École Normale Supérieure

Cursus Mathématiques-Physique

Internship report

Glassy dynamics: Trap and Random energy models

Author:
Ivailo Hartarsky
École Normale Supérieure

Supervisors:
Giulio Biroli
CEA Saclay, IPhT
Justin Salez
Paris Diderot, LPMA

Thursday 25th August, 2016

Abstract

In this report we outline the work accomplished during the internship. After giving the necessary prerequisites we focus mostly on the unexplored field of the maximal energy in the Random Energy Model under Metropolis dynamics and in other related settings. We present some quantitative non-rigorous results concerning it as well as their numerical aspects.

Contents

Foreword

I would like to address first of all future students taking the double diploma option, as I know from experience that they will not read on much further. The most important piece of advice I have for them is to follow closely the instructions of the tutors of the cursus Math-Physique and to get down to preparing their internships as early as possible, because the formalities can be quite slow to accomplish, not to speak of the amount of things one needs to learn in order to be able to do any contemporary theoretical physics (if at all possible at bachelor's level).

Thanks

First and foremost, I would like to express my utter gratitude for the guidance and help I recieved from both *Giulio Biroli* and *Justin Salez* as well as for their never waning enthusiasm and not least, for all they have taught me or given me the means to learn by myself. I thank the staff of *Institut de Physique Théorique* and more generally of the *Commissariat de l'Energie Atomique*, Saclay for the hospitality as well as the *École Normale Supérieure* for the flawless working conditions it provided me with during this academic year. I thank *Jéremie Bouttier* and *Kian-Amir Kashani-Poor* for their clear instructions and for their competent responses to my questions as well as for having the patience to study and evaluate this lengthy report. The aid of *Marco Baity Jesi*, with whom I discussed some numerical aspects of the REM dynamics and who kindly provided me with data from his simulations, should be acknowledged. I also thank *Vyara Hartarska* for her being an excellent and patient listener, which proved most helpful as always. I would also like to mention the aid of *Guilhem Semerjian*, *Francesco Zamponi* and *Cristina Toninelli* for finding the great supervisors I had. Finally, I would like to thank *Thomas Duquesne* for his simply perfect course, which could not have given me more solid a base for my future almost sure probabilistic endeavours, and *Jean-François Logeais* thanks to the no less perfect courses of whom I developed my taste for mathematical physics and my faith in its true existence.

Course of the internship

Let us first outline the work accomplished during the internship. We also discuss the content of some of the references.

Mathematical prerequisites

The strictly mathematical part of the internship which was supervised by Justin Salez took place during the second semester of the academic year in parallel with the coursework, namely from mid-February to mid-May. He suggested several references consistent with what Giulio Biroli had told him we would do in the "Physics" part as the author's stochastic process background was clearly insufficient if at all existent. More precisely:

[23] Namely the chapters 11-13 on conditionning, discrete martingales and discrete Markov chains, which were not covered in [14], which was already known and is a better reference on what can be found there.

[30] For the extensive treatment of Markov chains and Poisson processes.

[20] Recommended by Justin Salez due to the rather messy structure and hasty reasoning the author noticed in [30].

[15] For the Central limit theorems chapter and more precisely stable laws.

[21] For the concise and rigorous introduction to the REM (see 2.1.2).

Although those references left little ambiguity, he did discuss some problematic points with the author in order to clarify the uncertainties and to ensure the good comprehension of the material. This part of the work gave rise to most of chapters 1 and 4 and were, of course, of constant use throughout the rest of the present paper. The supervisor was always ready to suggest an adequate reference to reply to the trainee's needs.

Physics prerequisites and introduction to the area

Giulio Biroli suggested a few of the following references, which were to introduce the author to the area of spin glasses in general and the REM in particular and added a few others when prompted by my queries. The last three were stumbled upon rather accidentally and were thus less appropriate, but informative nonetheless. Their study was accomplished before the start of the internship in mid-June.

[24] A rather contemporary discussion of spin glasses and its links to information and computation. It provides a fair introduction of all those subjects including the REM.

[13] The original article introducing the REM and giving many of the fundamental results on it.

[8] The original article introducing trap models and discussing ageing from a phenomenological point of view.

[5] For a summary on aging in trap models and in particular [4] – namely section 5.

[3] For some mathematical results on aging in the REM.

[4],[11] For rather self-contained mathematical proofs of aging, namely on the complete graph, using different approaches. [11] gives as well a detailed explanation of the reduction of the REM to the BTM.

[22] For an excellent introduction to Monte Carlo methods from the practical viewpoint, namely chapters 1,2 and 7.

[10] For a physicist's view on the central limit theorems and extreme values – namely chapter 1.

[29] For a general qualitative overview of the area even though it is by far not an introduction for non-specialists.

[25] For a more comprehensible for beginners and more extensive coverage of the area despite the obscure reasoning.

[27] A general overview of phase transitions, which states rather what does not happen in glasses than what does.

Biroli was also available to discuss and suggest further reading on issues that were encountered in reading those, as they were much less clear than the mathematical ones at least to the student's impression. Although many of those references were there rather to provide cultural context than to be used directly, they supplied some much needed height on what was to be done during the internship.

Internship at IPhT

The internship at IPhT took place from Monday 20^{th} June, 2016 to Friday 22^{nd} July, 2016 under the supervision of Giulio Biroli. During this period the author worked in considerable autonomy, discussing with Biroli from time to time and yet often enough to keep the former occupied. At the beginning the supervisor was more directive, asking carefully chosen questions, whose answers he was well aware of, and giving the intern the time necessary for him to find the answers himself and to build up his understanding eventually arriving at the core matter of the internship. About that time Biroli gradually started suggesting only lines of thought for further development. He was always open to hearing out the ideas of the author and attentive when doing so, which allowed him to spot what sometimes had eluded his trainee.

Collaboration with Marco Baity Jesi, who is a post-doctoral student also under the supervision of Biroli, during this period also needs mentioning, as it clearly played an important role in the internship.

Impressions

In the author's opinion the whole idea of having a theoretical physics internship for the cursus Math-Pysique is very well conceived. Moreover, the mixed and co-supervised internship is very appropriate. It gave him the occasion to learn much Mathematics that will be of use for his future studies and Physics that is in close enough relation with Mathematics and therefore may provide a useful alternative point of view as well as a field of application and study. Furthermore, this first experience of actual research as it is done in a theoretical physics laboratory is quite an invaluable confirmation of how things work "in real life" and yet in a sufficiently supervised environment to ensure that work stays on the right track. Finally, the writing of the report, even though it is quite time consuming and of little interest for contemporary research, is a good opportunity to develop and refine writing skills that will be essential for the students' future careers.

An important aspect of the internship was the choice of the domain and afterwards of the theme. Most fortunately both went on splendidly, as the domain was precisely to the author's liking with its tight relations with probability and the theme was very successfully suggested by Biroli and turned out to be very accessible even at such low level but still by far not trivial and demanding good understanding of the context and the previously established results. Despite the fact that Biroli's choice was very appropriate it should be brought to the tutors of the cursus Math-Physique's attention that it might be a good idea to survey more closely the supervisors' ideas in order to ensure that they do not overestimate the modest knowledge of the students, which even their ultimate resolve may not be able to make up for. In this respect, Salez's selection of adapted and self-contained references that could make a smooth extension of already known material without need of searching elsewhere was no less crucial. Finally, both supervisors' pedagogical skills were much appreciated. The trainee's impressions could not easily be more positive. The only regrettable aspect is the fact that the current work remained almost entirely physical due to the author's lack of experience with processes with two levels of randomnees.

Chapter 1

Preliminaries and setting

1.1 Preliminary Mathematics

Even though the Mathematics used in this text is fascinating, it is out of the question to give here a self-contained work on stochastic processes. Yet, we will try to state in this section the main results and notions used throughout the text. Should the reader happen to be unfamiliar with any of these, it is highly recommended to consult for instance the references in Mathematical prerequisites if one intends to understand the matter on a non-superficial level. We will not give the "good" definitions, but rather stick to what is the most conveniently announced. Note that some objects are defined inside theorems and not restated as definitions though it is mentioned explicitly if we fix notations.

1.1.1 Generalities

In order to give a precise setting for what follows, we fix $\mathscr{U} := \left(\Omega, \mathscr{F}, \mathbb{P}, (\mathscr{F}_t)_{t\in\mathbb{R}_+}\right)$ a filtered probability space referred to as *the universe*. Unless otherwise mentioned any stochastic process[1] $(X_t)_{t\in\mathbb{R}_+}$[2] will be supposed adapted to $\mathscr{U}$. The process is valued in a measurable space $(S, \mathscr{S})$, $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$[3]. If the notation θ_t is used, we suppose that $\mathscr{U}$ is equipped with a commutative translation monoid $(\theta_t)_{t\in\mathbb{R}_+}$ such that $\forall t \geq 0, s \geq 0, X_t \circ \theta_s = X_{s+t}$ and we still denote $\mathscr{U} := \left(\Omega, \mathscr{F}, \mathbb{P}, (\mathscr{F}_t)_{t\in\mathbb{R}_+}, \theta, X\right)$. The same notations are used for discrete[4] time processes, substituting t by n and $\mathbb{R}_+$ by $\mathbb{N}$, where $\mathbb{N} := \{0, 1, \dots\}$.

Definition 1.1.1.1. We call *trajectories* of a process the functions $t \mapsto X_t(\omega)$ and we denote them by $X(\omega)$.

Definition 1.1.1.2.

- We call a *stopping time* a random variable[5] $T : \Omega \to \mathbb{R}_+ \cup \{\infty\}$[6] such that $\forall t \geq 0$, $(T < t) \in \mathscr{F}_t$.
- For a stopping time T we note $\mathscr{F}_T := \{A \in \mathscr{F} \mid \forall t \in \mathbb{R}_+, A \cap \{T < t\} \in \mathscr{F}_t\}$ the *σ-algebra of the past*.
- For a stopping time T we note $X_T(\omega) := X_{T(\omega)}(\omega)\mathbb{1}_{\{T<\infty\}}(\omega)$. We sometimes extend this definition to $\{T = \infty\}$ by setting a value depending on the context. If that set is negligible, that is of no importance.

In the discrete case we replace $T < t$ by $T \leq n$, which makes a difference.

Definition 1.1.1.3. For P a transition probability from S to $(F, \mathfrak{F})$ and $f \in \mathcal{F}_m(S, \mathbb{R})$[7] we note $Pf(y) := \int_S P(y, \mathrm{d}x)f(x)$, when defined, for instance for f positive or bounded. Hence $Pf \in \mathcal{F}_m(F, \mathbb{R})$ respectively positive or bounded.

Definition 1.1.1.4. We say that a process X equals X' in law if $\forall n \in \mathbb{N}, (t_1, \dots, t_n) \in \mathbb{R}_+^n$, $\mathscr{L}_{(X_{t_1},\dots,X_{t_n})} = \mathscr{L}_{(X'_{t_1},\dots,X'_{t_n})}$[8].

[1] We will say simply process.

[2] We will write simply X when there is no ambiguity. This same notation is used for other objects : we write for instance $(X_t)_{t\in\mathbb{R}_+}$ at the first occurrence and then use simply X without any mention afterwards.

[3] σ-algebras will be systematically implicit. $\mathbb{R}^n$ is equipped with the Lebesgue measure l unless otherwise stated.

[4] We will use countable and discrete interchangeably as any countable set will be equipped with the discrete topology and the algebra $\mathfrak{P}(S)$.

[5] We will say simply variable and it is implicit that it is measurable.

[6] We will write simply $[0; \infty]$.

[7] We will denote by $\mathcal{F}_m(S, F)$ measurable functions from S to F, $\mathcal{F}(S, F)$ – general and $\mathcal{F}_b(S)$ – measurable bounded real functions.

[8] We will denote by $\mathscr{L}_Y(\mathrm{d}y)$ the law of the variable Y and equality in law will be denoted by $\overset{law}{=}$.

1.1.2 Conditional expectation

Theorem 1.1.2.1. *For $X \in \mathrm{L}^1(\Omega, \mathscr{F}, \mathbb{P})$ and $\mathscr{G}$ a sub-algebra of $\mathscr{F}$ there is a unique variable $\mathbb{E}[X \mid \mathscr{G}] \in \mathrm{L}^1(\Omega, \mathscr{G}, \mathbb{P})$ called conditional expectation of X with respect to $\mathscr{G}$ such that*

$$\forall A \in \mathscr{G},\ \mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X \mid \mathscr{G}]\mathbb{1}_A].$$

Moreover, it verifies

$$\forall Z \in \mathcal{F}_b((\Omega, \mathscr{G})),\ \mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X \mid \mathscr{G}]Z]$$

and

$$X \in \mathcal{F}_m((\Omega, \mathscr{F}), [0; \infty]) \Rightarrow \mathbb{E}[X \mid \mathscr{G}] \in \mathcal{F}_m((\Omega, \mathscr{G}), [0; \infty]).$$

Definition 1.1.2.2. For $\mathscr{G}$ a sub-algebra of $\mathscr{F}$ we introduce the *conditional probability* knowing $\mathscr{G}$ by

$$\forall A \in \mathscr{F},\ \mathbb{P}(A|\mathscr{G}) := \mathbb{E}[\mathbb{1}_A|\mathscr{G}].$$

Remark 1.1.2.3. We will not take into account the fact that it is only defined a.e.[9] and treat it as a true probability.

Definition 1.1.2.4. For $(A, B) \in \mathscr{F}^2$, $\mathbb{P}(A) > 0$ and X a variable we introduce

$$\mathbb{P}(B \mid A) := \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}, \quad \mathbb{P}(\bullet \mid X) := \mathbb{P}(\bullet \mid \sigma(X)^{10}).$$

These definitions are coherent with for a.e. $\omega \in A$, $\mathbb{P}(B \mid A) = \mathbb{P}(B \mid \mathbb{1}_A)(\omega)$.

1.1.3 Markov chains and processes

Definition 1.1.3.1. We call a *Markov chain*[11] a process $(X_n)_{n\in\mathbb{N}}$, such that

$$\exists \Pi \in \mathcal{L}\left(\mathbb{R}^S\right)^{12} \forall n \in \mathbb{N}, (x_0, \ldots, x_n, y) \in S^{n+2},\ \mathbb{P}(X_{n+1} = y \mid X_0 = x_0, \ldots, X_n = x_n) = \Pi(x_n, y) =: \pi_{x_n,y}$$

when the conditional probability is defined. Π is called the *transition matrix* (it is a transition probability).

Definition 1.1.3.2. We call a *Markov process*[13] $(X_t)_{t\in\mathbb{R}_+}$ equiped with a transition probability monoid $(P_t)_{t\in\mathbb{R}_+}$ such that

$$\forall t \geq 0, s \geq 0, f \in \mathcal{F}_m(S, \mathbb{R}_+),\ \mathbb{E}[f(X_{t+s}) \mid \mathscr{F}_t] = P_s f(X_t)$$

and whose trajectories are right-continuous (this condition is not in the general definition of a Markov process).

Remark 1.1.3.3. We are already cheating a bit since X is a variable but we have not said in what measurable space it is valued. This problem is solved by the cylinder spanned sigma-algebra or, better yet, the Borel algebra of the Skorokhod topology (or metric) on the càdlàg space, but we will just pretend it did not matter.

Definition 1.1.3.4.

- We set the stopping times $T_n = \inf\{t > T_{n-1} \mid X_t \neq X_{T_{n-1}}\}$ eventually infinite with $T_0 = 0$, that we call *jump times*.
- We set $S_n = T_{n+1} - T_n$ and $\xi_n = X_{T_n}\mathbb{1}_{\{T_n<\infty\}} + \Delta\mathbb{1}_{\{T_n=\infty\}}$ with $\Delta \notin E$.
- We set $T_\infty := \lim T_n$, that we call the *first explosion* time. We say that there is *no explosion* when $T_\infty = \infty$ a.s.

Definition 1.1.3.5. Indifferently for a Markov chain or process (or its transition probability) we note for any $x \in S$ by $\mathbb{P}_x$ a probability such that $\mathbb{P}_x(X_0 = x) = 1$ (and X is Markov). We note for μ a probability measure on S by $\mathbb{P}_\mu$ a probability such that $\forall A \in \mathscr{S}$, $\mathbb{P}_\mu(X_0 \in A) = \mu(A)$. (Their existence is not trivial but holds.) We join $(\mathbb{P}_x)_{x\in S}$ to $\mathscr{U}$.

Theorem 1.1.3.6. *For a Markov process X, under $\mathbb{P}_\mu$ and conditionally to $\mathscr{F}_{T_n}$ the variables S_n and ξ_{n+1} are independent, there are $(a_i)_{i\in S}$ such that $S_{n|\mathscr{F}_{T_n}} \overset{law}{=} EXP(a_{\xi_n})$[14] and $\mathbb{P}_\mu(\xi_{n+1|\mathscr{F}_{T_n}} = j) =: \pi_{\xi_n,j}$ gives the transition matrix of ξ, which is a Markov chain under $\mathbb{P}_\mu$. We shall use these notations as well and we will call* transition rates $q_{i,j} = a_i\pi_{i,j}$.

[9] We will use a. for almost, e. for everywhere, every and s. for sure, surely according to context.
[10] $\sigma(\bullet)$ denotes the spanned sigma-algebra – for example for an event it is $\{\Omega, A, \Omega\backslash A, \emptyset\}$.
[11] Implicitly on a countable (i.e. which injects in $\mathbb{N}$) set S and in discrete time.
[12] Identified henceforth with $\mathfrak{M}_{|S|}$ by the means of the canonical basis, where $|S| \in \mathbb{N} \cup \{\mathbb{N}\}$ is the cardinal avoiding the ω notation.
[13] Implicitly on a countable set S and with time in $\mathbb{R}_+$.
[14] $\mathscr{L}_{EXP(a)}(\mathrm{d}x) = ae^{-ax}\mathbb{1}_{\mathbb{R}_+}\mathrm{d}x$ if $0 < a < \infty$ and δ_∞ otherwise.

Proposition 1.1.3.7. *If $sup_{x\in S} a_x < \infty$, then $\forall x \in S$ $\mathbb{P}_x$-a.s. there is no explosion. This is the case for $|S| < \infty$.*

Definition 1.1.3.8. A Markov chain is

- *reversible* with respect to a measure[15] μ if $\mu_i \pi_{i,j} = \mu_j \pi_{j,i}$. For a Markov process $\mu_i \pi_{i,j} a_i = \mu_j \pi_{j,i} a_j$.
- *invariant* with respect to a measure μ if $\mu\Pi = \mu$. For a Markov process $\mu Q = \mu$.
- *irreducible* if $\forall (x,y) \in S^2\ \exists n \in \mathbb{N},\ \Pi^n(x,y) > 0$. For a Markov process $\forall (x,y) \in S^2\ \exists t \in \mathbb{R}_+,\ P_t(x,y) > 0$.
- *d-periodic* if it is irreducible and $GCD(n \geq 1 \mid \Pi^n(x,x) > 0) = d$ for some (any) state x.
- *recurrent* if $\forall x \in S,\ \mathbb{P}_x(\exists n > 0,\ X_n = x) = 1$. For a Markov process replace X by ξ and n by t.
- *ergodic* if it is irreducible, 1-periodic, recurrent and has an invariant probability measure. For a Markov process remove 1-periodicity.

Remark 1.1.3.9. The chain associated to a Markov process reversible with respect to μ is $(a_i \mu_i)_i$-reversible.

Proposition 1.1.3.10. *A recurrent Markov chain has an invariant measure and if it is irreducible as well, the measure is unique up to a multiplicative constant.*

Theorem 1.1.3.11. *Let X be an ergodic Markov chain with invariant probability measure μ. Then*

$$\forall x,\ \sum_{y\in S} |\mathbb{P}_x(X_n = y) - \mu(y)| \overset{n\to\infty}{\longrightarrow} 0.$$

Theorem 1.1.3.12. *Let X be irreducible with invariant probability measure μ. Then $\forall f \in \mathcal{F}_m(S, \mathbb{R}_+) \forall x \in S$, $\mathbb{P}_x$-a.s.*

$$\frac{1}{n}\sum_0^n f(X_k) \to \int f \mathrm{d}\mu.$$

Theorem 1.1.3.13 (Kolmogorov equations)**.** *Set $p_t(i,j) := \mathbb{P}_i(X_t = j)$. The $t \mapsto p_t(i,j)$ are $\mathcal{C}^\infty$ and satisfy*

$$p_t'(i,j) = -a_i p(i,j) + \sum_k p(k,j) q_{i,k}.$$

If there is no explosion, then

$$p_t'(i,j) = -a_j p_t(i,j) + \sum_k p(i,k) q_{k,j}.$$

1.1.4 Poisson processes and subordinators

Definition 1.1.4.1.

- A *point process* is a random measure (measure-valued variable) on $\mathbb{R}$, that is the sum of a random number (at most a countable infinity) of Dirac masses at positions given by measurable functions i.e. $\mu(\omega, \mathrm{d}x) = \sum_{k=0}^{M(\omega)} \delta_{T_k(\omega)}(\mathrm{d}x)$.
- The *Laplace functional* of a point process μ is $\mathcal{F}_m(\mathbb{R}, \mathbb{R}_+) \to [0;1] : f \mapsto \mathbb{E}\left[\mathrm{e}^{-\mu(f)}\right]$, where $\mu(f) = \sum_{k=0}^{M} f(T_k)$.
- A *Poisson point process* is a point process whose Laplace functional is of the form $\exp\left(-\int \left(1 - \mathrm{e}^{-f(x)}\right) \nu(\mathrm{d}x)\right)$, where ν is the associated measure.
- The *intensity measure* of a point process μ is $\nu(\mathrm{d}x) = \mathbb{E}[\mu(\mathrm{d}x)]$.
- A *subordinator* is an increasing process X such that the increments $X_{t+s} - X_t$ are independent of the process $X_{[0;t]}$ and distributed as X_s.
- An *α-stable* subordinator is one whose Laplace transform is $\Psi(\lambda) = c\lambda^\alpha$.

Theorem 1.1.4.2. *The measure associated to a Poisson point process is its intensity measure.*

Theorem 1.1.4.3 (arcsine law)**.** *Let X be a stable subordinator. Then the variable $\frac{X(\inf\{t | X(t) > x\}^-)}{x}$ has the density $\frac{\sin(\pi\alpha)}{\pi} u^{\alpha-1}(1-u)^{-\alpha}$ on $[0;1]$.*

[15] All measures are supposed positive, never infinite and not identically zero.

