_{\sigma \in \{-1, +1\}^n} \exp \left\{ \sum_{A=1}^m J_A (\sigma_A - 1) + \sum_{i=1}^n \hat{H}_i \sigma_i \right\}.$$

The equalities (4.7)–(4.14) that are based on the gauge transformation are no longer true. Nevertheless, we still have the generic Nishimori identity (4.7). In order to use this identity, we expect that the overlap parameters should be re-defined with the one containing  $\sigma^0$ . How this new overlap would appear in the sum rule is unclear, as the expression (6.2) does not immediately contain  $\sigma^0$ .

### Q3: Can we extend Theorem 5.3 (concentration of multi-overlaps) to inference models?

A larger class in inference, such as the CBM in Chapter 4, under any symmetric channels satisfies a relaxed form of GKS inequalities [24]: for example, under suitable perturbation to the Hamiltonian and assuming that Nishimori identity (4.11) is satisfied, for any subsets of variable indices  $S, T \subset \{1 \dots n\}$  we have

$$\mathbb{E}\langle \sigma_S \rangle \geq 0 \quad \text{and} \quad \frac{d}{d\epsilon_T} \mathbb{E}\langle \sigma_S \rangle \geq 0,$$

where  $\epsilon_T$  is the mean and variance of the Gaussian coupling constant of the perturbation that applies to variables with indices  $T$ . These inequalities, however, are not strong enough to reproduce the proof of Theorem 5.3 for the corresponding models. It would be interesting to uncover if the proof approach of Theorem 5.3 is specific to the ferromagnetic spin models, or if any important identities in inference are still waiting to be discovered.

### Q4: Can we extend Theorem 4.1 (the replica prediction for the sparse CBM) to symmetric channels?

To extend Theorem 4.1 to symmetric channels by using adaptive interpolation, we need to be careful when addressing Q3. We need to check if the concentration of multi-overlaps still hold in the interpolating model with an adaptive choice of decoupled observations. This is currently addressed by showing that the Jacobian is lower bounded by a constant (this corresponds to  $dR^*/d\epsilon \geq 1$  in (2.15) in Chapter 2, and  $d\mathbb{E}[\bar{\epsilon}^{(t,s)}]/d\epsilon \geq 1 - \mathbb{E}[\bar{\epsilon}^{(t,s)}]$  in (4.71) in Chapter 3). For symmetric channels, it is not always clear what the distribution of the decoupled observations is, hence it is not clear what the notion of Jacobian should be.

**Q5: What is the replica symmetric formula for ferromagnetic models?**

We can replace all  $J_A$  in (4.1) by a constant  $J > 0$ . The resulting distribution defines a ferromagnetic spin model. We do not have one-sided bound as in 4.1 because we cannot show the remainder in the sum rule is either positive or negative. Although this time we have the full concentration of multi-overlaps due to Theorem 5.3, we are still unable to reproduce the other one-sided bound as in Sec. 4.4.4. This is because we are stuck at reproducing Lemma 4.1. In particular, the second equality of (4.96) does not hold. Note that this technical issue also appears when we want to prove the conjecture of replica symmetry formula for inference problems on sparse graphs with asymmetric channels.

