



**HAL**  
open science

# Algorithmic aspects of the continuous random energy model

Fu-Hsuan Ho

► **To cite this version:**

Fu-Hsuan Ho. Algorithmic aspects of the continuous random energy model. General Mathematics [math.GM]. Université Paul Sabatier - Toulouse III, 2023. English. NNT : 2023TOU30184 . tel-04347664

**HAL Id: tel-04347664**

**<https://theses.hal.science/tel-04347664v1>**

Submitted on 15 Dec 2023

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

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



# THÈSE

En vue de l'obtention du

**DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

---

---

Présentée et soutenue le *25 octobre 2023* par :

**FU-HSUAN HO**

**Aspects algorithmiques du modèle continu à énergie aléatoire**

---

---

## JURY

JEAN-MARC AZAÏS  
LISA HARTUNG  
YUEYUN HU  
MARC LELARGE  
PASCAL MAILLARD  
MATHIEU SABLİK

Université de Toulouse  
Universität Mainz  
Université Sorbonne Paris Nord  
INRIA  
Université de Toulouse  
Université de Toulouse

Président  
Examinatrice  
Rapporteur  
Examineur  
Directeur de thèse  
Examineur

---

**École doctorale et spécialité :**

*MITT : Domaine Mathématiques : Mathématiques appliquées*

**Unité de Recherche :**

*Institut de Mathématiques de Toulouse (UMR 5219)*

**Directeur de Thèse :**

*Pascal Maillard*

**Rapporteurs :**

*Aukosh Jagannath et Yueyun Hu*





# Remerciements

D'abord, je tiens à remercier chaleureusement mon Doktorvater, Pascal. Merci de m'avoir donné la liberté d'explorer le sujet de thèse avec ma vitesse, mais aussi m'avoir fourni les conseils et les soutiens nécessaires quand j'en avais besoin. J'apprécie également les efforts que tu as déployés pour créer une formidable équipe de recherche, dont j'ai tant bénéficié en interagissant avec les gens qui la composent.

Je tiens à remercier mes deux rapporteurs Aukosh Jagannath et Yueyun Hu ainsi que les examinateurs Jean-Marc Azaïs, Lisa Hartung, Marc Lelarge et Mathieu Sablik. C'est un grand plaisir et honneur que vous ayez accepté d'évaluer mes manuscrits de thèse.

I want to thank my research teammates, Justin and Alexander. It has been a pleasure working and spending time with you both. Justin, thank you for your patience and for sharing your knowledge of mean-field spin glasses with me. I have benefited immensely from our conversations. Alex, vielen Dank für jede Warum-Frage, die du gestellt hast. Diese Fragen haben mir geholfen, logische Lücken zu schließen, die ich sonst nicht hätte schließen können.

Grand merci à Michel pour les deux ans depuis tu es arrivé à Toulouse. J'admire ta capacité de trouver la façon plus claire de résumer et présenter les informations compliqués, ce que je n'ai pas encore complétement appris. J'apprécie non seulement tes compétences en tant que mathématicien professionnel, mais également ta qualité en tant que personne admirable.

C'était mon plaisir d'avoir rencontré Victor, Julie, Alexandre et Anthony. J'ai profité de mon temps avec vous, peu importe si c'était à Toulouse, à Saint-Flour, au Canada ou en Bretagne. Alors que je ne suis pas le meilleur joueur de Mario-Kart, ni le plus passionné de jeux de société ou de tarot, ni le meilleur nageur ou joueur de billard, mon expérience de doctorat perdra beaucoup de sa couleur sans tous ses activités que vous m'avez proposé.

Alberto, je suis ravi de t'avoir rencontré et d'avoir partagé ton énergie positive et tumultueuse pendant ces trois dernières années. Il y aurait eu beaucoup moins de rires sans toi comme voisin au deuxième étage du 1R1. Mon seul regret serait peut-être de ne pas avoir eu assez de temps pour apprendre suffisamment l'espagnol pour comprendre tes blagues, mais je suppose que ce sera pour un autre chapitre dans le monde où nous nous reverrons.

Nomination spéciale : Estelle, je n'aurais pas pu franchir la barrière de la langue française sans que tu fronces les sourcils à chaque fois que tu ne comprenais pas ce que je disais. Vivi, le temps que je passé avec toi et Estelle était super, le gâteau et le risotto étaient super également. Paola, Lucas, Virigle, Fanny, Mitja, Clément

B, Armand et Clément C : merci pour les soirées que vous avez organisé ou les événements que vous avez proposé. L'expérience que vous m'avez apportée a rendu ma vie à Toulouse merveilleuse.

Un grand merci aux anciens et nouveaux doctorants/postdocs du 1R1, 1R2 et 1R3 : Perla, Javi, Laëtitia, Jianyu, Simon, Michelle, Nicolas, Joachim, Mahmoud, Sophia, Mathis, Candice, Chifaa, Louis, Alain, Antonio, Jordi, Tuan, Anna, Irène, Julia, Julieta, Maxime, Diego, Etienne, Benjamin, Chung-Ming, Alejandro, Ming-min.<sup>1</sup> J'aimerais remercier les permanents : Reda, Guillaume, Pierre, et Mireille. J'ai tant d'histoires merveilleuses qu'elles ne pourraient pas tenir dans la marge de cette page ; mais en réalité, ces petites discussions sur les maths, la vie de labo ou simplement les choses futiles me font sentir inclus à l'IMT.

J'aimerais remercier mes collègues de l'INSA Toulouse, où j'ai enseigné pendant les deux dernières années. J'exprime ma gratitude tout particulièrement à Mélisande Albert et Olivier Mazet. Merci à Mélisande de m'avoir fait découvrir ton professionnalisme dans l'enseignement et ta considération pour les étudiants. Merci à Olivier de m'avoir proposé d'enseigner le cours sur le processus de Poisson. C'était un cours formidable qui m'a permis de rencontrer Mélisande et les étudiants motivés de 5ème année au GMM. J'aimerais également remercier mes étudiants : votre passion pour le métier et votre respect pour le cours ont rendu mes heures d'enseignement très agréables.

J'ai effectué deux séjours académiques durant ma thèse, qui n'auraient pas été possibles sans le soutien de Jih-Huang Li et Lisa Hartung, respectivement. Merci à Jih-Huang pour ton soutien lors de ma visite à l'NTU. Je suis content qu'une partie critique de cette thèse ait été réalisée à Taïpei, où j'ai passé mes années de licence. Vielen Dank, Lisa, für die Unterstützung bei meinem Besuch in Mainz und für das Feedback, das du Alex und mir für unser Projekt gegeben hast. Herzlichen Dank auch für die Gastfreundschaft meiner Kollegen und Kolleginnen in der Stochastik-Gruppe sowie einer Kollegin und eines Kollegen der Algebraischen Geometrie-Gruppe.

Je souhaite dire merci à mes ami.e.s qui n'ont pas fait partie du labo, mais qui ont rendu ma vie en Europe bien plus heureuse : Juan, Johannes, Agathe, Léa, Solenn, Leon, Laura, Sonja, Amélie, Xinyu, Reepa, Sambre, Camille, Tomas, Stella, Francesca, Justin... bien sûr, cette liste n'est pas exhaustive<sup>2</sup>.

Les dernières lignes de ces remerciements sont pour mes parents et pour mon frère : merci pour les sessions téléphoniques chaque week-end. Bien qu'elles aient consisté principalement en des conversations banales, elles m'ont aidé à surmonter tant de défis auxquels j'ai été confronté au cours des dernières années. Je vous porte dans mon cœur.

---

<sup>1</sup>Veuillez envoyer un e-mail pour vous plaindre si votre nom n'est pas répertorié ici ; idem si vous souhaitez que votre nom soit supprimé.

<sup>2</sup>D'autant plus qu'elle est faite à la dernière minute.

# Abstract

This thesis explores the algorithmic perspectives of the branching random walk and the continuous random energy model (CREM). Namely, we are interested in constructing polynomial-time algorithms that can sample the model's Gibbs measure with high probability, and to identify the hardness regime, which consists of any inverse temperature  $\beta$  such that such polynomial-time algorithms do not exist.

In Chapter 1, we provide a historical overview of the models and motivate the algorithmic problems under investigation. We also provide an overview on the mean-field spin glasses that motivates the line of our research.

In Chapter 2, we address the sampling problem of the Gibbs measure in the context of branching random walk. We identify a critical inverse temperature  $\beta_c$ , identical to the static critical point, that the a hardness transition occurs. In the subcritical regime  $\beta < \beta_c$ , we establish a recursive sampling algorithm is able to sample the Gibbs measure efficiently. In the supercritical regime  $\beta > \beta_c$ , we show that we cannot find polynomial-time algorithm that belongs to a certain class of algorithms.

In Chapter 3, we turn our attention to the same sampling problem for the continuous random energy model (CREM). For the case where the covariance function of this model is concave, we show that for any inverse temperature  $\beta < \infty$ , the recursive sampling algorithm considered in Chapter 2 is able to sample the Gibbs measure efficiently. For the non-concave case, we identify a critical point  $\beta_G$  that similar hardness transition as the one in Chapter 2 occurs. We also provide a lower bound of the CREM free energy that might be of independent interest.

In Chapter 4, we study the negative moment of the CREM partition function. While this is not connected directly to the main theme of the thesis, it spins off during the course of research.

In Chapter 5, we provide an outlook of some further directions that might be interesting to investigate.

**Keywords:** branching random walk ; continuous random energy model ; Gaussian process ; Gibbs measure ; Kullback–Leibler divergence ; negative moments; spin glass.

**MSC2020 subject classifications:** 60G15, 60J80, 60K35, 68Q17, 82D30.

# Résumé

Cette thèse explore les perspectives algorithmiques de la marche aléatoire branchante et du modèle continu d'énergie aléatoire (CREM). Nous nous intéressons notamment à la construction d'algorithmes en temps polynomial capables d'échantillonner la mesure de Gibbs du modèle avec une grande probabilité, et à identifier le régime de dureté, qui consiste en toute température inverse  $\beta$  telle que de tels algorithmes en temps polynomial n'existent pas.

Dans le Chapitre 1, nous fournissons un aperçu historique des modèles et motivations des problèmes algorithmiques étudiés. Nous donnons également un aperçu des verres de spin à champ moyen qui motive la ligne de notre recherche.

Dans le Chapitre 2, nous abordons le problème de l'échantillonnage de la mesure de Gibbs dans le contexte de la marche aléatoire branchante. Nous identifions une température inverse critique  $\beta_c$ , identique au point critique statique, où une transition de dureté se produit. Dans le régime sous-critique  $\beta < \beta_c$ , nous établissons qu'un algorithme d'échantillonnage récursif est capable d'échantillonner efficacement la mesure de Gibbs. Dans le régime supercritique  $\beta > \beta_c$ , nous montrons que nous ne pouvons pas trouver d'algorithme en temps polynomial qui appartienne à une certaine classe d'algorithmes.

Dans le Chapitre 3, nous portons notre attention sur le même problème d'échantillonnage pour le modèle continu d'énergie aléatoire (CREM). Dans le cas où la fonction de covariance de ce modèle est concave, nous montrons que pour toute température inverse  $\beta < \infty$ , l'algorithme d'échantillonnage récursif considéré au Chapitre 2 est capable d'échantillonner efficacement la mesure de Gibbs. Pour le cas non concave, nous identifions un point critique  $\beta_G$  où une transition de dureté similaire à celle du Chapitre 2 se produit. Nous fournissons également une borne inférieure de l'énergie libre du CREM qui pourrait être d'un intérêt indépendant.

Dans le Chapitre 4, nous étudions le moment négatif de la fonction de partition du CREM. Bien que cela ne soit pas directement lié au thème principal de la thèse, cela découle du cours de la recherche.

Dans le Chapitre 5, nous donnons un aperçu de certaines orientations futures qui pourraient être intéressantes à étudier.

**Mots-clé :** divergence de Kullback–Leibler ; marche aléatoire branchante ; mesure de Gibbs ; modèle continu à énergie aléatoire ; moments négatifs ; processus gaussien ; verre de spin.

**Classifications par métiers MSC 2020 :** 60G15, 60J80, 60K35, 68Q17, 82D30.

# Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Branching random walks . . . . .	1
1.1.1	Bienaymé–Galton–Watson tree . . . . .	1
1.1.2	Branching random walks . . . . .	3
1.1.3	Results on the supercritical branching random walks . . . . .	3
1.1.4	Other related models . . . . .	9
1.2	Derrida’s random energy model and its generalizations . . . . .	11
1.2.1	Random energy model . . . . .	11
1.2.2	Generalized random energy model . . . . .	12
1.2.3	Continuous random energy model . . . . .	13
1.2.4	Maximum asymptotics of time-inhomogeneous BRW . . . . .	14
1.3	Mean-field spin glasses . . . . .	15
1.3.1	Edwards–Anderson model . . . . .	15
1.3.2	Sherrington–Kirkpatrick model . . . . .	15
1.3.3	Chaos in mean-field spin glasses . . . . .	17
1.3.4	Thouless–Anderson–Palmer approach . . . . .	18
1.4	Algorithmic perspectives of spin glasses . . . . .	21
1.4.1	Glauber dynamics on mean-field spin glasses . . . . .	22
1.4.2	Optimization problem for mean-field spin glasses . . . . .	23
1.4.3	Optimization problem for the CREM . . . . .	24
1.5	Summary of contributions . . . . .	25
1.5.1	Sampling the Gibbs measure . . . . .	25
1.5.2	Recursive sampling algorithm . . . . .	28
1.5.3	Summary of Chapter 2 . . . . .	28
1.5.4	Branching random walk . . . . .	29
1.5.5	Contributions in Chapter 2 . . . . .	31
1.5.6	Summary of Chapter 3 . . . . .	34
1.5.7	Continuous random energy model . . . . .	34
1.5.8	Contributions in Chapter 3 . . . . .	36
1.5.9	Negative moments of the CREM partition function . . . . .	40
1.5.10	Contributions in Chapter 4 . . . . .	40
<b>2</b>	<b>Sampling the BRW Gibbs measure</b>	<b>45</b>
2.1	Introduction . . . . .	45
2.1.1	Notation and main results . . . . .	46
2.1.2	Related work . . . . .	51

2.2	Decomposition of the Kullback–Leibler divergence . . . . .	52
2.3	Some $L^p$ bounds . . . . .	55
2.4	Proof of Theorem 2.1.7 . . . . .	56
2.4.1	Proof of (2.1.9) . . . . .	56
2.4.2	Proof of (2.1.10) . . . . .	57
2.5	Proof of hardness result (Theorem 2.1.9) . . . . .	60
2.6	Proof of Lemma 2.1.4 . . . . .	62
2.7	Open questions and further directions . . . . .	63
<b>3</b>	<b>Sampling the CREM Gibbs measure</b>	<b>64</b>
3.1	Introduction . . . . .	64
3.1.1	Definitions and notation . . . . .	65
3.1.2	Main results . . . . .	68
3.1.3	Discussion and related work . . . . .	70
3.2	Decomposition of the KL divergence $\mathbf{d}(\mu_{\beta,M,N}  \mu_{\beta,N})$ . . . . .	73
3.2.1	Proof of Theorem 3.1.9 . . . . .	75
3.3	Asymptotics of the KL divergence $\mathbf{d}(\mu_{\beta,M,N}  \mu_{\beta,N})$ . . . . .	76
3.3.1	Proof of Proposition 3.3.1 . . . . .	77
3.4	A property of the Gibbs measure in the supercritical regime . . . . .	78
3.5	Hardness in the supercritical regime . . . . .	81
3.5.1	Proof of Proposition 3.5.4 . . . . .	83
3.5.2	Proof of Lemma 3.5.6 . . . . .	83
3.5.3	Proof of Proposition 3.5.5 . . . . .	85
3.A	A lower bound of the free energy $F_\beta$ . . . . .	87
3.B	Proof of Lemma 3.3.2 . . . . .	92
<b>4</b>	<b>Negative moments of the CREM partition function</b>	<b>97</b>
4.1	Introduction . . . . .	97
4.1.1	Historical background . . . . .	99
4.1.2	Proof strategy . . . . .	100
4.2	Initial estimate for the left tail of the partition function . . . . .	100
4.2.1	Useful properties of $A$ and $\hat{A}$ . . . . .	101
4.2.2	Moment estimates of the truncated partition function . . . . .	102
4.2.3	Proof of Proposition 4.2.1 . . . . .	109
4.3	Left tail estimates via a bootstrap argument . . . . .	110
4.4	Proof of Theorem 4.1.2 . . . . .	113
<b>5</b>	<b>Outlook</b>	<b>116</b>
5.1	A renormalized BRW free energy via the Ruelle cascade . . . . .	116
5.1.1	Case $r = 1$ of Problem 5.1.5 . . . . .	117
5.1.2	Case $r > 1$ of Problem 5.1.5 . . . . .	122
5.2	Fluctuation of the CREM free energy . . . . .	123
<b>A</b>	<b>Errata</b>	<b>127</b>
	<b>Bibliography</b>	<b>128</b>

# 1. Introduction

The goal of the introduction is to provide an overview of the models that we consider through the thesis, and to motivate the algorithmic problems that we study. Section 1.1 provides some relevant background on branching random walk. Section 1.2 focuses on the Derrida’s generalized random energy model and the continuous random energy. While the thesis does not concern directly about mean-field spin glasses, historically, the motivation to introduce the models mentioned in Section 1.2 were to have a deeper understanding of mean-field spin glasses. Therefore, we provide a historical review on this topic in Section 1.3. Finally, Section 1.5 summarizes the three articles that compose the thesis:

- *Efficient approximation of branching random walk Gibbs measures* in collaboration with Pascal Maillard, published in the Electronic Journal of Probability, 1–18, (2022).
- *Efficient sampling of the CREM Gibbs measure*, submitted with the name *Sampling from the Gibbs measure of the continuous random energy model and the hardness threshold*.
- *Negative moments for the continuous random energy model in the subcritical regime*, in preparation.

## 1.1 Branching random walks

In this section, we survey some results of the branching random walks that concern the thesis. We refer to Shi [168] for a modern treatment of the mathematical theory of branching random walks.

### 1.1.1 Bienaymé–Galton–Watson tree

The Bienaymé–Galton–Watson (BGW) tree (mostly known as the Galton–Watson tree) is one of the earliest and simplistic model for the study of population dynamics, and the model can be described informally as follows. Starting from one individual, the individual reproduces into a random number of children according to a given probability distribution  $L$ , called the offspring distribution. Next, individuals at generation  $n + 1$  independently give birth to random numbers of children with identical offspring distribution  $L$ , and for each individual at generation  $n + 1$ , the number of children is also independent of the previous generations. See Figure 1.1 for an example of the BGW trees.

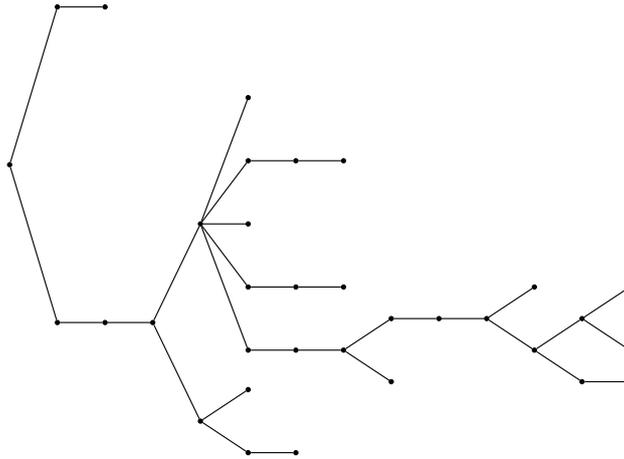


Figure 1.1: An example of a BGW tree with offspring distribution  $L$  that is a Poisson distribution with parameter 0.99. The branching random walks used as underlying models in this thesis do not possess such sparse genealogical trees. Indeed, the genealogical trees we consider are deterministic and have a fixed number of offspring  $d \geq 2$  (see Section 1.5.4).

It is well-known that the expectation of the offspring distribution dictates whether the size of tree is finite or not. This fact is well-established and the proof can be found in many place, such as in the lecture notes [168] by Shi.

**Fact 1.1.1** (Theorem 2.1, Shi [168]). *Let  $\mathbf{t}$  be a BGW tree with offspring distribution  $L$ . Denote by  $|\mathbf{t}|$  the size of the BGW tree. Then, the following dichotomy holds.*

- (i) *If  $\mathbb{E}[L] \leq 1$ , then  $\mathbb{P}(|\mathbf{t}| < \infty) = 1$ .*
- (ii) *If  $\mathbb{E}[L] > 1$ , then  $\mathbb{P}(|\mathbf{t}| = \infty) > 0$ .*

In the following, a BGW tree with  $\mathbb{E}[L] < 1$ ,  $\mathbb{E}[L] = 1$  and  $\mathbb{E}[L] > 1$  is refer to as a subcritical, critical and supercritical BGW tree, respectively.

The BGW trees were studied by Bienaymé [80] in 1845 and then independently by Galton and Watson [101] in 1875 in order to understand the growth and extinction of patronymics. While Galton and Watson adopted the correct method based on considering the generating function the offspring distribution, they wrongly concluded that the population always goes extinct almost surely. On the other hand, Bienaymé correctly stated that the extinction probability equals 1 if and only if the mean with respect to offspring distribution is at most 1, although a mathematical justification by himself was never found. Later in 1847, Cournot, a friend of Bienaymé, presented a problem in [81] where he acknowledged that the problem was equivalent to Bienaymé work. In modern language, Cournot considered the generating function of a BGW tree with at most  $m$  children. Unfortunately, Bienaymé contributions were forgotten until 1972, Heyde and Senata rediscovered his work

in the 1845 report of the Philomatic Society of Paris [80]. A more detailed version of the story above can be found in Kendall [120, 121] and Bacaër [24].

During the first half of the 20th century, the BGW trees and their generalizations regained attention and were extensively studied mathematically in tandem with the rapid development of genetics and nuclear physics. These models are referred to as the branching processes in the literature, and we refer the readers to the book by Harris [111] which presented the theoretical results as well as the applications of branching processes up until the early 1960s.

### 1.1.2 Branching random walks

Among all the generalizations of the BGW trees, we are interested in one particular example which is the branching random walks (BRW). Roughly speaking, a branching random walk can be constructed by placing i.i.d. random variables, called the increments, on a given BGW tree, where the random variables are also independent of the underlying BGW tree. The underlying BGW tree of a branching random walk is often referred to as the genealogical tree.

Formally, the branching random walk is defined as follows. Let  $\mathcal{L} = \sum_{i=1}^L \delta_{Y_i}$  be a point process, where  $(Y_i)_{i=1}^\infty$  is a sequence of i.i.d. random variables independent of  $L$ . Let  $(\mathcal{L}_u)_{u \in \mathcal{T}}$  be a collection of i.i.d. copies of  $\mathcal{L}$  indexed by the Ulam–Harris tree  $\mathcal{T}$ , where  $\mathcal{L}_u = \sum_{i=1}^{L_u} \delta_{Y_{ui}}$  provides the information of the amount of children of  $u$  and the increment of each child of  $u$ . A branching random walk  $(X_u)_{u \in \mathfrak{t}}$  indexed by the BGW tree  $\mathfrak{t}$  is a collection of random variables such that

$$\begin{cases} X_\emptyset = 0 \\ X_u = \sum_{\emptyset < w \leq v} Y_u, \quad u \neq \emptyset. \end{cases} \quad (1.1.1)$$

We provide two examples in Figure 1.2. One key quantity to study the branching random walk is the log-Laplace transform of  $\mathcal{L}$ :

$$\varphi(\beta) := \log \mathbb{E} \left[ \sum_{|u|=1}^L e^{\beta Y_i} \right], \quad \beta \in \mathbb{R}. \quad (1.1.2)$$

Let  $\mathcal{D}(\varphi) = \{\beta \in \mathbb{R} : \varphi(\beta) < \infty\}$ . The critical inverse temperature is defined as

$$\beta_c := \sup\{\beta \in \mathcal{D}(\varphi)^\circ \mid \beta \varphi'(\beta) < \varphi(\beta)\} \in (0, \infty], \quad (1.1.3)$$

where  $\mathcal{D}(\varphi)^\circ$  is the interior of  $\mathcal{D}(\varphi)$ .

### 1.1.3 Results on the supercritical branching random walks

In the following, we consider the supercritical branching random walk, i.e.,

$$\mathbb{E}[L] > 1.$$

See Figure 1.1.3 for an example.

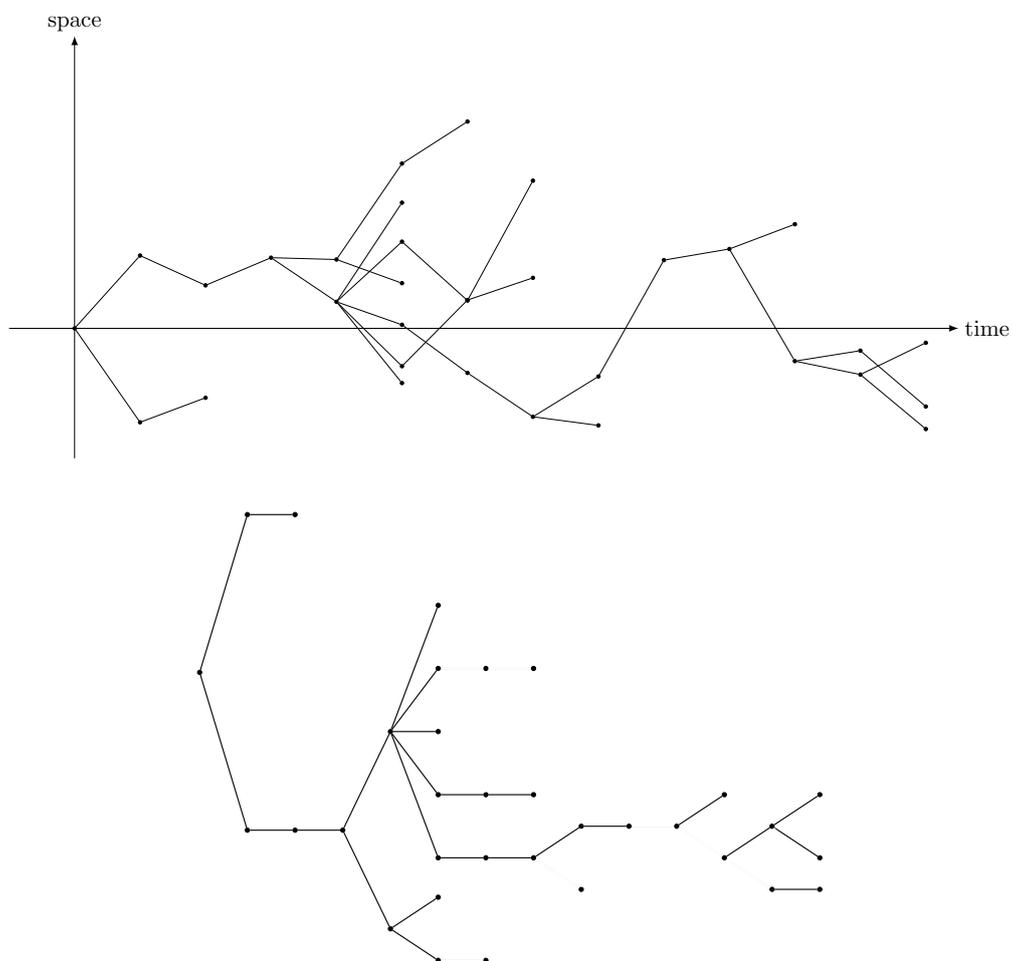


Figure 1.2: Two branching random walks realize on the BGW tree shown in Figure 1.1. The figure above represents the branching random walk with standard Gaussian increments, where the depths are represented as displacement in time and increments are represented as displacements in space. The figure below represents the branching random walk with increments obeying the Bernoulli distribution with parameter  $p = 0.5$ , where each edge remains if the value of the associating Bernoulli distribution equals 1 and is deleted if the value of the associating Bernoulli distribution equals 0.

We are in particular interesting in the following quantities: the maximum, the partition function and the Gibbs measure. For  $\beta \in \mathbb{R}$ , the partition function of a branching random walk is defined as

$$Z_{\beta,N} := \sum_{|u|=N} e^{\beta X_u}. \quad (1.1.4)$$

For  $\beta \in \mathbb{R}$  and  $N \in \mathbb{N}$ , the Gibbs measure  $\mu_{\beta,N}$  of a branching random walk is a (random) measure defined on  $\mathbf{t}_N$  such that for any  $u \in \partial \mathbf{t}_N$ ,

$$\mu_{\beta,N}(u) := \frac{Z_{\beta,N-|u|} e^{\beta X_u}}{Z_{\beta,N}}. \quad (1.1.5)$$

### Asymptotics of the maximum

The asymptotics of the maximum of supercritical branching random walks has been well-studied. Hammersley [110], Kingman [122] and Biggins [44] provided the first order asymptotics of the maximum of the branching random walk. Their discovery can be summarized as follows.

$$\frac{1}{N} \mathbb{E} \left[ \max_{|u|=N} X_u \right] \rightarrow \varphi'(\beta_c), \quad N \rightarrow \infty. \quad (1.1.6)$$

In [3], Addario-Berry and Reed expanded the expectation of the maximum up to  $O(1)$  precision. Combining with the result of Hammersley–Kingman–Biggins, we now know that

$$\mathbb{E} \left[ \max_{|u|=N} X_u \right] = \varphi'(\beta_c)N - \frac{3}{2\beta_c} \log N + O(1). \quad (1.1.7)$$

In the same paper, Addario-Berry and Reed also established tightness of the maximum around its expectation. Around the same time, Hu and Shi established in [114] a second order almost sure convergence theorem of the maximum. Among other things, they showed that conditioned on survival,

$$\lim_{N \rightarrow \infty} \frac{1}{\log N} \max_{|u|=N} X_u = \frac{3}{2\beta_c} \log N, \quad \text{in probability.} \quad (1.1.8)$$

Finally, Aïdékon showed in [6] that the maximum converges in distribution to a randomly shifted Gumbel distribution. More precisely, define the so-called derivative martingale.

$$D_{\beta,N} := \sum_{|u|=N} (X_u - \varphi'(\beta)N) e^{\beta X_u - \varphi(\beta)N} \quad (1.1.9)$$

It was proven in [45] that conditioned on survival,  $D_N$  converges almost surely to some limit  $D_\infty$  which is a strictly positive on the survival set  $\{|\mathbf{t}| = \infty\}$ . Aïdékon showed in [6] that there exists a constant  $C_* \in (0, \infty)$  such that for any  $x \in \mathbb{R}$ ,

$$\lim_{N \rightarrow \infty} \mathbb{P} \left( \max_{|u|=N} X_u \leq \varphi'(\beta_c)N - \frac{3}{2\beta_c} \log N + x \right) = \mathbb{E} \left[ e^{-C_* e^x D_\infty} \right]. \quad (1.1.10)$$

### Asymptotics of the partition function

Instead of considering its definition as in (1.1.4) directly, the partition function of a branching random walk is usually studied in its normalized form, called the *additive martingale*. The additive martingale was introduced by Mandelbrot [135] (see [138] for the English translation) in the context of multiplicative cascades and by Kingman [122]. For any  $\beta \in \mathbb{R}$ , the additive martingale is defined as follows.

$$W_{\beta,N} := \sum_{|u|=N} e^{\beta X_u - \varphi(\beta)N}, \quad N \geq 1.$$

This quantity provides the information of the typical behavior of a particle sampled by the Gibbs measure  $\mu_{\beta,N}$ . Since  $W_{\beta,N}$  is a positive martingale, it converges almost surely to a limit  $W_{\beta,\infty}$ , i.e.,

$$W_{\beta,N} \rightarrow W_{\beta,\infty}, \quad \text{almost surely } N \rightarrow \infty.$$

In [47], Biggins characterized the phase transition of  $W_{\beta,\infty}$  in terms of  $\beta$ . If  $\mathbb{E}[L \log L] < \infty$ , conditioned on survival, we have almost surely the following.

$$\begin{cases} W_{\beta,\infty} > 0, & \beta \in [0, \beta_c), \\ W_{\beta,\infty} = 0, & \beta \geq \beta_c. \end{cases} \quad (1.1.11)$$

The phase transition in (1.1.11) had already been proven by Kahane and Peyrière [118] in the context of multiplicative cascade, and Lyons simplified Biggins' proof in [129].

Beyond the subcritical regime, it is natural to ask what the correct normalization is such that the limit is non-trivial. In the critical case  $\beta = \beta_c$ , Aïdékon and Shi [7] showed that conditioned on survival,

$$\sqrt{N} Z_{\beta_c,N} \rightarrow \frac{1}{\sigma} \sqrt{\frac{2}{\pi}} D_{\beta_c,\infty}, \quad \text{in probability } N \rightarrow \infty, \quad (1.1.12)$$

where  $\sigma$  is a positive number defined by  $\sigma^2 := \mathbb{E} \left[ \sum_{|u|=1} X_u^2 e^{\beta X_u} \right]$ , and they showed that the mass of  $Z_{\beta_c,N}$  is mainly contributed by the particles of order  $\sqrt{N}$ . In the supercritical regime  $\beta > \beta_c$ , Madaule showed in Theorem 2.3 of his paper [130] (see also Theorem 1 of the note [27] by Barral, Rhodes and Vargas) that

$$N^{3\beta/2} Z_{\beta,N} \rightarrow T_\beta D_{\beta,\infty} \quad \text{in law, } N \rightarrow \infty \quad (1.1.13)$$

where  $T_\beta$  is a stable subordinator of index  $1/\beta$  (See Page 7 of [41] for the definition of a stable subordinator) independent of  $D_{\beta,\infty}$ . Moreover, the mass of  $Z_{\beta,N}$  is mainly contributed by the particles of order  $[\frac{3}{2} \log N - C, \frac{3}{2} \log N + C]$  where  $C$  is some large constant.

To understand the discontinuous transition of the size of the partition function, Alberts and Ortgièse [10] considered the near critical regime where  $\beta_N = \beta \pm N^{-\delta}$ ,

for  $\delta > 0$ , and they showed under strong assumptions that conditioned on survival,

$$W_{\beta_N, N} = \begin{cases} N^{2\delta-3/2+o(1)}, & \beta_N = \beta_c + N^{-\delta} \quad \text{with } 0 < \delta < 1/2, \\ N^{-1/2+o(1)}, & \beta_N = \beta_c \pm N^{-\delta} \quad \text{with } \delta \geq 1/2, \\ \exp\left(\frac{\sigma^2}{2}N^{1-2\delta}(1+o(1))\right), & \beta_N = \beta_c - N^{-\delta} \quad \text{with } 0 < \delta < 1/2. \end{cases} \quad (1.1.14)$$

On the other hand, Madaule [131] studied the near critical behavior of the additive martingale  $W_{\beta, N}$  as  $\beta \uparrow \beta_c$ , and he showed that conditioned on survival

$$\frac{W_{\beta, \infty}}{1 - \beta} \rightarrow 2D_{\beta_c, \infty} \quad \text{in probability, } N \rightarrow \infty. \quad (1.1.15)$$

Subsequently, with Madaule's result in mind, Pain [151] improved Alberts and Ortgiese's result to  $\beta_N = \beta_c \pm 1/\alpha_N$  for any  $\alpha_N \rightarrow \infty$ , and the result can be found in Theorem 1.1 in Pain's paper.

### Structures of the limiting Gibbs measure

The following questions concerning the Gibbs measure are particularly interested in the literature.

1. What is the limit of the Gibbs measure in the weak topology of measure?
2. What does the sampled trajectory by the Gibbs measure look like?
3. If we sample two particles from two Gibbs measures independently, where do the two lineages splits?

To present these results, we first introduce some notation. Let

$$\mathcal{T} = \bigcup_{n=0}^{\infty} \mathbb{N}^n, \quad \partial\mathcal{T} = \mathbb{N}^{\mathbb{N}} \quad \text{and} \quad \bar{\mathcal{T}} = \mathcal{T} \cup \partial\mathcal{T}. \quad (1.1.16)$$

For any  $u \in \mathcal{T}$  which is a finite sequence of integers, define the set (see Figure 1.4)

$$\bar{\mathcal{T}}^u = \{v \in \bar{\mathcal{T}} : v \wedge u = u\}. \quad (1.1.17)$$

Denote by  $\mathcal{D}([0, 1])$  be the set of real càdlàg functions defined on  $[0, 1]$  and  $\mathcal{C}_b(\mathcal{D}([0, 1]))$  the set of real, continuous and bounded function defined on  $\mathcal{D}([0, 1])$ . Define the rescaled trajectory of  $u$ 's lineage by

$$\mathbf{X}_u := \left( \frac{1}{\sigma\sqrt{N}} X_{u[tN]}, t \in [0, 1] \right). \quad (1.1.18)$$

For  $\beta \in \mathbb{R}$  and  $F \in \mathcal{C}_b(\mathcal{D}([0, 1]))$ , define

$$\mu_{\beta, N}(F) := \frac{1}{Z_{\beta, N}} \sum_{|u|=N} e^{\beta X_u} F(\mathbf{X}_u) \quad (1.1.19)$$

which encodes the information of the sampled trajectory by the Gibbs measure.

**Definition 1.1.2.** For any Borel set  $A \subseteq [0, 1]$ , the overlap distribution is defined by

$$\omega_{\beta,N}(A) := \mu_{\beta,N}^{\otimes 2} \left( \frac{|u \wedge w|}{N} \in A \right).$$

For any  $u \in \mathcal{T}$ , the Boltzmann–Gibbs measure is the unique measure where its value on the set  $\bar{\mathcal{T}}^u$  is given by

$$\mu_{\beta,\infty}(\bar{\mathcal{T}}^u) := \frac{W_{\beta,\infty}^u}{W_{\beta,\infty}} e^{\beta X_u - \varphi(\beta)|u|},$$

where  $W_{\beta,\infty}^u$  is the almost sure limit of

$$W_{\beta,N}^u = \sum_{|w|=N} e^{\beta(X_{uw} - X_u) - \varphi(\beta)N}. \quad (1.1.20)$$

In the subcritical regime  $\beta < \beta_c$ , Chauvin and Rouault mentioned that one of the consequences of (1.1.11) is that almost surely, the Gibbs measure  $\mu_{\beta,N}$  converges weakly to the Boltzmann–Gibbs measure. Combining with Theorem 6 in the paper [128] of Liu and Rouault which said that for  $\beta < \beta_c$ , the Boltzmann–Gibbs measure is non-atomic on  $\partial\mathcal{T}$ , Chauvin and Rouault showed that the overlap distribution  $\omega_{\beta,N}$  converges almost surely to the Dirac measure  $\delta_0$ . In Appendix C of his paper [151], Pain showed that for all  $F \in \mathcal{C}_b(\mathcal{D}([0, 1]))$ , conditioned on survival,

$$\frac{1}{Z_{\beta,N}} \sum_{|u|=N} e^{\beta X_u} F \left( \frac{tN\varphi'(\beta) - X_{u[tn]}}{\sigma_\beta \sqrt{N}}, t \in [0, 1] \right) \rightarrow \mathbb{E}[F(B)] \quad \text{in probability, } N \rightarrow \infty, \quad (1.1.21)$$

where  $B$  is the Brownian motion.

For the critical case  $\beta = \beta_c$ , Theorem 1.2 of Madaule’s paper [131] showed that for all  $F \in \mathcal{C}_b(\mathcal{D}([0, 1]))$ , conditioned on survival,

$$\mu_{\beta_c,N}(F) \rightarrow \mathbb{E}[F(\mathcal{M})] \quad \text{in probability, } N \rightarrow \infty, \quad (1.1.22)$$

where  $\mathcal{M}$  is the Brownian meander of length 1. He also showed that (1.1.22) implies that conditioned on survival, the overlap distribution  $\omega_{\beta_c,N}$  converges weakly to the Dirac measure  $\delta_0$  in probability. Pain [151] showed the same result for the near critical regime: he showed that for any  $\beta_N \rightarrow \beta_c$ , conditioned on survival, the overlap distribution converges to  $\delta_0$  in probability.

For the supercritical regime  $\beta > \beta_c$ , Chen, Madaule and Mallein [79] showed that for all *uniformly continuous*  $F \in \mathcal{C}_b(\mathcal{D}([0, 1]))$ ,

$$\mu_{\beta,N}(F) \rightarrow \sum_{n=1}^{\infty} p_n F(\epsilon_n). \quad (1.1.23)$$

where  $(\epsilon_n)_{n \geq 1}$  is a sequence of i.i.d. normalized Brownian excursion and  $(p_n)_{n \geq 1}$  is the atom of the Poisson–Dirichlet distribution with parameter  $(\beta_c/\beta, 0)$ . The

convergence in (1.1.23) is expected to be true for all  $F \in \mathcal{C}_b(\mathcal{D}([0, 1]))$ . On the other hand, for  $\beta = \infty$ , Chen showed in [78] that for all  $F \in \mathcal{C}_b(\mathcal{D}([0, 1]))$ , we have

$$\mathbb{E}^*[\mu_{\infty, N}(F)] \rightarrow \mathbb{E}[F(\mathfrak{e})], \quad N \rightarrow \infty, \quad (1.1.24)$$

where  $\mathfrak{e}$  denotes the normalized Brownian excursion.

Since (1.1.23) was only proven for uniformly continuous  $F \in \mathcal{C}_b(\mathcal{D}([0, 1]))$ , the convergence in law of the overlap distribution  $\omega_{\beta, N}$  cannot be derived from (1.1.23). Nevertheless, Mallein [133] showed that conditioned on survival, the overlap distribution converges in law to

$$(1 - \pi_\beta)\delta_0 + \pi_\beta\delta_1, \quad (1.1.25)$$

where  $\pi_\beta$  is the random variable defined by

$$\pi_\beta = \sum_{n=1}^{\infty} p_n^2, \quad (1.1.26)$$

with  $(p_n)_{n \geq 1}$  being the Poisson–Dirichlet distribution with parameter  $(\beta_c/\beta, 0)$  defined in Definition 5.1.3. Previously, Jagannath proved in [116] the same result for binary branching random walk with standard Gaussian increments with an alternative approach. Mallein’s result confirms the conjecture of Derrida and Spohn in [91], where they conjectured that the overlap distribution of branching random walks exhibits the 1-RSB.

### Overlap distribution between two different temperatures

There is also an interest to study the overlap distribution between two different temperatures for the hierarchical models, which originated from the study of chaos problem in spin glass models. A thorough survey on this topic was made by Rizzo in [162]. Some relevant results are also reviewed in Section 1.3.3 of this thesis.

For the discrete GFF, Pain and Zindy showed in [152] that mean of the overlap distribution of the discrete GFF is strictly smaller than the one for the random energy model. In a very recent article, Bonnefont showed in [51] the same result for the branching Brownian motion. It is not yet clear whether this result can be generalized to the branching random walks.

#### 1.1.4 Other related models

In this section, we briefly mention a few models that are related to the branching random walk. It is important to notice that the list is far from being comprehensive, as similar models have been reinvented and rediscovered by several times in history for various motivations. We refer to Biggins [43] for a list of these different traditions that are linked to branching random walks.

#### Branching Brownian motion

Branching Brownian motions (BBM) are the continuum counterpart of branching random walks which can be described informally as follows. At  $t = 0$ , a single

particle starts at the origin and moves as a one-dimensional Brownian motion until an exponential clock with parameter 1 rings. Then the particle branches into two children that move independently as one-dimensional Brownian motions and branch according to two independent exponential clocks with parameters 1. The definition then goes by induction. Let  $N(t)$  be the number of particles at time  $t$ , and let  $X_1(t), \dots, X_{N(t)}(t)$  be the positions of the particles at time  $t$ . Consider the following FKPP equation with the Heaviside initial condition.

$$\begin{cases} \partial_t u(t, x) = \frac{1}{2} \partial_x x u(t, x) + u(t, x)^2 - u(t, x), & (t, x) \in \mathbb{R}_+ \times \mathbb{R}, \\ u(0, x) = \mathbf{1}\{x \geq 0\}, & x \in \mathbb{R}. \end{cases} \quad (1.1.27)$$

The FKPP equation was studied by Fisher [100] and Kolmogorov, Petrovskii and Piskunov [124], and McKean [140] observed that the distribution function of the maximum of the BBM solves the FKPP equation (1.1.27), i.e.,

$$u(t, x) = \mathbb{P} \left( \max_{1 \leq i \leq N(t)} X_i(t) \leq x \right). \quad (1.1.28)$$

By leveraging on McKean's observation, Bramson [57, 56] proved that there exists a constant  $C > 0$  such that the maximum of the BBM satisfies the following asymptotic expansion

$$\max_{1 \leq i \leq N(t)} X_i(t) = \sqrt{2}t - \frac{3}{2\sqrt{2}} \log t + C + o(1), \quad t \rightarrow \infty. \quad (1.1.29)$$

In Section 1.1 of his lecture note [168], Shi briefly explained the traveling wave solution of (1.1.27) and its probabilistic interpretation. Berestycki's lecture note [40] gives a more detailed treatment between the BBM and the FKPP equation.

### Multiplicative cascade

The introduction of multiplicative cascades by Mandelbrot [135, 136, 137] was motivated by the study of intermittent turbulence. These cascades serve as an equivalent formulation to branching random walks, which we illustrate with a one-dimensional example. A multiplicative cascade on the interval  $[0, 1]$  is a sequence of measures  $(\mu_N)_{N \geq 0}$  defined recursively: we start with  $\mu_0$  as the Lebesgue measure. For all  $N \geq 1$ , for every dyadic interval  $I$  of length  $2^{-N}$ , the restriction of  $\mu_N$  to  $I$  is defined by  $Y_I \mu_{N-1}$ , where the random variables  $(Y_I)$ , indexed by dyadic interval  $I$ , are i.i.d. copies of a positive random variable  $Y$  with mean 1. This is equivalent to the exponential of a branching random walk, where each individual gives birth to 2 offspring with independent increments distributed as  $\log Y$ . The multiplicative cascades have been the subject of a separate literature. For more information on this topic, we refer the reader to the survey [25] by Barral and Mandelbrot.

### Two-dimensional discrete Gaussian free field

Let  $D_N$  be a subset of  $N^{-1}\mathbb{Z}^2$ . Denote by  $(z_n)_{n \geq 0}$  the simple random walk in  $N^{-1}\mathbb{Z}^2$  starting at  $x \in N^{-1}\mathbb{Z}^2$  with law denoted by  $\mathbf{P}_x$  and  $\mathbf{E}_x$  the corresponding

expectation. Let  $\tau := \{n \geq 0 : X_n \in D_N\}$  be the first exit time of the random walk from  $D_N$ . The Green's function is a function  $G_N$  defined on  $D_N \times D_N$  by

$$G_N(x, y) := \begin{cases} \mathbf{E}_x \left[ \sum_{k=0}^{\tau-1} \mathbf{1}\{z_k = y\} \right], & x, y \in D_N \\ 0, & \text{otherwise.} \end{cases} \quad (1.1.30)$$

The two-dimensional discrete Gaussian free field (GFF) on  $D_N$  is a centered Gaussian process  $(\Gamma_x)_{x \in D_N}$  with covariance function given by the Green's function  $G_N$  defined in (1.1.30). The two-dimensional GFF belongs to the family of log-correlated Gaussian field because for  $N$  sufficiently large and for  $x, y \in D_N$  away from the boundary, one has  $G_N(x, y) \approx \frac{2}{\pi} \log \frac{N}{|x-y|}$ . The two-dimensional GFF is known to satisfy a domain Markov property, which means that on a subset  $D'_N \subset D_N$ , the conditional law of the field  $(\Gamma_x)_{x \in D'_N}$  given  $(\Gamma_x)_{x \in D_N \setminus D'_N}$  is still a GFF with boundary condition on the  $\partial D'_N$  (in fact, this property is true in any dimension). The domain Markov property allows us to compare the GFF with the branching random walk via the following construction: let  $D_N$  be a cube. Decompose  $D_N$  into 4 smaller cubes separated by a cross-shaped domain. Then one can reconstruct the original GFF by the data on the smaller cubes approximately. This construction has led to a huge success of understanding the two-dimensional GFF, and we refer the readers to the thesis [150] of Pain that provided a comprehensive survey of the recent advances in this line of research.

## 1.2 Derrida's random energy model and its generalizations

In the 1980s, Derrida and his collaborators introduced a series of models that simplify mean-field spin glasses but still capture some essential properties of the mean-field models. These models are sometime referred to as Derrida's random energy model and its generalization or spin glasses in the Derrida class. The goal of this section is to collect the known mathematical results of these models, and we also provide the relevant historical backgrounds of these models. We focus on the random energy model (REM), the generalized random energy model (GREM), and the continuous random energy model (CREM).

### 1.2.1 Random energy model

The Random Energy Model (REM) is a simple spin glass model, first introduced by Derrida in 1980 in a series of two papers [86, 90]. The REM can be interpreted as a formal limit of the so-called pure  $p$ -spin model, with the SK model corresponding to  $p = 2$ . Mathematically, the REM, denoted as  $(X_i)_{i=1}^{2^N}$ , consists of  $2^N$  i.i.d. random variables with Gaussian distribution  $\mathcal{N}(0, N)$ . Despite its simplicity, the phase transition that the REM embodies qualitatively some interesting properties of spin glasses. This can be observed from the free energy and overlap of the REM.

Firstly, the REM free energy  $f_\beta$  exhibits a phase transition at  $\beta_c = \sqrt{2 \log 2}$ .

$$f_\beta := \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} [\log Z_{\beta, N}] = \begin{cases} \log 2 + \frac{\beta^2}{2}, & \beta \in [0, \sqrt{2 \log 2}], \\ \sqrt{2 \log 2} \beta, & \beta \geq \sqrt{2 \log 2}, \end{cases} \quad (1.2.1)$$

where

$$Z_{\beta, N} := \sum_{i=1}^{2^N} \exp(\beta X_i) \quad (1.2.2)$$

is called the partition function of the REM. Note that for  $\beta \geq \beta_c$ , the quantity  $f_\beta/\beta$  remains constant. This phenomenon is referred to in the physics literature as the freezing phase transition in the low temperature regime, which is a signature of spin glass models.

Secondly, the REM overlap experiences the so-called 1-RSB phase transition at  $\beta_c = \sqrt{2 \log 2}$ , which we will describe below. The Gibbs measure of the REM is defined as

$$\mu_{\beta, N}(i) = \frac{e^{\beta X_i}}{Z_{\beta, N}}, \quad \forall i = 1, \dots, 2^N. \quad (1.2.3)$$

The overlap of the REM is defined as  $q_N(i, j) := \mathbf{1}_{i=j}$ . The CDF the limiting overlap distribution of the REM is a function  $\alpha_\beta^{\text{REM}} : [0, 1] \rightarrow [0, 1]$  defined as

$$\alpha_\beta^{\text{REM}}(t) = \lim_{N \rightarrow \infty} \mathbb{E} \left[ \sum_{i=1}^{2^N} \sum_{j=1}^{2^N} \mu_{\beta, N}(i) \mu_{\beta, N}(j) \mathbf{1}_{q_N(i, j) \leq t} \right].$$

The function  $\alpha_\beta^{\text{REM}}$  exhibits the following phase transition: for all  $\beta < \beta_c$ , we have  $\alpha_\beta^{\text{REM}} \equiv 1$ . For all  $\beta \geq \beta_c$ , we have

$$\alpha_\beta^{\text{REM}}(t) = \begin{cases} \frac{\beta_c}{\beta}, & t \in [0, 1), \\ 1, & t = 1. \end{cases} \quad (1.2.4)$$

Intuitively, it means that the mass of the overlap distribution goes from concentrating at 0 to concentrating two points at 0 and 1. This means that the REM is a 1-RSB type model in the physicists' terminology.

## 1.2.2 Generalized random energy model

While the REM demonstrates several interesting properties, it does not replicate the ultrametric structure predicted by Parisi's ansatz in [158, 157]. Consequently, Derrida proposed the generalized random energy model (GREM) in [85] and studied subsequently the model with Gardner in [87, 88].

Formally, the GREM is defined as follows. Let  $(\alpha_i)_{i=1}^n$  be a sequence of non-negative numbers such that  $\prod_{i=1}^n \alpha_i = 2$ , and let  $\mathbb{T}_{n, N}$  be a finite tree of depth

$n$  rooted at  $\emptyset$  such that each vertex with depth  $k \in \llbracket 0, n-1 \rrbracket$  has  $\alpha_{k+1}^N$  vertices. Suppose that  $A : [0, 1] \rightarrow [0, 1]$  is a step function defined as

$$A(t) = a_1 + \cdots + a_k, \quad \forall t \in [\log_2(\alpha_1 \cdots \alpha_{k-1}), \log_2(\alpha_1 \cdots \alpha_k)],$$

where  $(a_i)_{i=1}^k$  is sequence of non-negative numbers such that  $a_1 + \cdots + a_k = 1$ . Then the GREM with  $n$  hierarchy is a centered Gaussian process with the covariance function

$$\mathbb{E}[X_u X_w] = N \cdot A\left(\frac{|u \wedge w|}{N}\right), \quad \forall u, w \in \mathbb{T}_{n,N}. \quad (1.2.5)$$

See Figure 1.5 for a schematic illustration of the model.

Many fundamental properties of the GREM are known. For example, Capocaccia, Cassandro and Picco computed rigorously in [59] an explicit formula of the GREM free energy in the 1980s. In the early 2000s, Bovier and Kurkova [53] made a fundamental contribution of understanding this model. We only mention part of their discoveries that are relevant to the thesis.

1. They characterized the fluctuation of the GREM free energy.
2. They characterized the limit of the GREM Gibbs measure as a Ruelle probability cascade, confirming Ruelle's observation in [165].
3. They provided a second way to characterize the Gibbs measure at the limit. Namely, they showed that the limit of the GREM Gibbs measure is characterized by the limiting overlap distribution and the so-called Ghirlanda–Guerra identities.

The Ruelle probability cascade is a hierarchical Poisson process introduced by Ruelle in [165] and then been further studied by Bolthausen and Sznitman in [49]. The definition of this object can be found in Definition 5.1.3.

One of the key properties of models in the Derrida class is that the Gibbs measure of these models is identical, up to a certain degree, to the limit of the SK Gibbs measure. This property was already predicted by several physicists. To name a few, Derrida and Toulouse argued in [89] that the Gibbs weights of the REM distribute identically in the thermodynamic limit as the Gibbs weights of the pure states in the SK model, and they showed the pure states of the SK model are organized in an ultrametric structure, and this result was previous argued by Mézard et al. in [142] using the replica method. Combining these evidences with Ruelle's observation, one should conjecture that one should be able to describe the limiting SK model as a Ruelle probability cascade. While nowadays this has been proven rigorously, the conjecture was only been settled recently by the seminal paper [155] of Panchenko in 2013. We refer to the book [156] of Panchenko for a complete version of this story.

### 1.2.3 Continuous random energy model

In companion of [53], Bovier and Kurkova introduced and studied in [54] and [55] the continuous random energy model (CREM). Formally, the CREM is defined as

follows. Let  $A : [0, 1] \rightarrow [0, 1]$  be a non-decreasing function with  $A(0) = 0$  and  $A(1) = 1$ . The CREM is a centered Gaussian process indexed by a binary tree with the covariance function

$$\mathbb{E}[X_u X_w] = NA \left( \frac{|u \wedge w|}{N} \right), \quad u, w \in \mathbb{T}_N.$$

Compared to the GREM, the number of hierarchies of the CREM grows is proportional to  $N$  which grows to infinity, while the number of hierarchies of the GREM stays finite as  $N \rightarrow \infty$ . See Figure 1.6 for an illustration of the covariance function.

We summarize the discoveries that Bovier and Kurkova have made in [54, 55].

1. They provided an explicit formula of the CREM free energy: let  $\hat{a}$  be the right derivative of the concave hull  $\hat{A}$  of  $A$  (see Figure 1.6). They showed that the CREM free energy  $F_\beta$  admits the following expression.

$$F_\beta := \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\log Z_{\beta, N}] = \int_0^1 f(\beta \sqrt{\hat{a}(s)}) ds, \quad (3.1.4)$$

where  $Z_{\beta, N}$  is the partition function defined as

$$Z_{\beta, N} := \sum_{|u|=N} e^{\beta X_u},$$

and the function  $f$  is the REM free energy defined in (1.2.1). However, they did not have results on the fluctuation of the CREM free energy.

2. They showed that the limit of the CREM Gibbs measure is characterized by the limiting overlap distribution and the so-called Ghirlanda–Guerra identities.
3. While we do not need the following result, we want to mention that in [55], Bovier and Kurkova further expressed the structure of the limiting CREM Gibbs measure via the genealogy of the so-called Neveu’s continuous state branching process (CSBP) with an appropriate time change. It was proven by Bertoin and Le Gall [42] that the genealogy of Neveu’s CSBP is the Bolthausen–Sznitman coalescent in [49].

#### 1.2.4 Maximum asymptotics of time-inhomogeneous branching random walks

The GREM and the CREM belong to a larger class of models, called the time-inhomogeneous branching random walks where the law of increments is no longer homogeneous over the depth on the genealogical tree. Recently, Mallein studied in [134] the asymptotic of the maximum for the time-inhomogeneous branching random walk. He proved that the first term of order  $N$  and the coefficient is given by an optimization problem, whereas the second term is of order  $N^{1/3}$ , in contrast with the homogeneous case. Maillard and Zeitouni [132] also drew similar conclusions for the branching Brownian motion with inhomogeneous variance, which is the continuum counterpart of the time-inhomogeneous branching random walk.

### 1.3 Mean-field spin glasses

While the mathematical tools to approach Derrida’s models are different the ones for the mean-field models, the introduction of Derrida’s models originates in the study of mean-field spin glass models. Therefore, we provide the historical context of the development of mean-field spin glass models.

#### 1.3.1 Edwards–Anderson model

The term “spin glass” was probably first coined by Anderson in 1970 [14] where he attempted to explain theoretically the magnetism of Cu-Mn alloys<sup>1</sup>. Later in 1975, Edwards and Anderson formulated the Edwards–Anderson model in their influential paper [96]. A formulation of their model can be described as follows: given a finite graph  $G = (V, E)$ <sup>2</sup>, let  $J = (J_e)_{e \in E}$  be a collection of i.i.d. random variables indexed by the edges. The Hamiltonian of the Edwards–Anderson model is a random function  $H(\sigma) : \{-1, +1\}^V \rightarrow \mathbb{R}$  defined as

$$H(\sigma) = \sum_{\{i,j\} \in E} J_{ij} \sigma_i \sigma_j.$$

Despite its seemingly simple formulation, the EA model continues to pose challenges both in mathematics and physics. The open problems and possible ways to attack these challenges can be found in [148] and the references therein. Very recently, Chatterjee [62] showed that the EA model has certain glassy behaviors at zero temperature, but the whole picture regarding the conjectured glassy phase of the EA model still widely open.

