Metastability of Markov Chains and in the Hopfield Model

vorgelegt von
Diplom-Mathematiker
Matthias an der Heiden

Von der Fakultät II - Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
Dr.rer.nat.

genehmigte Dissertation

Berichter:

Prof. Dr. Anton Bovier
Prof. Dr. Jürgen Gärtner
Prof. Dr. Matthias Löwe

Tag der wissenschaftlichen Aussprache: 2. November 2006

Berlin 2006

D83

für Wioletta

Danksagung

Ich möchte mich bei Prof. Anton Bovier für die fruchtbare Zusammenarbeit und die zur Selbständigkeit anleitende Betreuung bedanken.

Während meiner Promotion habe ich die angenehme Atmosphäre im Berliner Graduiertenkolleg "Stochastische Prozesse und Probabilistische Analysis" sehr geschätzt. Für die souveräne und stets freundliche Ausfüllung der Sprecherrolle, auch auf den zahlreichen Winterschulen, danke ich Prof. Michael Scheutzow.

Ich bin Prof. Jürgen Gärtner und Prof. Matthias Löwe sehr dankbar, dass sie zugesagt haben, diese Arbeit zu begutachten.

Sehr herzlich danke ich meinen Arbeitskollegen und Freunden für viele hilfreiche Diskussionen und die gemeinsame Zeit, insbesondere Stefan Ankirchner, Peter Arts, Andrej Depperschmidt, Steffen Dereich, Georgi Dimitroff, Alexander Drewitz, Felix Esche, Mesrop Janunts, Anton Klimovsky, Pierre-Yves Louis, Max von Renesse, Alla Slyncko, Florian Sobieczky, Michel Sortais, Stephan Sturm, Christian Vormoor, Stefan Weber, Wiebke Wittmüß und Henryk Zähle.

Schließlich danke ich der Deutschen Forschungsgemeinschaft für die finanzielle Unterstützung.

Zusammenfassung

Diese Dissertation behandelt das metastabile Verhalten von Markov Ketten mit abzählbarem diskreten Zustandsraum. Im ersten Teil betrachten wir Markov Ketten, die reversibel bezüglich eines vorgegebenen Wahrscheinlichkeitsmaßes π_ϵ sind. Der (kleine) Parameter $\epsilon \in (0,1)$ erlaubt es uns, im Rahmen des potentialtheoretischen Ansatzes von Bovier, Eckhoff, Gayrard und Klein Metastabilität rigoros zu definieren und nachzuweisen. Der wichtigste Begriff in diesem Ansatz ist die (Newtonsche) Kapazität einer Markov Kette. In einem ersten Schritt zeigen wir subexponentielle Abschätzungen dieser Größe unter sehr allgemeinen Bedingungen.

Das Hauptergebnis des ersten Teils liefert eine genaue Asymptotik der Kapazität unter restriktiveren Bedingungen an die Markov Kette und ihr reversibles Maß. Unter zu Hilfenahme bereits bekannter Ergebnisse können wir daraus die Eyring-Kramers Formel herleiten, die die Asymptotik bestimmter erwarteter Eintrittszeiten der Markov Kette angibt.

Im zweiten Teil werden diese Resultate auf das Hopfield Modell mit einer festen Anzahl M von gelernten Mustern angewandt. Für die Komponenten dieser Muster wählen wir unabhängige und gleichverteilte Zufallsvariablen. Wir möchten das Verhalten für große Anzahlen N von Neuronen beschreiben. Dabei modellieren wir die Dynamik mittels einer Markov Kette vom Glauber Typ, die reversibel bezüglich des Gibbsmaßes des Hopfield Modells ist.

Durch die Einführung von Blockspinvariablen erhalten wir eine Markov Kette ζ_N auf einer Teilmenge eines 2^M-dimensionalen Gitters. Für ζ_N können wir eine metastabile Menge bestehend aus $2M$ Punkten angeben, wobei jeder Punkt zu Konfigurationen in der Nähe eines der Muster oder seines Negativs gehört.

Wir zeigen, dass für Übergänge zwischen diesen metastabilen Punkten die Eyring-Kramers Formel gilt. Die asymptotisch erwarteten Eintrittszeiten können hierbei explizit angegeben werden, da wir in einem (sehr kleinen) Temperaturintervall alle essentiellen Sattelpunkte genau bestimmen können. Diese Punkte bleiben Kandidaten für die essentiellen Sattelpunkte bis zu einem bestimmten Temperatur-Schwellenwert.

Mit den gleichen Einschränkungen an die Temperatur können wir schließlich die genaue Struktur und Größe der kleinsten Eigenwerte des Generators von ζ_N bestimmen. Aufgrund der Spin-Flip Symmetrie und der anomal kleinen Schwankungen der Grundzustände des Hopfield Models muss die Täler Struktur des transformierten Hamiltonians berücksichtigt werden.

Summary

This thesis is concerned with the metastable behaviour of time homogeneous Markov chains evolving on a discrete countable set. In the first part, we consider Markov chains that are reversible with respect to a given probability measure π_ϵ. The small parameter $\epsilon \in (0,1)$ allows us to investigate metastability rigorously in the sense of the potential theoretic approach due to Bovier, Eckhoff, Gayrard and Klein. The main notion in this approach is the capacity of a Markov chain. We are able to show subexponential bounds on this quantity under very general assumptions and for a big class of discrete countable sets.

The main theorem in the first part yields, under more restrictive conditions, precise asymptotics of the capacity with multiplicative errors that tend to one. As a consequence we can prove the Eyring-Kramers formula providing sharp estimates for certain expected hitting times of our Markov chain. They exhibit the same form as in the case of a diffusion with small noise intensity on a subset of $\mathbb{R}^d$.

In the second part we apply our results to the Hopfield model with a fixed number, say M, of random patterns. We are interested in the behaviour for a large number, N, of neurons. The dynamics are modelled by a Markov chain of Glauber type on the set of all configurations, $\{-1,1\}^N$, which is reversible with respect to the Gibbs measure associated to the Hopfield Hamiltonian. With the help of a lumping procedure, we obtain a random Markov chain ζ_N on a subset of a lattice with dimension 2^M. We can construct a metastable set of ζ_N consisting of $2M$ points that correspond to configurations near one of the patterns or its negative.

Then we establish the Eyring-Kramers formula for transitions between these metastable points. We obtain a completely explicit expression since we can estimate precisely the (random) position and height of the relevant saddle points. However, this holds only in a very small intervall of the temperature, and it is an open question whether this result may be extended up to a certain temperature threshold. For temperatures that are even lower we are sure that the behaviour changes.

With the same restrictions on the temperature we are able to unravel the structure of the low lying eigenvalues of the generator of ζ_N. Due to the spin flip symmetry and the anomalously small random fluctuations of the ground states we have to take into account the valley structure of the transformed Hamiltonian.

Contents

Part I

Introduction

1 Metastability of Markov chains

This work is concerned with the metastable behaviour of time homogeneous Markov chains, $\zeta = (\zeta_n)_{n\in\mathbb{N}_0}$, evolving on a discrete countable set Y. We call Y the *state space* of ζ. Assume ζ is irreducible; then it is positive recurrent if and only if there exists a stationary probability distribution, π. In this case it is called ergodic. It then follows that π is unique and positive .[1] Assume now that ζ is ergodic. Then for any $f : Y \to \mathbb{R}$ such that

$$\pi\left(|f|\right) < \infty \tag{1.0.1}$$

and for any initial distribution μ, the pathwise ergodic theorem states that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} f\left(\zeta_k\right) = \pi\left(f\right). \tag{1.0.2}$$

One says that the Markov chain ζ converges to its equilibrium π. The main question we are confronted with is how long does this take and how does the Markov chain proceed in order to approach the equilibrium.

We will be investigating Markov chains that need an exponentially long time, measured on a certain scale, to come close to the equilibrium. For finite chains, the time can be measured on the scale of $|Y|$, the number of states, in general we will be using the inverse of a small parameter ϵ.

One of the main motivations behind such studies comes from the attempt to understand phenomena of non-equilibrium thermodynamics for (disordered) interacting particle systems: Consider an interacting particle system with N particles, whose equilibrium is described by the associated (random) Gibbs measure on an asymptotically (for $N \to \infty$) infinite dimensional space $\mathcal{S}^N$. For $\mathcal{S} = \{-1, 1\}$ one has an interacting spin system. In order to observe how this system converges to equilibrium, we introduce a particular kind of dynamics, namely a discrete time Markov chain that flips at most one spin per time step. As we shall see in part III, in the important case of the Hopfield model, one can use symmetries to map this Markov chain to another Markov chain on a subset of a finite dimensional lattice and apply our results.