1.1.5 Stable laws

Definition 1.1.5.1.

- We call a *stable law* the non-degenerate law of a variable $Y : \Omega \to \mathbb{R}$ such that if $Y_1, \dots, Y_n$ are IID[16] with the same law as Y, then $\exists((a_n),(b_n)) \in \left((\mathbb{R}_+^*)^{\mathbb{N}^*} \times \mathbb{R}^{\mathbb{N}^*}\right), \frac{Y_1+\dots+Y_n-b_n}{a_n} \overset{law}{=} Y$.
- L is a *slowly varying function* if $\forall t > 0, \lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$.

Theorem 1.1.5.2 (Gnedenko, Kolmogorov)**.** *The following assertions are equivalent for a variable Y that is not Gaussian.*

i Y *has a stable law.*

ii *There are (X_n) IID and $(a_n),(b_n)$ such that $\frac{\sum_{k=1}^n X_k - b_n}{a_n} \overset{law}{\to} Y$ (we can replace "law" by "a.s.").*

iii *The Fourier transform of Y is (after a translation and dilation) given by*

$$\Phi(t) = \mathrm{e}^{-|t|^\alpha(1+i\kappa.\mathrm{sgn}(t)w_\alpha(t))}, \tag{1.1}$$

where $0 < \alpha < 2$ and

$$w_\alpha = \begin{cases} -\tan\left(\frac{\pi\alpha}{2}\right) & \text{if } \alpha \neq 1 \\ \frac{2}{\pi}\ln|t| & \text{if } \alpha = 1. \end{cases}$$

Furthermore, the existence of a non-degenerate law for Y as in ii is equivalent to having

a $\exists \lim_{x\to\infty} \frac{\mathbb{P}(X_1>x)}{\mathbb{P}(|X_1|>x)} = \frac{\kappa+1}{2}$

b $\mathbb{P}(|X_1| > x) = x^{-\alpha}L(x)$ *with L slowly varying.*

Remark 1.1.5.3. As we will be mainly interested in $\alpha \in (0,1)$, it is practical to note that b_n are not necessary for ensuring the convergence to a non-degenerate limit if it is possible.

Definition 1.1.5.4.

- The *totally asymmetric* α-stable law is the law whose Fourier transform is the one from Equation 1.1 with $\kappa = 1$.
- We say that a law is in its *domain of attraction* if conditions a and b are fulfilled.

Remark 1.1.5.5. The unwary reader should be alert that such rescaling to a non-degenerate variable is not necessarily possible (be it for the mean or for the maximum – cf. next subsection) and thus the associated domains of attraction do not form a partition of the space of variable laws.

1.1.6 Extreme values

Though we will not use directly the convergence of types, it is useful to state it, as it gives an idea of what we might get as limit distribution of the extreme values. For our purposes what is of greater importance is the scale at which we are certain to find the max.

In this subsection only we shall denote by $(X_n)_{n\in\mathbb{N}}$ a sequence of IID real variables with law $\mathscr{L}$ and by $M_n := \max(X_0, \dots, X_{n-1})$.

Definition 1.1.6.1.

- The *Gumbel distribution* is given by $\forall x \in \mathbb{R}, \mathscr{L}_{GUM}((-\infty, x]) = \mathrm{e}^{-\mathrm{e}^{-x}}$.
- The *Weibull distribution* with parameter α is given by $\forall x \leq 0, \mathscr{L}_{WEI(\alpha)}((-\infty, x]) = \mathrm{e}^{-(-x)^\alpha}$ and $\mathscr{L}_{WEI(\alpha)}(\mathbb{R}_+) = 0$.
- The *Fréchet distribution* is given by $\forall x \geq 0, \mathscr{L}_{FRE(\alpha)}([0, x]) = \mathrm{e}^{-x^{-\alpha}}$ and $\mathscr{L}_{FRE(\alpha)}(\mathbb{R}_-) = 0$.

Theorem 1.1.6.2 (Convergence of types, Fisher-Tippett-Gnedenko)**.** *If $\frac{M_n - b_n}{a_n}$ has a non-trivial limit in law, then it is a translated and dilated (by constants not depending on n) GUM, WEI or FRE.*

The proofs of the following lemmas are very similar and we will only give the last one.

[16] We will say that a set of variables are IID if they are independent and identically distributed.

Lemma 1.1.6.3. *Set* $\mathscr{L} = \mathscr{L}_{EXP(a)}$ *and* $a_n = a^{-1}$, $b_n = a^{-1}\ln n$. *Then* $\frac{M_n - b_n}{a_n} \xrightarrow{law} GUM$.

Lemma 1.1.6.4. *Let* $\forall x \geq 0$, $\mathscr{L}([x,\infty)) = x^{-\alpha}$, $\alpha > 0$, $a_n = n^{\alpha^{-1}}$, $b_n = 0$. *Then* $\frac{M_n - b_n}{a_n} \xrightarrow{law} FRE(\alpha)$.

Lemma 1.1.6.5. *Let* $\mathscr{L}$ *be in the domain of attraction of the* α*-stable law and set* $b_n = 0$ *and* a_n *such that* $\mathbb{P}(X > a_n) = \frac{1}{n}$. *Then* $\frac{M_n - b_n}{a_n} \xrightarrow{law} FRE(\alpha)$.

Lemma 1.1.6.6. *Set* $\mathscr{L}(\mathrm{d}x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x$, $a_n = \frac{1}{\sqrt{2\ln n}}$ *and* $b_n = \sqrt{2\ln n} - \frac{\ln\ln n + \ln(4\pi)}{2\sqrt{2\ln n}}$. *Then* $\frac{M_n - b_n}{a_n} \xrightarrow{law} GUM$.

Proof. Set $F(x) = \mathbb{P}(X_1 \leq x)$. Then integrating by parts we get

$$1 - F(b_n + xa_n) = \frac{1}{\sqrt{2\pi}}\int_{b_n + xa_n}^{\infty} \mathrm{e}^{-\frac{y^2}{2}}\mathrm{d}y \overset{n\to\infty}{\sim} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(b_n+xa_n)^2}{2}}\frac{1}{b_n + xa_n} \sim \frac{1}{\sqrt{4\pi\ln n}}\exp\left(-x - \frac{b_n^2}{2}\right) \sim \frac{\mathrm{e}^{-x}}{n}. \qquad \square$$

Remark 1.1.6.7. Conditioning by M_n we can show that the difference with the runner-up M'_n scales as a_n.

1.2 Physical setting

We will not attempt to give here a self-contained work on spin glass theory (or worse yet, glass theory) and nor does the author believe in the existence of such a reference. In order to avoid a lengthy introduction, we will rather redirect the reader towards the references in Physics prerequisites and introduction to the area. A general qualitative overview of properties of structural glasses and particularly of the structural glass transition can be found in [27] – namely extensive remarks on experimental approaches, history and industrial applications. For a more concrete approach giving more insight of research results the talk [7] is a good start and the article [1] provides greater detail and relevant references. As for spin glasses, see [24]. We give here only a "definition" of spin glass in order to know what we are talking about even though we will exclusively focus on the REM.

Definition 1.2.0.1. A *spin glass* is a system of spins (implicitly 1/2), which interact in a frustrated manner i.e. which present both ferromagnetic and anti-ferromagnetic interactions (for 2 spin interactions) and that in a "disordered" way. We speak of a "glass" when the system is not in equilibrium and relaxation times are macroscopically large (or more).

Let us just mention a few of the key words found in the domain, which are of great interest and have seen notable development over the last 40 years:

- Glass transition – dramatic slowdown of dynamics along with emergence of complex types of order.
- Links to theoretical informatics, namely optimisation problems, for instance max cut.
- Annealing – relaxation and gradual switching on of constraints that are difficult to fulfil.
- Replicas – a mysterious but powerful method for tackling otherwise very complicated problems, which has proved to be particularly elusive for mathematical justification and is poorly understood.
- Belief propagation – using the locally tree-like percolation structure by assembling neighbours' "information".

We will not dwell more on context, as the next chapter will provide a bit more.

Chapter 2

Introduction

2.1 Definitions and notations

We introduce in this section the models and notations that will be used further.

2.1.1 Dynamics

Definition 2.1.1.1. We fix $n \in \mathbb{N}$ destined to go to ∞ and $G = (S, Ed)$ an n-regular simple graph.

- We call *Gibbs measure* (at temperature $T = \beta^{-1} \in [0; \infty]$) associated to an energy function $E : S \to \mathbb{R}$ the one given by $\mu_s = |S|^{-1} \mathrm{e}^{-\beta E_s}$ and $Z \in [0; \infty]$ the associated partition function. The only case that will be considered is $(E_s)_{s \in S}$ being IID (unless otherwise stated) with density ρ (if defined).
- *Glauber dynamics* is a Markov process (or chain) X reversible with respect to the Gibbs measure when conditioned by E. We actually need a product universe (product σ-algebra, product filtration, etc.) on which are defined both the random energies and the Markov process. We shall refer to $\mathbb{P}_E := \mathbb{P}(\bullet \mid E)$ as the *quenched probability*.
- *Metropolis dynamics* is given by the transition rates $q_{i,j} = n^{-1} \exp(-\beta \max(E(j) - E(i), 0)) \mathbb{1}_{\{i,j\} \in Ed}$.

Dynamics are used to study out-of-equilibrium properties of thermodynamic systems, in our case, aging, which is way out of reach for equilibrium analysis. What is more, the very definition of glasses often refers to relaxation, which is ill-defined without a microscopic dynamics mechanism[1].

Definition 2.1.1.2.

- t shall be the time scaling with n, ($t = n^\kappa$ or $\mathrm{e}^{\kappa n}$ by default). Do note that all times considered are only orders of magnitude – constants are irrelevant. Moreover, they are often correct just to the leading exponential order.
- N shall denote the number of states visited until t.
- E_{min}, E_{max} shall denote the extremes of the process $E \circ X$ until t. In text we will refer to them as *min* and *max* respectively. This should not be confused with the *global min* (*max*) on all the states.
- The dependence on t and therefore on N will be kept implicit. We may consider t as a function of N instead as well (fixing a number of jumps and setting $t = T_N$ or rather for the number of states)[2].

Remark 2.1.1.3. Although we will pretend that there is no difference between the number of jumps and the number of states visited (in our model from chapter 5 there would be just a $\mathrm{e}^{\beta\sigma} = \mathrm{e}^{\beta\sqrt{\frac{n}{\ln n}}}$ factor between them), if any rigour is attempted, it is crucial to pay the attention due to the difference. It actually turns out that in the REM with Metropolis dynamics the number of jumps is dominated by the rare event of neighbouring very low energy states.

[1] Some properties such as condensation of the Gibbs measure may be tackled without dynamics.
[2] This, however, could have consequences on aging functions, for instance.

2.1.2 REM

Definition 2.1.2.1.

- The *REM* (Derrida's Random Energy Model) is defined by its state space $S = \{0,1\}^n$ – the n-dimensional hypercube with its edges – and its IID energies distribution $E \overset{law}{=} \mathcal{N}(E_0)$ i.e. $\rho(x) = \frac{1}{\sqrt{2\pi n}} \mathrm{e}^{-\frac{x^2}{2n}}$.
- We shall denote by I the (typical) interval in which are located n such independent energies.[3]

The REM was introduced as a simplified version of usual spin glass models such as the p-spin (It can namely be regarded as its the infinite p limit.) and induced great interest due to its being analytically solvable (is some respects) and having the same qualitative behaviour as what was known for other models. One of its most important features is to be solvable *without* the use of questionable methods like the replica one and thus allowing a tractable way for studying the effects of the replica magic. The intuitive idea behind the REM is that we approximate the energy landscape of an n-spin (1/2) system with frustrated and complex interactions by a completely random one which neglects any correlation between close states – a single spin flip suffices for decorrelating from the initial energy. The normal distribution is there mostly for convenience though as we will see further the choice of the tail can lead to dramatic changes of the properties.

2.1.3 BTM

Definition 2.1.3.1.

- The *BTM* (Bouchaud Trap Model) is defined by a random *trapping times* function $\tau : S \to \mathbb{R}_+$ or rather by its marginal distribution, as it is assumed that τ_s are IID. The states are called *traps*.
- The associated *BTM dynamics* is defined by the transition rates $q_{i,j} = n^{-1}\tau_i^{-1}\mathbb{1}_{\{i,j\}\in Ed}$.
- If an energy function is available, we use $\tau(s) = \mathrm{e}^{-\beta E(s)}$.
- Unless otherwise stated the BTM is on the complete graph K_n and the distribution of τ is in the domain of attraction of the totally asymmetric α-stable law with $0 < \alpha < 1$. For technical reasons we shall assume that $\forall x \geq 1,\ \mathscr{L}_\tau([x;\infty)) = x^{-\alpha}$.
- We define τ_{max} and τ_{min} as in Definition 2.1.1.2.

The BTM was conceived as a simplification of the REM. The intuition goes as follows. In the random energy landscape there are only a few (macroscopically) low energy minima that become important below the critical temperature of condensation of the Gibbs measure to them (see chapter 4). Thus we need to get out of those "traps" in order to move to another one. It would be reasonable to think that getting out requires essentially climbing up to 0 energy (see the next chapter) and that once we are in those high energies, the system evolves very fast until it finds another trap and doing so unbiased by where it was or by the energy of the trap (taking the first found). We should mention that the corresponding n in the BTM is not the same as the dimension of the hypercube it corresponds to (see [11] for more detail).

2.1.4 Aging

Definition 2.1.4.1. We speak of *aging* if a certain two-time correlation function is not invariant by translation in time.

Example 2.1.4.2. One of the most common correlation functions studied in the literature is:

$$\Pi(t_w, t_w + t) := \mathbb{P}_E\left(\forall t' \in [t_w; t_w + t], X_{t'} = X_{t_w}{}^{4}\right).$$

We shall also use another common correlation function similar to Π:

$$R(t_w, t_w + t) := \mathbb{P}_E(X_{t_w} = X_{t_w+t}).$$

As we will see, the hallmark of aging is, loosely speaking, the dynamics becoming slower and slower as time passes up to a certain (very long) time scale (or forever in the step model, for instance).

[3] In fact we will rather use it as a variable depending on the vertex whose neighbours we are referring to, but we will not be that precise.

2.1.5 Additional models

We define here the Edwards-Anderson (EA), p-spin, the Sherrington-Kirkpatrick (SK) and the step models, for reference. Those will not be the object of this study, but as sporadic remarks mention them, they are given here for the reader's reference. We leave the choice of dynamics for these models free but mention that the BTM dynamics is rather irrelevant.

Definition 2.1.5.1.

- The *EA model* is given by a graph $G' = (S', Ed')$[5] (a lattice) and its IID interactions $(J_e)_{e\in Ed'}$ (their choice is not fixed but a standard choice is $\mathscr{L}_{J_e} = \frac{J}{2}(\delta_1 + \delta_{-1})$). Its energy is

$$E(\sigma) = -\sum_{\{i,j\}\in Ed'} J_{\{i,j\}}\sigma_i\sigma_j.$$

- The *p-spin model* is given by its interactions $(J_{i_1,\dots,i_p})_{1\leq i_1<\dots<i_p\leq n}$. It is defined on the hypercube and its energy is

$$E(\sigma) = -\sum_{1\leq i_1<\dots<i_p\leq n} J_{i_1,\dots,i_p}\prod_{j=1}^{p}\sigma_{i_j}.$$

  The distribution of the IID J_i is taken to be the following normal one (J is a constant)

$$\forall i,\ \rho(J_i) = \sqrt{\frac{n^{p-1}}{\pi p!}}\mathrm{e}^{-\frac{J_i^2 n^{p-1}}{J^2 p!}}.$$

- The *SK model* is the 2-spin model.

- The *step model* (*Barrat-Mézard model*) is given on an uncountable state space (we identify states and energies) with a certain density ρ' (often exponential) with the complete graph (no transition is forbidden). The original dynamics (others like Metropolis are also considered) has the following transition rates

$$q_{E,E'} = \frac{\rho(E')}{1+\mathrm{e}^{\beta(E'-E)}}.$$

For further reference on the latter the reader is redirected to [2, 6, 12]. For the others [25, 24] should be a good start.

Remark 2.1.5.2. As it was brought to the author's attention the model introduced in Chapter 5 and especially its simplified version in Appendix A may resemble the step one. However, usual viewpoints on it are extremely different from ours (including features like different graph, dynamics, energy distribution, zero temperature), the most important difference being the really infinite (uncountable) number of states – not just the infinite limit. Yet, what we will be regarding here is more of a mixture between the trap and step models and in the author's opinion is much closer to a modified trap model.

2.1.6 Models summary

It is interesting to notice how all of these models are linked and how that influences their invention and treatment. Denoting $A < B$ when A was invented chronologically *before* B, $A \subset B$ when B is a *generalisation* of A, $B \supset A$ when A is a *simplification* of B and $A \prec B$ when A has been treated rigorously *before* B (as it concerns aging), we get

$$\text{Ising[19]<EA[16]}\lesssim\text{SK[31]<}p\text{-spin[13, 18]<REM[13]<BTM[8]<step[2]},$$

$$\text{Ising} \overset{J_e=1}{\subset} \text{EA} \overset{\text{mean field}}{\supset} \text{SK} \overset{p=2}{\subset} p\text{-spin} \overset{p\to\infty}{\supset} \text{REM} \overset{\uparrow\{s|E(s)\ll\inf I\}}{\supset} \text{BTM}.$$

Finally, denoting (A, B) the dynamics B on the model A, we have (BTM,BTM)[9]$\prec$(REM,BTM)[3]$\prec$(REM,Metropolis)[17]. It is this last order that we will tend to follow in this paper, as it represents the increasing difficulty. We might also want to add to the picture the (BTM-hypercube,BTM) and (BTM-normal energies,BTM)[26] and the other variations whose combinatorics is left to the reader. The general pattern is that K_n is the "simplest" graph, the exponential energy distribution is the "simplest" distribution and BTM is the "simplest" dynamics.

[5] This is a geometrical graph representing the bonds between sites and not the abstract state space graph G we used earlier.

2.2 Basic observations and properties

2.2.1 BTM

Dynamics Notice that the Markov chain associated to the BTM dynamics is a simple random walk on G and $a_s = \tau_s^{-1}$, meaning that at each step we wait an exponentially distributed time with mean τ_s at the trap s and then jump to a randomly chosen neighbour completely disregarding the landscape (τ or E). This is somewhat the simplest reasonable dynamics possible. Hence, the process is often disassembled into a random walk and a clock process, both being simpler to apprehend. Furthermore, on the complete graph the random walk becomes completely trivial – it is a sequence of IID uniform variables on the states. As we will see further it is often possible to simplify even further, disregarding the walk and considering a clock that is the sum of IID variables drawn from the distribution of τ each multiplied by an independent $EXP(1)$ (step model with BTM dynamics).

Heavy tails Let us discuss the tails hypothesis, which might have evaded the reader's attention. The importance of the distribution tail being an α power law with $0 < \alpha < 1$ cannot be stressed enough. This artificial choice is absolutely essential and all of the model's properties are dictated by it. We will see that in certain regimes we can view the normal distribution as "locally an exponential one" (Subsection 4.2), thus making it more natural. This hypothesis is needed in order to have an infinite mean of the trapping times, thus causing the system to be impossible to explore entirely on the "macroscopic" time scale given by $|S|$ and resulting in aging. As we will see in 3 we need $|S|^{\frac{1}{\alpha}}$ instead. With that stated, we can make a handy observation.

Lemma 2.2.1.1. *Set $E \overset{law}{=} -EXP(\alpha)$. Then $\tau := \mathrm{e}^{-\beta E}$ has an α/β-exponent tail.*

Proof.

$$\forall f \in \mathcal{F}_b(\Omega),\ \int_{\mathbb{R}_-} f(\mathrm{e}^{-\beta E})\alpha \mathrm{e}^{\alpha E}\mathrm{d}E = \int_{[1;\infty)} f(\tau)\frac{\alpha}{\beta}\tau^{-\alpha/\beta}\frac{\mathrm{d}\tau}{\tau} \qquad \square$$

This property establishes the link between exponential variables and heavy tails. Thus whenever we want a heavy tail for τ we will take an exponential distribution for E. This property also has the virtue of being a "good" way of sampling heavy tails in "practice" (Monte Carlo based simulations), as exponential laws are readily sampled.

Non-returning An essential aspect of the BTM is its non-returning property. This states that on scales such that $N \ll |S|$ the process has no chance of returning to its starting point. This observation is completely trivial to prove but absolutely crucial for us. For the time being it is more important to see clearly what we are *not* stating here. We do *not* suggest that (and it is false) the process does not revisit states on those scales. It is actually easy to prove (bounding as crudely as possible) that it does not do so until $\sqrt{n}$ and is certain to do so afterwards.

Yet, we will extensively (and more often than not implicitly) use that it is as though we draw a new independent variable at each step. What gives us that right (in most contexts) is the fact that in total the revisits are rare (a zero fraction of N at $n \to \infty$) because the probability of going back to the initial state is small and the revisited states are chosen independently (this is needed because otherwise we might only revisit few states but many times amounting to a non-zero fraction). In this line of thought the revisits being independent and uniform (if we do not specify the order in which the states were visited), if we want to calculate an average on the walk, for example, means that even less corrections are needed (at sub-dominant orders), because the revisits average themselves out (they give the same mean as the rest). It is worth noticing that we are not being rigorous here, but nor could we at this degree of generality, as it is clear that whether we can act as though the process is equivalent to an IID with the original distribution or not depends on the property we are interested in – for a trivial counter example, given what was said, consider the probability of having the same trapping time at two different times separated by at least one jump. Ideally, whatever the property, at sufficiently short time scales (or N scales) the walk should not be able to determine that it is on a finite sample of the distribution and not on the distribution itself (this is somewhat a motivation for the step model).

2.2.2 REM

Non-returning We might want to mention as well that the non-returning holds for the BTM on the hypercube or, equivalently, the REM with BTM dynamics (as the energy landscape is irrelevant for returning with BTM dynamics) – it is, in fact, just a property of an ordinary unbiased random walk. That is a bit more tricky to establish (see [11]) but the idea is based on the fact that the binomial coefficients $\binom{n}{k}$ increase greatly up to $k = \lfloor \frac{n}{2} \rfloor$. This allows one to prove that it is very likely to increase the distance (induced by the 1-norm) to the initial state up to $\lfloor \frac{n}{2} \rfloor$ and then that it is very likely once we are at that distance to go back to a state at the same distance before returning to the origin. A

far more heuristic (and certainly very abusive) way of seeing it would be go through the non-returning on $\mathbb{Z}^n$, which is better-known.