#### 1.3.2 Sherrington–Kirkpatrick model

In contrast to the EA model, the mean-field spin glass models – which are models with complete underlying graph  $G$  – has witnessed several breakthroughs in the last decades. The most famous mean-field model might be the Sherrington–Kirkpatrick (SK) model [167] introduced in 1975. In the following, we will mainly focus on the SK model. The books by Talagrand [179, 180] and Panchenko [156] are excellent sources of the SK model.

Fixing  $N$  an positive integer, the Hamiltonian of the SK model  $H(\sigma) : \{-1, +1\}^N \rightarrow \mathbb{R}$  is defined as

$$H_N(\sigma) = \frac{1}{\sqrt{N}} \sum_{\{i,j=1\}}^N J_{ij} \sigma_i \sigma_j,$$

---

<sup>1</sup>However, we have to point out that there are earlier prototypes before Anderson. For example, Marshall proposed the Ruderman–Kittel–Kasuya–Yosida (RKKY) interaction to explain some exotic properties of the Cu-Mn and other alloys observed by experimental physicists. See Marshall [139] and the references therein.

<sup>2</sup>Actually, Edwards and Anderson considered the (inhomogeneous) Erdős–Rényi graph instead of a deterministic one.

where  $\{J_{ij}\}_{i=1,j=1}^N$  is a collection of i.i.d. standard Gaussian random variables. Therefore, the Hamiltonian  $(H_N(\sigma))_{\sigma \in \{-1,+1\}^N}$  forms a centered Gaussian process with covariance function

$$\mathbb{E}[H_N(\sigma^1)H_N(\sigma^2)] = N\xi(R_{1,2}), \quad \sigma^1, \sigma^2 \in \{-1,+1\}^N$$

where  $\xi(x) = x^2$  and  $R_{1,2} = \frac{1}{N} \sum_{i=1}^N \sigma_i^1 \sigma_i^2$ . The quantity  $R_{1,2}$  is called the overlap of two configurations  $\sigma^1$  and  $\sigma^2$ . The free energy of the SK model is defined to be

$$F_\beta = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\log Z_{\beta,N}],$$

where  $Z_{\beta,N}$  is the partition function of the SK model which equals

$$Z_{\beta,N} = \sum_{\sigma \in \{-1,+1\}^N} \exp(\beta H_N(\sigma)).$$

Finally, the Gibbs measure of the SK model is defined as

$$G_{\beta,N}(\sigma) = \frac{1}{Z_{\beta,N}} \exp(\beta H_N(\sigma)), \quad \sigma \in \{-1,+1\}^N.$$

The free energy  $F_\beta$  is a central theme of the study of mean-field spin glass. First of all, the free energy is naturally connected to the asymptotic of the maximum  $\mathbb{E}[\max_\sigma H_N(\sigma)]$  by taking  $\beta \rightarrow \infty$ . Secondly, it is known that the free energy encodes the geometry of the Gibbs measure. To illustrate this, we introduce the notion of overlap distribution which is the law of  $R_{1,2} = \frac{1}{N} \sum_{i=1}^N \sigma_i^1 \sigma_i^2$  of two spin configurations  $\sigma^1, \sigma^2 \in \{-1,+1\}^N$  sampled independently from the Gibbs measure  $G_{\beta,N}$ . It is known that when  $\beta$  is small, the overlap concentrates on one single value, while it can take several values when  $\beta$  is large enough. If the limit of the overlap distribution is supported on  $k$  values, we say the model exhibits  $k+1$  replica symmetry breaking ( $k+1$ -RSB), and if the support of the limiting overlap distribution contains an interval, we say that the model exhibits full replica symmetry breaking (full-RSB).

**Assertions from physicists.** In the same paper [167] where the SK model was introduced, Sherrington and Kirkpatrick proposed a formula of the free energy  $F_\beta$  based on the replica method, while they also observed that their replica symmetric solution implied the entropy became negative for  $\beta$  large, which is un-physical. Later, in two consecutive papers [158, 157] in 1979 and 1980, Parisi proposed a replica symmetric breaking solution also via the replica approach, which turned out to be the correct one. In Section V of scientific background provided in [182], the Nobel Prize Committee 2021 gave an overview of Parisi's studies of spin glass and its applications and implications in other disordered complex systems.

**Mathematical results for the SK model.** For clarity, we summarize Parisi's prediction for the SK model mathematically in the following conjecture, which was formulated in [21].

**Conjecture 1.3.1.** *Parisi predicted for the SK model that*

1. *The free energy  $F_\beta$  is given by the variational principle*

$$F_\beta = \inf\{\mathcal{P}_\beta(\alpha) : \alpha \text{ is a CDF on } [0, 1]\}, \quad a.s., \quad (1.3.1)$$

*where  $\mathcal{P}_\alpha$  is known as the Parisi functional. The measure corresponds to the minimizer (1.3.1) is called the Parisi measure.*

2. *The Gibbs measure is asymptotically ultrametric [142].*
3. *For  $\beta$  sufficiently large, the SK model exhibits full-RSB.*

The first two points of Conjecture 1.3.1 were proven. The first point in Conjecture 1.3.1 was verified for the SK model by Talagrand [181] in 2006 following the work of Guerra [109]. The second point in Conjecture 1.3.1 was established for the SK model by Panchenko [155] in 2012. Later Panchenko extended in [154] the Parisi formula for a larger class of spin glass models for which the ultrametric property holds. While the third point of Conjecture 1.3.1 is not yet being established, a very recently paper by Auffinger, Chen and Zeng [22] showed that the SK model is of  $\infty$ -RSB at zero temperature ( $\beta = \infty$ ), which is a first step toward proving Point 3 in Conjecture 1.3.1.

### 1.3.3 Chaos in mean-field spin glasses

As mentioned in Section 1.3.1, while the study of mean-field spin glasses have received huge progresses, EA models remains challenging to physicists and mathematicians up to today. In the late 1980s, Fisher and Huse [99] and Bray and Moore [58] proposed a framework to study the glassy nature of EA models through establishing the occurrence of chaos. Chaos phenomena in the context of disordered systems mean that the energy landscape is sensitive to small perturbations, and they are studies mainly from the two perspectives: a) the temperature chaos, where the energy landscape is sensitive to a small perturbation of the temperature. b) the disorder chaos, where the landscape energy is sensitive to a small perturbation of the disorder.

**Chaos in EA models.** Chatterjee showed in [61, 64] that the edge overlap of the EA model does not exhibit the disorder chaos in the sense that the bond overlap does not drop to 0 under a small disorder perturbation. His result was later generalized by Arguin and Hanson in [16] for non-Gaussian settings. However, in a very recent article, Chatterjee [62], among other other things, proved that the bond edge drops sharply to a value strictly less than 1. Moreover, he showed that, under certain conditions, the site overlap drops to 0 with high probability under small disorder perturbation.

**Disorder chaos in mean-field models.** Back to mean-field models. At positive temperature, Chatterjee [61] proved that the SK model exhibits disorder chaos. Subsequently, Chen [68] show that the SK model also exhibits disorder chaos even

with the presence of external field, and he extended these results in [69] and another joint work [71] with Panchenko to some other mixed  $p$ -spin models. In [70], Chen, Handschy and Lerman established disorder chaos for even  $p$ -spin models, based on the framework developed by Chen in [69]. Very recently, Eldan [97] proposed a simple approach to establish disorder chaos for mixed  $p$ -spin models.

For the spherical case, Chen, Hsieh, Hwang and Sheu [77] showed that the spherical mixed even  $p$ -spin models exhibit disorder chaos. In [75], Chen and Sen studied the ground state energy of mixed even  $p$ -spin models. In particular, they proved without external fields, the ground state energy of mixed even  $p$ -spin models exhibit disorder chaos in the sense that the overlap between ground states of two independent perturbed mixed even  $p$ -spin models converges to 0 in probability.

**Temperature chaos in mean-field models.** The first mathematically rigorous results of temperature chaos for mean-field models were established by Chen [67] and Chen and Panchenko [71]. Subsequently, Panchenko [153] extended these results and proved the occurrence of temperature chaos for generic mixed even  $p$ -spin models.

For the spherical case, temperature chaos was proven to be absent for pure spherical glasses at sufficiently low temperature by Subag [177]. In [72], Chen and Panchenko presented some examples of spherical mixed  $p$ -spin models where disorder chaos occurs. Finally, in [17], Arous, Subag and Zeitouni strengthened Chen and Panchenko's result and established temperature chaos for mixed spherical  $p$ -spin models at low temperature.

### 1.3.4 Thouless–Anderson–Palmer approach

In a quick succession of Sherrington and Kirkpatrick's work [167], Thouless, Anderson and Palmer [183] proposed another method to calculate the free energy and to avoid the negative entropy problem arise from [167], which is known as the TAP approach nowadays.

To present their result, let us introduce the following relevant notions. Given  $m \in [-1, 1]^N$ , the so-called TAP correction is defined as

$$\widetilde{\text{TAP}}_{\beta}(m) = -\frac{1}{N} \sum_{i=1}^N \left( \frac{1+m_i}{2} \log \frac{1+m_i}{2} + \frac{1-m_i}{2} \log \frac{1-m_i}{2} \right) + \frac{\beta^2}{2} \left( 1 - \frac{\|m\|_2^2}{N} \right)^2. \quad (1.3.2)$$

Thouless, Anderson and Palmer suggested that the SK free energy satisfies asymptotically the following

$$F_N(\beta) \approx \max_{m \in [-1, 1]^N} \left\{ \frac{\beta}{N} H_N(m) + \widetilde{\text{TAP}}_{\beta}(\mu_m) \right\}. \quad (1.3.3)$$

Therefore, the SK free energy can be calculated by finding the critical points of the expression at the right-handed side of (1.3.3). By differentiation, one can show

that any critical point  $m = (m_1, \dots, m_N) \in [-1, 1]^N$  satisfies

$$m_i = \tanh\left(\left(\beta \nabla H_N(m)\right)_i - 2\beta^2 m_i \left(1 - \frac{\|m\|_2^2}{N}\right)\right), \quad i \in \llbracket 1, N \rrbracket. \quad (1.3.4)$$

The high-dimensional system of equations in (1.3.4) is known to be the TAP equations, and there are currently two approaches to solve (1.3.4). The first approach is to take

$$m = (\langle \sigma_1 \rangle, \dots, \langle \sigma_N \rangle), \quad \text{where} \quad \langle \sigma_i \rangle = \sum_{|\sigma|=N} G_N(\sigma) \sigma_i, \quad i \in \llbracket 1, N \rrbracket.$$

This approach was done by Talagrand (Theorem 1.7.7 in [179]) and Chatterjee [63] which holds for sufficiently high temperature, or for  $\beta$  sufficiently small. The second approach was done by Bolthausen [50] where he introduced an iteration scheme that solves (1.3.4) asymptotically in the high temperature regime. The high temperature regime<sup>3</sup> can be characterized by the following replica symmetric condition, i.e., the high temperature regime consists of any  $\beta > 0$  such that

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[ \sum_{|\sigma^1|=N, |\sigma^2|=N} G_{\beta, N}(\sigma^1) G_{\beta, N}(\sigma^2) (R_{1,2} - q)^2 \right] = 0,$$

where  $q$  is the unique constant satisfying

$$q = \mathbb{E} \left[ \tanh^2(\beta z \sqrt{q}) \right], \quad z \sim \mathcal{N}(0, 1).$$

Bolthausen's scheme is now known as an example of the approximate message passing (AMP) type algorithms, which first appeared in the context of Bayesian inference (see Kabashima [117]) and in the context of compressed sensing (see Donoho, Maleki and Montanari [94]). The AMP type algorithms have been widely applied to solve optimization problem for high-dimensional random functions inspired from the spin glasses literature, and we provide a survey of this direction in Section 1.4.

Recently, Auffinger and Jagannath [23] showed that, for generic mixed  $p$ -spin glasses including the SK model, under a technical assumption that is conjectured to be true for a wide range of spin glass models, the TAP equation holds asymptotically in distribution.

Subag recently developed in [177] a generalized TAP approach to understand the structure of the Gibbs measure of the so-called pure spherical spin glass. This approach establishes a description of the free energy similar to that of (1.3.3). Subag and his collaborators have since extended this approach to several mean-field spin glasses, including the SK model, in a series of subsequent works [172, 73, 74, 176, 173, 174]. As explained in the introduction of [73], by choosing an

<sup>3</sup>In fact, Almeida and Thouless [12] conjectured that the high temperature regime can also be characterized by the so-called AT-line condition. While it has been shown that the replica symmetric condition, the converse direction has not been proven yet. See the introduction of the paper [76] by Chen and Tang for further discussions.

alternative TAP correction  $\text{TAP}_\beta(m)$ , for any  $q \in [0, 1)$  belongs to the support of the Parisi measure which appear in [1.3.1](#), one has for large  $N$  that

$$F_N(\beta) \approx \max_{m \in \mathbb{R}^N: \|m\|_2^2 = Nq} \left\{ \frac{\beta}{N} H_N(m) + \text{TAP}_\beta(\mu_m) \right\}. \quad (1.3.5)$$

The maximizers of the expression in the right-handed side of [\(1.3.5\)](#) carry non-trivial information of the structure of the Gibbs measure, which we explain as follows. Let

$$E_\star := \lim_{N \rightarrow \infty} \frac{1}{N} \max_{\sigma \in \{-1, +1\}^N} H_N(\sigma) = \lim_{\beta \rightarrow \infty} \frac{F_\beta}{\beta}. \quad (1.3.6)$$

Auffinger and Chen showed in [\[19\]](#) that for any  $E \in (0, E_\star)$ , if we choose  $\beta$  such that  $E = F'_\beta$ , then the cardinality of the corresponding approximate level set can be expressed via the free energy as follows.

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \# \left\{ \sigma \in \{-1, +1\}^N : \left| \frac{1}{N} H_N(\sigma) - E \right| < \varepsilon \right\} = F_\beta - \beta F'_\beta + O(\varepsilon). \quad (1.3.7)$$

Moreover, the Gibbs measure concentrates on the same set in the following sense.

$$\lim_{N \rightarrow \infty} \sum_{\sigma \in \{-1, +1\}^N} G_{\beta, N}(\sigma) \left\{ \sigma \in \{-1, +1\}^N : \left| \frac{1}{N} H_N(\sigma) - E \right| < \varepsilon \right\} = 1, \quad \text{almost surely.} \quad (1.3.8)$$

Then, if we choose  $m$  to be a maximizer of the expression in the right-handed side of [\(1.3.5\)](#), then for small  $\varepsilon > 0$ , the narrow band defined as

$$B(m, \varepsilon) := \left\{ \sigma \in \{-1, +1\}^N : \left| \sigma \cdot m - \|m\|_2^2 \right| = |m \cdot (\sigma - m)| < \varepsilon N \right\} \quad (1.3.9)$$

contains a large number of points in the approximate level set. Moreover, with high probability, the property for narrow bands can be applied simultaneously to all maximizers  $m$ .

### Energy landscape and the TAP representation

Recall that  $E_\star$  is defined in [\(1.3.6\)](#) as the maximum of the Hamiltonian, or the ground state energy ignoring the minus sign. Physicists predicted that (see [\[143\]](#)) that there are exponentially many spin configurations  $\sigma$  that reaches the ground state energy approximately, i.e.,  $\frac{1}{N} H_N(\sigma) \approx E_\star$ . Moreover, these maximizers are nearly orthogonal to each other. These predictions were proven rigorously for the mixed even  $p$ -spin models (with SK corresponding to  $p = 2$ ) by Chen, Handschy and Lerman in [\[70\]](#). Their results improved previous results by Chatterjee in [\[61\]](#) and result by Ding, Eldan and Zhai in [\[93\]](#), where the former established a logarithmic size of the number of the orthogonal peaks and the latter proved a polynomial size. On the other hand, for the spherical spin glasses, similar results on the highest critical points are known for the pure spin glasses [\[18, 82, 83, 175, 178\]](#), and some mixed spin glasses that are close to being pure [\[18, 17\]](#).

The TAP variational principle provided an intriguing description of the energy landscape at  $\beta = \infty$ . Namely, let  $\text{TAP}_\infty$  be a functional such that a large amount of maximizers that solves the following variational principle

$$E_\star \approx \max_{m \in \mathbb{R}^N: \|m\|_2^2 = Nq} \left\{ \frac{1}{N} H_N(m) + \text{TAP}_\infty(\mu_m) \right\} \quad (1.3.10)$$

can be arranged in a certain tree structure that is rooted at the origin and has the spin configurations as leaves. Moreover, the functional  $\text{TAP}_\infty$  is constant for  $\mu_\sigma$  on the configuration space, so the leaves are the maximizers of the variational principle (1.3.10). This picture was proven to be corrected for the spherical spin glasses by Subag in [172], and very recently for the  $p$ -spin models (including the SK model) by Chen, Panchenko and Subag in [74]. When full-RSB occurs, i.e., the support of the Parisi measure is equal to the interval  $[0, q_{EA}]$  as  $\beta \rightarrow \infty$ , the normalized radii  $\|m\|/\sqrt{N}$  of the inner vertices of the tree, which are points  $m \in (-1, 1)^N$ , are asymptotically dense in  $[0, 1]$  and the tree is asymptotically continuous in an appropriate sense, and this has deep connection with the recent progress of the optimization problem in mean-field spin glasses. We direct the readers to Section 1.4.2 for further discussion.

### Computing the free energy within the TAP framework

In this section, we primarily focus on the relationship between the TAP approach and the structure of the limiting Gibbs measure. However, it is important to note that the original paper by Thouless, Anderson, and Palmer [183] sparked other research directions as well. In particular, there have been efforts to compute the free energy for mean-field models using Thouless et al.'s original idea, without resorting to Parisi's ansatz. To our knowledge, there are currently only two mean-field spin glass models where this aim has been fully reached. The first one is the spherical SK model achieved by Belius and Kistler in [32], and the second one is the vector spin version of the spherical SK model, where the result was achieved very recently by Belius, Fröber and Ko in [31]. Also in a recent article [30], Belius provided an upper bound for the free energy of the SK model and its generalizations solely by the TAP approach, and he expected that the bound should be tight to leading order at high temperature and should be non-trivial in the presence of an external field.

## 1.4 Algorithmic perspectives of spin glasses

Disordered systems have been studied from the algorithmic points of view since the early days of the field. Their complicated energy landscape serve as an ideal playground for researchers to study the properties that prohibit computational problems to be solved by polynomial-time algorithms. In the context of mean-field spin glasses, a classical algorithm might be the Glauber dynamics, or the Langevin dynamics for the continuum models. One of the original motivations to introduce these dynamics was to observe the glassy phase transition by studying

its fast-to-slow convergence to equilibrium, without passing to Parisi's ansatz. The transition from fast convergence to slow convergence is however itself an intriguing computational problem, and has witnessed several progress recently. In recent years, the AMP algorithm, inspired by Bolthausen's solution to the TAP equation, has been proven to be the optimal algorithm to find the maximum for several mean-field models. Finally, we present the optimization problem for the CREM, which inspired the work in Chapter 2 and Chapter 3.

### 1.4.1 Glauber dynamics on mean-field spin glasses

One of the original motivations for introducing Glauber dynamics (or the Langevin dynamics) to spin glasses was to observe the glass phase transition. This approach focused on studying the dynamics' phase transition in terms of its convergence speed to equilibrium, thus bypassing the need for Parisi's ansatz. Early work on this topic can be found in Sherrington and Kirkpatrick's study [123], or in a more pedagogical presentation in [143]. According to physics literature (see [169, 143]), fast convergence to the Gibbs measure is expected throughout the high-temperature regime where  $\beta < 1$ . Over time, this line of research attracted further attention as a toy model for sampling from complicated probability distributions — here, the Gibbs measure of the underlying model — and as a way to explore how the hardness transition occurs. In this section, we survey early attempts to understand these phenomena where researchers opt for the so-called soft spin glasses, and also present recent advances made on the original mean-field models.

**Langevin dynamics on soft spin models.** The relaxational dynamics was proposed by the two physicists Sompolinsky and Zippelius [170] in 1982 where they introduced the soft SK model. Roughly speaking, in the soft SK model, the spins are continuous variables applied to a double well symmetric potential achieving global minimum at  $\pm 1$ ; moreover, each spin evolves in time according to a Langevin equation. This Langevin dynamics is called the relaxational dynamics.

In the late 1990s, Ben Arous and Guionnet studied mathematically the Langevin dynamics of the soft SK model in a series of papers [34, 35, 36] where they proved convergence and large deviation results as well as propagation of chaos. Later along with Dembo, they studied in [33] a mathematically more tractable model, the soft spherical SK (soft SSK) model, where the soft spins now belong to the  $N$ -dimensional sphere  $\mathbb{S}_N$  with radius  $\sqrt{N}$ . Among other things, they showed that the SSK model exhibits a dynamical phase transition at the static inverse temperature  $\beta_c$ , and an aging phenomenon occurs at the supercritical regime. Namely, when the relaxational dynamics starts from a i.i.d. initial condition, the limiting empirical covariance  $K(s, t)$  — a quantity that measures the correlations of soft spins at two different times  $s$  and  $t$  at the thermal dynamical limit (the number of spins  $N \rightarrow \infty$ ) — has the following phase transition: a) if  $\beta < \beta_c$ , the limiting empirical covariance decays exponentially to 0 for all  $s$  and  $t$ . b) if  $\beta = \beta_c$ , then limiting empirical covariance decays polynomially to 0 when  $t - s \rightarrow \infty$ . c) if  $\beta > \beta_c$  and  $t \gg s \gg 1$ , then  $K(s, t)(t/s)^{3/4} \in (0, \infty)$  is bounded away from 0 and

$\infty$ . In particular,  $K(s, t)$  converges to 0 if and only if  $t/s \rightarrow \infty$ . Result (c) is called the aging phenomenon: the correlation between the spin configuration at time  $s$  and the spin configuration at time  $t$  does not converge to 0 when both  $s$  and  $t - s$  converge to 0 but  $t - s$  is not large enough compared to  $s$ .

**Recent advances in Glauber dynamics.** Recently, it was shown by Bauerschmidt and Bodineau in [28] and by Eldan, Koehler and Zeitouni in [98] that fast mixing occurs when  $\beta < 1/4$ . Moreover, Eldan et al. showed in [98] that the Gibbs measure satisfies a Poincaré inequality for the Dirichlet form of Glauber dynamics, so the Glauber dynamics mixes in  $O(N^2)$  spin flips in total variation distance. Subsequently, this estimate was improved to  $O(N \log N)$  by Anari et al. in [13].

For the spherical spin glasses, Gheissari and Jagannath demonstrated in [108] that the Langevin dynamics (a continuum version of the Glauber dynamics) has a polynomial spectral gap for small values of  $\beta$ . On the other hand, Ben Arous and Jagannath proved in [37] that for sufficiently large  $\beta$ , the mixing time of the Glauber dynamics (as well as the Langevin dynamics) becomes exponentially large for both Ising and spherical spin glasses.

**A non MCMC algorithm outperforming the Glauber dynamics.** The Glauber dynamics and the Langevin dynamics belong the family of MCMC algorithms. In [9], Alaoui, Montanari and Sellke proposed an non MCMC type algorithm based on the stochastic localization for the SK model. They showed that for  $\beta < 1/2$ , there exists an algorithm with complexity  $O(N^2)$  with output law being close to the Gibbs measure in normalized Wasserstein distance. Moreover, for  $\beta > 1$ , they established a hardness result for the stable algorithms, which means that the output law of these algorithms are stable under small random perturbation of the defining matrix of the SK model. The hardness result for  $\beta > 1$  was proven by utilizing the disorder chaos, which means for them that Wasserstein distance between the Gibbs measure and the perturbed Gibbs measure is bounded from below by a positive constant for arbitrary small random perturbation.

### 1.4.2 Optimization problem for mean-field spin glasses

Since the late 1980s, optimization problems in disordered systems have been studied by researchers as paradigmatic examples to investigate why some computational problems cannot be solved using polynomial-time algorithms. Over time, this area of research has grown into a vast field that continues to expand. We do not intend to provide a comprehensive overview of this field here, and direct interested readers to the survey by Gamarnik, Moore, and Zdeborova [104] for more details.

In recent years, the so-called *overlap gap property* (OGP) has been identified as a significant barrier in the development of polynomial-time algorithms for optimization problems in disordered systems. The OGP property stems from the clustering properties observed by the physicists in [1, 141] in the context of random constraint satisfaction problems. We refer readers to [104] and the references therein. Moving forward, we will limit our focus to mean-field spin glasses.

In the context of mean-field spin glasses, the OGP can be informally described as the existence of an interval  $[\nu_1, \nu_2] \subset [0, 1]$  such that, with high probability and for sufficiently large  $N$ , any two near-ground state spin configurations  $\sigma^1$  and  $\sigma^2$  exhibit an overlap  $R_{1,2}$  that does not fall within the interval  $[\nu_1, \nu_2]$ . For a more rigorous definition of the OGP, we direct readers to Theorem 1 in [104]. By this definition, one can see that if a spin glass model exhibits the replica symmetric property or the full-RSB property, then the model does not exhibit the OGP; while if the model is  $k$ -RSB, with  $k \in \mathbb{N}$ , then the model exhibits the OGP.

For the Ising and spherical  $p$ -spin models, where  $p \geq 4$  is an even number, the OGP is known to hold. This was demonstrated in the Ising case by Theorem 2 of Chen, Gamarnik, Panchenko and Rahman's paper [20], while the spherical case can be deduced from Theorem 6 in the paper [20] by Auffinger and Chen.

It is conjectured that the OGP should be true for all  $p \geq 3$ . In contrast, the OGP is conjectured to be absent for the SK model, which corresponds to the full-RSB conjecture for the SK model at  $\beta = \infty$ . Under the assumption that the conjecture holds, Montanari has shown in [145] that an approximate message passing (AMP) algorithm can find a near-ground state in quadratic time. For the spherical model with  $p = 2$ , the OGP does not hold as the optimization problem trivially corresponds to optimizing a quadratic form over the sphere  $\mathbb{S}_N$ .

Recently, it is known that there exists an energy level  $E_{\text{ALG}}$  such that for any  $E < E_{\text{ALG}}$ , with high probability, there exists an algorithm that can find a spin configuration  $\sigma$  such that  $H(\sigma)/N \approx E$  in polynomial-time. This has been shown for full-RSB spherical spin glasses by Subag [171], for the SK model by Montanari [145], for Ising and spherical mixed  $p$ -sping models by El Alaoui, Montanari and Sellke, and for Ising and spherical mixed  $p$ -sping models with external field by Sellke [166]. The later three results rely on the AMP algorithm, and while Subag's result stems from ultrametric approximation of the Gibbs measure provided by the generalized TAP variational principle.

For models where the OGP is absent,  $E_{\text{ALG}} = E_*$ , indicating that the AMP algorithm successfully solves the optimization problem. However, in some models,  $E_{\text{ALG}} < E_*$  can occur. Very recently, for mixed even  $p$ -spin models where  $E_{\text{ALG}} < E_*$ , Huang and Sellke [115] introduced an ultrametric version of the OGP, termed the *branching-OGP*, which blocks a wide class of algorithms from finding an energy level greater than  $E_{\text{ALG}}$  in polynomial time. This class of algorithms includes the gradient descent algorithm, the AMP algorithm, and the Glauber dynamics.

### 1.4.3 Optimization problem for the CREM

As proven in [54] by Bovier and Kurkova, the maximum of the CREM satisfies

$$x_{GSE} := \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \left[ \max_{|u|=N} X_u \right] = \sqrt{2 \log 2} \int_0^1 \sqrt{\hat{a}(s)} \, ds. \quad (1.4.1)$$

With this result in mind, one can ask the following optimization problem in the context of the CREM.

**Problem 1.4.1.** *For what kind of  $A$  such that for all  $\varepsilon > 0$ , there exists a polynomial-time algorithm that can find a vertex  $|u| = N$  such that  $X_u \geq (x_{GSE} - \varepsilon)N$  with high probability?*

To respond to Problem 1.4.1, they showed the following phase transition: there exists a threshold

$$x_* = \sqrt{2 \log 2} \int_0^1 \sqrt{a(s)} \, ds$$

such that for any  $x < x_*$ , there exists a linear time algorithm that finds  $X_v \geq xN$  with high probability; for any  $x > x_*$ , there exists  $z > 0$  such that with high probability, it takes at least  $e^{zN}$  queries to find  $X_u \geq xN$ . Since  $x_{GSE} \geq x_*$  with equality holding if and only if  $A$  is concave, the near maximum can be found if and only if  $A$  is concave. Another remark is that their result correspond to the special case of the setting considered in Chapter 3 where  $\beta \rightarrow \infty$ , and linear algorithm they proposed is similar to Algorithm 1.1.

## 1.5 Summary of contributions

*For clarity, I adopt in this section the first-person single subject “I” instead of the conventional first-person subject “we” in mathematical writing to refer to the contributions I made, and I use the first-person plural subject “we” to refer the results achieved in collaboration with Maillard.*

In this section, I summarize the contributions of the thesis. The main research theme throughout my PhD has been the development of an efficient sampling algorithm for the CREM Gibbs measure and the investigation of whether a hardness threshold occurs. This led to the work in Chapter 2 and Chapter 3. An incidental result of generalizing the findings from Chapter 2 to the context of Chapter 3 was the estimation of the negative moments of the CREM partition function. Although this estimation was not needed in the final version of Chapter 3, Maillard suggested me to write it into a separate paper, presented in Chapter 4, as the estimate might be of value on its own.

### 1.5.1 Sampling the Gibbs measure

The goal of Chapter 2 and Chapter 3 is to answer the following Gibbs sampling problem raised by Addario-Berry and Maillard [2].

**Problem 1.5.1** (Problem 1 in Section 5 of [2]). *Show that there exists a threshold  $\beta_G$  such that the following transition occurs.*

- *For  $\beta < \beta_G$ , there exists an algorithm of polynomial-time such that its output law approximates the Gibbs measure.*
- *For  $\beta > \beta_G$ , there is no polynomial-time algorithm that can approximate the Gibbs measure.*

There are a few notions in Problem 1.5.1 that need to be specified. Firstly, the underlying model is the branching random walk in Chapter 2, where the detailed definition can be founded in Section 1.5.4 or Section 2.1.1. In Chapter 3, the underlying model is the continuous random energy model, and the detailed definition can be bounded in Section 1.5.7 or Section 3.1.1.

Secondly, the class of algorithms that we consider is defined as follows.

**Definition 1.5.2.** Let  $N \in \mathbb{N}$ . Let  $\tilde{\mathcal{F}}_k$  be a filtration defined by

$$\tilde{\mathcal{F}}_k = \sigma(v(1), \dots, v(k); X(v(1)), \dots, X(v(k)); U_1, \dots, U_{k+1})$$

where  $(U_k)_{k \geq 1}$  is a sequence of i.i.d. uniform random variables on  $[0, 1]$ , independent of the underlying model. A random sequence  $v = (v(k))_{k \geq 0}$  taking values in  $\mathbb{T}_N$  is called a (randomized) algorithm if  $v(0) = \emptyset$  and  $v(k+1)$  is  $\tilde{\mathcal{F}}_k$ -measurable for every  $k \geq 0$ . We further suppose that there exists a stopping time  $\tau$  with respect to the filtration  $\tilde{\mathcal{F}}$  and such that  $v(\tau) \in \partial\mathbb{T}_N$ . We call  $\tau$  the *running time* and  $v(\tau)$  the *output* of the algorithm. The *law of the output* is the (random) distribution of  $v(\tau)$ , conditioned on the underlying model.

Definition 1.5.2 appears in Section 2.1.1, where the underlying model is the branching random walk, and in Definition 3.1.2, where the underlying model is the continuous random energy model. It also appeared in similar forms in [2, 159]. Definition 1.5.2 includes a large class of algorithms. For example, The Metropolis algorithm that is defined below.

**Example 1.5.3.** The Metropolis algorithm is a Markov chain  $(v_n)_{n \in \mathbb{N}}$  on the vertices with transition probability

$$p(v, w) = \frac{e^{\beta(X_w - X_v)}}{\sum_{v \sim w'} e^{\beta(X_w - X_{v'})}}, \quad v \sim w,$$

where  $v \sim w$  means that the graph distance between  $v$  and  $w$  equals 1. One can show that Example 1.5.3 satisfies Definition 1.5.2.

Next, the notion of time complexity is defined as follows.

**Definition** (Definition 3.1.4). Let  $(\tau_N)$  be a sequence of running time corresponds to a sequence of algorithms indexed by  $N$ . Let  $h : \mathbb{N} \rightarrow \mathbb{N}$  be a function. We say that the sequence of running times is of order  $O(h(N))$  if almost surely, there exists  $N_0 \in \mathbb{N}$  such that  $\tau_N \leq h(N)$ . We say the running time is of polynomial order if there exists a polynomial  $P(N)$  such that almost surely, there exists  $N_0 \in \mathbb{N}$ .

Finally, two different notions of approximation are provided, which are adopted in Chapter 2 and in Chapter 3, respectively. Given two probability measures  $P$  and  $Q$  defined on a discrete space  $\Omega$ , the entropy of  $Q$  and the Kullback–Leibler divergence from  $Q$  to  $P$  are respectively defined by

$$H(Q) = \sum_{\omega \in \Omega} Q(\omega) \cdot \log \left( \frac{1}{Q(\omega)} \right) \quad (2.1.5)$$

$$\mathbf{d}(P||Q) = \sum_{\omega \in \Omega} P(\omega) \cdot \log \left( \frac{P(\omega)}{Q(\omega)} \right). \quad (2.1.6)$$

It should be noted that, due to Jensen's inequality, both the entropy and the Kullback-Leibler divergence are non-negative. In the subsequent discussions,  $P$  and/or  $Q$  are often selected as a Gibbs measure  $\mu_{\beta,N}$  for some  $N$ . Under these circumstances,  $\Omega$  is set to  $\partial\mathbb{T}_n$ , with  $n$  being the largest number that allows both  $P$  and  $Q$  to be defined on  $\partial\mathbb{T}_n$ .

The first notion of approximation is defined as follows.

**Definition** (Definition 2.1.5). Let  $(P_N)_{N \in \mathbb{N}}$  and  $(Q_N)_{N \in \mathbb{N}}$  are two sequences of *random* probability measures defined on a discrete space  $\Omega$ . We say that the sequence  $(P_N)_{N \in \mathbb{N}}$  approximates the sequence  $(Q_N)_{N \in \mathbb{N}}$  if

$$\frac{\mathbf{d}(P_N||Q_N)}{H(Q_N)} \rightarrow 0 \quad \text{in probability as } N \rightarrow \infty. \quad (1.5.1)$$

The second notion of approximation is given as follows.

**Definition** (Definition 3.1.6). Let  $(P_N)_{N \in \mathbb{N}}$  and  $(Q_N)_{N \in \mathbb{N}}$  are two sequences of *random* probability measures defined on a discrete space  $\Omega$ . We say that the sequence  $(P_N)_{N \in \mathbb{N}}$  approximates the sequence  $(Q_N)_{N \in \mathbb{N}}$  if

$$\frac{1}{N} \mathbf{d}(P_N||Q_N) \rightarrow 0 \quad \text{in probability as } N \rightarrow \infty.$$

Before closing this section, there are two remarks I want to make

- The following

$$\frac{1}{N} \mathbf{d}(P_N||Q_N) \rightarrow 0 \quad \text{as } N \rightarrow \infty. \quad (2.1.8)$$

is known as *measure equivalence* or *equivalence in the sense of specific relative entropy* in the physics literature. See [184] and the reference therein. Mathematically, Equation (2.1.8) implies the following: if  $(A_N)_{N \geq 1}$  is a sequence of sets such that  $Q_N(A_N)$  converges to 0 exponentially fast as  $N \rightarrow \infty$ , then we also have  $P_N(A_N) \rightarrow 0$ . Indeed, this is an easy consequence of Birgé's inequality (see e.g. Theorem 4.20 in [52]).

- Definition 3.1.6 is equivalent to saying that there exists  $\varepsilon_N \rightarrow 0$  such that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{P} \left( \frac{1}{N} \mathbf{d}(P_N||Q_N) < \varepsilon_N \right) = 0.$$

In the context of branching random walks, the following lemma on the branching random walk Gibbs measure is folklore, where a proof is provided in Section 2.6 for completeness.

**Lemma** (Lemma 2.1.4). *Let  $\beta_c$  be the critical inverse temperature defined in (1.5.2). Then the following phase transition occurs.*

1. *If  $\beta \in [0, \beta_c)$ , then  $H(\mu_{\beta,N})/N$  converges in probability to a positive constant as  $N \rightarrow \infty$ .*

2. If  $\beta > \beta_c$  and if  $\beta_c \in \mathcal{D}(\varphi)^\circ$ , then  $H(\mu_{\beta,N}) = O(1)$  in probability, as  $N \rightarrow \infty$ . In other words, the sequence of random variables  $(H(\mu_{\beta,N}))_{N \geq 1}$  is tight.

Therefore, in the context of branching random walks, for  $\beta < \beta_c$ , a sequence of (random) probability measure  $(P_N)_{N \in \mathbb{N}}$  approximates the BRW Gibbs measure  $\mu_{\beta,N}$  in Definition 2.1.5 is equivalent to saying that  $(P_N)_{N \in \mathbb{N}}$  approximates the BRW Gibbs measure  $\mu_{\beta,N}$  in Definition 3.1.6.

### 1.5.2 Recursive sampling algorithm

To address Problem 1.5.1, the following algorithm was proposed. Fix  $\beta > 0$ ,  $N \in \mathbb{N}$ , and  $M = M_N \in \llbracket 1, N \rrbracket$ . Given a configuration of the underlying model, consider the following algorithm, which is the same as Algorithm 2.1 and Algorithm 3.1:

---

**Algorithm 1.1:** Recursive sampling on renormalized tree

---

```

set  $v = \emptyset$    while  $|v| < N$  do
  | sample  $w$  with  $|w| = M \wedge (N - |v|)$  according to the Gibbs measure
  |  $\mu_{\beta, M \wedge (N - |v|)}^v$    replace  $v$  with  $vw$ 
output  $v$ 
    
```

---

The running time of Algorithm 1.1 is deterministic and bounded by  $\lceil N/M \rceil 2^M$ . The output law of Algorithm 1.1 is a random probability measure  $\mu_{\beta, M, N}$  on  $\partial \mathbb{T}_N$  that is recursively defined as follows:

$$\begin{aligned} \mu_{\beta, M, 0}(\emptyset) &= 1 \\ \mu_{\beta, M, N \wedge (k+1)M}(vw) &= \mu_{\beta, M, kM}(v) \cdot \mu_{\beta, M \wedge (N - kM)}^v(w) \end{aligned} \tag{3.1.9}$$

for all  $|v| = kM$ ,  $|w| = M \wedge (N - kM)$  and  $k \in \llbracket 0, \lfloor \frac{N}{M} \rfloor \rrbracket$ . It is not hard to see that

$$\mu_{\beta, M, N}(u) = \frac{e^{\beta X_u}}{Z_{\beta, M, N}(u)}, \tag{3.1.10}$$

where

$$Z_{\beta, M, N}(u) = \prod_{k=1}^{\lfloor N/M \rfloor} Z_{\beta, M \wedge (N - kM)}^{u[kM]}.$$

### 1.5.3 Summary of Chapter 2

In Chapter 2, Maillard and I answered Problem 1.5.1 partially in the setting of branching random walks, where the branching random walks have  $d \geq 2$  offspring in each generation. In summary, we showed that the following phase transition occurs at the critical inverse temperature  $\beta_c$ :

1. Let  $\beta < \beta_c$ . For any  $\varepsilon > 0$ , we constructed an  $O(N^{1+\varepsilon})$  algorithm that approximates the Gibbs measure in the sense of Definition 2.1.5. This description differs from the version provided in Chapter 2: there, the result was

showing the existence of a polynomial-time algorithm that approximates the Gibbs measure in the sense of Definition 2.1.5. However, I demonstrate in Corollary 1.5.4 that by choosing a different sequence  $M_N$ , the running time of the algorithm can indeed be reduced to  $O(N^{1+\varepsilon})$ .

2. For  $\beta > \beta_c$ , under mild regularity condition, we showed that the following hardness result: for any  $\delta > 0$ , there exists  $z > 0$  such that for any algorithm with output law approximating the Gibbs measure in the sense of Definition 2.1.5, the probability that the running time of the algorithm is greater than  $e^{z\sqrt{N}}$  is at least  $1 - \delta$ , as  $N \rightarrow \infty$ .

A detailed description of the contribution in Chapter 2 is provided in Section 1.5.5.

### 1.5.4 Branching random walk

To describe the results in details, I recall the content in Section 2.1.1 that provides the definition of the branching random walk that we considered in Chapter 2 and known facts of this model.

**Underlying tree.** Let  $\mathbb{T}$  be a rooted  $d$ -ary tree, where  $d \geq 2$ . The depth of a vertex  $v \in \mathbb{T}$  is denoted by  $|v|$ . The root is denoted by  $\emptyset$ , and any vertex  $v$  with depth  $n \geq 1$  is indexed by a string  $v_1 \cdots v_n \in \{0, \dots, d-1\}^n$ . For any  $v, w \in \mathbb{T}$ , the notation  $v \leq w$  means that  $v$  is a prefix of  $w$  and  $v < w$  means that  $v$  is a prefix of  $w$  strictly shorter than  $w$ . In the following, for any  $v \in \mathbb{T}_N$ , any vertex  $w$  is said to be an ancestor of  $v$  if  $w \leq v$ . For any  $v \in \mathbb{T}$  and  $m \in \llbracket 0, |v| \rrbracket$ , define  $v[m]$  to be the ancestor of  $v$  of depth  $m$ . For all  $v, w \in \mathbb{T}$ , let  $v \wedge w$  be the most recent common ancestor of  $v$  and  $w$ . Let  $\mathbb{T}_n$  be the subtree of  $\mathbb{T}$  containing vertices of depth less or equal to  $n$ , and  $\partial\mathbb{T}_n$  be the leaves of  $\mathbb{T}_n$ .

**Branching random walk.** Let  $\mathbf{Y} = (Y_0, \dots, Y_{d-1})$  be a  $d$ -dimensional random vector. Let  $(\mathbf{Y}^v)_{v \in \mathbb{T}}$  be i.i.d. copies of  $\mathbf{Y}$  where  $\mathbf{Y}^v = (Y_{v0}, \dots, Y_{v(d-1)})$  – this uniquely defines  $Y_u$  for every  $u \in \mathbb{T} \setminus \{\emptyset\}$ . Define the process  $\mathbf{X} = (X_v)_{v \in \mathbb{T}}$  by

$$\begin{cases} X_\emptyset = 0, \\ X_v = \sum_{\emptyset < w \leq v} Y_w, \quad |v| \geq 1. \end{cases}$$

The process  $\mathbf{X}$  is called the *branching random walk with increments  $\mathbf{Y}$* . See Figure 1.7 for an illustration of the definition.

It is well-known that  $\mathbf{X}$  has the branching property: let  $\mathcal{F} = (\mathcal{F}_n)_{n \geq 0}$  be its natural filtration. For any  $v \in \mathbb{T}$  with  $|v| = n$ , define

$$\mathbf{X}^v = (X_w^v)_{|w| \geq 0} = (X_{vw} - X_v)_{|w| \geq 0}.$$

Then  $(\mathbf{X}^v)_{|v|=n}$  are iid copies of  $\mathbf{X}$  and independent of  $\mathcal{F}_n$ .

**Gibbs measures.** Define the following function

$$\varphi(\beta) = \log \left( \mathbb{E} \left[ \sum_{i=0}^{d-1} e^{\beta Y_i} \right] \right) \in (-\infty, \infty], \quad \beta \in \mathbb{R},$$

and set  $\mathcal{D}(\varphi) = \{\beta \in \mathbb{R} : \varphi(\beta) < \infty\}$ . It is well known that  $\varphi$  is convex and that it is smooth on  $\mathcal{D}(\varphi)^\circ$ , the interior of  $\mathcal{D}(\varphi)$ . In Chapter 2, we assumed the following.

**Assumption** (Assumption 2.1.1).  $0 \in \mathcal{D}(\varphi)^\circ$ .

In the context of branching random walk, instead of considering directly the partition function, one usually opt to the additive martingale, which is defined as follows. For  $n \in \mathbb{N}$ , the additive martingales are defined as follows.

$$\begin{aligned} W_{\beta,n} &= \sum_{|v|=n} e^{\beta X_v - \varphi(\beta)n}, \quad n \geq 0, \\ W_{\beta,n}^u &= \sum_{|w|=n} e^{\beta X_w^u - \varphi(\beta)n}, \quad u \in \mathbb{T} \text{ and } n \geq 0. \end{aligned}$$

Note that for every  $m \leq n$ ,

$$W_{\beta,n} = \sum_{|u|=m} e^{\beta X_u - \varphi(\beta)m} \cdot W_{\beta,n-m}^u. \quad (2.1.1)$$

For  $\beta \in \mathbb{R}$  and  $n \in \mathbb{N}$ , the Gibbs measure of parameter  $\beta$  on  $\mathbb{T}_n$  is defined as

$$\mu_{\beta,n}(u) = e^{\beta X_u - \varphi(\beta)m} \cdot \frac{W_{\beta,n-m}^u}{W_{\beta,n}}, \quad |u| = m \leq n. \quad (2.1.2)$$

Note that  $\mu_{\beta,n}$  is usually defined on  $\partial\mathbb{T}_n$  only, but it will be helpful to define it on the whole tree  $\mathbb{T}_n$ . By (2.1.1), for every  $m \leq n$ , the restriction of  $\mu_{\beta,n}$  to  $\partial\mathbb{T}_m$  is a probability measure. Similarly, we can define

$$\mu_{\beta,n}^v(u) = e^{\beta X_u^v - \varphi(\beta)m} \cdot \frac{W_{\beta,n-m}^{vu}}{W_{\beta,n}^v}, \quad v \in \mathbb{T} \text{ and } |u| = m \leq n. \quad (2.1.3)$$

The free energy of the branching random walk has been calculated by Derrida and Spohn [91] (and can also be deduced from Biggins [46])

$$\lim_{n \rightarrow \infty} \frac{1}{\beta n} \log \sum_{|v|=n} e^{\beta X_v} = \begin{cases} \frac{1}{\beta} \varphi(\beta) & \text{if } \beta \in (0, \beta_c) \\ \frac{1}{\beta_c} \varphi(\beta_c) & \text{if } \beta \geq \beta_c, \end{cases}$$

where the limit is meant to be in probability and where the critical inverse temperature  $\beta_c$  is defined by

$$\beta_c = \sup\{\beta \in \mathcal{D}(\varphi)^\circ \mid \beta \varphi'(\beta) < \varphi(\beta)\} \in (0, \infty]. \quad (1.5.2)$$

We will mostly be interested in the phase  $\beta < \beta_c$ . In this phase, the following fact is known for the additive martingale, and the proof can be founded in [47, 118, 129].

**Fact** (Fact 2.1.2). *If  $\beta < \beta_c$ , then the martingale  $(W_{\beta,n})_{n \geq 0}$  is uniformly integrable. In fact, there exists a (strictly) positive random variable  $W_{\beta,\infty}$  with  $\mathbb{E}[W_{\beta,\infty}] = 1$  and such that  $W_{\beta,n} \rightarrow W_{\beta,\infty}$  almost surely and in  $L^1$  as  $n \rightarrow \infty$ .*

### 1.5.5 Contributions in Chapter 2

In the following, I give a more detailed description of the results in Chapter 2.

#### Subcritical regime $\beta < \beta_c$ : efficiency of recursive sampling

Firstly, Maillard and I showed that for all  $p \geq 1$ , the  $L^p$  norm of the KL divergence  $\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N})$  from  $\mu_{\beta,M,N}$  to the Gibbs measure  $\mu_{\beta,N}$  is bounded from above by the constant depending only on  $p$  times  $N/M$ . Moreover, we showed that the KL divergence concentrates to its expectation.

**Theorem** (Theorem 2.1.7). *Let  $N \in \mathbb{N}$ ,  $M \in \llbracket 1, N \rrbracket$ , and  $\beta \in [0, \beta_c)$ . Then for all  $p \geq 1$ , there exists a constant  $C_1(p) > 0$  such that*

$$\|\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N})\|_p \leq C_1(p) \cdot \left\lfloor \frac{N}{M} \right\rfloor. \quad (2.1.9)$$

Moreover, for all  $p \geq 1$ , there exists a constant  $C_1(p) > 0$  such that

$$\|\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N}) - \mathbb{E}[\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N})]\|_p \leq C_1(p). \quad (2.1.10)$$

I summarize the proof of (2.1.9) and (2.1.10) separately.

*Sketch of proof for (2.1.9).* The first step we took was to decompose the KL divergence as follows.

**Theorem** (Theorem 2.2.2). *For any two  $M$  and  $N$  integers such that  $M \leq N$ , we have*

$$\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N}) = \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot \mathbf{d}(\mu_{\beta,M}^u || \mu_{\beta,N-KM}^u). \quad (2.2.1)$$

Next, we wanted to bound the KL divergences between two Gibbs measures appearing in the decomposition of (2.2.1). To do so, we further decomposed the KL divergence between two Gibbs measures.

**Proposition** (Proposition 2.2.3). *For any two  $M$  and  $N$  integers such that  $M \leq N$ , we have*

$$\mathbf{d}(\mu_{\beta,M} || \mu_{\beta,N}) = \log W_{\beta,N} - \log W_{\beta,M} - \sum_{|u|=M} \mu_{\beta,M}(u) \cdot \log W_{\beta,N-M}^u.$$

Then, we showed that whenever  $\beta \in [0, \beta_c)$ ,  $(\log W_{\beta,n})_{n \geq 0}$  is bounded in  $L^p$  for all  $p > 1$ .

**Lemma** (Lemma 2.3.1). *Let  $\beta \in [0, \beta_c)$ . Then  $(\log W_{\beta,n})_{n \geq 0}$  is a supermartingale such that*

$$\sup_{n \geq 0} \|\log W_{\beta,n}\|_p < \infty, \quad p \geq 1.$$

Combining Proposition 2.2.3 and Lemma 2.3.1, we derived the following.

**Proposition** (Proposition 2.3.2). *For any  $p > 1$ , for any two integers  $M$  and  $N$  such that  $M \leq N$ , there exists a constant  $C(p) > 0$  such that*

$$\|\mathbf{d}(\mu_{\beta,M} \parallel \mu_{\beta,N})\|_p \leq C(p).$$

By Theorem 2.2.2, Minkowski's inequality, Jensen's inequality, the law of iterated expectation, the branching property and Proposition 2.3.2,

$$\begin{aligned} \|\mathbf{d}(\mu_{\beta,M,N} \parallel \mu_{\beta,N})\|_p &\leq \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \left\| \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot \mathbf{d}(\mu_{\beta,M}^u \parallel \mu_{\beta,N-KM}^u) \right\|_p \\ &\leq \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \left( \mathbb{E} \left[ \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot |\mathbf{d}(\mu_{\beta,M}^u \parallel \mu_{\beta,N-KM}^u)|^p \right] \right)^{1/p} \\ &\leq \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \|\mathbf{d}(\mu_{\beta,M} \parallel \mu_{\beta,N-KM})\|_p \\ &\leq C(p) \cdot \left\lfloor \frac{N}{M} \right\rfloor, \end{aligned}$$

and proof is completed.  $\blacksquare$

Next, I give the ideas to prove (2.1.10).

*Sketch of proof for (2.1.10).* First, we showed the following

**Lemma** (Lemma 2.4.1). *Let  $\beta \in [0, \beta_c)$ . For all  $M \in \mathbb{N}$ , there exists  $r \in (0, 1)$  independent of  $M$  such that,*

$$\mathbb{E} \left[ \sum_{|u|=M} \mu_{\beta,M}(u)^2 \right] \leq r \quad (2.4.4)$$

Moreover, for all  $K \in \mathbb{N}$ ,

$$\mathbb{E} \left[ \sum_{|u|=KM} \mu_{\beta,M,KM}(u)^2 \right] \leq r^K. \quad (2.4.5)$$

By Theorem 2.2.2 and Minkowski's inequality, we have

$$\begin{aligned} &\|\mathbf{d}(\mu_{\beta,M,N} \parallel \mu_{\beta,N}) - \mathbb{E} [\mathbf{d}(\mu_{\beta,M,N} \parallel \mu_{\beta,N})]\|_p \\ &\leq \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \left\| \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot (\mathbf{d}(\mu_{\beta,M}^u \parallel \mu_{\beta,N-KM}^u) - \mathbb{E} [\mathbf{d}(\mu_{\beta,M}^u \parallel \mu_{\beta,N-KM}^u)]) \right\|_p. \end{aligned} \quad (2.4.9)$$

Then by Lemma 2.4.1, (2.4.9), Proposition 2.3.2 and an argument involving concentration inequalities (see Section 2.4.2 for details), there exists a constant  $C > 0$  independent of  $K, M, N$  such that

$$\|\mathbf{d}(\mu_{\beta,M,N} \parallel \mu_{\beta,N}) - \mathbb{E} [\mathbf{d}(\mu_{\beta,M,N} \parallel \mu_{\beta,N})]\|_p \leq C \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} r^K \leq C \sum_{K=0}^{\infty} r^K,$$

and the proof is completed.  $\blacksquare$

By Theorem 2.1.7, Maillard and I derived Corollary 2.1.8, which states that with an appropriate choice of  $M_N$ , the algorithm approximates the Gibbs measure  $\mu_{\beta,N}$  with a running time of  $O(N^{1+\varepsilon})$ . It is important to note that the version of Corollary 2.1.8 presented below differs from the published version in Chapter 2: while the published version asserts the existence of a polynomial-time algorithm, I demonstrate here that by choosing a different sequence  $M_N$ , the running time of the algorithm can indeed be reduced to  $O(N^{1+\varepsilon})$ .

**Corollary 1.5.4** (Modified version of Corollary 2.1.8). *Suppose that  $\beta \in [0, \beta_c)$ . Let  $p > 1$ . Given  $\varepsilon > 0$ , let  $M_N = \lfloor \varepsilon \log_d N \rfloor \wedge N$  and  $\tilde{\mu}_{\beta,N} = \mu_{\beta, M_N, N}$  be the output law of Algorithm 3.1. Then,*

$$\frac{1}{N} \|\mathbf{d}(\tilde{\mu}_{\beta,N} \| \mu_{\beta,N})\|_p \rightarrow 0. \quad (2.1.11)$$

*In particular,  $\tilde{\mu}_{\beta,N}$  approximates the Gibbs measure  $\mu_{\beta,N}$  in the sense of Definition 2.1.5. Moreover, the running time is deterministic and of order  $O(N^{1+\varepsilon})$ .*

*Proof.* Let  $M_N = \lfloor \varepsilon \log_d N \rfloor \wedge N$  and set  $\tilde{\mu}_{\beta,N} = \mu_{\beta, M_N, N}$ . Equation (2.1.11) then follows from (2.1.9) in Theorem 2.1.7. Applying first part of Lemma 2.1.4,  $\tilde{\mu}_{\beta,N}$  approximates the Gibbs measure  $\mu_{\beta,N}$  in the sense of Definition 2.1.5. Finally, by our choice of  $M_N$ , Remark 2.1.3 implies that  $\tilde{\mu}_{\beta,N}$  can be computed by a  $O(N^{1+\varepsilon})$ -time algorithm.  $\blacksquare$

### Supercritical regime $\beta > \beta_c$ : hardness for generic algorithms

In the supercritical regime  $\beta > \beta_c$ , we provide a hardness result, assuming a mild regularity condition.

**Theorem** (Theorem 2.1.9). *Assume  $\beta_c \in \mathcal{D}(\varphi)^\circ$  (in particular,  $\beta_c < \infty$ ). Let  $\beta > \beta_c$ . Let  $v = (v(k))_{k \geq 0}$  be an algorithm which outputs a vertex of law  $\tilde{\mu}_N$  such that  $\tilde{\mu}_N$  approximates the Gibbs measure  $\mu_{\beta,N}$  in the sense of Definition 2.1.5. Let  $\tau$  be the running time of the algorithm. Then for every  $\delta > 0$ , there exists  $z > 0$ , such that for large enough  $N$ ,*

$$\mathbb{P} \left( \tau \geq e^{z\sqrt{N}} \right) \geq 1 - \delta.$$

*Sketch of proof for Theorem 2.1.9.* Assume that  $\beta > \beta_c$ . Let  $\delta > 0$ . For  $z > 0$ , call a vertex  $w \in \partial \mathbb{T}_{\lfloor N/2 \rfloor}$   $z$ -exceptional if it has a descendant  $u \in \partial \mathbb{T}_N$  such that  $X_u - X_w - mN/2 > z\sqrt{N}$ . For any  $v \in \mathbb{T}_N$ , define the following set of vertices:

$$\mathcal{V}_v = \begin{cases} w \in \mathbb{T}_N : |v \wedge w| \geq \lfloor N/2 \rfloor, & \text{if } |v| \geq \lfloor N/2 \rfloor \\ v, & \text{otherwise.} \end{cases}$$

Let  $(v(k))_{k \geq 0}$  be an algorithm that approximates the Gibbs measure in the sense of Definition 2.1.5. Define the stopping time

$$\tau' = \inf \{ k \geq 0 : \mathcal{V}_{v(k)} \text{ is } z\text{-exceptional} \}.$$

The key observations are listed as follows.

1. Let  $\tau$  be the running time of the algorithm. By Lemma 2.5.1 and the fact that the algorithm approximates the Gibbs measure, one can show that for a given  $\delta > 0$ , there exists  $z > 0$  such that

$$\mathbb{P}(\tau \geq \tau') \geq 1 - \delta$$

for  $N$  sufficiently large.

2. One can show that  $\tau'$  stochastically dominates a geometric random variable with parameter  $e^{-c\sqrt{N}}$ , for some  $c = c_z$ .
3. Combining Point 1 and Point 2, one can conclude that with probability at least  $1 - \delta$ ,  $\tau$  is bounded from below by  $e^{z'\sqrt{N}}$ , for some  $z' > 0$  independent of  $N$ .

This completes the proof. ■

### 1.5.6 Summary of Chapter 3

Subsequently to Chapter 2, I tried to generalize the results to general CREM without assuming homogeneity. In Chapter 3, I answered Problem 1.5.1 completely with assertion. In summary, I showed the following things.