[1] See e.g. Brémaud, [Bré99] Theorem 3.1, p. 104 and, for the next statement, Theorem 4.1, p. 111.

1.1 General methods

Define the *hitting time* of a subset $A \subset Y$ to be

$$\tau_A := \inf \{n > 0 \,|\, \zeta_n \in A\}. \tag{1.1.1}$$

To get a hold on the evolution of our Markov chain we isolate certain characteristic points of the state space Y and give precise estimates of the expected hitting times of these points. Working with points is certainly only possible in a space having at most countably many elements.

Let us say a word about the methods we are using. At the heart of our treatment lies the Dirichlet principle and a stochastic representation of certain harmonic functions. We will briefly introduce these concepts now:

Denote by P_μ the law of ζ with μ as starting distribution. If ζ starts at point $x \in Y$, we also write P_x. Let p be the transition probability of ζ and $L := p - \mathbb{1}$ the generator of ζ. Let A and B be disjoint compact subsets of Y. The *equilibrium potential* $h_{A,B} : Y \to [0,1]$ of ζ is defined to be the unique bounded solution of the Dirichlet problem

$$\begin{cases} Lh = 0 & \text{on } (A \cup B)^c, \\ h = 1 & \text{on } A, \\ h = 0 & \text{on } B. \end{cases} \tag{1.1.2}$$

We also say that $h_{A,B}$ is *harmonic* on $Y \setminus (A \cup B)$ with respect to L. It is well known (see e.g. [Bré99], Theorem 2.1, p. 181) that $h_{A,B}$ has the stochastic representation

$$h_{A,B}(x) = P_x(\tau_A < \tau_B) \qquad \text{for all } x \in Y \setminus (A \cup B). \tag{1.1.3}$$

The quadratic form associated with L, namely

$$\Phi(h) := -\langle h, Lh \rangle_\pi, \tag{1.1.4}$$

is called *Dirichlet form*. We now consider the space of $l_2(\pi)$-functions having the same boundary conditions as $h_{A,B}$, i.e.

$$\mathcal{H}_{A,B} := \left\{ h \in l_2(\pi) \;|\; h_{|A} = 0 \text{ and } h_{|B} = 1 \right\}. \tag{1.1.5}$$

Then the Dirichlet principle asserts that the infimum of Φ under all functions of $\mathcal{H}_{A,B}$ is attained by the equilibrium potential $h_{A,B}$. The minimum value is called the *(Newtonian) capacity between A and B*,

$$\operatorname{cap}(A, B) := \Phi(h_{A,B}). \tag{1.1.6}$$

This theorem is the analogue for reversible Markov chains of the classical Dirichlet principle from potential theory, which states the following: among all continuously differentiable functions h on a smooth bounded domain $D \subset \mathbb{R}^d$ taking specific boundary values, the integral

$$\int_D |\nabla h(x)|^2 \, dx \tag{1.1.7}$$

is minimised by the harmonic functions taking these boundary values. In the discrete setting, one can further show that

$$P_x(\tau_A < \tau_x) = \frac{\operatorname{cap}(x, A)}{\pi(x)}, \tag{1.1.8}$$

and this can be applied recursively together with the Dirichlet principle.

Since our techniques so much depend on the Dirichlet principle, we have to restrict ourselves to reversible Markov chains.

1.2 Defining metastability

A fundamental property of the Markov chains we investigate is their so called *metastability*. First of all, let us give an informal description of this phenomenon:

A Markov chain is said to exhibit metastable behaviour if, firstly, when starting in a certain subset of initial conditions, the chain remains for a "long" time in a limited subset of the state space. Secondly, this subspace has negligible measure in equilibrium. And thirdly, the transition to equilibrium or to another (larger) subspace occurs in an abrupt fashion.

Obviously metastability is a dynamical phenomenon that can only be observed on certain timescales. A dynamical definition of metastability has been suggested by Davies in [Dav82]. The requirement that the process spend a large time in a restricted subset domain, implies that the chain relaxes to a pseudo-equilibrium state. Thus, in a metastable state, the values of the macroscopic observables of interest will not show any systematic time-dependence, at least after some short initial transient effect.

Gaveau and Schulman revealed in [GS98] the intimate relation between metastable time scales and the low lying eigenvalues of the generator of a

Markov chain. In a series of papers starting with [BEGK01] Bovier, Eckhoff, Gayrard and Klein could rigorously verify a very precise form of this relation for reversible Markov chains and trace it back to their definition of metastability. They developed the so called potential theoretic approach to metastability.

In order to come to a precise mathematical description, we introduce a small positive parameter ϵ, that enables us to zoom into this picture and amplify the details we are interested in.

We let $\Lambda \subset \mathbb{R}^d$, whereas $(\Lambda_\epsilon)_{0<\epsilon<1} \subset \Lambda$ is a family of countable discrete sets. Let $(\xi^\epsilon)_{0<\epsilon<1}$ be a family of irreducible homogeneous Markov chains such that ξ^ϵ is positive recurrent on Λ_ϵ. Denote by P_μ the law of ξ^ϵ conditioned to have μ as starting distribution. If ξ^ϵ starts in a point $x \in \Lambda_\epsilon$, we also write P_x. We denote the stationary distribution of ξ^ϵ by π_ϵ. Such Markov chains can be fully characterised by specifying their transition matrix p_ϵ.

In the following, we will often be dealing with probabilities like $P_x[\tau_A < \tau_x]$ which we call *escape probability* from x to A.

Following Bovier, Eckhoff, Gayrard and Klein [BEGK02], we define metastability in the following way:

Definition 1.1 (metastability) Let $\mathscr{M}_\epsilon$ be a finite subset of Λ_ϵ such that the cardinality $|\mathscr{M}_\epsilon|$ is independent of ϵ. Let $\rho : (0,1) \to (0,1)$ be a monotone increasing function with $\lim_{\epsilon \downarrow 0} \rho(\epsilon) = 0$.

Then the family of Markov processes $(\xi^\epsilon)_{\epsilon \in (0,1)}$ is said to be *ρ-metastable* with respect to $(\mathscr{M}_\epsilon)$, if

$$\max_{m \in \mathscr{M}_\epsilon} P_m(\tau_{\mathscr{M}_\epsilon} < \tau_m) \leq \rho(\epsilon) \inf_{x \notin \mathscr{M}_\epsilon} P_x(\tau_{\mathscr{M}_\epsilon} < \tau_x). \tag{1.2.1}$$

The elements of $\mathscr{M}_\epsilon$ are called *ρ-metastable points* of (ξ^ϵ).

We say (ξ^ϵ) is metastable with respect to $(\mathscr{M}_\epsilon)$ if there exists a function ρ for which (ξ^ϵ) is ρ-metastable.

This definition suggest a decomposition of the state space into a finite collection of subsets. We define for each point $m \in \mathscr{M}_\epsilon$ the domain of attraction of m by

$$A(m, \mathscr{M}_\epsilon) := \left\{ x \in \Lambda_\epsilon \,|\, P_x(\tau_m = \tau_{\mathscr{M}_\epsilon}) \geq \max_{n \in \mathscr{M}_\epsilon} P_x(\tau_n = \tau_{\mathscr{M}_\epsilon}) \right\}. \tag{1.2.2}$$

In words, Definition 1.1 states the following: The infimum of the escape probabilities from any point $x \in \mathscr{M}_\epsilon^c$ to $\mathscr{M}_\epsilon$ is much bigger than the escape probability from a point, m, in $\mathscr{M}_\epsilon$ to another one. The function ρ in Definition 1.1 describes the factor, by which the escape probabilities between metastable points is smaller compared to the escape probability of any point with respect to the set of metastable points.

Therefore we have at least two different time scales: One that measures the time required for a typical excursion away from m that stays inside $A(m, \mathscr{M}_\epsilon)$ and another one on which we expect a changeover to $\mathscr{M}_\epsilon \setminus m$. This type of behaviour has been studied for a long time and is rigorously treated on the level of large deviations, in particular in the book of Freidlin and Wentzell [FW84].

The benefit of Definition 1.1 is that we only have to control hitting times of points or finite sets of points on the state space. In the analogues situation of a Diffusion in $\mathbb{R}^d$, one can deal with small balls around these points (see [BEGK04]).