# Bibliography

---

- [1] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.
- [2] M. E. J. Newman, “Communities, modules and large-scale structure in networks,” *Nature Physics*, vol. 8, pp. 25–31, 12 2011.
- [3] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi, “A survey of statistical network models,” *Foundations and Trends in Machine Learning*, vol. 2, no. 2, pp. 129–233, 2010.
- [4] S. Fortunato, “Community detection in graphs,” *Physics Reports*, vol. 486, no. 3, pp. 75–174, 2010.
- [5] M. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Physical Review E*, vol. 69, p. 026113, 2004.
- [6] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, “Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications,” *Physical Review E*, vol. 84, no. 6, p. 066106, 2011.
- [7] P. W. Holland, K. B. Laskey, and S. Leinhardt, “Stochastic blockmodels: First steps,” *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983.
- [8] T. N. Bui, S. Chaudhuri, F. T. Leighton, and M. Sipser, “Graph bisection algorithms with good average case behavior,” in *25th Annual Symposium on Foundations of Computer Science*, 1984, pp. 181–192.
- [9] B. Söderberg, “General formalism for inhomogeneous random graphs,” *Physical Review E*, vol. 66, p. 066121, 2002.
- [10] B. Bollobás, S. Janson, and O. Riordan, “The phase transition in inhomogeneous random graphs,” *Random Struct. Algorithms*, vol. 31, no. 1, pp. 3–122, 2007.
- [11] E. Abbe, “Community detection and stochastic block models: Recent developments,” *Journal of Machine Learning Research*, vol. 18, 2018.

- [12] E. Abbe and A. Montanari, “Conditional random fields, planted constraint satisfaction and entropy concentration,” in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 2013, pp. 332–346.
- [13] E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer, “Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery,” *IEEE Transactions on Network Science and Engineering*, vol. 1, no. 1, pp. 10–22, Jan 2014.
- [14] P. Chin, A. Rao, and V. Vu, “Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery,” in *Proceedings of The 28th Conference on Learning Theory*, vol. 40, Jul 2015, pp. 391–423.
- [15] A. Saade, M. Lelarge, F. Krzakala, and L. Zdeborová, “Spectral detection in the censored block model,” in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1184–1188.
- [16] Y. Deshpande, E. Abbe, and A. Montanari, “Asymptotic mutual information for the balanced binary stochastic block model,” *Information and Inference: A Journal of the IMA*, vol. 6, no. 2, pp. 125–170, 2017.
- [17] M. Lelarge and L. Miolane, “Fundamental limits of symmetric low-rank matrix estimation,” *Probability Theory and Related Fields*, Apr 2018.
- [18] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [19] S. Nicolas, “Spin glass models as error-correcting codes,” *Nature*, vol. 339, p. 693, June 1989.
- [20] T. Tanaka, “A statistical-mechanics approach to large-system analysis of cdma multiuser detectors,” *IEEE Transactions on Information Theory*, vol. 48, no. 11, pp. 2888–2910, Nov 2002.
- [21] Y. Fu and P. W. Anderson, “Application of statistical mechanics to NP-complete problems in combinatorial optimisation,” *Journal of Physics A Mathematical General*, vol. 19, pp. 1605–1620, Jun. 1986.
- [22] D. J. Amit, H. Gutfreund, and H. Sompolinsky, “Spin-glass models of neural networks,” *Phys. Rev. A*, vol. 32, pp. 1007–1018, Aug 1985.
- [23] R. B. Griffiths, *Phase transitions and critical phenomena*, ser. Phase Transitions and Critical Phenomena, C. Domb and M. S. Green, Eds. Academic Press, 1972, vol. 1.
- [24] N. Macris, “Griffith-Kelly-Sherman correlation inequalities: A useful tool in the theory of error correcting codes,” *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 664–683, Feb 2007.