1. When the covariance function  $A$  of the CREM is concave, for all  $\beta > 0$ , the recursive sampling algorithm approximates the Gibbs measure in the sense of Definition 3.1.6 with running time of order  $O(N^{1+\varepsilon})$ .
2. When  $A$  is non-concave, we identify a threshold  $\beta_G < \infty$  such that the following hardness transition occurs:
  - (a) For every  $\beta \leq \beta_G$ , the recursive sampling algorithm approximates the Gibbs measure in the sense of Definition 3.1.6 with running time of order  $O(N^{1+\varepsilon})$ .
  - (b) For every  $\beta > \beta_G$ , we prove a hardness result for a generic class of algorithms. Namely, there exists  $\gamma > 0$  such that for any algorithm in this class that approximates the Gibbs measure in the sense of Definition 3.1.6, the running time of this algorithm is at least  $e^{\gamma N}$  with probability approaching 1.

A detailed descriptions of the result in Chapter 3 is provided in Section 1.5.8.

### 1.5.7 Continuous random energy model

Let  $\mathbb{N} = \{1, 2, \dots\}$  be the set of positive integer. For each pair of integers  $n$  and  $m$  such that  $n \leq m$ , let  $\llbracket n, m \rrbracket$  be the set of integers between  $n$  and  $m$ .

**Binary tree.** Fixing  $N \in \mathbb{N}$ , let  $\mathbb{T}_N = \{\emptyset\} \cup \bigcup_{n=1}^N \{0, 1\}^n$  be the binary tree rooted at  $\emptyset$ . The depth of a vertex  $v \in \mathbb{T}_N$  is denoted by  $|v|$ . For any  $v, w \in \mathbb{T}_N$ ,  $v \leq w$  means that  $v$  is a prefix of  $w$  and  $v < w$  means that  $v$  is a prefix of  $w$  strictly shorter than  $w$ . In the following, for any  $v \in \mathbb{T}_N$ , a vertex  $w$  is said to be an ancestor of  $v$  if  $w \leq v$ . For any  $v \in \mathbb{T}_N$  and  $n \in \llbracket 0, |v| \rrbracket$ , define  $v[n]$  to be the ancestor of  $v$  of depth  $n$ . For all  $v, w \in \mathbb{T}_N$ , let  $v \wedge w$  be the most recent common ancestor of  $v$  and  $w$ . Let  $\partial\mathbb{T}_N$  be the set of leaves of  $\mathbb{T}_N$ , and for any  $v \in \mathbb{T}_N$ , let  $\mathbb{T}_n^v$  be the subtree of  $\mathbb{T}_N$  rooted at  $v$  with depth  $n$ .

**The model.** Let  $A$  be a non-decreasing function defined on an interval  $[0, 1]$  such that  $A(0) = 0$  and  $A(1) = 1$ . In Chapter 3, it is assumed that there exists a bounded Riemann integrable function  $a$  such that  $A$  for all  $t \in [0, 1]$ ,

$$A(t) = \int_0^t a(s) ds.$$

Let  $\hat{A}$  be the concave hull of  $A$  and let  $\hat{a}$  be the right derivative of  $\hat{A}$ . Note that the  $\hat{A}$  is also equals to the Riemann integral of  $\hat{a}$ , i.e., for all  $t \in [0, 1]$ ,

$$\hat{A}(t) = \int_0^t \hat{a}(s) ds.$$

The CREM is now defined as follows.

**Definition** (Definition 3.1.1). Given  $N \in \mathbb{N}$ , the CREM is a centered Gaussian process  $\mathbf{X} = (X_u)_{u \in \mathbb{T}_N}$  indexed by the binary tree  $\mathbb{T}_N$  of depth  $N$  with covariance function

$$\mathbb{E}[X_v X_w] = N \cdot A\left(\frac{|v \wedge w|}{N}\right), \quad \forall v, w \in \mathbb{T}_N, \quad (3.1.1)$$

where  $|v \wedge w|$  is the depth of the most recent common ancestor of  $v$  and  $w$ .

In Chapter 3, I considered a sequence of CREM  $(\mathbf{X}_N)_{N \in \mathbb{N}}$  defined on the same underlying probability space. For simplicity, I dropped  $N$  as long as it causes no ambiguity.

**Branching property.** The CREM can be viewed as an inhomogeneous binary branching random walk with Gaussian increments. In particular, it has the following branching property: let  $(\mathcal{F}_k)_{k=0}^N$  be the natural filtration of the CREM. For any  $u \in \mathbb{T}_N$  with  $|u| = n \in \llbracket 0, N \rrbracket$ , call the process

$$\mathbf{X}^u = (X_w^u)_{w \in \mathbb{T}_{N-n}^u}$$

the CREM indexed by the subtree  $\mathbb{T}_{N-n}^u$ , where  $X_w^u = X_{uw} - X_u$ . For any  $n \in \llbracket 0, N \rrbracket$ , let  $\mathbf{X}^{(n)} = (X_u^{(n)})_{u \in \mathbb{T}_{N-n}}$  be a centered Gaussian process with covariance function

$$\mathbb{E}\left[X_{w_1}^{(n)} X_{w_2}^{(n)}\right] = N \cdot A\left(\frac{n + |w_1 \wedge w_2|}{N}\right), \quad \forall w_1, w_2 \in \mathbb{T}_{N-n}.$$

Then, the branching property states that collection of processes  $\{\mathbf{X}^u : |u| = n\}$  are independent and have the identical distribution of  $\mathbf{X}^{(n)}$ , and they are independent of  $\mathcal{F}_n$ .

**Partition function and Gibbs measure.** Given a subtree  $\mathbb{T}_M^v$  rooted at  $v$  and of depth  $M \in \llbracket 0, N - |v| \rrbracket$ , the *Gibbs measure with inverse temperature*  $\beta > 0$  is defined by

$$\mu_{\beta, M}^v(u) = \frac{1}{Z_{\beta, M}^v} e^{\beta X_u^v}, \quad \forall vu \in \mathbb{T}_M^v, \quad (3.1.2)$$

where

$$Z_{\beta, M}^v = \sum_{vu \in \mathbb{T}_M^v} e^{\beta X_u^v}, \quad (3.1.3)$$

is the partition function on the subtree  $\mathbb{T}_M^v$ . In particular, the following conventions are adopted.

$$\mu_{\beta, M} = \mu_{\beta, M}^\emptyset \quad \text{and} \quad Z_{\beta, M} = Z_{\beta, M}^\emptyset$$

for any  $M \in \llbracket 0, N \rrbracket$ . For completeness, define  $Z_{\beta, M}^{(n)} = \sum_{|u|=M} e^{\beta X_u^{(n)}}$ .

**Free energy and its lower bound.** For  $v \in \mathbb{T}_N$ , the logarithm of the partition function  $\log Z_{\beta, M}^v$  is referred to as the free energy on the subtree  $\mathbb{T}_M^v$ . The free energy  $F_\beta$  of the CREM is defined as follows, and  $F_\beta$  admits an explicit expression.

$$F_\beta := \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\log Z_{\beta, N}] = \int_0^1 f(\beta \sqrt{\hat{a}(s)}) ds, \quad (3.1.4)$$

where the function  $f$  is defined as

$$f(\beta) = \begin{cases} \log 2 + \frac{\beta^2}{2}, & \beta < \sqrt{2 \log 2} \\ \sqrt{2 \log 2} \beta, & \beta \geq \sqrt{2 \log 2}. \end{cases} \quad (3.1.5)$$

For completeness, the proof of (3.1.4) is included in Fact 3.A.2. When clear,  $F_\beta$  is abbreviated as the free energy. I also introduced a related quantity  $\tilde{F}_\beta$  defined as

$$\tilde{F}_\beta := \int_0^1 f(\beta \sqrt{a(s)}) ds. \quad (3.1.6)$$

In Proposition 3.A.1, I showed that  $F_\beta \geq \tilde{F}_\beta$ , and I characterized the condition where the equality holds.

### 1.5.8 Contributions in Chapter 3

Now, I describe in details the contribution in Chapter 3.

#### Subcritical and critical regime $\beta \leq \beta_G$ : optimality of recursive sampling

The first theorem states that the KL divergence from the output law of Algorithm 3.1 to the Gibbs measure concentrates in the following sense.

**Theorem** (Theorem 3.1.9). *Let  $\beta > 0$ ,  $N \in \mathbb{N}$  and  $M \in \llbracket 1, N \rrbracket$ . Then for all  $p \geq 1$ , there exists a constant  $C_p > 0$  depending only on  $p$  such that*

$$\frac{1}{N} \|\mathbf{d}(\mu_{\beta, M, N} \| \mu_{\beta, N}) - \mathbb{E}[\mathbf{d}(\mu_{\beta, M, N} \| \mu_{\beta, N})]\|_p \leq \frac{\beta C_p}{\sqrt{M}}.$$

*Sketch of proof for Theorem 3.1.9.* The structure of the proof is similar to the proof of (2.1.10) but with several modifications. Firstly, a more natural way to decompose the KL divergence then Proposition 2.2.3 is the following.

**Proposition** (Proposition 3.2.1). *For all  $\beta > 0$  and for any two integers  $M, N \in \mathbb{N}$  such that  $M \leq N$ , we have*

$$\mathbf{d}(\mu_{\beta, M, N} \| \mu_{\beta, N}) = \log Z_{\beta, N} - \sum_{k=0}^{\lfloor N/M \rfloor} \sum_{|u|=kM} \mu_{\beta, M, kM}(u) \cdot \log Z_{\beta, M \wedge (N-kM)}^u.$$

By the branching property, one can show that

**Proposition** (Proposition 3.2.2). *For all  $\beta > 0$  and for any two integers  $M, N \in \mathbb{N}$  such that  $M \leq N$ , the expectation of the KL divergence from  $\mu_{\beta, M, N}$  to  $\mu_{\beta, N}$  admits the following decomposition.*

$$\mathbb{E}[\mathbf{d}(\mu_{\beta, M, N} \| \mu_{\beta, N})] = \mathbb{E}[\log Z_{\beta, N}] - \sum_{k=1}^{\lfloor N/M \rfloor} \mathbb{E}\left[\log Z_{\beta, M \wedge (N-kM)}^{(kM)}\right].$$

By a similar argument appeared in the proof of (2.1.10), one can derive that

$$\begin{aligned} & \frac{1}{N} \|\mathbf{d}(\mu_{\beta, M, N} \| \mu_{\beta, N}) - \mathbb{E}[\mathbf{d}(\mu_{\beta, M, N} \| \mu_{\beta, N})]\|_p \\ & \leq \frac{1}{N} \sum_{k=0}^{\lfloor N/M \rfloor} \left\| \log Z_{\beta, M \wedge (N-kM)}^{(kM)} - \mathbb{E}\left[\log Z_{\beta, M \wedge (N-kM)}^{(kM)}\right] \right\|^p. \end{aligned} \quad (1.5.3)$$

Since the CREM is a Gaussian process, one can then apply the concentration inequality of free energies (Theorem 1.2 in [156]) to conclude the proof.  $\blacksquare$

Next, I proved that with a suitable choice of  $M_N$ , the expectation of the KL divergence renormalized by  $N$  converges to the difference between  $F_\beta$  and  $\tilde{F}_\beta$ .

**Theorem** (Theorem 3.1.10). *Let  $\beta > 0$ ,  $N \in \mathbb{N}$ , and  $M_N$  be a sequence such that  $M_N \in \llbracket 1, N \rrbracket$  and  $M_N \rightarrow \infty$ . Let  $\tilde{\mu}_{\beta, N} = \mu_{\beta, M_N, N}$  be the output law of Algorithm 3.1. Then,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\mathbf{d}(\tilde{\mu}_{\beta, N} \| \mu_{\beta, N})] = F_\beta - \tilde{F}_\beta \geq 0,$$

*with equality holding if and only if  $\beta \leq \beta_G$ .*

*Sketch of proof for Theorem 3.1.10.* It suffices to show that

**Proposition** (Proposition 3.3.1). *Let  $M_N$  be a sequence such that  $M_N \in \llbracket 1, N \rrbracket$  and  $M_N \rightarrow \infty$ . Then,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^{\lfloor N/M_N \rfloor} \mathbb{E}\left[\log Z_{\beta, M_N \wedge (N-kM_N)}^{(kM_N)}\right] = \tilde{F}_\beta.$$

The proof of Proposition 3.3.1 is essentially by sandwiching the covariance function CREM with the covariance functions of two GREM. ■

As a corollary of Theorem 3.1.10, in the subcritical regime, with a good choice of  $M_N$ , the mean of the KL divergence divided by  $N$  converges to 0 when  $N \rightarrow \infty$ . Moreover, for  $\varepsilon > 0$ , with a good choice of  $M_N$ , the running time is of  $O(N^{1+\varepsilon})$ .

**Corollary** (Corollary 3.1.11). *Fix  $\beta \in [0, \beta_G]$ . Given  $\varepsilon > 0$ , let  $M_N = \lfloor \varepsilon \log_2 N \rfloor \wedge N$  and  $\tilde{\mu}_{\beta,N} = \mu_{\beta,M_N,N}$  be the output law of Algorithm 3.1. Then,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} [\mathbf{d}(\tilde{\mu}_{\beta,N} || \mu_{\beta,N})] = 0. \quad (3.1.11)$$

Moreover, the running time is deterministic and of order  $O(N^{1+\varepsilon})$ .

*Remark 1.5.5.* Note that for  $A$  concave, Corollary 3.1.11 yields that Algorithm 3.1 approximates the Gibbs measure for all  $\beta \in (0, \infty)$  as  $\beta_G = \infty$  for  $A$  concave.

Corollary 3.1.11 implies in particular that the algorithm approximates the Gibbs measure with probability approaching 1. Indeed, by choosing, e.g.,  $\varepsilon_N = (\frac{1}{N} \mathbb{E} [\mathbf{d}(\tilde{\mu}_{\beta,N} || \mu_{\beta,N})])^{1/2}$ , Corollary 3.1.11 and Markov's inequality then yield

$$\mathbb{P} \left( \frac{1}{N} \mathbf{d}(\tilde{\mu}_{\beta,N} || \mu_{\beta,N}) \leq \varepsilon_N \right) \geq 1 - \varepsilon_N^{1/2} \rightarrow 1, \quad \text{as } N \rightarrow \infty. \quad (3.1.12)$$

I present the proof of Corollary 3.1.11 below as it is short.

*Proof of Corollary 3.1.11.* Note that the choice of  $M_N$  satisfies the assumption of Theorem 3.1.10, so the first statement follows directly from Theorem 3.1.10. Next, as mentioned in Remark 3.1.8, the running time of Algorithm 3.1 is deterministic and is bounded by  $\lceil N/M_N \rceil 2^{M_N}$ . With our choice of  $M_N$ , we conclude that

$$\lceil N/M_N \rceil 2^{M_N} \leq N \cdot 2^{\varepsilon \log_2 N} \leq N^{1+\varepsilon},$$

and the proof is completed. ■

### Supercritical regime $\beta > \beta_G$ : hardness for generic algorithms

Assuming that  $A$  is non-concave,  $\beta_G < \infty$ . For  $\beta > \beta_G$ , I proved the following hardness result for the class of algorithms satisfying Definition 3.1.2.

**Theorem** (Theorem 3.1.13). *Suppose that  $A$  is non-concave. Let  $\beta > \beta_G$ . For any algorithm satisfying Definition 3.1.2 that approximates the Gibbs measure with probability approaching 1, there exists  $\gamma > 0$  such that*

$$\lim_{N \rightarrow \infty} \mathbb{P} (\tau \geq e^{\gamma N}) = 1,$$

where  $\tau$  is the running time of the algorithm.

Before diving in the proof of Theorem 3.1.13, a few definitions are required. The first is a chain of subtrees defined as follows.

**Definition** (Definition 3.5.1). For  $v \in \mathbb{T}$ , let  $\mathcal{C}_v$  be a chain of subtrees containing  $v$  and all its ancestors defined by

$$\mathcal{C}_v = \bigcup_{k=0}^{\lfloor N|v|/K \rfloor} \mathbb{T}_{\lfloor N(k+1)/K \rfloor - \lfloor Nk/K \rfloor}^{v[\lfloor Nk/K \rfloor]}.$$

Next, the following stopping time is considered.

**Definition** (Definition 3.5.3). Let  $(v(n))_{n \in \mathbb{N}}$  be an algorithm. Define the stopping time  $\tau'$  as the first time  $n$  when the algorithm finds a vertex in  $\mathcal{C}_{v(n)}$  with a  $(z, K, \mathbf{X})$ -steep ancestor, given by:

$$\tau' = \inf \{n \in \mathbb{N} : \exists w \in \mathcal{C}_{v(n)} \text{ such that } w \text{ has a } (z, K, \mathbf{X})\text{-steep ancestor}\}.$$

Back to Sketch of proof for Theorem 3.1.13. The proof is based on the following two propositions. The first proposition asserts that the running time of an algorithm that approximates the Gibbs measure dominates the stopping time  $\tau'$  with probability approaching 1.

**Proposition** (Proposition 3.5.4). *Suppose  $A$  to be non-concave. Let  $\beta > \beta_G$ . If  $\tau$  is the running time of an algorithm that approximates the Gibbs measure, then*

$$\lim_{N \rightarrow \infty} \mathbb{P}(\tau \geq \tau') = 1.$$

The proof of Proposition 3.5.4 is provided in Section 3.5.1. Note that Addario-Berry and Maillard proved in [2] a hardness result of finding a vertex  $v \in \partial \mathbb{T}_N$  such that  $X_v$  lies in a level set above a critical level, denoted by  $x_*N$ . In their case,  $\tau \geq \tau'$  holds deterministically because they showed in Lemma 3.1 in their paper that any for any vertex  $v \in \partial \mathbb{T}_N$  such that  $X_v$  lies in a level set above  $xN$ , where  $x > x_*$ ,  $v$  must have a  $(z, K, \mathbf{X})$ -ancestor. Nevertheless, Proposition 3.5.4 is sufficient to prove Theorem 3.1.13.

The second proposition to prove Theorem 3.1.13 is the following. The proposition asserts that the  $\tau'$  is exponentially large with probability approaching 1.

**Proposition** (Proposition 3.5.5). *There exists  $\gamma > 0$  such that*

$$\lim_{N \rightarrow \infty} \mathbb{P}(\tau' > e^{\gamma N}) = 1.$$

Proposition 3.5.5 is proven following the same argument as in [2], and the proof is included in Section 3.5.3 for completeness.

Conditioned on Proposition 3.5.4 and Proposition 3.5.5, the proof of Theorem 3.1.13 is fairly short, which is provided here.

*Proof of Theorem 3.1.13.* Fix  $\beta > \beta_G$ . Let  $(v(n))_{n \in \mathbb{N}}$  be an algorithm that approximates the Gibbs measure with probability approaching 1. Suppose that  $\tau$  is its running time and  $\tilde{\mu}_N$  is its output law, which is the law of  $v(\tau)$  conditioned on the CREM.

Now, combining Proposition 3.5.4 with Proposition 3.5.5, we conclude that there exists  $\gamma > 0$  such that

$$\lim_{N \rightarrow \infty} \mathbb{P}(\tau \geq e^{\gamma N}) \geq \lim_{N \rightarrow \infty} \mathbb{P}(\{\tau \geq \tau'\} \cap \{\tau' \geq e^{\gamma N}\}) = 1,$$

and the proof is completed. ■

### 1.5.9 Negative moments of the CREM partition function

In Chapter 4, I studied the partition function of the CREM in the high temperature regime  $\beta < \beta_c$ . I showed that for all  $\beta < \beta_c$  and for all  $s > 0$ , the negative  $s$  moment of the CREM partition function is comparable with the expectation of the CREM partition function to the power of  $-s$ , up to constants that are independent of  $N$ .

In the same paper, Bovier and Kurkova also showed that the maximum of the CREM satisfies the following.

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \left[ \max_{|u|=N} X_u \right] = \sqrt{2 \log 2} \int_0^1 \sqrt{\hat{A}'(s)} \, ds. \quad (4.1.3)$$

Combining (4.1.2) and (4.1.3) indicates that there exists

$$\beta_c := \frac{\sqrt{2 \log 2}}{\sqrt{\hat{A}'(0)}} \quad (1.5.4)$$

such that the following phase transition occurs. a) For all  $\beta < \beta_c$ , the main contribution to the partition function comes from an exponential amount of the particles. b) For all  $\beta > \beta_c$ , the maximum starts to contribute significantly to the partition function. The quantity  $\beta_c$  is sometimes referred to as the static critical inverse temperature of the CREM. In the following, the subcritical regime  $\beta < \beta_c$  is referred to as the high temperature regime.

### 1.5.10 Contributions in Chapter 4

Now I present the theorem that I obtained regarding the negative moments of the partition function  $Z_{\beta,N}$  in the high temperature regime.

**Theorem** (Theorem 4.1.2). *Suppose Assumption 4.1.1 is true. Let  $\beta < \beta_c$ . For all  $s > 0$ , there exist  $N_0 \in \mathbb{N}$  and a constant  $C = C(A, \beta, s)$ , independent of  $N$ , such that for all  $N \geq N_0$ ,*

$$\mathbb{E} [(Z_{\beta,N})^{-s}] \leq C \mathbb{E} [Z_{\beta,N}]^{-s}.$$

*Remark 1.5.6.* For all  $\beta > 0$ ,  $N \in \mathbb{N}$  and  $s > 0$ , there is the following trivial lower bound provided by the Jensen inequality and the convexity of  $x \mapsto x^{-s}$ ,

$$\mathbb{E} [(Z_{\beta,N})^{-s}] \geq \mathbb{E} [Z_{\beta,N}]^{-s}. \quad (4.1.5)$$

Thus, combining (4.1.5) with Theorem 4.1.2, one sees that for all  $\beta < \beta_c$  in the high temperature regime and for all  $s > 0$ ,  $\mathbb{E} [(Z_{\beta,N})^{-s}]$  is comparable with  $\mathbb{E} [Z_{\beta,N}]^{-s}$ .

**Summarizing the proof of Theorem 4.1.2.** To prove Theorem 4.1.2, I adapted the proof of Lemma A.3 in the paper [38] by Benjamini and Schramm so that it works not only for homogeneous models, but also for general CREM. In particular, I showed that for all  $s > 0$ , there exist two positive sequences  $\varepsilon_k$  and  $\eta_k$  that both decay double exponentially to 0 as  $k \rightarrow \infty$  such that for  $N$  sufficiently large,

$$\mathbb{P} (Z_{\beta,N} \leq \varepsilon_k \mathbb{E} [Z_{\beta,N}]) \leq \eta_k, \quad (4.1.7)$$

and that for all  $k \in \llbracket 1, C' \log N \rrbracket$ , where  $C' > 0$ , there exist  $C = C(A, \beta, s) > 0$  and  $c = c(A, \beta, s) > 0$  such that

$$\varepsilon_k^{-s} \eta_k \leq C e^{-ce^{ck}}. \quad (4.1.8)$$

The proof of (4.1.7) and (4.1.8) is by established an initial left tail estimate (Section 4.2), and then use the branching property to bootstrap the estimate (Section 4.3).

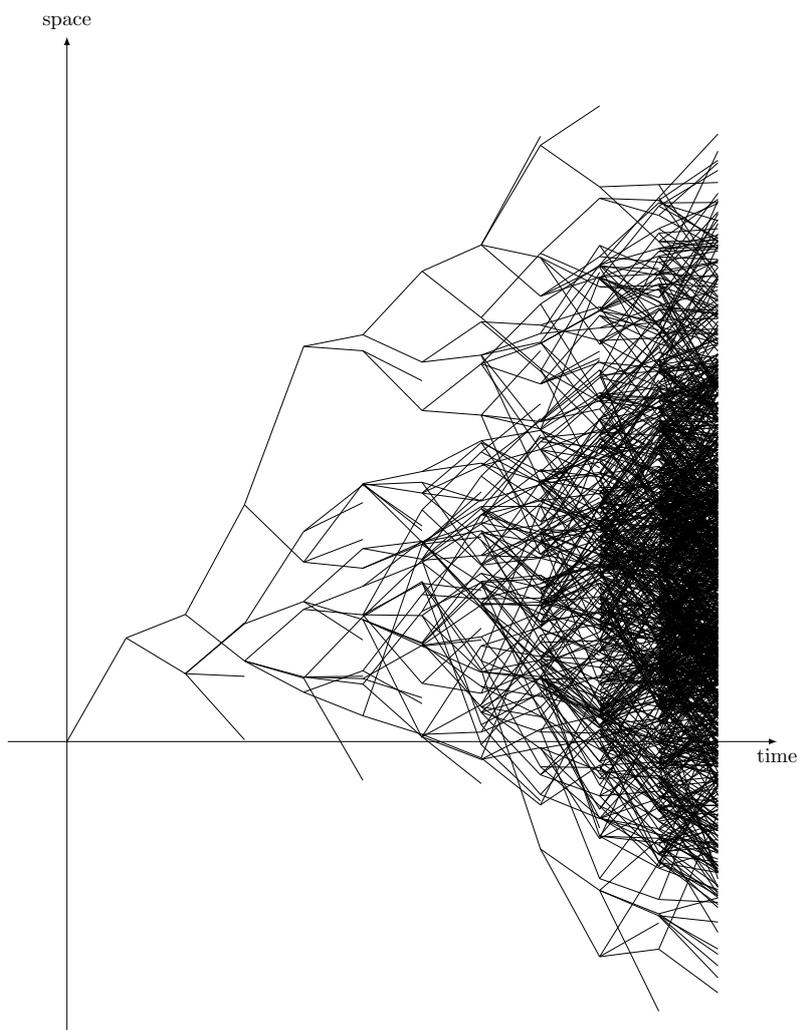


Figure 1.3: An example of a supercritical branching random walk with depth equal to 10. The law of increments follows a standard Gaussian distribution. Each vertex gives birth to 0, 2, 3 and 4 with probability 0.15, 0.65, 0.1 and 0.1, respectively, so  $EXL = 2$ .

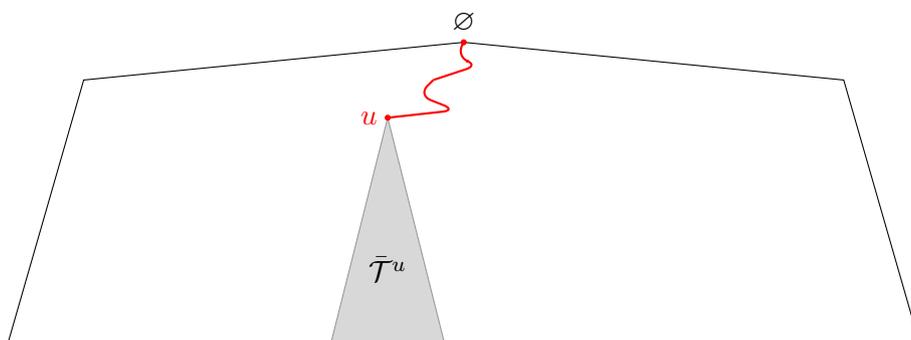


Figure 1.4: Schematic illustration of the set  $\bar{\mathcal{T}}^u$ .

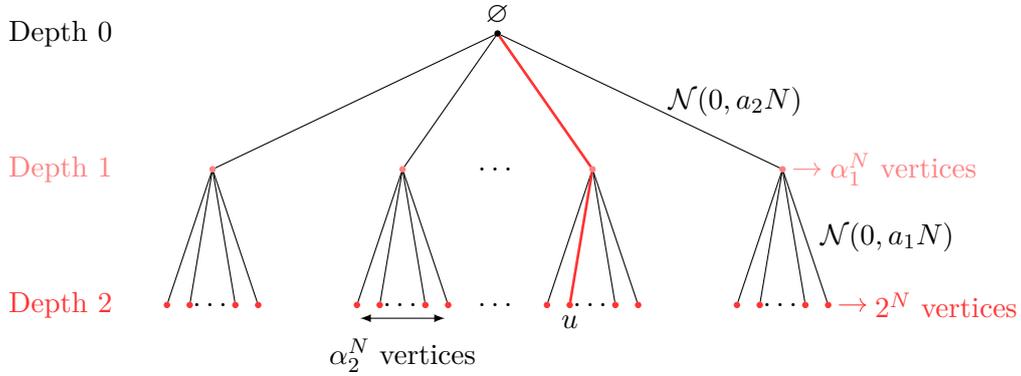


Figure 1.5: Schematic illustration of the covariance structure of the GREM with 2 hierarchies. Here, the parameters satisfy  $\alpha_1\alpha_2 = 2$  and  $a_1 + a_2 = 1$ . There are  $\alpha_1^N$  vertices with depth 1, and each vertex with depth 1 has  $\alpha_2^N$  children. Each edge represents an independent Gaussian random variables with zero mean and variance  $a_1N$  for the edges between depth 0 and depth 1, and variance  $a_2N$  for the edges between depth 1 and depth 2. For each  $|u| = N$ , the value of  $X_u$  is equal to the sum of Gaussian variables along the path from the root  $\emptyset$  to  $u$ , represented in red in the illustration.

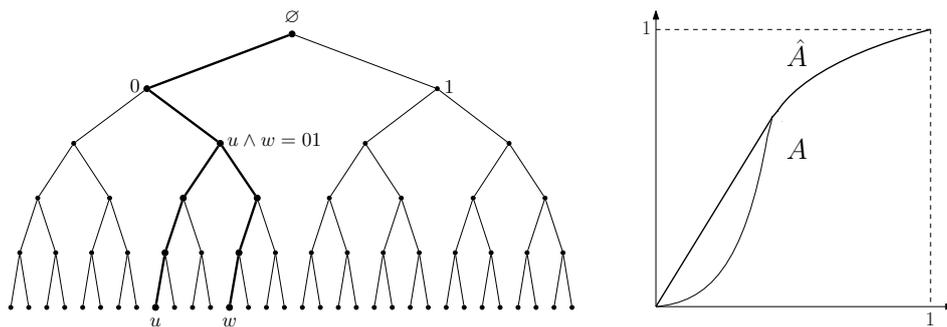


Figure 1.6: The covariance function of a CREM is determined by the underlying binary tree (left) and a function  $A$  (right). The concave hull  $\hat{A}$  of the function  $A$  is also shown on the right, which determines the CREM free energy and the asymptotics of the maximum of the CREM.

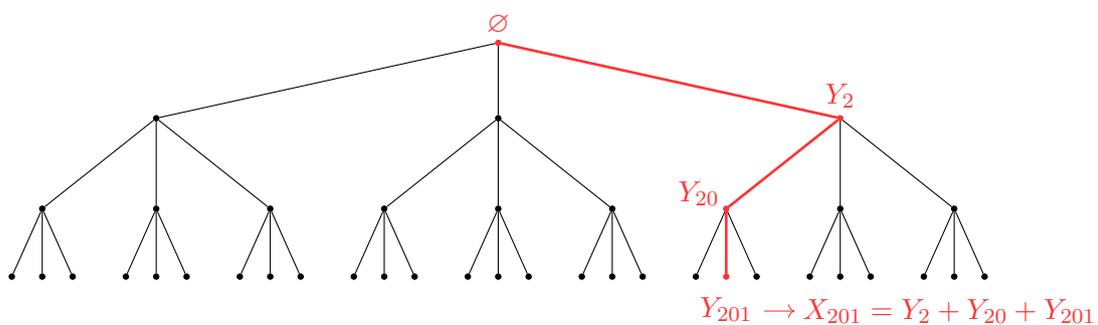


Figure 1.7: Schematic illustration of a branching random walk indexed by the ternary tree ( $d = 3$ ). Here, the value of  $X_{201}$  is calculated by summing the increments along the path from the root to the node indexed by 201. In other words,  $X_{201} = Y_2 + Y_{20} + Y_{201}$ .

# 2. Efficient sampling of branching random walk Gibbs measures

*The material in this chapter largely draws upon the joint paper [112] co-authored with Maillard. The title has been slightly adjusted and a few typos have been corrected. Please refer to Appendix A for a comprehensive list of these corrections.*

Disordered systems such as spin glasses have been used extensively as models for high-dimensional random landscapes and studied from the perspective of optimization algorithms. In a recent paper by L. Addario-Berry and the Maillard, the *continuous random energy model (CREM)* was proposed as a simple toy model to study the efficiency of such algorithms. The following question was raised in that paper: what is the threshold  $\beta_G$ , at which sampling (approximately) from the Gibbs measure at inverse temperature  $\beta$  becomes algorithmically hard?

This chapter is a first step towards answering this question. We consider the branching random walk, a time-homogeneous version of the continuous random energy model. We show that a simple greedy search on a renormalized tree yields a polynomial-time algorithm which approximately samples from the Gibbs measure, for every  $\beta < \beta_c$ , the (static) critical point. More precisely, we show that for every  $\varepsilon > 0$ , there exists such an algorithm such that the specific relative entropy between the law sampled by the algorithm and the Gibbs measure of inverse temperature  $\beta$  is less than  $\varepsilon$  with high probability.

In the supercritical regime  $\beta > \beta_c$ , we provide the following hardness result. Under a mild regularity condition, for every  $\delta > 0$ , there exists  $z > 0$  such that the running time of any given algorithm approximating the Gibbs measure stochastically dominates a geometric random variable with parameter  $e^{-z\sqrt{N}}$  on an event with probability at least  $1 - \delta$ .

## 2.1 Introduction

We consider the following family of branching random walks. An initial particle is located at the origin. It gives birth to  $d$  child particles,  $d \geq 2$ , scattering on the real line, and each of the child particles produces  $d$  child particles again, and so on. The displacement of each particle is independent of the past of the process and of

the displacements of its siblings. The genealogy of the particles can be represented by a  $d$ -ary tree  $\mathbb{T}_N$ , identifying particles with the vertices of the tree. We denote by  $X_v$  the location of a particle  $v \in \mathbb{T}_N$ .

Addario-Berry and Maillard [2] considered algorithms that, for a given  $x > 0$ , find a leaf  $v$  of  $\mathbb{T}_N$  such that  $X_v \geq xN$  in the framework of the continuous random energy model (CREM). This is a binary time-inhomogeneous branching random walk with Gaussian displacements. More precisely, the CREM is a Gaussian process whose covariance function is given by

$$\mathbb{E}[X_v X_w] = A\left(\frac{|v \wedge w|}{N}\right), \quad \forall v, w \in \mathbb{T}_n$$

where  $A : [0, 1] \rightarrow [0, 1]$  is an increasing function with  $A(0) = 0$  and  $A(1) = 1$  and  $|v \wedge w|$  is the depth of the most recent common ancestor of  $v$  and  $w$ . The authors proved the existence of a threshold  $x_*$  such that the following holds: a) for every  $x < x_*$ , there exists a polynomial-time algorithm that can accomplish the task with high probability, b) for every  $x > x_*$ , every such algorithm has a running time which is at least exponential in  $N$  with high probability. The authors also raised the question of the complexity of sampling a *typical* vertex of value roughly  $xN$ , which can be interpreted as sampling a vertex according to a Gibbs measure with a certain parameter  $\beta$  depending on  $x$ .

The present work attacks this problem in the simpler setting of the (homogeneous) branching random walk, corresponding to the case  $A(x) = x$  of the CREM in the special case of Gaussian displacements. The Gibbs measure is a probability measure on the leaves  $v$  of  $\mathbb{T}_N$  with weight proportional to  $e^{\beta X_v}$ , where  $\beta > 0$  is a given parameter called the *inverse temperature*. We show that there exists a threshold  $\beta_c > 0$  such that the following holds: a) in the subcritical regime  $\beta < \beta_c$ , there exists a polynomial-time algorithm such that with high probability, the specific relative entropy between the law sampled by the algorithm and the Gibbs measure of inverse temperature  $\beta$  is arbitrarily small, b) in the supercritical regime  $\beta > \beta_c$ , under a mild regularity condition, we show that with high probability, the running time of any given algorithm approximating the Gibbs measure in this sense is at least stretched exponential in  $N$ .

### 2.1.1 Notation and main results

Let  $\mathbb{T}$  be a rooted  $d$ -ary tree, where  $d \geq 2$ . The depth of a vertex  $v \in \mathbb{T}$  is denoted by  $|v|$ . We denote the root by  $\emptyset$ , and any vertex  $v$  with depth  $n \geq 1$  is indexed by a string  $v_1 \cdots v_n \in \{0, \dots, d-1\}^n$ . For any  $v, w \in \mathbb{T}$ , we write  $v \leq w$  if  $v$  is a prefix of  $w$  and write  $v < w$  if  $v$  is a prefix of  $w$  strictly shorter than  $w$ . We denote  $\mathbb{T}_n$  to be the subtree of  $\mathbb{T}$  containing vertices of depth less or equal to  $n$  and  $\partial\mathbb{T}_n$  to be the leaves of  $\mathbb{T}_n$ .

Let  $\mathbf{Y} = (Y_0, \dots, Y_{d-1})$  be a  $d$ -dimensional random vector. Let  $(\mathbf{Y}^v)_{v \in \mathbb{T}}$  be i.i.d. copies of  $\mathbf{Y}$  where  $\mathbf{Y}^v = (Y_{v0}, \dots, Y_{v(d-1)})$  — this uniquely defines  $Y_u$  for

every  $u \in \mathbb{T} \setminus \{\emptyset\}$ . Define the process  $\mathbf{X} = (X_v)_{v \in \mathbb{T}}$  by

$$\begin{cases} X_\emptyset = 0, \\ X_v = \sum_{\emptyset < w \leq v} Y_w, \quad |v| \geq 1. \end{cases}$$

The process  $\mathbf{X}$  is called the *branching random walk with increments*  $\mathbf{Y}$ .

It is well-known that  $\mathbf{X}$  has the branching property: let  $\mathcal{F} = (\mathcal{F}_n)_{n \geq 0}$  be its natural filtration. For any  $v \in \mathbb{T}$  with  $|v| = n$ , define

$$\mathbf{X}^v = (X_w^v)_{|w| \geq 0} = (X_{vw} - X_v)_{|w| \geq 0}.$$

Then  $(\mathbf{X}^v)_{|v|=n}$  are i.i.d. copies of  $\mathbf{X}$  and independent of  $\mathcal{F}_n$ .

**Gibbs measures.** Define the following function

$$\varphi(\beta) = \log \left( \mathbb{E} \left[ \sum_{i=0}^{d-1} e^{\beta Y_i} \right] \right) \in (-\infty, \infty], \quad \beta \in \mathbb{R},$$

and set  $\mathcal{D}(\varphi) = \{\beta \in \mathbb{R} : \varphi(\beta) < \infty\}$ . It is well known that  $\varphi$  is convex and that it is smooth on  $\mathcal{D}(\varphi)^\circ$ , the interior of  $\mathcal{D}(\varphi)$ . Throughout the article, we assume the following.

**Assumption 2.1.1.**  $0 \in \mathcal{D}(\varphi)^\circ$ .

For  $n \in \mathbb{N}$ , define the following (normalized) partition functions

$$\begin{aligned} W_{\beta,n} &= \sum_{|v|=n} e^{\beta X_v - \varphi(\beta)n}, \quad n \geq 0, \\ W_{\beta,n}^u &= \sum_{|w|=n} e^{\beta X_w^u - \varphi(\beta)n}, \quad u \in \mathbb{T} \text{ and } n \geq 0. \end{aligned}$$

Note that for every  $m \leq n$ ,

$$W_{\beta,n} = \sum_{|u|=m} e^{\beta X_u - \varphi(\beta)m} \cdot W_{\beta,n-m}^u. \quad (2.1.1)$$

For  $\beta \in \mathbb{R}$  and  $n \in \mathbb{N}$ , we now define the Gibbs measure of parameter  $\beta$  on  $\mathbb{T}_n$  to be

$$\mu_{\beta,n}(u) = e^{\beta X_u - \varphi(\beta)m} \cdot \frac{W_{\beta,n-m}^u}{W_{\beta,n}}, \quad |u| = m \leq n. \quad (2.1.2)$$

Note that  $\mu_{\beta,n}$  is usually defined on  $\partial\mathbb{T}_n$  only, but it will be helpful to define it on the whole tree  $\mathbb{T}_n$ . By (2.1.1), for every  $m \leq n$ , the restriction of  $\mu_{\beta,n}$  to  $\partial\mathbb{T}_m$  is a probability measure. Similarly, we can define

$$\mu_{\beta,n}^v(u) = e^{\beta X_u^v - \varphi(\beta)m} \cdot \frac{W_{\beta,n-m}^{vu}}{W_{\beta,n}^v}, \quad v \in \mathbb{T} \text{ and } |u| = m \leq n. \quad (2.1.3)$$

The free energy of the branching random walk has been calculated by Derrida and Spohn [91] (and can also be deduced from Biggins [46])

$$\lim_{n \rightarrow \infty} \frac{1}{\beta n} \log \sum_{|v|=n} e^{\beta X_v} = \begin{cases} \frac{1}{\beta} \varphi(\beta) & \text{if } \beta \in (0, \beta_c) \\ \frac{1}{\beta_c} \varphi(\beta_c) & \text{if } \beta \geq \beta_c, \end{cases}$$

where the limit is meant to be in probability and where the critical inverse temperature  $\beta_c$  is defined by

$$\beta_c = \sup\{\beta \in \mathcal{D}(\varphi)^\circ \mid \beta \varphi'(\beta) < \varphi(\beta)\} \in (0, \infty].$$

We will mostly be interested in the phase  $\beta < \beta_c$ . In this phase, we recall the following result.

**Fact 2.1.2** ([118, 47, 129]). *If  $\beta < \beta_c$ , then the martingale  $(W_{\beta,n})_{n \geq 0}$  is uniformly integrable. In fact, there exists a (strictly) positive random variable  $W_{\beta,\infty}$  with  $\mathbb{E}[W_{\beta,\infty}] = 1$  and such that  $W_{\beta,n} \rightarrow W_{\beta,\infty}$  almost surely and in  $L^1$  as  $n \rightarrow \infty$ .*

**Algorithms.** We define an algorithmic model similar to the ones in [159, 2]. Let  $N \in \mathbb{N}$ . A random sequence  $\mathbf{v} = (v(k))_{k \geq 0}$  taking values in  $\mathbb{T}_N$  is called a (randomized) algorithm if  $v(0) = \emptyset$  and  $v(k+1)$  is  $\tilde{\mathcal{F}}_k$ -measurable for every  $k \geq 0$ . Here,

$$\tilde{\mathcal{F}}_k = \sigma(v(1), \dots, v(k); X_{v(1)}, \dots, X_{v(k)}; U_1, \dots, U_{k+1})$$

where  $(U_k)_{k \geq 1}$  is a sequence of i.i.d. uniform random variables on  $[0, 1]$ , independent of the branching random walk  $\mathbf{X}$ . Roughly speaking, the filtration  $\tilde{\mathcal{F}} = (\tilde{\mathcal{F}}_k)_{k \geq 0}$  contains all information about everything we have queried so far, as well as the additional randomness needed to choose the next vertex. We further suppose that there exists a stopping time  $\tau$  with respect to the filtration  $\tilde{\mathcal{F}}$  and such that  $v(\tau) \in \partial \mathbb{T}_N$ . We call  $\tau$  the *running time* and  $v(\tau)$  the *output* of the algorithm. The *law of the output* is the (random) distribution of  $v(\tau)$ , conditioned on the branching random walk.

Often, we will consider a family of algorithms indexed by  $N$ , which we also call an algorithm by abuse of notation. We say that the algorithm is a *polynomial-time* algorithm if there exists a (deterministic) polynomial  $P(N)$  such that  $\tau \leq P(N)$  almost surely.

Fix  $M \in \mathbb{N}$ . Given a configuration of the branching random walk of depth  $N$ , consider the following algorithm (See Figure 2.1.1):

---

**Algorithm 2.1:** Recursive sampling on  $M$ -renormalized tree

---

```

set  $v = \emptyset$    while  $|v| < N$  do
  | choose  $w$  with  $|w| = M \wedge (N - |v|)$  according to the Gibbs measure
  |  $\mu_{\beta, M \wedge (N - |v|)}^v$    replace  $v$  with  $vw$ 
output  $v$ 

```

---

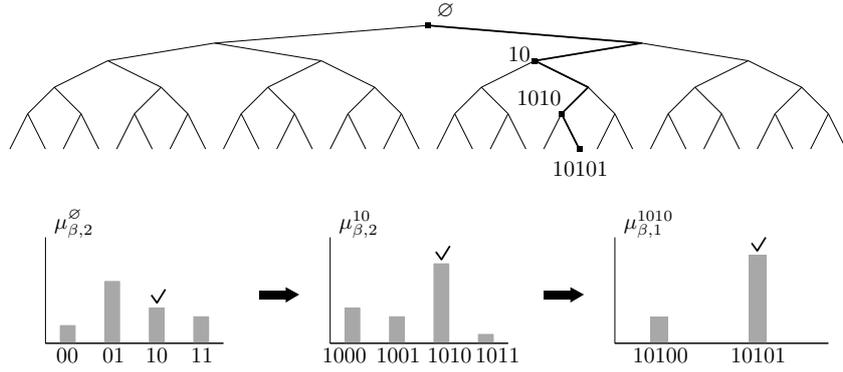


Figure 2.1: Schematic illustration of Algorithm 2.1 with  $N = 5$  and  $M = 2$  on a given configuration of the binary branching random walk. The squares are the vertices that the algorithm chooses at each step, and the thick line is the path from the root to the sampled vertex. The bar charts represent the Gibbs measures from which the algorithm samples at each step, and the check mark indicates which vertex is chosen. Note that the histograms are not drawn to scale.

*Remark 2.1.3.* It is easy to see that Algorithm 2.1 can be formally written as a randomized algorithm according to the above algorithmic model. Furthermore, its running time is deterministic and bounded by  $\lceil N/M \rceil d^M$ . The law of its output is a random probability measure  $\mu_{\beta,M,N}$  on  $\partial\mathbb{T}_N$  that can be recursively defined as follows:

$$\begin{aligned} \mu_{\beta,M,0}(\emptyset) &= 1 \\ \mu_{\beta,M,N \wedge (K+1)M}(vw) &= \mu_{\beta,M,KM}(v) \cdot \mu_{\beta,M \wedge (N-KM)}^v(w) \end{aligned} \tag{2.1.4}$$

for all  $|v| = KM$ ,  $|w| = M \wedge (N - KM)$  and  $0 \leq K \leq \lfloor \frac{N}{M} \rfloor$ .

**Approximation and threshold.** Given two probability measures  $P$  and  $Q$  defined on a discrete space  $\Omega$ , the entropy of  $Q$  and the Kullback–Leibler divergence from  $Q$  to  $P$  are respectively defined by

$$H(Q) = \sum_{\omega \in \Omega} Q(\omega) \cdot \log \left( \frac{1}{Q(\omega)} \right) \tag{2.1.5}$$

$$\mathbf{d}(P||Q) = \sum_{\omega \in \Omega} P(\omega) \cdot \log \left( \frac{P(\omega)}{Q(\omega)} \right). \tag{2.1.6}$$

Note that by Jensen’s inequality, the entropy and the Kullback–Leibler divergence are non-negative. In what follows, we will often take  $P$  and/or  $Q$  to be a Gibbs measure  $\mu_{\beta,N}$  for some  $N$ . In that case, we set  $\Omega = \partial\mathbb{T}_n$ , where  $n$  is the largest number such that both  $P$  and  $Q$  are defined on  $\partial\mathbb{T}_n$ .

The following lemma is folklore. For completeness, we provide a proof in Section 2.6.

**Lemma 2.1.4.**

1. If  $\beta \in [0, \beta_c)$ , then  $H(\mu_{\beta,N})/N$  converges in probability to a positive constant as  $N \rightarrow \infty$ .
2. If  $\beta > \beta_c$  and if  $\beta_c \in \mathcal{D}(\varphi)^\circ$ , then  $H(\mu_{\beta,N}) = O(1)$  in probability, as  $N \rightarrow \infty$ . In other words, the sequence of random variables  $(H(\mu_{\beta,N}))_{N \geq 1}$  is tight.

We are interested in the *algorithmically efficient approximation* of the Gibbs measure  $\mu_{\beta,N}$ . The notion of approximation we will use is the following.

**Definition 2.1.5.** Let  $\beta \geq 0$ . We say that a sequence of random probability measures  $(\tilde{\mu}_{\beta,N})_{N \geq 1}$  *approximates* the Gibbs measure  $\mu_{\beta,N}$  if

$$\frac{\mathbf{d}(\tilde{\mu}_{\beta,N} || \mu_{\beta,N})}{H(\mu_{\beta,N})} \rightarrow 0 \quad \text{in probability as } N \rightarrow \infty. \quad (2.1.7)$$

*Remark 2.1.6.* More generally, assume that we are given two sequences of probability measures  $(P_N)_{N \geq 1}$  and  $(Q_N)_{N \geq 1}$  satisfying  $H(Q_N)/N \rightarrow C \in (0, \infty)$  and

$$\frac{1}{N} \mathbf{d}(P_N || Q_N) \rightarrow 0 \quad \text{as } N \rightarrow \infty. \quad (2.1.8)$$

This has been called *measure equivalence* or *equivalence in the sense of specific relative entropy* in the physics literature [184]. Mathematically, Equation (2.1.8) implies the following: if  $(A_N)_{N \geq 1}$  is a sequence of sets such that  $Q_N(A_N)$  converges to 0 exponentially fast as  $N \rightarrow \infty$ , then we also have  $P_N(A_N) \rightarrow 0$ . Indeed, this is an easy consequence of Birgé's inequality (see e.g. Theorem 4.20 in [52]).

**Main results.** We now state the main theorem of this paper.

**Theorem 2.1.7** (Approximation bounds). *Let  $N \in \mathbb{N}$ ,  $M \in \llbracket 1, N \rrbracket$ , and  $\beta \in [0, \beta_c)$ . Then for all  $p \geq 1$ , there exists a constant  $C_1(p) > 0$  such that*

$$\|\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N})\|_p \leq C_1(p) \cdot \left\lfloor \frac{N}{M} \right\rfloor. \quad (2.1.9)$$

Moreover, for all  $p \geq 1$ , there exists a constant  $C_1(p) > 0$  such that

$$\|\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N}) - \mathbb{E}[\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N})]\|_p \leq C_1(p). \quad (2.1.10)$$

By Theorem 2.1.7, we can derive the following corollary which states the existence of an algorithm that approximates the Gibbs measure  $\mu_{\beta,N}$  efficiently.

**Corollary 2.1.8** (Complexity upper bound). *If  $\beta \in [0, \beta_c)$ , then there exists a polynomial-time algorithm such that for every  $p > 1$ , denoting by  $\tilde{\mu}_{\beta,N}$  the law of its output,*

$$\frac{1}{N} \|\mathbf{d}(\tilde{\mu}_{\beta,N} || \mu_{\beta,N})\|_p \rightarrow 0 \quad (2.1.11)$$

as  $N \rightarrow \infty$ . In particular,  $\tilde{\mu}_{\beta,N}$  approximates the Gibbs measure  $\mu_{\beta,N}$  in the sense of Definition 2.1.5.

*Proof.* Let  $M = M(N)$  be a sequence that goes to infinity as  $N \rightarrow \infty$ , and set  $\tilde{\mu}_{\beta,N} = \mu_{\beta,M,N}$ . Equation (2.1.11) then follows from (2.1.9) in Theorem 2.1.7. Assuming moreover that  $M = O(\log N)$ , Remark 2.1.3 implies that  $\tilde{\mu}_{\beta,N}$  can be computed by a polynomial-time algorithm. The second statement follows from the first part of Lemma 2.1.4. ■

Finally, we also provide a hardness result, assuming a mild regularity condition.

**Theorem 2.1.9** (Complexity lower bound). *Assume  $\beta_c \in \mathcal{D}(\varphi)^\circ$  (in particular,  $\beta_c < \infty$ ). Let  $\beta > \beta_c$ . Let  $v = (v(k))_{k \geq 0}$  be an algorithm which outputs a vertex of law  $\tilde{\mu}_N$  such that  $\tilde{\mu}_N$  approximates the Gibbs measure  $\mu_{\beta,N}$  in the sense of Definition 2.1.5. Let  $\tau$  be the running time of the algorithm. Then for every  $\delta > 0$ , there exists  $z > 0$ , such that for large enough  $N$ ,*

$$\mathbb{P}\left(\tau \geq e^{z\sqrt{N}}\right) \geq 1 - \delta.$$

### 2.1.2 Related work

An early study on searching algorithms on the branching random walk can be found in Karp and Pearl [119]. They considered the binary branching random walk with Bernoulli increments  $\text{Ber}(p)$ , and they showed that for  $p > 1/2$  and  $p = 1/2$ , they gave an algorithm that can find an exact maximal vertex in linear and quadratic expected time, respectively. While for  $p < 1/2$ , it is possible to find an approximate maximal vertex in linear time with high probability using a depth-first search on a renormalized tree. Aldous [11] gave a different algorithm and, among other things, extended the result of Karp and Pearl [119] to general increment distributions. A hardness result was obtained by Pemantle [159]. Among other things, he showed for the binary branching random walk with Bernoulli increments with mean  $p < 1/2$  that any search algorithm which finds a vertex within a  $(1 - \varepsilon)$  factor of the maximum with high probability needs at least  $N \times \exp(\Theta(\varepsilon^{-1/2}))$  with high probability.<sup>1</sup>

As mentioned above, Addario-Berry and Maillard [2] considered this optimization problem for the continuous random energy model (CREM), which is a binary time-inhomogeneous branching random walk with Gaussian displacements, proving the existence of an threshold  $x_*$  such that the following holds: a) for every  $x < x_*$ , there exists a polynomial-time algorithm that finds a vertex with  $X_v \geq xN$  with high probability, b) for every  $x > x_*$ , every such algorithm has a running time which is at least exponential in  $N$  with high probability.

The CREM, introduced by Bovier and Kurkova [54] based on previous work by Derrida and Spohn [91], is a toy model of a disordered system in statistical physics, i.e. a model where the Hamiltonian — the function that assigns energies to the states of the system — is itself random. These systems have recently seen a lot of interest in the mathematical literature with regards to efficient algorithms for

<sup>1</sup>To be precise, this explicit bound relied on a conjecture on branching random walk killed at a linear space-time barrier (Conjecture 1 in [159]), which was subsequently proven to be true [39, 106].

finding low-energy states. A key quantity of importance in these models is the so-called *overlap* between two states, a measure of their correlation. In the case of the CREM, it is equal to the depth of the most recent common ancestor of two vertices, divided by  $N$ . Then, for a given  $\beta > 0$ , the *overlap distribution* is the limiting law (as  $N \rightarrow \infty$ ) of the overlap of two vertices sampled independently according to the Gibbs measure with inverse temperature  $\beta$ . A picture that has emerged is that the existence of a gap in the support of the overlap distribution, the so-called “overlap gap property”, is an obstruction to the existence of efficient algorithms finding approximate minimizers of the Hamiltonian. This has been rigorously proven for a certain class of algorithms by Gamarnik and Jagannath [103] in the case of the Ising  $p$ -spin model, with  $p \geq 4$ . On the other hand, Montanari [145] showed that for  $p = 2$ , the Sherrington–Kirkpatrick model, there exists a quadratic time algorithm that can find near optimal state with high probability, assuming a widely believed conjecture that this model does not exhibit an overlap gap. A similar result for spherical spin glass models (for which the overlap distribution is explicitly known) has been obtained by Subag [171].

The question of efficient *sampling* of the Gibbs measure of a disordered system seems to have been considered mostly under the angle of Glauber or Langevin dynamics. See e.g. [37, 108] for the spherical spin glass model. We restrict ourselves here to the case of the Sherrington–Kirkpatrick model. For this model, it has been recently obtained that fast mixing occurs for  $\beta < 1/4$  [28, 98]. However, very recently, El Alaoui, Montanari and Sellke [9] have provided another algorithm which yields fast mixing for  $\beta < 1/2$ , and they conjecture that this in fact holds for all  $\beta < 1$ . They also provide a hardness result for  $\beta > 1$  for a certain class of algorithms. Their algorithm for the  $\beta < 1/2$  phase belongs to the class of *approximate message passing* algorithms, which is also the case for Montanari’s algorithm for the optimization problem [145]. This illustrates the fact that Glauber or Langevin dynamics may in general not be optimal sampling algorithms, and that algorithms which exploit the underlying tree structure of the model may be efficient in a wider range of the parameters. For a discussion of this question in the context of statistical inference problems, see e.g. [15]. Altogether, this motivates the study of tree-based models as a toy problem, such as the one from the present article.

**Outline.** The paper is organized as follows. In Section 2.2, we prove that the Kullback–Leibler divergence can be decomposed into a weighted sum of the Kullback–Leibler divergences on subtrees. In Section 2.3, we give  $L^p$  bounds of the logarithm of Biggins’ martingales and  $L^p$  bounds of the Kullback–Leibler divergences between two Gibbs measures. Theorem 2.1.7 is proven in Section 2.4 and Theorem 2.1.9 in Section 2.5. Section 2.6 provides the proof of Lemma 2.1.4. Finally, we state in Section 2.7 some open questions that might interest the readers.

## 2.2 Decomposition of the Kullback–Leibler divergence

The main goal of this section is to prove Theorem 2.2.2. Before proving the theorem, we need the following lemma, which states that the weight of  $u_1 u_2$  with

respect to the Gibbs measure  $\mu_{\beta,m}$  can be decomposed into the product of the weights of  $u_1$  and  $u_2$  with respect to another two Gibbs measures.

**Lemma 2.2.1.** *For any  $|u_1| = m_1$  and  $|u_2| = m_2$ , we have the decomposition*

$$\mu_{\beta,m}(u_1 u_2) = \mu_{\beta,m}(u_1) \cdot \mu_{\beta,m-m_1}^{u_1}(u_2)$$

for all  $m \geq m_1 + m_2$ .

*Proof.* By computation, we have

$$\begin{aligned} \mu_{\beta,m}(u_1 u_2) &= \frac{e^{\beta X_{u_1 u_2} - \varphi(\beta)m} \cdot W_{\beta,m-m_1-m_2}^{u_1 u_2}}{W_{\beta,m}} \\ &= \frac{e^{\beta X_{u_1} - \varphi(\beta)m_1} \cdot W_{\beta,m-m_1}^{u_1}}{W_{\beta,m}} \cdot \frac{e^{\beta X_{u_2}^{u_1} - \varphi(\beta)(m-m_1)} \cdot W_{\beta,m-m_1-m_2}^{u_1 u_2}}{W_{\beta,m-m_1}^{u_1}} \\ &= \mu_{\beta,m}(u_1) \cdot \mu_{\beta,m-m_1}^{u_1}(u_2), \end{aligned}$$

where the last equality is by (2.1.2) and (2.1.3). ■

Now we can decompose the Kullback–Leibler divergence as follows.