In the Metropolis REM the situation with returning is more complex. This is natural, as the states visited become dependent on the energies, which they are not with BTM dynamics. Here we can expect for low energy states to be "attractive" and so making returning easier (provided we start in such a state). Generally, as we are mostly in low energy states (as we shall discuss later), we can easily find our way back, given that the dynamics is constantly pulled down. Namely, if the trap we are in is the only lower neighbour of its own lowest neighbour, then we are nearly certain to jump to that neighbour and then go back (a few times). As this scenario is not unlikely, once we are likely to be in a trap (as we shall discuss in chapter 5, that is very soon after the start), returning to the current state is not unlikely.

However, as we defined it (returning to the *starting* state) returning is still very unlikely in this setting because of the initial uniform distribution – there is very little chance that we start at a trap and the odds of finding exactly the same high energy state any time soon after we left it are very small as well (only traps are potentially easily regained). However, that should not make us overconfident, as it is purely a property of the initial distribution (in the BTM dynamics case the initial distribution was completely irrelevant). What is important is that returning to the state we are in at t for t not very small is not impossible, as we said. In particular we cannot use the same arguments as before, saying that we draw a new energy at every new jump – we need to retain a probability of going back to the last pool we were in.

Yet, as this property is very handy, we might want to find a way to retrieve it in an "effective" sense. Namely, if we group together low energy states and their "domains of attraction" (we remain intentionally vague), we can hope that the resulting "traps" will not favour returning (by construction). This heuristics is more or less the same one as the one permitting to reduce the REM to the BTM.

Remark 2.2.2.1. Let us repeat that the fact (which is true) that there is no return i.e. returns to the starting position for a simple random walk on the hypercube does not at all state (and it is false) that the dynamics does not self-intersect. Moreover, typically after n (compare with the $\sqrt{|S|}$ for K_n) jumps we already have jumped right back to a state we had just left - the thing is that we do not specify when.

Neighbours In the infinite n limit by the extremes we have $I \approx \left[-\sqrt{2n\ln n}; \sqrt{2n\ln n}\right]$ and typical differences between different $\inf I$ are $\simeq$ [6] $\sqrt{\frac{n}{\ln n}}$ as well as the difference between the lowest and the second lowest energies. Thus, most of the states see exactly I as their surrounding – with one minimal neighbour at $\inf I$ (with uncertainty of order $\sqrt{\frac{n}{\ln n}}$), a few other nearly lowest energies spaced by about $\sqrt{\frac{n}{\ln n}}$ as well, many states sampling the Gaussian from there up to about $\sup I$ and, finally, the max and the few highest energies (behaving at the inverse of the lowest ones). Of course, there are rare exceptions, but as long as we do not push things too far considering events that concentrate on such rare cases, we will argue as thought all the states saw this same scenery.

[6] $\simeq$ shall denote "is of the same order in n as" i.e. "$\ldots = \Theta(\ldots)$".

Chapter 3

BTM

3.1 Preliminary remarks

As the BTM's original purpose is to effectively describe the dynamics of other models, it would be out of place to look at its equilibrium properties, which are rather clear anyway. From Lemma 2.2.1.1 we see that a phase transition will occur at $\beta = \alpha$, as the Gibbs measure will stop giving a strictly positive fraction of the mass to single states.

Trapping times In order to see why we associate a trapping time that is given by $\tau = \mathrm{e}^{-\beta E}$, we will consider the Metropolis dynamics on the REM (for simplicity, set $\rho(x) = \mathrm{e}^{-\frac{x^2}{2}}$). Let $i \in S$ be a local minimum of the energy which is low in energy (what we would call a trap). Then the exit rate is the sum of the transition rates to its neighbours, that is

$$a_i = \frac{1}{n} \sum_{j,\{i,j\}\in Ed} \mathrm{e}^{-\beta(E(j)-E(i))} \overset{n\to\infty}{\approx} \int_{E>E(i)} \mathrm{e}^{-\beta(E-E(i))}\rho(E)\mathrm{d}E = \mathrm{e}^{\beta E(i)}\mathrm{e}^{\frac{\beta^2}{2}} \int_{E>E(i)} \rho(E+\beta)\mathrm{d}E \overset{E(i)\ll -1}{\approx} \mathrm{e}^{\beta E(i)}\mathrm{e}^{\frac{\beta^2}{2}}$$

and thus the average waiting time at the trap is proportional to τ.

Remark 3.1.0.1. Note that this is only true for traps and not at all for high energy states. Therefore, it is not viable to use the BTM to study the max (we will look into this further on).

3.2 τ_{max} and aging

3.2.1 τ_{max}

Let us discuss the evolution of $\tau_{max}(t)$. First of all, we have already seen that $\tau_{max}N^{-\frac{1}{\alpha}} \overset{law}{\to} FRE(\alpha)$. Technically, we actually proved that for independent variables drawn from the distribution of τ and not from the discrete realisation of the sampling we have (recall that at fixed n we have n values of τ and we indifferently chose one of them at each step rather than drawing a new independent τ). However, as long as we are on a scale $N \simeq n^\gamma$, $0 < \gamma < 1$ the non-returning property assures that only a zero fraction of the jumps were towards already visited states and thus the convergence still holds.

We now show that t and τ_{max} are of the same order in the sense $\frac{\ln \tau_{max}}{\ln t} \overset{\mathbb{P}}{\to} 1$, where $\overset{\mathbb{P}}{\to}$ denotes convergence in probability.

Proof. Suppose that $\tau_{max} > n^\varepsilon t$ with non-zero probability (at $n \to \infty$ after subsequence extraction) by which we condition. Then $N^{\frac{1}{\alpha}} > n^\varepsilon t\epsilon$ with high probability for ϵ sufficiently small. Consider the first $Nn^{-\varepsilon'}$ jumps with $\varepsilon' < \frac{\alpha\varepsilon}{2}$. Among them we should have $\tau'_{max} \simeq \left(Nn^{-\varepsilon'}\right)^{\frac{1}{\alpha}}$ and with high probability the process spent at least ϵ' of the expected time in the state corresponding to τ'_{max}. Hence $t > \epsilon'\tau'_{max} \gtrsim n^{\varepsilon/2}t\epsilon\epsilon'$ – contradiction.

Suppose that $\tau_{max} < n^{-\varepsilon}t$ with non-zero probability (at $n \to \infty$ after subsequence extraction) by which we condition. We then have $N \simeq (tn^{-\varepsilon})^\alpha$ and thus $t \lesssim (tn^{-\varepsilon})^\alpha \int_1^t \tau \mathrm{d}\mathscr{L}_\tau = \alpha t n^{-\varepsilon\alpha}$ – contradiction. □

This means that a macroscopic part of the time is spent in a single state and that holds at any time scale (before exploring the whole system). In particular the exploration of the whole system (in the sense of discovering that it is finite by finding the global minimum) takes about the global max of τ. Hence, according to the extremes it is of order $n^{\frac{1}{\alpha}}$. Moreover, we get $\tau_{max}(t) \approx t$ and so $E_{min} \approx -\beta^{-1}\ln t$.

3.2.2 Aging

The previous subsection's arguments should be sufficiently convincing for one to believe that until that time scale we will observe aging – if the deepest trap takes us a time of order t to get out, then there should be a strictly positive probability that we are still there and thus a strictly positive probability that we stay there another θt for θ a constant (of order 1). This intuition is indeed right – more precisely, one can show the following[4].

Theorem 3.2.2.1. *If $t_w = n^\kappa$, $0 < \kappa < \frac{1}{\alpha}$ and $\theta \in \mathbb{R}_+$, then for a.e. E*

$$\lim_{n\to\infty} \Pi(t_w, t_w(1+\theta)) = \int_0^{\frac{1}{1+\theta}} \frac{\sin(\alpha\pi)}{\pi} u^{\alpha-1}(1-u)^{-\alpha}\mathrm{d}u.$$

Though we will not prove this last statement, in order to give the reader an idea, we give the main steps (we follow the proof in [4] and not the initial renewal theory one). At the fixed time scale the traps are divided into three groups – the shallow ones (shallower than the deep ones), the deep ones (of the order of the time scale with a factor between $\epsilon \to 0$ and $M \to \infty$) and very deep ones (deeper than the deep ones). The very deep traps are shown to be too unlikely to have been visited and the shallow ones are shown not to contribute. Finally, the deep ones are shown to account for a clock process which is well approximated by an α-stable subordinator, which yields the result using the arcsine law for stable subordinators.

There is actually not much else to say about this model (besides on the longest and shortest time scales for example, but that does not concern us much) once we know that it is approximately a stable subordinator, most of the questions we may ask are reduced to classically known ones.

3.3 Bare, process and chain measures

We turn our attention to an aspect that is not particularly attached to the BTM, but we need to make the distinction between the three natural measures we have at our disposal, as we constantly use all of them and the sooner we make a clear point on them, the better. We will come back to this in subsection 5.3.2 in a more precise setting, but let us give some generalities here.

3.3.1 Preliminaries

Definition 3.3.1.1. In this subsection only we shall use the following names for the measures.

- The *bare* one is the measure μ_b from which we have drawn E independently for each state.
- The *process* one is the measure (depending on t) given by $\mu_p(\{s\}) := \int_0^t \mathrm{d}t' \mathbb{1}_{\{s\}}(X(t'))$.
- The *chain* one is the measure (depending on N thus implicitly on t) given by $\mu_c(\{s\}) := \sum_{k=0}^N \mathbb{1}_{\{s\}}(\xi_k)$.

We call $\mathscr{L}_p := E(\mu_p)$ the image (push forward) measure by E and the same with index c.

Example 3.3.1.2. The *bare* measure is the one whose density we called ρ and in our setting it is exponential or Gaussian.

Remark 3.3.1.3. μ_p can also be seen as the image measure of $l_{\restriction[0;t]}$ by the dynamics X. Similarly, μ_c is the image of $\#$[1] on $[0;N]$ by the discrete dynamics chain ξ.

Remark 3.3.1.4. Another way to put it is that μ_p is the weight as seen by the Markov process – the time sum, while μ_c is the associated Markov chain – the jump sum. Thus the first one may be seen as the second weighted by the (random) mean (on the occurrences in the dynamics up to t) waiting times.

Remark 3.3.1.5. The weary reader has doubtlessly noticed the link with the problem from Remark 2.1.1.3 and is right to note that we will be talking rather of yet another measure given by $\mu_n(\{s\}) := \mathbb{1}_{(0;\infty)}(\mu_c(\{s\}))$ – the new state measure, whose $\mathscr{L}_n$ is simply $\#(\bullet \cap E(X([0;t])))$[2]. However, we will stick to our convention not to consider the difference between the n and c indices.

[1] $\#$ shall henceforth denote the counting measure ($A \mapsto |A|$) on any space and in any context.

[2] This is true in the a.s. event of no energy coincidences. Moreover, it is of finite mass in the a.s. event of no explosion of the dynamics process, which ensures a finite number of jumps and a good definition of X at all t.

3.3.2 General properties

Let us outline the general (true for both REM and BTM, both Metropolis and BTM dynamics) properties of those measures. The very definition of Glauber dynamics demands for the process measure to converge towards the Gibbs one μ_g on the longest time scale consisting in taking $\lim_{t\to\infty}$ first and (eventually) $\lim_{n\to\infty}$ only afterwards. Clearly, in the part that has not yet been accessed we have $\mathscr{L}_c = \mathscr{L}_p = 0$. We also know that at the part G of the energy spectrum that is in equilibrium (visited many times close to all its energies) at time t the laws verify $\mathscr{L}_{g\restriction G} = \mathscr{L}_{p\restriction G} = a_E^{-1}\mathscr{L}_{c\restriction G} = \mathscr{L}_b$, where a_E^{-1} is the mean waiting time at energy E. This can be seen to hold by a law of large numbers argument – the number of states visited with energy close to E is big so we have had the chance to mean out fluctuations of the state dependency of the exit rates as well as the fluctuations of the exponentially distributed exit times in order to find only the mean (on the states and on the exit times) a_E. Moreover, still in G but also in the zone H where $a_E \propto \mathrm{e}^{\beta E}$, we have $\mathscr{L}_c \propto \mathscr{L}_b$. Finally, the area close to the borders E_{max} and E_{min} is more complicated – there we do not have many states yet – the measures are what is called *granular* – and thus we need a more specific approach. Essentially the energies visited form a Poisson process, but we will look into this further, as the consideration in subsection 4.1.3 will certainly make things clear enough. Yet, if we allow ourselves to take the mean over the realisations of the energy landscape E and we cut at a jump time, we should have $\mathbb{E}[\mathscr{L}_p(\mathrm{d}E')] = \mathbb{E}[\mathscr{L}_c(\mathrm{d}E')a_{E'}^{-1}]$ everywhere.

3.3.3 Model specific features

Let us now give the model dependent aspects.

- The a_E we mentioned is, of course, dependent on the dynamics we chose. Namely, for BTM dynamics it is $\mathrm{e}^{\beta E}$, while for Metropolis it is (as we shall see in chapters 5 and 6) $\frac{1}{n}\mathrm{e}^{-\beta(\inf I - E)}$ for energies lower that $\inf I$, 1 for energies higher than $\sup I$ and of order 1 in between (with the exception of a transitory zone from $\inf I$ to energy scales of order $\sqrt{n}$).
- H is all $\mathbb{R}$ in the BTM dynamics, while with Metropolis dynamics on the REM $H = (-\infty; \inf I]$.
- For the BTM dynamics (where this is of little importance), we can use the non-returning to justify that $\mathscr{L}_c = \mathscr{L}_n$.

Chapter 4

REM

In this chapter we summarise a few classical properties of the REM, that will be of use for the understanding and the motivation of what is to follow.

Definition 4.0.3.1. We define the critical inverse temperature by $\beta_c := \sqrt{2\ln 2}$ and set $\beta_i = \min(\beta_c, \beta)$ for further use.

The first thing to notice is that according to the extremes the global min is $\approx -\beta_c n$.

4.1 Thermodynamic properties

As the reader has doubtlessly noticed we are not particularly interested in the equilibrium of the models, as is natural for glasses, which are for all practical purposes out of it and usually far from it. Yet, we will take advantage from the fact that we will need some calculations in the dynamics regime which yield the equilibrium values as particular cases. Thus, we regroup what is available throughout the text (mostly in chapter 5), but also give a more rigorous approach.

4.1.1 Phase transition

Let us begin with two results from chapter 5. We will not seek to understand how they are established just yet, but only interpret them.

- According to Equation 5.1 the mass given by the Gibbs measure to the ground state is around

$$\frac{\mathrm{e}^{\beta\beta_c n}}{\frac{2^n}{\sqrt{2\pi n}}\int_{-\beta_c n}^{\infty}\mathrm{e}^{-\beta E}\mathrm{e}^{-\frac{E^2}{2n}}\mathrm{d}E} \approx \begin{cases} \exp\left(\left(\beta_c\beta - \left(\ln 2 + \frac{\beta^2}{2}\right)\right)n\right) & \text{if } \beta < \beta_c \\ \frac{1}{\sqrt{2\pi n}(\beta+\beta_c)} & \text{if } \beta > \beta_c. \end{cases}$$

  This can also be seen directly, considering that the energies are given by the Gaussian truncated at $-\beta_c n$, which is where the ground state is. We should keep account of the fact that this cut-off is rather imprecise as is the usage of the value $-\beta_c n$ as value for the minimum (recall the corrective terms from the extremes). That imprecision leads to the incorrect result at $\beta > \beta_c$, as we can only argue that it is $\Omega\left(\mathrm{e}^{an}\right)$ for all $a < 0$. What we actually show in this way is that

$$\frac{\ln\max\mu}{n} = \begin{cases} -\beta_c\beta + \ln 2 + \frac{\beta^2}{2} & \text{if}\beta < \beta_c \\ 0 & \text{if}\beta > \beta_c \end{cases}$$

  is not twice differentiable at β_c. We will see this more neatly and in more detail in the next subsection.

- According to Equation 5.2 the average energy is given by $E_{mean} = -n\min(\beta, \beta_c)$. Again this allows to see the second order transition, as the mean intensive energy $\frac{E_{mean}}{n}$ is not differentiable at β_c. This is what is called the glass transition of the REM (the glass is the low temperature phase).

It is clear that from a physical point of view the second discontinuity is more important than the first one, while the concentration of the measure on single states is certainly more exciting mathematically. Needless to say, we will emphasise on the concentration phenomenon.

4.1.2 Condensation of the Gibbs measure

We follow closely or even just detail the proof of [21] for a rigorous approach to the condensation. Let $\mathscr{N}(\beta) = \#\{s \in S \mid E(s) \leq -\beta n\}$ be the number of states with energies $\leq -\beta n$. We have $\mathbb{E}[\mathscr{N}(\beta)] \sim \frac{|S|}{\beta} \mathrm{e}^{-\frac{n\beta^2}{2}}$ for $\beta > 0$. Hence,

$$\sum_{n=1}^{\infty} \mathbb{P}(\mathscr{N}(\beta) \geq 1) \leq \sum_{n=0}^{\infty} \mathbb{E}[\mathscr{N}(\beta)] < \infty$$

for $\beta > \beta_c$ and by Borel-Cantelli a.s. $\mathscr{N}(\beta) = 0$ for large n.

Furthermore, for $0 < \beta \leq \beta_c$ we get by independence

$$V(\mathscr{N}(\beta)) = 2^n V(\mathbb{1}_{E(s)\leq -\beta n}) = 2^n \left(\mathbb{P}(E(s) \leq -\beta n) - (\mathbb{P}(E(s) \leq -\beta n))^2\right) \sim \frac{1}{\beta} 2^n \mathrm{e}^{-\frac{n\beta^2}{2}}.$$

By Chebyshev's inequality for large n

$$\mathbb{P}(|\mathscr{N}(\beta) - \mathbb{E}[\mathscr{N}(\beta)]| > \varepsilon) \leq C \frac{2^n \mathrm{e}^{-\frac{n\beta^2}{2}}}{\varepsilon^2}$$

and so by Borel-Cantelli a.s. $\ln \mathscr{N}(\beta) = \ln \mathbb{E}[\mathscr{N}(\beta)] + o(1) = \frac{n}{2}(\beta_c^2 - \beta^2) - \ln(\beta) + o(1)$, taking $\varepsilon = \left(2^n \mathrm{e}^{-\frac{n\beta^2}{2}}\right)^{\frac{2}{3}}$. These results should be understood as a condensation of the Gibbs measure to a small (not exponential with n) number of states (recall that if we introduce the Boltzmann factor in the Gaussian density, we get another Gaussian centered at $-\beta n$) – in the high temperature phase many states are involved (those around $-\beta n$ are about as many as those below $-\beta n$, as we can see if we apply two times the result for $\mathscr{N}$), while in the low temperature one only the lowest energies count.

4.1.3 Lower edge

Let us now consider the lower edge of the distribution of the energies more precisely (we still keep close to [21]). To that end, using what we saw at the extremes, we introduce the point process given by the energies and center it where the lower edge is i.e. we set $\Xi = \sum_{s\in S} \delta_{E(s)-a_n}$ with $a_n = \beta_c n - \frac{\ln n}{2\beta_c}$. We want to prove that its Laplace functional converges to a Poisson point process' (which implies process convergence). We have for $f \in \mathcal{F}_m(S, \mathbb{R}_+)$ with compact support

$$\mathbb{E}\left[\mathrm{e}^{-\Xi(f)}\right] = \left(1 - \mathbb{E}\left[1 - \exp(E(s) - a_n)\right]\right)^{2^n} \sim \left(1 - 2^{-n} \int \left(1 - \mathrm{e}^{f(x)}\right) \mathrm{e}^{-\beta_c x} \mathrm{d}x\right)^{2^n} \to \exp\left(-\int \left(1 - \mathrm{e}^{f(x)}\right) \mathrm{e}^{-\beta_c x} \mathrm{d}x\right),$$

which is exactly the Laplace functional of a Poisson point process with intensity measure $\mathrm{e}^{-\beta_c x}\mathrm{d}x$.

Note that this result zooms around the lower edge on a scale of order 1, which is very small compared to the usual n scale – the order of the position of the edge. What it means is that a_n is indeed the typical position of the min up to a fluctuation of order 1, the next lowest energies are also around the same distance away and that we find an exponentially increasing number of states.

4.2 Dynamics

As we will discuss the dynamics of the max in length in the next chapter, we will not say too much about the dynamics in general here but a few things that are good to keep in mind for what follows.

An important point is to explain why we expect to observe aging despite the different setting. Recall that in the BTM an essential feature that induced aging was the heavy tailed waiting time or, otherwise stated, the exponential tail of the energies. In the REM, the Gaussian will, of course, not produce a heavy tail and one might expect not to observe aging as the lowest energy should not dominate the Gibbs measure and we should not spend most of the time there. Yet, perhaps surprisingly, we find nearly the same aging results for the REM as for the BTM. More precisely, we observe "interrupted aging" with an effective time-dependent exponent, as we will explain.

Let us explain how this fact should be understood heuristically. With time the dynamics explores more and more of the configuration space and consequently of the energy distribution available in the sample. Therefore, at a fixed time scale the dynamics is only aware of what it has seen of the energy distribution and ignores the real Gaussian tail, as it has not yet "seen" it. That is why the dynamics behaves as though we had an exponential tail but with an exponent that depends on the cut-off up to where we have explored the energy distribution and thereby depends on time – the Gaussian is "locally exponential-tailed". Yet, this dependence is slow and allows us to see the dynamics given by the current exponent establish well before we go to the "next" one.

We should note that this mechanism cannot go on forever. There are two potential reasons for it to be "interrupted" – either we reach the end of the distribution i.e. we find the global min while aging and clearly it cannot continue, or we explore enough of the Gaussian to understand that the tail is not heavy. This reasoning turns out to be correct, as the Gaussian "becomes less heavy-tailed as we explore more of it" and, as we will see in the next chapter, depending on temperature, aging can continue as long as we have not explored all the negative energy states or until a scale determined by the temperature itself.