- [25] —, “Sharp bounds on generalized exit functions,” *IEEE Transactions on Information Theory*, vol. 53, no. 7, pp. 2365–2375, July 2007.
- [26] M. Mezard and A. Montanari, *Information, Physics, and Computation*. New York, NY, USA: Oxford University Press, Inc., 2009.
- [27] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin-Glass Theory and Beyond*. Singapore: World Scientific, 1987, vol. 9.
- [28] M. Mézard and A. Montanari, *Information, Physics and Computation*. Oxford Press, 2009.
- [29] J. S. Yedidia, W. T. Freeman, and Y. Weiss, “Constructing free-energy approximations and generalized belief propagation algorithms,” *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2282–2312, July 2005.
- [30] F. Krzakala, J. Xu, and L. Zdeborová, “Mutual information in rank-one matrix estimation,” in *2016 IEEE Information Theory Workshop, ITW 2016*, 2016, pp. 71–75.
- [31] S. B. Korada and A. Montanari, “Applications of the Lindeberg principle in communications and statistical learning,” *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2440–2450, Apr 2011.
- [32] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister, “Threshold saturation for spatially coupled ldpc and ldgm codes on bms channels,” *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7389–7415, Dec 2014.
- [33] M. Bayati and A. Montanari, “The dynamics of message passing on dense graphs, with applications to compressed sensing,” *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, Feb 2011.
- [34] G. Parisi, “A sequence of approximated solutions to the S-K model for spin glasses,” *Journal of Physics A Mathematical General*, vol. 13, pp. L115–L121, Apr. 1980.
- [35] M. Talagrand, “The parisi formula,” *Annals of Mathematics*, vol. 163, no. 1, pp. 221–263, 2006.
- [36] F. Guerra and F. L. Toninelli, “The thermodynamic limit in mean field spin glass models,” *Communications in Mathematical Physics*, vol. 230, no. 1, pp. 71–79, 2002.
- [37] M. Aizenman, R. Sims, and S. L. Starr, “Extended variational principle for the sherrington-kirkpatrick spin-glass model,” *Phys. Rev. B*, vol. 68, p. 214403, Dec 2003.

- [38] J. Barbier, M. Dia, N. Macris, and F. Krzakala, “The Mutual Information in Random Linear Estimation,” in *in the 54th Annual Allerton Conference on Communication, Control, and Computing*, September 2016.
- [39] G. Reeves and H. D. Pfister, “The replica-symmetric prediction for compressed sensing with gaussian matrices is exact,” in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 665–669.
- [40] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, “Optimal errors and phase transitions in high-dimensional generalized linear models,” in *Proceedings of the 31st Conference On Learning Theory*, vol. 75, Jul 2018, pp. 728–731.
- [41] M. Gabri  , A. Manoel, C. Luneau, J. Barbier, N. Macris, F. Krzakala, and L. Zdeborov  , “Entropy and mutual information in models of deep neural networks,” in *Advances in Neural Information Processing Systems 31*, 2018, pp. 1826–1836.
- [42] B. Aubin, A. Maillard, J. Barbier, F. Krzakala, N. Macris, and L. Zdeborov  , “The committee machine: Computational to statistical gaps in learning a two-layers neural network,” in *Advances in Neural Information Processing Systems 31*, 2018, pp. 3227–3238.
- [43] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborov  , “Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula,” in *Advances in Neural Information Processing Systems (NIPS) 29*, 2016, pp. 424–432.
- [44] T. Lesieur, L. Miolane, M. Lelarge, F. Krzakala, and L. Zdeborov  , “Statistical and computational phase transitions in spiked tensor estimation,” in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 511–515.
- [45] J. Barbier, N. Macris, and L. Miolane, “The Layered Structure of Tensor Estimation and its Mutual Information,” in *47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2017.
- [46] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborov  , “Information-theoretic thresholds from the cavity method,” *Advances in Mathematics*, vol. 333, pp. 694–795, 2018.
- [47] J. Barbier and N. Macris, “The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference,” *Probability Theory and Related Fields*, Oct 2018.
- [48] —, “The adaptive interpolation method for proving replica formulas. applications to the curie-weiss and wigner spike models,” *Journal of Physics A: Mathematical and Theoretical*, vol. 52, no. 29, p. 294002, jun 2019.