Observe that Definition 1.1 does not determine a unique family $(\mathscr{M}_\epsilon)$ even for fixed ρ. Indeed, having isolated a very large set $\mathscr{M}_\epsilon$, in many cases one can find a subset $\mathscr{N}_\epsilon \subset \mathscr{M}_\epsilon$ such that the Markov chain also exhibits a metastable behaviour with respect to $\mathscr{N}_\epsilon$. We formulate this important property of Definition 1.1 in

Proposition 1.2 *Assume we have choose the set $\mathscr{M}_\epsilon$ such that*

$$\pi_\epsilon(m) = \max_{x \in A(m, \mathscr{M}_\epsilon)} \pi_\epsilon(x). \tag{1.2.3}$$

Let I_ϵ be the set of all $i \in \mathscr{M}_\epsilon$ for which there exists $c > 0$, independent of ϵ, such that

$$P_i(\tau_{\mathscr{M}_\epsilon} < \tau_i) \geq c \max_{m \in \mathscr{M}_\epsilon} P_m(\tau_{\mathscr{M}_\epsilon} < \tau_m). \tag{1.2.4}$$

Then we can construct a minimal set $J_\epsilon \subset I_\epsilon$ such that (ξ^ϵ) is metastable with respect to $\hat{\mathscr{M}}_\epsilon \equiv \mathscr{M}_\epsilon \setminus J_\epsilon$.

We will use this reduction mechanism in part III to find the low lying eigenvalues for the generator. A similar argument has been used by Bovier et al. in [BEGK02].

A striking example of the power of Definition 1.1, and the associated potential theoretic approach, is the recent work of Bovier, den Hollander and Nardi [BdHN06], about the metastable behaviour of a lattice gas subject to Kawasaki dynamics in two or three dimensions in the limit of low temperature and low density.

1.3 Estimation of the capacity

We need to introduce some notions about the structural properties of the equilibrium measure π_ϵ.

Definition 1.3 Since π_ϵ is positive, we can define the *potential* $F_\epsilon : \Lambda_\epsilon \to \mathbb{R}_{>0}$ by

$$F_\epsilon(x) := -\epsilon \ln \pi_\epsilon(x). \tag{1.3.1}$$

We now assume that (F_ϵ) converges uniformly to a unique continuous function $F : \Lambda \to \mathbb{R}_{\geq 0}$, i.e. for all $\kappa > 0$ there exists $\epsilon_0 > 0$ such that for all $\epsilon < \epsilon_0$ we have

$$\sup_{x \in \Lambda_\epsilon} |F_\epsilon(x) - F(x)| < \kappa. \tag{1.3.2}$$

Moreover, we assume that F has compact level sets, i.e.

$$\{F \leq b\} \subset\subset \Lambda \quad \text{for all } b \geq 0. \tag{1.3.3}$$

Hence, for small ϵ the potential will be the essential object, while the invariant measure degenerates in the limit.

The key result that we prove for reversible Markov chains ξ^ϵ on a uniformly locally finite graph establishes a connection between the dynamical behaviour of the chain and the geometry of its potential F. Similar versions have been shown e.g. in [BEGK01].

To do this we describe the geometry of F with the help of the following notions: A *path* γ is a finite sequence $(\gamma_1, \dots, \gamma_k)$ of *communicating points*, i.e. $p_\epsilon(\gamma_i, \gamma_{i+1}) > 0$ for all $1 \leq i \leq k-1$. We write $x \in \gamma$ when γ visits the point x. Let A and B be disjoint compact subsets of Λ_ϵ. We denote by $\mathcal{P}_{A,B}$ the set of paths starting in A and ending in B. We define the *communication height* between A and B to be

$$\hat{F}_\epsilon(A, B) := \min_{\gamma \in \mathcal{P}_{A,B}} \max_{x \in \gamma} F_\epsilon(x). \tag{1.3.4}$$

We denote the lower level set of $\hat{F}_\epsilon(A, B)$ by

$$W_\epsilon(A, B) := \left\{ x \in \Lambda_\epsilon \,|\, F_\epsilon(x) < \hat{F}_\epsilon(A, B) \right\}. \tag{1.3.5}$$

Assume $A \subset W_\epsilon(A, B)$. Then the connected component of $W_\epsilon(A, B)$ containing A is called the *valley of A with respect to B* and is denoted by $V_B^{(\epsilon)}(A)$.

Under a mild condition on the transition probabilities p_ϵ (see section II.4) we obtain then

Proposition 1.4 *Let (ξ^ϵ) be a family of ergodic and reversible Markov chains. Let A and B be disjoint compact sets of Λ_ϵ such that $F_\epsilon(x) < \hat{F}_\epsilon(A,B)$ for all $x \in A$.*

Then, under some regularity conditions there exist a constants $c_1, c_2 > 0$ such that

$$c_1 \epsilon^d \leq \frac{\mathrm{cap}\,(A,B)}{\exp\left(-\hat{F}_\epsilon(A,B)/\epsilon\right)} \leq c_2 \epsilon^{-d}. \tag{1.3.6}$$

This property shows already how the potential theoretic approach works: The capacity, which gives us information about the generator of the Markov chain ξ^ϵ, and therefore about the dynamics of our process, can be estimated by quantities reflecting the geometry of the potential.

For example given $m \in \mathscr{M}_\epsilon$ it always holds true that the valley of m with respect fo $\mathscr{M}_\epsilon$ is a subset of the domain of attraction $A(m, \mathscr{M}_\epsilon)$.

To further illustrate the usage of Proposition 1.4 let $\mathcal{M}$ be the set of all local minima of F, and assume $\mathcal{M}$ consists of finitely many points. Then Proposition 1.4 implies that there exist subsets $\mathcal{M}_\epsilon$ of Λ_ϵ with $|\mathcal{M}_\epsilon| = |\mathcal{M}|$ and such that ξ^ϵ is metastable with respect to $(\mathcal{M}_\epsilon)$. (See Example 4.10)

Observe that we are not assuming that the limiting function F is differentiable. Bovier and Faggionato used similar results to prove metastability in the sense of Definition 1.1 for Sinai's random walk in a random potential and gave precise estimates for the associated capacity [BF05].

Let us now assume that $\Lambda_\epsilon \equiv \Lambda \cap \epsilon\mathbb{Z}^d$. Under some more restrictive assumptions on the potentials F_ϵ, and assuming the limiting potential F is in $C^3(\Lambda)$, we provide matching upper and lower bounds of the capacity up to multiplicative errors that tend to one. To state the result precisely we define the *set of optimal paths* between two minima $m, n \in \mathscr{M}_\epsilon$ by

$$\mathcal{O}_{m,n} := \left\{ \gamma \in \mathcal{P}_{m,n} \mid \max_{x\in\gamma} F_\epsilon(x) = \hat{F}_\epsilon(m,n) \right\}. \tag{1.3.7}$$

For simplicity, we assume here that there is a unique point, $s \equiv s^*(m,n)$, that is visited by all paths of $\mathcal{O}_{m,n}$. This point is called the relevant saddle point between m and n. Our basic example for ξ^ϵ is the Metropolis sampler of the measure π_ϵ. In this case we have

$$p_\epsilon(x,y) = \begin{cases} \frac{1}{2d}\min\left(1, \frac{\pi_\epsilon(y)}{\pi_\epsilon(x)}\right) & \text{if } y \in \mathcal{N}_x, \\ 1 - \sum_{z\in\mathcal{N}_x} p_\epsilon(x,z) & \text{if } y = x, \\ 0 & \text{else.} \end{cases} \tag{1.3.8}$$

We then prove the following

Theorem 1.5 *Let* (ξ^ϵ) *be a family of ergodic Markov chains with reversible measure* π_ϵ*. Let* $m, n \in \mathcal{M}_\epsilon$*, and assume* $s \equiv s^*_\epsilon(m,n)$ *is the unique relevant saddle point between them. Then, under some regularity conditions,*

$$\begin{aligned} \operatorname{cap}(m,n) &= \frac{1}{2d}\left(\frac{2\pi}{\epsilon}\right)^{d/2-1} \frac{\lambda_d}{\sqrt{|\det \nabla^2 F_\epsilon(s)|}} \times \\ &\quad \times \exp\left(-F_\epsilon(s)/\epsilon\right)\left(1+\mathcal{O}\left(\sqrt{\epsilon}\,|\ln \epsilon|^{3/2}\right)\right), \end{aligned} \tag{1.3.9}$$

where $-\lambda_d$ *is the unique negative eigenvalue of the Hessian matrix* $\nabla^2 F_\epsilon(s)$ *at the relevant saddle point.*

The general strategy to prove this result is the same as in [BEGK04]: First, we will establish a direct connection between return probabilities and the capacity $\operatorname{cap}(A,B)$ between disjoint subsets A and B of Λ_ϵ (see Definition 3.6), namely

$$\operatorname{cap}(A,B) = \sum_{x\in B} \pi_\epsilon(x)\, P_x(\tau_A < \tau_x). \tag{1.3.10}$$

To obtain estimates for the capacity, we then use the Dirichlet principle. In the reversible setting, one can rewrite the Dirichlet form as a sum of positive terms, and this in turn yields a priori bounds on the capacity.