Now let us explain what has been established to get an idea of the potential difficulies. In [26] one can find physicist's proofs of aging on the complete graph with exponential tails and effective aging with gaussian distributions but still on the complete graph. Afterwards were proved results on the REM with BTM dynamics in [3], for example. The only change being the underlying graph (and the more rigorous treatment), the difficulties have "topological" roots, as results such as the non-returning we discussed need to be re-established in this new and more complex context. Finally, [17] proved aging in the REM with Metropolis dynamics. Here the main problem was to handle the numerous "0-energy" states (those close to zero at a scale at most about $\sqrt{n}$), where the dynamics is quite complicated. In all of these settings the same results hold in the limits set by the heuristic considerations we outlined above. More insight will be gained in the next chapter, but the interested reader should consider referring to the bibliography.

4.3 Essentials

We recapitulate here the toolbox we dispose of as the REM is concerned in order to use it in the next central chapter.

- The equilibrium REM goes through a phase transition at β_c. This transition leads to concentration of the Gibbs measure on single states – the lowest energy ones.
- At the (lower) edge the energies available are distributed like a Poisson point process on the scale 1 around the position of the minimum (see the extremes).
- The BTM dynamics REM does not return, while the Metropolis one can only be considered non-returning once we group states together to form "domains of attraction" or once we restrict the dynamics to low energy states.
- If we restrict the dynamics to low energy states, the Metropolis dynamics on the REM behaves like the BTM.
- The REM exhibits interrupted aging corresponding to the Gaussian being "locally exponentially-tailed", where the exponent depends on the exploration of the state space.

Chapter 5

Dynamics of the maximum

This chapter addresses the main theme studied during the internship. We will begin with simple or even trivial cases before we move on to the main object of interest, namely the maximum of the energy as a function of time in the REM with Metropolis dynamics. The second section can well be viewed as a warm-up, but before we start, let us explain why we might be interested in the max of the Metropolis REM.

5.1 Motivation

Let us recall how the different spin glass models were coined in the first place. Globally, the movement was from complicated and realistic to simple and easy to deal with models, culminating with the REM. As equilibrium was concerned it was possible (by means of weird techniques or much more elaborate tools recently developed) to solve even the complex models, leaving the easier ones a test ground for methods. Yet, when the sight was turned towards dynamics even the simplest model turned out to be in need of simplifications, as even for it things were very unclear. That gave rise to the BTM and after its rather quick resolution, the inverse path was to be taken. We may safely say that for at least 10 years now the BTM is very well understood and we dispose of varied different techniques for rigorously proving the interesting results. Then it was the REM's turn. As a realistic one, Metropolis dynamics was the physically interesting choice. Yet, the difficulties were only just overcome with [17] and are certainly not yet well comprehended. Nevertheless, now that aging is understood in the Metropolis REM, we might want to expand on that dynamics by looking at more complex features than aging, as the REM is much richer than the BTM, where aging is about all there is to it (without being trivial). Besides, we could hope that the new aspects studied in the REM will behave qualitatively as in more realistic models, but even if that is not the case, they remain interesting, as they would provide an idea of the limitations of the simplification towards the REM.

That is where Biroli's interest in the max comes into play. As we shall see, the max contains all the information about aging and the min as well as much more (the max reflects both tails of the energy distribution), as it continues to vary beyond the usual aging stage. Moreover, we might see the variations of the max as some "extended aging", as we will note deliberately only in the conclusion. Finally, as always with extremes, the interest can be justified with estimation of risk. Let us imagine for a second that the REM were a realistic model representing a physical system (this will be justified if the qualitative behaviour of more realistic models is the same). Then very high energy states would often be undesirable, as they often lead to degrading the system or simply going out of the model. Furthermore, those sudden (as we will see) jumps to and from very high energies would be accompanied by abrupt transfers of much energy from and to the thermostat, which could perturb its normal functioning. In this sense, this may be a way of calculating the lifetime of the system.

To sum it all up, the Metropolis REM is the most realistic model where aging is well understood, the max contains most of the features we are already familiar with, as well as new ones, and may be used to estimate the risk of breaking of the model or waiting times for very rare events.

5.2 BTM

First of all, we should note that for all practical purposes the only important thing about the distribution of τ is its tail. However, obviously, studying the evolution of the max requires knowledge of the high energy end as well ($\tau \to 0^+$). We thus specify the model by assuming that the energies are exponentially distributed in both the positive and the negative regions with parameters giving τ exponents α_+ and $0 < \alpha_- < 1$ at 0 and ∞ respectively. We assume that the number of positive (negative) energy states is n_+ (resp. n_-), both being of order n.

5.2.1 Usual BTM

The conduct under the BTM dynamics is particularly simple once one has understood the evolution and influence of the min. We shall assume that $\beta = 1$, as β only changes α (see 2.2.1.1). Notice that the positive energy states have a mean trapping time and it is of order 1, so as soon as many have been visited, the time spent there is of the same order as the number of such visits. Moreover, by the law of large numbers $N_- = N\frac{n_-}{n} + o(N)$ a.s. (and similarly for N_+).

- If $\alpha_- < 1$, we know that the number of jumps to negative energy states N_- satisfies that $\frac{N_-}{\tau_{max}^{\alpha_-}}$ is of order one by the extremes (those are IID power law variables). Now recall that t and τ_{max} are of the same order, which still holds in this context as we spend less time in positive energy states than in negative[1]. Hence, N_+ is also of order t^{α_-} (passing by N_- and $\tau_{max}^{\alpha_-}$). Thus, once more by the extremes of power laws $\boxed{E_{max} \approx \frac{\alpha_-}{\alpha_+} \ln t}$.
- If $\alpha_- > 1$, the whole distribution has a mean trapping time, so after $t \gg \mathbb{E}[\tau] \simeq 1$ the law of large numbers gives $N_+ \frac{n}{n_+} \approx N \approx \frac{t}{\mathbb{E}[\tau]}$. Moreover, by extremes $E_{max}\alpha_+ \approx \ln N_+$ and combining the two we get $\boxed{E_{max} \approx \alpha_+^{-1} \ln t}$.

We might as well consider a more complicated distribution for positive energies and all we would need to determine in order to find the evolution of the max is its limit extreme value distribution scaling and plug N_+ in it. A final thing to bear in mind is that those expressions are only valid as long as $N \ll n$, when the max stations at the global max.

5.2.2 An additional model

Note that the main reason for the simplicity in the considerations above is that the time spent in positive energies is necessarily negligible compared to that in negative energy states. This property makes the positive energies rather irrelevant to the dynamics itself – they only appear because the question we are asking is about them. We might then want to explore what happens when this is not the case and the positive energy states are the important ones determining the speed of the dynamics.

A way to achieve that is to assume that there are much fewer negative energy states than positive energy ones. Such a setting might suggest to view the negative energy states as impurities, but as will be discussed in the next subsection the physical significance of this model should not be overestimated. We suppose that $n_+ \approx n$, $0 < \alpha_- < \alpha_+$, $\frac{n}{n_-} \ll t \ll n$ (as might be expected, we lose the over-macroscopic relaxation times in this setting) and we consider the BTM dynamics. Within this frame we stay at a state of positive energy an average time of order 1 (it has a mean of order 1). We still have $\frac{N_-}{\tau_{max}^{\alpha_-}}$ of order 1, τ_{max} of order t_- and $N_- = N\frac{n_-}{n}$ (provided the time scale satisfies $N \gg \frac{n}{n_-}$, which we suppose, as otherwise we are certain not to have visited many negative energy states). But then $N_+ \approx N$ and by the law of large numbers $t_+ \sim N \gg (N\frac{n_-}{n})^{\frac{1}{\alpha_-}} \simeq t_-$ (we use $N \ll n$, as $t \ll n$). Then $E_{max} \approx \alpha_+^{-1} \ln t$. However, we do not see any influence of the positive energies exponent on the minimum, but it no longer varies as $\alpha_-^{-1} \ln t$ but as $\alpha_-^{-1} \ln\left(t\frac{n_-}{n}\right)$.

5.2.3 Critics

Except for the simplicity of this section, its main drawback is that it lacks just about any physical sense. That is because the BTM is a good approximation of the REM (which is an approximation of more realistic models) when we are interested in the dynamics in the traps of the REM. One should bear in mind the vertices of the complete graph in the BTM correspond to the deepest wells in the energy landscape of the REM and the maxima of the REM are completely forgotten in the BTM. Therefore, we cannot really expect similar results in the REM. The results on the BTM might still be interpreted as the evolution of the depth of the least deep among the deep states visited[2], even though a precise definition, a way for experimental or numerical detection or an interest of these results seem rather elusive.

5.3 REM

We now consider the REM.

[1] In fact, we spend no time in positive energy states in the sense that over $[0, t]$ the time spent in positive states is on a lower time scale than t – it is in fact of order N_+ and thus of order $t^{\alpha_-} \ll t$.

[2] It is in fact worse than that, as the limiting process leading from the REM to the BTM does not produce positive energies in the BTM. We would then be better off considering a scaling of convergence towards Weibull at 0 with no positive energy state, but the considerations given are closer to what we will see in the next section.

5.3.1 BTM dynamics

If we use BTM dynamics, things are not really complicated. We are, as usual, drawing independent energies from the Gaussian distribution, which means $n_- = n_+$. Like in the previous section the time scale sets the number of jumps via the negative energy tail (this has nothing to do with the max). We then only need to consider the extreme value of N Gaussian variables, which we have already done. What is more, we do not even need to do that, as the Gaussian is symmetric and so the max is just the inverse of the min at a fixed number of trails. Thus, studying the max does not add anything compared to the min.

5.3.2 Metropolis dynamics

It should be noted that Biroli had already obtained Equations 5.4 and 5.5 (among else), perhaps more heuristically.

Let us now examine Metropolis dynamics. The difference is that now the time spent in high energy states is of order 1 (all their neighbours have lower energy and thus the exit rate is 1). What is more, the high energy states are harder to get to, which we would expect to decrease the max.

Transitions Let us first find out what are the different regimes that we should expect in the dynamics that will potentially affect the behaviour of the max (as a function of time). As always, the jump speed will be dictated by negative energy states and given what we have seen on aging, we might expect that any qualitative change of the minimum as a function of time should imply such in the max.

Definitions We consider the following events (*a priori* in an unknown order) that may be relevant to the evolution of the max (depending on the regime in which we are):

a The min becomes smaller than $\inf I$, more precisely the dynamics visits a local minimum of the energy (the second happens a time $\simeq n$ after we go from a state to its lowest neighbour).

b The max becomes larger than $\sup I$.

c The *aging* regime begins (most of the time is spent in the min).

d The aging regime ends.

e The mean energy becomes stationary – the system reaches *macroscopic equilibrium*.

f The system visits the global minimum.

g The system visits the global maximum (the dynamics process has discovered all the states) – the system reaches *microscopic equilibrium*.

Interpretations Let us explain why is each of those changes important:

a From then on the min typically has no lower neighbour and thus is obliged to go up to I, which makes the trap model relevant (strictly speaking it becomes relevant slightly later). Namely from this time on, we can use the trapping time $\mathrm{e}^{-\beta E}$ as an estimate of the time spent in the min.

b Before this time typically every state has a neighbour of energy as high as the max and thus the time required for attaining it is only the time needed for accepting the transition, while afterwards the dynamics needs to explore first in order to find a high energy state and only then wait to make the transition.

c Aging means that we need to investigate only the behaviour of the min and not that of all the states.

d We have to average (in an appropriate sense) over all the states and not only consider the min.

e From then on we can just use the stationary distribution to determine the evolution of the max and no longer need to consider what happens with the min, even if it is still moving.

f From then on the min is constant and equal to the global min.

g From then on the max is constant and equal to the global max.

Estimates We now estimate the times at which these events come about in order to know at each time scale what regime we are observing depending on the time scale (and the temperature as we will see).

a We typically always have exactly one neighbour of energy equal to the typical minimum of n independent energies (it is tautological). As the rate of transiting to it is $1/n$, we need a time of order $\boxed{n}$ to get there. This is discussed in a more serious way in Appendix A.

b As in the previous case we always have a neighbour of that order. Recall that the energy of the max of m independent energies is $\sqrt{n2\ln m}$ as we have seen. We need to visit about $\mathrm{e}^{2\beta\sqrt{2n\ln n}}$ local minima in order to get to the typical max of n independent states (every time we are in a local minimum we are nearly sure we will go to the lowest among the neighbours and the probability of going to the highest among them is the same divided by this factor). Among that many local minima the lowest energy attained should be about

$$\sqrt{2n\ln\left(n\mathrm{e}^{2\beta\sqrt{2n\ln n}}\right)} = \sqrt{2n\left(\ln n + 2\beta\sqrt{2n\ln n}\right)} \approx 2\sqrt{\beta n\sqrt{2n\ln n}} =: \sqrt{2}E_1.$$

We know that we wait at each one a time such that $\ln t = -\beta\left(E + \sqrt{2n\ln n}\right)$ – that is $\simeq \beta\sqrt{\frac{n}{\ln n}}$ at the shallowest ones and $\sqrt{2}\beta E_1$ at the deepest one. We then notice that as most of them are shallow and that $\mathrm{e}^{2\beta\sqrt{2\ln n}}\mathrm{e}^{\beta\sqrt{\frac{n}{\ln n}}} \ll \mathrm{e}^{\sqrt{2}\beta E_1}$ (we had better use c and d). Hence, the time scale we are after is $\boxed{\mathrm{e}^{2\beta\sqrt{\beta n\sqrt{2n\ln n}}}}$.

c According to [26][3] the corresponding time scale is $\boxed{\ln t \simeq \beta\sqrt{n}}$. We will actually show this in our own way later.

d According to [26][3] the corresponding time scale is $\ln t \simeq \beta^2 n$. We will not use the result but rather establish it for high temperatures as well as its low temperature counterpart. In order to determine it, it suffices to compare the time spent in the deepest trap with the time spent in all the others. That is $\mathrm{e}^{-\beta E_{min}}$ and (as in Equation 5.2)

$$\frac{N}{\sqrt{2\pi n}}\int_{E_{min}}^{0}\mathrm{e}^{-\beta E}\mathrm{e}^{-\frac{E^2}{2n}}\mathrm{d}E \approx \begin{cases} N\mathrm{e}^{\frac{\beta^2 n}{2}} & \text{if } \beta n < -E_{min} \\ \frac{nN}{\sqrt{2\pi n}}\mathrm{e}^{-\beta E_{min}}\mathrm{e}^{-\frac{E_{min}^2}{2n}}\frac{1}{E_{min}+\beta n} & \text{if } \beta n > -E_{min}, \end{cases} \tag{5.1}$$

where $N = \mathrm{e}^{\frac{E_{min}^2}{2n}}$ is the number of states visited until we find E_{min}. Hence

- If $\beta < \beta_c$, the system stops aging when $E_{min} = -\beta n$. This gives the time scale $\boxed{\mathrm{e}^{\beta^2 n}}$ (deepest trap time).
- If $\beta > \beta_c$, the system ages until the min reaches the global minimum (after that no aging can take place anyhow). This gives the time scale $\boxed{\mathrm{e}^{\beta\beta_c n}}$ (deepest trap time).

e Knowing that all the states between E_{min} and E_{max} are at equilibrium and the others have not been discovered, we get the following for the mean energy.

$$E_{mean} = Z^{-1}\int_{E_{min}}^{E_{max}} E\mathrm{e}^{-\beta E}\mathrm{e}^{-\frac{E^2}{2n}}\mathrm{d}E = \frac{\int_{E_{min}+\beta n}^{\infty}(E-\beta n)\mathrm{e}^{-\frac{E^2}{2}}\mathrm{d}E}{\int_{E_{min}+\beta n}^{\infty}\mathrm{e}^{-\frac{E^2}{2}}\mathrm{d}E} \approx \max(-\beta n, E_{min}). \tag{5.2}$$

We deduce from there that there are two different scenarios depending on the temperature, both identical to d.

- If $\beta < \beta_c$, macroscopic equilibrium is reached when $E_{min} = -\beta n$ and afterwards the minimum keeps going downwards without impact on macroscopic properties. The corresponding time scale is $\boxed{\mathrm{e}^{\beta^2 n}}$.
- If $\beta > \beta_c$, macroscopic equilibrium is reached when $E_{min} = -\beta_c n$, which is the global minimum – in this case the transition coincides with f and the time scale is $\boxed{\mathrm{e}^{\beta\beta_c n}}$.

f
- If $\beta > \beta_c$ we just saw that the time scale is $\boxed{\mathrm{e}^{n\beta\beta_c}}$.
- If $\beta < \beta_c$ after the aging stage equilibrium settles and the mean time per state stations at $\mathrm{e}^{\beta^2 n}\mathrm{e}^{-n\frac{\beta^2}{2}} = \mathrm{e}^{\beta^2\frac{n}{2}}$ (cf. d). That determines N and using $N = \mathrm{e}^{\frac{E_{min}^2}{2n}}$, we get $E_{min} = -\sqrt{2n\ln t - \beta^2 n^2}$. Thus the time scale is $\boxed{\mathrm{e}^{(\beta^2+\beta_c^2)\frac{n}{2}}}$.

[3] The setting considered there is the complete graph with BTM dynamics and with Gaussian energies rather than the REM with Metropolis.

g For symmetry reasons the system discovers (and not visits) the positive energies just like the negative energy ones. Thus the time needed to visit the global max is the one for visiting the global min multiplied by the inverse of the transition rate towards the global max (from I i.e. 0 energy as we will discuss further and provided that we neglect the small correction that was paid attention to at d), that is (compare with Equation 5.6)

- $\boxed{\mathrm{e}^{2n\beta\beta_c}}$, if $\beta > \beta_c$.
- $\boxed{\mathrm{e}^{\frac{n}{2}(\beta+\beta_c)^2}}$, if $\beta < \beta_c$.

Remark 5.3.2.1. We should mention that given that after f the dynamics starts systematically returning to already explored states (a non-zero fraction of the states have been explored already), the implicit effective independence hypothesis might cause potential difficulty in establishing related statements rigorously if they do indeed hold.

Preliminary remarks Before describing the max, we should note two facts that will be of use in what is to follow.

Markov chain presence probabilities The Markov process we call dynamics has, as usual, an associated chain (see 1.1.3). Observe that at traps $a_i = \mathrm{e}^{\beta E}$, at high energies $a_i = 1$ and around I $a_i \simeq 1$. Moreover, the Gibbs measure is reversible for the continuous time process, so the measure given by the Markov chain presence probabilities ($\mu_m(\{i\}) := \mathbb{P}_E(\xi_m = i)$ for $i \in S$)[4] is uniform on the traps ($\simeq 2^{-n}$) and decreases as the Boltzmann factor on the high energy states ($\simeq 2^{-n}\mathrm{e}^{-\beta E}$). This justifies our considering that newly drawn trap energies are IID Gaussians cut above 0 ($\inf I$) – the chain presence probability is uniform on them and thus in terms of energy it is Gaussian.

Remark 5.3.2.2. A point that needs attention if we aim at rigour is that the chain we are describing is *not* ergodic, as it is (clearly) 2-periodic. Knowing that the initial uniform probability gives half of the mass to "odd" and to "even" states (imagine a chessboard coloured hypercube), that remains so for all times. Recall that (the chain is irreducible positive recurrent) the chain measure (see Section 3.3) converges towards the invariant measure we discussed above and that the invariant measure in question gives necessarily equal masses to even and odd states (as it is invariant). Therefore, the initial condition is nice enough to ensure that the 2-periodicity does not interfere and the chain behaves like an ergodic one, the chain presence probabilities converging towards the invariant measure. Do note that this is the second time (along with returning) when the seemingly innocent and arbitrary uniform initial distribution turns out to give interesting consequences and this time it is not just a curious fact but an important feature of the dynamics.

New records for the max always jump from 0 Recall that the Gibbs measure decreases like the inverse Boltzmann factor and consider jumping to a new record for the max. As the Botzmann factor is exactly compensated (it is easier to go up to the new record from a state that is already high but such a state is unlikely to be in such a state), the distribution of the energy, from which we attain the new record, is the bare Gaussian. Therefore, it is legitimate to consider that we reach records from 0.

Evolution of the max Now that we have discussed (most of) the transitions that are supposed to take place, we are ready to look into the evolution of the max itself as a function of the time scale. We can expect the following regimes:

1. At the start of the process the max is certain to be of order $\sqrt{n}$ (Avoid confusion with I, as we will use $\sqrt{n} \ll \sqrt{n \ln n}$) i.e. the standard deviation of the Gaussian (just like the min), as we take a uniform initial distribution.
2. We will explain in Appendix A why until a the maximum is strictly constant. Moreover, the dynamics goes down rapidly.
3. The extremes of the Gaussian show that at a the typical energy barrier that the dynamics has to overcome is $\simeq \sqrt{\frac{n}{\ln n}}$. Therefore, up until the time scale (for once we do not even have an accurate leading order but just the order of $\ln t$) $\mathrm{e}^{\beta\sqrt{\frac{n}{\ln n}}}$ the system is stuck in the same state (note that this is far more than just to say that it stays there the majority of the time) – in particular aging has already begun, after which it climbs to the lowest energy neighbour, then looks for a lower energy among his neighbours (thus jumping back with a probability $\simeq 1$). Thus until that time scale both the min and the max stay constant (exactly and not just in order) and equal to their values at the time scale n.
4. Before the time scale $\mathrm{e}^{\beta\sqrt{2n\ln n}}$ (on any strictly shorter scale) it is certain that the max cannot increase, as the system has not been given enough time to return to 0 energy (in the sense energy $\lesssim \sqrt{n}$) from the energy it had discovered at a. Moreover, with the same reasoning as in b one can show that there is no movement of the max until $\mathrm{e}^{\beta E_1}$ either. It is worth noticing that we are already way past the point at which the trapping times become relevant (strictly speaking that is when $E_{min} \ll -\sqrt{2n\ln n}$).

[4] Take heed that we consider states and not energies – the latter would require to include the Gaussian density.

5. Using the same reasoning as in b we get

$$\forall t = \mathrm{e}^{\sqrt{1+\gamma}\beta E_1},\ \gamma \in (0;1) \Rightarrow E_{max} \approx \gamma\sqrt{2n\ln n}. \tag{5.3}$$

Note that we have constrained ourselves not to *use* that there is aging but we rather *show* that there is.