- [49] T. Lesieur, F. Krzakala, and L. Zdeborová, “Phase transitions in sparse pca,” in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 1635–1639.
- [50] J. Barbier, C. Chan, and N. Macris, “Mutual information for the stochastic block model by the adaptive interpolation method,” in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019.
- [51] J. Barbier, C. L. Chan, and N. Macris, “Mutual information for the stochastic block model by the adaptive interpolation method,” *CoRR*, 2019. [Online]. Available: <http://arxiv.org/abs/1902.07273>
- [52] J. Barbier, C. Chan, and N. Macris, “Adaptive path interpolation for sparse systems: Application to a simple censored block model,” in *2018 IEEE International Symposium on Information Theory (ISIT)*, 2018.
- [53] J. Barbier, C. L. Chan, and N. Macris, “Adaptive path interpolation for sparse systems: Application to a simple censored block model,” *CoRR*, 2018. [Online]. Available: <http://arxiv.org/abs/1806.05121>
- [54] —, “Concentration of multi-overlaps for random ferromagnetic spin models,” *CoRR*, 2019. [Online]. Available: <http://arxiv.org/abs/1901.06521>
- [55] J. Barbier and N. Macris, “The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference,” *Probability Theory and Related Fields*, Oct 2018.
- [56] J. Barbier and N. Macris, “The adaptive interpolation method for proving replica formulas. Applications to the Curie-Weiss and Wigner spike models,” *Journal of Physics A: Mathematical and General*, vol. JPhysA-111295.R1, 2019.
- [57] H. Nishimori, “Internal energy, specific heat and correlation function of the bond-random ising model,” *Progress of Theoretical Physics*, vol. 66, no. 4, pp. 1169–1181, 10 1981.
- [58] —, *Statistical Physics of Spin Glasses and Information Processing: an Introduction*. Oxford University Press, 2001.
- [59] P. Hartman, *Ordinary Differential Equations*, 2nd ed. Society for Industrial and Applied Mathematics, 2002.
- [60] M. Aizenman, R. Sims, and S. L. Starr, “Extended variational principle for the Sherrington-Kirkpatrick spin-glass model,” *Physical Review B*, vol. 68, no. 21, p. 214403, 2003.
- [61] S. B. Korada and N. Macris, “Exact solution of the gauge symmetric p-spin glass model on a complete graph,” *Journal of Statistical Physics*, vol. 136, no. 2, pp. 205–230, 2009.

- [62] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, “Optimal errors and phase transitions in high-dimensional generalized linear models,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 12, pp. 5451–5460, 2019.
- [63] M. Talagrand, *Spin Glasses: A Challenge for Mathematicians - Cavity and Mean Field Models*. Cambridge University Press, 2004.
- [64] P. Carmona and Y. Hu, “Universality in Sherrington-Kirkpatrick’s spin glass model,” *Annales de l’Institut Henri Poincaré (B) Probability and Statistics*, vol. 42, no. 2, pp. 215 – 222, 2006.
- [65] S. Franz and M. Leone, “Replica bounds for optimization problems and diluted spin systems,” *Journal of Statistical Physics*, vol. 111, no. 3, pp. 535–564, 2003.
- [66] M. Aizenman, R. Sims, and S. L. Starr, “Extended variational principle for the Sherrington-Kirkpatrick spin-glass model,” *Physical Review B*, vol. 68, no. 21, p. 214403, 2003.
- [67] A. Giurgiu, N. Macris, and R. Urbanke, “Spatial coupling as a proof technique and three applications,” *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5281–5295, Oct 2016.
- [68] E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer, “Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery,” *IEEE Transactions on Network Science and Engineering*, vol. 1, no. 1, pp. 10–22, Jan 2014.
- [69] S. Kudekar and N. Macris, “Sharp bounds for optimal decoding of low-density parity-check codes,” *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4635–4650, Oct 2009.
- [70] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.
- [71] A. Montanari, “Tight bounds for ldpc and ldgm codes under map decoding,” *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3221–3246, Sept 2005.
- [72] N. Macris, “Griffith–kelly–sherman correlation inequalities: A useful tool in the theory of error correcting codes,” *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 664–683, Feb 2007.
- [73] T. Richardson and R. Urbanke, *Modern Coding Theory*. New York, NY, USA: Cambridge University Press, 2008.
- [74] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. New York, NY, USA: Cambridge University Press, 2005.