**Theorem 2.2.2.** *For any two  $M$  and  $N$  integers such that  $M \leq N$ , we have*

$$\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N}) = \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot \mathbf{d}(\mu_{\beta,M}^u || \mu_{\beta,N-KM}^u). \quad (2.2.1)$$

*Proof.* Denote  $N' = \lfloor \frac{N}{M} \rfloor \cdot M$  for simplicity. By (2.1.4) and Lemma 2.2.1, for all  $|u_1| = N'$  and  $|u_2| = N - N'$ , we have

$$\log \left( \frac{\mu_{\beta,M,N}(u_1 u_2)}{\mu_{\beta,N}(u_1 u_2)} \right) = \log \left( \frac{\mu_{\beta,M,N'}(u_1) \cdot \mu_{\beta,N-N'}^{u_1}(u_2)}{\mu_{\beta,N}(u_1) \cdot \mu_{\beta,N-N'}^{u_1}(u_2)} \right) = \log \left( \frac{\mu_{\beta,M,N'}(u_1)}{\mu_{\beta,M,N}(u_1)} \right). \quad (2.2.2)$$

Thus, the Kullback–Leibler divergence can be rewritten as

$$\begin{aligned} &\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N}) \\ &= \sum_{|u|=N} \mu_{\beta,M,N}(u) \cdot \log \left( \frac{\mu_{\beta,M,N}(u)}{\mu_{\beta,N}(u)} \right) \\ &= \sum_{|u_1|=N'} \sum_{|u_2|=N-N'} \mu_{\beta,M,N}(u_1 u_2) \cdot \log \left( \frac{\mu_{\beta,M,N}(u_1 u_2)}{\mu_{\beta,N}(u_1 u_2)} \right) \\ &= \sum_{|u_1|=N'} \sum_{|u_2|=N-N'} \mu_{\beta,M,N'}(u_1) \cdot \mu_{\beta,N-N'}^{u_1}(u_2) \cdot \log \left( \frac{\mu_{\beta,M,N'}(u_1)}{\mu_{\beta,M,N}(u_1)} \right) \quad (\text{by (2.1.4) and (2.2.2)}) \\ &= \sum_{|u_1|=N'} \mu_{\beta,M,N'}(u_1) \cdot \log \left( \frac{\mu_{\beta,M,N'}(u_1)}{\mu_{\beta,M,N}(u_1)} \right) \\ &= \mathbf{d}(\mu_{\beta,M,N'} || \mu_{\beta,N}). \end{aligned} \quad (2.2.3)$$

Next, by (2.1.4) and Lemma 2.2.1, for all  $|u_1| = KM$ ,  $|u_2| = M$  and  $0 \leq K \leq \lfloor \frac{N}{M} \rfloor - 1$ , we have

$$\begin{aligned} \log \left( \frac{\mu_{\beta, M, (K+1)M}(u_1 u_2)}{\mu_{\beta, N}(u_1 u_2)} \right) &= \log \left( \frac{\mu_{\beta, M, KM}(u_1) \cdot \mu_{\beta, M}^{u_1}(u_2)}{\mu_{\beta, N}(u_1) \cdot \mu_{\beta, N-KM}^{u_1}(u_2)} \right) \\ &= \log \left( \frac{\mu_{\beta, M, KM}(u_1)}{\mu_{\beta, N}(u_1)} \right) + \log \left( \frac{\mu_{\beta, M}^{u_1}(u_2)}{\mu_{\beta, N-KM}^{u_1}(u_2)} \right). \end{aligned} \quad (2.2.4)$$

Thus,

$$\begin{aligned} &\mathbf{d}(\mu_{\beta, M, (K+1)M} \| \mu_{\beta, N}) \\ &= \sum_{|u|=(K+1)M} \mu_{\beta, M, (K+1)M}(u) \cdot \log \left( \frac{\mu_{\beta, M, (K+1)M}(u)}{\mu_{\beta, N}(u)} \right) \\ &= \sum_{|u_1|=KM} \sum_{|u_2|=M} \mu_{\beta, M, KM}(u_1) \\ &\quad \cdot \mu_{\beta, M}^{u_1}(u_2) \cdot \left[ \log \left( \frac{\mu_{\beta, M, KM}(u_1)}{\mu_{\beta, N}(u_1)} \right) + \log \left( \frac{\mu_{\beta, M}^{u_1}(u_2)}{\mu_{\beta, N-KM}^{u_1}(u_2)} \right) \right] \quad (\text{by (2.1.4) and (2.2.4)}) \\ &= \mathbf{d}(\mu_{\beta, M, KM} \| \mu_{\beta, N}) + \sum_{|u_1|=KM} \mu_{\beta, M, KM}(u_1) \cdot \mathbf{d}(\mu_{\beta, M}^{u_1} \| \mu_{\beta, N-KM}^{u_1}). \end{aligned} \quad (2.2.5)$$

Finally, by (2.2.3), (2.2.5) and the fact that  $\mathbf{d}(\mu_{\beta, M, 0} \| \mu_{\beta, N}) = 0$ , we derive (2.2.1).  $\blacksquare$

Next, we show that the Kullback–Leibler divergence between two Gibbs measures can be written in terms of the logarithms of the partition functions.

**Proposition 2.2.3.** *For any two  $M$  and  $N$  integers such that  $M \leq N$ , we have*

$$\mathbf{d}(\mu_{\beta, M} \| \mu_{\beta, N}) = \log W_{\beta, N} - \log W_{\beta, M} - \sum_{|u|=M} \mu_{\beta, M}(u) \cdot \log W_{\beta, N-M}^u. \quad (2.2.6)$$

*Proof.* For any  $|u| = M$ , we have

$$\frac{\mu_{\beta, M}(u)}{\mu_{\beta, N}(u)} = \frac{e^{\beta X_u - \varphi(\beta)M} \cdot \frac{1}{W_{\beta, M}}}{e^{\beta X_u - \varphi(\beta)M} \cdot \frac{W_{\beta, N-M}^u}{W_{\beta, N}}} = \frac{W_{\beta, N}}{W_{\beta, M} \cdot W_{\beta, N-M}^u}.$$

Thus,

$$\begin{aligned} \mathbf{d}(\mu_{\beta, M} \| \mu_{\beta, N}) &= \sum_{|u|=M} \mu_{\beta, M}(u) \cdot \log \left( \frac{\mu_{\beta, M}(u)}{\mu_{\beta, N}(u)} \right) \\ &= \sum_{|u|=M} \mu_{\beta, M}(u) \cdot (\log W_{\beta, N} - \log W_{\beta, M} - \log W_{\beta, N-M}^u) \\ &= \log W_{\beta, N} - \log W_{\beta, M} - \sum_{|u|=M} \mu_{\beta, M}(u) \cdot \log W_{\beta, N-M}^u. \end{aligned}$$

This completes the proof.  $\blacksquare$

### 2.3 Some $L^p$ bounds

We first show that whenever  $\beta \in [0, \beta_c)$ ,  $(\log W_{\beta,n})_{n \geq 0}$  is bounded in  $L^p$  for all  $p > 1$ .

**Lemma 2.3.1.** *Let  $\beta \in [0, \beta_c)$ . Then  $(\log W_{\beta,n})_{n \geq 0}$  is a supermartingale such that*

$$\sup_{n \geq 0} \|\log W_{\beta,n}\|_p < \infty, \quad p \geq 1.$$

*Proof.* The supermartingale property of  $(\log W_{\beta,n})_{n \geq 0}$  follows from the fact that  $(W_{\beta,n})_{n \geq 0}$  is a martingale and  $x \mapsto \log x$  is a concave function.

Now let  $p \geq 1$ . Since  $\beta < \beta_c$ , by Fact 2.1.2, we derive that

$$\sup_{n \geq 0} \|W_{\beta,n}\|_1 < \infty$$

Furthermore, using Assumption 2.1.1 and Liu [127, Theorem 2.4], we have for some  $s > 0$ ,

$$\sup_{n \geq 0} \|W_{\beta,n}^{-s}\|_1 < \infty.$$

Then by the fact that

$$|\log x|^p \leq C (|x| + |x|^{-s})$$

for some constant  $C > 0$ , we have

$$\sup_{n \geq 0} \|\log W_{\beta,n}\|_p^p \leq C \cdot \sup_{n \geq 0} \left( \|W_{\beta,n}\|_1 + \|W_{\beta,n}^{-s}\|_1 \right) < \infty.$$

This proves the lemma. ■

Lemma 2.3.1 implies the following proposition about the boundedness of the Kullback–Leibler divergence between two Gibbs measures.

**Proposition 2.3.2.** *For any  $p > 1$ , for any two integers  $M$  and  $N$  such that  $M \leq N$ , there exists a constant  $C(p) > 0$  such that*

$$\|\mathbf{d}(\mu_{\beta,M} \|\mu_{\beta,N})\|_p \leq C(p).$$

*Proof.* By Minkowski's inequality and Proposition 2.2.3, we have

$$\|\mathbf{d}(\mu_{\beta,M} \|\mu_{\beta,N})\|_p \leq \|\log W_{\beta,N}\|_p + \|\log W_{\beta,M}\|_p + \left\| \sum_{|u|=M} \mu_{\beta,M}(u) \cdot \log W_{\beta,N-M}^u \right\|_p.$$

Since the first and the second term above are bounded by Lemma 2.3.1, it suffices to prove that the third term above is bounded. By the branching property and Jensen's inequality, we have

$$\mathbb{E} \left[ \left\| \sum_{|u|=M} \mu_{\beta,M}(u) \cdot \log W_{\beta,M}^u \right\|_p^p \middle| \mathcal{F}_M \right] \leq \mathbb{E} [|\log W_{\beta,N-M}|^p]. \quad (2.3.1)$$

Therefore, by (2.3.1),

$$\left\| \sum_{|u|=M} \mu_{\beta,M}(u) \cdot \log W_{\beta,N-M}^u \right\|_p \leq \|\log W_{\beta,N-M}\|_p. \quad (2.3.2)$$

Finally, we conclude by (2.3.2) and Lemma 2.3.1 that

$$\|\mathbf{d}(\mu_{\beta,M} \parallel \mu_{\beta,N})\|_p \leq 3 \cdot \sup_{n \geq 0} \|\log W_{\beta,n}\|_p < \infty,$$

and the proof is completed.  $\blacksquare$

## 2.4 Proof of Theorem 2.1.7

In this section, we prove (2.1.9) and (2.1.10) of Theorem 2.1.7. The proof of (2.1.9) relies essentially on the decomposition theorem of the Kullback–Leibler divergence (Theorem 2.2.2) and Proposition 2.3.2. The proof of (2.1.10) needs more precise moment estimates.

### 2.4.1 Proof of (2.1.9)

Let  $p \geq 1$ . By Theorem 2.2.2 and Minkowski's inequality, we have

$$\begin{aligned} \|\mathbf{d}(\mu_{\beta,M,N} \parallel \mu_{\beta,N})\|_p &= \left\| \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot \mathbf{d}(\mu_{\beta,M}^u \parallel \mu_{\beta,N-KM}^u) \right\|_p \\ &\leq \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \left\| \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot \mathbf{d}(\mu_{\beta,M}^u \parallel \mu_{\beta,N-KM}^u) \right\|_p. \end{aligned} \quad (2.4.1)$$

Let  $K \leq \lfloor N/M \rfloor - 1$ . Applying Jensen's inequality to  $\mu_{\beta,M,KM}$ , we have

$$\begin{aligned} &\mathbb{E} \left[ \left| \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot \mathbf{d}(\mu_{\beta,M}^u \parallel \mu_{\beta,N-KM}^u) \right|^p \right] \\ &\leq \mathbb{E} \left[ \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot |\mathbf{d}(\mu_{\beta,M}^u \parallel \mu_{\beta,N-KM}^u)|^p \right]. \end{aligned} \quad (2.4.2)$$

Then by the law of iterated expectation and the branching property, (2.4.2) is equal to

$$\begin{aligned} &\mathbb{E} \left[ \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot \mathbb{E} [|\mathbf{d}(\mu_{\beta,M}^u \parallel \mu_{\beta,N-KM}^u)|^p \mid \mathcal{F}_{KM}] \right] \\ &= \mathbb{E} \left[ \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \right] \cdot \mathbb{E} [|\mathbf{d}(\mu_{\beta,M} \parallel \mu_{\beta,N-KM})|^p] = \mathbb{E} [|\mathbf{d}(\mu_{\beta,M} \parallel \mu_{\beta,N-KM})|^p]. \end{aligned} \quad (2.4.3)$$

Combining (2.4.1), (2.4.3) and Proposition 2.3.2, we conclude that

$$\|\mathbf{d}(\mu_{\beta,M,N} \parallel \mu_{\beta,N})\|_p \leq \left\lfloor \frac{N}{M} \right\rfloor \cdot C(p).$$

### 2.4.2 Proof of (2.1.10)

In this section, we prove (2.1.10) which gives a tighter control on the Kullback–Leibler divergence between  $\mu_{\beta,M,N}$  and the Gibbs measure  $\mu_{\beta,N}$ . We start with the following simple lemma.

**Lemma 2.4.1.** *Let  $\beta \in [0, \beta_c)$ . For all  $M \in \mathbb{N}$ , there exists  $r \in (0, 1)$  independent of  $M$  such that,*

$$\mathbb{E} \left[ \sum_{|u|=M} \mu_{\beta,M}(u)^2 \right] \leq r \quad (2.4.4)$$

Moreover, for all  $K \in \mathbb{N}$ ,

$$\mathbb{E} \left[ \sum_{|u|=KM} \mu_{\beta,M,KM}(u)^2 \right] \leq r^K. \quad (2.4.5)$$

*Proof of Lemma 2.4.1.* For  $x \in (0, 1)$ ,  $x^2 < x$ . Thus, by the fact that  $\mu_{\beta,M}(u) \in (0, 1)$  for all  $|u| = M$ , we derive that

$$\sum_{|u|=M} \mu_{\beta,M}(u)^2 < \sum_{|u|=M} \mu_{\beta,M}(u) = 1,$$

for every  $M \in \mathbb{N}$ . This shows that (2.4.4) holds for every fixed  $M \in \mathbb{N}$  and with  $r < 1$  possibly depending on  $M$ . Uniformity in  $M$  follows as soon as we show that

$$\limsup_{M \rightarrow \infty} \mathbb{E} \left[ \sum_{|u|=M} \mu_{\beta,M}(u)^2 \right] < 1.$$

To this end, recall that  $W_{\beta,M} \rightarrow W_{\beta,\infty}$  almost surely as  $M \rightarrow \infty$  and that  $\mathbb{E}[W_{\beta,\infty}] = 1$ . Hence, there exist  $a < 1$  and  $M_0 \in \mathbb{N}$  such that

$$\forall M \geq M_0 : \mathbb{P}(W_{\beta,M} < 1/2) \leq a. \quad (2.4.6)$$

Now fix  $p \in (1, 2]$  such that  $\varphi(p\beta) < p\varphi(\beta)$ , which exists because  $\beta < \beta_c$ . Then decompose:

$$\begin{aligned} \mathbb{E} \left[ \sum_{|u|=M} \mu_{\beta,M}(u)^2 \right] &\leq \mathbb{E} \left[ \sum_{|u|=M} \mu_{\beta,M}(u)^p \right] \\ &\leq \mathbb{E} \left[ \left( \sum_{|u|=M} \mu_{\beta,M}(u)^p \right) \mathbf{1}_{W_{\beta,M} < 1/2} \right] + \mathbb{E} \left[ \left( \sum_{|u|=M} \mu_{\beta,M}(u)^p \right) \mathbf{1}_{W_{\beta,M} \geq 1/2} \right] \\ &\leq \mathbb{P}(W_{\beta,M} < 1/2) + 2^p e^{(\varphi(p\beta) - p\varphi(\beta))M}, \end{aligned}$$

using the definition of  $\mu_{\beta,M}(u)$  for the last inequality. This shows that

$$\limsup_{M \rightarrow \infty} \mathbb{E} \left[ \sum_{|u|=M} \mu_{\beta,M}(u)^2 \right] \leq a,$$

which concludes the proof of (2.4.4).

We now show (2.4.5) by induction. The case  $K = 1$  follows directly from (2.1.4) and (2.4.4). For  $K > 1$ , by (2.1.4) and branching property, we obtain that

$$\begin{aligned} & \mathbb{E} \left[ \sum_{|u|=KM} \mu_{\beta,M,KM}(u)^2 \middle| \mathcal{F}_{(K-1)M} \right] \\ &= \mathbb{E} \left[ \sum_{|u_1|=(K-1)M} \sum_{|u_2|=M} \mu_{\beta,M,(K-1)M}(u_1)^2 \cdot \mu_{\beta,M}^{u_1}(u_2)^2 \middle| \mathcal{F}_{(K-1)M} \right] \\ &= \sum_{|u_1|=(K-1)M} \mu_{\beta,M,(K-1)M}(u_1)^2 \cdot \mathbb{E} \left[ \sum_{|u_2|=M} \mu_{\beta,M}(u_2)^2 \right]. \end{aligned} \quad (2.4.7)$$

Taking expectations, we derive that

$$\mathbb{E} \left[ \sum_{|u|=KM} \mu_{\beta,M,KM}(u)^2 \right] = \mathbb{E} \left[ \sum_{|u|=(K-1)M} \mu_{\beta,M,(K-1)M}(u)^2 \right] \cdot \mathbb{E} \left[ \sum_{|u|=M} \mu_{\beta,M}(u)^2 \right]. \quad (2.4.8)$$

The equation (2.4.5) then follows from the induction hypothesis.  $\blacksquare$

We now proceed with the proof of (2.1.10). Without loss of generality, using the fact that  $\|\cdot\|_p \leq \|\cdot\|_2$  for every  $p \in [1, 2]$ , we assume that  $p \geq 2$ . By Theorem 2.2.2 and Minkowski's inequality, we have

$$\begin{aligned} & \|\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N}) - \mathbb{E}[\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N})]\|_p \\ & \leq \sum_{K=0}^{\lfloor \frac{N}{M} \rfloor - 1} \left\| \sum_{|u|=KM} \mu_{\beta,M,KM}(u) \cdot (\mathbf{d}(\mu_{\beta,M}^u || \mu_{\beta,N-KM}^u) - \mathbb{E}[\mathbf{d}(\mu_{\beta,M}^u || \mu_{\beta,N-KM}^u)]) \right\|_p. \end{aligned} \quad (2.4.9)$$

We now introduce some notation. For all  $0 \leq K \leq \lfloor \frac{N}{M} \rfloor - 1$ , denote

$$d_K^u = \mathbf{d}(\mu_{\beta,M}^u || \mu_{\beta,N-KM}^u) \quad \text{and} \quad d_K = \mathbf{d}(\mu_{\beta,M} || \mu_{\beta,N-KM})$$

and

$$Z_u = \mu_{\beta,M,KM}(u) \cdot (d_K^u - \mathbb{E}[d_K^u]).$$

We claim that for all  $p \geq 1$ , the sequence

$$a_K = \left\| \sum_{|u|=KM} Z_u \right\|_p \quad (2.4.10)$$

is summable, with a bound independent of  $M$ . This will imply that the right-hand side of (2.4.9) is bounded by the same quantity, which completes proof. To prove this, first observe that  $(Z_u)_{|u|=KM}$  is a sequence of i.i.d. random variables having

zero mean and finite  $p$ -th moments, for any  $p \geq 1$ , with respect to  $\mathbb{E}[\cdot | \mathcal{F}_{KM}]$ , by Proposition 2.3.2 and the branching property. Denote by  $C_1, C_2, \dots$  some constants possibly depending on  $p$  (and  $\beta$  and the law of  $\mathbf{Y}$ ). By Rosenthal's inequality [164, Theorem 3], we have

$$\mathbb{E} \left[ \left| \sum_{|u|=KM} Z_u \right|^p \middle| \mathcal{F}_{KM} \right] \leq C_1 \cdot \left\{ \left( \sum_{|u|=KM} \mathbb{E} [|Z_u|^p | \mathcal{F}_{KM}] \right) + \left( \sum_{|u|=KM} \mathbb{E} [|Z_u|^2 | \mathcal{F}_{KM}] \right)^{p/2} \right\} \quad (2.4.11)$$

By the branching property and Proposition 2.3.2, the first term of (2.4.11) can be bounded by

$$\begin{aligned} \sum_{|u|=KM} \mathbb{E} [|Z_u|^p | \mathcal{F}_{KM}] &= \sum_{|u|=KM} \mu_{\beta, M, KM}(u)^p \cdot \mathbb{E} [|d_K - \mathbb{E}[d_K]|^p] \\ &\leq \sum_{|u|=KM} \mu_{\beta, M, KM}(u)^p \cdot C_2 \\ &\leq \sum_{|u|=KM} \mu_{\beta, M, KM}(u)^2 \cdot C_2, \end{aligned} \quad (2.4.12)$$

using that  $p \geq 2$  in the last line. Taking expectations and applying Lemma 2.4.1, we get

$$\mathbb{E} \left[ \sum_{|u|=KM} \mathbb{E} [|Z_u|^p | \mathcal{F}_{KM}] \right] \leq \mathbb{E} \left[ \sum_{|u|=KM} \mu_{\beta, M, KM}(u)^2 \right] \cdot C_2 \leq r^K \cdot C_2, \quad (2.4.13)$$

with  $r < 1$  as in Lemma 2.4.1.

We now estimate the second term of (2.4.11). By the branching property and Lemma 2.3.1,

$$\begin{aligned} \left( \sum_{|u|=KM} \mathbb{E} [|Z_u|^2 | \mathcal{F}_{KM}] \right)^{p/2} &= \left( \sum_{|u|=KM} \mu_{\beta, M, KM}(u)^2 \cdot \mathbb{E} [|d_K - \mathbb{E}[d_K]|^2] \right)^{p/2} \\ &\leq \left( \sum_{|u|=KM} \mu_{\beta, M, KM}(u)^2 \right)^{p/2} \cdot C_3 \\ &\leq \left( \sum_{|u|=KM} \mu_{\beta, M, KM}(u)^2 \right) \cdot C_3. \end{aligned} \quad (2.4.14)$$

The inequality (2.4.14) is because

$$\sum_{|u|=KM} \mu_{\beta, M, KM}(u)^2 \in [0, 1]$$

and  $x^{p/2} \leq x$ , for all  $x \in [0, 1]$  and  $p \geq 2$ . Taking expectations and applying Lemma 2.4.1, we get

$$\mathbb{E} \left[ \left( \sum_{|u|=KM} \mathbb{E} [|Z_u|^2 | \mathcal{F}_{KM}] \right)^{p/2} \right] \leq \mathbb{E} \left[ \sum_{|u|=KM} \mu_{\beta, M, KM}(u)^2 \right] \cdot C_3 \leq r^K \cdot C_3. \quad (2.4.15)$$

Combining (2.4.11), (2.4.13) and (2.4.15), we conclude that

$$\left\| \sum_{|u|=KM} Z_u \right\|_p \leq C_1(C_2 + C_3) \cdot r^K.$$

This implies that (2.4.10) is summable and finishes the proof.

## 2.5 Proof of hardness result (Theorem 2.1.9)

In this section, we prove Theorem 2.1.9. To do this, we recall two results of asymptotic behaviors of the maximal particle of a branching random walk.

Under the assumption of the theorem, it is known in Corollary (3.4) of [46] that

$$\frac{\max_{|u|=N} X_u}{N} \rightarrow m := \varphi'(\beta_c), \quad \text{a.s. as } N \rightarrow \infty. \quad (2.5.1)$$

Furthermore, we have the following tail estimate, which easily follows from a union bound together with Chernoff's bound (see e.g. the proof of Theorem 2 in [185]): there exists a constant  $c > 0$  such that

$$\forall x \geq 0 : \mathbb{P} \left( \max_{|u|=N} X_u \geq mN + x \right) \leq e^{-cx}. \quad (2.5.2)$$

The key to Theorem 2.1.9 is the following observation:

**Lemma 2.5.1.** *Let  $\beta > \beta_c$  and assume  $\beta_c \in \mathcal{D}(\varphi)^\circ$ . Let  $u$  be a particle sampled according to the Gibbs measure  $\mu_{\beta,N}$  and let  $w$  be its ancestor at generation  $\lfloor N/2 \rfloor$ . Then there exists a positive random variable  $Z$  with continuous distribution function such that*

$$\frac{X_u - X_w - mN/2}{\sqrt{N}} \rightarrow Z, \quad \text{in law as } N \rightarrow \infty.$$

*Proof.* This is a consequence of a result by Chen, Madaule and Mallein [79]. These authors show the following fact: if  $u$  is sampled according to the Gibbs measure  $\mu_{\beta,N}$ , and  $X_u(t)$  denotes the position of its ancestor at generation  $\lfloor tN \rfloor$ , and if we define

$$Z_t^N := \frac{mtN - X_u(t)}{\sqrt{N}}, \quad t \in [0, 1],$$

then  $(Z_t^N)_{t \in [0,1]}$  converges in law (w.r.t. Skorokhod's topology) to a multiple of a Brownian excursion as  $N \rightarrow \infty$ . Note that the assumptions in their article are implied by our hypothesis that  $\beta_c \in \mathcal{D}(\varphi)^\circ$  and the fact that in our branching random walk, the number of offspring of a particle is deterministic. Now, we also have that  $X_u - mN = O(\log N)$  in probability (see e.g. [6]), so that

$$\frac{X_u - mN}{\sqrt{N}} \xrightarrow{\text{law}} 0, \quad \text{as } N \rightarrow \infty.$$

Together, both results imply the lemma. ■

*Proof of Theorem 2.1.9.* Assume that we are given an algorithm  $(v(k))_{k \geq 1}$  that samples a vertex according to a random probability measure  $\tilde{\mu}_{\beta, N}$  approximating the Gibbs measure  $\mu_{\beta, N}$ . By Lemma 2.1.4, it follows that  $\mathbf{d}(\tilde{\mu}_{\beta, N} || \mu_{\beta, N}) \rightarrow 0$  in probability as  $N \rightarrow \infty$ . Hence, by Pinsker's inequality (see e.g. Theorem 4.19 in [52]), the total variation distance between  $\tilde{\mu}_{\beta, N}$  and  $\mu_{\beta, N}$  goes to 0 as well in probability, as  $N \rightarrow \infty$ . It follows that Lemma 2.5.1 holds as well for  $u$  sampled according to  $\tilde{\mu}_{\beta, N}$ .

Let  $\delta > 0$ . For  $z > 0$ , call a vertex  $w \in \partial\mathbb{T}_{\lfloor N/2 \rfloor}$  *z-exceptional* if it has a descendant  $u \in \partial\mathbb{T}_N$  such that  $X_u - X_w - mN/2 > z\sqrt{N}$ . By the preceding paragraph, there exists  $z > 0$  such that for large enough  $N$ , the algorithm finds a vertex  $u$  whose ancestor  $w$  at generation  $\lfloor N/2 \rfloor$  is *z-exceptional* with probability at least  $1 - \delta$ . Hence, it is enough to show that any algorithm which solves the simpler problem of finding a *z-exceptional* vertex at generation  $\lfloor N/2 \rfloor$  has a running time at least  $e^{z'\sqrt{N}}$  with probability  $1 - \delta$ , for some  $z' > 0$ . We will now show that the statement of the theorem holds even for this simpler problem. For this, we use an argument similar to the one in Section 3 of [2]. We first present the argument in an informal way.

Denote by  $E_w$  the event that a given vertex  $w \in \partial\mathbb{T}_{\lfloor N/2 \rfloor}$  is *z-exceptional*. Note that this event only depends on the displacements of the descendants of  $w$ . Hence, the events  $(E_w)_{w \in \partial\mathbb{T}_{\lfloor N/2 \rfloor}}$  are independent by the branching property. Furthermore, by (2.5.2), for each  $w \in \partial\mathbb{T}_{\lfloor N/2 \rfloor}$ , we have  $\mathbb{P}(E_w) \leq e^{-cz\sqrt{N/2}}$  for some  $c > 0$ . Finally, in order to determine whether a vertex  $w$  is *z-exceptional*, the algorithm has to explore at least one vertex in the subtree of the vertex  $w$ . Hence, the running time of the algorithm is bounded from below by the number of vertices  $w$  that have to be probed in order to find a *z-exceptional* vertex. But this quantity follows the geometric distribution with success probability  $\mathbb{P}(E_w) \leq e^{-cz\sqrt{N/2}}$ . Altogether, for any  $z' < cz/\sqrt{2}$  and for  $N$  sufficiently large, this shows that the running time  $\tau$  of any algorithm solving the simpler problem is at least  $e^{z'\sqrt{N}}$  with probability  $1 - \delta$ . The statement readily follows.

We now make this argument formal. Recall that, by definition, an algorithm is a stochastic process  $(v(n))_{n \geq 0}$  previsible with respect to the filtration  $\tilde{\mathcal{F}}$ , defined by

$$\tilde{\mathcal{F}}_k = \sigma(v(1), \dots, v(k); X_{v(1)}, \dots, X_{v(k)}; U_1, \dots, U_{k+1})$$

where  $(U_k)_{k \geq 1}$  is a sequence of i.i.d. uniform random variables on  $[0, 1]$ , independent of the branching random walk  $\mathbf{X}$ . We now define a larger filtration  $\mathcal{G}$ . For this, define for any  $v \in \mathbb{T}_N$  the following set of vertices:

$$\mathcal{V}_v = \begin{cases} w \in \mathbb{T}_N : |v \wedge w| \geq \lfloor N/2 \rfloor, & \text{if } |v| \geq \lfloor N/2 \rfloor \\ v, & \text{otherwise.} \end{cases}$$

Note that  $v \in \mathcal{V}_v$  for every  $v \in \mathbb{T}_N$ . We then set

$$\mathcal{G}_k = \sigma(v(1), \dots, v(k); (X_w)_{w \in \mathcal{V}_{v(1)}}, \dots, (X_w)_{w \in \mathcal{V}_{v(k)}}; U_1, \dots, U_{k+1}).$$

Note that  $\tilde{\mathcal{F}}_k \subset \mathcal{G}_k$  for all  $k \geq 0$  — heuristically,  $\mathcal{G}_k$  adds to  $\tilde{\mathcal{F}}_k$  the information about the values in the branching random walk of all vertices contained in  $\mathcal{V}_{v(i)}$ ,

$i = 1, \dots, k$ . Note that trivially, the stochastic process  $v(n)_{n \geq 0}$  is still previsible with respect to this larger filtration  $\mathcal{G}$ .

Now say that  $\mathcal{V}_v$  is  $z$ -exceptional if  $|v| \geq \lfloor N/2 \rfloor$  and the ancestor of  $v$  at generation  $\lfloor N/2 \rfloor$  is  $z$ -exceptional in the sense defined above — note that this definition does not depend on the choice of  $v$ . Define

$$\tau' = \inf\{k \geq 0 : \mathcal{V}_{v(k)} \text{ is } z\text{-exceptional}\},$$

and note that  $\tau'$  is a stopping time with respect to the filtration  $\mathcal{G}$ . Now, by the equality of events

$$\{\tau' = k\} = \{\mathcal{V}_{v(k)} \text{ is } z\text{-exceptional}\} \cap \{\tau' > k - 1\},$$

and since  $\{\tau' > k - 1\} \in \mathcal{G}_{k-1}$  and  $v(k)$  is  $\mathcal{G}_{k-1}$ -measurable, we have

$$\mathbb{P}(\tau' = k | \mathcal{G}_{k-1}) = \sum_{v \in \mathbb{T}_N} \mathbb{P}(\mathcal{V}_v \text{ is } z\text{-exceptional} | \mathcal{G}_{k-1}) \mathbf{1}_{(v(k)=v, \tau' > k-1)}.$$

Now, for any  $v \in \mathbb{T}_N$ , if  $|v| < \lfloor N/2 \rfloor$ , or if  $v \in \bigcup_{i=0}^{k-1} \mathcal{V}_{v(i)}$ , the above probability is zero, because none of  $\mathcal{V}_{v(i)}$ ,  $i = 0, \dots, k - 1$  are  $z$ -exceptional on the event  $\{\tau' > k - 1\}$ . On the other hand, if  $v \notin \bigcup_{i=0}^{k-1} \mathcal{V}_{v(i)}$ , then, by the branching property, the above conditional probability is equal to the unconditioned probability that  $\mathcal{V}_v$  is  $z$ -exceptional, which is bounded by  $e^{-cz\sqrt{N/2}}$  by (2.5.2). Hence, we get in total that

$$\mathbb{P}(\tau' = k | \tau' > k - 1) \leq e^{-cz\sqrt{N/2}},$$

and  $\tau'$  is dominated from below by a geometric random variable with success probability  $e^{-cz\sqrt{N/2}}$ . The proof now continues as above.  $\blacksquare$

## 2.6 Proof of Lemma 2.1.4

We first consider the case  $\beta \in [0, \beta_c)$ . Define

$$D_{\beta, N} = \frac{d}{d\beta} W_{\beta, N} = \sum_{|u|=n} (X_u - \varphi'(\beta)N) e^{\beta X_u - \varphi(\beta)N}.$$

We express the entropy by

$$\begin{aligned} H(\mu_{\beta, N}) &= \frac{1}{W_{\beta, N}} \sum_{|u|=N} (\varphi(\beta)N - \beta X_u) e^{\beta X_u - \varphi(\beta)N} + \log W_{\beta, N} \\ &= (\varphi(\beta) - \beta \varphi'(\beta))N - \beta \frac{D_{\beta, N}}{W_{\beta, N}} + \log W_{\beta, N}. \end{aligned}$$

By the assumption on  $\beta$ , we have  $\varphi(\beta) - \beta \varphi'(\beta) > 0$ . Furthermore,  $W_{\beta, N}$  converges almost surely to a positive random variable as  $N \rightarrow \infty$  by Fact 2.1.2 and  $D_{\beta, N}$  converges almost surely as well as  $N \rightarrow \infty$ , see [48]. The first statement follows.

Now let  $\beta > \beta_c$  and assume  $\beta_c \in \mathcal{D}(\varphi)^\circ$ . Define

$$\widetilde{W}_{\beta, N} = \sum_{|u|=N} e^{\beta(X_u - \varphi'(\beta_c)N) - \frac{3}{2\beta_c} \log N}.$$

We now write

$$H(\mu_{\beta,N}) = \frac{1}{\widetilde{W}_{\beta,N}} \sum_{|u|=N} \left( -\beta \left( X_u - \varphi'(\beta_c)N - \frac{3}{2\beta_c} \log N \right) \right) e^{\beta(X_u - \varphi'(\beta_c)N - \frac{3}{2\beta_c} \log N)} + \log \widetilde{W}_{\beta,N}.$$

Fix  $\beta' \in (\beta_c, \beta)$ . Then there exists  $C > 0$ , such that  $xe^{-\beta x} \leq Ce^{-\beta' x}$  for every  $x \in \mathbb{R}$ . We have

$$H(\mu_{\beta,N}) \leq C \frac{\widetilde{W}_{\beta',N}}{\widetilde{W}_{\beta,N}} + \log \widetilde{W}_{\beta,N}.$$

Now,  $\widetilde{W}_{\beta,N}$  and  $\widetilde{W}_{\beta',N}$  converge almost surely as  $N \rightarrow \infty$  to positive random variables, see [130]. The second statement follows.

## 2.7 Open questions and further directions

In this section, we state a few questions and further directions related to the CREM and the branching random walks for future study.

1. It was conjectured in [2] that for the CREM, there exists a threshold  $\beta_G \geq 0$  such that its Gibbs measure with parameter  $\beta$  can be efficiently approximated if  $\beta < \beta_G$  and cannot if  $\beta > \beta_G$ . A conjectured explicit expression<sup>2</sup> of  $\beta_G$  appears in Item 1, Section 5 of [2]. Our results confirm the conjecture in the case where the CREM has correlation function  $A(x) = x$ , with the notion of *approximation* from Definition 2.1.5. Moreover, our results imply that  $\beta_G = \beta_c = \sqrt{2 \log 2}$ , the *(static) critical inverse temperature*. One can check that  $\beta_c$  equals the expression of  $\beta_G$  from [2]. Ongoing work of the authors is trying to generalize the result to the CREM with a general correlation function.
2. Back to the branching random walk, one might be interested in the near critical regime to understand how the transition happens near  $\beta_c$ . To do this, one can take a sequence  $\beta(N) = \beta_c - N^{-\delta}$  for some  $\delta > 0$ . It might be interesting to study the time complexity of any algorithm approximating the Gibbs measure  $\mu_{\beta(N),N}$  in the sense of Definition 2.1.5. One should expect a phase transition at  $\delta = 1/2$ , in line with a phase transition for the asymptotics of the partition function obtained by Alberts and Ortgiese [10]. See also the introduction of Pain [151]. This should be related to Pemantle's [159] study of optimization algorithms discussed in Section 2.1.2.

---

<sup>2</sup>There is a typo in the expression of  $\beta_G$  in the case of the CREM in Item 1, Section 5 of [2]. The definition of  $t_G$  should be replaced by the following one:  $t_G = \sup\{t \in [0, 1] : A(s) = \hat{A}(s) \text{ for all } s \leq t\}$ .

# 3. Efficient sampling of the CREM Gibbs measure

The continuous random energy model (CREM) is a toy model of disordered system introduced by Bovier and Kurkova in 2004 based on previous work by Derrida and Spohn in the 80s. In a recent paper by Addario-Berry and Maillard, they raised the following question: what is the threshold  $\beta_G$ , at which sampling approximately the Gibbs measure at inverse temperature  $\beta_G$  becomes algorithmically hard? Here, sampling approximately means that the Kullback–Leibler divergence from the output law of the algorithm to the Gibbs measure is of order  $o(N)$  with probability approaching 1, as  $N \rightarrow \infty$ , and algorithmically hard means that the running time, the numbers of vertices queries by the algorithms, is beyond of polynomial order.

The present work shows that when the covariance function  $A$  of the CREM is concave, for all  $\beta > 0$ , a recursive sampling algorithm on a renormalized tree approximates the Gibbs measure with running time of order  $O(N^{1+\varepsilon})$ . For  $A$  non-concave, the present work exhibits a threshold  $\beta_G < \infty$  such that the following hardness transition occurs: a) For every  $\beta \leq \beta_G$ , the recursive sampling algorithm approximates the Gibbs measure with running time of order  $O(N^{1+\varepsilon})$ . b) For every  $\beta > \beta_G$ , a hardness result is established for a large class of algorithms. Namely, for any algorithm from this class that samples the Gibbs measure approximately, there exists  $z > 0$  such that the running time of this algorithm is at least  $e^{zN}$  with probability approaching 1. In other words, it is impossible to sample approximately in polynomial-time the Gibbs measure in this regime.

Additionally, we provide a lower bound of the free energy of the CREM that could hold its own value.

## 3.1 Introduction

The continuous random energy model (CREM) is a toy model of a disordered system in statistical physics, i.e. a model where the Hamiltonian — the function that assigns energies to the states of the system — is itself random. The CREM was introduced by Bovier and Kurkova [54] based on previous work by Derrida and Spohn [91]. Mathematically, the model is defined as follows. For a given integer  $N \in \mathbb{N}$ , the CREM is a centered Gaussian process  $\mathbf{X} = (X_u)_{u \in \mathbb{T}_N}$  indexed by the

binary tree  $\mathbb{T}_N$  of depth  $N$  with covariance function

$$\mathbb{E}[X_v X_w] = N \cdot A\left(\frac{|v \wedge w|}{N}\right), \quad \forall v, w \in \mathbb{T}_N.$$

Here,  $|v \wedge w|$  is the depth of the most recent common ancestor of  $v$  and  $w$ , and the function  $A$  is assumed to be a non-decreasing function defined on an interval  $[0, 1]$  such that  $A(0) = 0$  and  $A(1) = 1$ . An essential quantity of this model is the Gibbs measure, which is a probability measure defined on the set of leaves  $\partial\mathbb{T}_N$  where the weight of  $v \in \partial\mathbb{T}_N$  is proportional to  $e^{\beta X_v}$ .

The present work consider the sampling problem of the Gibbs measure. We say that a (randomized) algorithm approximates the Gibbs measure if the Kullback–Leibler divergence from the output law of this algorithm to the Gibbs measure is of order  $o(N)$  with probability approaching 1. The present work considers a recursive sampling algorithm that is similar to the one appearing in [2] and [112]. We shows that when the covariance function  $A$  of the CREM is concave, for all  $\beta > 0$ , the recursive sampling algorithm approximates the Gibbs measure with running time of order  $O(N^{1+\varepsilon})$ . Moreover, when  $A$  is non-concave, we identify a threshold  $\beta_G < \infty$  such that the following hardness transition occurs: a) For every  $\beta \leq \beta_G$ , the recursive sampling algorithm approximates the Gibbs measure with running time of order  $O(N^{1+\varepsilon})$ . b) For every  $\beta > \beta_G$ , we prove a hardness result for a generic class of algorithms. Namely, there exists  $\gamma > 0$  such that for any algorithm in this class that approximates the Gibbs measure, the running time of this algorithm is at least  $e^{\gamma N}$  with probability approaching 1.

### 3.1.1 Definitions and notation

Throughout this paper, we denote by  $\mathbb{N} = \{1, 2, \dots\}$  the set of positive integer. For each pair of integers  $n$  and  $m$  such that  $n \leq m$ , we denote by  $[[n, m]]$  the set of integers between  $n$  and  $m$ .

**Binary tree.** Fixing  $N \in \mathbb{N}$ , we denote by  $\mathbb{T}_N = \{\emptyset\} \cup \bigcup_{n=1}^N \{0, 1\}^n$  the binary tree rooted at  $\emptyset$ . The depth of a vertex  $v \in \mathbb{T}_N$  is denoted by  $|v|$ . For any  $v, w \in \mathbb{T}_N$ , we write  $v \leq w$  if  $v$  is a prefix of  $w$  and write  $v < w$  if  $v$  is a prefix of  $w$  strictly shorter than  $w$ . In the following, for any  $v \in \mathbb{T}_N$ , we refer to any vertex  $w$  with  $w \leq v$  as an ancestor of  $v$ . For any  $v \in \mathbb{T}_N$  and  $n \in [[0, |v|]]$ , define  $v[n]$  to be the ancestor of  $v$  of depth  $n$ . For all  $v, w \in \mathbb{T}_N$ , we denote by  $v \wedge w$  the most recent common ancestor of  $v$  and  $w$ . We denote by  $\partial\mathbb{T}_N$  the set of leaves of  $\mathbb{T}_N$ , and for any  $v \in \mathbb{T}_N$ , let  $\mathbb{T}_n^v$  be the subtree of  $\mathbb{T}_N$  rooted at  $v$  with depth  $n$ .

**The model.** Let  $A$  be a non-decreasing function defined on an interval  $[0, 1]$  such that  $A(0) = 0$  and  $A(1) = 1$ . For the sake of this paper, we assume that there exists a bounded Riemann integrable function  $a$  such that  $A$  for all  $t \in [0, 1]$ ,

$$A(t) = \int_0^t a(s) ds.$$

We denote by  $\hat{A}$  the concave hull of  $A$  and by  $\hat{a}$  the right derivative of  $\hat{A}$ . Note that the  $\hat{A}$  is also equals to the Riemann integral of  $\hat{a}$ , i.e., for all  $t \in [0, 1]$ ,

$$\hat{A}(t) = \int_0^t \hat{a}(s) ds.$$

We now introduce the CREM.

**Definition 3.1.1.** Given  $N \in \mathbb{N}$ , the CREM is a centered Gaussian process  $\mathbf{X} = (X_u)_{u \in \mathbb{T}_N}$  indexed by the binary tree  $\mathbb{T}_N$  of depth  $N$  with covariance function

$$\mathbb{E}[X_v X_w] = N \cdot A \left( \frac{|v \wedge w|}{N} \right), \quad \forall v, w \in \mathbb{T}_N, \quad (3.1.1)$$

where  $|v \wedge w|$  is the depth of the most recent common ancestor of  $v$  and  $w$ .

Throughout this paper, we consider a sequence of CREM  $(\mathbf{X}_N)_{N \in \mathbb{N}}$  defined on the same underlying probability space. For simplicity, we drop  $N$  as long as it causes no ambiguity.

**Branching property.** The CREM can be viewed as an inhomogeneous binary branching random walk with Gaussian increments. In particular, it has the following branching property: let  $(\mathcal{F}_k)_{k=0}^N$  be the natural filtration of the CREM. For any  $u \in \mathbb{T}_N$  with  $|u| = n \in \llbracket 0, N \rrbracket$ , call the process

$$\mathbf{X}^u = (X_w)_{w \in \mathbb{T}_{N-n}^u}$$

the CREM indexed by the subtree  $\mathbb{T}_{N-n}^u$ , where  $X_w^u = X_{uw} - X_u$ . For any  $n \in \llbracket 0, N \rrbracket$ , let  $\mathbf{X}^{(n)} = (X_u^{(n)})_{u \in \mathbb{T}_{N-n}}$  be a centered Gaussian process with covariance function

$$\mathbb{E}[X_{w_1}^{(n)} X_{w_2}^{(n)}] = N \cdot A \left( \frac{n + |w_1 \wedge w_2|}{N} \right), \quad \forall w_1, w_2 \in \mathbb{T}_{N-n}.$$

Then, the branching property states that collection of processes  $\{\mathbf{X}^u : |u| = n\}$  are independent and have the identical distribution of  $\mathbf{X}^{(n)}$ , and they are independent of  $\mathcal{F}_n$ .

**Partition function and Gibbs measure.** Given a subtree  $\mathbb{T}_M^v$  rooted at  $v$  and of depth  $M \in \llbracket 0, N - |v| \rrbracket$ , the *Gibbs measure with inverse temperature*  $\beta > 0$  is defined by

$$\mu_{\beta, M}^v(u) = \frac{1}{Z_{\beta, M}^v} e^{\beta X_u^v}, \quad \forall u \in \mathbb{T}_M^v, \quad (3.1.2)$$

where

$$Z_{\beta, M}^v = \sum_{u \in \mathbb{T}_M^v} e^{\beta X_u^v}, \quad (3.1.3)$$

is the partition function on the subtree  $\mathbb{T}_M^v$ . In particular, we adopt the conventions

$$\mu_{\beta, M} = \mu_{\beta, M}^\emptyset \quad \text{and} \quad Z_{\beta, M} = Z_{\beta, M}^\emptyset$$

for any  $M \in \llbracket 0, N \rrbracket$ . For completeness, we also define  $Z_{\beta, M}^{(n)} = \sum_{|u|=M} e^{\beta X_u^{(n)}}$ .

**Free energy and its lower bound.** For  $v \in \mathbb{T}_N$ , we refer to the logarithm of the partition function  $\log Z_{\beta, M}^v$  as the free energy on the subtree  $\mathbb{T}_M^v$ . The free energy  $F_\beta$  of the CREM is defined as follows, and  $F_\beta$  admits an explicit expression.

$$F_\beta := \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} [\log Z_{\beta, N}] = \int_0^1 f(\beta \sqrt{\hat{a}(s)}) ds, \quad (3.1.4)$$

where the function  $f$  is defined as

$$f(\beta) = \begin{cases} \log 2 + \frac{\beta^2}{2}, & \beta < \sqrt{2 \log 2} \\ \sqrt{2 \log 2} \beta, & \beta \geq \sqrt{2 \log 2}. \end{cases} \quad (3.1.5)$$

For completeness, we include the proof of (3.1.4) in Fact 3.A.2. When clear, we also simply refer to  $F_\beta$  as the free energy. We introduce a related quantity  $\tilde{F}_\beta$  defined as

$$\tilde{F}_\beta := \int_0^1 f(\beta \sqrt{a(s)}) ds. \quad (3.1.6)$$

In Proposition 3.A.1, we show that  $F_\beta \geq \tilde{F}_\beta$  and characterize the condition where the equality holds.

**Algorithms.** We follow the same definition of randomized algorithms as in [2, 112], which also appeared in similar forms in [159].

**Definition 3.1.2** (Algorithm). Let  $N \in \mathbb{N}$ . Let  $\tilde{\mathcal{F}}_k$  be a filtration defined by

$$\tilde{\mathcal{F}}_k = \sigma(v(1), \dots, v(k); X(v(1)), \dots, X(v(k)); U_1, \dots, U_{k+1})$$

where  $(U_k)_{k \geq 1}$  is a sequence of i.i.d. uniform random variables on  $[0, 1]$ , independent of the continuous random energy model  $\mathbf{X}$ . A random sequence  $v = (v(k))_{k \geq 0}$  taking values in  $\mathbb{T}_N$  is called a (*randomized*) *algorithm* if  $v(0) = \emptyset$  and  $v(k+1)$  is  $\tilde{\mathcal{F}}_k$ -measurable for every  $k \geq 0$ . We further suppose that there exists a stopping time  $\tau$  with respect to the filtration  $\tilde{\mathcal{F}}$  and such that  $v(\tau) \in \partial \mathbb{T}_N$ . We call  $\tau$  the *running time* and  $v(\tau)$  the *output* of the algorithm. The *law of the output* is the (random) distribution of  $v(\tau)$ , conditioned on  $\mathbf{X}$ .

*Remark 3.1.3.* Roughly speaking, the filtration  $\tilde{\mathcal{F}} = (\tilde{\mathcal{F}}_k)_{k \geq 0}$  contains all the information about everything the algorithm has queried so far, as well as the additional randomness needed to choose the next vertex.

Throughout the paper, the notion of time complexity is given by the following definition.

**Definition 3.1.4** (Time complexity). Let  $(\tau_N)$  be a sequence of running time corresponds to a sequence of algorithms indexed by  $N$ . Let  $h : \mathbb{N} \rightarrow \mathbb{N}$  be a function. We say that the sequence of running times is of order  $O(h(N))$  if almost surely, there exists  $N_0 \in \mathbb{N}$  such that  $\tau_N \leq h(N)$ . We say the running time is of polynomial order if there exists a polynomial  $P(N)$  such that almost surely, there exists  $N_0 \in \mathbb{N}$ .

*Remark 3.1.5.* In the rest of the paper, when we say that the running time of an algorithm is of order  $O(h(N))$ , we implicitly assume that there is an underlying sequence of algorithms indexed by  $N$ , which we also refer to as an algorithm by abuse of notation.

**Kullback–Leibler divergence.** Given two probability measures  $P$  and  $Q$  defined on a discrete space  $\Omega$ , the Kullback–Leibler divergence (also known as the relative entropy) from  $Q$  to  $P$  is defined by

$$\mathbf{d}(P||Q) = \sum_{\omega \in \Omega} P(\omega) \cdot \log \left( \frac{P(\omega)}{Q(\omega)} \right). \quad (3.1.7)$$

From now on, we abbreviate the Kullback–Leibler divergence as the KL divergence.

The notion of approximation in the present work is the following.

**Definition 3.1.6.** Let  $(P_N)_{N \in \mathbb{N}}$  and  $(Q_N)_{N \in \mathbb{N}}$  are two sequences of *random* probability measures defined on a discrete space  $\Omega$ . We say that the sequence  $(P_N)_{N \in \mathbb{N}}$  approximates the sequence  $(Q_N)_{N \in \mathbb{N}}$  with probability approaching 1 if

$$\lim_{N \rightarrow \infty} \mathbb{P} \left( \frac{1}{N} \mathbf{d}(P_N||Q_N) < \varepsilon_N \right) = 0.$$

*Remark 3.1.7.* Note that Definition 3.1.6 is equivalent to saying that

$$\frac{1}{N} \mathbf{d}(P_N||Q_N) \xrightarrow{\mathbb{P}} 0, \quad \text{as } N \rightarrow \infty.$$

### 3.1.2 Main results

Recall that  $a$  is the derivative of  $A$ . By the Lebesgue criterion of Riemann integrability, the function  $a$  is continuous almost everywhere on  $[0, 1]$ . If  $A$  is non-concave, define the threshold

$$\beta_G = \frac{\sqrt{2 \log 2}}{\operatorname{ess\,sup}_{t \in \{A \neq \hat{A}\}} \sqrt{a(t)}}, \quad (3.1.8)$$

For completeness, we define  $\beta_G = \infty$  when  $A$  is concave. We now state the main results.

#### Subcritical and critical regime $\beta \leq \beta_G$ : optimality of recursive sampling

Fix  $\beta > 0$ ,  $N \in \mathbb{N}$ , and  $M = M_N \in \llbracket 1, N \rrbracket$ . Given a configuration of the continuous random energy model with depth  $N$ , consider the following algorithm:

---

**Algorithm 3.1:** Recursive sampling on renormalized tree

---

```

set  $v = \emptyset$    while  $|v| < N$  do
  | sample  $w$  with  $|w| = M \wedge (N - |v|)$  according to the Gibbs measure
  |  $\mu_{\beta, M \wedge (N - |v|)}^v$    replace  $v$  with  $vw$ 
output  $v$ 

```

---

*Remark 3.1.8.* This is the same algorithm as the one in [112] except that now the law of the Gibbs measure  $\mu_{\beta, M \wedge (N-|v|)}^v$  depends on the depth of  $v$ . Again, its running time is deterministic and bounded by  $\lceil N/M \rceil 2^M$ . The output law of Algorithm 3.1 is a random probability measure  $\mu_{\beta, M, N}$  on  $\partial\mathbb{T}_N$  that is recursively defined as follows:

$$\begin{aligned} \mu_{\beta, M, 0}(\emptyset) &= 1 \\ \mu_{\beta, M, N \wedge (k+1)M}(vw) &= \mu_{\beta, M, kM}(v) \cdot \mu_{\beta, M \wedge (N-kM)}^v(w) \end{aligned} \quad (3.1.9)$$

for all  $|v| = kM$ ,  $|w| = M \wedge (N - kM)$  and  $k \in \llbracket 0, \lfloor \frac{N}{M} \rfloor \rrbracket$ . It is not hard to see that

$$\mu_{\beta, M, N}(u) = \frac{e^{\beta X_u}}{Z_{\beta, M, N}(u)}, \quad (3.1.10)$$

where

$$Z_{\beta, M, N}(u) = \prod_{k=1}^{\lfloor N/M \rfloor} Z_{\beta, M \wedge (N-kM)}^{u[kM]}.$$

The first theorem states that the KL divergence from the output law of Algorithm 3.1 to the Gibbs measure concentrates in the following sense.

**Theorem 3.1.9** (Concentration bounds). *Let  $\beta > 0$ ,  $N \in \mathbb{N}$  and  $M \in \llbracket 1, N \rrbracket$ . Then for all  $p \geq 1$ , there exists a constant  $C_p > 0$  depending only on  $p$  such that*

$$\frac{1}{N} \|\mathbf{d}(\mu_{\beta, M, N} \| \mu_{\beta, N}) - \mathbb{E}[\mathbf{d}(\mu_{\beta, M, N} \| \mu_{\beta, N})]\|_p \leq \frac{\beta C_p}{\sqrt{M}}.$$

Next, we show that with a suitable choice of  $M_N$ , the expectation of the KL divergence renormalized by  $N$  converges to the difference between  $F_\beta$  and  $\tilde{F}_\beta$ .

**Theorem 3.1.10** (Convergence of the KL divergence). *Let  $\beta > 0$ ,  $N \in \mathbb{N}$ , and  $M_N$  be a sequence such that  $M_N \in \llbracket 1, N \rrbracket$  and  $M_N \rightarrow \infty$ . Let  $\tilde{\mu}_{\beta, N} = \mu_{\beta, M_N, N}$  be the output law of Algorithm 3.1. Then,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\mathbf{d}(\tilde{\mu}_{\beta, N} \| \mu_{\beta, N})] = F_\beta - \tilde{F}_\beta \geq 0,$$

with equality holding if and only if  $\beta \leq \beta_G$ .

As a corollary of Theorem 3.1.10, in the subcritical regime, with a good choice of  $M_N$ , the mean of the KL divergence divided by  $N$  converges to 0 when  $N \rightarrow \infty$ . Moreover, for  $\varepsilon > 0$ , with a good choice of  $M_N$ , the running time is of  $O(N^{1+\varepsilon})$ .

**Corollary 3.1.11** (Efficient sampling). *Fix  $\beta \in [0, \beta_G]$ . Given  $\varepsilon > 0$ , let  $M_N = \lceil \varepsilon \log_2 N \rceil \wedge N$  and  $\tilde{\mu}_{\beta, N} = \mu_{\beta, M_N, N}$  be the output law of Algorithm 3.1. Then,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\mathbf{d}(\tilde{\mu}_{\beta, N} \| \mu_{\beta, N})] = 0. \quad (3.1.11)$$

Moreover, the running time is deterministic and of order  $O(N^{1+\varepsilon})$ .

*Remark 3.1.12.* Note that for  $A$  concave, Corollary 3.1.11 yields that Algorithm 3.1 approximates the Gibbs measure for all  $\beta \in (0, \infty)$  as  $\beta_G = \infty$  for  $A$  concave.

Corollary 3.1.11 implies in particular that the algorithm approximates the Gibbs measure with probability approaching 1. Indeed, if we choose, e.g.,  $\varepsilon_N = (\frac{1}{N} \mathbb{E} [\mathbf{d}(\tilde{\mu}_{\beta, N} || \mu_{\beta, N})])^{1/2}$ , Corollary 3.1.11 and Markov's inequality then yield

$$\mathbb{P} \left( \frac{1}{N} \mathbf{d}(\tilde{\mu}_{\beta, N} || \mu_{\beta, N}) \leq \varepsilon_N \right) \geq 1 - \varepsilon_N^{1/2} \rightarrow 1, \quad \text{as } N \rightarrow \infty. \quad (3.1.12)$$

We provide the proof of Corollary 3.1.11 below as it is short.

*Proof of Corollary 3.1.11.* Note that the choice of  $M_N$  satisfies the assumption of Theorem 3.1.10, so the first statement follows directly from Theorem 3.1.10. Next, as mentioned in Remark 3.1.8, the running time of Algorithm 3.1 is deterministic and is bounded by  $\lceil N/M_N \rceil 2^{M_N}$ . With our choice of  $M_N$ , we conclude that

$$\lceil N/M_N \rceil 2^{M_N} \leq N \cdot 2^{\varepsilon \log_2 N} \leq N^{1+\varepsilon},$$

and the proof is completed. ■

### Supercritical regime $\beta > \beta_G$ : hardness for generic algorithms

Now we assume that  $A$  is non-concave, so  $\beta_G < \infty$ . For  $\beta > \beta_G$ , we provide the following hardness result for the class of algorithms satisfying Definition 3.1.2.

**Theorem 3.1.13** (Hardness). *Suppose that  $A$  is non-concave. Let  $\beta > \beta_G$ . For any algorithm satisfying Definition 3.1.2 that approximates the Gibbs measure with probability approaching 1, there exists  $\gamma > 0$  such that*

$$\lim_{N \rightarrow \infty} \mathbb{P} (\tau \geq e^{\gamma N}) = 1,$$

where  $\tau$  is the running time of the algorithm.