In a second step, we use a renewal equation for ξ^ϵ to obtain

$$h_{A,B}(x) \le \frac{\operatorname{cap}(x,A)}{\operatorname{cap}(x,B)}, \tag{1.3.11}$$

so that the a priori bounds for the capacity yield upper bounds for $h_{A,B}$ and $h_{B,A} = 1 - h_{A,B}$. The form of these bounds suggests, as we will see, that only a neighbourhood of the relevant saddle points (see Definition 4.3) needs to be investigated in detail. Just like in the continuous setting, a precise upper bound for the capacity can be achieved by choosing a function h^+ that is nearly optimal in a certain neighbourhood of the relevant saddles and inserting it in the Dirichlet form Φ. But the lower bound is more intricate. A special problem in the discrete setting is that the instable direction of a relevant saddle need not to be one of the lattice directions. To overcome this difficulty, we partition the lattice in a neighbourhood of a relevant saddle into parallel "strings", each string pointing in the right direction and having some microscopic structure. In particular, these strings are in general non-disjoint.

1.4 Expected hitting times

Expected hitting times are interesting quantities not only for themselves, but also because of their connection to the eigenvalues of the generator of ξ^ϵ, see e.g. [BEGK02]. We will discuss this point in part III in the context of the Hopfield model.

In [BEGK02] (Corollary 3.3, p.230) it has been shown that the expected hitting times of reversible Markov chains can be expressed by quantities we already know, namely

$$\mathbb{E}_x \tau_A = \frac{\pi_\epsilon (h_{x,A})}{\mathrm{cap}\,(x, A)}. \tag{1.4.1}$$

In the context of a *finite* state space it was also established, (see [BEGK02], Theorem 3.5, p. 231) that if (ξ^ϵ) is metastable with respect to $\mathscr{M}_\epsilon$, then for $m \in \mathscr{M}_\epsilon$

$$E_m \tau_{\mathscr{M}_\epsilon \setminus m} = \frac{\pi_\epsilon (A_\epsilon (m))}{\mathrm{cap}\,(m, \mathscr{M}_\epsilon \setminus m)} (1 + \mathcal{O} (\rho (\epsilon) |\Lambda_\epsilon|)). \tag{1.4.2}$$

Furthermore, according to their Corollary 3.4 (p. 230) one has

$$E_x (\tau_{\mathscr{M}_\epsilon}) \leq \frac{|\Lambda_\epsilon|}{a_\epsilon}, \tag{1.4.3}$$

where

$$a_\epsilon := \inf_{x \in E_\epsilon \setminus \mathscr{M}_\epsilon} P_x (\tau_{\mathscr{M}_\epsilon} < \tau_x). \tag{1.4.4}$$

By using formula (1.4.2) for the expected hitting time, one obains that a family of reversible Markov chains (ξ^ϵ) on a finite state space is ρ-metastable with respect to $\mathscr{M}_\epsilon$ iff

$$\inf_{m \in \mathscr{M}_\epsilon} E_m \left(\tau_{\mathscr{M}_\epsilon \setminus m}\right) \geq \frac{1}{\overline{\rho} (\epsilon)} \sup_{x \in \Lambda_\epsilon \setminus \mathscr{M}_\epsilon} E_x (\tau_{\mathscr{M}_\epsilon}), \tag{1.4.5}$$

where

$$\overline{\rho} (\epsilon) = \rho (\epsilon) |\Lambda_\epsilon| \frac{\pi_\epsilon (m)}{\pi_\epsilon (A_\epsilon (m))}. \tag{1.4.6}$$

Observe that (1.4.5) is useless for countable state spaces Λ_ϵ, since then $\overline{\rho} (\epsilon) = \infty$. This can not be repaired easily, because

$$\sup_{x \in \Lambda_\epsilon \setminus \mathscr{M}_\epsilon} E_x (\tau_{\mathscr{M}_\epsilon}) \tag{1.4.7}$$

can also be infinity in this case.

The main theorem of this part is the Eyring-Kramers formula, which we state here only for the Metropolis algorithm and in the case of a unique relevant saddle point, for simplicity.

Theorem 1.6 *(Eyring-Kramers formula) Let $m \in \mathcal{M}_\epsilon$ and assume s is the unique relevant saddle point between m and $\mathcal{M}_\epsilon \setminus m$. Denote by $-\hat{\lambda}_d$ the unique negative eigenvalue of $p(s) \cdot \nabla^2 F_\epsilon(s)$. Then for ξ^ϵ starting in m, the expected time needed to reach another point of $\mathcal{M}_\epsilon$ is given by*

$$\begin{aligned} \mathbb{E}_m\left(\tau_{\mathcal{M}_\epsilon \setminus m}\right) &= \frac{2\pi}{\epsilon}\frac{1}{\hat{\lambda}_d}\sqrt{\frac{\left|\det \nabla^2 F_\epsilon(s)\right|}{\left|\det \nabla^2 F_\epsilon(m)\right|}} \exp\left\{\left(F_\epsilon(s) - F_\epsilon(m)\right)/\epsilon\right\} \times \\ &\quad \times \left(1 + \mathcal{O}\left(\sqrt{\epsilon}\left|\ln \epsilon\right|^{3/2}\right)\right). \end{aligned} \tag{1.4.8}$$

We are left with the

Open Question How could these expected hitting times be given precise estimates in the case of a non-reversible Markov chain?

1.5 The pathwise approach

In the recent treatise "Large Deviations and Metastability" by Enzo Olivieri and Maria Eulalia Vares, [OV05], metastability is discussed in great detail from the point of view of a pathwise approach. Let us transfer in our setting the two asymptotic properties of Metastability which are emphasised in this book.

A point $m \in \Lambda_\epsilon$ is called *metastable in the sense of Olivieri and Vares* iff the following two properties hold:

1. *Unpredictability of the tunneling time.*
   Assume ξ^ϵ starts in m. Then $\tau_{\mathcal{M}_\epsilon \setminus m}$ is called *unpredictable* if it converges in distribution to an exponential random variable, i.e.

   $$\frac{\tau_{\mathcal{M}_\epsilon \setminus m}}{E_m\left(\tau_{\mathcal{M}_\epsilon \setminus m}\right)} \xrightarrow{\mathcal{D}} \mathcal{E} \quad \text{for } \epsilon \downarrow 0, \tag{1.5.1}$$

   where $\mathcal{E}$ is a unit mean exponential random variable.

2. *Thermalisation.*
   Let $s, t \in \mathbb{N}_0$. We define the empirical average measure of ξ^ϵ between the times s and $s+t$ as

$$\mu_{s,t} := \frac{1}{t} \sum_{k=s+1}^{s+t} \delta_{\xi_k^\epsilon}. \tag{1.5.2}$$

   Hence $\mu_{s,t}(B)$ is the fraction of time ξ^ϵ spends in $B \subset \Lambda_\epsilon$ between time s and $s+t$. Let $V := V^{(\epsilon)}_{\mathscr{M}_\epsilon \setminus m}(m)$ be the valley of m with respect to $\mathscr{M}_\epsilon \setminus m$.
   Let ξ^ϵ again start at m. We say ξ^ϵ *thermalises at* m if there exists a deterministic time scale t_ϵ such that $\lim_{\epsilon \downarrow 0} t_\epsilon = \infty$, but $t_\epsilon = o\left(E_m\left(\tau_{\partial^+ V}\right)\right)$ and for every open set $B \subset \mathbb{R}^d$ containing m and every $\kappa > 0$

$$\lim_{\epsilon \downarrow 0} P_m \left(\tau_{\partial^+ V} > t_\epsilon \text{ and } \sup_{s < \tau_{\partial^+ V} - t_\epsilon} \mu_{s,t_\epsilon}(B) > 1 - \kappa \right) = 1. \tag{1.5.3}$$

Bovier, Eckhoff, Gayrard and Klein showed indeed that the unpredictability can be seen as a consequence of Definition 1.1, see [BEGK02], Theorem 1.3 (iv), p. 223.