6. As long as aging lasts (depending on temperature) we have $-\beta E_{min} = \ln t$, $\ln N = \frac{E_{min}^2}{2n}$ and thus with the reasoning from b we get $N = \mathrm{e}^{\frac{E_{max}^2}{2n}}\mathrm{e}^{\beta E_{max}}$ (we take into account that we only have a neighbour of energy E_{max} once out of $\mathrm{e}^{\frac{E_{max}^2}{2n}}$ states). Hence,

$$E_{max} = -\beta n + \sqrt{n^2\beta^2 + \beta^{-2}(\ln t)^2}. \tag{5.4}$$

Note that at short time scales we find $E_{max} = \frac{(\ln t)^2}{2\beta^2 n}$ and in particular in the setting of Equation 5.3 $(1+\gamma)\beta\sqrt{2n\ln n}$, which is of the right order but contains an error due to the fact that we do not satisfy the (implicit) hypothesis that $E_{max} \gg \sqrt{2n\ln n}$. In order to get a correct result for the scales $\gamma > 1$ one needs to mix this point's reasoning with that of b, but that only leads to needless complications. Another interesting value is the longest aging time scale at high and low temperature, namely $E_{max} = \beta n(\sqrt{2}-1)$ and $n(\sqrt{\beta^2+\beta_c^2}-\beta)$ respectively.

7. After the end of aging we have (the only change compared to the previous case is that the time per state is stationary instead of determined by the min) $t = N\mathrm{e}^{\beta\beta_i n}\mathrm{e}^{-n\frac{\beta_i^2}{2}}$, $N = \mathrm{e}^{\frac{E_{max}^2}{2n}}\mathrm{e}^{\beta E_{max}}$. Hence,

$$E_{max} = -\beta n + \sqrt{(\beta-\beta_i)^2 n^2 + 2n\ln t}. \tag{5.5}$$

Of course, this is only true as long as the max is smaller than the global max i.e. until

$$\ln t \leq (2\beta - \beta_i + \beta_c)(\beta_c + \beta_i)\frac{n}{2} = \begin{cases} n\frac{(\beta+\beta_c)^2}{2} & \text{if } \beta < \beta_c \\ 2n\beta_c\beta & \text{if } \beta > \beta_c. \end{cases} \tag{5.6}$$

Note that this gives a (perhaps) more convincing proof of the scales found in g.

8. The system has reached full microscopic equilibrium by visiting all of its states. The max is constant ever after.

Remark 5.3.2.3. As always with phase transitions, the analysis near the critical temperature needs special attention and some of the results may not hold there.

Summary To put it all clearly, we reassemble the previous results here.

- Initial value $\simeq \sqrt{n}$.
- Constant until $t \simeq \mathrm{e}^{\beta\sqrt{2\beta n\sqrt{2n\ln n}}} = \mathrm{e}^{\beta E_1}$.
- An increase leading to $E_{max} \approx \sqrt{2n\ln n}$ at $\ln t \approx \sqrt{2}\beta E_1$ according to Equation 5.3.
- As long as there is aging (cf. d), the max evolves according to Equation 5.4.
- After macroscopic equilibrium (cf. d) the evolution is given by Equation 5.5 until the scales given by Equation 5.6.
- Constant afterwards.

Underlying model As it may have remained unclear exactly what approximations and hypotheses we made, we recapitulate here the "effective" model we described instead of the strict Metropolis on the REM.

- The energies in I are treated apart. We assume that there are always neighbours of all those energies.
- For $E < \inf I$ we say that the state holds the process for the time necessary for it to have a probability of order 1 to go up to a state in I (to a neighbour essentially). However, we do not assume that it always goes to $\inf I$, but has a chance $\mathrm{e}^{-\beta(E'-\inf I)}$ to go to a state of energy E' instead, as long as such a state is available among the neighbours. We neglect the possibility of having two such "deep" states next to each other.
- For $E > \sup I$ the process stays there for a time 1 (clearly this is a.s. the mean waiting time). We neglect the possibility of having two such "high" states next to each other as well as any other way of accessing them other than what was described in the previous point.

- When we are "far" from I, we often neglect the number of neighbours and set it to 1 (this only simplifies calculations a bit and the hypothesis can be easily discarded).
- Finally, we suppose that the system becomes independent from its past at each jump, namely that the neighbours are dawn anew (which is clearly only approximately true and completely false beyond f for example, but might be true in an effective fashion). Yet, we keep account of a global max and min somewhat artificially and say that we cannot go beyond them even though by drawing new variables we will eventually go past them.

Now, this might seem (or not) very close to the original Metropolis dynamics on the REM but if a rigorous treatment is attempted the strict equivalence is needed and is not necessarily obvious.

Records approach We now present an approach which allows us to obtain a few additional results for free. Even though it might seem much more rigorous, all the dirt is just concentrated in one place at the very beginning so finally we are doing basically the same thing but in a way that ensures we find ourselves exactly in the framework of the common theory of IID records (and extremes) and thus permitting the use of all the classical results. In order to do that, we need to rephrase our max in terms of an IID.

We suppose that at every jump (recall Remark 2.1.1.3) to a state of energy $> \inf I$ (at least half the jumps according to our hypothesis that there are no adjacent traps, so we will say simply that there are N of those jumps) in an effective fashion we draw an independent variable from the law $\mathscr{L}$ that we will give explicitly in Equation 5.7. This distribution should incorporate the original energy distribution and the acceptance rate. For C a normalisation constant we have

$$\mathscr{L}(\mathrm{d}x) = C_n \left(\mathrm{e}^{-\frac{x^2}{2n}} \mathrm{e}^{-\beta(x-\inf I)} \mathbb{1}_{[\inf I, \beta_c n]}(x)\mathrm{d}x + \mathrm{e}^{-\beta(\beta_c n - \inf I)} 2^{-n} \delta_{\beta_c n}(\mathrm{d}x) \right). \tag{5.7}$$

We had best disregard the Dirac and the cut, as they only ensure the correctness of the final value. Given that $\mathscr{L}$ is just a cut Gaussian centered at $-\beta n$, when we are still far from the global max we find a GUM max and afterwards identically the global max.

Now in order to get as many results as we want all we need to do is look up a result from records theory (for example in [28]), apply it to this law and then go back to an expression in terms of t instead of N using what happens with the min. For this last purpose, let us sum up what we establish in this chapter and the next one.

- During the aging stage we have $\ln N \approx \frac{(\beta^{-1}\ln t)^2}{2n}$ and if we include finite n corrections, that becomes
$$\ln N \approx \frac{(\inf I - \beta^{-1}\ln t)^2}{2n}.$$
- The end of aging is given by $\ln t \approx n\beta\beta_i$ and if we include finite n corrections, that becomes $\ln t \approx n\beta\beta_i + \beta \inf I$.
- After the end of aging $\ln N \approx \ln t + n\beta_i \left(\frac{\beta_i}{2} - \beta\right)$ and if we include finite n corrections, that becomes
$$\ln N \approx \ln t - \beta(\beta_i n + \inf I) + \frac{n\beta_i^2}{2}.$$

Example 5.3.2.4. Using this recipe we obtain that the expected number of records is given by $\approx \min\left(\ln N, n\beta_c\left(\beta + \frac{\beta_c}{2}\right) - \beta \inf I\right)$.

Remark 5.3.2.5. We can also minorate the standard deviation (there are other additional fluctuations) by $\sqrt{\ln N}$.

Yet, we need to take heed not to use directly results on the records' lifetimes. For instance, if we are interested in the longest lifetime of a record, is not (necessarily) the rescaled lifetime of the record which survived the most jumps, as the jumps become more scarce as time goes. Nevertheless, we can potentially get correct results by modifying accordingly the proofs of standard records results. As this is not of central concern to us, we will not investigate this further.

Chapter 6

Numerical results

It should be noted that many of the observations in this chapter had already been done by Baity Jesi.

6.1 Motivation

Though the theory given in the previous chapter should hold under the assumptions given, it is good to note that those are far from the "reality" of numerical simulations of the REM, currently possible. The content of this chapter is influenced by discussions with Marco Baity Jesi, who was doing simulations of the Metropolis dynamics on the REM and also provided the author with some priceless data to examine. Most unfortunately, the limited numerical capacities do not permit going over $t = 10^{12}$ (do note that for such t it is already impossible to naively save the trajectory, as the size of such data is overwhelming) and so most of the regimes we discussed remain beyond their reach. What is more, the max as it was defined remains constant and equal to its initial value most of the time and if we are a bit unlucky we may even see it constant over the whole simulation which is hardly of much use.

We can, however calculate other types of max, that convey more information. As was explained, the max remains constant for a very long while mostly because increasing it requires going out of the first trap visited. Then, we come to some natural ideas that Baity Jesi used to suppress the initial value or even make the max lose its memory. Namely, we can consider the two-time max in two manners – $M(t_w, t)$ and $M(\frac{t}{2}, t)$ (note that this computationally very particular choice corresponds to calculating the max on disjoint intervals, as the times considered in the simulation are powers of 2) with a particular choice of t_w of order n (16), which ensures that we will be in the first trap by then. With this trick one manages to get much more information concerning the max out of the sample and we will use these functions to compare theory and simulation results.

Clearly, if we desire access to exponential time scales like g, for example, we cannot afford to have very large n. However, at small n much of what we have stated fails to be true and sometimes by far (Nevertheless, in physical systems with n being a number of spins in a sample, we should not have this kind of problems). Typically the possible n for which one is able to reach time scales like e^n and to operate on $\{0,1\}^n$ is at most about 20 (in numerical applications this is the value we will use, as well as $\beta = \frac{4}{3}$). Now, clearly we have to reconsider $n^{1/4} \gg 1$ and worse yet, $\sqrt{\ln n} \gg 1$.

In order to see that something more needs to be done, take a look at Figures 6.1 and 6.2. For all figures we systematically omit $\mathbb{E}[\bullet]$ and we write $M_{a,n}(t_1, t)$ for the mean of the max on the interval $[t_1, t]$ for the given n (if ∞, it is without corrections) and a is among s, t, f, standing respectively for simulation, theory and fit (we will discuss those further). As we will see in Appendix B, 6.2 is more important and 6.1 is somewhat of a consequence. Indeed the tendency of simulation data is towards the infinite n prediction but we are quite far away.

Therefore, we will attempt to give an adapted theory for finite n in the spirit of the one for infinite n, so that the results may be applicable to the case studied in simulations. However, no justification will be attempted, for one because a moment's thought shows that most of the hypotheses that will be made are false. We only seek to find a correct result, acting as if n were big when needed but keeping negligible corrections, that will not be negligible in the end.

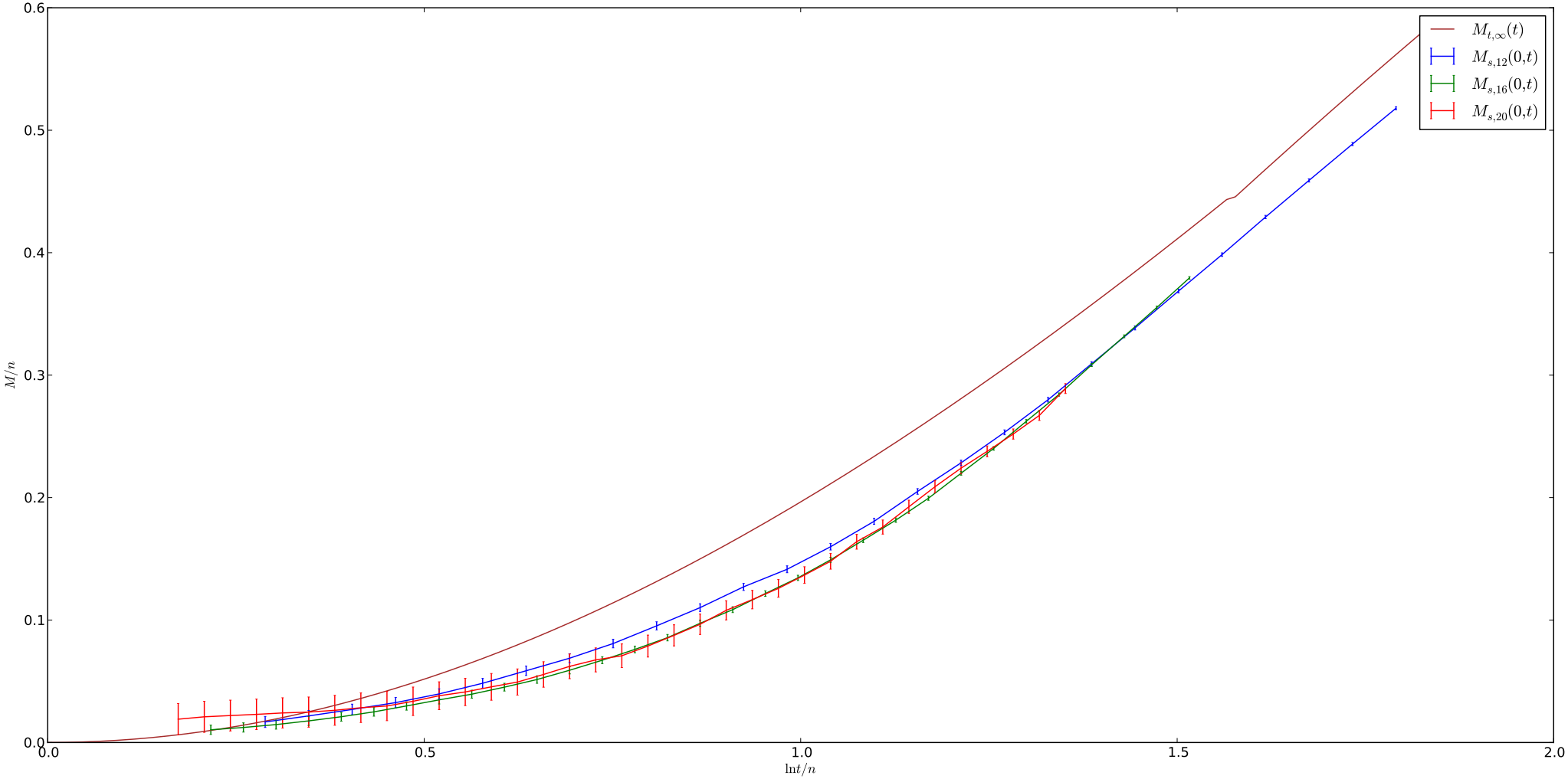

Figure 6.1: Starting at 0

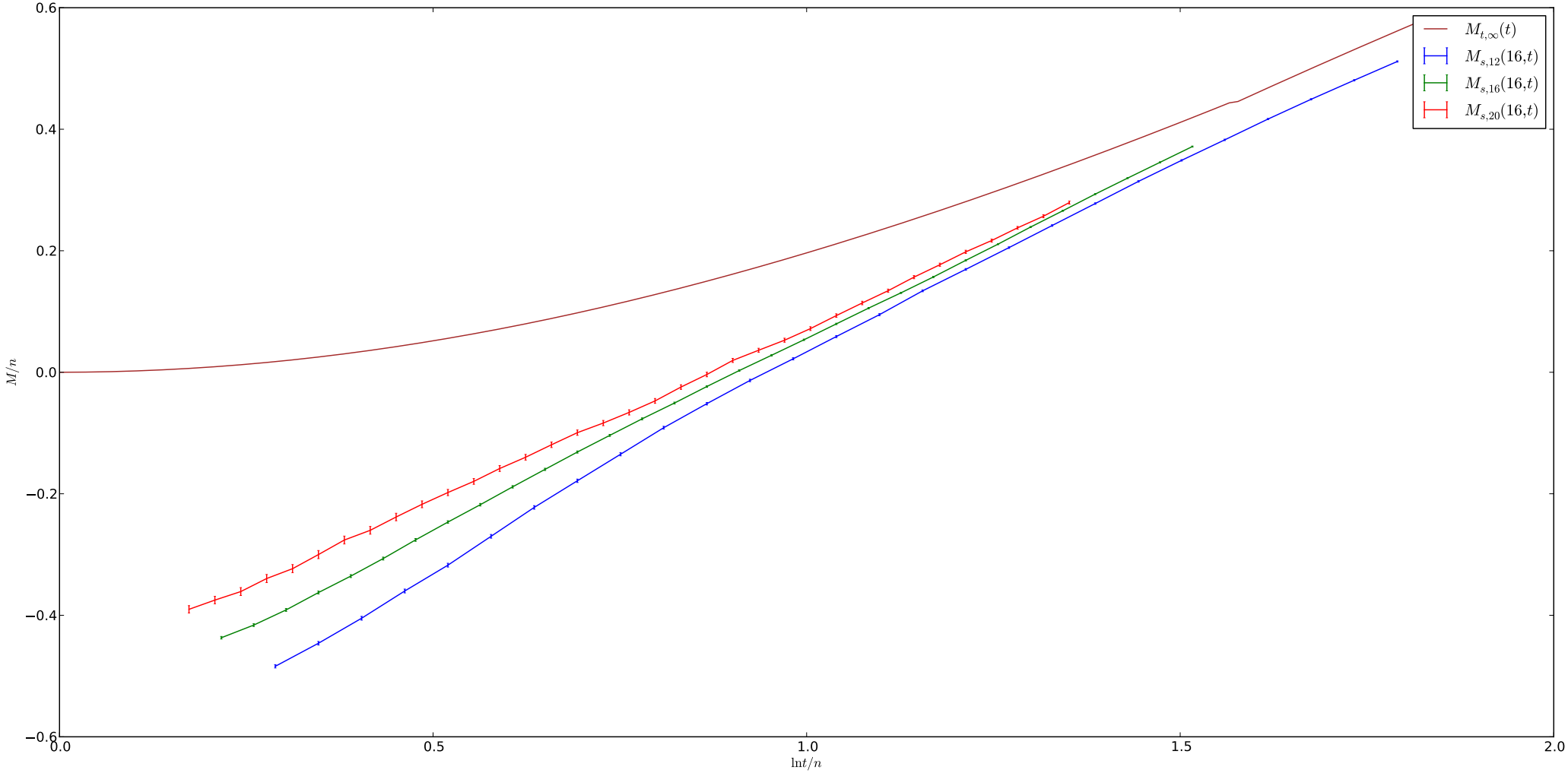

Figure 6.2: Starting at 16

6.2 Finite n corrections

This section summarises the useful results of Appendix B.[1] Most of what will be said is false at $t < n$. Before we start, let us mention what is taken into account. The most important is that we do not neglect $\inf I$ before anything and we use the true value of $\inf I$ instead of the asymptotic equivalent. The global min is not corrected.

6.2.1 Notations

$$\beta_c := \sqrt{2\ln 2} \tag{6.1}$$

$$\beta_i := \min(\beta, \beta_c) \tag{6.2}$$

$$\operatorname{erfc} := \frac{1}{\sqrt{2\pi}} \int_\bullet^\infty \mathrm{e}^{-\frac{E^2}{2}} \mathrm{d}E \tag{6.3}$$

$$\inf I := -\sqrt{n} \int E \mathrm{d}(1-\operatorname{erfc})^n \tag{6.4}$$

$$\sigma := \sqrt{n \int E^2 \mathrm{d}(1-\operatorname{erfc})^n - (\inf I)^2} \tag{6.5}$$

$$S := \mathrm{e}^{\beta(\inf I + n\beta_i)} \tag{6.6}$$

$$\ln F := (2\beta - \beta_i + \beta_c)(\beta_c + \beta_i)\frac{n}{2} = \begin{cases} n\frac{(\beta+\beta_c)^2}{2} & \text{if } \beta < \beta_c \\ 2n\beta_c\beta & \text{if } \beta > \beta_c. \end{cases} \tag{6.7}$$

That is $\inf I \approx -8.34053$ (all numbers are at $n = 20$, $\beta = \frac{4}{3}$). It has a statistical error (it may vary from one point of the same instance to another) of order $\sigma \approx 2.34241$, which decreases after $t = S$ at $\beta < \beta_c$ and stations at some instance-dependant value (but no longer time-dependant) at $\beta > \beta_c$.

6.2.2 $M(16, 16+t)$

After the initial phase, at the end of which both the min and the max are well out of the zone $[\inf I - \sigma, \inf I + \sigma]$, things get more robust. If we call this time t_0, what we will refer to by t is in fact $t - t_0$ (if it is large, this does not matter).

- For $t < S$

$$E_{max}{}^{2} \approx -\beta n + \sqrt{(\beta n + \inf I)^2 - 2\beta^{-1} \inf I \ln t + \beta^{-2}(\ln t)^2}. \tag{6.8}$$

  While we are here, we also have

$$E_{min} \approx \inf I - \beta^{-1} \ln t. \tag{6.9}$$

- For $S < t < F$

$$\frac{E_{max}}{n} = -\beta + \sqrt{(\beta - \beta_i)^2 + 2\frac{\ln t}{n}}. \tag{6.10}$$

- For $t > F$

$$E_{max} = \beta_c n. \tag{6.11}$$

6.2.3 $M(\frac{t}{2}, t)$

$$t = \mathrm{e}^{\beta(\inf I + n\beta_c)} \tag{6.12}$$

(times a few $\mathrm{e}^{\beta\sigma}$, say 1) is the time at which $M(\frac{t}{2}, t)$ joins the other two definitions of the max (up to small fluctuations).

6.2.4 $\mathbb{E}[M(0,t)]$

In this subsection E_m is E_{max} from the previous one, corresponding to $M(16, t+16)$. From here on we discuss the mean max energy.

[1] The interested reader is invited to study the appendix before continuing.

[2] We shall refer to $M_{|E_0=\inf I}(0,t)$ as $M(16, 16+t)$ or E_{max} in this subsection.

- At very short times $t \ll e^{\beta\sqrt{n}}$ (particularly unimportant regime that can safely be ignored)

$$\mathbb{E}[E_{max}] \approx \left(\operatorname{erfc}\left(\frac{\inf I}{\sqrt{n}}\right) - \operatorname{erfc}\left(\frac{E_m}{\sqrt{n}}\right)\right) E_m \approx \ln t \frac{\inf I}{\beta(\beta n + \inf I)^2} \frac{\inf I}{\sqrt{2\pi n}} e^{-\frac{(I)^2}{2n}}. \tag{6.13}$$

  Either one of the expressions can be used – the first is more accurate, the second is simpler...

- Afterwards

$$\mathbb{E}[E_{max}] \approx \left(1 - \operatorname{erfc}\left(\frac{E_m}{\sqrt{n}}\right)\right) E_m + \sqrt{\frac{n}{2\pi}} e^{-\frac{E_m^2}{2n}}. \tag{6.14}$$

  To be more explicit, we give the E_m that should be plugged in Equation 6.14

  - Up to $t = S$ see E_m in Equation 6.8.
  - Up to $t = F$ see E_m in Equation 6.10.
  - After $t = F$

$$\mathbb{E}[E_{max}] \approx \beta_c n \tag{6.15}$$

    or, if you are very pedantic, see E_m in Equation 6.11 – it is not exact.