- [75] L. A. Pastur and M. V. Shcherbina, “Absence of self-averaging of the order parameter in the sherrington-kirkpatrick model,” *Journal of Statistical Physics*, vol. 62, no. 1, pp. 1–19, Jan 1991.
- [76] F. Guerra and F. L. Toninelli, “The thermodynamic limit in mean field spin glass models,” *Communications in Mathematical Physics*, vol. 230, no. 1, pp. 71–79, Sep 2002.
- [77] —, “Quadratic replica coupling in the sherrington–kirkpatrick mean field spin glass model,” *Journal of Mathematical Physics*, vol. 43, no. 7, pp. 3704–3716, 2002.
- [78] F. Guerra, “An introduction to mean field spin glass theory: methods and results,” *Mathematical Statistical Physics: Lecture Notes of the Les Houches Summer School 2005*, pp. 243–271, 2005.
- [79] S. Ghirlanda and F. Guerra, “General properties of overlap probability distributions in disordered spin systems. towards parisi ultrametricity,” *Journal of Physics A: Mathematical and General*, vol. 31, no. 46, p. 9149, 1998.
- [80] M. Talagrand, *Mean field models for spin glasses: Volume I: Basic examples*. Springer Science & Business Media, 2010, vol. 54.
- [81] —, *Mean field models for spin glasses: Volume II: Advanced Replica-Symmetry and Low Temperature*. Springer Science & Business Media, 2011, vol. 55.
- [82] S. Chatterjee, “Absence of replica symmetry breaking in the random field ising model,” *Communications in Mathematical Physics*, vol. 337, no. 1, pp. 93–102, Jul 2015.
- [83] A. Barra and L. De Sanctis, “Overlap fluctuations from the boltzmann random overlap structure,” *Journal of mathematical physics*, vol. 47, no. 10, p. 103305, 2006.
- [84] —, “Stability properties and probability distributions of multi-overlaps in dilute spin glasses,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2007, no. 08, p. P08025, 2007.
- [85] S. Franz, M. Leone, and F. L. Toninelli, “Replica bounds for diluted non-poissonian spin systems,” *Journal of Physics A: Mathematical and General*, vol. 36, no. 43, p. 10967, 2003.
- [86] L. De Sanctis and S. Franz, “Self-averaging identities for random spin systems,” in *Spin Glasses: Statics and Dynamics*. Birkhäuser Basel, 2009, pp. 123–142.

- [87] M. Aizenman and P. Contucci, “On the stability of the quenched state in mean-field spin-glass models,” *Journal of Statistical Physics*, vol. 92, no. 5, pp. 765–783, Sep 1998.
- [88] D. Panchenko, *The Sherrington-Kirkpatrick Model*. Springer Monographs in Mathematics, 2013.
- [89] M. Bayati, D. Gamarnik, and P. Tetali, “Combinatorial approach to the interpolation method and scaling limits in sparse random graphs,” in *Proceedings of the forty-second ACM symposium on Theory of computing*. ACM, 2010, pp. 105–114.
- [90] J. Salez, “The interpolation method for random graphs with prescribed degrees,” *Combinatorics, Probability and Computing*, vol. 25, no. 3, pp. 436–447, 2016.
- [91] T. M. Liggett, “T. E. Harris’ contributions to interacting particle systems and percolation,” *Ann. Probab.*, vol. 39, no. 2, pp. 407–416, 03 2011. [Online]. Available: <https://doi.org/10.1214/10-AOP593>

EDUCATION

- Ph.D. in Computer and Communication Sciences**    Sep 2015 – Dec 2019  
École Polytechnique Fédérale de Lausanne
- M.Phil. in Information Engineering**    Aug 2013 – Jul 2015  
**B.Eng. in Information Engineering**    Aug 2009 – Jan 2013  
The Chinese University of Hong Kong