Here we show:

Theorem 1.7 *Let (ξ^ϵ) be a family of ergodic and reversible Markov chains. Let $\mathcal{M}_\epsilon$ be the set of local minima of F_ϵ. Assume that $\mathcal{M}_\epsilon$ is a finite set and $|\mathcal{M}_\epsilon|$ is independent of ϵ. Choose $m \in \mathcal{M}_\epsilon$ and let $V := V^{(\epsilon)}_{\mathcal{M}_\epsilon \setminus m}(m)$ be the valley of m with respect to $\mathcal{M}_\epsilon \setminus m$. Assume ξ^ϵ starts at m, then it thermalises at m.*

It should be possible to show thermalisation for more general sets. Therefore we have

Open Question Show the thermalisation property for a general metastable set in the sense of Definition 1.1.

At least up to now, Definition 1.1 can describe more general situations. It focuses not only on a single metastable state and the ensuing transition to equilibrium, but describes a consistent set of metastable points. Moreover, it seems to be easier to check the criterion of Definition 1.1 than the thermalisation property mentioned in the pathwise approach.

2 Metastability in the Hopfield model

2.1 The Hopfield model

In the second part of this work, we apply the general results on metastability of countable Markov chains to investigate the metastable behaviour of the Hopfield model.

A famous interpretation of the Hopfield model is to view it as a model for a neural network. Basically we mean by a neural network model a labeled and possibly oriented graph $\Gamma = (\Lambda, \mathcal{E})$ together with a set $\mathcal{S}$ with at least two elements; Λ is the set of neurons and $\mathcal{E}$ the set of synapses connecting these neurons. The activity of each of the neurons is described by a variable σ_i taking its values in $\mathcal{S}$, for all $i \in \Lambda$). We will model the dynamics of this network by a Markov chain $\sigma_G = \{\sigma_G(t)\}_{t \in \mathbb{N}_0}$ on $\mathcal{S}^\Lambda$.

One of the most important advances due to Hopfield ([Hop82]) has been to understand that these dynamics correspond to a Hamiltonian H_N. Assume the information to be stored is encoded in so-called patterns ξ^μ, $\mu \in \{1, \ldots, M(N)\}$, each of the ξ^μ itself being a sequence of $\xi_i^\mu \in \mathcal{S}$ for $i \in \{1, \ldots, N\}$. To make the neural net capable of adapting to different sequences of patterns, we have to introduce a set of variables J_{ij} for all $\{i, j\} \in \mathcal{E}$ called the synaptic efficacy and describing the strength of interaction between the neurons at sites i and j. It is commonly assumed that the variable J_{ij} is measurable with respect to the set $\left\{\xi_i^\mu, \xi_j^\mu \,|\, \mu \in \{1, \ldots, M(N)\}\right\}$. This is then called locality of the weights J_{ij}. The associated Hamiltonian H_N is given by

$$H_N(\sigma) := -\frac{1}{2} \sum_{\{i,j\} \in E} J_{ij} \sigma_i \sigma_j. \tag{2.1.1}$$

The Hopfield model ([Hop82]) is among the most classical and best understood models of neural network. Although originally introduced by Pastur and Figotin, [FP77], as a simplified model of a spin glass, this model earned much of its success through its reinterpretation as an auto-associative memory by Hopfield and may therefore by right be called the Hopfield model. Here the graph G is the complete graph K_N on the vertex set $\Lambda = \{1, \ldots, N\}$ and $\mathcal{S} = \{-1, +1\}$ corresponds to a neuron being switched either 'on' or 'off', and the weights J_{ij} are given by 'Hebb's learning rule', i.e. by the formula

$$J_{ij} := \frac{1}{N} \sum_{\mu=1}^{M(N)} \xi_i^\mu \xi_j^\mu. \tag{2.1.2}$$

Note that (2.1.1) may be rewritten in the convenient form

$$H_N [\xi] (\sigma) = -\frac{1}{2N} \sum_{\mu=1}^{M(N)} \langle \xi^\mu, \sigma \rangle^2 . \tag{2.1.3}$$

The scalar product $\frac{1}{N} \langle \xi^\mu, \sigma \rangle$ is the so called *overlap between ξ^μ and σ*. Note that this scalar product may be regarded as an index for how similar σ is to either ξ^μ or $-\xi^\mu$, because its absolute value can be written as

$$\frac{1}{N} |\langle \xi^\mu, \xi^\nu \rangle| = 1 - 2 \min \{ d_H (\xi^\mu, \xi^\nu), d_H (\xi^\mu, -\xi^\nu) \} , \tag{2.1.4}$$

where d_H is the normalised Hamming distance, namely

$$d_H (\sigma, \tau) = \frac{1}{N} \sum_{i=1}^{N} 1 (\sigma_i \neq \tau_i) . \tag{2.1.5}$$

At this point, one may notice the spin-flip symmetry

$$H_N (-\sigma) = H_N (\sigma) , \tag{2.1.6}$$

showing that the Hopfield model can not distinguish between a spin configuration and its negative.

Observe also that (2.1.3) makes it plausible that - at least for $M(N)$ small enough - the minima of H_N are located close to the patterns ξ^μ. (Actually this is trivially fulfilled if the patterns are orthogonal, i.e. if $\langle \xi^\mu, \xi^\nu \rangle = \delta_{\mu\nu}$).

Let $\beta \in \mathbb{R}_{\geq 0}$ be a non negative parameter; in the context of statistical mechanics it plays the role of an inverse temperature. The Hamiltonian H_N determines a finite volume Gibbs measure $\pi_N \equiv \pi_{N,\beta} [\xi]$ given by

$$\pi_N (\sigma) := \frac{1}{Z_{N,\beta}} \exp (-\beta H_N (\sigma)) . \tag{2.1.7}$$

Here, $Z_{N,\beta} \equiv \sum_{\sigma \in \mathcal{S}^N} \exp (-\beta H_N (\sigma))$, the *partition function*, is a normalising factor assuring that π_N is a probability measure.

From now on we will refer to the Hopfield model as a Markov chain $\sigma_{N,\beta} \equiv (\sigma_{N,\beta}(t))_{t\in\mathbb{N}_0}$ on the configuration space $\mathcal{S}^N$ that is reversible with respect to the Gibbs measure π_N. We consider *Glauber dynamics*, so that during each time step at most one spin is flipped.

Now we choose the components of the patterns, ξ_i^μ, uniformly at random in $\{-1,1\}$ and independently of each other. Of course, $\sigma_{N,\beta}$ is then a Markov chain with random rates. The dependence on

$$\xi \equiv (\xi^\mu)_{1\le\mu\le M} \tag{2.1.8}$$

will be indicated explicitly whenever we want to stress it. Otherwise, we will frequently drop it to simplify the notation.

There exists a threshold value for the number of patterns such that the memory works for low temperatures i.e. $\beta > 1$. The critical dependence is $M(N) \sim \alpha N$ with $\alpha \approx 0,138$. (see e.g. [AGS85],[AGS87],[BG94]).

We assume $M(N) \equiv M$ to be a finite number, independent of N, and therefore we are in the regime of perfect memory. We will analyse the long time behaviour of $\sigma_{\beta,N}$. As we will see this can be described in the general framework of metastability.

The following two papers have dealt with several aspects of the problem.

- V.A. Malyshev, F.M. Spieksma "Dynamics in Binary Neural Networks with a Finite Number of Patterns" ([MS97]) treats the case of zero temperature, i.e. $\beta = \infty$. In this setting the phenomenon of metastability does not occur. If the process reaches one of the local minima of the effective energy $\bar{H}$, it stays there forever. The stochastic behaviour they investigate is localised at the boundaries of the domains of attraction of different minima.

- In G. Biroli and R. Monasson, "Relationship between Long Time Scales and the Static Free-Energy in the Hopfield Model", ([BM98]), contrary to the announcement in the title, the authors do not really investigate the long time behaviour of the Hopfield model. They only show that the Hopfield model behaves in the neighbourhood of a critical point like a quantum mechanical harmonic oscillator, i.e. that the effective energy can be approximated by a quadratic function near the critical points.

We use the symmetry of the model to reduce the dimension of the state space from N to $d \equiv 2^M$. This is done by a transformation invented by

Grensing and Kühn [GK86] that lumps together certain groups of spins. The randomness of the pattern ξ^μ is then encoded in the size $\ell_k \equiv \frac{N}{d}(1+\frac{1}{\sqrt{N}}\lambda_k)$ of these groups. In the following we restrict ourselves to the set of patterns Ξ such that $\lambda_k = \mathcal{O}[\ln N]$ for all k. Observe that due to the law of the iterated logarithm Ξ has asymptotically full measure. The transformed process is a Markov chain, $\zeta_{N,\beta}$, on the (random) d-dimensional lattice $\mathcal{L}_{N,\beta} = \times_{k=1}^{d} \frac{2}{\ell_k}\mathbb{Z}$ intersected with the hypercube $[-1,1]^d$. $\zeta_{N,\beta}$ is again reversible with respect to a Gibbs measure $\varrho_{N,\beta}$, which is characterised by a modified Hamiltonian $\mathcal{H}_{N,\beta}$.