### 3.1.3 Discussion and related work

A natural way to sample from the Gibbs measure is via the Markov chain Monte Carlo (MCMC) method. In [147], Nascimento and Fontes studied a Metropolis dynamics on the GREM, where the state space of this dynamics is the set of leaves. They showed that for all  $\beta > 0$ , the spectral gap of the Metropolis dynamics decays exponentially to 0 as  $N \rightarrow \infty$  almost surely, which hinted that the MCMC method might not be the best way to approximate the Gibbs measure efficiently.

The current work is largely inspired by the previous work of Addario-Berry and Maillard [2] on finding the near maximum (ground state) of the CREM. Bovier and Kurkova showed in [54] that the maximum of the CREM satisfies

$$x_{GSE} := \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \left[ \max_{|u|=N} X_u \right] = \sqrt{2 \log 2} \int_0^1 \sqrt{\hat{a}(s)} \, ds. \quad (3.1.13)$$

With this result in mind, the problem that Addario-Berry and Maillard addressed can be phrased as the following optimization problem.

**Problem 3.1.14.** *For what kind of  $A$  such that for all  $\varepsilon > 0$ , there exists a polynomial-time algorithm that can find a vertex  $|u| = N$  such that  $X_u \geq (x_{GSE} - \varepsilon)N$  with high probability?*

To respond to Problem 3.1.14, they showed the following phase transition: there exists a threshold

$$x_* = \sqrt{2 \log 2} \int_0^1 \sqrt{a(s)} \, ds$$

such that for any  $x < x_*$ , there exists a linear time algorithm that finds  $X_v \geq xN$  with high probability; for any  $x > x_*$ , there exists  $z > 0$  such that with high probability, it takes at least  $e^{zN}$  queries to find  $X_u \geq xN$ . Since  $x_{GSE} \geq x_*$  with equality holding if and only if  $A$  is concave, the near maximum can be found if and only if  $A$  is concave. Another remark is that their result correspond to the special case of our result where  $\beta \rightarrow \infty$ , and linear algorithm they proposed is similar to Algorithm 3.1.

Problem 3.1.14 also appeared in the context of mean-field spin glass. It is known that a generalized Thouless–Anderson–Palmer approach proposed by Subag in [172] gives a tree structure from the origin to the spin space when  $\beta = \infty$ . This picture allows Subag to show in [171] that for the full-RSB spherical spin glasses, a greedy type algorithm that exploits this tree structure gives an efficient way to find a near maximum of these models. On the other hand, it was conjectured by the physicists that when  $\beta \rightarrow \infty$ , the SK model also exhibits the full-RSB property (see [143]). By assuming this conjecture, Montanari solved Problem 3.1.14 for the SK model in [145] via the so-called approximate message passing (AMP) type algorithm, where one example of the AMP algorithms was Bolthausen’s iteration scheme [50] which solves the so-called TAP equation. Later, Alaoui, Montanari and Sellke [8] extended Montanari’s previous result to other mean-field spin glasses that do not exhibit the overlap gap property.

The problem of sampling from the Gibbs measure was also considered the context of mean-field spin glasses. This problem was usually attacked by introducing the Glauber dynamics, which also belongs to the MCMC method. For the Sherrington–Kirkpatrick model, physicists (see [169, 143]) expected fast convergence to the Gibbs measure in the whole high temperature regime  $\beta < 1$ . Recently, it was shown by Bauerschmidt and Bodineau in [28] and by Eldan, Koehler and Zeitouni in [98] that fast mixing occurs when  $\beta < 1/4$ . Moreover, Eldan et al. showed in [98] that the Gibbs measure satisfies a Poincaré inequality for the Dirichlet form of Glauber dynamics, so the Glauber dynamics mixes in  $O(N^2)$  spin flips in total variation distance. Subsequently, this estimate was improved to  $O(N \log N)$  by Anari et al. in [13].

For spherical spin glasses, Gheissari and Jagannath in [108] that Langevin dynamics (continuum version of Glauber dynamics) have a polynomial spectral gap for  $\beta$  small. On the other hand, Ben Arous and Jagannath proved in [37] that for  $\beta$  sufficiently large, the mixing times of Glauber and Langevin dynamics are exponentially large in Ising and spherical spin glasses, respectively.

In [9], Alaoui, Montanari and Sellke proposed a non MCMC type algorithm based on the stochastic localization for the SK model. They showed that for  $\beta < 1/2$ , there exists an algorithm with complexity  $O(N^2)$  with output law being close to the Gibbs measure in normalized Wasserstein distance. Moreover, for  $\beta > 1$ , they established a hardness result for the stable algorithms, which means that the output law of these algorithms are stable under small random perturbation of the defining matrix of the SK model. The hardness result for  $\beta > 1$  was proven by utilizing the disorder chaos, which means for them that Wasserstein distance between the Gibbs measure and the perturbed Gibbs measure is bounded from below by a positive constant for arbitrary small random perturbation.

**Overlap gap property and algorithmic hardness.** The overlap gap property, emerging from studying of mean-field spin glasses, seems to be an obstruction of many optimization algorithms for random structures. See Gamarnik [102] for a survey. In the context of the CREM, for a given  $\beta > 0$ , the *overlap distribution* is the limiting law (as  $N \rightarrow \infty$ ) of the overlap  $\frac{|u \wedge w|}{N}$  of two vertices  $u$  and  $w$  sampled independently according to the Gibbs measure with inverse temperature  $\beta$ . The CDF of the limiting overlap distribution  $\alpha_\beta : [0, 1] \rightarrow [0, 1]$  is defined as

$$\alpha_\beta(t) := \lim_{N \rightarrow \infty} \mathbb{E} \left[ \sum_{|u|=N} \sum_{|w|=N} \mu_{\beta,N}(u) \mu_{\beta,N}(w) \mathbf{1}_{|u \wedge w|/N \leq t} \right],$$

The overlap gap property in the context of the CREM means that  $\alpha_\beta(t)$  is equal to a constant strictly less than 1 in an interval  $[t_1, t_2] \subseteq [0, 1]$ . On the other hand, it is known in [54] that  $\alpha_\beta(t)$  satisfies the following

$$\alpha_\beta(t) = \begin{cases} \frac{\sqrt{2 \log 2}}{\beta \sqrt{\hat{a}(t)}}, & t \leq t_0(\beta), \\ 1, & t > t_0(\beta). \end{cases}$$

When  $A$  is concave, the CREM does not exhibit the overlap gap property for any  $\beta > 0$ , which does not contradict the picture mentioned in the previous paragraph. On the other hand, when  $A$  is non-concave, the CREM has the overlap gap property if and only if  $\beta > \beta'_G = \sqrt{2 \log 2} / \sqrt{\hat{a}(t_G)}$ . Comparing with the hardness threshold  $\beta_G$  defined in (3.1.8), we see  $\beta_G < \beta'_G$ , which means that some extra ingredients are needed to explain the algorithmic hardness we observe in the present work.

**Further direction.** Corollary 3.1.11 implies that if  $A$  is concave, for any  $\beta \in [0, \infty)$ , then the sequence of algorithm constructed from Algorithm 3.1 can approximate the Gibbs measure in the sense of Definition 3.1.6. One might ask whether a higher precision is achievable. Namely, let  $\alpha \in [0, 1]$ . Given  $\beta > 0$ , does there exist a sequence of algorithms with corresponding output laws  $\tilde{\mu}_{\beta,N}$  such that

$$\lim_{N \rightarrow \infty} \frac{1}{N^\alpha} \mathbf{d}(\tilde{\mu}_{\beta,N} || \mu_{\beta,N}) = 0,$$

with probability approaching 1? Note that for a general class of branching random walks, with  $\alpha = 0$  and  $\beta > \beta_c$ , it is shown in [112] that with positive probability, this task has a running time of stretched exponential. The result in [112] is derived from the fluctuation of the sampled path of supercritical Gibbs measure done by [79].

**Outline.** The paper is organized as follows. In Section 3.2, we prove that the KL divergence  $\mathbf{d}(\mu_{\beta,M,N}||\mu_{\beta,N})$  can be decomposed into weight sums of free energies on subtrees, and we also compute its expectation. This information readily allows us to prove Theorem 3.1.9 which is provided in Section 3.2.1. Building on the decomposition of the KL divergence provided in Section 3.2.1, we study in Section 3.3 the renormalized limit of  $\mathbb{E}[\mathbf{d}(\mu_{\beta,M,N}||\mu_{\beta,N})]$ . This leads to the proof of Theorem 3.1.10 which is provided at the end of the introduction of Section 3.3. In Section 3.4, we show that for  $A$  non-concave and  $\beta > \beta_G$ , the Gibbs measure tends to sample a rare event. Based on this observation, Theorem 3.1.13 is proven in Section 3.5, where the details are provided at the end of the introduction of Section 3.5. In Appendix 3.A, we provide a lower bound of the free energy  $F_\beta$  that may be of independent interest. Finally, in Appendix 3.B, we provide the details of the proof of Lemma 3.3.2.

## 3.2 Decomposition of the KL divergence $\mathbf{d}(\mu_{\beta,M,N}||\mu_{\beta,N})$

In this section, we provide in the following proposition a simple decomposition of the KL divergence  $\mathbf{d}(\mu_{\beta,M,N}||\mu_{\beta,N})$  in terms of a sum of free energies on subtrees.

**Proposition 3.2.1.** *For all  $\beta > 0$  and for any two integers  $M, N \in \mathbb{N}$  such that  $M \leq N$ , we have*

$$\mathbf{d}(\mu_{\beta,M,N}||\mu_{\beta,N}) = \log Z_{\beta,N} - \sum_{k=0}^{\lfloor N/M \rfloor} \sum_{|u|=kM} \mu_{\beta,M,kM}(u) \cdot \log Z_{\beta,M \wedge (N-kM)}^u.$$

*Proof.* By (3.1.7) the definition of the KL divergence,

$$\begin{aligned}
 & \mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N}) \\
 &= \sum_{|u|=N} \mu_{\beta,M,N}(u) \cdot (\log Z_{\beta,N} - \log Z_{\beta,M,N}(u)) \quad (\text{By (3.1.2) and (3.1.10)}) \\
 &= \sum_{|u|=N} \mu_{\beta,M,N}(u) \cdot \left( \log Z_{\beta,N} - \sum_{k=0}^{\lfloor N/M \rfloor} \log Z_{\beta,M \wedge (N-kM)}^{u[kM]} \right) \\
 &= \log Z_{\beta,N} - \sum_{|u|=N} \sum_{k=0}^{\lfloor N/M \rfloor} \mu_{\beta,M,N}(u) \cdot \log Z_{\beta,M \wedge (N-kM)}^{u[kM]} \\
 &= \log Z_{\beta,N} - \sum_{k=0}^{\lfloor N/M \rfloor} \sum_{|u|=N} \mu_{\beta,M,N}(u) \cdot \log Z_{\beta,M \wedge (N-kM)}^{u[kM]} \\
 &= \log Z_{\beta,N} - \sum_{k=0}^{\lfloor N/M \rfloor} \sum_{|w|=kM} \sum_{|w'|=N-kM} \mu_{\beta,M,N}(ww') \cdot \log Z_{\beta,M \wedge (N-kM)}^w \\
 &= \log Z_{\beta,N} - \sum_{k=0}^{\lfloor N/M \rfloor} \sum_{|w|=kM} \mu_{\beta,M,kM}(w) \cdot \log Z_{\beta,M \wedge (N-kM)}^w, \quad (\text{By (3.1.9)})
 \end{aligned}$$

the proof is completed.  $\blacksquare$

The next proposition asserts that the expectation of the KL divergence  $\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N})$  can be written as the difference between the free energy of the CREM and the sum of free energies on the subtrees.

**Proposition 3.2.2.** *For all  $\beta > 0$  and for any two integers  $M, N \in \mathbb{N}$  such that  $M \leq N$ , the expectation of the KL divergence from  $\mu_{\beta,M,N}$  to  $\mu_{\beta,N}$  admits the following decomposition.*

$$\mathbb{E} [\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N})] = \mathbb{E} [\log Z_{\beta,N}] - \sum_{k=1}^{\lfloor N/M \rfloor} \mathbb{E} \left[ \log Z_{\beta,M \wedge (N-kM)}^{(kM)} \right].$$

*Proof.* Combining Proposition 3.2.1, the branching property and the law of iterated expectation, we have

$$\begin{aligned}
 & \mathbb{E} [\mathbf{d}(\mu_{\beta,M,N} || \mu_{\beta,N})] \\
 &= \mathbb{E} [\log Z_{\beta,N}] - \mathbb{E} \left[ \sum_{k=0}^{\lfloor N/M \rfloor} \sum_{|u|=kM} \mu_{\beta,M,kM}(u) \cdot \log Z_{\beta,M \wedge (N-kM)}^u \right] \\
 &= \mathbb{E} [\log Z_{\beta,N}] - \mathbb{E} \left[ \sum_{k=0}^{\lfloor N/M \rfloor} \sum_{|u|=kM} \mu_{\beta,M,kM}(u) \cdot \mathbb{E} \left[ \log Z_{\beta,M \wedge (N-kM)}^u \mid \mathcal{F}_{kM} \right] \right] \\
 &= \mathbb{E} [\log Z_{\beta,N}] - \sum_{k=0}^{\lfloor N/M \rfloor} \mathbb{E} \left[ \sum_{|u|=kM} \mu_{\beta,M,kM}(u) \right] \mathbb{E} \left[ \log Z_{\beta,M \wedge (N-kM)}^{(kM)} \right] \\
 &= \mathbb{E} [\log Z_{\beta,N}] - \sum_{k=0}^{\lfloor N/M \rfloor} \mathbb{E} \left[ \log Z_{\beta,M \wedge (N-kM)}^{(kM)} \right], \quad (3.2.1)
 \end{aligned}$$

where (3.2.1) follows from the fact that  $\mu_{\beta,M,kM}$  is a probability measure.  $\blacksquare$

### 3.2.1 Proof of Theorem 3.1.9

The following argument is similar to the proof of (1.9) in [112], where the difference is that we use the concentration inequalities of free energies to control certain terms.

Let  $p \geq 1$ . By Proposition 3.2.1 and Minkowski's inequality, we have

$$\begin{aligned} & \frac{1}{N} \|\mathbf{d}(\mu_{\beta,M,N} \|\mu_{\beta,N}) - \mathbb{E}[\mathbf{d}(\mu_{\beta,M,N} \|\mu_{\beta,N})]\|_p \\ & \leq \frac{1}{N} \sum_{k=0}^{\lfloor N/M \rfloor} \left\| \sum_{|u|=kM} \mu_{\beta,M,kM}(u) \cdot \left( \log Z_{\beta,M \wedge (N-kM)}^u - \mathbb{E} \left[ \log Z_{\beta,M \wedge (N-kM)}^u \right] \right) \right\|_p. \end{aligned} \quad (3.2.2)$$

Applying Jensen's inequality to  $\mu_{\beta,M,kM}$  and the fact that  $x \mapsto |x|^p$  is convex for all  $p \geq 1$ , we obtain

$$\begin{aligned} & \mathbb{E} \left[ \left| \sum_{|u|=kM} \mu_{\beta,M,kM}(u) \cdot \left( \log Z_{\beta,M \wedge (N-kM)}^u - \mathbb{E} \left[ \log Z_{\beta,M \wedge (N-kM)}^u \right] \right) \right|^p \right] \\ & \leq \mathbb{E} \left[ \sum_{|u|=kM} \mu_{\beta,M,kM}(u) \cdot \left| \log Z_{\beta,M \wedge (N-kM)}^u - \mathbb{E} \left[ \log Z_{\beta,M \wedge (N-kM)}^u \right] \right|^p \right]. \end{aligned} \quad (3.2.3)$$

Then by the law of iterated expectation and the branching property, the expectation (3.2.3) above equals

$$\begin{aligned} & \mathbb{E} \left[ \sum_{|u|=kM} \mu_{\beta,M,kM}(u) \cdot \mathbb{E} \left[ \left| \log Z_{\beta,M \wedge (N-kM)}^u - \mathbb{E} \left[ \log Z_{\beta,M \wedge (N-kM)}^u \right] \right|^p \mid \mathcal{F}_{kM} \right] \right] \\ & = \mathbb{E} \left[ \left| \log Z_{\beta,M \wedge (N-kM)}^{(kM)} - \mathbb{E} \left[ \log Z_{\beta,M \wedge (N-kM)}^{(kM)} \right] \right|^p \right]. \end{aligned} \quad (3.2.4)$$

Now, the concentration inequality of free energies (see, Theorem 1.2 in [156]) implies that for all  $p \geq 1$ ,

$$\begin{aligned} & \mathbb{E} \left[ \left| \log Z_{\beta,M \wedge (N-kM)}^{(kM)} - \mathbb{E} \left[ \log Z_{\beta,M \wedge (N-kM)}^{(kM)} \right] \right|^p \right] \\ & \leq \int_0^\infty 2 \exp \left( - \frac{x^{2/p}}{4\beta^2 N \left( A \left( \frac{kM + M \wedge (N-kM)}{N} \right) - A \left( \frac{kM}{N} \right) \right)} \right) dx \\ & = \beta^p N^{p/2} \left( A \left( \frac{kM + M \wedge (N-kM)}{N} \right) - A \left( \frac{kM}{N} \right) \right)^{p/2} \cdot \underbrace{2^{p/2+1} p \int_0^\infty \exp \left( - \frac{y^2}{2} \right) y^{p-1} dy}_{=: C_1(p)} \\ & = \beta^p N^{p/2} \left( A \left( \frac{kM + M \wedge (N-kM)}{N} \right) - A \left( \frac{kM}{N} \right) \right)^{p/2} \cdot C_1(p). \end{aligned} \quad (3.2.5)$$

Combining (3.2.2), (3.2.4) and (3.2.5) and letting  $C_2(p) = C_1(p)^{1/p}$ , we derive that

$$\begin{aligned}
 & \frac{1}{N} \|\mathbf{d}(\mu_{\beta, M, N} \|\mu_{\beta, N}) - \mathbb{E}[\mathbf{d}(\mu_{\beta, M, N} \|\mu_{\beta, N})]\|_p \\
 & \leq \frac{\beta C_2(p)}{\sqrt{N}} \sum_{k=0}^{\lfloor N/M \rfloor} \sqrt{\left( A\left(\frac{kM + M \wedge (N - kM)}{N}\right) - A\left(\frac{kM}{N}\right) \right)} \\
 & \leq \frac{\beta C_2(p)}{\sqrt{N}} \underbrace{\sqrt{\sum_{k=0}^{\lfloor N/M \rfloor} \left( A\left(\frac{kM + M \wedge (N - kM)}{N}\right) - A\left(\frac{kM}{N}\right) \right)}}_{=1} \sqrt{\sum_{k=0}^{\lfloor N/M \rfloor} 1} \quad (3.2.6) \\
 & = \frac{\beta C_2(p)}{\sqrt{N}} \sqrt{\left\lceil \frac{N}{M} \right\rceil} \\
 & \leq \frac{\beta C_2(p) \sqrt{2}}{\sqrt{M}}, \quad (3.2.7)
 \end{aligned}$$

where (3.2.6) is derived from the Cauchy–Schwarz inequality, and (3.2.7) follows from bounding  $\sqrt{\lfloor N/M \rfloor / N}$  by  $\sqrt{2/M}$ . By choosing  $C_p = C_2(p)\sqrt{2}$ , the proof of Theorem 3.1.9 is completed.

### 3.3 Asymptotics of the KL divergence $\mathbf{d}(\mu_{\beta, M, N} \|\mu_{\beta, N})$

The goal of this section is to prove Theorem 3.1.10. In view of (3.1.4) and Proposition 3.2.2, it remains to show the following proposition.

**Proposition 3.3.1.** *Let  $M_N$  be a sequence such that  $M_N \in \llbracket 1, N \rrbracket$  and  $M_N \rightarrow \infty$ . Then,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^{\lfloor N/M_N \rfloor} \mathbb{E} \left[ \log Z_{\beta, M_N \wedge (N - kM_N)}^{(kM_N)} \right] = \tilde{F}_\beta.$$

The proof of Proposition 3.3.1 is postponed to Section 3.3.1. Conditioned on Proposition 3.3.1, we are now ready to prove Theorem 3.1.10.

*Proof of Proposition 3.1.10.* Fix  $\beta < \beta_G$ . Let  $M_N$  be a sequence such that  $M_N \in \llbracket 1, N \rrbracket$ , and  $M_N \rightarrow \infty$ . By Fact 3.A.2 and Proposition 3.3.1,

$$\begin{aligned}
 \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\mathbf{d}(\mu_{\beta, M_N, N} \|\mu_{\beta, N})] &= \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\log Z_{\beta, N}] - \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^{\lfloor N/M_N \rfloor} \mathbb{E} \left[ \log Z_{\beta, M_N \wedge (N - kM_N)}^{(kM_N)} \right] \\
 &= F_\beta - \tilde{F}_\beta.
 \end{aligned}$$

Now, by Proposition 3.A.1,  $F_\beta - \tilde{F}_\beta = 0$  for all  $\beta \leq \beta_G$  and  $F_\beta - \tilde{F}_\beta > 0$  for all  $\beta > \beta_G$ . This completes the proof.  $\blacksquare$

### 3.3.1 Proof of Proposition 3.3.1

The proof of Proposition 3.3.1 is based on the following lemma.

**Lemma 3.3.2.** *Let  $M_N$  be a sequence such that  $M_N \in \llbracket 1, N \rrbracket$  and  $M_N \rightarrow \infty$ . For all  $k \in \llbracket 0, \lfloor N/M_N \rfloor \rrbracket$ , define*

$$a_k^- := \operatorname{ess\,inf}_{t \in [\frac{kM_N}{N}, \frac{(k+1)M_N}{N}]} a(t) \quad \text{and} \quad a_k^+ := \operatorname{ess\,sup}_{t \in [\frac{kM_N}{N}, \frac{(k+1)M_N}{N}]} a(t). \quad (3.3.1)$$

Then for all  $\varepsilon > 0$ , there exists  $N_0 \in \mathbb{N}$  such that for all  $N \geq N_0$ ,

$$f(\beta\sqrt{a_k^-}) - \varepsilon\beta\sqrt{a_k^-} \leq \frac{1}{M_N} \mathbb{E} \left[ \log Z_{\beta, M}^{(kM)} \right] \leq f(\beta\sqrt{a_k^+}) + \varepsilon\beta\sqrt{a_k^+},$$

for all  $k \in \llbracket 0, \lfloor N/M_N \rfloor \rrbracket$ .

The proof of Lemma 3.3.2 is based on comparing the free energy of the CREM with the free energy of the so-called branching random walk, which is a CREM with  $A$  equal to the identity function. While the proof of Lemma 3.3.2 is rather standard, the proof requires some standard properties of the free energy of the branching random walk, so we postpone the proof of Lemma 3.3.2 to Appendix 3.B.

We now proceed to the proof of Proposition 3.3.1.

*Proof of Proposition 3.3.1.* Fix  $\varepsilon > 0$ . Fix  $M_N$  being a sequence such that  $M_N \in \llbracket 1, N \rrbracket$  and  $M_N \rightarrow \infty$ . We denote  $K_N = \lfloor N/M_N \rfloor$  for simplicity. First of all, note that

$$\frac{1}{N} \sum_{k=1}^{K_N} \mathbb{E} \left[ \log Z_{\beta, M_N}^{(kM_N)} \right] = \frac{1}{N} \sum_{k=1}^{K_N-1} \mathbb{E} \left[ \log Z_{\beta, M_N}^{(kM_N)} \right] + \frac{1}{N} \mathbb{E} \left[ \log Z_{\beta, N - K_N M_N}^{(K_N M_N)} \right]. \quad (3.3.2)$$

We claim that the second term of (3.3.2) converges to 0. For any  $|u| = N - \lfloor N/M_N \rfloor M_N$ ,

$$\mathbb{E} \left[ \log Z_{\beta, N - \lfloor N/M_N \rfloor M_N}^{(\lfloor N/M_N \rfloor M_N)} \right] \geq \mathbb{E} \left[ \beta X_u^{(\lfloor N/M_N \rfloor M_N)} \right] = 0.$$

Now, we turn to the upper bound. By Jensen's inequality,

$$\begin{aligned} \frac{1}{N} \mathbb{E} \left[ \log Z_{\beta, N - \lfloor N/M_N \rfloor M_N}^{\lfloor N/M_N \rfloor M_N} \right] &\leq \frac{1}{N} \log \mathbb{E} \left[ Z_{\beta, N - \lfloor N/M_N \rfloor M_N}^{\lfloor N/M_N \rfloor M_N} \right] \\ &= \log 2 \left( 1 - \left\lfloor \frac{N}{M_N} \right\rfloor \frac{M_N}{N} \right) + \left( A(1) - A \left( \left\lfloor \frac{N}{M_N} \right\rfloor \frac{M_N}{N} \right) \right) \rightarrow 0, \end{aligned}$$

as  $N \rightarrow \infty$ , which proves that the second term of (3.3.2) converges to 0.

It remains to show that the first term of (3.3.2) converges to  $\tilde{F}_\beta$ . By Lemma 3.3.2,

$$\frac{M_N}{N} \sum_{k=1}^{K_N-1} f(\beta\sqrt{a_k^-}) - \varepsilon\beta\sqrt{a_k^-} \leq \frac{1}{N} \sum_{k=1}^{K_N-1} \mathbb{E} \left[ \log Z_{\beta, M}^{(kM)} \right] \leq \frac{M_N}{N} \sum_{k=1}^{K_N-1} f(\beta\sqrt{a_k^+}) + \varepsilon\beta\sqrt{a_k^+}. \quad (3.3.3)$$

Because the function  $a(\cdot)$  is Riemann integrable and  $f(\beta\sqrt{\cdot})$  is continuous, their composition  $f(\beta\sqrt{a(\cdot)})$  is also Riemann integrable. Similarly,  $\sqrt{a(\cdot)}$  is also Riemann integrable. Thus, by taking  $N \rightarrow \infty$ , (3.3.3) yields

$$\lim_{N \rightarrow \infty} \left| \frac{1}{N} \sum_{k=1}^{K_N-1} \mathbb{E} \left[ \log Z_{\beta, M}^{(kM)} \right] - \int_0^1 f(\beta\sqrt{a(s)}) ds \right| \leq \varepsilon \beta \int_0^1 \sqrt{a(s)} ds. \quad (3.3.4)$$

Since  $\varepsilon > 0$  is arbitrary chosen, (3.3.4) implies that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^{K_N-1} \mathbb{E} \left[ \log Z_{\beta, M}^{(kM)} \right] = \int_0^1 f(\beta\sqrt{a(s)}) ds = \tilde{F}_\beta,$$

as desired. ■

### 3.4 A property of the Gibbs measure in the supercritical regime

From now on, we assume that  $A$  is non-concave, so  $\beta_G < \infty$ . Also, we suppose that  $\beta > \beta_G$ . The goal of this section is to show that the Gibbs measure tends to sample a vertex that has an ancestor that jumps exceptionally high. The meaning of having an ancestor that jumps exceptionally high is quantified in the following definition.

**Definition 3.4.1.** Given  $z > 0$ ,  $K \in \mathbb{N}$  and a CREM  $\mathbf{X}$ , a vertex  $v \in \mathbb{T}_N$  with  $|v| = n \in \llbracket 1, N \rrbracket$  is said to have a  $(z, K, \mathbf{X})$ -steep ancestor if there exists  $k \in \llbracket 1, \lfloor nK/N \rfloor \rrbracket$  such that

$$X_{v[\lfloor Nk/K \rfloor]} - X_{v[\lfloor N(k-1)/K \rfloor]} > N \sqrt{2 \log 2(1+z)a_k},$$

where  $a_k = (A(k/K) - A((k-1)/K))/K$ .

The goal of this section can now be phrased as the following proposition.

**Proposition 3.4.2.** *Let  $\beta > \beta_G$ . There exist  $z > 0$ ,  $K \in \mathbb{N}$  such that, for all  $\delta > 0$  sufficiently small,*

$$\lim_{N \rightarrow \infty} \mathbb{P} \left( \sum_{|u|=N} \mu_{\beta, N}(u) \mathbf{1}\{u \text{ has a } (z, K, \mathbf{X})\text{-steep ancestor}\} > 1 - e^{-\delta N} \right) = 1.$$

The proof of Proposition 3.4.2 is based on the following lemma which states the free energy converges to  $F_\beta$  in probability.

**Lemma 3.4.3.** *For all  $\beta > 0$ , for all  $\varepsilon > 0$ , we have*

$$\lim_{N \rightarrow \infty} \mathbb{P} \left( \left| \frac{1}{N} \log Z_{\beta, N} - F_\beta \right| > \varepsilon \right) = 0.$$

*Proof.* For all  $\beta > 0$ , for all  $\varepsilon > 0$ , the concentration inequality of free energies (see, Theorem 1.2 in [156]) states that

$$\mathbb{P} \left( \left| \frac{1}{N} \log Z_{\beta, N} - \frac{1}{N} \mathbb{E} [\log Z_{\beta, N}] \right| > \varepsilon \right) \leq 2 \exp \left( -\frac{\varepsilon^2}{4\beta^2} N \right).$$

Then, the proof is completed by incorporating Fact 3.A.2. ■

We now prove Proposition 3.4.2.

*Proof of Proposition 3.4.2.* For all  $u \in \partial\mathbb{T}_N$ , let  $A_u$  be the set where  $u$  does not have a  $(z, K, \mathbf{X})$ -steep ancestor defined as

$$A_u := \left\{ \forall k \in \llbracket 1, K \rrbracket : X_{u[\lfloor Nk/K \rfloor]} - X_{u[\lfloor N(k-1)/K \rfloor]} \leq N \sqrt{2 \log 2 (1+z) a_k} \right\}. \quad (3.4.1)$$

To prove Proposition 3.4.2, it suffices to show that there exist  $K \in \mathbb{N}$  and  $z > 0$  such that for all  $\delta > 0$  sufficiently small,

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left( \sum_{|u|=N} \frac{e^{\beta X_u}}{Z_{\beta, N}} \mathbf{1}_{A_u} \geq e^{-\delta N} \right) = 0,$$

Because the function  $a(\cdot)$  is Riemann integrable and  $f(\beta\sqrt{\cdot})$  is continuous, their composition  $f(\beta\sqrt{a(\cdot)})$  is also Riemann integrable. On the other hand, since  $\beta > \beta_G$ , Proposition 3.A.1 implies that  $F_\beta - \tilde{F}_\beta > 0$ . Therefore, we can choose  $z > 0$  sufficiently small and  $K \in \mathbb{N}$  sufficiently large such that

$$F_\beta - (1+z) \frac{1}{K} \sum_{k=1}^K \max_{s \in \llbracket (k-1)/K, k/K \rrbracket} f(\beta\sqrt{a(s)}) \geq C > 0, \quad (3.4.2)$$

for some  $C > 0$ . In the rest of the proof, we fix our choice of  $z$  and  $K$ . We also fix  $\delta > 0$  and  $c > 0$  sufficiently small such that  $C - \delta - c > 0$ .

Now,

$$\begin{aligned} & \mathbb{P} \left( \sum_{|u|=N} \frac{e^{\beta X_u}}{Z_{\beta, N}} \mathbf{1}_{A_u} \geq e^{-\delta N} \right) \\ & \leq \mathbb{P} (Z_{\beta, N} < \exp(F_\beta(1-c)N)) \\ & + \mathbb{P} \left( \left\{ \sum_{|u|=N} \frac{e^{\beta X_u}}{Z_{\beta, N}} \mathbf{1}_{A_u} \geq e^{-\delta N} \right\} \cap \left\{ Z_{\beta, N} \geq \exp(F_\beta(1-c)N) \right\} \right) \\ & \leq \mathbb{P} (Z_{\beta, N} < \exp(F_\beta(1-c)N)) + \mathbb{P} \left( \sum_{|u|=N} e^{\beta X_u} \mathbf{1}_{A_u} \geq e^{-\delta N} e^{F_\beta(1-c)N} \right). \quad (3.4.3) \end{aligned}$$

By Lemma 3.4.3, the first term in (3.4.3) tends to 0 as  $N \rightarrow \infty$ . Thus, it remains to prove the second probability in (3.4.3) converges to 0 as  $N \rightarrow \infty$ . Let  $Y_k \sim$

$N(0, NKa_{k,N})$  and  $a_{k,N} := (A(\lfloor kN/K \rfloor/N) - A(\lfloor (k-1)N/K \rfloor/N))/K$  for all  $k \in \llbracket 1, K \rrbracket$ . By completing the square, we have

$$\begin{aligned} & \mathbb{E} \left[ \sum_{|u|=N} e^{\beta X_u} \mathbf{1}_{A_u} \right] \\ & \leq 2^N \prod_{k=1}^K e^{\beta^2 NKa_{k,N}/2} \mathbb{P} \left( Y_k \leq N(\sqrt{2 \log 2(1+z)a_k} - \beta K a_{k,N}) \right) \\ & = \prod_{k=1}^K 2^{N/K} e^{\beta^2 NKa_{k,N}/2} \mathbb{P} \left( Y_k \leq N(\sqrt{2 \log 2(1+z)a_k} - \beta K a_{k,N}) \right). \end{aligned} \quad (3.4.4)$$

**Case 1.** If  $\sqrt{2 \log 2(1+z)a_k} < \beta K a_{k,N}$ , the Chernoff bound yields

$$\begin{aligned} & 2^{N/K} e^{\beta^2 NKa_{k,N}/2} \mathbb{P} \left( Y_k \leq N(\sqrt{2 \log 2(1+z)a_k} - \beta K a_{k,N}) \right) \\ & \leq 2^{N/K} e^{\beta^2 NKa_{k,N}/2} \exp \left( -\frac{N^2(\sqrt{2 \log 2(1+z)a_k} - \beta K a_{k,N})^2}{2NKa_{k,N}} \right) \\ & = 2^{N/K} e^{\beta^2 NKa_{k,N}/2} \exp(-N \log 2(1+z)a_k/Ka_{k,N}) \exp \left( \beta N \sqrt{2 \log 2(1+z)a_k} \right) \exp(-N\beta^2 Ka_{k,N}/2) \\ & = \underbrace{\exp(-N(\log 2)za_k/Ka_{k,N})}_{\leq 1} \exp(N \log 2(1 - a_k/a_{k,N})/K) \exp \left( N\beta \sqrt{2 \log 2(1+z)a_k} \right) \\ & \leq \exp(N \log 2(1 - a_k/a_{k,N})/K) \exp \left( N\beta \sqrt{2 \log 2(1+z)a_k} \right). \end{aligned}$$

**Case 2.** If  $\sqrt{2 \log 2(1+z)a_k} \geq \beta K a_{k,N}$ , we simply bound the probability in (3.4.4) by 1 and obtain

$$\begin{aligned} & 2^{N/K} e^{\beta^2 NKa_{k,N}/2} \mathbb{P} \left( Y_k \leq N(\sqrt{2 \log 2(1+z)a_k} - \beta K a_{k,N}) \right) \\ & \leq 2^{N/K} e^{\beta^2 NKa_{k,N}/2} \end{aligned}$$

Then, by (3.4.4) and the two cases above, we have

$$\begin{aligned} & \limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E} \left[ \sum_{|u|=N} e^{\beta X_u} \mathbf{1}_{A_u} \right] \\ & \leq \sum_{\sqrt{2 \log 2(1+z)a_k} < \beta K a_k} \beta \sqrt{2 \log 2(1+z)a_k} + \sum_{\sqrt{2 \log 2(1+z)a_k} \geq \beta K a_k} \frac{\log 2}{K} + \frac{\beta^2 K a_k}{2} \\ & \leq (1+z) \frac{1}{K} \sum_{k=1}^K f(\beta/(1+z)\sqrt{a_k}K) \\ & \leq (1+z) \frac{1}{K} \sum_{k=1}^K \max_{s \in \llbracket (k-1)/K, k/K \rrbracket} f(\beta \sqrt{a(s)}), \end{aligned} \quad (3.4.5)$$

where (3.4.5) follows from monotonicity of the function  $f$ . By the Markov inequality, (3.4.2), the second term in (3.4.3) satisfies the following

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} \left( \sum_{|u|=N} e^{\beta X_u} \mathbf{1}_{A_u} \geq e^{-\delta N} e^{F_\beta(1-c)N} \right) \leq -C + c + \delta < 0. \quad (3.4.6)$$

where  $C$ ,  $c$  and  $\delta$  are chosen as in the first paragraph of the proof. Combining (3.4.6) and Proposition 3.A.1, we conclude that

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left( \sum_{|u|=N} e^{\beta X_u} \mathbf{1}_{A_u} \geq e^{-\delta N} e^{F_\beta(1-c)N} \right) = 0,$$

and the proof is completed. ■

### 3.5 Hardness in the supercritical regime

Assume that  $A$  is non-concave and  $\beta > \beta_G$ . The goal of this section is to prove Theorem 3.1.13. Before we dive in the section, we introduce a few definitions. The first is a chain of subtrees defined as follows.

**Definition 3.5.1.** For  $v \in \mathbb{T}$ , let  $\mathcal{C}_v$  be a chain of subtrees containing  $v$  and all its ancestors defined by

$$\mathcal{C}_v = \bigcup_{k=0}^{\lfloor N|v|/K \rfloor} \mathbb{T}_{\lfloor N(k+1)/K \rfloor - \lfloor Nk/K \rfloor}^{v[\lfloor Nk/K \rfloor]}.$$

*Remark 3.5.2.* See Figure 3.1 for an illustration of Definition 3.5.1. Also, note that in particular,  $v \in \mathcal{C}_v$  for every  $v \in \mathbb{T}_N$ .

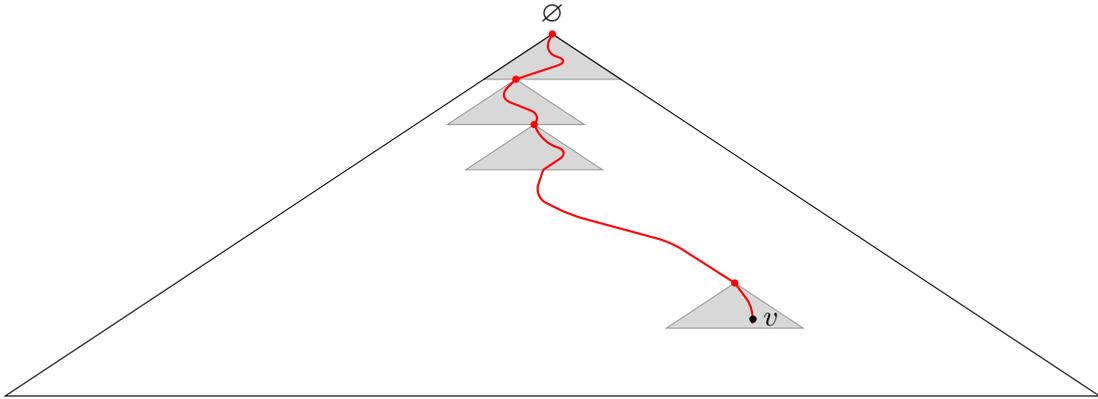


Figure 3.1: A schematic illustration of the set  $\mathcal{C}_v$  appearing in Definition 3.5.1. The largest isosceles triangle represents the binary tree  $\mathbb{T}_N$ . The chain of isosceles triangles colored in gray represent the set  $\mathcal{C}_v$ , where the  $k$ -th subtree has depth  $\lfloor N(k+1)/K \rfloor - \lfloor Nk/K \rfloor$ . The red dots represent  $v[\lfloor Nk/K \rfloor]$ , the ancestors of  $v$  at depth  $\lfloor Nk/K \rfloor$ .

Next, we introduce the following stopping time.

**Definition 3.5.3.** Let  $(v(n))_{n \in \mathbb{N}}$  be an algorithm. Define the stopping time  $\tau'$  as the first time  $n$  when the algorithm finds a vertex in  $\mathcal{C}_{v(n)}$  with a  $(z, K, \mathbf{X})$ -steep ancestor, given by:

$$\tau' = \inf \{n \in \mathbb{N} : \exists w \in \mathcal{C}_{v(n)} \text{ such that } w \text{ has a } (z, K, \mathbf{X})\text{-steep ancestor}\}.$$

Now, we come back to the proof of Theorem 3.1.13. The proof is based on the following two propositions. The first proposition asserts that the running time of an algorithm that approximates the Gibbs measure dominates the stopping time  $\tau'$  with probability approaching 1.

**Proposition 3.5.4.** *Suppose  $A$  to be non-concave. Let  $\beta > \beta_G$ . If  $\tau$  is the running time of an algorithm that approximates the Gibbs measure, then*

$$\lim_{N \rightarrow \infty} \mathbb{P}(\tau \geq \tau') = 1.$$

The proof of Proposition 3.5.4 is provided in Section 3.5.1. Note that Addario-Berry and Maillard proved in [2] a hardness result of finding a vertex  $v \in \partial\mathbb{T}_N$  such that  $X_v$  lies in a level set above a critical level, denoted by  $x_*N$ . In their case,  $\tau \geq \tau'$  holds deterministically because they showed in Lemma 3.1 in their paper that any for any vertex  $v \in \partial\mathbb{T}_N$  such that  $X_v$  lies in a level set above  $xN$ , where  $x > x_*$ ,  $v$  must have a  $(z, K, \mathbf{X})$ -ancestor. Nevertheless, Proposition 3.5.4 is sufficient for our purpose.

The second proposition to prove Theorem 3.1.13 is the following. The proposition asserts that the  $\tau'$  is exponentially large with probability approaching 1.

**Proposition 3.5.5.** *There exists  $\gamma > 0$  such that*

$$\lim_{N \rightarrow \infty} \mathbb{P}(\tau' > e^{\gamma N}) = 1.$$

Proposition 3.5.5 is proven following the same argument as in [2], and the proof is included in Section 3.5.3 for completeness.

Conditioned on Proposition 3.5.4 and Proposition 3.5.5, the proof of Theorem 3.1.13 is fairly short, so we provide it here.

*Proof of Theorem 3.1.13.* Fix  $\beta > \beta_G$ . Let  $(v(n))_{n \in \mathbb{N}}$  be an algorithm that approximates the Gibbs measure with probability approaching 1. Suppose that  $\tau$  is its running time and  $\tilde{\mu}_N$  is its output law, which is the law of  $v(\tau)$  conditioned on the CREM.

Now, combining Proposition 3.5.4 with Proposition 3.5.5, we conclude that there exists  $\gamma > 0$  such that

$$\lim_{N \rightarrow \infty} \mathbb{P}(\tau \geq e^{\gamma N}) \geq \lim_{N \rightarrow \infty} \mathbb{P}(\{\tau \geq \tau'\} \cap \{\tau' \geq e^{\gamma N}\}) = 1,$$

and the proof is completed. ■

### 3.5.1 Proof of Proposition 3.5.4

The proof of Proposition 3.5.4 follows from the following lemma which states if an algorithm approximates the Gibbs measure, with probability approaching 1, its output law also tends to sample a vertex with a  $(z, K, \mathbf{X})$ -steep ancestor.

**Lemma 3.5.6.** *Let  $\beta > \beta_G$ . Suppose that  $\tilde{\mu}_N$  is the output law of an algorithm that approximates the Gibbs measure. Then, there exist  $z > 0$ ,  $K \in \mathbb{N}$  such that there exists  $\varepsilon_N \rightarrow 0$  such that, with probability approaching 1,*

$$\lim_{N \rightarrow \infty} \mathbb{P} \left( \sum_{|u|=N} \tilde{\mu}_N(u) \mathbf{1}\{u \text{ has a } (z, K, \mathbf{X})\text{-steep ancestor}\} > 1 - \varepsilon_N \right) = 1,$$

We now proceed to the proof of Proposition 3.5.4.

*Proof of Proposition 3.5.4.* Fix  $\beta > \beta_G$ . Let  $(v(n))_{n \in \mathbb{N}}$  be an algorithm that approximates the Gibbs measure with probability approaching 1. Suppose that  $\tau$  is its running time and  $\tilde{\mu}_N$  is its output law, which is the law of  $v(\tau)$  conditioned on the CREM. Recall that  $\tau'$  defined in Definition 3.5.3 is the first time where the algorithm finds a vertex with a  $(z, K, \mathbf{X})$ -steep ancestor. Therefore, the output  $v(\tau)$  has a  $(z, K, \mathbf{X})$ -steep ancestor implies that  $\tau \geq \tau'$ . Defining  $\mathcal{G}_N$  the event

$$\mathcal{G}_N := \left\{ \sum_{|u|=N} \tilde{\mu}_N(u) \mathbf{1}\{u \text{ has a } (z, K, \mathbf{X})\text{-steep ancestor}\} > 1 - \varepsilon_N \right\},$$

we have

$$\begin{aligned} \mathbb{P}(\tau \geq \tau') &\geq \mathbb{P}(v(\tau) \text{ has a } (z, K, \mathbf{X})\text{-steep ancestor}) \\ &= \mathbb{E} \left[ \sum_{|u|=N} \tilde{\mu}_N(u) \mathbf{1}\{u \text{ has a } (z, K, \mathbf{X})\text{-steep ancestor}\} \right] && \text{(Definition of } \tilde{\mu}_N) \\ &\geq \mathbb{E} \left[ \mathbf{1}_{\mathcal{G}_N} \sum_{|u|=N} \tilde{\mu}_N(u) \mathbf{1}\{u \text{ has a } (z, K, \mathbf{X})\text{-steep ancestor}\} \right] \\ &\geq \mathbb{P}(\mathcal{G}_N) (1 - \varepsilon_N) \rightarrow 1, \quad N \rightarrow \infty, && \text{(By Lemma 3.5.6)} \end{aligned}$$

and the proof is completed. ■

### 3.5.2 Proof of Lemma 3.5.6

This section is devoted to the proof of Proposition 3.5.6. The proof relies on Proposition 3.4.2 and the following lemma which states that if the KL divergence between two sequences of random probability measures are close to each other with probability approaching 1, and if the measures of certain events in the second sequence decay exponentially to 0 with probability approaching 1, then the measures of the corresponding events in the first sequence also converge to 0 with probability approaching 1.

**Lemma 3.5.7.** *Suppose  $(P_N)_{N \in \mathbb{N}}$  and  $(Q_N)_{N \in \mathbb{N}}$  be two sequences of random probability measures defined on a discrete space  $S$  such that the sequence  $(P_N)_{N \in \mathbb{N}}$  approximates the sequence  $(Q_N)_{N \in \mathbb{N}}$  with probability approaching 1. If  $(A_N)_{N \in \mathbb{N}}$  is a sequence of events on  $S$  such that with probability approaching 1,  $Q_N(A_N)$  converges to 0 exponentially fast as  $N \rightarrow \infty$ , i.e., there exists  $c > 0$  such that*

$$\lim_{N \rightarrow \infty} \mathbb{P}(Q_N(A_N) \leq e^{-cN}) = 1,$$

*then  $P_N(A_N)$  converges to 0 with probability approaching 1 as  $N \rightarrow \infty$ , i.e., there exists  $\varepsilon_N \rightarrow 0$  such that*

$$\lim_{N \rightarrow \infty} \mathbb{P}(P_N(A_N) \leq \varepsilon_N) = 1.$$

Lemma 3.5.7 follows from the so-called Birgé's inequality which, roughly speaking, says that if for two probability measures  $P$  and  $Q$  defined on the same probability space such that  $P$  is dominated by  $Q$ , for any event  $A$ , the difference of between  $P(A)$  and  $Q(A)$  is gauged by the KL divergence from  $P$  to  $Q$ .

**Fact 3.5.8** (Birgé's inequality, Theorem 4.20 in [52]). *Let  $P$  and  $Q$  be two probability measures defined on probability space  $(S, \mathcal{S})$  such that  $P$  is dominated by  $Q$ , i.e., for all event  $A \in \mathcal{S}$ ,  $Q(A) = 0$  implies  $P(A) = 0$ . Then,*

$$\sup_{A \in \mathcal{A}} h(P(A), Q(A)) \leq \mathbf{d}(P||Q),$$

where  $h(p, q) = p \log(p/q) + (1 - p) \log((1 - p)/(1 - q))$  is the relative entropy between two Bernoulli distribution with parameters  $p$  and  $q$ , respectively.

*Remark 3.5.9.* Note that the positions of  $P$  and  $Q$  are swapped comparing with the statement of Theorem 4.20 in [52].

Next, we state a simple but handy fact of the function  $x \mapsto x \log x$  where the proof is omitted.

**Fact 3.5.10.** *The range of the function  $g$  defined by  $g(0) = 0$  and  $g(x) = x \log x$  on  $(0, 1]$  equals  $[-e^{-1}, 0]$ .*

We are now ready to prove Lemma 3.5.7.

*Proof of Lemma 3.5.7.* Let  $(P_N)_{N \in \mathbb{N}}$  and  $(Q_N)_{N \in \mathbb{N}}$  be two sequences of random probability measures defined on a discrete space  $S$ . Suppose that the sequence  $(P_N)_{N \in \mathbb{N}}$  approximates the sequence  $(Q_N)_{N \in \mathbb{N}}$  with probability approaching 1, and there exists a sequence of event  $(A_N)_{N \in \mathbb{N}}$  such that there exists  $c > 0$  such that for all  $N \in \mathbb{N}$ ,

$$\mathbb{P}(Q_N(A_N) \leq e^{-cN}) = 1 - o_N(1). \quad (3.5.1)$$

For all  $N \in \mathbb{N}$ , on the event  $\{Q_N(A_N) \leq e^{-cN}\}$ , Fact 3.5.10 yields

$$\begin{aligned} h(P_N(A_N), Q_N(A_N)) &= P_N(A_N)(\log P_N(A_N) - \log Q_N(A_N)) \\ &\quad + (1 - P_N(A_N))(\log(1 - P_N(A_N)) - \underbrace{\log(1 - Q_N(A_N))}_{\leq 0}) \\ &\geq P_N(A_N)cN - 2e^{-1}. \end{aligned} \quad (3.5.2)$$

Therefore, combining (3.5.2) and Birgé's inequality, we have

$$P_N(A_N) \leq \frac{1}{cN} h(P_N(A_N), Q_N(A_N)) + \frac{1}{cN} 2e^{-1} \leq \frac{1}{cN} \mathbf{d}(P_N \| Q_N) + \frac{1}{cN} 2e^{-1}. \quad (3.5.3)$$

On the other hand, since  $P_N$  approximates  $Q_N$  with probability approaching 1, there exists  $\varepsilon_N \rightarrow 0$  such that

$$\mathbb{P} \left( \frac{1}{N} \mathbf{d}(P_N \| Q_N) \leq \varepsilon_N \right) = 1 - o_N(1).$$

Thus, by (3.5.3), with probability approaching 1,

$$P_N(A_N) \leq \frac{\varepsilon_N}{c} + \frac{1}{cN} 2e^{-1} \rightarrow 0, \quad N \rightarrow \infty$$

as desired. ■

We now prove Proposition 3.5.6.

*Proof of Proposition 3.5.6.* Fix  $\beta > \beta_G$ . Let  $\tilde{\mu}_N$  be the output law of an algorithm that approximates the Gibbs measure. We apply Lemma 3.5.7 with  $P_N := \mu_{\beta, N}$ ,  $Q_N := \tilde{\mu}_N$  and  $A_N$  defined in (3.4.1) where its complement equals

$$A_N^c := \{u \in \partial \mathbb{T}_N : u \text{ has a } (z, K, \mathbf{X})\text{-steep ancestor}\}.$$

Since Proposition 3.4.2 implies that there exists  $\delta > 0$  such that

$$\lim_{N \rightarrow \infty} \mathbb{P} \left( P_N(A_N^c) \geq 1 - e^{-\delta N} \right) = 1,$$

we then conclude from Lemma 3.5.7 that there exists  $\varepsilon_N \rightarrow 0$  such that

$$\begin{aligned} & \lim_{N \rightarrow \infty} \mathbb{P} \left( \sum_{|u|=N} \tilde{\mu}_N(u) \mathbf{1}\{u \text{ has a } (z, K, \mathbf{X})\text{-steep ancestor}\} > 1 - \varepsilon_N \right) \\ &= \lim_{N \rightarrow \infty} \mathbb{P} (Q_N(A_N^c) \geq 1 - \varepsilon_N) = 1, \end{aligned}$$

and proof is completed. ■

### 3.5.3 Proof of Proposition 3.5.5

This section is devoted to the proof of Proposition 3.5.5. As the proof is modified from the proof for the second part of Theorem 1.1 in [2], we start by recalling some relevant notation and lemmas from that article.

**Notation.** For  $v \in \mathbb{T}_N$ , recall the definition of  $\mathcal{C}_v$  in Definition 3.5.1. We then define the filtration

$$\mathcal{G}_k = \sigma \left( v(1), \dots, v(k); (X_w)_{w \in \mathcal{C}_{v(1)}}, \dots, (X_w)_{w \in \mathcal{C}_{v(k)}}; U_1, \dots, U_{k+1} \right).$$

Note that  $\tilde{\mathcal{F}}_k \subset \mathcal{G}_k$  for all  $k \geq 0$  — heuristically,  $\mathcal{G}_k$  adds to  $\tilde{\mathcal{F}}_k$  the information about the values in the branching random walk of all vertices contained in  $\mathcal{C}_{v(i)}$ ,

$i = 1, \dots, k$ . Note that trivially, the stochastic process  $v(n)_{n \geq 0}$  is still measurable with respect to this larger filtration  $\mathcal{G}$ . For  $n \geq 1$ , let  $\mathcal{R}_n = \bigcup_{i=1}^n \mathcal{C}_{v(i)}$  be the union of  $\mathcal{C}_{v(i)}$ ,  $i = 1, \dots, k$ . Also, let  $\hat{v}(n)$  be the most recent ancestor of  $v(n)$  in  $\mathcal{R}_{n-1}$  if  $n > 1$ , and let  $\hat{v}(n)$  be the root of  $\mathbb{T}_n$  if  $n = 1$ . Finally,  $\mathbf{X}' := (X'_v)_{v \in \mathbb{T}_n}$  is a i.i.d. copy of  $\mathbf{X} = (X_v)_{v \in \mathbb{T}_n}$  and is independent of  $\mathcal{G}_{n-1}$ .

Now, we recall the statements of two lemmas in [2] which will be useful in the proof of Proposition 3.5.5. The first lemma is a direct implication of the branching property.

**Lemma 3.5.11** (Lemma 3.2 in [2]). *Fix any randomized search algorithm  $v = (v(n))_{n \geq 1}$ . Then conditioned on  $\mathcal{G}_{n-1}$ , the family of random variables  $(X_v - X_{\hat{v}(n)})_{v \in \mathcal{R}_n \setminus \mathcal{R}_{n-1}}$  has the same law as  $(X'_v - X'_{\hat{v}(n)})_{v \in \mathcal{R}_n \setminus \mathcal{R}_{n-1}}$ .*

The next lemma states, roughly speaking, that  $(z, K, \mathbf{X})$ -steep vertices are rare.

**Lemma 3.5.12** (Lemma 3.3 in [2]). *For all  $K \in \mathbb{N}$  and  $z > 0$ , for any  $\gamma \in (0, (z \log 2)/K)$ , for all  $N$  sufficiently large, for any  $w \in \mathbb{T}_N$ ,*

$$\mathbb{P}(\exists v \in \mathcal{C}_w : v \text{ is } (z, K, \mathbf{X})\text{-steep}) \leq e^{-\gamma N}$$

We now proceed to the proof of Proposition 3.5.5.

*Proof of Proposition 3.5.5.* The goal is to show that  $\tau'$  stochastically dominates a geometric random variable with an exponentially small parameter, which follows from an argument slightly adapted from the proof for the second part of Theorem 1.1 in [2]. The argument goes as follows.

By Lemma 3.5.11,

$$\begin{aligned} & \mathbb{P}(\exists v \in \mathcal{R}_n \setminus \mathcal{R}_{n-1} : v \text{ is } (z, K, \mathbf{X})\text{-steep} \mid \mathcal{G}_{n-1}) \\ &= \mathbb{P}(\exists v \in \mathcal{R}_n \setminus \mathcal{R}_{n-1} : v \text{ is } (z, K, \mathbf{X}')\text{-steep} \mid \mathcal{G}_{n-1}) \\ &\leq \mathbb{P}(\exists v \in \mathcal{C}_{v(n)} : v \text{ is } (z, K, \mathbf{X}')\text{-steep} \mid \mathcal{G}_{n-1}) \\ &\leq \sup_{w \in \mathbb{T}_N} \mathbb{P}(\exists v \in \mathcal{C}_w : v \text{ is } (z, K, \mathbf{X}')\text{-steep}). \end{aligned} \tag{3.5.4}$$

The first inequality uses the fact that  $\mathcal{R}_n \setminus \mathcal{R}_{n-1} \subset \mathcal{C}_{v(n)}$  and the second inequality uses the independence of  $\mathbf{X}'$  and  $\mathcal{G}_{n-1}$ . Since  $\mathbf{X}'$  and  $\mathbf{X}$  have the same law, by Lemma 3.5.12, with  $\gamma = \gamma(K, z)$  as in that lemma, (3.5.4) yields that

$$\mathbb{P}(\exists v \in \mathcal{R}_n \setminus \mathcal{R}_{n-1} : v \text{ is } (z, K, \mathbf{X})\text{-steep} \mid \mathcal{G}_{n-1}) \leq e^{-\gamma N}.$$

We thus obtain

$$\begin{aligned} \mathbb{P}(\tau' = n) &= \mathbb{E}[\mathbf{1}_{\tau > n-1} \cdot \mathbb{P}(\tau = n \mid \mathcal{G}_{n-1})] \\ &= \mathbb{E}[\mathbf{1}_{\tau > n-1} \cdot \mathbb{P}(\exists v \in \mathcal{R}_n \setminus \mathcal{R}_{n-1} : v \text{ is } (z, K, \mathbf{X})\text{-steep} \mid \mathcal{G}_{n-1})] \\ &\leq \mathbb{P}(\tau > n-1) \cdot e^{-\gamma N}, \end{aligned}$$

from which we conclude that  $\tau'$  stochastically dominates a geometric random variable with success probability  $e^{-\gamma N}$ . In particular, this implies that for any positive

constant  $\gamma' \in (0, \gamma)$ ,

$$\mathbb{P}\left(\tau' \geq e^{\gamma' N}\right) \geq \sum_{n=\lceil e^{\gamma' N} \rceil}^{\infty} e^{-\gamma N} (1 - e^{-\gamma N})^n \sim \exp\left(-e^{-(\gamma-\gamma')N}\right) \rightarrow 1,$$

as  $N \rightarrow \infty$ . ■

### 3.A A lower bound of the free energy $F_\beta$

Recall that the free energy of the CREM is defined in (3.1.4) as

$$F_\beta := \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\log Z_{\beta, N}].$$

Recall also that we defined in (3.1.6) the quantity

$$\tilde{F}_\beta = \int_0^1 f(\beta \sqrt{a(s)}) ds,$$

where

$$f(\beta) = \begin{cases} \log 2 + \frac{\beta^2}{2}, & \beta < \sqrt{2 \log 2} \\ \sqrt{2 \log 2} \beta, & \beta \geq \sqrt{2 \log 2}. \end{cases}$$

The main goal of this section is to prove the following proposition, which asserts that  $F_\beta \geq \tilde{F}_\beta$ , and that equality holds if and only if  $\beta \leq \beta_G$ .