JOURNAL  
PREPRINTS

- [1] J. Barbier, **C. L. Chan** and N. Macris, “Concentration of multi-overlaps for random ferromagnetic spin models,” Submitted to Journal of Statistical Physics, [Online]. Available: <https://arxiv.org/abs/1901.06521>
- [2] J. Barbier, **C. L. Chan** and N. Macris, “Adaptive path interpolation for sparse systems: application to a simple censored block model,” Submitted to IEEE Transactions on Information Theory, [Online]. Available: <https://arxiv.org/abs/1806.05121>
- [3] J. Barbier, **C. L. Chan** and N. Macris, “Mutual information for the stochastic block model by the adaptive interpolation method,” Submitted to Information and Inference: A Journal of the IMA, [Online]. Available: <https://arxiv.org/abs/1902.07273>

JOURNAL  
PUBLICATIONS

- [J4] D. Cruz *et al.* “Efficient quantum algorithms for GHZ and W states, and implementation on the IBM quantum computer”, Advanced Quantum Technologies, vol. 2, no. 5 – 6, 2019 (**2nd place of IBM Q Best Paper Award**)
- [J5] **C. L. Chan**, W. Fernandes, N. Kashyap, M. Krishnapur, “Phase transitions for the uniform distribution in the pattern maximum likelihood problem and its Bethe approximation,” SIAM J. Discrete Math., vol. 31, no. 1, pp. 597 – 631, 2017
- [J6] **C. L. Chan**, S. Jaggi, V. Saligrama, and S. Agnihotri, “Non-adaptive group testing: explicit bounds and novel algorithms,” IEEE Transactions on Information Theory, vol. 60, no. 5, pp. 3019 – 3035, May 2014

CONFERENCE  
PUBLICATIONS

- [C7] J. Barbier, **C. L. Chan** and N. Macris, “Mutual information for the stochastic block model by the adaptive interpolation method,” IEEE International Symposium on Information Theory (ISIT), Paris, France, 2019.
- [C8] J. Barbier, **C. L. Chan** and N. Macris, “Adaptive path interpolation for sparse systems: application to a simple censored block model,” IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 2018 (**in the TPC choice session**)
- [C9] **C. L. Chan** and N. Macris, “Stability threshold and phase transition of generalized censored block models,” IEEE Information Theory Workshop (ITW), Kaohsiung, Taiwan, 2017.
- [C10] **C. L. Chan**, M. J. Siavoshani, S. Jaggi, N. Kashyap and P. O. Vontobel, “Generalized belief propagation for estimating the partition function of the 2D Ising model,” IEEE International Symposium on Information Theory (ISIT), Hong Kong, China, 2015.
- [C11] T. Li, **C. L. Chan**, W. Huang, T. Kaced, and S. Jaggi, “Group testing with prior statistics,” IEEE International Symposium on Information Theory (ISIT), Honolulu, HI, USA, 2014.
- [C12] **C. L. Chan**, S. Cai, M. Bakshi, S. Jaggi and V. Saligrama, “Stochastic threshold group testing,” IEEE Information Theory Workshop (ITW), Seville, Spain, 2013.
- [C13] **C. L. Chan**, S. Jaggi, V. Saligrama, and S. Agnihotri, “Non-adaptive group testing: explicit bounds and novel algorithms,” IEEE International Symposium on Information Theory (ISIT), Boston, MA, USA, 2012.

- [C14] **C. L. Chan**, P. H. Che, S. Jaggi, V. Saligrama, “Non-adaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms,” 49th Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 2011.

**TEACHING**

TA for bachelor courses: quantum computation (Fall & Spring 2016, Fall 2017, Spring 2018)

TA for master courses: machine learning (Fall 2018), learning theory (Spring 2019)

**SERVICE**

Reviewer for ITW, ISIT, and IEEE Trans. Inform. Theory

Committee member of EPFL Quantum Computing Association