We can think of $\zeta_{\beta,N}$ as a process exploring a landscape given by the random function $f_{\beta,\lambda}$ that equals up to a constant $\frac{1}{N}\mathcal{H}_{N,\beta}$. The ground states corresponds in this picture to the global minima of $f_{\beta,\lambda}$.

Let $\{b_1,\dots,b_d\}$ be an enumeration of all vectors in $\{-1,1\}^M$. Hence $b^\mu \in \{-1,1\}^d$, and we denote $b^{-\mu} := -b^\mu$. Moreover, we introduce the graph $G=(V,E)$, where

$$V := \{-M,\dots,M\} \setminus \{0\}$$

and

$$E := \{\{\mu,\nu\} \in V \times V \,|\, \mu \notin \{-\nu,\nu\}\}\,.$$

Let m^* denote the unique positive solution of the 'mean field equation'

$$m = \tanh(\beta m)\,. \tag{2.1.9}$$

Similarly to e.g. Genz, ([Gen96]), we show that for all $\beta > 1$ the global minima of $f_{\beta,\lambda}$ have positions that are small random perturbations of the points

$$\overline{m}_\mu = m^* b^\mu \quad \text{for } \mu \in V. \tag{2.1.10}$$

Therefore the set of global minima of $f_{\beta,\lambda}$ can be written as

$$\mathcal{M}_N := (m_\mu \,|\, \mu \in V)\,. \tag{2.1.11}$$

The minimum $m_{\pm\mu}$ corresponds to a spin configuration near the μ-th pattern or its negative, $-\xi^\mu$.

We wil show that the Hopfield model exhibits metastable behaviour. Therefore, as we saw in part II the long time evolution of $\zeta_{N,\beta}$ is controlled by the position and height of the so called *relevant saddle points* between the minima. To determine them, we have to be very careful. In a quite small interval of the temperature, namely $1 < \beta < 1 + (9d + 500M^8)^{-1}$, Koch and Piasko showed that the so called "symmetric solutions" provide the only critical points of the deterministic function $f_\beta \equiv f_{\beta,0}^{(N)}$.

Since all these critical points are non degenerate ($\det \nabla^2 f_\beta(s) \neq 0$), the only candidates for relevant saddle points are the critical points of f_β with a Hessian matrix with one negative and $(d-1)$ positive eigenvalues (1-saddles). We show that $f_{\beta,\lambda}$ has a unique critical point in a small neighbourhood of each 1-saddle of the symmetric solutions

$$\overline{s}_{\mu,\nu} = \frac{1}{2} m^* \left(b^\mu + b^\nu\right) \quad \text{ for } \{\mu, \nu\} \in E, \tag{2.1.12}$$

and these are the only 1-saddles. Hence the set of 1-saddles can be represented by

$$\left(s_{\mu\nu} \mid \{\mu, \nu\} \in E\right). \tag{2.1.13}$$

These points are 1-saddles for all $1 < \beta < \beta_c$, where $\beta_c \approx 1.7$ is the unique positive solution of $\beta = \frac{2}{2-m^*(\beta)^2}$. This leads to the following

Open Question (a) Is it true that for all $1 < \beta < \beta_c$ the relevant saddles between the global minima of $f_{\beta,\lambda}$, namely between the elements in $\mathcal{M}_N$, are contained in

$$\left(s_{\mu\nu} \mid \{\mu, \nu\} \in E\right)? \tag{2.1.14}$$

(b) What are the relevant saddles between these global minima for $\beta \geq \beta_c$?

In contrast to the heights of the lowest minima of $f_{\beta,\lambda}$ the heights of the 1-saddles perform random fluctuations with an amplitude of order $1/\sqrt{N}$. To give the precise form of these fluctuations, we denote the free energy of the Curie-Weiss model by

$$f_{CW}(\beta) := \frac{1}{2} m^{*2} - \frac{1}{\beta} I(m^*). \tag{2.1.15}$$

We introduce the symmetric matrix A_N given by

$$A_N^{\mu\nu} := \frac{1}{\sqrt{N}} \langle \xi^\mu, \xi^\nu \rangle \quad \text{for all } \mu \neq \nu \tag{2.1.16}$$

and $A_N^{\mu\mu} := 0$. As Külske pointed out ([Kül97]), the matrix A_N has asymptotically standard normal entries outside the diagonal.

Proposition 2.1 *For all $\xi \in \Xi$ and $N \geq N_0[\xi]$, we obtain*

$$\begin{aligned} & f_{\beta,\lambda}(m_{\pm\mu}) \\ & = -f_{CW}(\beta) + \frac{k_0}{N}\left(A_N^2\right)^{\mu,\mu} + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right)^3 \end{aligned} \tag{2.1.17}$$

and

$$\begin{aligned} & f_{\beta,\lambda}(s_{\mu,\pm\nu}) \\ & = -\frac{1}{2}f_{CW}(\beta) \mp \frac{k_1}{\sqrt{N}}A_N^{\mu,\nu} + \\ & \quad -\frac{k_2}{N}\sum_{\alpha}\left(A_N^{\alpha\mu} \pm A_N^{\alpha\nu}\right)^2 - \frac{k_3}{N}\left(A_N^{\mu\nu}\right)^2 + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right)^3. \end{aligned} \tag{2.1.18}$$

The constants can be given explicitly in terms of β, M and m^*.

We now have all the ingredients enabling us to apply the Eyring-Kramers formula proved in part 1 in order to give a precise estimate for the expected time needed by $\zeta_{\beta,N}$ to change over from one ground state to another one. In the context of a neural network we can say we are associating another pattern to the one we remembered first. Despite the mean field nature of the Hopfield model and the i.i.d. choice of the patterns, this will be for all $\xi \in \Xi$ and $N \geq N_0[\xi]$ uniquely determined.

We state our result for the (random) Markov chain $\zeta_{\beta,N}[\xi]$ on the compact state space $\mathcal{X}_N[\xi] \equiv [-1,1]^d \cap \mathcal{L}_N[\xi]$.

We assume that the values $(A_n^{\mu\nu})_{1\leq\mu<\nu\leq M}$ are all sufficiently different. Therefore we define

$$J_{N,\delta}[\xi] := \left\{ n \leq N \mid \min_{a,b\in\overline{E}}\left(A_n^a - A_n^b\right) < n^{-\frac{1}{2}+\delta} \right\}. \tag{2.1.19}$$

We can show that this set has cardinality

$$|J_{N,\delta}[\xi]| = o(N) \qquad \text{for all } \xi \in Z'_\delta. \tag{2.1.20}$$

We denote now

$$J_\delta[\xi] := \left\{ n \in \mathbb{N} \mid \min_{a,b\in\overline{E}}\left(A_n^a - A_n^b\right) < n^{-\frac{1}{2}+\delta} \right\}. \tag{2.1.21}$$

For simplicity we assume that the original Markov chain $\sigma_{N,\beta}$ is the (Glauber) Metropolis algorithm for π_N.

Theorem 2.2 *We assume $1 < \beta < 1 + (9d + 500M^8)^{-1}$. Choose $\delta \in (0, \frac{1}{2})$ and assume $\xi \in Z'_\delta$ and $N \geq N_0[\xi]$, as well as $N \in J_\delta$. Let I and J be disjoint subsets of $\mathcal{M}_N$.*

If $s \in S_N(I, J)$ is a relevant saddle point between I and J we obtain

$$\begin{aligned} \operatorname{cap}(I, J) &= k_4 |S_N(I, J)| N^{(d-2)/2} \varrho_{N,\beta}(s) \times \\ &\quad \times \left(1 + \mathcal{O}\left(\sqrt{\ln^3 N}/\sqrt{N}\right)\right). \end{aligned} \tag{2.1.22}$$

Starting in $m \in \mathcal{M}_N \backslash I$ the expected (quenched) hitting time of J satisfies

$$\begin{aligned} \mathrm{E}_m(\tau_I) &= \frac{k_5 N}{|S_N(m, J)|} \sum_{n \in V_J(m)} \exp(N b_N(n, J)) \times \\ &\quad \times \left(1 + \mathcal{O}\left(\sqrt{\ln^3 N}/\sqrt{N}\right)\right), \end{aligned} \tag{2.1.23}$$

where

$$b_N(n, J) := \beta\left(\hat{f}_{\beta,\lambda}(n, I) - f_{\beta,\lambda}(n)\right) \tag{2.1.24}$$

is the barrier between n and J. The constants k_4 and k_5 can be given explicitly.