**Proposition 3.A.1.** *Suppose that  $A$  is non-concave. For all  $\beta \in [0, \infty)$ , define*

$$G_\beta := F_\beta - \tilde{F}_\beta.$$

Then,

- (i) For all  $\beta \in [0, \beta_G]$ ,  $G_\beta = 0$ .
- (ii) For all  $\beta > \beta_G$ ,  $G'_\beta > 0$ . In particular, this implies that  $G_\beta > 0$  for all  $\beta > \beta_G$ .

Before starting the proof, we recall that the free energy of the CREM has the following formula which can be found in Bovier and Kurkova [54] based on previous results by Capocaccia et al. [59].

**Fact 3.A.2.** *Given  $\beta > 0$ , let  $t_0(\beta) = \sup\{t \in [0, 1] : \hat{a}(t) > 2 \log 2 / \beta^2\}$ . Then, the free energy of the CREM is given as follows*

$$F_\beta = \beta \sqrt{2 \log 2} \int_0^{t_0(\beta)} \sqrt{\hat{a}(t)} dt + \frac{\beta^2}{2} (1 - \hat{A}(t_0(\beta))) + \log 2 (1 - t_0(\beta)) \quad (3.A.1)$$

$$= \int_0^1 f(\beta \sqrt{\hat{a}(s)}) ds, \quad (3.A.2)$$

where  $f(\beta)$  is defined as (3.1.5).

*Proof.* The proof of (3.A.1) can be found in Theorem 3.3 of Bovier and Kurkova [54]. While the authors of that paper assumed that the function  $A$  has to be continuously differentiable, their result can be extended to the case where  $A$  is merely Riemann integrable. This extension is possible because their argument is based on the following two ingredients. The first ingredient is the free energy formula for the GREM, given by Capocaccia et al. [59]. As a remark, the definition of the GREM is identical to the definition of the CREM except that the function  $A$  is a step function. The second ingredient is a Gaussian comparison argument that only requires the Riemann integrability of  $a$ .

Now, by (3.A.1) and the fact that  $\hat{a}$  is non-increasing,

$$\begin{aligned} & \beta\sqrt{2\log 2} \int_0^{t_0(\beta)} \sqrt{\hat{a}(t)} dt + \frac{\beta^2}{2}(1 - \hat{A}(t_0(\beta))) + \log 2(1 - t_0(\beta)) \\ &= \beta\sqrt{2\log 2} \int_0^{t_0(\beta)} \sqrt{\hat{a}(t)} dt + \int_{t_0(\beta)}^1 \left( \frac{\beta^2}{2}\hat{a}(t) + \log 2 \right) dt \\ &= \int_0^{t_0(\beta)} f(\beta\sqrt{\hat{a}(t)}) dt + \int_{t_0(\beta)}^1 f(\beta\sqrt{\hat{a}(t)}) dt = \int_0^1 f(\beta\sqrt{\hat{a}(t)}) dt, \end{aligned}$$

which proves (3.A.2). ■

To prove Proposition 3.A.1, we require the following three lemmas. The first lemma provides some useful properties of the function  $A$  and its concave hull  $\hat{A}$ .

**Lemma 3.A.3.** *The following are true.*

- (i) *On the set  $\{A = \hat{A}\}$ ,  $a = \hat{a}$  almost everywhere.*
- (ii) *Suppose that  $A$  is non-concave. Let  $\mathcal{I}$  be a connected component of  $\{t \in [0, 1] : A(t) < \hat{A}(t)\}$ . Then,  $\hat{a}$  is equal to a positive constant on the interior of  $\mathcal{I}$ , denoted by  $\hat{a}_{\mathcal{I}}$ . Moreover,*

$$\int_{\mathcal{I}} a(s) ds = \int_{\mathcal{I}} \hat{a}(s) ds = \hat{a}_{\mathcal{I}}|\mathcal{I}|,$$

where  $|\mathcal{I}|$  denotes the Lebesgue measure of  $\mathcal{I}$ .

- (iii) *With the same assumptions as in (ii), we have*

$$\int_{\mathcal{I}} \sqrt{a(s)} ds < \int_{\mathcal{I}} \sqrt{\hat{a}(s)} ds = \sqrt{\hat{a}_{\mathcal{I}}}|\mathcal{I}|.$$

*Proof.* We prove this lemma by addressing each point separately.

Proof of (i). The set  $\{A = \hat{A}\}$  is Lebesgue measurable because  $\{A = \hat{A}\} = (\hat{A} - A)^{-1}(\{0\})$  and the function  $\hat{A} - A$  is continuous. If  $\{A = \hat{A}\}$  is of measure zero, the statement trivially holds

Suppose now that  $\{A = \hat{A}\}$  has positive measure. Note that  $\{A = \hat{A}\}$  contains all the global maximum points of  $\hat{A} - A$  as  $\hat{A} \geq A$ . Thus, by Fermat's theorem of stationary points, for all  $t \in \{A = \hat{A}\} \cap \{\hat{A} - A \text{ is differentiable}\}$ , we have  $\hat{a}(t) = a(t)$ . It remains to show that  $t \in \{A = \hat{A}\} \cap \{\hat{A} - A \text{ is not differentiable}\}$

is of measure zero. By the fundamental theorem of calculus and the fact that  $\hat{a} - a$  is continuous almost everywhere on  $[0, 1]$ , the function  $\hat{A} - A$  is differentiable almost everywhere on  $[0, 1]$ . Therefore,  $\{A = \hat{A}\} \cap \{\hat{A} - A \text{ is not differentiable}\}$  is of measure zero.

Proof of (ii). Let  $\mathcal{I}$  be a connected component of  $\{A \neq \hat{A}\}$  with endpoints  $t_1$  and  $t_2$ . By the continuity of  $A$  and  $\hat{A}$ ,  $A(t_1) = \hat{A}(t_1)$  and  $A(t_2) = \hat{A}(t_2)$ . By the minimality of  $\hat{A}$ , for all  $t \in \text{int}(\mathcal{I})$ ,  $\hat{A}(t)$  is equal to the linear interpolation between  $A(t_1)$  and  $A(t_2)$ . In particular, this implies that the  $\hat{a}$  is constant on  $\text{int}(\mathcal{I})$ . Moreover,  $\hat{a}$  has to be positive. Otherwise, by the fundamental theorem of calculus,  $A(t) = \hat{A}(t)$  for any  $t \in \text{int}(\mathcal{I})$  which contradicts the assumption that  $\mathcal{I}$  is a connected component of  $\{A \neq \hat{A}\}$ .

To prove the second statement of (ii), note that

$$\int_{\mathcal{I}} a(s) \, ds = A(t_2) - A(t_1) = \hat{A}(t_2) - \hat{A}(t_1) = \int_{\mathcal{I}} \hat{a}(s) \, ds = \hat{a}_{\mathcal{I}} |\mathcal{I}|.$$

Proof of (iii). By (ii),  $\hat{a}_{\mathcal{I}}$  is positive. Then, by the Cauchy–Schwarz inequality,

$$\int_{\mathcal{I}} \sqrt{a(s)} \, ds = \int_{\mathcal{I}} \frac{\sqrt{a(s)}}{\sqrt{\hat{a}_{\mathcal{I}}}} \sqrt{\hat{a}_{\mathcal{I}}} \, ds < \sqrt{\int_{\mathcal{I}} \frac{a(s)}{\hat{a}_{\mathcal{I}}} \, ds} \sqrt{\int_{\mathcal{I}} \hat{a}_{\mathcal{I}} \, ds} = \sqrt{\int_{\mathcal{I}} \frac{a(s)}{\hat{a}_{\mathcal{I}}} \, ds} \sqrt{\hat{a}_{\mathcal{I}} |\mathcal{I}|}. \quad (3.A.3)$$

Note that the inequality above is strict as the equality holds if and only if there exists  $c \in \mathbb{R}$  such that  $a = c\hat{a}_{\mathcal{I}}$ . If that was the case, then by (ii),  $c = 1$ , and therefore  $A = \hat{A}$  on  $\mathcal{I}$  which is a contradiction.

Now, by (ii) and (3.A.3),

$$\int_{\mathcal{I}} \sqrt{a(s)} \, ds < \underbrace{\sqrt{\frac{\int_{\mathcal{I}} a(s) \, ds}{\hat{a}_{\mathcal{I}} |\mathcal{I}|}}}_{=1} \sqrt{\hat{a}_{\mathcal{I}} |\mathcal{I}|} = \int_{\mathcal{I}} \sqrt{\hat{a}(s)} \, ds,$$

and the proof is completed. ■

The second lemma collects two useful implications from the definition of  $\beta_G$ . The first one characterizes the  $\beta$  such that  $\beta\sqrt{a(t)} \leq \sqrt{2\log 2}$  for almost every  $t \in \{A \neq \hat{A}\}$ , and the second one shows that when  $\beta \leq \beta_G$ ,  $\beta\sqrt{\hat{a}(t)} \leq \sqrt{2\log 2}$  for all  $t \in \{A \neq \hat{A}\}$ .

**Lemma 3.A.4.** *Suppose that  $A$  is non-concave. Then the following statements hold.*

(i)  $\beta \leq \beta_G$  if and only if the set

$$\{A \neq \hat{A}\} \cap \{s \in [0, 1] : \beta\sqrt{a(s)} > \sqrt{2\log 2}\}$$

is of measure zero.

(ii) If  $\beta \leq \beta_G$ , then for every connected component  $\mathcal{I}$  of  $\{A \neq \hat{A}\}$ , we have  $\beta\sqrt{\hat{a}_{\mathcal{I}}} \leq \sqrt{2\log 2}$ .

*Proof.* We start with the proof of (i). By the definition of  $\beta_G$ ,  $\beta \leq \beta_G$  is true if and only if almost every  $s \in \{A \neq \hat{A}\}$ ,

$$\beta\sqrt{a(s)} \leq \sqrt{2\log 2}.$$

This immediately implies that the set  $\beta \leq \beta_G$  if and only if  $\{A \neq \hat{A}\} \cap \{s \in [0, 1] : \beta\sqrt{a(s)} > \sqrt{2\log 2}\}$  is of measure zero.

We proceed to the proof of (ii), and our strategy is to prove it by contradiction. Suppose that there exists a connected component  $\mathcal{I}$  of  $\{A \neq \hat{A}\}$  such that

$$\beta\sqrt{\hat{a}_{\mathcal{I}}} > \sqrt{2\log 2}. \quad (3.A.4)$$

Then,

$$\begin{aligned} \int_{\mathcal{I}} \hat{a}_{\mathcal{I}} \, ds &> \int_{\mathcal{I}} \frac{2\log 2}{\beta^2} \, ds && \text{(By (3.A.4) and the fact that } |\mathcal{I}| > 0) \\ &\geq \int_{\mathcal{I}} a(s) \, ds && \text{(By (i) and the assumption that } \beta \leq \beta_G) \\ &= \int_{\mathcal{I}} \hat{a}_{\mathcal{I}} \, ds, && \text{(By (ii) of Lemma 3.A.3)} \end{aligned}$$

which yields a contradiction. ■

The third lemma compares the difference between two integrals, one using  $a$  and the other using  $\hat{a}$ .

**Lemma 3.A.5.** *Recall that the derivative of  $f$  equals*

$$f'(x) = \begin{cases} x, & x < \sqrt{2\log 2} \\ \sqrt{2\log 2}, & x \geq \sqrt{2\log 2}. \end{cases} \quad (3.A.5)$$

(i) *Suppose that  $\mathcal{I}$  is a connected component of  $\{A \neq \hat{A}\}$ . Then for all  $\beta \geq 0$ ,*

$$\int_{\mathcal{I}} \left( f'(\beta\sqrt{\hat{a}(s)})\sqrt{\hat{a}(s)} - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds \geq 0.$$

(ii) *Moreover, if  $\beta > \beta_G$ , there exists a connected component  $\mathcal{I}$  of  $\{A \neq \hat{A}\}$  such that*

$$\int_{\mathcal{I}} \left( f'(\beta\sqrt{\hat{a}(s)})\sqrt{\hat{a}(s)} - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds > 0.$$

*Proof.* We prove this lemma by addressing each point separately.

Proof of (i). Let  $\mathcal{I}$  be a connected component of  $\{A \neq \hat{A}\}$ . By (ii) of Lemma 3.A.3,  $\hat{a}$  is equal to a positive constant  $\hat{a}_{\mathcal{I}}$  on  $\text{int}(\mathcal{I})$ . We now distinguish the two cases of  $\hat{a}_{\mathcal{I}}$ .

**Case 1:**  $\beta\sqrt{\hat{a}_{\mathcal{I}}} < \sqrt{2\log 2}$ . We have

$$\begin{aligned}
& \int_{\mathcal{I}} \left( f'(\beta\sqrt{\hat{a}(s)})\sqrt{\hat{a}(s)} - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds \\
&= \int_{\mathcal{I}} \left( \beta\hat{a}_{\mathcal{I}} - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds && \text{(By (3.A.5))} \\
&= \int_{\mathcal{I}} \left( \beta a(s) - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds && \text{(By (ii) of Lemma 3.A.3)} \\
&\geq \int_{\mathcal{I}} \underbrace{(\beta a(s) - \beta a(s))}_{=0} ds && \text{(Because } f'(x) \leq x \text{ for all } x \geq 0) \\
&= 0.
\end{aligned}$$

**Case 2:**  $\beta\sqrt{\hat{a}_{\mathcal{I}}} \geq \sqrt{2\log 2}$ . We have

$$\begin{aligned}
& \int_{\mathcal{I}} \left( \sqrt{2\log 2}\sqrt{\hat{a}(s)} - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds \\
&> \int_{\mathcal{I}} \left( \sqrt{2\log 2}\sqrt{a(s)} - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds && \text{(By (iii) of Lemma 3.A.3)} \\
&\geq \int_{\mathcal{I}} \underbrace{\left( \sqrt{2\log 2}\sqrt{a(s)} - \sqrt{2\log 2}\sqrt{a(s)} \right)}_{=0} ds && \text{(Because } f'(x) \leq \sqrt{2\log 2} \text{ for all } x \geq 0) \\
&= 0.
\end{aligned}$$

Proof of (ii). Suppose that  $\beta > \beta_G$ . We distinguish again the two cases of  $\hat{a}_{\mathcal{I}}$ .

**Case 1:**  $\beta\sqrt{\hat{a}_{\mathcal{I}}} < \sqrt{2\log 2}$ . By Lemma 3.A.4, there exists a connected component  $\mathcal{I}$  of  $\{A \neq \hat{A}\}$  such that

$$|\mathcal{I} \cap \{\beta\sqrt{a(s)} > \sqrt{2\log 2}\}| > 0. \quad (3.A.6)$$

Then we have

$$\begin{aligned}
& \int_{\mathcal{I}} \left( f'(\beta\sqrt{\hat{a}(s)})\sqrt{\hat{a}(s)} - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds \\
&= \int_{\mathcal{I}} \left( \beta\hat{a}_{\mathcal{I}} - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds && \text{(By (3.A.5))} \\
&= \int_{\mathcal{I}} \left( \beta a(s) - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds && \text{(By (ii) of Lemma 3.A.3)} \\
&= \int_{\mathcal{I} \cap \{\beta\sqrt{a(s)} \leq \sqrt{2\log 2}\}} \underbrace{(\beta a(s) - \beta a(s))}_{=0} ds \\
&+ \int_{\mathcal{I} \cap \{\beta\sqrt{a(s)} > \sqrt{2\log 2}\}} \underbrace{(\beta a(s) - \sqrt{2\log 2}\sqrt{a(s)})}_{>0} ds > 0. && \text{(By (3.A.6))}
\end{aligned}$$

**Case 2:**  $\beta\sqrt{\hat{a}_{\mathcal{I}}} \geq \sqrt{2\log 2}$ . In this case, as shown in Case 2 in the proof of (i), for any connected component  $\mathcal{I}$  of  $\{A \neq \hat{A}\}$ ,

$$\int_{\mathcal{I}} \left( f'(\beta\sqrt{\hat{a}(s)})\sqrt{\hat{a}(s)} - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds > 0.$$

This completes the proof. ■

We are now ready to prove Proposition 3.A.1.

*Proof of Proposition 3.A.1.* In the following, let  $\{\mathcal{I}_i\}_{i=1}^\infty$  be the collection of the connected components of  $\{A \neq \hat{A}\}$ .

We start with the proof of (i). Assume that  $\beta \leq \beta_G$ . We have

$$\begin{aligned} G_\beta &= \int_0^1 \left( f(\beta\sqrt{\hat{a}(s)}) - f(\beta\sqrt{a(s)}) \right) ds \\ &= \sum_{i=1}^\infty \int_{\mathcal{I}_i} \left( f(\beta\sqrt{\hat{a}_{\mathcal{I}_i}}) - f(\beta\sqrt{a(s)}) \right) ds \quad (\text{By Lemma 3.A.3}) \\ &= \sum_{i=1}^\infty \int_{\mathcal{I}_i} \left( \frac{\beta^2}{2} \hat{a}_{\mathcal{I}_i} - \frac{\beta^2}{2} a(s) \right) ds \quad (\text{By (i) and (ii) of Lemma 3.A.4}) \\ &= 0, \quad (\text{By (ii) of Lemma 3.A.3}) \end{aligned}$$

and the proof of (i) is completed.

We now proceed to the proof of (ii). Assume that  $\beta > \beta_G$ . Differentiating  $G_\beta$  yields

$$G'_\beta = \int_0^1 \left( f'(\beta\sqrt{\hat{a}(s)})\sqrt{\hat{a}(s)} - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds. \quad (3.A.7)$$

Again by Lemma 3.A.3,  $G'_\beta$  satisfies the following

$$\begin{aligned} &\int_0^1 \left( f'(\beta\sqrt{\hat{a}(s)})\sqrt{\hat{a}(s)} - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds \\ &= \sum_{i=1}^\infty \int_{\mathcal{I}_i} \left( f'(\beta\sqrt{\hat{a}_{\mathcal{I}_i}})\sqrt{\hat{a}_{\mathcal{I}_i}} - f'(\beta\sqrt{a(s)})\sqrt{a(s)} \right) ds > 0, \quad (3.A.8) \end{aligned}$$

where (3.A.8) is true because of Lemma 3.A.5 and the assumption that  $\beta > \beta_G$ . This proves (ii).  $\blacksquare$

### 3.B Proof of Lemma 3.3.2

This section is devoted to the proof of Lemma 3.3.2, and the strategy is to compare the free energy of the CREM with the free energy of the branching random walk, which is defined as follows.

**The branching random walk.** Let  $(\tilde{X}_u)_{u \in \mathbb{T}_M}$  be a centered Gaussian process indexed by  $\mathbb{T}_M$  with the covariance function

$$\mathbb{E}[\tilde{X}_u \tilde{X}_w] = |u \wedge w|$$

for all  $u, w \in \mathbb{T}_M$ . This Gaussian process is called the branching random walk with standard Gaussian increments, which will be abbreviated as the branching random walk. Define

$$f_M(\beta) = \frac{1}{M} \mathbb{E}[\log \tilde{Z}_{\beta, M}], \quad \text{where} \quad \tilde{Z}_{\beta, M} = \sum_{|u|=M} e^{\beta \tilde{X}_u}.$$

It is known that (see, [65]) the function (3.1.5) is the pointwise limit of  $f_M(\beta)$ , i.e., for all  $\beta \in [0, \infty)$ ,

$$f = \lim_{M \rightarrow \infty} f_M(\beta).$$

The proof of Lemma 3.3.2 relies on a quantitative estimate of the convergence above, which is stated in detail in Lemma 3.B.3. Before we proceed to Lemma 3.B.3, we state the following lemma that is handy to prove Lemma 3.B.3.

**Lemma 3.B.1.** *Define  $g_M : [0, \infty) \rightarrow \mathbb{R}$  as  $g_M(0) := 0$  and  $g_M(\beta) := f_M(\beta)/\beta - 2 \log 2/\beta$ . Define  $g(\beta) := f(\beta)/\beta - 2 \log 2/\beta$  which equals*

$$g(\beta) := \begin{cases} \frac{\beta}{2}, & \beta \in [0, \sqrt{2 \log 2}] \\ \sqrt{2 \log 2} - \frac{\log 2}{\beta}, & \beta > \sqrt{2 \log 2} \end{cases}$$

*Then, the following statements are true.*

- (i) *For all  $\beta \in [0, \infty)$ ,  $\lim_{M \rightarrow \infty} g_M(\beta) = g(\beta)$ .*
- (ii) *For all  $\beta \in [0, \infty)$  and  $M \in \mathbb{N}$ ,  $g_M(\beta) \leq g(\beta)$ .*
- (iii) *For all  $M \in \mathbb{N}$ , the function  $g_M$  is non-decreasing. Moreover,  $g_M(\infty) := \lim_{\beta \rightarrow \infty} g_M(\beta)$  exists and  $\lim_{M \rightarrow \infty} g_M(\infty) = \sqrt{2 \log 2} = g(\infty)$ , where  $g(\infty) := \lim_{\beta \rightarrow \infty} g(\beta)$ .*
- (iv) *The sequence of functions  $g_M$  converges uniformly to  $g$ .*

*Remark 3.B.2.* As we will see below, the proof of (ii) in Lemma 3.B.1 is a standard argument in the context of statistical physics. The argument to prove (iv) is a slight modification of the proof of Dini's second theorem<sup>1</sup> which states that if a sequence of monotone (continuous or discontinuous) functions converges on a closed interval to a continuous function, the sequence converges uniformly. The second statement of (iii) allows us to generalize Dini's second theorem to our setting.

*Proof of Lemma 3.B.1.* We prove this lemma by addressing each point separately. Proof of (i). This follows directly from the definition of  $g_M$  and  $g$ , and the pointwise convergence of  $f_M$  to  $f$ .

Proof of (ii). It suffices to show that for all  $\beta \in [0, \infty)$  and  $M \in \mathbb{N}$ ,  $f_M(\beta) \leq f(\beta)$ . To this purpose, we claim that for all  $\beta \in [0, \infty)$  the sequence  $M f_M$  is super-additive. If this is true, then Fekete's lemma implies that  $f_M(\beta) \leq f(\beta)$ .

<sup>1</sup>This simple but handy result appears in some French textbooks under the name "deuxième théorème de Dini". One can find the proof in Solution 127 in Part II, Chapter 3 of [160].

Now, fixing  $M_1, M_2 \in \mathbb{N}$ , we have

$$\begin{aligned}
 \mathbb{E} [\log \tilde{Z}_{\beta, M_1+M_2}] &= \mathbb{E} \left[ \log \sum_{|u|=M_1+M_2} e^{\beta \tilde{X}_u} \right] \\
 &= \mathbb{E} \left[ \log \sum_{|u_1|=M_1} \sum_{|u_2|=M_2} e^{\beta(\tilde{X}_{u_1} + \tilde{X}_{u_2}^{u_1})} \right] \\
 &= \mathbb{E} \left[ \log \sum_{|u_1|=M_1} e^{\beta \tilde{X}_{u_1}} \tilde{Z}_{\beta, M_2}^{u_1} \right] \\
 &= \mathbb{E} [\log \tilde{Z}_{\beta, M_1}] + \mathbb{E} \left[ \log \sum_{|u_1|=M_1} \frac{e^{\beta X_{u_1}}}{\tilde{Z}_{\beta, M_1}} \tilde{Z}_{\beta, M_2}^{u_1} \right].
 \end{aligned}$$

By Jensen's inequality and the branching property, we have

$$\mathbb{E} \left[ \log \sum_{|u_1|=M_1} \frac{e^{\beta X_{u_1}}}{\tilde{Z}_{\beta, M_1}} \tilde{Z}_{\beta, M_2}^{u_1} \right] \geq \mathbb{E} \left[ \sum_{|u_1|=M_1} \frac{e^{\beta X_{u_1}}}{\tilde{Z}_{\beta, M_1}} \log \tilde{Z}_{\beta, M_2}^{u_1} \right] = \mathbb{E} [\log \tilde{Z}_{\beta, M_2}].$$

Therefore, we conclude that

$$\begin{aligned}
 (M_1 + M_2)g_{M_1+M_2} &= \mathbb{E} [\log \tilde{Z}_{\beta, M_1+M_2}] \\
 &\geq \mathbb{E} [\log \tilde{Z}_{\beta, M_1}] + \mathbb{E} [\log \tilde{Z}_{\beta, M_2}] = M_1 g_{M_1} + M_2 g_{M_2}.
 \end{aligned}$$

Proof of (iii). Fix  $M \in \mathbb{N}$ . For all  $\beta > 0$ , by Jensen's inequality and the fact that  $x \mapsto \log x$  is concave,

$$\log \sum_{|u|=M} \frac{1}{2^M} e^{\beta \tilde{X}_u} \geq \sum_{|u|=M} \frac{1}{2^M} \log e^{\beta \tilde{X}_u} = \sum_{|u|=M} \frac{1}{2^M} \beta \tilde{X}_u.$$

Thus, by the fact that  $(\tilde{X}_u)_{|u|=M}$  is centered,

$$g_M(\beta) \geq \frac{1}{\beta M} \mathbb{E} \left[ \log \sum_{|u|=M} \frac{1}{2^M} e^{\beta \tilde{X}_u} \right] \geq \mathbb{E} \left[ \sum_{|u|=M} \frac{1}{2^M} \beta \tilde{X}_u \right] = 0 = g(0).$$

For all  $0 < \beta < \beta'$ , the function  $x \mapsto x^{\beta'/\beta}$  is convex. Therefore, by Jensen's inequality,

$$\left( \sum_{|u|=M} \frac{1}{2^M} e^{\beta \tilde{X}_u} \right)^{\beta'/\beta} \leq \sum_{|u|=M} \frac{1}{2^M} e^{\beta' \tilde{X}_u}.$$

It then yields immediately that

$$g_M(\beta) = \frac{1}{\beta M} \mathbb{E} \left[ \log \left( \sum_{|u|=M} \frac{1}{2^M} e^{\beta \tilde{X}_u} \right) \right] \leq \frac{1}{\beta' M} \mathbb{E} \left[ \log \left( \sum_{|u|=M} \frac{1}{2^M} e^{\beta' \tilde{X}_u} \right) \right] = g_M(\beta'),$$

which proves that  $g_M$  is non-decreasing. Now, by (ii) and the fact that  $g \leq \sqrt{2 \log 2}$ , monotone convergence theorem implies that  $\lim_{\beta \rightarrow \infty} g_M(\beta)$  exists and is bounded from above by  $\sqrt{2 \log 2}$ . Finally, note that

$$\frac{1}{M} \mathbb{E} \left[ \max_{|u|=M} X_u \right] - \frac{\log 2}{\beta} \leq g_M(\beta) \leq \frac{1}{M} \mathbb{E} \left[ \max_{|u|=M} X_u \right].$$

Taking  $\beta \rightarrow \infty$ , we obtain the equality

$$g_M(\infty) = \frac{1}{M} \mathbb{E} \left[ \max_{|u|=M} X_u \right].$$

It is well-known that

$$\lim_{M \rightarrow \infty} \frac{1}{M} \mathbb{E} \left[ \max_{|u|=M} X_u \right] = \sqrt{2 \log 2},$$

which is an implication of Theorem 3.1 in [54] by letting the covariance function to be the identity function. Therefore, taking  $M \rightarrow \infty$ , we conclude that

$$\lim_{M \rightarrow \infty} g_M(\infty) = \sqrt{2 \log 2} = \lim_{\beta \rightarrow \infty} g(\beta) = g(\infty).$$

Proof of (iv). Fix  $\varepsilon > 0$ . By its definition, the function  $g$  is continuous and non-decreasing on  $[0, \infty)$ . Moreover, by (iii),  $g(\infty) := \lim_{\beta \rightarrow \infty} g(\beta)$  and  $g_M(\infty) := \lim_{\beta \rightarrow \infty} g_M(\beta)$  exist and  $\lim_{M \rightarrow \infty} g_M(\infty) = g(\infty)$ . By the intermediate value theorem, there exists a subdivision  $0 = \beta_0 < \beta_1 < \dots < \beta_{k-1} < \beta_k = \infty$  such that  $g(\beta_{i+1}) - g(\beta_i) < \varepsilon$ , for all  $i = 0, \dots, k-1$ . Thus, for all  $\beta \in [0, \infty)$  and  $i = 0, \dots, k-1$ , we have

$$\begin{aligned} g_M(\beta) - g(\beta) &\leq g_M(\beta_{i+1}) - g(\beta_i) && \text{(Because } g_M \text{ and } g \text{ are non-decreasing)} \\ &\leq g_M(\beta_{i+1}) - g(\beta_{i+1}) + \varepsilon && \text{(By the choice of subdivision)} \end{aligned} \quad (3.B.1)$$

and

$$\begin{aligned} g_M(\beta) - g(\beta) &\geq g_M(\beta_i) - g(\beta_{i+1}) && \text{(Because } g_M \text{ and } g \text{ are non-decreasing)} \\ &\geq g_M(\beta_i) - g(\beta_i) - \varepsilon. && \text{(By the choice of subdivision)} \end{aligned} \quad (3.B.2)$$

By (i) and (iii), there exists  $M_\varepsilon \in \mathbb{N}$  such that for all  $M \geq M_\varepsilon$  and  $i = 0, \dots, k-1$ ,

$$|g_M(\beta_i) - g(\beta_i)| < \varepsilon. \quad (3.B.3)$$

Combining (3.B.1), (3.B.2) and (3.B.3), we conclude that for all  $M \geq M_\varepsilon$ ,

$$|g_M(\beta) - g(\beta)| < 2\varepsilon,$$

which proves that  $g_M$  converges to  $g$  uniformly as, desired.  $\blacksquare$

Lemma 3.B.1 implies the following quantitative convergence of  $f_M$ .

**Lemma 3.B.3.** *For all  $\varepsilon > 0$ , there exists  $M \in \mathbb{N}$  independent of  $\beta \in [0, \infty)$  such that*

$$|f_M(\beta) - f(\beta)| \leq \beta \varepsilon. \quad (3.B.4)$$

*Proof.* By (iv) of Lemma 3.B.1,  $g_M$  converges uniformly to  $g$  on  $[0, \infty)$ . Combining this with the fact

$$|f_M(\beta) - f(\beta)| = \beta |g_M(\beta) - g(\beta)|.$$

the proof is completed. ■

We now proceed to the proof of Lemma 3.3.2.

*Proof of Lemma 3.3.2.* Fix  $M_N$  a sequence such that  $M_N \in \llbracket 1, N \rrbracket$  and  $M_N \rightarrow \infty$ .

For all  $N \in \mathbb{N}$ , by Kahane's inequality (see, Theorem 3.11 in [125]), for all  $k \in \llbracket 0, \lfloor N/M_N \rfloor \rrbracket$ ,

$$\mathbb{E} \left[ \log \tilde{Z}_{\beta\sqrt{a_k^-}, M_N} \right] \leq \mathbb{E} \left[ \log Z_{\beta, M_N}^{(kM_N)} \right] \leq \mathbb{E} \left[ \log \tilde{Z}_{\beta\sqrt{a_k^+}, M_N} \right], \quad (3.B.5)$$

where

$$a_k^- := \operatorname{ess\,inf}_{t \in [\frac{kM_N}{N}, \frac{(k+1)M_N}{N}]} a(t) \quad \text{and} \quad a_k^+ := \operatorname{ess\,sup}_{t \in [\frac{kM_N}{N}, \frac{(k+1)M_N}{N}]} a(t).$$

Now, fix  $\varepsilon > 0$ . By Lemma 3.B.3, there exists  $N_0 \in \mathbb{N}$  such that for all  $N \geq N_0$ ,

$$\frac{1}{M_N} \mathbb{E} \left[ \log \tilde{Z}_{\beta\sqrt{a_k^-}, M} \right] \geq f(\beta\sqrt{a_k^-}) - \varepsilon\beta\sqrt{a_k^-} \quad \text{and} \quad \frac{1}{M_N} \mathbb{E} \left[ \log \tilde{Z}_{\beta\sqrt{a_k^+}, M} \right] \leq f(\beta\sqrt{a_k^+}) + \varepsilon\beta\sqrt{a_k^+}. \quad (3.B.6)$$

Combining (3.B.5) and (3.B.6), for all  $N \geq N_0$ , we conclude that

$$\mathbb{E} \left[ \log Z_{\beta, M_N}^{(kM_N)} \right] \geq \mathbb{E} \left[ \log \tilde{Z}_{\beta\sqrt{a_k^-}, M_N} \right] \geq f(\beta\sqrt{a_k^-}) - \varepsilon\beta\sqrt{a_k^-}$$

and

$$\mathbb{E} \left[ \log Z_{\beta, M_N}^{(kM_N)} \right] \leq \mathbb{E} \left[ \log \tilde{Z}_{\beta\sqrt{a_k^+}, M_N} \right] \leq f(\beta\sqrt{a_k^+}) + \varepsilon\beta\sqrt{a_k^+}.$$

These complete the proof. ■

**Acknowledgments.** I want to thank Pascal Maillard for his guidance throughout the whole project. I am also grateful to Alexandre Legrand and Michel Pain for stimulating discussions.

# 4. Negative moments of the CREM partition function in the high temperature regime

The continuous random energy model (CREM) was introduced by Bovier and Kurkova in 2004 as a toy model of disordered systems. Among other things, their work indicates that there exists a critical point  $\beta_c$  such that the partition function exhibits a phase transition. The present work focuses on the high temperature regime where  $\beta < \beta_c$ . We show that for all  $\beta < \beta_c$  and for all  $s > 0$ , the negative  $s$  moment of the CREM partition function is comparable with the expectation of the CREM partition function to the power of  $-s$ , up to constants that are independent of  $N$ .

## 4.1 Introduction

In this paper, we consider the continuous random energy model (CREM) introduced by Bovier and Kurkova in [54] based on previous work by Derrida and Spohn [91]. Formally, the model is defined as follows. Let  $N \in \mathbb{N}$ . Denote by  $\mathbb{T}_N$  the binary tree with depth  $N$ . Given  $u \in \mathbb{T}_N$ , we denote by  $|u|$  the depth of  $u$ . For all  $u, v \in \mathbb{T}_N$ , let  $u \wedge v$  be the most recent common ancestor of  $u$  and  $v$ . The CREM is a centered Gaussian process  $(X_u)_{u \in \mathbb{T}_N}$  indexed by the binary tree  $\mathbb{T}_N$  of depth  $N$  with covariance function

$$\mathbb{E}[X_v X_w] = N \cdot A\left(\frac{|v \wedge w|}{N}\right), \quad \forall v, w \in \mathbb{T}_N.$$

For the purpose of this paper, we require that the function  $A$  satisfies the following assumption.

**Assumption 4.1.1.** We suppose that the function  $A$  is a non-decreasing function defined on the interval  $[0, 1]$  such that  $A(0) = 0$  and  $A(1) = 1$ . Let  $\hat{A}$  be the concave hull of  $A$ , and we denote by  $\hat{A}'$  the right derivative of  $\hat{A}$ . Throughout this paper, we assume the following regularity conditions.

- (i) The function  $A$  is differentiable in a neighborhood of 0, i.e., there is an  $x_0 \in (0, 1]$  such that  $A$  is differentiable on the interval  $(0, x_0)$ . Furthermore, we assume that  $A$  has a finite right derivative at 0.

- (ii) There exists  $\alpha \in (0, 1)$  such that the derivative of  $A$  is locally Hölder continuous with exponent  $\alpha$  in a neighborhood of 0, i.e., there exists  $x_1 \in (0, 1]$  such that

$$\sup_{\substack{x, y \in [0, x_1] \\ x \neq y}} \frac{|A'(x) - A'(y)|}{|x - y|^\alpha} < \infty.$$

- (iii) The right derivative of  $\hat{A}$  at  $x = 0$  is finite, i.e.,  $\hat{A}'(0) < \infty$ .

To study the CREM, one of the key quantity is the partition function. The partition function of the CREM is defined as

$$Z_{\beta, N} := \sum_{|u|=N} e^{\beta X_u}. \quad (4.1.1)$$

The free energy of the CREM is defined as

$$F_\beta := \lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} [\log Z_{\beta, N}].$$

The free energy  $F_\beta$  admits an explicit expression. Namely, for all  $\beta \geq 0$ ,

$$F_\beta = \int_0^1 f\left(\beta \sqrt{\hat{A}'(s)}\right) ds, \quad \text{where } f(x) := \begin{cases} \frac{x^2}{2}, & x < \sqrt{2 \log 2}, \\ \sqrt{2 \log 2}, & x \geq \sqrt{2 \log 2}. \end{cases} \quad (4.1.2)$$

Formula (4.1.2) was proven by Bovier and Kurkova in [54], based on a Gaussian comparison argument and previous work of Capocaccia, Cassandro and Picco [59] in the 1980s. While Bovier and Kurkova required the function  $A$  to be continuously differentiable, their result can be extended to the class of  $A$  that can be approximated pointwise by step functions from above and below.

In the same paper, Bovier and Kurkova also showed that the maximum of the CREM satisfies the following.

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \left[ \max_{|u|=N} X_u \right] = \sqrt{2 \log 2} \int_0^1 \sqrt{\hat{A}'(s)} ds. \quad (4.1.3)$$

Combining (4.1.2) and (4.1.3) indicates that there exists

$$\beta_c := \frac{\sqrt{2 \log 2}}{\sqrt{\hat{A}'(0)}} \quad (4.1.4)$$

such that the following phase transition occurs. a) For all  $\beta < \beta_c$ , the main contribution to the partition function comes from an exponential amount of the particles. b) For all  $\beta > \beta_c$ , the maximum starts to contribute significantly to the partition function. The quantity  $\beta_c$  is sometimes referred to as the static critical inverse temperature of the CREM. In the following, we refer to the subcritical regime  $\beta < \beta_c$  as the high temperature regime.

Our goal is to study the negative moments of the partition function  $Z_{\beta, N}$  in the high temperature regime, and we obtain the following result.

**Theorem 4.1.2.** *Suppose Assumption 4.1.1 is true. Let  $\beta < \beta_c$ . For all  $s > 0$ , there exist  $N_0 \in \mathbb{N}$  and a constant  $C = C(A, \beta, s)$ , independent of  $N$ , such that for all  $N \geq N_0$ ,*

$$\mathbb{E}[(Z_{\beta,N})^{-s}] \leq C\mathbb{E}[Z_{\beta,N}]^{-s}.$$

*Remark 4.1.3.* For all  $\beta > 0$ ,  $N \in \mathbb{N}$  and  $s > 0$ , we have the trivial lower bound provided by the Jensen inequality and the convexity of  $x \mapsto x^{-s}$ ,

$$\mathbb{E}[(Z_{\beta,N})^{-s}] \geq \mathbb{E}[Z_{\beta,N}]^{-s}. \quad (4.1.5)$$

Thus, combining (4.1.5) with Theorem 4.1.2, we see that for all  $\beta < \beta_c$  in the high temperature regime and for all  $s > 0$ ,  $\mathbb{E}[(Z_{\beta,N})^{-s}]$  is comparable with  $\mathbb{E}[Z_{\beta,N}]^{-s}$ .

### 4.1.1 Historical background

The continuous random energy model (CREM) can be viewed as a time-inhomogeneous branching random walk. For the (homogeneous) branching random walk, the typical approach is not to study the partition function directly but to consider the additive martingale  $W_{\beta,N} = Z_{\beta,N}/\mathbb{E}[Z_{\beta,N}]$ , which converges pointwise to  $W_{\beta,\infty}$ . A standard method to establish that  $W_{\beta,\infty}$  has negative moments involves the following observation. Suppose that  $Y'$ ,  $Y''$ ,  $W'$  and  $W''$  are independent random variables such that  $Y'$  and  $Y''$  share the law of the increment of the branching random walk, and each of  $W'$  and  $W''$  follows the same distribution as  $W_{\beta,\infty}$ . Then, the limit of the additive martingale  $W_{\beta,\infty}$  satisfies the following fixed point equation in distribution:

$$W_{\beta,\infty} \stackrel{(d)}{=} Y'W' + Y''W''. \quad (4.1.6)$$

Formula (4.1.6) is a special case of the so-called smoothing transform, which has been studied extensively in the literature. In the context of multiplicative cascades, which is an equivalent model of the branching random walk, Molchan showed in [144] that if  $\mathbb{E}[(Y')^{-s}] < \infty$  for some  $s > 0$ , then  $\mathbb{E}[(W_{\beta,\infty})^{-s}] < \infty$ . Subsequently, Molchan's result was extended by Liu in [127, 126]. More recently, Hu studied in [113] the small deviation of the maximum of the branching random walk based on Liu's result. On the other hand, Nikula provided small deviations for lognormal multiplicative cascades in [149], thereby refining Molchan's result.

In the context of Gaussian multiplicative chaos (GMC), Garban, Holden, Sepúlveda and Sun showed in [107] that given a subcritical GMC measure with mild conditions on the base measure, the total mass of this GMC measure has negative moments of all orders. Their works expanded on the findings of Robert and Vargas [163], Duplantier and Sheffield [95] and Remy [161], with the base measure taken as the Lebesgue measure restricted to some open set, where these authors were interested in quantifying the left tail behavior of the total mass of the GMC near 0. A final remark is that Robert and Vargas referred the proof of Theorem 3.6 in their paper [163], which concerned the negative moment of the total mass of the GMC, to Barral and Mandelbrot [26], where they studied a log-Poisson cascade.

### 4.1.2 Proof strategy

For the CREM in general,  $W_{\beta,N}$  is not a martingale, so there is no obvious way to show that  $W_{\beta,\infty}$  exists. To prove Theorem 4.1.2, we adapt the proof of Lemma A.3 in the paper [38] by Benjamini and Schramm. While their argument also involved the smoothing transform, it can be adjusted for general CREM. In particular, we show that for all  $s > 0$ , there exist two positive sequences  $\varepsilon_k$  and  $\eta_k$  that both decay double exponentially to 0 as  $k \rightarrow \infty$  such that for  $N$  sufficiently large,

$$\mathbb{P}(Z_{\beta,N} \leq \varepsilon_k \mathbb{E}[Z_{\beta,N}]) \leq \eta_k, \quad (4.1.7)$$

and that for all  $k \in \llbracket 1, C' \log N \rrbracket$ , where  $C' > 0$ , there exist  $C = C(A, \beta, s) > 0$  and  $c = c(A, \beta, s) > 0$  such that

$$\varepsilon_k^{-s} \eta_k \leq C e^{-ce^{ck}}. \quad (4.1.8)$$

The proof of (4.1.7) and (4.1.8) is by established an initial left tail estimate, and then use the branching property to bootstrap the estimate. Before ending this section, we want to mention that Ding showed in [92] that maximum the two-dimensional discrete Gaussian free field has double exponential left tail, and the proof for the lower bound also involved a bootstrap argument, where the bootstrap was done by initial estimate in small boxes plus the Fortuin–Kasteleyn–Ginibre (FKG) inequality.

**Outline.** The rest of the paper is organized as follows. We provide in Section 4.2 an initial estimate of the left tail of the partition function. Next, in Section 4.3, we improve the estimate obtained in the previous section using a bootstrap argument. Finally, Theorem 4.1.2 is proven in Section 4.4.

## 4.2 Initial estimate for the left tail of the partition function

The main goal of this section is to estimate the left tail of  $Z_{\beta,N}$ . Before stating the proposition, we first introduce the necessary notation.

**Notation.** Fix  $k \in \llbracket 0, N \rrbracket$ . Define  $(X_u^{(k)})_{|u|=N-k}$  to be the centered Gaussian process with covariance function

$$\mathbb{E}[X_u^{(k)} X_w^{(k)}] = N \left( A \left( \frac{|v \wedge w| + k}{N} \right) - A \left( \frac{k}{N} \right) \right).$$

We also introduce the corresponding partition function.

$$Z_{\beta,N-k}^{(k)} = \sum_{|u|=N-k} e^{\beta X_u^{(k)}}.$$

We are ready to state the main proposition of this section, which gives an initial bound, independent of  $N$  and  $k$ , for the left tail of the partition function  $Z_{\beta,N-k}^{(k)}$  in the high temperature regime.

**Proposition 4.2.1.** *Let  $\beta < \beta_c$ . Let  $K = \lfloor \frac{2}{\log(19/10)} \log N \rfloor$ . Then, there exist  $N_0 = N_0(A, \beta) \in \mathbb{N}$  and a constant  $\eta_0 = \eta_0(A, \beta) < 1$  such that for all  $N \geq N_0$  and  $k \in \llbracket 0, K \rrbracket$ ,*

$$\mathbb{P} \left( Z_{\beta, N-k}^{(k)} \leq \frac{1}{2} \cdot \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \right) \leq \eta_0.$$

The proof of Proposition 4.2.1 requires the following four lemmas.

#### 4.2.1 Useful properties of $A$ and $\hat{A}$

The first lemma gives a trivial bound of the difference between  $A(y)$  and  $A(z)$  by the concave hull.

**Lemma 4.2.2.** *For all  $y, z \in [0, 1]$  with  $y > z$ , we have*

$$A(y) - A(z) \leq \hat{A}'(0)y.$$

*Proof.* Since  $A(0) = 0$  and  $A$  is non-decreasing,  $A$  is non-negative on  $[0, 1]$ . On the other hand, by the definition of the concave hull,  $A(y) \leq \hat{A}'(0)y$  for all  $y \in [0, 1]$ . Combining these two facts together, we conclude that for all  $y \in [0, 1]$ ,

$$A(y) - A(z) \leq A(y) \leq \hat{A}'(0)y,$$

and this completes the proof. ■

The second lemma estimates the difference between  $A(y)$  and  $A(z)$  when  $y$  and  $z$  are close to 0.

**Lemma 4.2.3.** *There exists  $x_0 \in (0, 1]$  such that for all  $y, z \in [0, x_0]$ , there exists a constant  $C = C(A, x_0) > 0$  such that,*

$$|A(y) - A(z)| \leq \hat{A}'(0)|y - z| + Cy^\alpha|y - z| + C|y - z|^{1+\alpha}. \quad (4.2.1)$$

*Proof.* Fix  $y, z \in [0, x_1]$ , where  $x_1$  is given in (i) of Assumption 4.1.1. Then by applying the 1st order Taylor expansion of  $A$  at  $y$  with Lagrange remainder, we have

$$A(z) = A(y) + A'(y)(z - y) + (z - y)(A'(\xi) - A'(y)),$$

where  $\xi$  is between  $z$  and  $y$ . Then, the triangle inequality and the fact that  $A$  is non-decreasing yield

$$|A(z) - A(y)| \leq A'(y)|z - y| + |z - y||A'(\xi) - A'(y)|. \quad (4.2.2)$$

By the local Hölder continuity of  $A$  in a neighborhood around 0 provided by Assumption 4.1.1, there exists  $C = C(A, x_1) > 0$  such that

$$A'(y) \leq A'(0) + Cy^\alpha. \quad (4.2.3)$$

Similarly, by the local Hölder continuity of  $A$  in a neighborhood around 0 provided by Assumption 4.1.1, we have

$$|A'(\xi) - A'(y)| \leq C|z - y|^\alpha, \quad (4.2.4)$$

where  $C$  is the same constant as in (4.2.3). Combining (4.2.2), (4.2.3) and (4.2.4), we conclude that

$$|A(z) - A(y)| \leq \hat{A}'(0)|z - y| + Cy^\alpha|z - y| + C|z - y|^{1+\alpha},$$

and this completes the proof.  $\blacksquare$

## 4.2.2 Moment estimates of the truncated partition function

Our strategy of proving Proposition 4.2.1 is to adopt a truncated moment estimate. To perform the estimate, we introduce the relevant notation regarding the underlying binary tree of the CREM.

**Definition 4.2.4.** For any  $v \in \mathbb{T}_N$  with prefix  $w$ , let  $v \setminus w$  be the suffix of  $v$  such that  $v = w(v \setminus w)$ .

The truncated moment estimate requires introducing a good truncation which is defined as follows. For all  $|u| = N$ ,  $k \in \llbracket 0, N \rrbracket$ ,  $a > 0$  and  $b > 0$ , define the truncating set

$$G_{u,k} := \left\{ \forall n \in \llbracket 1, N - k \rrbracket : X_u^{(k)}(n) \leq \beta \cdot N \cdot \left( A\left(\frac{n+k}{N}\right) - A\left(\frac{k}{N}\right) \right) + a \cdot n + b \right\}.$$

Define the truncated partition function

$$Z_{\beta, N-k}^{(k), \leq} := \sum_{|u|=N-k} e^{\beta X_u^{(k)}} \mathbf{1}_{G_{u,k}}. \quad (4.2.5)$$

We explain roughly why truncated partition function defined in (4.2.5) is good. Firstly, In Lemma 4.2.5, we show that with  $a > 0$  and with  $b > 0$  sufficiently large, the first moment of truncated partition function is approximately the same as the first moment of the untruncated one. Secondly, in Lemma 4.2.6, we show that with  $b > 0$  and with  $a > 0$  sufficiently small, the second moment of truncated partition function is comparable with the square of the first moment of the untruncated one. Once we have these two estimates, by passing to a Paley–Zygmund type argument, we are able to estimate left tail of the untruncated partition function.

We now proceed with the first moment estimate.

**Lemma 4.2.5** (First moment estimate). *For all  $a > 0$  and  $b \geq \frac{\hat{A}'(0)}{a} \log \left( 10 \max \left\{ \frac{e^{-a^2/(2\hat{A}'(0))}}{1-a^2/(2\hat{A}'(0))}, 1 \right\} \right)$ , there exist  $N_0 = N_0(A, \beta, a, b) \in \mathbb{N}$  such that for all  $N \geq N_0$  and  $k \in \llbracket 0, K \rrbracket$  where  $K := \left\lfloor \frac{2}{\log(19/10)} \log N \right\rfloor$ ,*

$$\mathbb{E} \left[ Z_{\beta, N-k}^{(k), \leq} \right] \geq \frac{7}{10} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right].$$

*Proof.* By exponential tilting, we have

$$\begin{aligned}
 \mathbb{E} \left[ Z_{\beta, N-k}^{(k), \leq} \right] &= \mathbb{E} \left[ \sum_{|u|=N-k} e^{\beta X_u^{(k)}} \mathbf{1}_{G_{u,k}} \right] \\
 &= 2^{N-k} \mathbb{E} \left[ e^{\beta X_{N-k}^{(k)}} \mathbf{1}_{G_{N-k}} \right] \\
 &= 2^{N-k} \frac{\beta^2 N}{2} \left( 1 - A \left( \frac{k}{N} \right) \right) \cdot \mathbb{P} \left( \forall n \in \llbracket 1, N-k \rrbracket : X_{N-k}^{(k)}(n) \leq a \cdot n + b \right) \\
 &= \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \cdot \mathbb{P} \left( \forall n \in \llbracket 1, N-k \rrbracket : X_{N-k}^{(k)}(n) \leq a \cdot n + b \right). \quad (4.2.6)
 \end{aligned}$$

We want to give a lower bound of the probability in (4.2.6). The union bound and the Chernoff bound yield

$$\begin{aligned}
 \mathbb{P} \left( \exists n \in \llbracket 1, N-k \rrbracket : X_{N-k}^{(k)}(n) > a \cdot n + b \right) &\leq \sum_{n=1}^{N-k} \mathbb{P} \left( X_{N-k}^{(k)}(n) > a \cdot n + b \right) \\
 &\leq \sum_{n=1}^{N-k} \exp \left( - \frac{(a \cdot n + b)^2}{2N \cdot (A(\frac{n+k}{N}) - A(\frac{k}{N}))} \right). \quad (4.2.7)
 \end{aligned}$$

Letting  $K_1 := \left\lfloor \frac{2}{\log(19/10)} (\log N)^2 \right\rfloor$ , we separate (4.2.7) into two terms

$$(4.2.7) = \sum_{n=1}^{K_1} \exp \left( - \frac{(a \cdot n + b)^2}{2N \cdot (A(\frac{n+k}{N}) - A(\frac{k}{N}))} \right) + \sum_{n=K_1+1}^{N-k} \exp \left( - \frac{(a \cdot n + b)^2}{2N \cdot (A(\frac{n+k}{N}) - A(\frac{k}{N}))} \right). \quad (4.2.8)$$

We start with bounding the first term of (4.2.8). Define  $h(N) := C \frac{K_1^{1+\alpha}}{N} + C \frac{K_1 k^\alpha}{N} + C \frac{k^{1+\alpha}}{N^\alpha}$ . By Lemma 4.2.3, for all  $n \in \llbracket 1, K_1 \rrbracket$ , we have

$$\begin{aligned}
 &N \cdot \left( A \left( \frac{n+k}{N} \right) - A \left( \frac{k}{N} \right) \right) \\
 &\leq \hat{A}'(0)n + C \frac{n(n+k)^\alpha}{N} + C \frac{n^{1+\alpha}}{N^\alpha} \\
 &\leq \hat{A}'(0)n + C \frac{n(n^\alpha + k^\alpha)}{N} + C \frac{n^{1+\alpha}}{N^\alpha} \quad (\text{By subadditivity of } x \mapsto x^\alpha) \\
 &\leq \hat{A}'(0)n + C \frac{K_1^{1+\alpha}}{N} + C \frac{K_1 k^\alpha}{N} + C \frac{k^{1+\alpha}}{N^\alpha} \\
 &= \hat{A}'(0)n + h(N). \quad (4.2.9)
 \end{aligned}$$

By (4.2.9), the first term of (4.2.8) is bounded from above by

$$\begin{aligned}
 & \sum_{n=1}^{K_1} \exp\left(-\frac{(a \cdot n + b)^2}{2N \cdot (A(\frac{n+k}{N}) - A(\frac{k}{N}))}\right) \\
 & \leq \sum_{n=1}^{K_1} \exp\left(-\frac{(a \cdot n + b)^2}{2(\hat{A}'(0)n + h(N))}\right) \\
 & = \sum_{n=1}^{K_1} \exp\left(-\frac{a^2}{2(\hat{A}'(0) + \frac{1}{n}h(N))}\right) \exp\left(-\frac{ab}{\hat{A}'(0) + \frac{1}{n}h(N)}\right) \underbrace{\exp\left(-\frac{b^2}{2N(\hat{A}'(0) + h(N))}\right)}_{\leq 1}
 \end{aligned} \tag{4.2.10}$$

Since for all  $n \geq 1$ ,

$$\hat{A}'(0) + \frac{1}{n}h(N) \geq \hat{A}'(0) + h(N),$$

(4.2.10) is bounded from above by

$$\begin{aligned}
 (4.2.10) & \leq \sum_{n=1}^{K_1} \exp\left(-\frac{a^2}{2(\hat{A}'(0) + h(N))}\right) \exp\left(-\frac{ab}{\hat{A}'(0) + h(N)}\right) \\
 & \leq \sum_{n=1}^{\infty} \exp\left(-\frac{a^2}{2(\hat{A}'(0) + h(N))}\right) \exp\left(-\frac{ab}{\hat{A}'(0) + h(N)}\right) \\
 & \leq \frac{\exp\left(-\frac{a^2}{2(\hat{A}'(0) + h(N))}\right)}{1 - \exp\left(-\frac{a^2}{2(\hat{A}'(0) + h(N))}\right)} \exp\left(-\frac{ab}{\hat{A}'(0) + h(N)}\right).
 \end{aligned} \tag{4.2.11}$$

Since  $h(N) \rightarrow 0$  as  $N \rightarrow \infty$ , by the choice of parameters  $a$  and  $b$ , we have

$$\lim_{N \rightarrow \infty} \frac{\exp\left(-\frac{a^2}{2(\hat{A}'(0) + h(N))}\right)}{1 - \exp\left(-\frac{a^2}{2(\hat{A}'(0) + h(N))}\right)} \exp\left(-\frac{ab}{\hat{A}'(0) + h(N)}\right) = \frac{e^{-a^2/(2\hat{A}'(0))}}{1 - e^{-a^2/(2\hat{A}'(0))}} e^{-ab/\hat{A}'(0)} \leq \frac{1}{10}. \tag{4.2.12}$$

Combining (4.2.11) and (4.2.12), there exists  $N_1 = N_1(A, \beta) \in \mathbb{N}$  such that for all  $N \in N_1$ , the first term of (4.2.8) is bounded from above by

$$\sum_{n=1}^{K_1} \exp\left(-\frac{(a \cdot n + b)^2}{2N \cdot (A(\frac{n+k}{N}) - A(\frac{k}{N}))}\right) \leq \frac{2}{10}. \tag{4.2.13}$$

It remains to bound the second term of (4.2.8). By Lemma 4.2.2,

$$\begin{aligned}
 & \sum_{n=K_1+1}^{N-k} \exp\left(-\frac{(a \cdot n + b)^2}{2N \cdot (A(\frac{n+k}{N}) - A(\frac{k}{N}))}\right) \\
 & \leq \sum_{n=K_1+1}^{N-k} \exp\left(-\frac{a^2}{2(1 + k/n)\hat{A}'(0)}n\right) \\
 & \leq \exp\left(-\frac{a^2}{2(1 + K_1/(K_1 + 1))\hat{A}'(0)}(K_1 + 1)\right) \frac{1}{1 - \exp\left(-\frac{a^2}{2(1 + K_1/(K_1 + 1))\hat{A}'(0)}\right)}.
 \end{aligned} \tag{4.2.14}$$

By our choice of  $K_1$ , (4.2.14) converges to 0 as  $N \rightarrow \infty$ . In particular, there exists  $N_2 = N_2(A, \beta)$  such that for all  $N \geq N_2$ ,

$$\sum_{n=K_1+1}^{N-k} \exp\left(-\frac{(a \cdot n + b)^2}{2N \cdot (A(\frac{n+k}{N}) - A(\frac{k}{N}))}\right) \leq \frac{1}{10}. \quad (4.2.15)$$

Let  $N_3 := \max\{N_1, N_2\}$ . Combining (4.2.8), (4.2.13) and (4.2.15), for all  $N \geq N_3$  and for all  $r \in \llbracket 0, K \rrbracket$ ,

$$\mathbb{P}\left(\exists n \in \llbracket 1, N-k \rrbracket : X_{N-k}^{(k)}(n) > a \cdot n + b\right) \leq \frac{3}{10}. \quad (4.2.16)$$

Combining (4.2.6) and (4.2.16), we conclude that for all  $N \geq N_3$  and for all  $k \in \llbracket 0, K \rrbracket$ ,

$$\mathbb{E}\left[Z_{\beta, N-k}^{(k), \leq}\right] \geq \frac{7}{10} \mathbb{E}\left[Z_{\beta, N-k}^{(k)}\right],$$

and this completes the proof.  $\blacksquare$

It remains to provide a second moment estimate of the truncated partition function.