6.3 Comparison

Using the simulation of Baity Jesi we can now investigate to what extent this theory reflects the (numerical) "truth".

6.3.1 Qualitative

To begin with, we outline some of the aspects predicted by the previous chapter, which are confirmed by the simulations. We will mostly comment on the memory-loosing max $M\left(\frac{t}{2}, t\right)$ as it contains the most information.

- The max from $t = 0$ is constant and equal to its initial value for a very long time.
- The first trap is reached very quickly.
- The process does exactly as we assumed in our model – when it reaches a trap it starts jumping eventually back an forth between it and its lowest neighbour. This is testified to by the long time regions where $M\left(\frac{t}{2}, t\right)$ takes the same value around $\inf I$, while it is more than clear that the process can escape from such a state quickly, were it a trap and not the lowest neighbour of a much deeper trap.
- $M\left(\frac{t}{2}, t\right)$ joins the other two at some point, namely when the dynamics stop taking values about (or below) $\inf I$.

An important aspect that is foreign to our model (especially as it is in Appendix A) is the presence of short cycles. This might seem innocent, but it actually shows that we are overlooking some things. In order to emphasise this, we argue below that the frequency of appearance of 3-cycles remains large even at infinite systems and is not just a small n problem. As already discussed, we often find a deep trap A with a lowest neighbour B around $\inf I$. Then all we need (it suffices) in order to observe a 3-cycle is that B be the lowest neighbour of C and that C have exactly one neighbour lower than C, where C is the lowest neighbour of B excluding A. Under these circumstances we will typically jump A-B-C-B-A...[3] That is rewritten as $\inf E(\Gamma(C)) \geq \inf E(\Gamma(A)) =: E(B)$, $\inf E(\Gamma(C)\backslash B) > \inf E(\Gamma(B)\backslash A) =: E(C)$ and $E(B) < E(C)$. There is exactly one common neighbour of two among A, B and C, which we neglect to say that we independently draw the energies of the neighbours of A, the remaining ones of B (save A) and finally, the remaining ones of C (including the one that is technically already known but excluding B). In this setting the first inequality is "the lowest among n is lower than another $n-1$" – its probability is about $\frac{1}{2}$. The second one is roughly $\frac{1}{2}$ as well, because C is the min of $n-1$ variables just like $\inf E(\Gamma(C)\backslash B))$. The same holds for the third one. Although this does not prove anything at all especially as there is no independence, it should be enough to convince that the event of having a 3 cycle is far from being rare and that is all we were aiming at.

While this can be handled at infinite n by artificially adding an "effective trap" (much like what is done in the BTM) containing the trap itself and the lowest neighbours and eventually their lowest neighbours and so on as long as we are nearly certain to come back to the original trap. At infinite n this is all right as that will not get the threshold energy up, at small n we might find the discrete energies influencing the effective $\inf I$.

[3] Read the last three sentences two more times and draw a figure.

6.3.2 Quantitative and finite n corrections

General agreement As it is easily seen from Figures 6.3 and 6.4 (recall that at infinite n the prediction for $M(n,t)$ is the same as $M(0,t)$ at t such that $\ln t \simeq n$, as we are way past b and thus way past the initial value) the corrections at finite n are crucial and give exactly what they should during the aging stage (after all that is the most important period in practice). After the aging regime the theory becomes the same with or without corrections as we saw and it is quite close to the truth (see rather Figure 6.9). Yet, there is a seemingly constant asymptotic error.

Additional fitted correction Let us attempt to understand this error and to see if it means that the theory is somewhat false or if it is within the "errorbar of the theory". Firstly, it is nice to note that it is approximately constant (up to a transitory phase) after the aging stage and it starts being visible exactly where we expect (with finite n corrections) that the end of aging will be. We happen to have actually noted a feature that might intervene in this fashion in Remark B.1.0.2. If the global min is not what we took it to be, corrective terms will appear in 6.10, leading to[4]

$$\frac{E_{max}}{n} = -\beta + \sqrt{\beta^2 + \beta_c^2 - 2\beta\beta_m + 2\frac{\ln t}{n}} \tag{6.16}$$

at $\beta < \beta_c$, where β_m (it is a modified β_c) is an adjustable parameter corresponding to $-\frac{E_{min}}{n}$ for the global min and the end of aging is shifted to $\mathrm{e}^{\beta(\inf I+\beta_m)}$. If we allow ourselves to vary β_m so as to fit the obtained prediction to the simulation data, we get Figure 6.5 and the resulting (mimicking the way we obtained Equation B.5) Figure 6.6. For verifiability, we note that the value used is $\beta_c + 0.055$.

In should be noted that we prefer to keep the discontinuity for visibility. Clearly in the real system there should be a smooth joint, which works to our advantage, as obviously combining the two regimes will result to a serious improvement around the jump, as they are on opposite sides of the simulation. The reason not to make a smooth curve is to avoid choosing arbitrarily what convolution kernel to use, for example.

Let us give three reasons as to why we are entitled to this adjustable parameter.

- We *know* that the global min is not $-\beta_c n$ according to the extremes. Moreover, the corrective terms (for our values of n we have $\ln(n\ln 2) \not\gg \ln(4\pi)$) are of order $\frac{1}{n}$ (for $\frac{E}{n}$), which is just the order of the deviation (yet it should be noted that it goes in the wrong direction).
- Still according to extremes, we know that the (random) global min has the Gumbel distribution, whose mean is not 0 but γ (Euler's constant) and so we should correct the global min by $\frac{\gamma}{n}$ if we want the average.
- In the same line of thoughts, we are not at all obliged to use the mean, as that would be arbitrary. All we can be certain of is that the effective global minimum that we need to consider (taking into account the distribution of typical size $\frac{1}{n}$) differs from the one we used by an order $\frac{1}{n}$ i.e. exactly the order of the difference we need.

Records Let us look into the number of records. For Figure 6.7 we theoretically expect about 14 records up to the end of aging and we count 8-9 of them. We should take into account that the standard deviation is still rather big (square of the mean) and that it is possible that some records be hidden by the simulation (especially when we see successive records), as we calculate the max only on huge intervals (what we see is a minorant of the number of records). Thus, the result can be deemed satisfactory.

Remark 6.3.2.1. Although the analysis carried out does not at all permit us to claim that the corrections added are, say the sub-leading orders up to a yet smaller error, the remarkable agreement with numerical data might suggest that such is the case.

$M\left(\frac{t}{2},t\right)$ Finally, we discuss the only observable where the previous considerations fail to give an adequate result – $\mathbb{E}\left[M\left(\frac{t}{2},t\right)\right]$. This is the reason why we do not show any theoretical prediction in Figure 6.8. We might well try Equation B.6 and even envision a correction to $\Pi\left(\frac{t}{2},t\right)$ – seeing it as a parameter to be fitted between 0 and 1, but that does not at all allow us to obtain the dip clearly visible in Figure 6.8 (or obtain a "not very straight" curve). We might as well try the very sensitive expression from the crude approach, but that is rather guessing than doing anything serious and in any case, no satisfactory result was found. This can only lead to the conclusion that this observable requires more attention and going back over all the difficulties of determining the energy distribution at a given time, namely close to the edge.

[4] We assume that the number of visited states before the global min is unhindered, but the time needed to do it is altered by its depth.