Of course, the assumption of independence of the pattern components is only one possible choice. Indeed, there are at least two sensible ways of introducing correlations among the patterns. One is to consider spatial correlation, i.e. to choose the patterns correlated in i but independent in μ, which may be interesting when e.g. thinking about the patterns as images to be stored. The other way is to choose sequentially or semantically correlated patterns, which means that the dependency now enters via μ only. This situation might be useful as a very simple model for patterns with some sort of causal relations, as in the storage of films for example. The dependence can be modelled e.g. via a Markov chain, i.e. in the case of spatial correlation, ξ^μ_{i+1} taking with probability $p \in (0, 1)$ the same value as ξ^μ_i and with probability $(1 - p)$ the value of $-\xi^\mu_i$. (See for example Löwe [Löw98]). This leads to

Open Question Is it possible to compute the (Newtonian) capacity and the expected hitting times separating ground states in a Hopfield model with spatially or semantically correlated patterns?

2.2 Structure of the ground states

We can represent the structure of $f_{\beta,\lambda}$ (given by the minima and the 1-saddles) through a weighted graph (V, E, w). The weights are given by

$$w_{\mu,\nu} = \exp\left(-k_1 \langle \xi^\mu, \xi^\nu \rangle\right) \quad \text{for all } \{\mu, \nu\} \in E. \tag{2.2.1}$$

Due to the spin flip symmetry, we obtain the same weights between the negatives, $-m_\mu$, i.e. $w_{-\mu,-\nu} = w_{\mu\nu}$. The cross weights are given by $w_{-\mu,\nu} = w_{\mu,-\nu} = 1/w_{\mu\nu}$ for $\mu \neq \nu$. There is no connection between m_μ and $-m_\mu$, i.e. $w_{\mu,-\mu} = \infty$.

The form of these weights implies that it is much easier to use several edges (μ, ν) with smaller values of $A_N^{\mu\nu}$ than one with a larger matrix entry.

Consider the simplified weighted graph $\left(\overline{V}, \overline{E}, \overline{w}\right)$, where we identify μ and its negative $-\mu$, i.e. the set of vertices is $\overline{V} := \{1, \dots, M\}$, the edge set is $\overline{E} = \left\{\{\mu, \nu\} \in \overline{V} \times \overline{V} \mid \mu \neq \nu\right\}$, and define

$$\overline{w}_{\mu\nu} := \min\left(w_{\mu\nu}, w_{\mu,-\nu}\right) \quad \text{for } \{\mu, \nu\} \in \overline{E}. \tag{2.2.2}$$

Hence we obtain

$$\overline{w}_{\mu\nu} = \exp\left(-k_1 \left|\langle \xi^\mu, \xi^\nu \rangle\right|\right). \tag{2.2.3}$$

This graph induces a tree structure appearing in the following way: We arrange the edges linearly as $(s_1, s_2, \dots)$ in such a way that

$$\overline{w}_{s_i} < \overline{w}_{s_{i+1}} \quad \text{for all } 1 \leq i \leq \binom{M}{2}. \tag{2.2.4}$$

Now we start with M single vertices and then merge together classes of vertices according to this order until all vertices are in one class. Using the representation (2.1.4), we see that the distance of two leaves in this tree, say m_μ and m_ν, is determined by the minimal number of spins one has to change in ξ^μ in order to reach either ξ^ν or $-\xi^\nu$.

Since every connected graph includes all edges corresponding to essential saddles, these are included in particular in the edgeset of the minimal connected graph. We then apply a theorem of Erdös and Rényi from the theory of random graphs to get the desired estimate.

Theorem 2.3 *Let $\xi \in \Xi$ and $N \geq N_0[\xi]$, and assume $1 < \beta < 1 + \left(9d + 500M^8\right)^{-1}$. Then asymptotically almost surely (for $M \to \infty$), the communication height between two disjoint subsets of $\mathcal{M}_N$, say I and J, can be* *estimated by*

$$\hat{f}_{\beta,\lambda}(I,J) \quad \leq \quad \frac{1}{2} f_{CW}(\beta) - \frac{k_1}{\sqrt{N}}\sqrt{2\ln M}. \tag{2.2.5}$$

We now want to determine the low lying eigenvalues of the Hopfield model. Let $\{\lambda_0, \ldots, \lambda_{2M-1}\}$ with $0 = \lambda_0 \leq \ldots \leq \lambda_{2M-1}$ be the smallest eigenvalues of the generator $-L_{N,\beta}[\xi]$ of the transformed Markov chain $\zeta_{N,\beta}[\xi]$.

Due to the symmetry under total spin flip and the unusually small fluctuations of the heights of the minima in $\mathcal{M}_N$, we cannot directly use the results of Bovier, Eckhoff, Gayrard and Klein in [BEGK02], but we can apply similar methods.

The weighted graph structure (V, E, w) governs the form of the small eigenvalues of the generator $L \equiv L_{N,\beta}[\xi]$. Let $\mathcal{T} = (t_1, \ldots, t_{2M-1})$ be a minimal spanning tree of (V, E, w) such that

$$w^{t_{2M-1}} \leq w^{t_{2M-2}} \leq \ldots \leq w^{t_1} < 0.$$

Notice that (up to the order and sometimes choice of equally weighted edges) Kruskal's algorithm to construct a minimal spanning tree starts with t_{2M-1} and adds along our enumeration edges to the spanning tree until it ends with t_1. Let $I_{\mathcal{T}} \subset \{1, \ldots, 2M-1\}$ denote the set of indices such that $w^{t_i} < w^{t_{i-1}}$.

Using the exception set J_δ defined by equation (2.1.21) we obtain

Theorem 2.4 *Let $\xi \in Z'_\delta$ and $N \geq N_0[\xi]$. There exists an increasing sequence $(\mathscr{M}_i \,|\, i \in I_{\mathcal{T}})$ of metastable sets of $\zeta_{N,\beta}$. We define*

$$E_i^* = \arg \min_{\{m,n\} \in \mathscr{M}_i \times \mathscr{M}_i} \left(\hat{f}_{\beta,\lambda}(m,n) \right). \tag{2.2.6}$$

Denote for all $m \in \mathcal{M}_i$

$$\gamma_{m,i} = E_m \left(\tau_{\mathscr{M}_i \setminus m} \right)^{-1}. \tag{2.2.7}$$

We distinguish three cases:

- *Assume $E_i^* = \{\{m,n\}, \{-m,-n\}\}$, then*

$$\lambda_{i-1} = \lambda_i = (\gamma_{m,i} + \gamma_{n,i}) \left(1 + \mathcal{O}\left(e^{-\delta N}\right)\right). \tag{2.2.8}$$

- *Assume* $E_i^* = \{\{m,n\},\{-m,n\}\}$, *then*

$$\lambda_i = (2\gamma_{m,i} + \gamma_{n,i}) \left(1 + \mathcal{O}\left(e^{-\delta N}\right)\right) \tag{2.2.9}$$

  and

$$\lambda_{i-1} = \gamma_{m,i} \left(1 + \mathcal{O}\left(e^{-\delta N}\right)\right). \tag{2.2.10}$$

- *Assume* $E_i^* = \{m,n\}$, *then*

$$\lambda_i = (\gamma_{m,i} + \gamma_{n,i}) \left(1 + \mathcal{O}\left(e^{-\delta N}\right)\right). \tag{2.2.11}$$

Together with Theorem 2.2 this yields explicit estimates for the low lying spectrum of the generator of $\zeta_{N,\beta}$.

Part II

Metastability of Markov Chains

3 Equilibrium potential and capacity

This section describes the potential theoretic approach to metastability developed by Bovier, Eckhoff, Gayrard and Klein. A review of this can also be found in [Bov04]. We use this here mostly to introduce the notation. The potential theoretic approach works for ergodic Markov processes on connected locally finite graphs. The most results require a reversible process.