**Lemma 4.2.6** (Second moment estimate). *Let  $\beta < \beta_c$ . For all  $a \in (0, \frac{\log 2}{\beta} - \frac{1}{2}\beta\hat{A}'(0))$ , where and  $b > 0$ , there exist  $N_0 = N_0(A, \beta) \in \mathbb{N}$  and  $C = C(A, \beta) > 0$  such that for all  $N \geq N_0$  and  $k \in \llbracket 0, K \rrbracket$ , where  $K := \lfloor \frac{2}{\log(19/10)} \log N \rfloor$ , we have*

$$\mathbb{E}\left[(Z_{\beta, N-k}^{(k), \leq})^2\right] \leq C \mathbb{E}\left[Z_{\beta, N-k}^{(k)}\right]^2.$$

*Proof.* Fix  $\beta < \beta_c$  and  $a \in (0, \frac{\log 2}{\beta} - \frac{1}{2}\beta\hat{A}'(0))$ . Define

$$c = c(A, \beta) := \log 2 - \frac{1}{2}\beta^2\hat{A}'(0) - a > 0. \quad (4.2.17)$$

By the many-to-two lemma, we have

$$\mathbb{E}\left[(Z_{\beta, N-k}^{(k), \leq})^2\right] = \mathbb{E}\left[Z_{2\beta, N-k}^{(k), \leq}\right] + \sum_{\ell=0}^{N-k-1} \sum_{\substack{|u \wedge w| = \ell, \\ |u| = |w| = N-k}} \mathbb{E}\left[e^{\beta(X_u^{(k)} + X_w^{(k)})} \mathbf{1}_{G_{u,k} \cap G_{w,k}}\right]. \quad (4.2.18)$$

We start with bounding the first term of (4.2.18). By exponential tilting, the first term of (4.2.18) is bounded from above by

$$\begin{aligned} \mathbb{E}\left[Z_{2\beta, N-k}^{(k), \leq}\right] &= 2^{N-k} \exp\left(2\beta^2 N \left(1 - A\left(\frac{k}{N}\right)\right)\right) \\ &\quad \cdot \mathbb{P}\left(X_u^{(k)} \leq -\beta \cdot N \cdot \left(1 - A\left(\frac{k}{N}\right)\right) + a(N-k) + b\right) \\ &\leq 2^{N-k} \exp\left(\frac{3}{2}\beta^2 N \left(1 - A\left(\frac{k}{N}\right)\right)\right) \exp(a(N-k) + b) \\ &= \mathbb{E}\left[Z_{\beta, N-k}^{(k)}\right]^2 2^{-(N-k)} \exp\left(\frac{1}{2}\beta^2 N \left(1 - A\left(\frac{k}{N}\right)\right)\right) \exp(a(N-k) + b) \end{aligned} \quad (4.2.19)$$

By Lemma 4.2.2, there exists  $N_1 = N_1(A, \beta) \in \mathbb{N}$  such that for all  $N \geq N_1$  and  $k \in \llbracket 0, K \rrbracket$ , (4.2.19) is bounded from above by

$$\begin{aligned}
 (4.2.19) &\leq \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 2^{-(N-k)} \exp \left( \frac{1}{2} \beta^2 \hat{A}'(0) N \right) \exp(a(N-k) + b) \\
 &= \exp \left( \beta b + \frac{1}{2} \hat{A}'(0) k \right) \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \exp \left( \left( -\log 2 + \frac{1}{2} \beta^2 \hat{A}'(0) + a \right) (N-k) \right) \\
 &\leq \exp \left( \beta b + \frac{1}{2} \hat{A}'(0) K \right) \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \exp(-c(N-k)), \tag{4.2.20}
 \end{aligned}$$

where the constant  $c$  is defined in (4.2.17).

It remains to bound the second term of (4.2.18). For all  $|u| = |w| = N-k$  with  $|u \wedge w| = \ell \in \llbracket 0, N-k-1 \rrbracket$ , we have

$$G_{u,k} \cap G_{w,k} \subseteq \left\{ X_{u \wedge w}^{(k)} \leq \beta \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) + a \cdot \ell + b \right\}. \tag{4.2.21}$$

Combining (4.2.21) and the branching property, we have

$$\begin{aligned}
 &\mathbb{E} \left[ e^{\beta(X_u^{(k)} + X_w^{(k)})} \mathbf{1}_{G_{u,k} \cap G_{w,k}} \right] \\
 &\leq \mathbb{E} \left[ e^{2\beta X_{u \wedge w}^{(k)}} \mathbf{1} \left\{ X_{u \wedge w}^{(k)} \leq \beta \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) + a \cdot \ell + b \right\} \right] \\
 &\quad \cdot \mathbb{E} \left[ e^{\beta(X_u^{(k)} + X_w^{(k)} - 2X_{u \wedge w}^{(k)})} \right] \\
 &= \mathbb{E} \left[ e^{2\beta X_{u \wedge w}^{(k)}} \mathbf{1} \left\{ X_{u \wedge w}^{(k)} \leq \beta \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) + a \cdot \ell + b \right\} \right] \mathbb{E} \left[ e^{\beta X_{u \setminus (u \wedge w)}^{(k+\ell)}} \right]^2, \tag{4.2.22}
 \end{aligned}$$

where  $u \setminus (u \wedge w)$  is defined in Definition 4.2.4, meaning the suffix of  $u$  such that  $u = (u \wedge w)(u \setminus (u \wedge w))$ . The second term of (4.2.22) is equal to

$$\mathbb{E} \left[ e^{\beta X_{u \setminus (u \wedge w)}^{(\ell+k)}} \right]^2 = \exp \left( \beta^2 \cdot N \cdot \left( 1 - A \left( \frac{\ell+k}{N} \right) \right) \right). \tag{4.2.23}$$

By exponential tilting, the first term of (4.2.22) is equal to

$$\begin{aligned}
 &\mathbb{E} \left[ e^{2\beta X_{u \wedge w}^{(k)}} \right] \mathbb{P} \left( X_{u \wedge w}^{(k)} \leq -\beta \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) + a \cdot \ell + b \right) \\
 &= \exp \left( 2\beta^2 \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) \right) \\
 &\quad \cdot \mathbb{P} \left( X_{u \wedge w}^{(k)} \leq -\beta \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) + a \cdot \ell + b \right) \tag{4.2.24}
 \end{aligned}$$

**Case 1:**  $-\beta \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) + a \cdot \ell + b \geq 0$ . In this case, by bounding the probability in (4.2.24) by 1, we obtain

$$(4.2.24) \leq \exp \left( 2\beta^2 \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) \right). \tag{4.2.25}$$

Combining (4.2.22), (4.2.23) and (4.2.25), we see that

$$\begin{aligned}
 & \mathbb{E} \left[ e^{\beta(X_u^{(k)} + X_w^{(k)})} \mathbf{1}_{G_{u,k} \cap G_{w,k}} \right] \\
 & \leq \exp \left( \beta^2 \cdot N \cdot \left( 1 - A \left( \frac{k}{N} \right) \right) \right) \cdot \exp \left( \beta^2 \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) \right) \\
 & \leq \exp \left( \beta^2 \cdot N \cdot \left( 1 - A \left( \frac{k}{N} \right) \right) \right) \cdot \exp(\beta \cdot (a \cdot \ell + b)). \tag{4.2.26}
 \end{aligned}$$

Since there are  $2^{2(N-k)-\ell-1}$  pairs of vertices  $u$  and  $w$  such that  $|u| = |w| = N - k$  with  $|u \wedge w| = \ell \in \llbracket 0, N - k - 1 \rrbracket$ , we conclude that

$$\begin{aligned}
 & \sum_{\substack{|u \wedge w| = \ell, \\ |u| = |w| = N - k}} \mathbb{E} \left[ e^{\beta(X_u^{(k)} + X_w^{(k)})} \mathbf{1}_{G_{u,k} \cap G_{w,k}} \right] \\
 & \leq 2^{2(N-k)-\ell-1} \exp \left( \beta^2 \cdot N \cdot \left( 1 - A \left( \frac{k}{N} \right) \right) \right) \cdot \exp(\beta \cdot (a \cdot \ell + b)) \\
 & \leq \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \cdot 2^{-\ell-1} \cdot \exp(\beta \cdot (a \cdot \ell + b)) \\
 & = \frac{1}{2} \exp(\beta b) \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \cdot \exp((-\log 2 + \beta a)\ell) \\
 & \leq \frac{1}{2} \exp(\beta b) \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \cdot \exp(-c\ell), \tag{4.2.27}
 \end{aligned}$$

where the constant  $c$  is defined in (4.2.17).

**Case 2:**  $-\beta \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) + a \cdot \ell + b < 0$ . In this case, the Chernoff bound yields

$$\begin{aligned}
 & \tag{4.2.24} \\
 & \leq \exp \left( 2\beta^2 \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) \right) \\
 & \quad \cdot \exp \left( - \frac{(\beta \cdot N \cdot (A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right)) - (a \cdot \ell + b))^2}{2N \cdot (A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right))} \right) \\
 & = \exp \left( 2\beta^2 \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) \right) \cdot \exp \left( -\frac{1}{2} \beta^2 \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) \right) \\
 & \quad \cdot \exp(\beta \cdot (a \cdot \ell + b)) \cdot \underbrace{\exp \left( - \frac{(a \cdot \ell + b)^2}{2N \cdot (A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right))} \right)}_{\leq 1} \\
 & \leq \exp \left( \frac{3}{2} \beta^2 \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) \right) \cdot \exp(\beta \cdot (a \cdot \ell + b)). \tag{4.2.28}
 \end{aligned}$$

Combining (4.2.22) and (4.2.28), we conclude that

$$\begin{aligned}
 & \mathbb{E} \left[ e^{\beta(X_u^{(k)} + X_w^{(k)})} \mathbf{1}_{G_{u,k} \cap G_{w,k}} \right] \\
 & \leq \exp \left( \beta^2 \cdot N \cdot \left( 1 - A \left( \frac{k}{N} \right) \right) \right) \cdot \exp \left( \frac{1}{2} \beta^2 \cdot N \cdot \left( A \left( \frac{\ell+k}{N} \right) - A \left( \frac{k}{N} \right) \right) \right) \cdot \exp(\beta \cdot (a \cdot \ell + b)). \tag{4.2.29}
 \end{aligned}$$

We distinguish the two cases of  $\ell$ .

- (i) If  $\ell \in \llbracket 0, K_1 \rrbracket$ , by Lemma 4.2.3, there exists  $N_2 = N_2(A, \beta) \in \mathbb{N}$  such that for all  $N \geq N_2$  and  $k \in \llbracket 0, K \rrbracket$ , (4.2.29) is bounded from above by

$$(4.2.29) \leq \exp\left(\beta^2 \cdot N \cdot \left(1 - A\left(\frac{k}{N}\right)\right)\right) \exp\left(\frac{1}{2}\beta^2 \cdot \left(\hat{A}'(0)\ell + h(N)\right)\right) \cdot \exp(\beta \cdot (a \cdot \ell + b)).$$

Since there are  $2^{2(N-k)-\ell-1}$  pairs of vertices  $u$  and  $w$  such that  $|u| = |w| = N - k$  with  $|u \wedge w| = \ell \in \llbracket 0, N - k - 1 \rrbracket$ , we conclude that

$$\begin{aligned} & \sum_{\substack{|u \wedge w| = \ell, \\ |u| = |w| = N - k}} \mathbb{E} \left[ e^{\beta(X_u^{(k)} + X_w^{(k)})} \mathbf{1}_{G_{u,k} \cap G_{w,k}} \right] \\ & \leq 2^{2(N-k)-\ell-1} \exp\left(\beta^2 \cdot N \cdot \left(1 - A\left(\frac{k}{N}\right)\right)\right) \cdot \exp\left(\frac{1}{2}\beta^2 \cdot \left(\hat{A}'(0)\ell + h(N)\right)\right) \cdot \exp(\beta \cdot (a \cdot \ell + b)) \\ & = \frac{1}{2} \exp\left(\beta b + \frac{1}{2}\beta^2 h(N)\right) \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \exp\left(\left(-\log 2 + \frac{1}{2}\beta^2 \hat{A}'(0) + \beta a\right)\ell\right) \\ & \leq \frac{1}{2} \exp\left(\beta b + \frac{1}{2}\beta^2 h(N)\right) \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \exp(-c\ell), \end{aligned} \quad (4.2.30)$$

where the constant  $c$  is defined in (4.2.17).

- (ii) If  $\ell \in \llbracket K_1 + 1, N - k - 1 \rrbracket$ , by Lemma 4.2.2, there exists  $N_3 = N_3(A, \beta) \in \mathbb{N}$  such that for all  $N \geq N_3$  and  $k \in \llbracket 0, K \rrbracket$ , (4.2.29) is bounded from above by

$$(4.2.29) \leq \exp\left(\beta^2 \cdot N \cdot \left(1 - A\left(\frac{k}{N}\right)\right)\right) \exp\left(\frac{1}{2}\beta^2 \cdot \left(\hat{A}'(0)(\ell + k)\right)\right) \cdot \exp(\beta \cdot (a \cdot \ell + b)).$$

Since there are  $2^{2(N-k)-\ell-1}$  pairs of vertices  $u$  and  $w$  such that  $|u| = |w| = N - k$  with  $|u \wedge w| = \ell \in \llbracket 0, N - k - 1 \rrbracket$ , we conclude that

$$\begin{aligned} & \sum_{\substack{|u \wedge w| = \ell, \\ |u| = |w| = N - k}} \mathbb{E} \left[ e^{\beta(X_u^{(k)} + X_w^{(k)})} \mathbf{1}_{G_{u,k} \cap G_{w,k}} \right] \\ & \leq 2^{2(N-k)-\ell-1} \exp\left(\beta^2 \cdot N \cdot \left(1 - A\left(\frac{k}{N}\right)\right)\right) \exp\left(\frac{1}{2}\beta^2 \cdot \left(\hat{A}'(0)(\ell + k)\right)\right) \cdot \exp(\beta \cdot (a \cdot \ell + b)) \\ & = \frac{1}{2} \exp\left(\beta b + \frac{1}{2}\beta^2 \hat{A}'(0)k\right) \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \exp\left(\left(-\log 2 + \frac{1}{2}\beta^2 \hat{A}'(0) + \beta a\right)\ell\right) \\ & \leq \frac{1}{2} \exp\left(\beta b + \frac{1}{2}\beta^2 \hat{A}'(0)K\right) \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \exp(-c\ell), \end{aligned} \quad (4.2.31)$$

where the constant  $c$  is defined in (4.2.17).

Combining (4.2.18), (4.2.20), (4.2.27) in Case 1, (4.2.30) in (i) of Case 2 and

(4.2.31) in (ii) of Case 2, we have that for all  $N \geq \max\{N_1, N_2, N_3\}$  and  $k \in \llbracket 0, K \rrbracket$ ,

$$\begin{aligned}
 & \mathbb{E} \left[ (Z_{\beta, N-k}^{(k), \leq})^2 \right] \\
 &= \mathbb{E} \left[ Z_{2\beta, N-k}^{(k), \leq} \right] + \sum_{\ell=0}^{N-k-1} \sum_{\substack{|u \wedge w| = \ell, \\ |u| = |w| = N-k}} \mathbb{E} \left[ e^{\beta(X_u^{(k)} + X_w^{(k)})} \mathbf{1}_{G_{u,k} \cap G_{w,k}} \right] \\
 &\leq \exp\left(\beta b + \frac{1}{2} \hat{A}'(0) K\right) \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \exp(-c(N-k)) \\
 &+ \sum_{\ell=0}^{K_1} \frac{1}{2} \exp\left(\beta b + \frac{1}{2} \beta^2 h(N)\right) \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \exp(-c\ell) \\
 &+ \sum_{\ell=K_1+1}^{N-k-1} \frac{1}{2} \exp\left(\beta b + \frac{1}{2} \beta^2 \hat{A}'(0) K\right) \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \exp(-c\ell) \\
 &\leq \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \left[ \frac{1}{2} \exp\left(\beta b + \frac{1}{2} \beta^2 h(N)\right) \sum_{\ell=0}^{\infty} \exp(-c\ell) \right. \\
 &\quad \left. + \frac{1}{2} \exp\left(\beta b + \frac{1}{2} \beta^2 \hat{A}'(0) K\right) \sum_{\ell=K_1+1}^{\infty} \exp(-c\ell) \right] \\
 &= \leq \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \left[ \frac{1}{2} \exp\left(\beta b + \frac{1}{2} \beta^2 h(N)\right) \frac{1}{1-e^{-c}} + \frac{1}{2} \exp\left(\beta b + \frac{1}{2} \beta^2 \hat{A}'(0) K\right) \frac{e^{-c(K_1+1)}}{1-e^{-c}} \right] \\
 &= \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2 \frac{1}{2} \frac{e^{\beta b}}{1-e^{-c}} \underbrace{\left[ e^{\frac{1}{2} \beta^2 h(N)} + e^{\frac{1}{2} \beta^2 \hat{A}'(0) K - c(K_1+1)} \right]}_{\rightarrow 2, \quad N \rightarrow \infty}
 \end{aligned}$$

Thus, there exist  $N_0 = N_0(A, \beta) \in \mathbb{N}$  with  $N_0 \geq \max\{N_1, N_2, N_3\}$  and  $C = C(A, \beta) := \frac{3}{2} \frac{e^{\beta b}}{1-e^{-c}}$  such that for all  $N \geq N_0$  and  $k \in \llbracket 0, K \rrbracket$ ,

$$\mathbb{E} \left[ (Z_{\beta, N-k}^{(k), \leq})^2 \right] \leq C \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2,$$

and this completes the proof.  $\blacksquare$

### 4.2.3 Proof of Proposition 4.2.1

We are now ready to prove Proposition 4.2.1.

*Proof of Proposition 4.2.1.* Fix  $\beta < \beta_c$ . By Lemma 4.2.5 and Lemma 4.2.6, if we choose

$$a = \frac{1}{2\beta} (\log 2 - \frac{1}{2} \beta^2 \hat{A}'(0)) \quad \text{and} \quad b = \frac{\hat{A}'(0)}{a} \log \left( 10 \max \left\{ \frac{e^{-a^2/(2\hat{A}'(0))}}{1 - a^2/(2\hat{A}'(0))}, 1 \right\} \right), \tag{4.2.32}$$

then there exist  $N_1 = N_1(A, \beta) \in \mathbb{N}$ ,  $N_2 = N_2(A, \beta) \in \mathbb{N}$  and  $C = C(A, \beta) > 0$  such that for all  $N \geq N_1$  and  $k \in \llbracket 0, K \rrbracket$ ,

$$\mathbb{E} \left[ Z_{\beta, N-k}^{(k), \leq} \right] \geq \frac{7}{10} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right], \tag{4.2.33}$$

and for all  $N \geq N_2$  and  $k \in \llbracket 0, K \rrbracket$ ,

$$\mathbb{E} \left[ (Z_{\beta, N-k}^{(k), \leq})^2 \right] \leq C \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]^2. \quad (4.2.34)$$

Now, since  $Z_{\beta, N-k}^{(k), \leq} \leq Z_{\beta, N-k}^{(k)}$ , we have

$$\mathbb{P} \left( Z_{\beta, N-k}^{(k)} > \frac{1}{2} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \right) \geq \mathbb{P} \left( Z_{\beta, N-k}^{(k), \leq} > \frac{1}{2} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \right). \quad (4.2.35)$$

On the other hand, by the Cauchy–Schwarz inequality, we have

$$\begin{aligned} & \mathbb{E} \left[ Z_{\beta, N-k}^{(k), \leq} \right] \\ &= \mathbb{E} \left[ Z_{\beta, N-k}^{(k), \leq} \mathbf{1} \left\{ Z_{\beta, N-k}^{(k), \leq} > \frac{1}{2} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \right\} \right] + \underbrace{\mathbb{E} \left[ Z_{\beta, N-k}^{(k), \leq} \mathbf{1} \left\{ Z_{\beta, N-k}^{(k), \leq} \leq \frac{1}{2} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \right\} \right]}_{\leq \frac{1}{2} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]} \\ &\leq \mathbb{E} \left[ (Z_{\beta, N-k}^{(k), \leq})^2 \right]^{1/2} \mathbb{P} \left( Z_{\beta, N-k}^{(k), \leq} > \frac{1}{2} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \right)^{1/2} + \frac{1}{2} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]. \end{aligned} \quad (4.2.36)$$

Let  $N_0 = N_0(A, \beta) := \max\{N_1, N_2\}$ . Combining (4.2.33), (4.2.34) and (4.2.36), we derive that for all  $N \geq N_0$  and  $k \in \llbracket 0, K \rrbracket$ ,

$$\mathbb{P} \left( Z_{\beta, N-k}^{(k), \leq} \geq \frac{1}{2} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \right)^{1/2} \geq \frac{\mathbb{E} \left[ Z_{\beta, N-k}^{(k), \leq} \right] - \frac{1}{2} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right]}{\mathbb{E} \left[ (Z_{\beta, N-k}^{(k), \leq})^2 \right]^{1/2}} \geq \frac{2}{10\sqrt{C}}. \quad (4.2.37)$$

Combining (4.2.35) and (4.2.37), we conclude that for all  $N \geq N_0$  and  $r \in \llbracket 0, K \rrbracket$ ,

$$\begin{aligned} \mathbb{P} \left( Z_{\beta, N-k}^{(k)} \leq \frac{1}{2} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \right) &= 1 - \mathbb{P} \left( Z_{\beta, N-k}^{(k)} > \frac{1}{2} \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \right) \\ &\leq 1 - \mathbb{P} \left( Z_{\beta, N-k}^{(k), \leq} > \frac{1}{2} \cdot \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \right) \\ &\leq 1 - \frac{4}{100C}. \end{aligned}$$

By taking  $\eta_0 = \eta_0(A, \beta) = 1 - \frac{4}{100C}$ , the proof is completed.  $\blacksquare$

### 4.3 Left tail estimates via a bootstrap argument

The main goal of this section is to provide a finer estimate of the left tail of  $Z_{\beta, N}$  than the one provided by Proposition 4.2.1. Namely, we want to an exponentially decaying sequence that satisfies (4.1.8), and this is manifested in the following proposition.

**Proposition 4.3.1.** *Let  $K = \lfloor \frac{2}{\log(19/10)} \log N \rfloor$ . There exist  $N_0 = N_0(A, \beta) \in \mathbb{N}$  and two sequences  $(\eta_k)_{k=0}^K$  and  $(\varepsilon_k)_{k=0}^K$  such that*

(i) *for all  $N \geq N_0$  and  $k \in \llbracket 0, K \rrbracket$ ,*

$$\mathbb{P} (Z_{\beta, N} \leq \varepsilon_k \mathbb{E} [Z_{\beta, N}]) \leq \eta_k.$$

(ii) for all  $s > 0$ , there exists a constant  $C = C(A, \beta, s) > 0$ , independent of  $K$ , such that

$$\sum_{r=0}^K (\varepsilon_{r+1})^{-s} \eta_r \leq C.$$

The proof of Proposition 4.3.1 requires the following two lemmas. For all  $k \in \llbracket 0, N-1 \rrbracket$ , define

$$M_{\beta,1}^{(k)} = \frac{e^{\beta X_1^{(k)}}}{\mathbb{E} \left[ Z_{\beta,1}^{(k)} \right]}.$$

The first lemma states the bootstrap inequality.

**Lemma 4.3.2.** *For all  $c > 0$ ,  $\delta > 0$ ,  $N \in \mathbb{N}$ , and  $k \in \llbracket 0, N-1 \rrbracket$ , we have*

$$\mathbb{P} \left( Z_{\beta, N-k}^{(k)} \leq c\delta \cdot \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \right) \leq \left[ \mathbb{P} \left( Z_{\beta, N-k-1}^{(k+1)} \leq c \cdot \mathbb{E} \left[ Z_{\beta, N-k-1}^{(k+1)} \right] \right) + \mathbb{P} \left( M_{\beta,1}^{(k)} \leq \delta \right) \right]^2.$$

*Proof.* Fix  $N \in \mathbb{N}$ . For every  $c > 0$ ,  $\delta > 0$  and  $k \in \llbracket 0, N-1 \rrbracket$ , we have

$$\begin{aligned} & \mathbb{P} \left( Z_{\beta, N-k}^{(k)} \leq c\delta \cdot \mathbb{E} \left[ Z_{\beta, N-k}^{(k)} \right] \right) \\ &= \mathbb{P} \left( \sum_{|u|=1} M_{\beta,u}^{(k)} \cdot Z_{\beta, N-k-1}^{(k),u} \leq c\delta \cdot \mathbb{E} \left[ Z_{\beta, N-k-1}^{(k+1)} \right] \right) \\ &\leq \mathbb{P} \left( \forall |u|=1 : M_{\beta,u}^{(k)} \cdot Z_{\beta, N-k-1}^{(k),u} \leq c\delta \cdot \mathbb{E} \left[ Z_{\beta, N-k-1}^{(k+1)} \right] \right) \\ &= \mathbb{P} \left( M_{\beta,1}^{(k)} \cdot Z_{\beta, N-k-1}^{(k+1)} \leq c\delta \cdot \mathbb{E} \left[ Z_{\beta, N-k-1}^{(k+1)} \right] \right)^2 \quad (\text{independence}) \\ &\leq \left[ \mathbb{P} \left( Z_{\beta, N-k-1}^{(k+1)} \leq c \cdot \mathbb{E} \left[ Z_{\beta, N-k-1}^{(k+1)} \right] \right) + \mathbb{P} \left( M_{\beta,1}^{(k)} \leq \delta \right) \right]^2. \end{aligned}$$

This completes the proof. ■

The next lemma provides a uniform estimate for the left tail of  $M_{\beta,1}^{(k)}$  when  $k$  is small compared to  $N$ .

**Lemma 4.3.3.** *Let  $N \in \mathbb{N}$  and  $K = \lfloor \frac{2}{\log(19/10)} \log N \rfloor$ . For all  $\beta > 0$ ,  $k \in \llbracket 0, K \rrbracket$  and  $s > 0$ , there exist  $N_0 = N_0(A, \beta, s) \in \mathbb{N}$  and a constant  $C = C(A, \beta, s) > 0$ , independent of  $N$ , such that for all  $N \geq N_0$ ,*

$$\mathbb{E} \left[ (M_{\beta,1}^{(k)})^{-s} \right] \leq C.$$

*In particular, by Markov's inequality, this implies that for all  $\delta > 0$  and for all  $N \geq N_0$ ,*

$$\mathbb{P} \left( M_{\beta,1}^{(k)} \leq \delta \right) \leq C_{\beta,s} \cdot \delta^s.$$

*Proof.* Let  $k \in \llbracket 0, K \rrbracket$ . By Lemma 4.2.3, there exist  $N_1 = N_1(A, \beta) \in \mathbb{N}$  and  $C_1 > 0$  such that for all  $N \geq N_1$ ,

$$N \left( A \left( \frac{k+1}{N} \right) - A \left( \frac{k}{N} \right) \right) \leq \hat{A}'(0) + \underbrace{C_1 \frac{(k+1)^\alpha}{N^\alpha} + C_1 \frac{1}{N^\alpha}}_{\rightarrow 0, N \rightarrow \infty}.$$

In particular, there exists  $N_2 = N_2(A, \beta) \in \mathbb{N}$  such that for all  $N \geq N_2$ ,

$$N \left( A \left( \frac{k+1}{N} \right) - A \left( \frac{k}{N} \right) \right) \leq \hat{A}'(0) + 1.$$

Let  $N_0 := \max\{N_1, N_2\}$  and  $C = C(A, \beta, s) := \exp\left(\frac{\beta^2 s^2}{2}(\hat{A}'(0) + 1)\right) \cdot 2^s \exp\left(\frac{\beta^2 s}{2}(\hat{A}'(0) + 1)\right)$ . For all  $N \geq N_0$ , we conclude that

$$\begin{aligned} & \mathbb{E} \left[ (M_{\beta,1}^{(k)})^{-s} \right] \\ &= \mathbb{E} \left[ e^{-\beta s X_1^{(k)}} \right] \cdot \mathbb{E} \left[ Z_{\beta,1}^{(k)} \right]^s \\ &= \exp\left(\frac{\beta^2 s^2}{2} N \left( A \left( \frac{k+1}{N} \right) - A \left( \frac{k}{N} \right) \right)\right) \cdot 2^s \exp\left(\frac{\beta^2 s}{2} N \left( A \left( \frac{k+1}{N} \right) - A \left( \frac{k}{N} \right) \right)\right) \\ &\leq \exp\left(\frac{\beta^2 s^2}{2}(\hat{A}'(0) + 1)\right) \cdot 2^s \exp\left(\frac{\beta^2 s}{2}(\hat{A}'(0) + 1)\right) = C, \end{aligned}$$

and this completes the proof.  $\blacksquare$

We now turn to the proof of Proposition 4.3.1.

*Proof of Proposition 4.3.1.* Let  $\eta_0 = \eta(A, \beta) < 1$  be the same as in Proposition 4.2.1, and for all  $k \in \llbracket 1, K \rrbracket$ , define  $\eta_k = \eta_{k-1}^{19/10}$ . Let  $\varepsilon_0 = \frac{1}{2}$ . For all  $k \in \llbracket 1, K \rrbracket$  and  $s > 0$ , define with  $C = C(A, \beta, \frac{19}{17}s)$  being the same constant appeared in Lemma 4.3.3

$$\varepsilon_k = \frac{1}{2} \left( \frac{1}{C^k} \prod_{n=0}^{k-1} (\eta_n^{19/20} - \eta_n) \right)^{\frac{17}{19} \frac{1}{s}}.$$

Firstly, we prove by induction that for all  $k \in \llbracket 0, K \rrbracket$ ,

$$\mathbb{P}(Z_{\beta,N} \leq \varepsilon_k \mathbb{E}[Z_{\beta,N}]) \leq \eta_k. \quad (4.3.1)$$

The case where  $k = 0$  follows from Proposition 4.2.1. Suppose that (4.3.1) is true for  $k = \ell - 1 \in \llbracket 0, K - 1 \rrbracket$ . Then, by induction hypothesis, Lemma 4.3.2 and Lemma 4.3.3,

$$\begin{aligned} & \mathbb{P} \left( Z_{\beta, N-\ell}^{(\ell)} \leq \varepsilon_{\ell-1} \cdot \left( \frac{1}{C} (\eta_{\ell-1}^{19/20} - \eta_{\ell-1}) \right) \mathbb{E} \left[ Z_{\beta, N-\ell}^{(\ell)} \right] \right) \\ &\leq \left( \mathbb{P} \left( Z_{\beta, N-\ell-1}^{(\ell+1)} \leq \varepsilon_{\ell-1} \mathbb{E} \left[ Z_{\beta, N-\ell-1}^{(\ell+1)} \right] \right) + \mathbb{P} \left( M_{\beta,1}^{(\ell)} \leq \left( \frac{1}{C} (\eta_{\ell-1}^{19/20} - \eta_{\ell-1}) \right)^{\frac{17}{19} \frac{1}{s}} \right) \right)^2 \\ &\leq (\eta_{\ell-1} + \eta_{\ell-1}^{19/20} - \eta_{\ell-1})^2 = (\eta_{\ell-1})^{19/10} = \eta_{\ell}. \end{aligned} \quad (4.3.2)$$

By induction, (4.3.1) holds for all  $k \in \llbracket 0, K \rrbracket$ .

It remains to show that for all  $s > 0$ ,  $\sum_{k=0}^K (\varepsilon_{k+1})^{-s} \eta_k$  is uniformly bounded in  $K$ . Since  $(\eta_k)$  is a decreasing sequence, for all  $k \geq 1$ ,

$$\begin{aligned}
 \varepsilon_k &= \frac{1}{2} \left( \frac{1}{C^k} \prod_{n=0}^{k-1} (\eta_n^{\frac{19}{20}} - \eta_n) \right)^{\frac{17}{19} \frac{1}{s}} \\
 &= \frac{1}{2} \left( \frac{1}{C^k} \prod_{n=0}^{k-1} \eta_n^{\frac{19}{20}} (1 - \eta_n^{\frac{1}{20}}) \right)^{\frac{17}{19} \frac{1}{s}} \\
 &\geq \frac{1}{2} \left( \frac{1}{C^k} (1 - \eta_0^{\frac{1}{20}})^k \prod_{n=0}^{k-1} \eta_n^{\frac{19}{20}} \right)^{\frac{17}{19} \frac{1}{s}} \\
 &= \frac{1}{2} \left( \frac{1}{C^k} (1 - \eta_0^{\frac{1}{20}})^k \prod_{n=0}^{k-1} \eta_0^{\frac{19}{20} (\frac{19}{10})^n} \right)^{\frac{17}{19} \frac{1}{s}} \quad (\text{Definition of } \eta_k) \\
 &= \frac{1}{2} \left( \frac{1}{C^k} (1 - \eta_0^{\frac{1}{20}})^k \eta_0^{\frac{19}{18} ((\frac{19}{10})^k - 1)} \right)^{\frac{17}{19} \frac{1}{s}}. \tag{4.3.3}
 \end{aligned}$$

Applying (4.3.3),

$$\begin{aligned}
 &\sum_{k=0}^K (\varepsilon_{k+1})^{-s} \eta_k \\
 &\leq \sum_{k=1}^{K-1} 2^s \left( \frac{1}{C^k} (1 - \eta_0^{\frac{1}{20}})^k \eta_0^{\frac{19}{18} ((\frac{19}{10})^k - 1)} \right)^{-\frac{17}{19}} \cdot \eta_k \\
 &= \sum_{n=1}^{K-1} 2^s \left( \frac{1}{C^k} (1 - \eta_0^{\frac{1}{20}})^k \eta_0^{\frac{19}{18} ((\frac{19}{10})^k - 1)} \right)^{-\frac{17}{19}} \cdot \eta_0^{(\frac{19}{10})^k} \quad (\text{Definition of } \eta_k) \\
 &\leq 2^s \eta_0^{\frac{19}{18}s} \sum_{k=1}^{K-1} \eta_0^{\frac{1}{18} (\frac{19}{10})^k} \cdot \left( \frac{C}{(1 - \eta_0^{\frac{1}{20}})} \right)^k \\
 &= 2^s \eta_0^{\frac{17}{18}} \sum_{k=1}^{K-1} \exp \left( (\log \eta_0) \frac{1}{18} \left( \frac{19}{10} \right)^k + \log \left( \frac{C}{(1 - \eta_0^{\frac{1}{20}})} \right) k \right) \\
 &\leq 2^s \eta_0^{\frac{17}{18}} \sum_{k=1}^{\infty} \exp \left( \underbrace{(\log \eta_0)}_{< 0} \frac{1}{18} \left( \frac{19}{10} \right)^k + \log \left( \frac{C}{(1 - \eta_0^{\frac{1}{20}})} \right) k \right) \tag{4.3.4}
 \end{aligned}$$

Since (4.3.4) is a finite sum depending only on  $s$  and  $\eta_0 = \eta_0(A, \beta)$ , we conclude that  $\sum_{k=0}^K (\varepsilon_{k+1})^{-s} \eta_k$  is bounded from above by a constant  $C_1 = C_1(A, \beta, s) > 0$  as desired.  $\blacksquare$

## 4.4 Proof of Theorem 4.1.2

The goal of this section is to prove Theorem 4.1.2. Before proceeding with the proof, we introduce the relevant notation for this section.

**Notation.** For given functions  $f : \mathbb{N} \rightarrow \mathbb{N}$  and  $g : \mathbb{N} \rightarrow \mathbb{N}$ , we write  $f(N) = \Theta(g(N))$  if there exist positive constants  $C_1 > 0$ ,  $C_2 > 0$ , and  $N_0 \in \mathbb{N}$  such that  $C_1|g(n)| \leq |f(n)| \leq C_2|g(n)|$  for all  $N \geq N_0$ . Also, for simplicity, we define  $W_{\beta,N} := Z_{\beta,N}/\mathbb{E}[Z_{\beta,N}]$ .

We now begin the proof of Theorem 4.1.2.

*Proof of Theorem 4.1.2.* Suppose that  $\beta < \beta_c$ . Let  $N \in \mathbb{N}$ . Fix  $K = \lfloor \frac{2}{\log(19/10)} \log N \rfloor$  and  $s < 18/19$ . It suffice to prove that there exists a constant  $C = C(A, \beta, s) > 0$  such that

$$\mathbb{E} [(W_{\beta,N})^{-s}] \leq C.$$

We have

$$\begin{aligned} & \mathbb{E} [(W_{\beta,N})^{-s}] \\ & \leq \varepsilon_0^{-s} \mathbb{P}(W_{\beta,N} > \varepsilon_0) + \sum_{k=1}^{K-1} (\varepsilon_{k+1})^{-s} \cdot \mathbb{P}(\varepsilon_{k+1} \leq W_{\beta,N} \leq \varepsilon_k) + \mathbb{E} [(W_{\beta,N})^{-s} \mathbf{1}_{W_{\beta,N} \leq \varepsilon_K}] \\ & \leq 2^s + \sum_{n=1}^{K-1} (\varepsilon_{k+1})^{-s} \cdot \mathbb{P}(W_{\beta,N} \leq \varepsilon_k) + \mathbb{E} [(W_{\beta,N})^{-s} \mathbf{1}_{W_{\beta,N} \leq \varepsilon_K}]. \end{aligned} \quad (4.4.1)$$

By Proposition 4.3.1, there exist  $N_0 = N(A, \beta, s) \in \mathbb{N}$  and  $C_1 = C_1(A, \beta, s) > 0$  such that for all  $N \geq N_0$  the first term of (4.4.1) is bounded by  $C_1$ . Thus, it remains to bound the second term of (4.4.1).

For all  $|u| = N$ ,

$$Z_{\beta,N} \geq e^{\beta X_u}. \quad (4.4.2)$$

Fix a vertex  $u$  with  $|u| = N$ . By (4.4.2) and exponential tilting, the second term of (4.4.1) is bounded from above by

$$\begin{aligned} \mathbb{E} [(W_{\beta,N})^{-s} \mathbf{1}_{W_{\beta,N} \leq \varepsilon_K}] & \leq \mathbb{E} [Z_{\beta,N}]^s \cdot \mathbb{E} \left[ (e^{\beta X_u})^{-s} \mathbf{1}_{e^{\beta X_u} \leq \varepsilon_K} \mathbb{E}[Z_{\beta,N}] \right] \\ & = \mathbb{E} [Z_{\beta,N}]^s \cdot \mathbb{E} \left[ e^{-\beta s X_u} \mathbf{1}_{X_u \leq \frac{1}{\beta} (\log \varepsilon_K + \log \mathbb{E}[Z_{\beta,N}])} \right] \\ & = \mathbb{E} [Z_{\beta,N}]^s \cdot \mathbb{E} \left[ e^{-\beta s X_u} \mathbf{1}_{X_u \leq \frac{1}{\beta} (\log \varepsilon_K + N \log 2 + N \frac{\beta^2}{2})} \right] \\ & = \mathbb{E} [Z_{\beta,N}]^s \cdot \exp \left( -\frac{\beta^2 s}{2} N \right) \\ & \quad \cdot \int_{-\infty}^{\frac{1}{\beta N} (\log \varepsilon_K + N \log 2 + N \frac{\beta^2}{2}) - \beta s} e^{-x^2/2} dx \\ & = \mathbb{E} [Z_{\beta,N}]^s \cdot \exp \left( -\frac{\beta^2 s}{2} N \right) \\ & \quad \cdot \int_{-\frac{1}{\beta N} (\log \varepsilon_K + N \log 2 + N \frac{\beta^2}{2}) + \beta s}^{\infty} e^{-x^2/2} dx \end{aligned} \quad (4.4.3)$$

Recall the definition of  $(\varepsilon_k)_{k=0}^K$  in the proof of Proposition 4.3.1, we have

$$\begin{aligned}
 \varepsilon_K &= \frac{1}{2} \left( \frac{1}{C^K} \prod_{n=0}^{K-1} (\eta_n^{\frac{19}{20}} - \eta_n) \right)^{\frac{17}{19} \frac{1}{s}} \\
 &\leq \frac{1}{2} \left( \frac{1}{C^K} \prod_{n=0}^{K-1} \eta_n^{\frac{19}{20}} \right)^{\frac{17}{19} \frac{1}{s}} = \frac{1}{2} \left( \frac{1}{C^K} \prod_{n=0}^{K-1} \eta_0^{\frac{19}{18} \left(\frac{19}{10}\right)^n} \right)^{\frac{17}{19} \frac{1}{s}} \\
 &= \frac{1}{2} \left( \frac{1}{C^K} \eta_0^{\frac{19}{18} \left(\left(\frac{19}{10}\right)^K - 1\right)} \right)^{\frac{17}{19} \frac{1}{s}} \\
 &= \frac{1}{2} \cdot \left( N^{-\frac{2 \log C}{\log(19/10)}} \cdot \eta_0^{\frac{19}{18} (N^2 - 1)} \right)^{\frac{17}{19} \frac{1}{s}}. \tag{Definition of  $K$ }
 \end{aligned} \tag{4.4.4}$$

Thus,

$$\begin{aligned}
 & - \frac{1}{\beta N} (\log \varepsilon_K + N \log 2 + N \frac{\beta^2}{2}) + \beta s \\
 & \geq - \frac{1}{\beta N} \left( -\log 2 - \frac{17}{19} \frac{1}{s} \frac{2 \log C_{\beta,1}}{\log(19/10)} \log N + \underbrace{\frac{17}{19} \frac{1}{s} \log \eta_0 \frac{19}{18} (N^2 - 1)}_{<0} + N \log 2 + N \frac{\beta^2}{2} \right) + \beta s \\
 & = \Theta(N).
 \end{aligned} \tag{4.4.5}$$

Thus, by (4.4.5) and the tail bound of the standard Gaussian distribution, (4.4.3) is bounded from above by

$$(4.4.3) \leq 2^s \int_{\Theta(N)}^{\infty} e^{-x^2/2} dx \leq \frac{1}{\Theta(N)} e^{-\Theta(N^2)}$$

which converges to 0 as  $N \rightarrow \infty$ . In particular, this implies that there exists a constant  $C_2 = C_2(A, \beta, s) > 0$ , independent of  $N$ , such that the second term of (4.4.1) is bounded from above by  $C_2$ . Therefore, we conclude that for all  $N \geq N_0$ ,

$$\mathbb{E} [(W_{\beta, N})^{-s}] \leq C_1 + C_2.$$

By taking  $C := C_1 + C_2$ , the proof is completed. ■

# 5. Outlook

In this chapter, we provide some further directions that might be interesting to investigate.

## 5.1 A renormalized BRW free energy via the Ruelle cascade

In this section, we consider a renormalized BRW free energy via the Poisson–Dirichlet cascade. This quantity arises naturally from computing the initial condition of the Hamilton–Jacobi equation for the enriched CREM free energy, inspired by the program initiated by Mourrat [146] in the context of the Sherrington–Kirkpatrick model. The goal of this section is to state the research problem of finding an explicit formula of this quantity. As a final remark, the section is extract from a on-going project with Alexander Alban and Justin Ko.

We start by introducing the relevant notation.

**Ulam–Harris tree.** For any integer  $r \geq 0$ , we denote by  $\mathcal{T}_r = \mathbb{N}^0 \cup \mathbb{N}^1 \cup \dots \cup \mathbb{N}^r$  the Ulam–Harris tree rooted at  $\emptyset$  of depth  $r$ , and we set the convention that  $\mathbb{N}^0 = \{\emptyset\}$ . For every  $\alpha \in \mathbb{N}^r$ , we denote by  $|\alpha|$  the depth of  $\alpha$ . For every vertex  $\alpha = n_1 \cdots n_k \in \mathcal{T}_r$  with  $k \in \llbracket 1, r \rrbracket$ , we define  $\alpha[0] = \emptyset$  and  $\alpha[m] = n_1 \cdots n_m$  for  $m \in \llbracket 1, k \rrbracket$ . For any  $\alpha^1, \alpha^2 \in \mathcal{T}_r$ , we denote by  $\alpha^1 \wedge \alpha^2$  the most recent common ancestor of  $\alpha^1$  and  $\alpha^2$ .

We now introduce the following double-indexed Gaussian process.

**Definition 5.1.1.** Let  $\mu$  be a discrete probability measure on the half line  $[0, \infty)$  defined as

$$\mu = \sum_{k=0}^r (\zeta_{k+1} - \zeta_k) \delta_{q_k}$$

which is encoded by the two sequences

$$0 = \zeta_0 < \zeta_1 \leq \zeta_2 \leq \dots \leq \zeta_{r-1} \leq \zeta_r < \zeta_{r+1} = 1$$

and

$$0 = q_{-1} \leq q_0 \leq \dots \leq q_{r-1} \leq q_r < q_{r+1} = \infty.$$

For  $\alpha \in \mathbb{N}^r$  and  $\sigma \in \{-1, +1\}^N$ , we define  $Y(\sigma, \alpha) := Y_\mu(\sigma, \alpha)$  to be a centered Gaussian process with covariance function

$$\mathbb{E}[Y_\mu(\sigma^1, \alpha^1) Y_\mu(\sigma^2, \alpha^2)] = q_{|\alpha^1 \wedge \alpha^2|} \cdot |\sigma^1 \wedge \sigma^2|.$$

*Remark 5.1.2.* The Gaussian process  $Y(\sigma, \alpha)$  can be constructed explicitly as follows

$$Y(\sigma, \alpha) = \sum_{i=1}^N \sum_{\ell=0}^r (q_\ell - q_{\ell-1})^{1/2} z_{\alpha[\ell], \sigma[i]}.$$

where  $z_{\alpha[\ell], \sigma[i]}$  are i.i.d. standard Gaussian random variables.

The goal of this section is to compute the free energy of this double-indexed model. However, as  $\alpha$  ranges over an countably infinite set  $\mathbb{N}^r$ , we need to introduce a finite measure on  $\mathbb{N}^r$  such that the summation makes sense. The measure we consider here is called the Ruelle probability cascades introduced by Ruelle in [165]. See also Figure 5.1 for an illustration of the definition.

**Definition 5.1.3** (Ruelle cascade). Fix an integer  $r \geq 0$ . Let sequences

$$0 = \zeta_0 < \zeta_1 \leq \zeta_2 \leq \dots \leq \zeta_{r-1} \leq \zeta_r < \zeta_{r+1} = 1.$$

For each  $|\alpha| = k$  with  $k \in \llbracket 0, r-1 \rrbracket$ , we associated  $\alpha$  with an independent Poisson point process with intensity measure  $\zeta_{k+1} x^{-1-\zeta_{k+1}} dx$ , and the points of the Poisson point process are arranged in decreasing order as

$$u_{\alpha 1} > u_{\alpha 2} > \dots.$$

The Ruelle probability cascade is a probability measure defined on  $\mathbb{N}^r$  with weight

$$v_\alpha = \frac{\prod_{k=1}^r u_{\alpha[k]}}{\sum_{\alpha' \in \mathbb{N}^r} \prod_{k=1}^r u_{\alpha'[k]}}, \quad \alpha \in \mathbb{N}^r.$$

When  $r = 0$ , then we simply define  $v_\alpha = 1$ .

*Remark 5.1.4.* The case where  $r = 1$  is called the Poisson–Dirichlet cascade in the literature.

We now state the research problem

**Problem 5.1.5.** For  $0 \leq q_1 \leq \dots \leq q_r < \infty$ , compute the free energy

$$F(q_1, \dots, q_r) := \lim_{N \rightarrow \infty} \mathbb{E} \left[ \log \sum_{|\sigma|=N} \sum_{|\alpha|=r} e^{Y(\sigma, \alpha)} v_\alpha \right].$$

### 5.1.1 Case $r = 1$ of Problem 5.1.5

Currently, we are able to compute the the free energy for  $r = 1$ . In this case, we can apply Lemma 2.2 from Panchenko’s book [156] to simplify the computation

**Lemma 5.1.6** (Lemma 2.2 in [156]). Let  $(v_n)_{n \in \mathbb{N}}$  be a Ruelle cascade with  $r = 1$  with parameter  $\zeta \in (0, 1)$ . Suppose that  $(Z_n)_{n \in \mathbb{N}}$  is a sequence of i.i.d. random variables with the same distribution as the random variable  $Z$ . If  $\mathbb{E}[Z^\zeta] < \infty$ , then

$$\mathbb{E} \left[ \log \sum_{n \in \mathbb{N}} v_n Z_n \right] = \frac{1}{\zeta} \log \mathbb{E} [Z^\zeta].$$

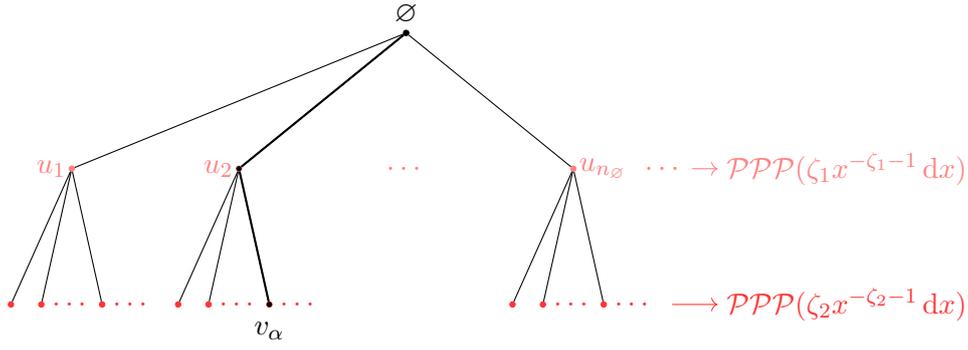


Figure 5.1: Schematic illustration of a Ruelle cascade with  $r = 2$ . The vertices at depth 1 are associated with weights  $u_1 > u_2 > \dots$ , sampled from a Poisson point process with an intensity measure given by  $\zeta_1 x^{-\zeta_1-1} dx$ . The children of each vertex  $n_\emptyset \in \mathbb{N}$  at depth 1 are associated with weights  $u_{n_\emptyset 1} > u_{n_\emptyset 2} > \dots$ , sampled from a Poisson process with an intensity measure given by  $\zeta_2 x^{-\zeta_2-1} dx$ . All the Poisson point processes are independent. A Ruelle cascade with  $r = 2$  is a probability measure defined on  $\mathbb{N}^2$ , such that the weight  $v_\alpha$  for each  $\alpha \in \mathbb{N}^2$  is the product of the Poissonian weights  $u_{\alpha[1]}$  and  $u_\alpha$  along the path from  $\emptyset$  to  $\alpha$ , normalized by the sum of these products over  $\mathbb{N}^2$ . See Definition 5.1.3 for a precise description of the Ruelle cascade.

By Lemma 2.2, Problem 5.1.5 is reduced to computing the  $\zeta$ -moments of the BRW partition function, with  $\zeta \in (0, 1)$ .

**Proposition 5.1.7.** *For  $q > 0$ , define*

$$Z_q := \sum_{|\sigma|=N} e^{\sqrt{q}X_\sigma}, \quad (5.1.1)$$

where  $(X(\sigma))_{\sigma \in \mathbb{T}}$  is the branching random walk with  $N(0, 1)$  increments constructed by  $X(\sigma) = \sum_{i=1}^N z_{\sigma[i]}$  and the set  $(z_{\sigma[i]})_{\sigma[i] \in \mathbb{T}_N \setminus \emptyset}$  is a collection of i.i.d.  $N(0, 1)$  random variables. Then, we have

$$F(q) = \lim_{N \rightarrow \infty} \frac{1}{\zeta N} \log \mathbb{E} \left[ (Z_q)^\zeta \right] = \begin{cases} \frac{q}{2} + \log 2, & q < 2 \log 2, \\ \sqrt{2 \log 2q}, & 2 \log 2 \leq q \leq \frac{2 \log 2}{\zeta^2}, \\ \frac{1}{2} \zeta q + \frac{\log 2}{\zeta}, & q > \frac{2 \log 2}{\zeta^2}. \end{cases}$$

*Proof.* We distinguish the two cases where  $q < 2 \log 2$  and  $q \geq 2 \log 2$ .

**Case 1:**  $q < 2 \log 2$ . The upper bound follows simply from the Jensen inequality:

$$\frac{1}{\zeta N} \log \mathbb{E} \left[ (Z_q)^\zeta \right] \leq \frac{1}{N} \log \mathbb{E} [Z_q] = \frac{q}{2} + \log 2.$$

It remains to show the lower bound. By the Jensen inequality, we have

$$\frac{1}{\zeta N} \log \mathbb{E} \left[ (Z_q)^\zeta \right] \geq \frac{1}{\zeta} \log \mathbb{E} \left[ (Z_q)^{\zeta/N} \right]. \quad (5.1.2)$$

Combining (5.1.2) and Fatou's lemma, we derive the lower bound

$$\frac{1}{\zeta N} \log \mathbb{E} \left[ (Z_q)^\zeta \right] \geq \liminf_{N \rightarrow \infty} \frac{1}{\zeta} \log \mathbb{E} \left[ (Z_q)^{\zeta/N} \right] \geq \frac{1}{\zeta} \log \mathbb{E} \left[ \liminf_{N \rightarrow \infty} (Z_q)^{\zeta/N} \right].$$

Thus, if we can show that for all  $q < 2 \log 2$ ,

$$\liminf_{N \rightarrow \infty} (Z_q)^{\zeta/N} = e^{\zeta(\frac{q}{2} + \log 2)} \quad \text{almost surely,}$$

the proof of Case 1 is completed. By the fact that  $Z_q \geq Z_q^\leq$  and the Paley–Zygmund inequality,

$$\begin{aligned} & \mathbb{P} \left( (Z_q)^{\zeta/N} \geq \frac{1}{2} \exp \left( \zeta \left( \frac{q}{2} + \log 2 \right) \right) \right) \\ & \geq \mathbb{P} \left( Z_q^\leq \geq \left( \frac{1}{2} \right)^{N/\zeta} \exp \left( N \left( \frac{q}{2} + \log 2 \right) \right) \right) \\ & \geq \frac{\left( \mathbb{E} [Z_q^\leq] - \left( \frac{1}{2} \right)^{N/\zeta} \exp \left( N \left( \frac{q}{2} + \log 2 \right) \right) \right)^2}{\mathbb{E} \left[ (Z_q^\leq)^2 \right]} \end{aligned} \quad (5.1.3)$$

Recall the additive martingale of the branching random walk is defined as

$$W_{q,N} := \frac{Z_q}{\exp \left( N \left( \frac{q}{2} + \log 2 \right) \right)}.$$

Moreover, for  $q < 2 \log 2$ ,  $(W_{q,N})$  is a positive martingale. Then, (5.1.3) implies that there exists a uniform constant  $c_1 > 0$  such that

$$\mathbb{P} \left( W_{q,N} \geq \frac{1}{2} \right) \geq c_1$$

for all  $N \in \mathbb{N}$ . Then by the zero–one law of the limit  $W_{q,\infty}$ , we conclude that

$$\mathbb{P} (W_{q,\infty} > 0) = 1. \quad (5.1.4)$$

Therefore, by (5.1.4), for almost all configuration  $\omega$ , there exists  $N_0(\omega)$  and  $c_2(\omega) > 0$  such that

$$W_{q,N}(\omega) \geq c_2(\omega)$$

for all  $N \geq N_0(\omega)$ . Thus, we have

$$\liminf_{N \rightarrow \infty} (W_{q,N}(\omega))^{\zeta/N} \geq \liminf_{N \rightarrow \infty} c_2(\omega)^{1/N} \geq 1.$$

Therefore, we conclude that

$$\liminf_{N \rightarrow \infty} (Z_q)^{\zeta/N} \geq \exp \left( \zeta \left( \frac{q}{2} + \log 2 \right) \right) \quad \text{almost surely,}$$

and this completes the proof of Case 1.

**Case 2:**  $q \geq 2 \log 2$ . We start with the upper bound. For all  $\varepsilon > 0$ , choose  $B > 0$  such that

$$e^{(-\sqrt{q} + \sqrt{2 \log 2})B} < \varepsilon.$$

Define the truncated partition function

$$\begin{aligned} Z_q^> &:= \sum_{|\sigma|=N} e^{\sqrt{q}X(\sigma)} \mathbf{1}\{X(\sigma) > \sqrt{2 \log 2}N - B\}, \\ Z_q^{\leq} &:= \sum_{|\sigma|=N} e^{\sqrt{q}X(\sigma)} \mathbf{1}\{X(\sigma) \leq \sqrt{2 \log 2}N - B\}. \end{aligned}$$

By the subadditivity of  $x \mapsto x^\zeta$ , we have

$$\begin{aligned} \mathbb{E} \left[ (Z_q)^\zeta \right] &= \mathbb{E} \left[ (Z_q^> + Z_q^{\leq})^\zeta \right] \\ &\leq \mathbb{E} \left[ (Z_q^>)^\zeta \right] + \mathbb{E} \left[ (Z_q^{\leq})^\zeta \right] \quad (\text{Subadditivity}) \\ &\leq \mathbb{E} \left[ (Z_q^>)^\zeta \right] + \mathbb{E} \left[ (Z_q^{\leq})^\zeta \right]. \quad (\text{Jensen's inequality}) \end{aligned} \quad (5.1.5)$$

We first bound the second term of (5.1.5). By exponential tilting and the Chernoff bound, we have

$$\begin{aligned} \mathbb{E} \left[ Z_q^{\leq} \right] &= \mathbb{E} \left[ \sum_{|\sigma|=N} e^{\sqrt{q}X_\sigma} \mathbf{1}_{X_\sigma \leq \sqrt{2 \log 2}N - B} \right] \\ &= 2^N \mathbb{E} \left[ e^{\beta X_\sigma} \mathbf{1}_{X_\sigma \leq \sqrt{2 \log 2}N - B} \right] \\ &= 2^N e^{qN/2} \mathbb{P} \left( X_\sigma < \sqrt{2 \log 2}N - \sqrt{q}N - B \right) \\ &\leq 2^N e^{qN/2} \exp \left( -\frac{(\sqrt{q}N - \sqrt{2 \log 2}N + B)^2}{2N} \right) \quad (5.1.6) \\ &= 2^N e^{qN/2} \exp \left( -\frac{qN}{2} \right) \exp \left( \sqrt{q}(\sqrt{2 \log 2}N - B) \right) \exp \left( -\frac{(\sqrt{2 \log 2}N - B)^2}{2N} \right) \\ &= 2^N e^{\sqrt{2 \log 2}qN - \sqrt{q}B} \exp \left( -\frac{(\sqrt{2 \log 2}N - B)^2}{2N} \right) \\ &= 2^N e^{\sqrt{2 \log 2}qN - \sqrt{q}B} \exp \left( -\frac{2 \log 2N}{2} \right) \exp \left( \sqrt{2 \log 2}B \right) \underbrace{\exp \left( -\frac{B^2}{2N} \right)}_{\leq 1} \\ &= \exp \left( \sqrt{2 \log 2}qN \right) \underbrace{\exp \left( (-\sqrt{q} + \sqrt{2 \log 2})B \right)}_{< \varepsilon} \\ &\leq \varepsilon \exp \left( \sqrt{2 \log 2}qN \right), \quad (5.1.7) \end{aligned}$$

where (5.1.6) follows from the assumption that  $q \geq 2 \log 2$ .