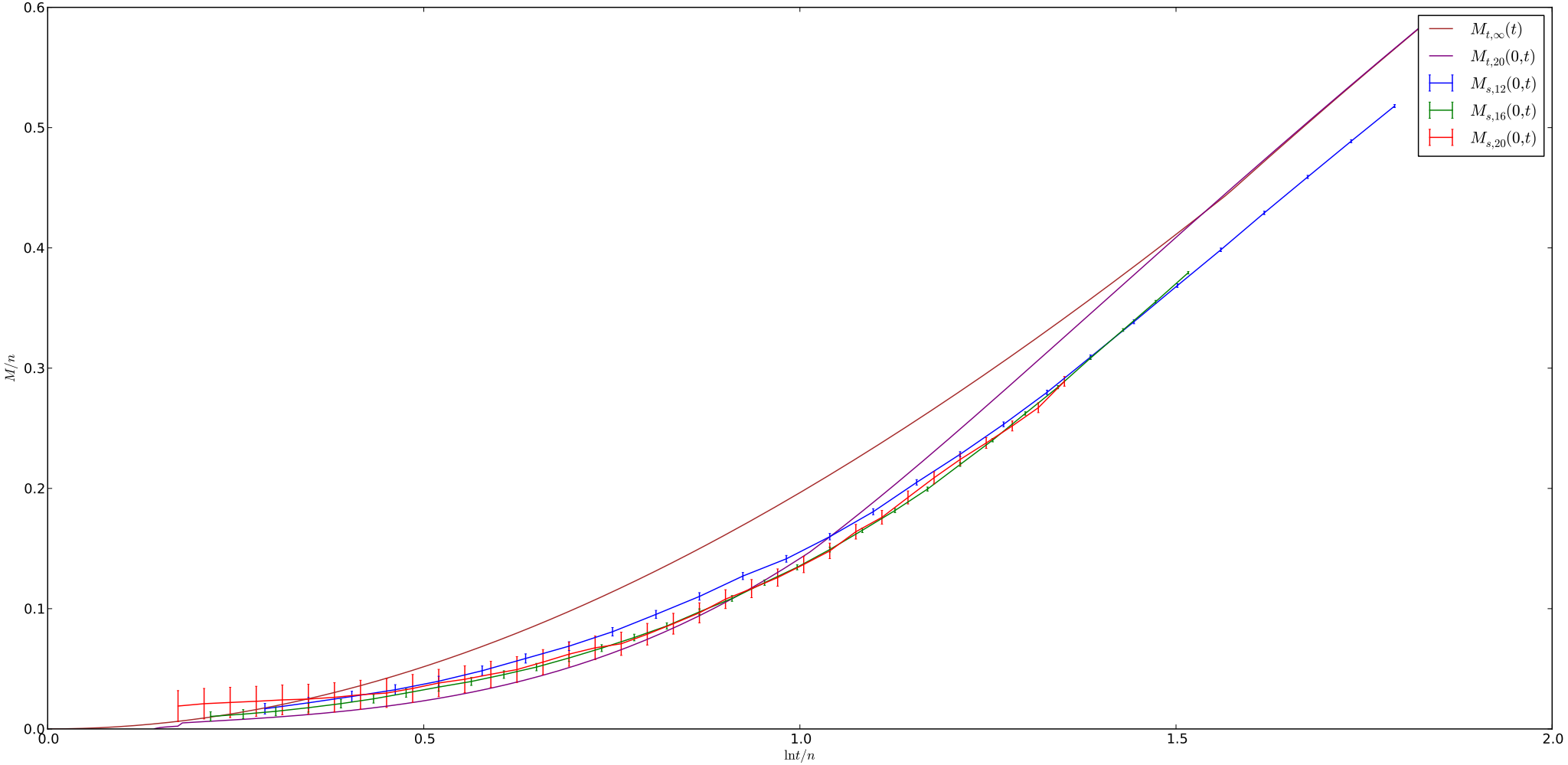

Figure 6.3: Starting at 0

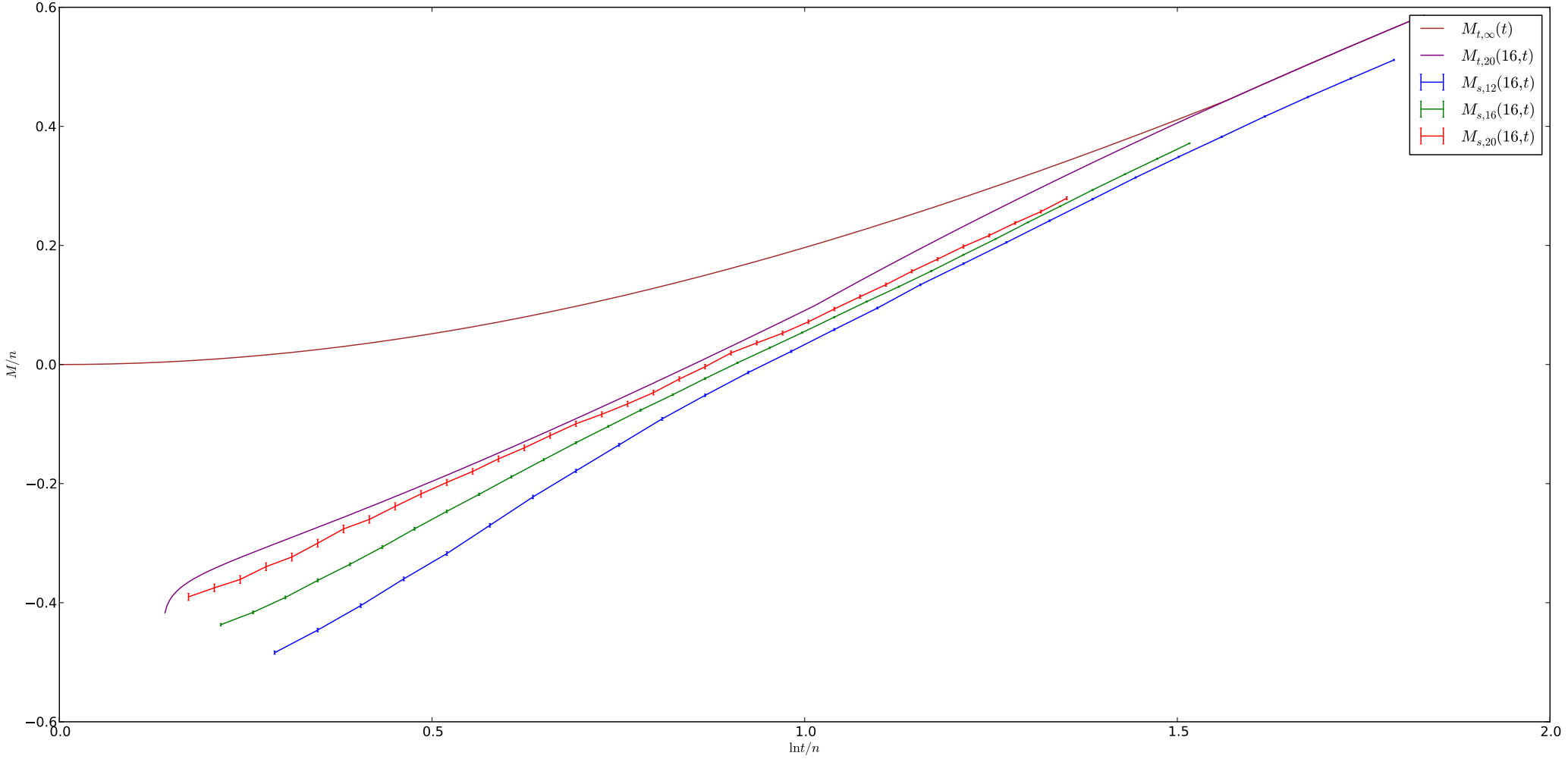

Figure 6.4: Starting at 16

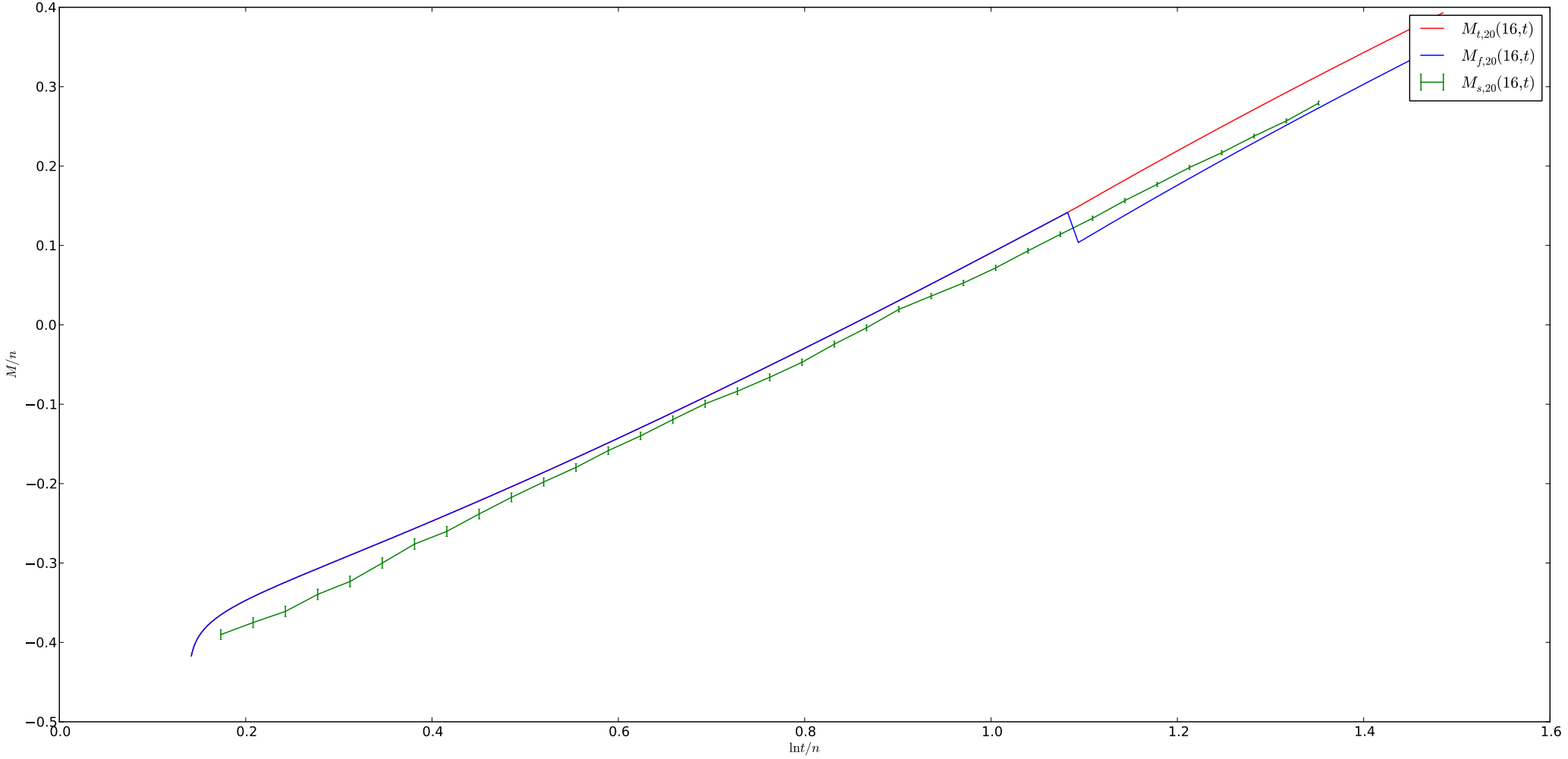

Figure 6.5: $n = 20$, starting at 16

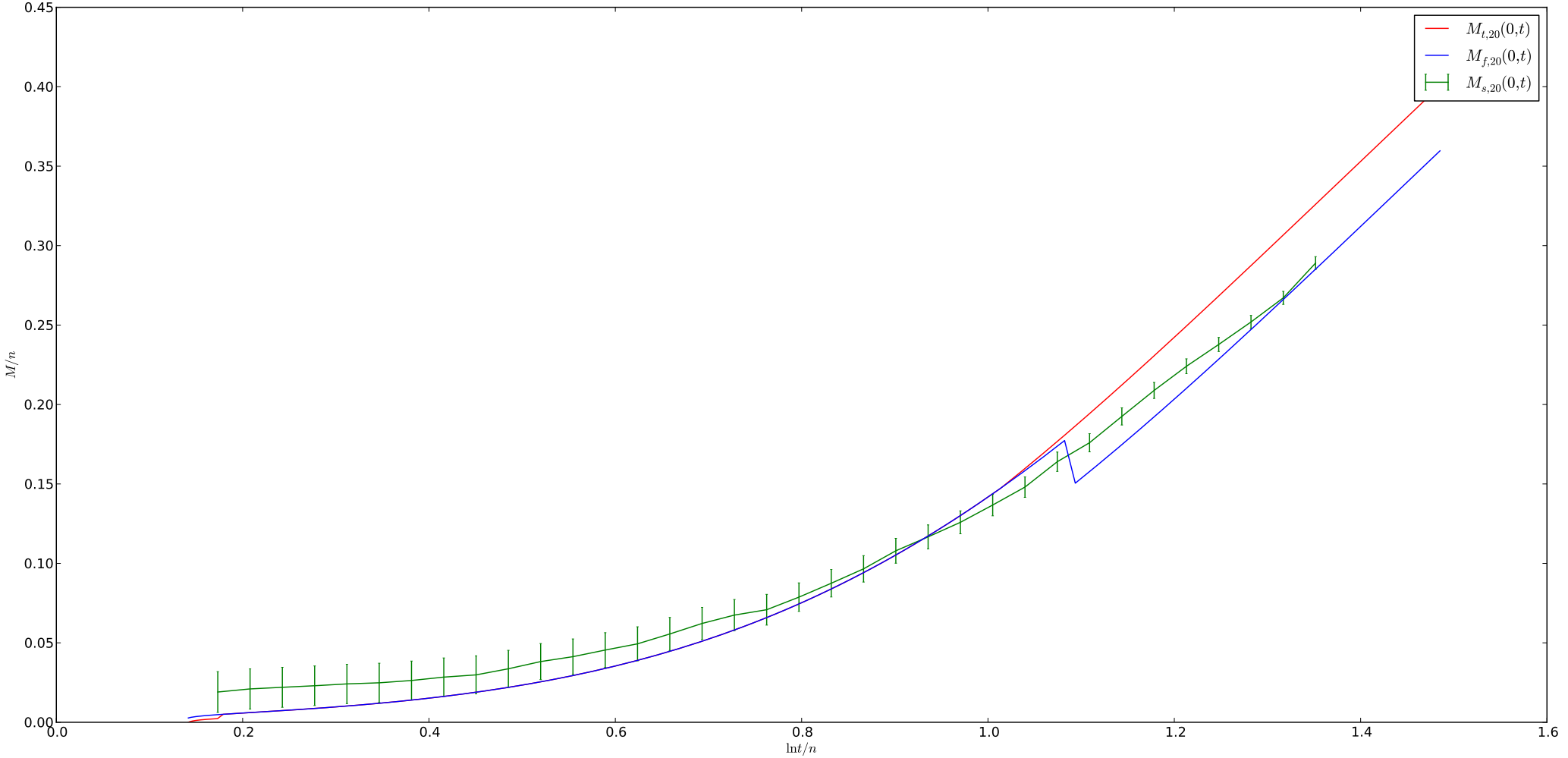

Figure 6.6: $n = 20$, starting at 0

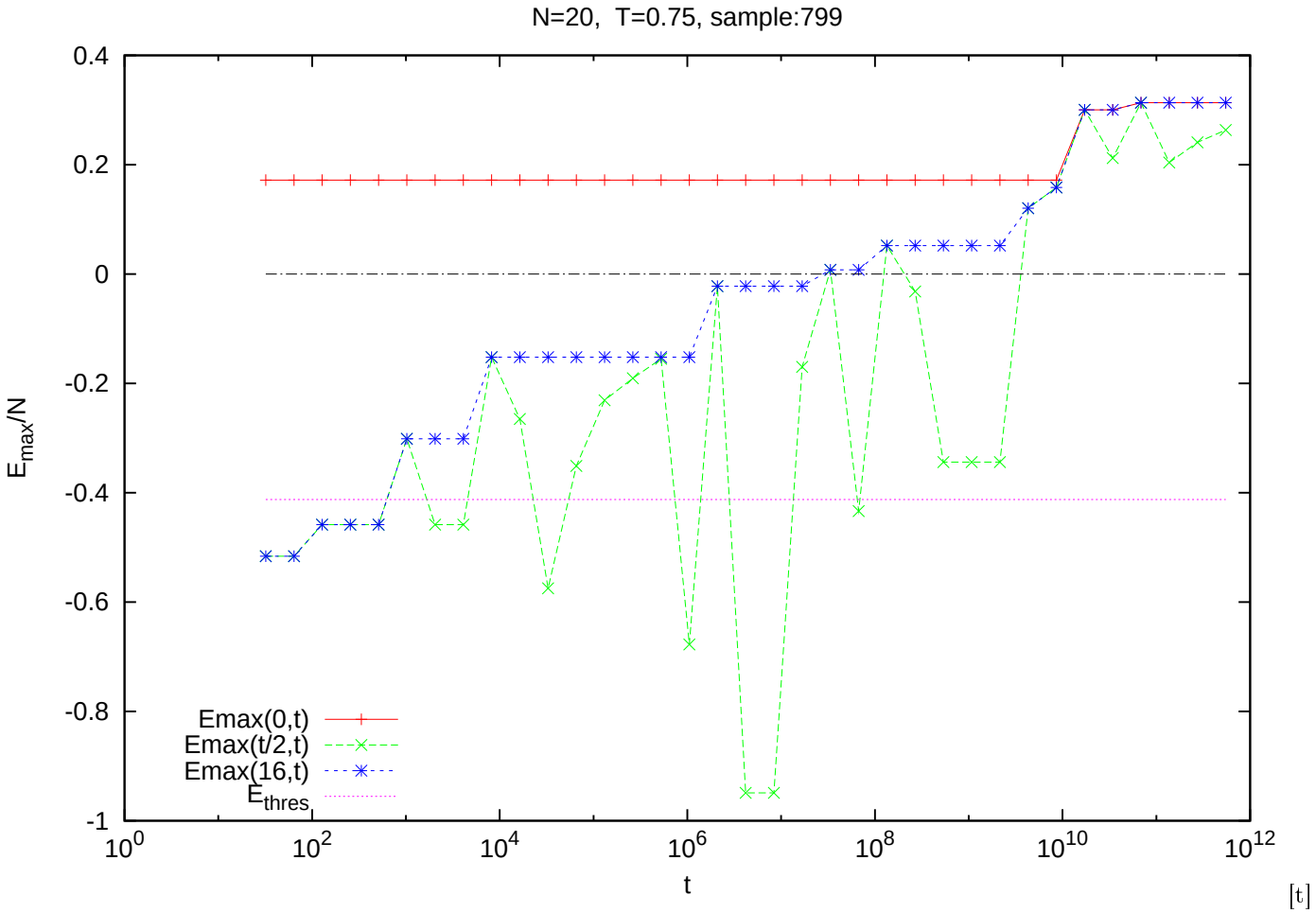

Figure 6.7: A randomly chosen sample

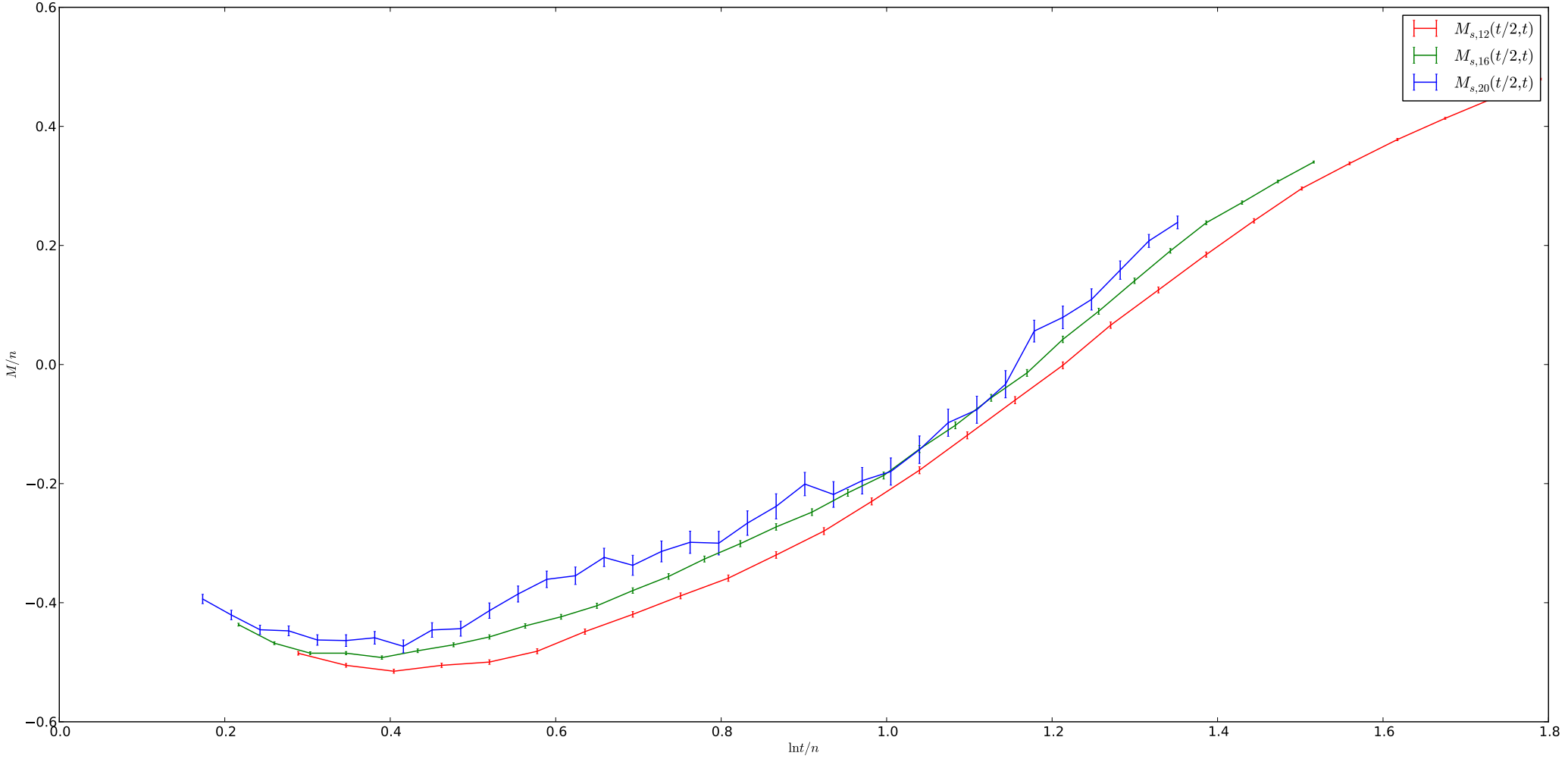

Figure 6.8: Starting at $\frac{t}{2}$

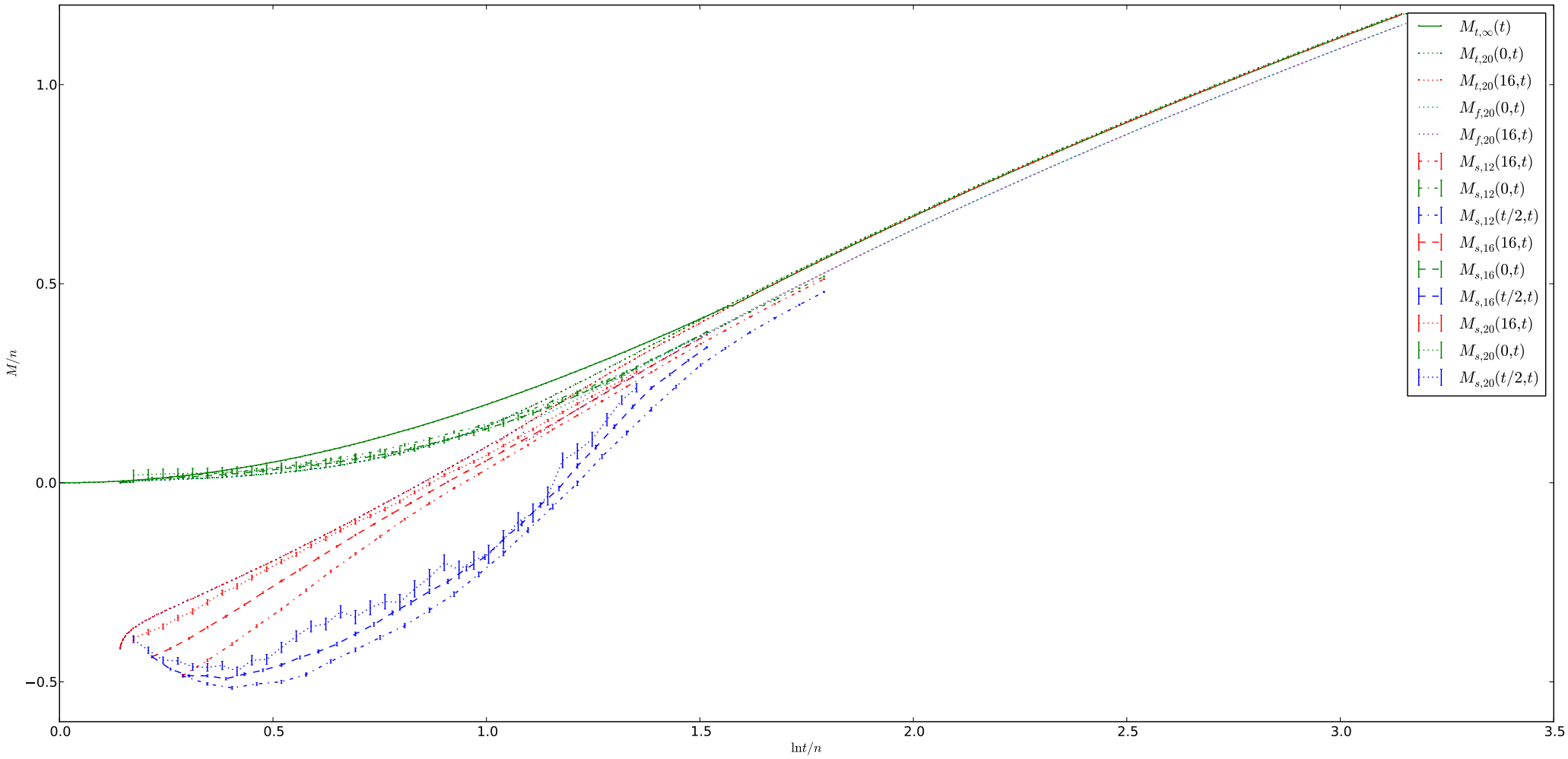

Figure 6.9: All plots

Chapter 7

Conclusion

To conclude, let us first recapitulate what we have seen throughout this work.

The REM is a spin glass model representing a random energy environment where the energy has no spacial correlation in the state space. It can be related to other more realistic models, but its simplicity allows analytical equilibrium treatment as well as mathematically tractable dynamics. One of the natural Glauber dynamics associated with it is the Metropolis one, which favours transitions towards low energies. As we have mentioned in this setting the dynamics restricted to "traps" behaves like the BTM one, which is much simpler, at least as far as aging is concerned. This last very important feature of glasses represents the lack of time invariance of a correlation function. In both models it boils down to the dynamics being trapped progressively in lower and lower energy states for longer and longer lapses so that at all times up to a certain scale where aging eventually ends (for the REM that is the point where the Gaussian is explored sufficiently to find that its tail is not heavy enough to induce aging on the long run) the dynamics has spent much of its time in the deepest trap and is likely to still be there. Furthermore, thanks to non-returning not moving at all is practically the only way for the dynamics to find itself in the same state. This also makes the walk act as though neighbours were drawn on the go. Finally, we saw a few classical results like the convergence of Π towards the arcsine law in different contexts, which establishes aging (and gives much more information). We also discussed the phase transition in the REM at equilibrium and the related condensation of the Gibbs measure to a few states.

From then on we focused exclusively on the evolution of the maximal energy visited rather than the minimal one. In order to justify this we might want to notice that if we look at a "low correlation" function like $\mathbb{P}_E(\exists t' \in [t_w; t_w + t),\ d(X_{t'}, X_t) > an)$, the max becomes significant. Similarly to what we have seen, we should expect that once the global min is attained and setting, for example, $t = t_w$, the function is nearly 0 above the current max $+ \beta_c n$ and nearly 1 below it (for the REM with Metropolis). That is why according to Definition 2.1.4.1 we will also have aging depending on the max. This "extended" aging will last until all the energies are at equilibrium. The reason not to mention this earlier is that this "aging" is somewhat a mathematical artefact caused by the strict interpretation of the definition of aging rather than a physically observable reality, as the physical properties of the system become the equilibrium ones beyond the end of (normal) aging. This idea corresponds to the notion of "microscopic" equilibrium we used. Though this is not of major physical interest, it should be enough to justify our study.

Besides for completeness the only setting of interest for the max is the REM with Metropolis dynamics. For it we need to consider the aging and post-aging stages separately and differentiate between before and after the max outgrows the max of n Gaussians. Then the simplified model we described gives adequate results. Yet, those results are only good for the infinite n limit and numerical implementations require additional corrective terms. The corrections we gave are not redundant and agree almost up to statistical fluctuations with the simulations data, thus suggesting that the underlying model is sufficiently truthful. Yet, it is not precise enough to cover one of the observable max, which we finally left out. If there is one thing to remember from chapter 6, it is that the dynamics is accelerated during the aging stage at finite n in a way dictated by $\inf I$.

Finally, we could suggest a few routes for further development. We already mentioned some features that we left open like $\mathbb{E}[M(t/2, t)]$, the longest life of records and the corrections needed for the expression of the global min for finite n, as well as the extensions mentioned in Appendix A. Furthermore, as always the behaviour near the critical temperature and further corrective terms are worth investigating, although we should recall that the physical content of the max is perhaps too limited to justify such pursuits.

Appendices

Appendix A

Very short time scales

In this appendix we discuss the very short time scales dynamics – before reaching the first trap.

A.1 Results

We consider here the oriented graph

$$G = \left(\mathbb{N} \times \{1, \dots, n\}, \left\{((m,l),(m+1,l')) \mid (l,l') \in \{1,\dots,n\}^2\right\}\right) =: (S, Ed)$$

with Metropolis dynamics (the only possible transitions go in the direction of the edges) associated to E Gaussian IID on S with uniform initial distribution on $\{0\} \times \{1, \dots, n\}$. This corresponds only to the last approximation in the "Underlying model" paragraph of chapter 5. We introduce $T_{\mathcal{N}} := \mathcal{T} := \min\{t \mid E(X_t) < \min E(\Gamma^+(X_t))\}$[1], where $\Gamma^+((m,l)) := \{(m+1,l') \in S\}$ denotes the outbound neighbourhood of (m,l) in G. We also consider its associated Markov chain using the notations we have seen[2]. Note that a.s. in E there are no two equal energies – we restrict ourselves to this event.

We will explain (in a way that can doubtlessly be turned without too much difficulty into a rigorous proof, although we will not give one) why the energy is strictly decreasing until $\mathcal{T}$ and that $\mathcal{T} \simeq n$, while $\mathcal{N} \simeq \ln n$.

A.1.1 Monotony

Consider a current energy E. The states that are n^ϵ above E, where ϵ is small and positive, are impossible to reach, as (we have not yet reached a local min, so we have a lower neighbour) the total transition rate towards them is at most $\mathrm{e}^{-\beta n^\epsilon}$. Therefore, we only need to compare the number of states with lower energies to that of those differing from E by at most n^ϵ, which is much smaller. Hence, it is very likely to go down. This reasoning is valid for one transition and, as we want to show that the dynamics decreases all the time, we need to have few enough jumps for it to still be unlikely to go up even once. That turns out to correspond to a number of jumps about $\sqrt{n}$. It is good to note that this reasoning also shows that the dynamics decreases by at least n^ϵ per jump.

Now that we know that over $\sqrt{n}$ jumps we never go up in energy (provided that we do not reach the first trap before that), we can show that the dynamics goes down fast, which will yield that we reach the first trap far before $\sqrt{n}$ jumps and will thereby prove that the energy actually decreases (fast) all the time (until the first trap). We know that the transitions towards lower energy states are equally probable. As the states of a generation are distributed in the same way for all generations, we tend to go halfway down the distribution (i.e. dividing the rank by 2 in the ordered statistics or, equivalently, dividing by two the remainder of the distribution function below the current energy). That way in around $\log_2 n$ jumps we already find ourselves at a state at $\inf I$, which is the lowest one among its generation (which does not make it a trap, as that depends on the next generation). Once we are there, there is a non-zero probability that the next generation will be entirely comprised of higher energies, so (recall, that we know that we decrease for about $\sqrt{n}$ jumps) in a finite number of jumps (not going to ∞ with n) we will finally reach the first trap without having ascended.

Remark A.1.1.1. We can clearly not state that a.s. for large n the process is decreasing, as is readily seen using that the different n processes are independent and applying Borel-Cantelli only to the first transition, which has a probability of order $\frac{1}{\sqrt{n}}$ to go up.

[1] The first min exists because the trajectories are càdlàg and it is clearly a.s. finite. This is the "a local min is visited" stopping time.
[2] This chain can be seen as defined on $\{1, \dots, n\}$ with time-dependant transition probabilities but we will not take this point of view.

A.1.2 Time for reaching the first trap

We already said that $\mathcal{N} \simeq \ln n$, but in Chapter 5 we were interested in $\mathcal{T}$. It is very easy to see that it takes a time of order n to go to a state of energy $\inf I$, by introducing another Markov process, with two states, one being absorbing and the transition corresponding to going to the lowest neighbour for the first time. Then, the reasoning from above shows that in a few jumps, which take a time of order n each, we get to a trap. Hence, $\mathcal{T} \simeq n$, which is indeed the result we stated in Chapter 5. This means that aging is already in place (recall that towards the last transition we have only one lower neighbour, which means that we take a time of order n to transit to it, which is of the same order as the total time elapsed), even though we have not even reached a trap!

A.2 Extensions

We might want to note a few extensions of the preceding results. To begin with, it is natural to try to make them work on our original model – the REM. To that end, we could add a state to the next generation with the energy of the previous state visited and prove that it is unlikely to be used, which is the case because we go down quickly and so the energy difference to overcome will be too large. Furthermore, we would need to take into account the second common neighbour of two vertices of the hypercube. That is without impact, as it comes down to choosing randomly and uniformly a neighbour of the m-th state, which is not the $m+1$-th and adding it as a neighbour of the $m+2$-th. Finally, we would need a non-returning property to be sure to avoid having any other common neighbours throughout the walk.

Furthermore, it is more than clear that the reasoning is not optimal in the sense that the first phase gives a rather weak result, that we need to use in the second one, which inevitably leads to a weak final result. Hence, if a precise statement was to be made, it could be made reinforced if we limit the number of jumps to $C \ln n$ instead of $\sqrt{n}$ (and take $\epsilon \to 0^+$), as it is all we need and it increases dramatically the probability of the event of decrease.

Finally, we have already mentioned that we show more than mere monotony, as we bounded the increments away from 0. In this line of thought, we might want to attempt to prove stronger results, assuring a certain rate of decrease.

Appendix B

Finite n corrections

B.1 Overall corrections

The overall model remains the same as in chapter 5. However, we give up some approximations – we now say that:

- No state is far from I or deep.
- The trapping times are never relevant – not even for the deepest traps.
- The bounds of I are no longer given by their asymptotic values, but taken "exactly" – that is

$$\pm\sqrt{n}\int E\mathrm{d}(1-\mathrm{erfc})^n, \text{ where } \mathrm{erfc} := \frac{1}{\sqrt{2\pi}}\int_{\bullet}^{\infty} \mathrm{e}^{-\frac{E^2}{2}}\mathrm{d}E. \tag{B.1}$$

  Attention: In what follows erfc denotes this function and not a standard error function, should it differ!

  Remark B.1.0.1. Note that the value of the max of n variables can be very uncertain due to its high variance (for $n = 20$ we get a standard deviation of $0.524\sqrt{n}$, while the mean is $1.865\sqrt{n}$). Hence we need to have already discovered a good number of deep traps before we can identify $\inf I$ with its mean.

  Remark B.1.0.2. We need not consider corrections of the global max and min, as for them we have 2^n variables, which is sufficient for approximating them by their asymptotic values. Yet, the variance is not to be underestimated – it is of order 1 (in fact at infinite n it is of order $\frac{\ln n}{n}$, but we shall disregard that as 1) even for rather small n but the problem is that the smaller n is, the less negligible 1 becomes. Nevertheless, we will not take this into account (we should add the corresponding fluctuations to the scales at which aging ends and at which the global max is attained, as well as to the whole regime between the two for low temperatures.). The effect of that will be discussed in chapter 6.

- We fix the starting energy at E_0 instead of the Gaussian distribution.

  Remark B.1.0.3. In order to get all three max functions with one hit, we only need to consider the usual one (beginning at $t = 0$ without memory loss), but with atypical initial energy E_0 not necessarily of order $\sqrt{n}$ (we forgot all about orders anyway, so this is somewhat redundant) and use the fact that the process is Markovian. Then the delayed max is just the original one starting at the energy we expect to have at the starting time (determined by aging if relevant) and the $\frac{t}{2} - t$ max is the max starting at the energy we expect at $\frac{t}{2}$ and lasting for $\frac{t}{2}$ (we do not actually need to take the expected energy, if we allow ourselves to access the true one).