It remains to compute the first term of (5.1.5). Again, by the subadditivity of

$x \mapsto x^\zeta$ , the many-to-one formula and exponential tilting, we have

$$\begin{aligned} \mathbb{E} \left[ (Z_q^\geq)^\zeta \right] &\leq \mathbb{E} \left[ \sum_{|\sigma|=N} e^{\zeta \sqrt{q} X(\sigma)} \mathbf{1}_{\{X(\sigma) > \sqrt{2 \log 2N} - B\}} \right] \\ &= 2^N \mathbb{E} \left[ e^{\zeta \sqrt{q} X(\sigma)} \mathbf{1}_{\{X(\sigma) > \sqrt{2 \log 2N} - B\}} \right] \\ &= 2^N e^{\zeta^2 q N/2} \mathbb{P} \left( X(\sigma) > \sqrt{2 \log 2N} - \zeta \sqrt{q} N - B \right). \end{aligned} \quad (5.1.8)$$

We distinguish the following two cases.

- If  $q \leq 2 \log 2 / \zeta^2$ , then by (5.1.8) and the Chernoff bound,

$$\begin{aligned} \mathbb{E} \left[ (Z_q^\geq)^\zeta \right] &\leq 2^N e^{\zeta^2 q N/2} \exp \left( - \frac{((\sqrt{2 \log 2} - \zeta \sqrt{q})N - B)^2}{2N} \right) \\ &\leq 2^N e^{\zeta^2 q N/2} \exp \left( - \frac{(\sqrt{2 \log 2} - \zeta \sqrt{q})^2 N}{2} \right) \\ &\quad \cdot \underbrace{\exp \left( (\sqrt{2 \log 2} - \zeta \sqrt{q}) B \right)}_{\leq 1} \underbrace{\exp \left( - \frac{B^2}{2N} \right)}_{\leq 1} \\ &= \exp \left( \zeta \sqrt{2 \log 2q} N \right). \end{aligned} \quad (5.1.9)$$

Combining (5.1.5), (5.1.7), (5.1.8) and (5.1.9), we conclude that

$$\begin{aligned} \frac{1}{\zeta N} \log \mathbb{E} \left[ (Z_q)^\zeta \right] &\leq \frac{1}{\zeta N} \log \left( (1 + \varepsilon^\zeta) e^{\zeta \sqrt{2 \log 2q} N} \right) \\ &= \sqrt{2 \log 2q} + \frac{1}{\zeta N} \log \left( 1 + \varepsilon^\zeta \right) \\ &\rightarrow \sqrt{2 \log 2q}, \quad N \rightarrow \infty. \end{aligned}$$

- If  $q > 2 \log 2 / \zeta^2$ , then we simply bound the probability in (5.1.8) by 1 and obtain

$$\mathbb{E} \left[ (Z_q^\geq)^\zeta \right] \leq 2^N e^{\zeta^2 q N/2}. \quad (5.1.10)$$

Combining (5.1.5), (5.1.7) and (5.1.10), we obtain

$$\begin{aligned} \frac{1}{\zeta N} \log \mathbb{E} \left[ (Z_q)^\zeta \right] &\leq \frac{1}{\zeta N} \log \left( \varepsilon^\zeta e^{\zeta \sqrt{2 \log 2q} N} + 2^N e^{\zeta^2 q N/2} \right) \\ &= \frac{1}{\zeta N} \log \left( e^{\zeta^2 q N/2 + \log 2N} (1 + \varepsilon^\zeta e^{\zeta \sqrt{2 \log 2q} N - \zeta^2 q N/2 - \log 2N}) \right) \\ &\rightarrow \frac{1}{2} \zeta q + \frac{\log 2}{\zeta}, \end{aligned}$$

where the last line above follows from the assumption that  $q > 2 \log 2 / \zeta^2$ , and thus

$$e^{\zeta \sqrt{2 \log 2q} N - \zeta^2 q N/2 - \log 2N} = e^{-\frac{N}{2} (\zeta \sqrt{q} - \sqrt{2 \log 2})^2} \rightarrow 0, \quad N \rightarrow \infty. \quad (5.1.11)$$

It remains to show the lower bound. Let  $M_N = \max_{|\sigma|=N} X_\sigma$ . By bounding the partition function by one single particle, we have

$$\mathbb{E} \left[ (Z_q)^\zeta \right] \geq \mathbb{E} \left[ e^{\zeta q^{1/2} M_N} \right]. \quad (5.1.12)$$

Then, Varadhan's Lemma (see Theorem 4.3.1 in Dembo and Zeitouni's book [84]) states that

$$\lim_{N \rightarrow \infty} \frac{1}{\zeta N} \log \mathbb{E} \left[ e^{\zeta q^{1/2} M_N} \right] = \sup_{c \in \mathbb{R}} \{ q^{1/2} c - h(c)/\zeta \}, \quad (5.1.13)$$

where  $h(c)$  is the rate function of  $M_N$  defined as

$$h(c) := \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} (M_N > cN).$$

Applying Theorem 3.2 in Gantert and Höfelsauer [105] gives the formula of the rate function

$$h(c) = \frac{1}{2} (c^2 - 2 \log 2)_+.$$

See also the work [66] by Chauvin and Rouault on similar results for the branching Brownian motion.

We distinguish the two cases.

- If  $q < 2 \log 2 / \zeta^2$ , then the maximum of (5.1.13) is achieved at  $\sqrt{2 \log 2}$  with value  $\sqrt{2 \log 2} q$ .
- If  $q \geq 2 \log 2 / \zeta^2$ , then the maximum of (5.1.13) is achieved at  $\zeta \sqrt{q}$  with the value  $\log 2 / \zeta + \frac{1}{2} \zeta q$ .

Since the lower bound matches the upper bound for the both cases, the proof is now completed. ■

### 5.1.2 Case $r > 1$ of Problem 5.1.5

We also compute the limiting free energy  $F(q_1, q_2)$  for  $r = 2$ , while the proof is omitted. See Figure 5.1.2 for the phase diagram.

**Proposition 5.1.8.** *In the case  $r = 2$ , the limiting free energy  $F(q_1, q_2)$  has the*

following description:

$$F(q_1, q_2) = \begin{cases} \frac{q_2}{2} + \log 2, & \text{if } q_1 < \frac{2 \log 2}{\zeta_1^2} \text{ and } q_2 < 2 \log 2, \\ \sqrt{2 \log 2 q_2}, & \text{if } q_1 < \frac{2 \log 2}{\zeta_1^2} \text{ and } q_2 \in \left[ 2 \log 2, \frac{2 \log 2}{\zeta_1^2} \right], \\ \frac{1}{2} \zeta_1 q_2 + \frac{\log 2}{\zeta_1}, & \text{if } q_1 < \frac{2 \log 2}{\zeta_1^2} \text{ and } q_2 > \frac{2 \log 2}{\zeta_1^2}, \\ \sqrt{2 \log 2 q_1} + \frac{q_2 - q_1}{2} \zeta_1, & \text{if } q_1 \in \left[ \frac{2 \log 2}{\zeta_1^2}, \frac{2 \log 2}{\zeta_0^2} \right], \\ \frac{1}{2} \zeta_0 q_1 + \frac{\log 2}{\zeta_0} + \frac{q_2 - q_1}{2} \zeta_1, & \text{if } q_1 > \frac{2 \log 2}{\zeta_0^2}. \end{cases}$$

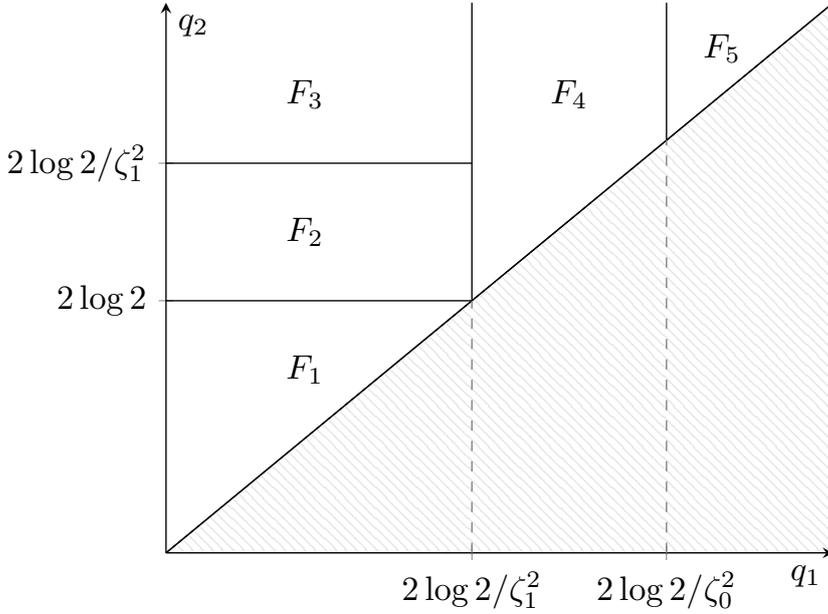


Figure 5.2: Phase diagram of  $F(q_1, q_2)$ . In the figure,  $F_1 = \frac{q_2}{2} + \log 2$ ,  $F_2 = \sqrt{2 \log 2 q_2}$ ,  $F_3 = \frac{1}{2} \zeta_1 q_2 + \frac{\log 2}{\zeta_1}$ ,  $F_4 = \sqrt{2 \log 2 q_1} + \frac{q_2 - q_1}{2} \zeta_1$  and  $F_5 = \frac{1}{2} \zeta_0 q_1 + \frac{\log 2}{\zeta_0} + \frac{q_2 - q_1}{2} \zeta_1$ .

## 5.2 Fluctuation of the CREM free energy

The limiting free energy of the continuous random energy model has an explicit formula given by Bovier and Kurkova [54], while the authors were not able to show fluctuation results as in their previous work [53] on the GREM. As a first step toward characterizing the fluctuation, we propose the following problem.

**Problem 5.2.1.** For all  $\beta > 0$  and  $p \geq 1$ , study the asymptotics of  $\|\log Z_{\beta,N} - \mathbb{E}[\log Z_{\beta,N}]\|_p$ .

The work in Chapter 4 implies that for all  $\beta < \beta_c$ , for all  $p \geq 1$ , there exists a constant  $C = C(A, \beta, p) > 0$  such that

$$\|\log Z_{\beta,N} - \mathbb{E}[\log Z_{\beta,N}]\|_p \leq C.$$

This implies that in the high temperature regime  $\beta < \beta_c$ , the fluctuation of the unnormalized free energy  $\log Z_{\beta,N}$  is of  $O(1)$ .

To tackle Problem 5.2.1 for  $\beta$  beyond the high temperature regime, we argue to it is sufficient to study the fluctuation of  $\log Z_{\beta,N}$  when  $\beta > \sqrt{2 \log 2 / \hat{a}(1)}$ . Let  $N_0 = \lfloor t_0 N \rfloor$ . We have the following decomposition of the  $L^p$  norm

$$\begin{aligned} & \|\log Z_{\beta,N} - \mathbb{E}[\log Z_{\beta,N}]\|_p \\ & \leq \|\log_{\beta,N_0} - \mathbb{E}[\log_{\beta,N_0}]\|_p \\ & + \left\| \log \sum_{|u|=N_0} \mu_{\beta,N_0}(u) Z_{\beta,N-N_0}^u - \mathbb{E} \left[ \log \sum_{|u|=N_0} \mu_{\beta,N_0}(u) Z_{\beta,N-N_0}^u \right] \right\|_p. \end{aligned} \quad (5.2.1)$$

We argue that the second term of (5.2.1) can be reduced to the high temperature regime. First, for all  $x > 0$  and  $s > 0$ , there exists a constant  $C = C(p, s) > 0$  such that

$$|\log x|^p \leq C(|x| + |x|^{-s}).$$

Therefore,

$$\begin{aligned} & \mathbb{E} \left[ \left\| \log \sum_{|u|=N_0} \mu_{\beta,N_0}(u) Z_{\beta,N-N_0}^u - \mathbb{E} \left[ \log \sum_{|u|=N_0} \mu_{\beta,N_0}(u) Z_{\beta,N-N_0}^u \right] \right\|_p^p \right] \\ & \leq C \mathbb{E} \left[ \frac{\sum_{|u|=N_0} \mu_{\beta,N_0}(u) Z_{\beta,N-N_0}^u}{\exp\left(\mathbb{E} \left[ \log \sum_{|u|=N_0} \mu_{\beta,N_0}(u) Z_{\beta,N-N_0}^u \right]\right)} \right] \\ & + C \mathbb{E} \left[ \frac{\left(\sum_{|u|=N_0} \mu_{\beta,N_0}(u) Z_{\beta,N-N_0}^u\right)^{-s}}{\exp\left(-s \mathbb{E} \left[ \log \sum_{|u|=N_0} \mu_{\beta,N_0}(u) Z_{\beta,N-N_0}^u \right]\right)} \right]. \end{aligned} \quad (5.2.2)$$

By the branching property and the Jensen inequality, the first term of (5.2.2) is

bounded from above by

$$\begin{aligned}
 & \mathbb{E} \left[ \frac{\sum_{|u|=N_0} \mu_{\beta, N_0}(u) Z_{\beta, N-N_0}^u}{\exp\left(\mathbb{E} \left[ \log \sum_{|u|=N_0} \mu_{\beta, N_0}(u) Z_{\beta, N-N_0}^u \right]\right)} \right] \\
 &= \frac{\mathbb{E} \left[ Z_{\beta, N-N_0}^{(N_0)} \right]}{\exp\left(\mathbb{E} \left[ \log \sum_{|u|=N_0} \mu_{\beta, N_0}(u) Z_{\beta, N-N_0}^u \right]\right)} \\
 &\leq \frac{\mathbb{E} \left[ Z_{\beta, N-N_0}^{(N_0)} \right]}{\exp\left(\mathbb{E} \left[ \sum_{|u|=N_0} \mu_{\beta, N_0}(u) \log Z_{\beta, N-N_0}^u \right]\right)} \\
 &= \frac{\mathbb{E} \left[ Z_{\beta, N-N_0}^{(N_0)} \right]}{\exp\left(\mathbb{E} \left[ \log Z_{\beta, N-N_0}^{(N_0)} \right]\right)}. \tag{5.2.3}
 \end{aligned}$$

If we are able to show that

$$\exp\left(\mathbb{E} \left[ \log Z_{\beta, N-N_0}^{(N_0)} \right]\right) = \mathbb{E} \left[ Z_{\beta, N-N_0}^{(N_0)} \right] (1 + o(1)) = 2^N \exp\left(\frac{\beta^2 N}{2}\right) (1 + o(1)),$$

we should be able to show that (5.2.3) converges to 1 as  $N \rightarrow \infty$ , which implies in particular that (5.2.3) is uniformly bounded in  $N$ .

On the other hand, by the branching property and the Jensen inequality, the second term of (5.2.2) is bounded from above by

$$\begin{aligned}
 & \mathbb{E} \left[ \frac{\left(\sum_{|u|=N_0} \mu_{\beta, N_0}(u) Z_{\beta, N-N_0}^u\right)^{-s}}{\exp\left(-s \mathbb{E} \left[ \log \sum_{|u|=N_0} \mu_{\beta, N_0}(u) Z_{\beta, N-N_0}^u \right]\right)} \right] \\
 &\leq \frac{\mathbb{E} \left[ (Z_{\beta, N-N_0})^{-s} \right]}{\exp\left(-s \mathbb{E} \left[ \log Z_{\beta, N-N_0}^{(N_0)} \right]\right)} \\
 &\leq \frac{\mathbb{E} \left[ (Z_{\beta, N-N_0}^{(N_0)})^{-s} \right]}{\mathbb{E} \left[ Z_{\beta, N-N_0}^{(N_0)} \right]^{-s}}. \tag{5.2.4}
 \end{aligned}$$

The Gaussian process  $(X_u^{(N_0)})_{|u|=N-N_0}$  has zero mean and covariance function

$$\mathbb{E} \left[ X_u^{(N_0)} X_w^{(N_0)} \right] = N \left( A \left( \frac{|u \wedge w| + N_0}{N} \right) - A \left( \frac{N_0}{N} \right) \right).$$

By adapting the result in Chapter 4, we should be able to show that there exists a constant  $C_2 = C_2(A, \beta, s) > 0$  such that

$$\frac{\mathbb{E} \left[ (Z_{\beta, N-N_0}^{(N_0)})^{-s} \right]}{\mathbb{E} \left[ Z_{\beta, N-N_0}^{(N_0)} \right]^{-s}} \leq C_2.$$

Controlling the first term in (5.2.1) continues to pose a challenge. Although the asymptotics of the maximum of the CREM can be deduced from Mallein's work

[134] (also refer to Maillard and Zeitouni’s corresponding continuum result [132] for the branching Brownian motion with inhomogeneous variance), it remains uncertain whether the  $o(N)$  asymptotics of  $\log Z_{\beta,N}$  aligns with that of the maximum. Through the use of a Gaussian comparison argument, this issue could potentially be reduced to studying the asymptotics of  $\log Z_{\beta,N}$  when  $A$  is concave. As an initial step, the following problem should be investigated:

**Problem 5.2.2.** *Let  $A$  be a concave function. For all  $\beta > 0$ , study the asymptotics of  $\text{VAR}[Z_{\beta,N}]$ .*

The techniques employed to solve Problem 5.2.2 should also be applicable for deriving similar estimates for the  $p$ -norm.

To conclude this section, we mention that in the spin glass literature, there is a method that was developed by Chatterjee [61] to estimate the variance of the unnormalized free energy  $\log Z_{\beta,N}$ . The idea is to apply Gaussian integration by parts to represent the variance of  $\log Z_{\beta,N}$  as an integral involving the “overlap” of two spins. These spins are drawn from the original Gibbs measure and the perturbed Gibbs measure. More precisely, it can be shown that

$$\text{VAR}[\log Z_{\beta,N}] = \beta^2 N \int_0^1 \mathbb{E} \left[ \sum_{|u_1|=N} \sum_{|u_2|=N} \frac{e^{\beta X_{u_1}}}{Z_{\beta,N}} \frac{e^{\beta X_{u_2}^t}}{Z_{\beta,N}^t} A\left(\frac{|u_1 \wedge u_2|}{N}\right) \right] dt$$

Then, if one can demonstrate that the underlying model exhibits what is known as *disorder chaos*, it can be shown that  $\text{VAR}[\log Z_{\beta,N}] = o(N)$ . Very recently, the concept of disorder chaos was extended to the context of first-passage/last-passage percolation by Ahlberg, Deijfen and Sfragata [5, 4].

# A. Errata

The material in Chapter 2 mostly follows the published paper [112]. During the course of writing the thesis, I have discovered a few typos, and I list them below.

- The original title of the published version was *Efficient approximation of branching random walk Gibbs measures*. Although not technically an error, I have adjusted it to *Efficient sampling of the CREM Gibbs measure* for better alignment with the title of Chapter 3 and to emphasize our aim of efficiently sampling the model’s Gibbs measure.
- Line 3, Paragraph 2 of the abstract of the published version: “linear-time” should be replaced by “polynomial-time”. That sentence referred to Corollary 2.1.8, where I showed that by choosing a sequence  $M_N = O(\log N)$ , the algorithm can be of polynomial-time. But in fact, as presented in the proof of Corollary 1.5.4, for all  $\varepsilon > 0$ , I am able to modify the choice of the sequence  $M_N$  such that the algorithm is of  $O(N^{1+\varepsilon})$  time. I decide to keep the statement and the proof Corollary 2.1.8 as it was, so that the readers can also see both the modified version in Corollary 1.5.4 and the original version in Corollary 2.1.8.
- The same correction from “linear-time” to “polynomial-time” applies to Line 6 of the introductory text of Section 1.
- Line 3, Section 2.1.1: the index set should be  $\{0, \dots, d - 1\}^n$  instead of  $\{1, \dots, d\}$ .
- Remark 1.3 in the published version incorrectly states that the running time of the algorithm should be bounded by  $\lfloor N/M \rfloor 2^M$ , when it should actually be  $\lfloor N/M \rfloor d^M$ . This error stems from the fact that the branching random walk we studied has  $d$  offspring in each generation, not two.
- Reference 8 in the published version inaccurately cites a paper [29], authored by Bauerschmidt and Bodineau. The correct reference should be [28], a paper also authored by Bauerschmidt and Bodineau.

# Bibliography

- [1] Dimitris Achlioptas and Federico Ricci-Tersenghi. “On the Solution-Space Geometry of Random Constraint Satisfaction Problems”. In: ().
- [2] Louigi Addario-Berry and Pascal Maillard. “The Algorithmic Hardness Threshold for Continuous Random Energy Models”. In: *Mathematical Statistics and Learning* 2 (Feb. 25, 2020), pp. 77–101.
- [3] Louigi Addario-Berry and Bruce Reed. “Minima in Branching Random Walks”. In: *Annals of Probability* 37 (May 2009), pp. 1044–1079.
- [4] Daniel Ahlberg, Maria Deijfen, and Matteo Sfragara. *Chaos, Concentration and Multiple Valleys in First-Passage Percolation*. Feb. 22, 2023. URL: <http://arxiv.org/abs/2302.11367> (visited on 07/05/2023). preprint.
- [5] Daniel Ahlberg, Maria Deijfen, and Matteo Sfragara. *From Stability to Chaos in Last-Passage Percolation*. Feb. 22, 2023. URL: <http://arxiv.org/abs/2302.11379> (visited on 07/05/2023). preprint.
- [6] Elie Aïdékon. “Convergence in Law of the Minimum of a Branching Random Walk”. In: *Annals of Probability* 41 (3A May 2013), pp. 1362–1426.
- [7] Elie Aidekon and Zhan Shi. “The Seneta–Heyde Scaling for the Branching Random Walk”. In: *The Annals of Probability* 42 (May 2014), pp. 959–993.
- [8] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. “Optimization of Mean-Field Spin Glasses”. In: *The Annals of Probability* 49 (Nov. 2021), pp. 2922–2960.
- [9] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. “Sampling from the Sherrington-Kirkpatrick Gibbs Measure via Algorithmic Stochastic Localization”. Mar. 9, 2022.
- [10] Tom Alberts and Marcel Ortgiese. “The Near-Critical Scaling Window for Directed Polymers on Disordered Trees”. In: *Electronic Journal of Probability* 18 (none Jan. 2013), pp. 1–24.
- [11] David Aldous. “Greedy Search on the Binary Tree with Random Edge-Weights”. In: *Combinatorics, Probability and Computing* 1 (Dec. 1992), pp. 281–293.
- [12] J. R. L. de Almeida and D. J. Thouless. “Stability of the Sherrington-Kirkpatrick Solution of a Spin Glass Model”. In: *Journal of Physics A: Mathematical and General* 11 (May 1978), p. 983.

- [13] Nima Anari et al. *Entropic Independence I: Modified Log-Sobolev Inequalities for Fractionally Log-Concave Distributions and High-Temperature Ising Models*. Nov. 4, 2021. (Visited on 07/30/2023). preprint.
- [14] P. W. Anderson. “Localisation Theory and the Cu-Mn Problem: Spin Glasses”. In: *Materials Research Bulletin* 5 (Aug. 1, 1970), pp. 549–554.
- [15] Fabrizio Antenucci et al. “Glassy Nature of the Hard Phase in Inference Problems”. In: *Physical Review X* 9 (Jan. 31, 2019), p. 011020.
- [16] Louis-Pierre Arguin and Jack Hanson. “On Absence of Disorder Chaos for Spin Glasses on  $\mathbb{Z}^d$ ”. In: *Electronic Communications in Probability* 25 (none Jan. 2020), pp. 1–12.
- [17] Gérard Ben Arous, Eliran Subag, and Ofer Zeitouni. “Geometry and Temperature Chaos in Mixed Spherical Spin Glasses at Low Temperature: The Perturbative Regime”. In: *Communications on Pure and Applied Mathematics* 73 (2020), pp. 1732–1828.
- [18] Antonio Auffinger, Gérard Ben Arous, and Jiří Černý. “Random Matrices and Complexity of Spin Glasses”. In: *Communications on Pure and Applied Mathematics* 66 (2013), pp. 165–201.
- [19] Antonio Auffinger and Wei Kuo Chen. “On Concentration Properties of Disordered Hamiltonians”. In: *Proceedings of the American Mathematical Society* 146 (2018), pp. 1807–1815.
- [20] Antonio Auffinger and Wei-Kuo Chen. “On the Energy Landscape of Spherical Spin Glasses”. In: *Advances in Mathematics* 330 (May 25, 2018), pp. 553–588.
- [21] Antonio Auffinger and Wei-Kuo Chen. “The Parisi Formula Has a Unique Minimizer”. In: *Communications in Mathematical Physics* 335 (May 1, 2015), pp. 1429–1444.
- [22] Antonio Auffinger, Wei-Kuo Chen, and Qiang Zeng. “The SK Model Is Infinite Step Replica Symmetry Breaking at Zero Temperature”. In: *Communications on Pure and Applied Mathematics* 73 (May 2020), pp. 921–943.
- [23] Antonio Auffinger and Aukosh Jagannath. “Thouless–Anderson–Palmer Equations for Generic  $\mathbb{S}$ -Spin Glasses”. In: *The Annals of Probability* 47 (July 2019), pp. 2230–2256.
- [24] Nicolas Bacaër. *A Short History of Mathematical Population Dynamics*. London: Springer, 2011.
- [25] Julien Barral and Benoît B. Mandelbrot. “Introduction to Infinite Products of Independent Random Functions (Random Multiplicative Multifractal Measures. I)”. In: *Fractal geometry and applications: A jubilee of Benoit Mandelbrot, Part 2. Proc. Sympos. Pure Math* 72 (2004), pp. 3–16.
- [26] Julien Barral and Benoît B. Mandelbrot. “Multifractal Products of Cylindrical Pulses”. In: *Probability Theory and Related Fields* 124 (Nov. 1, 2002), pp. 409–430.

- [27] Julien Barral, Rémi Rhodes, and Vincent Vargas. *Limiting Laws of Super-critical Branching Random Walks*. May 27, 2012. URL: <http://arxiv.org/abs/1203.5445> (visited on 07/24/2023). preprint.
- [28] Roland Bauerschmidt and Thierry Bodineau. “A Very Simple Proof of the LSI for High Temperature Spin Systems”. In: *Journal of Functional Analysis* 276 (Apr. 15, 2019), pp. 2582–2588.
- [29] Roland Bauerschmidt and Thierry Bodineau. “Spectral Gap Critical Exponent for Glauber Dynamics of Hierarchical Spin Models”. In: *Communications in Mathematical Physics* 373 (Feb. 2020), pp. 1167–1206.
- [30] David Belius. *High Temperature TAP Upper Bound for the Free Energy of Mean Field Spin Glasses*. Apr. 1, 2022. (Visited on 04/10/2023). preprint.
- [31] David Belius, Leon Fröber, and Justin Ko. *TAP Variational Principle for the Constrained Overlap Multiple Spherical Sherrington-Kirkpatrick Model*. Apr. 8, 2023. (Visited on 07/31/2023). preprint.
- [32] David Belius and Nicola Kistler. “The TAP–Plefka Variational Principle for the Spherical SK Model”. In: *Communications in Mathematical Physics* 367 (May 1, 2019), pp. 991–1017.
- [33] Gérard Ben Arous, A. Dembo, and A. Guionnet. “Aging of Spherical Spin Glasses”. In: *Probability Theory and Related Fields* 120 (May 1, 2001), pp. 1–67.
- [34] Gérard Ben Arous and Alice Guionnet. “Langevin Dynamics for Sherrington-Kirkpatrick Spin Glasses”. In: *Mathematical Aspects of Spin Glasses and Neural Networks*. Ed. by Anton Bovier and Pierre Picco. Progress in Probability. Boston, MA: Birkhäuser, 1998, pp. 323–353.
- [35] Gérard Ben Arous and Alice Guionnet. “Large Deviations for Langevin Spin Glass Dynamics”. In: *Probability Theory and Related Fields* 102 (Dec. 1, 1995), pp. 455–509.
- [36] Gérard Ben Arous and Alice Guionnet. “Symmetric Langevin Spin Glass Dynamics”. In: *The Annals of Probability* 25 (1997), pp. 1367–1422.
- [37] Gérard Ben Arous and Aukosh Jagannath. “Spectral Gap Estimates in Mean Field Spin Glasses”. In: *Communications in Mathematical Physics* 361 (July 2018), pp. 1–52.
- [38] Itai Benjamini and Oded Schramm. “KPZ in One Dimensional Random Geometry of Multiplicative Cascades”. In: *Communications in Mathematical Physics* 289 (July 1, 2009), pp. 653–662.
- [39] Jean Bérard and Jean-Baptiste Gouéré. “Survival Probability of the Branching Random Walk Killed Below a Linear Boundary”. In: *Electronic Journal of Probability* 16 (none Jan. 2011), pp. 396–418.
- [40] Julien Berestycki. *Topics on Branching Brownian Motion*. 2014.

- [41] Jean Bertoin. “Subordinators: Examples and Applications”. In: *Lectures on Probability Theory and Statistics: Ecole d’Eté de Probabilités de Saint-Flour XXVII - 1997*. Ed. by Jean Bertoin et al. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer, 1999, pp. 1–91.
- [42] Jean Bertoin and Jean-François Le Gall. “The Bolthausen–Sznitman Coalescent and the Genealogy of Continuous-State Branching Processes”. In: *Probability Theory and Related Fields* 117 (June 1, 2000), pp. 249–266.
- [43] J. D. Biggins. “Branching Out”. Mar. 24, 2010.
- [44] J. D. Biggins. “The First- and Last-Birth Problems for a Multitype Age-Dependent Branching Process”. In: *Advances in Applied Probability* 8 (1976), pp. 446–459.
- [45] J. D. Biggins and A. E. Kyprianou. “Measure Change in Multitype Branching”. In: *Advances in Applied Probability* 36 (2004), pp. 544–581.
- [46] John D. Biggins. “Chernoff’s Theorem in the Branching Random Walk”. In: *Journal of Applied Probability* 14 (1977), pp. 630–636.
- [47] John D. Biggins. “Martingale Convergence in the Branching Random Walk”. In: *Journal of Applied Probability* 14 (Mar. 1977), pp. 25–37.
- [48] John D. Biggins. “Uniform Convergence of Martingales in the Branching Random Walk”. In: *Annals of Probability* 20 (Jan. 1992), pp. 137–151.
- [49] E. Bolthausen and A.-S. Sznitman. “On Ruelle’s Probability Cascades and an Abstract Cavity Method”. In: *Communications in Mathematical Physics* 197 (Oct. 1, 1998), pp. 247–276.
- [50] Erwin Bolthausen. “An Iterative Construction of Solutions of the TAP Equations for the Sherrington–Kirkpatrick Model”. In: *Communications in Mathematical Physics* 325 (Jan. 1, 2014), pp. 333–366.
- [51] Benjamin Bonnefont. “The Overlap Distribution at Two Temperatures for the Branching Brownian Motion”. In: *Electronic Journal of Probability* 27 (none Jan. 2022), pp. 1–21.
- [52] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Feb. 7, 2013.
- [53] Anton Bovier and Irina Kurkova. “Derrida’s Generalised Random Energy Models 1: Models with Finitely Many Hierarchies”. In: *Annales de l’Institut Henri Poincaré (B) Probability and Statistics* 40 (July 1, 2004), pp. 439–480.
- [54] Anton Bovier and Irina Kurkova. “Derrida’s Generalized Random Energy Models 2: Models with Continuous Hierarchies”. In: *Annales de l’Institut Henri Poincaré (B) Probability and Statistics* 40 (July 1, 2004), pp. 481–495.
- [55] Anton Bovier and Irina Kurkova. “Much Ado about Derrida’s GREM”. In: *Spin Glasses*. Ed. by Erwin Bolthausen and Anton Bovier. Vol. 1900. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 81–115.

- [56] Maury Bramson. “Convergence of Solutions of the Kolmogorov Equation to Travelling Waves”. In: *Memoirs of the American Mathematical Society* 44 (1983), pp. 0–0.
- [57] Maury D. Bramson. “Maximal displacement of branching brownian motion”. In: *Communications on Pure and Applied Mathematics* 31 (1978), pp. 531–581.
- [58] A. J. Bray and M. A. Moore. “Chaotic Nature of the Spin-Glass Phase”. In: *Physical Review Letters* 58 (Jan. 5, 1987), pp. 57–60.
- [59] D. Capocaccia, M. Cassandro, and P. Picco. “On the Existence of Thermodynamics for the Generalized Random Energy Model”. In: *Journal of Statistical Physics* 46 (Feb. 1, 1987), pp. 493–505.
- [60] David Carpentier and Pierre Le Doussal. “Glass Transition of a Particle in a Random Potential, Front Selection in Nonlinear Renormalization Group, and Entropic Phenomena in Liouville and Sinh-Gordon Models”. In: *Physical Review E* 63 (Jan. 24, 2001), p. 026110.
- [61] Sourav Chatterjee. *Disorder Chaos and Multiple Valleys in Spin Glasses*. Nov. 23, 2009. (Visited on 01/06/2023). preprint.
- [62] Sourav Chatterjee. *Spin Glass Phase at Zero Temperature in the Edwards-Anderson Model*. Feb. 28, 2023. (Visited on 04/08/2023). preprint.
- [63] Sourav Chatterjee. “Spin Glasses and Stein’s Method”. In: *Probability Theory and Related Fields* 148 (Nov. 1, 2010), pp. 567–600.
- [64] Sourav Chatterjee. *Superconcentration and Related Topics*. Springer Monographs in Mathematics. Cham: Springer International Publishing, 2014.
- [65] B. Chauvin and A. Rouault. “Boltzmann-Gibbs Weights in the Branching Random Walk”. In: *Classical and Modern Branching Processes. Proceedings of the IMA Workshop, Minneapolis, MN, USA, June 13–17, 1994*. New York, NY: Springer, 1997, pp. 41–50.
- [66] Brigitte Chauvin and Alain Rouault. “KPP Equation and Supercritical Branching Brownian Motion in the Subcritical Speed Area. Application to Spatial Trees”. In: *Probability Theory and Related Fields* 80 (Dec. 1, 1988), pp. 299–314.
- [67] Wei-Kuo Chen. “Chaos in the Mixed Even-Spin Models”. In: *Communications in Mathematical Physics* 328 (June 1, 2014), pp. 867–901.
- [68] Wei-Kuo Chen. “Disorder Chaos in the Sherrington–Kirkpatrick Model with External Field”. In: *The Annals of Probability* 41 (Sept. 2013), pp. 3345–3391.
- [69] Wei-Kuo Chen. “Variational Representations for the Parisi Functional and the Two-Dimensional Guerra–Talagrand Bound”. In: *The Annals of Probability* 45 (6A 2017), pp. 3929–3966.
- [70] Wei-Kuo Chen, Madeline Handschy, and Gilad Lerman. “On the Energy Landscape of the Mixed Even P-Spin Model”. In: *Probability Theory and Related Fields* 171 (June 1, 2018), pp. 53–95.

- [71] Wei-Kuo Chen and Dmitry Panchenko. “An Approach to Chaos in Some Mixed  $p$ -Spin Models”. In: *Probability Theory and Related Fields* 157 (Oct. 1, 2013), pp. 389–404.
- [72] Wei-Kuo Chen and Dmitry Panchenko. “Temperature Chaos in Some Spherical Mixed  $p$ -Spin Models”. In: *Journal of Statistical Physics* 166 (Mar. 1, 2017), pp. 1151–1162.
- [73] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. “Generalized TAP Free Energy”. In: *Communications on Pure and Applied Mathematics* 76 (2023), pp. 1329–1415.
- [74] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. “The Generalized TAP Free Energy II”. In: *Communications in Mathematical Physics* 381 (Jan. 1, 2021), pp. 257–291.
- [75] Wei-Kuo Chen and Arnab Sen. “Parisi Formula, Disorder Chaos and Fluctuation for the Ground State Energy in the Spherical Mixed  $p$ -Spin Models”. In: *Communications in Mathematical Physics* 350 (Feb. 1, 2017), pp. 129–173.
- [76] Wei-Kuo Chen and Si Tang. “On Convergence of the Cavity and Bolthausen’s TAP Iterations to the Local Magnetization”. In: *Communications in Mathematical Physics* 386 (Sept. 1, 2021), pp. 1209–1242.
- [77] Wei-Kuo Chen et al. “Disorder Chaos in the Spherical Mean-Field Model”. In: *Journal of Statistical Physics* 160 (July 1, 2015), pp. 417–429.
- [78] X. Chen. “Scaling Limit of the Path Leading to the Leftmost Particle in a Branching Random Walk”. In: *Theory of Probability & Its Applications* 59 (Jan. 2015), pp. 567–589.
- [79] Xinxin Chen, Thomas Madaule, and Bastien Mallein. “On the Trajectory of an Individual Chosen According to Supercritical Gibbs Measure in the Branching Random Walk”. In: *Stochastic Processes and their Applications* 129 (Oct. 2019), pp. 3821–3858.
- [80] William L. Corbin et al. *Extraits Des Procès-Verbaux Des Séances / Société Philomathique de Paris*. Vol. 10. Paris: A. René, 1845, pp. 1–440. 440 pp.
- [81] Antoine Augustin Cournot. *De l’origine et des limites de la correspondance entre l’algèbre et la géométrie*. L. Hachette, 1847. 426 pp.
- [82] A. Crisanti, L. Leuzzi, and T. Rizzo. “The Complexity of the Spherical  $p$ -Spin Spin Glass Model, Revisited”. In: *The European Physical Journal B - Condensed Matter and Complex Systems* 36 (Nov. 1, 2003), pp. 129–136.
- [83] A. Crisanti and H.-J. Sommers. “Thouless-Anderson-Palmer Approach to the Spherical  $p$ -Spin Spin Glass Model”. In: *Journal de Physique I* 5 (July 1, 1995), pp. 805–813.
- [84] Dembo, Amir and Zeitouni, Ofer. *Large Deviations Techniques and Applications*. Stochastic Modelling and Applied Probability. Springer-Verlag Berlin Heidelberg, 2010.

- [85] B. Derrida. “A Generalization of the Random Energy Model Which Includes Correlations between Energies”. In: *Journal de Physique Lettres* 46 (1985), pp. 401–407.
- [86] B. Derrida. “Random-Energy Model: Limit of a Family of Disordered Models”. In: *Physical Review Letters* 45 (July 14, 1980), pp. 79–82.
- [87] B. Derrida and E. Gardner. “Magnetic Properties and the Function  $q(x)$  of the Generalised Random-Energy Model”. In: *Journal of Physics C: Solid State Physics* 19 (Oct. 1986), p. 5783.
- [88] B. Derrida and E. Gardner. “Solution of the Generalised Random Energy Model”. In: *Journal of Physics C: Solid State Physics* 19 (May 1986), p. 2253.
- [89] B. Derrida and G. Toulouse. “Sample to Sample Fluctuations in the Random Energy Model”. In: *Journal de Physique Lettres* 46 (Mar. 1, 1985), pp. 223–228.
- [90] Bernard Derrida. “Random-Energy Model: An Exactly Solvable Model of Disordered Systems”. In: *Physical Review B* 24 (Sept. 1, 1981), pp. 2613–2626.
- [91] Bernard Derrida and Herbert Spohn. “Polymers on Disordered Trees, Spin Glasses, and Traveling Waves”. In: *Journal of Statistical Physics* 51 (June 1, 1988), pp. 817–840.
- [92] Jian Ding. “Exponential and Double Exponential Tails for Maximum of Two-Dimensional Discrete Gaussian Free Field”. In: *Probability Theory and Related Fields* 157 (Oct. 1, 2013), pp. 285–299.
- [93] Jian Ding, Ronen Eldan, and Alex Zhai. “On Multiple Peaks and Moderate Deviations for the Supremum of a Gaussian Field”. In: *The Annals of Probability* 43 (Nov. 2015), pp. 3468–3493.
- [94] David L. Donoho, Arian Maleki, and Andrea Montanari. “Message-Passing Algorithms for Compressed Sensing”. In: *Proceedings of the National Academy of Sciences* 106 (Nov. 10, 2009), pp. 18914–18919.
- [95] Bertrand Duplantier and Scott Sheffield. “Liouville Quantum Gravity and KPZ”. In: *Inventiones mathematicae* 185 (Aug. 1, 2011), pp. 333–393.
- [96] S. F. Edwards and P. W. Anderson. “Theory of Spin Glasses”. In: *Journal of Physics F: Metal Physics* 5 (May 1975), p. 965.
- [97] Ronen Eldan. “A Simple Approach to Chaos For P-Spin Models”. In: *Journal of Statistical Physics* 181 (Nov. 1, 2020), pp. 1266–1276.
- [98] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. “A Spectral Condition for Spectral Gap: Fast Mixing in High-Temperature Ising Models”. Aug. 6, 2021.
- [99] Daniel S. Fisher and David A. Huse. “Ordered Phase of Short-Range Ising Spin-Glasses”. In: *Physical Review Letters* 56 (Apr. 14, 1986), pp. 1601–1604.

- [100] R. A. Fisher. “The Wave of Advance of Advantageous Genes”. In: *Annals of Eugenics* 7 (1937), pp. 355–369.
- [101] Francis Galton and Henry William Watson. “On the Probability of the Extinction of Families”. In: *The Journal of the Anthropological Institute of Great Britain and Ireland* 4 (1875), pp. 138–144.
- [102] David Gamarnik. “The Overlap Gap Property: A Topological Barrier to Optimizing over Random Structures”. In: *Proceedings of the National Academy of Sciences* 118 (Oct. 12, 2021), e2108492118.
- [103] David Gamarnik and Aukosh Jagannath. “The Overlap Gap Property and Approximate Message Passing Algorithms for  $\mathbb{Z}_2$ -Spin Models”. In: *The Annals of Probability* 49 (Jan. 2021), pp. 180–205.
- [104] David Gamarnik, Cristopher Moore, and Lenka Zdeborová. “Disordered Systems Insights on Computational Hardness”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2022 (Jan. 2022), p. 114015.
- [105] Nina Gantert and Thomas Höfelsauer. “Large Deviations for the Maximum of a Branching Random Walk”. In: *Electronic Communications in Probability* 23 (none Jan. 2018), pp. 1–12.
- [106] Nina Gantert, Yueyun Hu, and Zhan Shi. “Asymptotics for the Survival Probability in a Killed Branching Random Walk”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 47 (Feb. 2011), pp. 111–129.
- [107] Christophe Garban et al. “Negative Moments for Gaussian Multiplicative Chaos on Fractal Sets”. In: *Electronic Communications in Probability* 23 (none Jan. 2018), pp. 1–10.
- [108] Reza Gheissari and Aukosh Jagannath. “On the Spectral Gap of Spherical Spin Glass Dynamics”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 55 (May 2019), pp. 756–776.
- [109] Francesco Guerra. “Broken Replica Symmetry Bounds in the Mean Field Spin Glass Model”. In: *Communications in Mathematical Physics* 233 (Feb. 1, 2003), pp. 1–12.
- [110] J. M. Hammersley. “Postulates for Subadditive Processes”. In: *The Annals of Probability* 2 (1974), pp. 652–680.
- [111] Theodore Edward Harris. *The Theory of Branching Processes*. 1st ed. Springer Berlin Heidelberg, 1963. XVI, 232.
- [112] Fu-Hsuan Ho and Pascal Maillard. “Efficient Approximation of Branching Random Walk Gibbs Measures”. In: *Electronic Journal of Probability* 27 (none Jan. 2022), pp. 1–18.
- [113] Yueyun Hu. “How Big Is the Minimum of a Branching Random Walk?” In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 52 (Feb. 2016), pp. 233–260.
- [114] Yueyun Hu and Zhan Shi. “Minimal Position and Critical Martingale Convergence in Branching Random Walks, and Directed Polymers on Disordered Trees”. In: *The Annals of Probability* 37 (Mar. 2009), pp. 742–789.

- [115] Brice Huang and Mark Sellke. *Tight Lipschitz Hardness for Optimizing Mean Field Spin Glasses*. Sept. 11, 2022. (Visited on 08/04/2023). preprint.
- [116] Aukosh Jagannath. “On the Overlap Distribution of Branching Random Walks”. In: *Electronic Journal of Probability* 21 (none Jan. 2016), pp. 1–16.
- [117] Yoshiyuki Kabashima. “A CDMA Multiuser Detection Algorithm on the Basis of Belief Propagation”. In: *Journal of Physics A: Mathematical and General* 36 (Oct. 2003), p. 11111.
- [118] Jean-Pierre Kahane and Jacques Peyrière. “Sur certaines martingales de Benoit Mandelbrot”. In: *Advances in Mathematics* 22 (Nov. 1, 1976), pp. 131–145.
- [119] Richard M. Karp and Judea Pearl. “Searching for an Optimal Path in a Tree with Random Costs”. In: *Artificial Intelligence. Search and Heuristics* 21 (Mar. 1, 1983), pp. 99–116.
- [120] David G. Kendall. “Branching Processes Since 1873”. In: *Journal of the London Mathematical Society* s1-41 (1966), pp. 385–406.
- [121] David G. Kendall. “The Genealogy of Genealogy Branching Processes before (and after) 1873”. In: *Bulletin of the London Mathematical Society* 7 (1975), pp. 225–253.
- [122] J. F. C. Kingman. “The First Birth Problem for an Age-dependent Branching Process”. In: *The Annals of Probability* 3 (Oct. 1, 1975).
- [123] Scott Kirkpatrick and David Sherrington. “Infinite-Ranged Models of Spin-Glasses”. In: *Physical Review B* 17 (June 1, 1978), pp. 4384–4403.
- [124] Andrey Kolmogorov, Ivan Petrovsky, and Nikolai Piskunov. “Étude de l’équation de La Diffusion Avec Croissance de La Quantité de Matière et Son Application à Un Problème Biologique”. In: *Moscow Univ. Bull. Ser. Internat. Sect. A* 1 (1937), pp. 1–25.
- [125] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1991.
- [126] Quansheng Liu. “An Extension of a Functional Equation of Poincaré and Mandelbrot”. In: *Asian Journal of Mathematics* 6 (Mar. 2002), pp. 145–168.
- [127] Quansheng Liu. “Asymptotic Properties and Absolute Continuity of Laws Stable by Random Weighted Mean”. In: *Stochastic Processes and their Applications* 95 (Sept. 1, 2001), pp. 83–107.
- [128] Quansheng Liu and Alain Rouault. “On Two Measures Defined on the Boundary of a Branching Tree”. In: *Classical and Modern Branching Processes*. Ed. by Krishna B. Athreya and Peter Jagers. Red. by Avner Friedman and Robert Gulliver. Vol. 84. New York, NY: Springer New York, 1997, pp. 187–201.
- [129] Russell Lyons. “A Simple Path to Biggins’ Martingale Convergence for Branching Random Walk”. In: *Classical and Modern Branching Processes*. Ed. by Krishna B. Athreya and Peter Jagers. The IMA Volumes in Mathematics and Its Applications. New York, NY: Springer, 1997, pp. 217–221.

- [130] Thomas Madaule. “Convergence in Law for the Branching Random Walk Seen from Its Tip”. In: *Journal of Theoretical Probability* 30 (Mar. 1, 2017), pp. 27–63.
- [131] Thomas Madaule. “First Order Transition for the Branching Random Walk at the Critical Parameter”. In: *Stochastic Processes and their Applications* 126 (Feb. 2016), pp. 470–502.
- [132] Pascal Maillard and Ofer Zeitouni. “Slowdown in Branching Brownian Motion with Inhomogeneous Variance”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 52 (Aug. 2016), pp. 1144–1160.
- [133] Bastien Mallein. “Genealogy of the Extremal Process of the Branching Random Walk”. In: *Latin American Journal of Probability and Mathematical Statistics* 15 (2018), p. 1065.
- [134] Bastien Mallein. “Maximal Displacement of a Branching Random Walk in Time-Inhomogeneous Environment”. In: *Stochastic Processes and their Applications* 125 (Oct. 1, 2015), pp. 3958–4019.
- [135] Benoit Mandelbrot. “Multiplications Aléatoires Itérées et Distributions Invariantes Par Moyenne Pondérée Aléatoire”. In: *CR Acad. Sci. Paris* 278 (1974), pp. 355–358.
- [136] Benoit Mandelbrot. “Multiplications Aléatoires Itérées et Distributions Invariantes Par Moyenne Pondérée Aléatoire, II Quelques Extensions.” In: *C. R. Acad. Sci. Paris. A* 278 (1974), pp. 355–358.
- [137] Benoit B. Mandelbrot. “Intermittent Turbulence in Self-Similar Cascades: Divergence of High Moments and Dimension of the Carrier”. In: *Journal of Fluid Mechanics* 62 (Jan. 1974), pp. 331–358.
- [138] Benoit B. Mandelbrot. “Iterated Random Multiplications and Invariance under Randomly Weighted Averaging”. In: *Multifractals and 1/f Noise: Wild Self-Affinity in Physics (1963–1976)*. Ed. by Benoit B. Mandelbrot. New York, NY: Springer, 1999, pp. 358–371.
- [139] W. Marshall. “Specific Heat of Dilute Alloys”. In: *Physical Review* 118 (June 15, 1960), pp. 1519–1523.
- [140] H. P. McKean. “Application of Brownian Motion to the Equation of Kolmogorov-Petrovskii-Piskunov”. In: *Communications on Pure and Applied Mathematics* 28 (1975), pp. 323–331.
- [141] M. Mézard, T. Mora, and R. Zecchina. “Clustering of Solutions in the Random Satisfiability Problem”. In: *Physical Review Letters* 94 (May 19, 2005), p. 197205.
- [142] M. Mézard et al. “Nature of the Spin-Glass Phase”. In: *Physical Review Letters* 52 (Mar. 26, 1984), pp. 1156–1159.
- [143] Marc Mezard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin Glass Theory And Beyond: An Introduction To The Replica Method And Its Applications*. World Scientific Publishing Company, Nov. 1, 1987. 477 pp.

- [144] G. M. Molchan. “Scaling Exponents and Multifractal Dimensions for Independent Random Cascades”. In: *Communications in Mathematical Physics* 179 (Jan. 1996), pp. 681–702.
- [145] Andrea Montanari. “Optimization of the Sherrington-Kirkpatrick Hamiltonian”. In: *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. 2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS). Nov. 2019, pp. 1417–1433.
- [146] Jean-Christophe Mourrat and Dmitry Panchenko. “Extending the Parisi Formula along a Hamilton-Jacobi Equation”. In: *Electronic Journal of Probability* 25 (none Jan. 2020), pp. 1–17.
- [147] A. M. B. Nascimento and L. R. Fontes. “Convergence Time to Equilibrium of the Metropolis Dynamics for the GREM”. In: *Journal of Statistical Physics* 178 (Jan. 1, 2020), pp. 297–317.
- [148] Charles M. Newman and Daniel L. Stein. “Thermodynamic Chaos and the Structure of Short-Range Spin Glasses”. In: *Mathematical Aspects of Spin Glasses and Neural Networks*. Ed. by Anton Bovier and Pierre Picco. Boston, MA: Birkhäuser Boston, 1998, pp. 243–287.
- [149] Miika Nikula. “Small Deviations in Lognormal Mandelbrot Cascades”. In: *Electronic Communications in Probability* 25 (none Jan. 2020), pp. 1–12.
- [150] Michel Pain. “Mouvement brownien branchant et autres modèles hiérarchiques en physique statistique”. PhD thesis. Sorbonne Université, June 5, 2019.
- [151] Michel Pain. “The Near-Critical Gibbs Measure of the Branching Random Walk”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 54 (Aug. 2018), pp. 1622–1666.
- [152] Michel Pain and Olivier Zindy. “Two-Temperatures Overlap Distribution for the 2D Discrete Gaussian Free Field”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 57 (May 2021), pp. 685–699.
- [153] Dmitry Panchenko. “Chaos in Temperature in Generic 2p-Spin Models”. In: *Communications in Mathematical Physics* 346 (Sept. 1, 2016), pp. 703–739.
- [154] Dmitry Panchenko. “The Parisi Formula for Mixed  $p$ -Spin Models”. In: *The Annals of Probability* 42 (May 1, 2014).
- [155] Dmitry Panchenko. “The Parisi Ultrametricity Conjecture”. In: *Annals of Mathematics* 177 (Jan. 1, 2013), pp. 383–393.
- [156] Dmitry Panchenko. *The Sherrington-Kirkpatrick Model*. Springer Monographs in Mathematics. New York, NY: Springer New York, 2013.
- [157] G. Parisi. “A Sequence of Approximated Solutions to the S-K Model for Spin Glasses”. In: *Journal of Physics A: Mathematical and General* 13 (Apr. 1980), p. L115.
- [158] G. Parisi. “Infinite Number of Order Parameters for Spin-Glasses”. In: *Physical Review Letters* 43 (Dec. 3, 1979), pp. 1754–1756.

- [159] Robin Pemantle. “Search Cost for a Nearly Optimal Path in a Binary Tree”. In: *Annals of Applied Probability* 19 (Aug. 2009), pp. 1273–1291.
- [160] George Pólya and Gabor Szegő. *Problems and Theorems in Analysis I*. Berlin, Heidelberg: Springer, 1972.
- [161] Guillaume Remy. “The Fyodorov–Bouchaud Formula and Liouville Conformal Field Theory”. In: *Duke Mathematical Journal* 169 (Jan. 2020), pp. 177–211.
- [162] Tommaso Rizzo. “Chaos in Mean-field Spin-glass Models”. In: *Spin Glasses: Statics and Dynamics*. Ed. by Anne Boutet de Monvel and Anton Bovier. Progress in Probability. Basel: Birkhäuser, 2009, pp. 143–157.
- [163] Raoul Robert and Vincent Vargas. “Gaussian Multiplicative Chaos Revisited”. In: *The Annals of Probability* 38 (Mar. 2010), pp. 605–631.
- [164] Haskell P. Rosenthal. “On the Subspaces of  $L_p$  ( $p > 2$ ) Spanned by Sequences of Independent Random Variables”. In: *Israel Journal of Mathematics* 8 (Sept. 1, 1970), pp. 273–303.
- [165] David Ruelle. “A Mathematical Reformulation of Derrida’s REM and GREM”. In: *Communications in Mathematical Physics* 108 (June 1, 1987), pp. 225–239.
- [166] Mark Sellke. *Optimizing Mean Field Spin Glasses with External Field*. May 21, 2023. (Visited on 08/04/2023). preprint.
- [167] David Sherrington and Scott Kirkpatrick. “Solvable Model of a Spin-Glass”. In: *Physical Review Letters* 35 (Dec. 29, 1975), pp. 1792–1796.
- [168] Zhan Shi. *Branching Random Walks*. Vol. 2151. Lecture Notes in Mathematics. Cham: Springer International Publishing, 2015.
- [169] H. Sompolinsky and Annette Zippelius. “Dynamic Theory of the Spin-Glass Phase”. In: *Physical Review Letters* 47 (Aug. 3, 1981), pp. 359–362.
- [170] H. Sompolinsky and Annette Zippelius. “Relaxational Dynamics of the Edwards-Anderson Model and the Mean-Field Theory of Spin-Glasses”. In: *Physical Review B* 25 (June 1, 1982), pp. 6860–6875.
- [171] Eliran Subag. “Following the Ground States of Full-RSB Spherical Spin Glasses”. In: *Communications on Pure and Applied Mathematics* 74 (2021), pp. 1021–1044.
- [172] Eliran Subag. *Free Energy Landscapes in Spherical Spin Glasses*. Apr. 27, 2018. (Visited on 06/26/2023). preprint.
- [173] Eliran Subag. *TAP Approach for Multi-Species Spherical Spin Glasses I: General Theory*. Nov. 13, 2021. (Visited on 07/30/2023). preprint.
- [174] Eliran Subag. *TAP Approach for Multi-Species Spherical Spin Glasses II: The Free Energy of the Pure Models*. May 3, 2022. (Visited on 07/30/2023). preprint.

- [175] Eliran Subag. “The Complexity of Spherical  $p$ -Spin Models—A Second Moment Approach”. In: *The Annals of Probability* 45 (Sept. 2017), pp. 3385–3450.
- [176] Eliran Subag. “The Free Energy of Spherical Pure  $p$ -Spin Models: Computation from the TAP Approach”. In: *Probability Theory and Related Fields* 186 (Aug. 1, 2023), pp. 715–734.
- [177] Eliran Subag. “The Geometry of the Gibbs Measure of Pure Spherical Spin Glasses”. In: *Inventiones mathematicae* 210 (Oct. 1, 2017), pp. 135–209.
- [178] Eliran Subag and Ofer Zeitouni. “The Extremal Process of Critical Points of the Pure  $p$ -Spin Spherical Spin Glass Model”. In: *Probability Theory and Related Fields* 168 (Aug. 1, 2017), pp. 773–820.
- [179] Michel Talagrand. *Mean Field Models for Spin Glasses*. Vol. 1. 2 vols. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
- [180] Michel Talagrand. *Mean Field Models for Spin Glasses*. Vol. 2. 2 vols. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
- [181] Michel Talagrand. “The Parisi Formula”. In: *Annals of Mathematics* 163 (Jan. 1, 2006), pp. 221–263.
- [182] *The Nobel Prize in Physics 2021*. NobelPrize.org. Oct. 5, 2021. URL: <https://www.nobelprize.org/prizes/physics/2021/press-release/> (visited on 04/10/2023).
- [183] D. J. Thouless, P. W. Anderson, and R. G. Palmer. “Solution of ‘Solvable Model of a Spin Glass’”. In: *The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics* 35 (Mar. 1, 1977), pp. 593–601.
- [184] Hugo Touchette. “Equivalence and Nonequivalence of Ensembles: Thermodynamic, Macrostate, and Measure Levels”. In: *Journal of Statistical Physics* 159 (June 2015), pp. 987–1016.
- [185] Ofer Zeitouni. “Branching Random Walks and Gaussian Fields”. In: *Probability and statistical physics in St. Petersburg* 91 (2016), pp. 437–471.