B.2 Corrected results

- Assume that $E_0 > \inf I$. Then the dynamics goes down to the first trap quickly as in chapter 5 i.e. in $t \simeq n$.
- Let us assume that $E_0 < \inf I$. Then the max goes to $\inf I$ at about $\mathrm{e}^{\beta(\inf I - E_0)}$. Then, as soon as we get out of the initial trap (the time is of the same order), it starts behaving as though we had begun at $E_0 = \inf I$. Note that the time scale we just gave is extremely fluctuating.
- We are thus left with the case $E_0 = \inf I$ and we can assume that we are in a local minimum (otherwise, we say that $E_0 > \inf I$). The typical waiting time there is $\ln t \simeq \beta\sigma$ (with a factor up to 2, say).

Remark B.2.0.1. We should mention that this is rather vague, as we have constantly used $\inf I$, while its value is unstable.

After this initial phase, at the end of which both the min and the max are well out of the zone $[\inf I - \sigma, \inf I + \sigma]$, things get more robust. If we call this time t_0, what we will refer to by t in what follows is in fact $t - t_0$ (if it is large, this does not matter).

- We can approximately say that most of the time is spent in traps and that almost all of the remaining time is spent at $\inf I$ and the time spent at $\inf I$ while at a trap E is the fraction $\mathrm{e}^{\beta(E-\inf I)}$. As far as the min is concerned, nothing changes in the result of d from chapter 5 except for the fact that the time in a trap of energy E is $\mathrm{e}^{\beta(\inf I - E)}$ and so $t \approx \mathrm{e}^{\beta(\inf I - E_{min})}$ (as in d – it is not a tautology). In particular aging ends at $\boxed{\mathrm{e}^{\beta(\inf I + n\beta_c)}}$ and $\boxed{\mathrm{e}^{\beta(\beta n + \inf I)}}$ at low and high temperatures respectively. To be honest, we should multiply the time spent by n, as we only look for the deepest neighbour, but that is compensated by the fact that we should also divide N by n, as we somewhat draw new trap energies only among those lower than $\inf I$. We should be aware that there is a fluctuating factor about $\mathrm{e}^{\pm\beta\sigma}$ until the end of aging (as the deepest trap dominates and all that matters is the single deepest neighbour it has, which does not mean out σ). However, after this regime, this factor gradually dies out to 1 in the high temperature phase and stations at some value (which no longer fluctuates) in the low temperature phase (corresponding to the fluctuation of the lowest neighbour of the global min).

  About half of the states visited are around $\inf I$ with a trap neighbour and without another lower neighbour. The time spent in a state like that is of order n and so we have spent about $\mathrm{e}^{\frac{E_{min}^2}{2n}}$. Then[1][2]

  $$\mathbb{P}_E(E_{max} > E_1) \approx \mathrm{e}^{\frac{E_{min}^2}{2n}} \sqrt{\frac{n}{2\pi}} \int_{E_1}^{\infty} \mathrm{e}^{\beta(\inf I - E)} \mathrm{e}^{-\frac{E^2}{2n}} \mathrm{d}E \approx \mathrm{e}^{\beta(\inf I - E_1)} \mathrm{e}^{-\frac{E_1^2}{2n}} \mathrm{e}^{\frac{E_{min}^2}{2n}}$$

  and so

  $$E_{max} \approx -\beta n + \sqrt{\beta^2 n^2 + 2n\beta \inf I + E_{min}^2} \approx -\beta n + \sqrt{(\beta n + \inf I)^2 - 2\beta^{-1} \inf I \ln t + \beta^{-2} (\ln t)^2}. \tag{B.2}$$

  Once again, note that $\inf I$ may fluctuate by about σ.

- After the end of aging E_{min} is replaced by $n\beta_i$ in the following way:

  $$\mathbb{P}_E(E_{max} > E_1) \approx t \mathrm{e}^{-\beta(\beta_i n + \inf I)} \mathrm{e}^{\frac{n\beta_i^2}{2}} \sqrt{\frac{n}{2\pi}} \int_{E_1}^{\infty} \mathrm{e}^{\beta(\inf I - E)} \mathrm{e}^{-\frac{E^2}{2n}} \mathrm{d}E \approx t \mathrm{e}^{-\beta(n\beta_i + E_1)} \mathrm{e}^{-\frac{E_1^2}{2n}} \mathrm{e}^{\frac{n^2\beta_i^2}{2n}}$$

  and so

  $$\frac{E_{max}}{n} = -\beta + \sqrt{(\beta - \beta_i)^2 + 2\frac{\ln t}{n}}. \tag{B.3}$$

  This is, of course, the same as Equation 5.5, so the bounds given by Equation 5.6 are still valid.

- A final feature that we might want to investigate is when do all three max functions start giving the same thing. The coincidence of the normal max and the delayed one is trivial – when the delayed one gets to the initial value of the process, they stick together. The memory loosing one, however presents an additional aspect – it keeps taking dips from time to time until a certain point and from then on it is equal to the others up to fluctuations[3]. With what we have said it is easy to quantify when is that point – it is when the min is no longer deep enough to hold the process, that is $\ln t = \beta(\inf I + n\beta_c)$ (+a few $\beta\sigma$) and that no matter the temperature.

- In order to be able to produce a full-fledged prediction for the mean (over different energies or, which is the same for the main part, over different trajectories) of the max (starting from 0), we will need to mean out the initial regime of waiting until the max starting at the first trap (we will call it delayed max and denote E_m) reaches the initial value. For a start, we will neglect the probability to have an initial value lower than $\inf I$, the time needed to reach the first trap (of order n) and assume that it is always of energy $\inf I$. Thus,

  $$\mathbb{E}[E_{max}(t)] \approx \frac{1}{\sqrt{2\pi n}} \int_{\inf I}^{\infty} \max(E_m(t), E) \mathrm{e}^{-\frac{E^2}{2n}} \mathrm{d}E = \left(\operatorname{erfc}\left(\frac{\inf I}{\sqrt{n}}\right) - \operatorname{erfc}\left(\frac{E_m}{\sqrt{n}}\right)\right) E_m + \sqrt{\frac{n}{2\pi}} \mathrm{e}^{-\frac{E_m^2}{2n}}.$$

  We see clearly that at E_m well over $\sqrt{n}$ we get E_m. Now let us evaluate the term given by low initial states that we neglected. For them $E_{max}\left(t + \mathrm{e}^{\beta(\inf I - E_0)}\right) \approx E_m(t)$, which gives E_m when the trap is sufficiently shallow and E_0 if it is deep, hence a contribution about $E_m \operatorname{erfc}\left(\frac{-\inf I}{\sqrt{n}}\right)$. Therefore, except at very small t,

  $$\mathbb{E}[E_{max}] \approx \left(1 - \operatorname{erfc}\left(\frac{E_m}{\sqrt{n}}\right)\right) E_m + \sqrt{\frac{n}{2\pi}} \mathrm{e}^{-\frac{E_m^2}{2n}} \tag{B.4}$$

[1] Needless to say, this is a linear approximation, that ceases to be valid before reaching 1.
[2] We get the same result if we say that we go up from E_{min} in average, as the time increase compensates the rate decrease exactly.
[3] Of course, if we insist on exact equality, we will have to wait until a bit after the global max is attained.

and

$$\mathbb{E}[E_{max}] \approx \left(\operatorname{erfc}\left(\frac{\inf I}{\sqrt{n}}\right) - \operatorname{erfc}\left(\frac{E_m}{\sqrt{n}}\right)\right) E_m \approx \ln t \frac{\inf I}{\beta(\beta n + \inf I)^2} \frac{\inf I}{\sqrt{2\pi n}} \mathrm{e}^{-\frac{(\inf I)^2}{2n}} \tag{B.5}$$

at very small t (typically $n < t < \mathrm{e}^{\beta\sqrt{n}}$).

Remark B.2.0.2. It is good to notice that the differences from chapter 5's results are not only corrective terms in complicated expressions. There are serious qualitative differences – for instance at $n = 20$, $\beta = \frac{4}{3}$ we get to the end of aging (and reach macroscopic equilibrium) *before* the max has reached $\sup I$ – it is still just over 0 and in some cases we are even still in the phase when the max has not moved from its initial value (this is because $\sqrt[4]{20} \not\gg 1$, among else)!

B.3 $\mathbb{E}_E\left[M\left(\frac{t}{2}, t\right)\right]$

We will give this last max function the special attention it deserves, as it is (perhaps surprisingly) much more complex than the others (before the end of aging, as we already noted that afterwards it joins the others). Its consideration was useful in samples, as it conveys the most information, but when we average it this diverse information makes it difficult to handle. We give a natural and naïve attempt to treat it, which allows us to see why it requires much more effort.

B.3.1 Decomposition

We consider the events $A := \left\{\forall t' \in \left[\frac{t}{2}, t\right],\ X_{t'} = X_t\right\}$ and $B := \left\{X_t = X_{\frac{t}{2}}\right\}$ (this second one will be interpreted as remaining in the "domain of attraction" of the $\frac{t}{2}$ trap) to get

$$\mathbb{E}_E\left[M\left(\frac{t}{2}, t\right)\right] = \mathbb{E}_E\left[E\left(\frac{t}{2}\right)\mathbb{1}_A\right] + \mathbb{E}_E\left[M\left(\frac{t}{2}, t\right)\mathbb{1}_{B\setminus A}\right] + \mathbb{E}_E\left[M\left(\frac{t}{2}, t\right)\mathbb{1}_{\Omega\setminus B}\right].$$

In the recent paper [17] Gayrard proves the long conjectured fact that[4] $R(t, t+s)$ has the non-trivial limit at $n \to \infty$:

$$\frac{\sin(\pi\alpha)}{\pi}\int_0^{\frac{t}{t+s}} u^{\alpha-1}(1-u)^{-\alpha}\mathrm{d}u,$$

where $\alpha = \frac{\beta_c}{\beta}$ (see the paper for more precision and for the 40-page proof). What is more, her proof seems to imply the same result for Π, which means that $B\setminus A$ is negligible and so the second term is zero. The third one can be dealt with conditioning by the time $t' + \frac{t}{2}$ of exiting the "domain of attraction" (whatever its definition this is a stopping time). Thus it is written

$$-\int_{t'=0}^{\frac{t}{2}} \mathbb{E}_E\left[M_{|E_0=\inf I}\left(0, \frac{t}{2} - t'\right)\right] \mathrm{d}R\left(\frac{t}{2}, \frac{t}{2} + t'\right) =$$

$$= \frac{\sin(\pi\alpha)}{\pi}\int_{t'=0}^{\frac{t}{2}} \left(-\beta n + \sqrt{(\beta n + \inf I)^2 - 2\beta^{-1}\inf I \ln\left(\frac{t}{2} - t'\right) + \beta^{-2}\left(\ln\left(\frac{t}{2} - t'\right)\right)^2}\right)\left(\frac{t}{2t'}\right)^{\alpha}\frac{2}{t+2t'}\mathrm{d}t'.$$

As $\mathbb{E}_E\left[M\left(16, \frac{t}{2} - t'\right)\right]$[5] varies slowly with t' save for $2t' \approx t$ (which is also diminished by the $\frac{t}{2t'}$ factor), that simplifies to

$$\mathbb{E}_E\left[M\left(16, \frac{t}{2}\right)\right]\frac{\sin(\pi\alpha)}{\pi}\int_0^1 \frac{1}{(1+x)x^{\alpha}}\mathrm{d}x$$

B.3.2 No jump term

A crude approach Now let us address the most problematic term $\mathbb{E}_E\left[E\left(\frac{t}{2}\right)\mathbb{1}_A\right]$. To begin with, we note that the two factors are by far not independent (clearly if the energy is low, we are unlikely to jump and vice versa). For this one we condition by $E\left(\frac{t}{2}\right)$ in order to obtain

$$\int_{\mathbb{R}} E'\mathbb{E}_E\left[\mathrm{e}^{-a(E')\frac{t}{2}}\right]\mathrm{d}\mathscr{L}_{\frac{t}{2}}(E') \approx \int_{-\infty}^{\inf I} E'\mathrm{e}^{-\mathrm{e}^{-\beta(\inf I - E')}\frac{t}{2n}}\mathrm{d}\mathscr{L}_{\frac{t}{2}}(E') + O\left(\mathrm{e}^{-\frac{t}{2n}\inf I}\right),$$

where $\mathscr{L}_t$ denotes the law of the energy at time t conditional on the random environment and $a(E')$ (we assume it to be a function only of the energy, which it is not) is the exit rate from a state of energy E (conditional on the environment).

[4] Please do note that what we announce here is not her result but rather a loose interpretation of it. For a simpler statement, see 3.2.2.1.
[5] We shall refer to $M_{|E_0=\inf I}(0, t')$ as $M(16, 16+t')$ and eventually omit the second 16, as for $n = 20$ it is what we use in practice.

The problem is that we do not know that law. We might be tempted to approximate it by a natural candidate – the equilibrium distribution between $\mathbb{E}_E\left[E_{min}\left(\frac{t}{2}\right)\right] = \inf I - \beta^{-1}\ln\frac{t}{2}$ and $\mathbb{E}_E\left[M\left(16,\frac{t}{2}\right)\right]$ and 0 elsewhere. Thus we get

$$\frac{\int_{\inf I-\beta^{-1}\ln\frac{t}{2}}^{\inf I} E'\mathrm{e}^{-\mathrm{e}^{-\beta(\inf I-E')}\frac{t}{2n}}\mathrm{e}^{-\beta E'}\mathrm{e}^{-\frac{E'^2}{2n}}\mathrm{d}E'}{\int_{\inf I-\beta^{-1}\ln\frac{t}{2}}^{\mathbb{E}_E[M(16,\frac{t}{2})]}\mathrm{e}^{-\beta E'}\mathrm{e}^{-\frac{E'^2}{2n}}\mathrm{d}E'} \approx \left(\inf I-\beta^{-1}\ln\frac{t}{2}\right)\frac{\int_{\inf I-\beta^{-1}\ln\frac{t}{2}+\beta n}^{\infty}\mathrm{e}^{-\frac{E'^2}{2n}}\mathrm{e}^{-\mathrm{e}^{-\beta(\inf I-E'+\beta n)}\frac{t}{2n}}\mathrm{d}E'}{\exp\left(-\frac{(\inf I-\beta^{-1}\ln\frac{t}{2}+\beta n)^2}{2n}\right)\frac{n}{\inf I-\beta^{-1}\ln\frac{t}{2}+\beta n}},$$

which simplifies to $\inf I - \beta^{-1}\ln\frac{t}{2}$, because the double exponential factor is 1 near the lower bound where most of the Gaussian is concentrated (we use $\ln n \gg 1$ and $\ln t \simeq n$).

Problems Now let us explain why this reasoning is not legitimate. The problem is that we used in a crucial way our "knowledge" of $\mathbb{E}_E\left[E_{min}\left(\frac{t}{2}\right)\right]$ with precision we do not have (we only know an equivalent of it but we used the same expression up to a relative error of order $\frac{\ln n}{n}$). What is more, we somewhat arbitrarily postulated that the lower bound is the mean of the minimum rather than another similar quantity – the mode for instance. If we suppose (reasonably) that $\frac{2n}{t}\mathrm{e}^{\beta\left(\inf I-E_{min}\left(\frac{t}{2}\right)\right)}$ has a non-trivial law, this last choice can lead to a multiplicative factor in the final result, which we do not tolerate. Hence, the usage of this method requires far better knowledge of the law of the energy at $\frac{t}{2}$. Logically enough, this is confirmed numerically – the result above is very different from the one obtained if we leave out the n from $\frac{t}{2n}$ in the exponential (which is precisely a relative error of order $\frac{\ln n}{n}$ on E_{min}).

Yet another startling issue is the upper bound in the denominator integral. Numerically we find out that at rather short (but still not very short) scales it has an influence on the result (recall that we are at finite n). We then need to remain with the first and more complicated expression above (the integral quotient). Clearly, this problem should disappear at big n but in our case it is still very visible.

To sum up, $\mathbb{E}_E\left[M\left(\frac{t}{2},t\right)\right]$ remains beyond our reach (even at $n\to\infty$) with the methods outlined due to insufficient precision on the law of the energy at $\frac{t}{2}$ near its lower "edge" (we should use a result like the one in subsection 4.1.3, which we will not attempt here).

An alternative idea Let us now describe another idea suggested by Biroli, which manages to exploit the result of Gayrard in a way that circumvents any finer knowledge of $\mathscr{L}_{\frac{t}{2}}$ and does not use the crude cut-off approximation. To that end we turn back to $\int_{\mathbb{R}} E'\mathbb{E}_E\left[\mathrm{e}^{-a(E')\frac{t}{2}}\right]\mathrm{d}\mathscr{L}_{\frac{t}{2}}(E')$ and notice that, as we saw before the double exponential behaves roughly as 1 up to about E_{min} and 0 afterwards, while $\mathscr{L}_{\frac{t}{2}}$ should be concentrated around E_{min} (recall that we expect for E_{min} to have a variance $\simeq 1$ at time scales $\ln t \simeq n$ by the extremes). Then we can approximate the integral by

$$\mathbb{E}_E[E_{min}]\int\mathbb{E}_E[\mathrm{e}^{-a(E')\frac{t}{2}}]\mathrm{d}\mathscr{L}_{\frac{t}{2}}(E') \approx \left(\inf I-\ln t\beta^{-1}\right)\Pi\left(\frac{t}{2},t\right) \approx \left(\inf I-\ln t\beta^{-1}\right)\frac{\sin(\pi\alpha)}{\pi}\int_0^{\frac{1}{2}}\left(\frac{u}{1-u}\right)^{\alpha}\frac{\mathrm{d}u}{u}.$$ [6]

We should note that all of this boils down to writing (in a slightly justified manner)

$$\mathbb{E}_E[M(t/2,t)] \approx \mathbb{E}_E\left[M\left(16,\frac{t}{2}\right)\right]\left(1-\Pi\left(\frac{t}{2},t\right)\right)+\Pi\left(\frac{t}{2},t\right)\mathbb{E}_E\left[E_{min}\left(\frac{t}{2}\right)\right]. \tag{B.6}$$

[6] Note that we ought not use the convergence result established at $n\to\infty$ in conjunction with the finite n correction but it is our best bet.

Bibliography

[1] Stephan Albert, Thomas Bauer, Marion Michl, Giulio Biroli, Jean-Philippe Bouchaud, Alois Loidl, Peter Lunkenheimer, Roland Tourbot, Cecile Wiertel-Gasquet, and François Ladieu. Fifth-order susceptibility unveils growth of thermodynamic amorphous order in glass-formers. *Science*, 352(6291):1308–1311, 2016.

[2] Alain Barrat and Marc Mézard. Phase space diffusion and low temperature aging. *J. Phys. I France*, 5(8):941–947, 1995.

[3] Gérard Ben Arous, Anton Bovier, and Véronique Gayard. Aging in the Random Energy Model. *Physical Review Letters*, 88(087201), February 2002.

[4] Gérard Ben Arous and Jiří Černý. Dynamics of trap models, February 2006.

[5] Gérard Ben Arous and Alexey Kuptsov. *Spin Glasses: Statistics and Dynamics*, volume 62 of *Progress in Probability*, chapter REM Universality for Random Hamiltonians, pages 45–84. Birkhäuser Verlag AG, 2009.

[6] Eric Bertin. Entropic aging and extreme value statistics. *Journal of Physics A: Mathematical and Theoretical*, 43(34):345002, 2010.

[7] Giulio Biroli. Non-linear responses, soft-modes and the true nature of glasses, December 2015. Talk available at: http://www.phys.ens.fr/spip.php?article2214.

[8] Jean-Philippe Bouchaud. Weak ergodicity breaking and aging in disordered systems. *Journal de Physique I*, 2:1705–1713, September 1992.

[9] Jean-Philippe Bouchaud and David Dean. Aging on Parisi's Tree. *Journal de Physique I*, 5(3):265–286, 1995.

[10] Jean-Philippe Bouchaud and Marc Mézard. Complex systems, 2013. Lecture notes.

[11] Anton Bovier. An introduction to ageing, 2007. Lecture notes.

[12] Chiara Cammarota and Enzo Marinari. Spontaneous energy-barrier formation in entropy-driven glassy dynamics. *Phys. Rev. E*, 92:010301, Jul 2015.

[13] Bernard Derrida. Random-energy model: An exactly solvable model of disordered systems. *Physical Review B*, 24:2613, September 1981.

[14] Thomas Duquesne. Théorie de l'intégration. probabilités de base, 2015.

[15] Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 4.1 edition, 2013.

[16] Samuel Edwards and Philip Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965, 1975.

[17] Véronique Gayrard. Aging in Metropolis dynamics of the REM: a proof. working paper or preprint, February 2016.

[18] David Gross and Marc Mézard. The simplest spin glass. *Nuclear Physics B*, 240, November 1984.

[19] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.

[20] Jean Jacod. Chaînes de Markov, processus de Poisson et applications, 2004.

[21] Nicola Kistler. Derrida's random energy models: From spin glass to the extremes of correlated random fields, December 2014.

[22] Werner Krauth. *Statistical Mechanics: Algorithms and Computations*, volume 13 of *Oxford Master Series in Statistical, Computational, and Theoretical Physics*. Oxford University Press, 2006.

[23] Jean-François Le Gall. Intégration, probabilités et processus aléatoires, September 2006.

[24] Marc Mézard and Andrea Montanari. *Information, Physics and Computation*. Oxford Graduate Texts. Oxford Universitiy Press, 2009.

[25] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific, 1987.

[26] Cécile Monthus and Jean-Philippe Bouchaud. Models of traps and glass phenomenology. *Journal of Physics A: Mathematical and General*, 29(14):3847, 1996.

[27] Piere Papon, Jacques Leblond, and Paul H.E. Meijer. *The Physics of phase transitions: Concepts and applications*. Springer, 2 edition, 2006.

[28] Grégory Schehr and Satya Majumdar. *Exact Record and Order Statistics of Random Walks via First-Passage Ideas*, pages 226–251. World Scientific, 5 2014.

[29] Guilhem Semerjian. Mean-field disordered systems: glasses and optimization problems, classical and quantum, January 2013. Thèse d'habilitation à diriger des recherches.

[30] Richard Serfozo. *Basics of Applied Stochastic Processes*. Probability and Its Apllications. Springer, 2009.

[31] David Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35:1792–1795, 1975. doi:10.1103/PhysRevLett.35.1792.