A *graph*, Γ, consists of a countable discrete set, Y, that has no cluster points and a nonempty set, $G \subset Y \times Y$, of ordered pairs of points, such that $(x, y) \in G$ if and only if $(y, x) \in G$. Without restriction of generality we assume that all self edges (x, x) are in G. We say y is in the *neighbourhood of* x, i.e. $y \in \mathcal{N}_x$ if $(x, y) \in G$ and $x \neq y$. The family $\mathcal{N} = \{\mathcal{N}_x\}_{x \in Y}$ is called the *neighbourhood system* of Γ associated to G. We say Γ is *locally finite*, if the number of neighbours of each point is finite, i.e. $|\mathcal{N}_x| < \infty$ for all $x \in Y$.

For $A \subset Y$ we define the *external boundary* to be

$$\partial^+ A := \left(\bigcup_{s \in A} \mathcal{N}_s \right) \setminus A \tag{3.0.1}$$

and the *internal boundary* to be

$$\partial^- A := \partial^+ (A^c) . \tag{3.0.2}$$

Here A^c denotes the complement of A. Define moreover the *thickened set*

$$A^+ := A \cup \partial^+ A. \tag{3.0.3}$$

Let $\Gamma = (Y, G)$ be a locally finite connected graph and $\zeta \equiv (\zeta_t)_{t \in \mathbb{T}}$ a homogeneous Markov process on Γ with time set $\mathbb{T}$. We consider the cases of continuous time set, i.e. $\mathbb{T} = \mathbb{R}_{\geq 0}$ and of discrete time, i.e. $\mathbb{T} = \mathbb{N}_0$. In the discrete time case we call ζ a *Markov chain.* Here, ζ is characterised by the starting distribution and the transition probability, p. By ζ being a Markov process on Γ we mean that $p(x, y) > 0$ if and only if $(x, y) \in G$. For continuous time a Markov process on Γ has the property $(\xi_t, \xi_{t-}) \in G$ for all $t \in \mathbb{R}_{\geq 0}$.

We assume that ζ is ergodic. Hence the whole space Y is a positive recurrent class of ζ and there exists a unique invariant probability measure π. For $x \in Y$ we denote by P_x the law of ζ with starting point x and by E_x the associated expectation. Since some statements of this section do not use reversibility, we indicate the places where it enters.

Definition 3.1 We call a homogeneous Markov process, ζ, with continuous time set *regular* iff it is stable, conservative and nonexplosive, i.e. its infinitesimal generator, L, is of the form

$$Lf(x) = \sum_{y \in \mathcal{N}_x} L(x,y)(f(y) - f(x)) \tag{3.0.4}$$

with non negative finite *rates* $(L(x,y))_{x \neq y}$ and ζ has a.s. only finitely many jumps in a finite interval of time. The waiting time of ζ at a point $x \in Y$ is an exponential distributed random variable with parameter

$$r(x) := \sum_{y \in \mathcal{N}_x} L(x,y). \tag{3.0.5}$$

Remark 3.2 The Criterion of Reuter says that a stable and conservative generator L is nonexplosive iff it admits no non-negative bounded eigenvectors with positive eigenvalue (see [Bré99], Theorem 4.4, p. 351).

The *embedded Markov chain* forgets about the waiting times of ζ and notices only the jumps while taking the number of jumps as time. We define

Definition 3.3 Let ζ be a regular Markov process with continuous time parameter and generator L. We denote $r(x) := -L(x,x)$. The *embedded Markov chain* is defined to have the same starting distribution and a transition matrix, p defined by

$$p(x,y) := \frac{L(x,y)}{r(x)} \quad \text{for } y \in \mathcal{N}_x \tag{3.0.6}$$

and zero otherwise.

Therefore the generator of the embedded Markov chain, $L^{(d)}$, has the form

$$L^{(d)}(x,y) = \frac{1}{r(x)} L(x,y). \tag{3.0.7}$$

For the associated invariant probability measure, $\pi^{(d)}$ we obtain

$$\pi^{(d)}(x) = \frac{r(x)\,\pi(x)}{\sum_{z \in Y} r(z)\,\pi(z)}. \tag{3.0.8}$$

3.1 The equilibrium potential

We only consider the discrete time setting, i.e. $\mathbb{T} = \mathbb{N}_0$ and will use the embedded Markov chain in the case of continuous time (see Definition 3.3).

Given two disjoint subsets A and B of Y and $x \in Y$, what can we say about the properties of

$$P_x (\tau_A < \tau_B)? \tag{3.1.1}$$

To answer this question we use

Proposition 3.4 *Let $\Gamma = (Y, G)$ be a locally finite graph, $D \subsetneq Y$ a (nonempty) connected set and L the generator of an ergodic Markov chain ζ on Γ. Suppose $f : \partial^+ D \to \mathbb{R}$ and $g : D \to \mathbb{R}$ are bounded functions. If h is a bounded solution of the Dirichlet-Poisson problem*

$$\begin{cases} -Lh &= g \quad \text{on } D, \\ \quad h &= f \quad \text{on } \partial^+ D, \end{cases} \tag{3.1.2}$$

then $\tau := \tau_{\partial^+ D}$ is P-a.s. finite and

$$h(x) = E_x (f(\xi_\tau)) + E_x \left[\sum_{n=0}^{\tau-1} g(\xi_n) \right] \tag{3.1.3}$$

for all $x \in D$.

Proof. Due to the ergodicity of ζ we have for every $x \in Y$ that $\mathbb{E}_x \tau_x = \frac{1}{\pi(x)} < \infty$. Due to the irreducibility also τ is almost surely finite independent of the starting point $x \in D$. Now we can apply Theorem 2.1, p. 181 in Brémaud [Bré99]. □

Now we look, more specifically, at

Definition 3.5 Let L be the generator of the ergodic Markov chain ζ. The *equilibrium potential* $h_{A,B} : Y \to [0,1]$ of ζ is defined to be the unique bounded solution of the boundary value problem

$$\begin{cases} Lh = 0 & \text{on } (A \cup B)^c, \\ h = 1 & \text{on } A, \\ h = 0 & \text{on } B. \end{cases} \tag{3.1.4}$$

We also say that $h_{A,B}$ is *harmonic* on $Y \setminus (A \cup B)$ with respect to L.

Then Proposition 3.4 tells us that

$$h_{A,B}(x) = P_x(\tau_A < \tau_B) \qquad \text{for all } x \in Y \setminus (A \cup B). \tag{3.1.5}$$

To treat the case when the starting point of ζ lies inside $A \cup B$, we use the following reasoning to find an equation for $P_x(\tau_A < \tau_B)$. The first step of ζ leads either to B, and the event $\{\tau_A < \tau_B\}$ fails to happen, or to A, in which case the event happens, or to another point $y \notin A \cup B$, in which case the event happens with probability $P_y(\tau_A < \tau_B)$. Thus for all $x \in Y$

$$\begin{aligned} P_x[\tau_A < \tau_B] &= \sum_{y \in A} p(x,y) + \sum_{y \notin A \cup B} p(x,y) P_y[\tau_A < \tau_B]. \\ &= p h_{A,B}(x) = L h_{A,B}(x) + 1_A(x), \end{aligned} \tag{3.1.6}$$

since $h_{A,B}|_{A \cup B} = 1_A$.

In the case of continuous time we use the embedded Markov chain, that has a transition probability matrix given in Definition 3.3. Therefore we obtain

$$P_x(\tau_A < \tau_B) = p h_{A,B}(x) = \frac{1}{r(x)} L h_{A,B}(x) + 1_A(x). \tag{3.1.7}$$

This result suggests to introduce the following notion that originates from the theory of electromagnetism.

Definition 3.6 Let A and B be disjoint subsets of Y and L the generator of a Markov process ζ on Y. We call $e_{A,B} := L h_{A,B}$ the *equilibrium measure* for the *capacitor* A, B.

Now we can answer the question of the beginning of this section. Namely we have proved the following

Proposition 3.7 *Let L be the generator of the ergodic Markov chain ζ. In the case of continuous time we define $r(x) \equiv -L(x,x)$, whereas in the case of discrete time we put $r(x) \equiv 1$. Then we can conclude that (3.1.1) can be written in the form*

$$P_x[\tau_A < \tau_B] = \begin{cases} h_{A,B}(x) & , x \in Y \setminus (A \cup B), \\ 1 + \frac{1}{r(x)} e_{A,B}(x) & , x \in A, \\ \frac{1}{r(x)} e_{A,B}(x) & , x \in B. \end{cases} \tag{3.1.8}$$

The next definition introduces the essential object that will allow us to estimate the capacity. In particular, it will allow us to treat simultaneously Markov processes with discrete and continuous time set $\mathbb{T}$.

Definition 3.8 The *Dirichlet form* Φ associated to a reversible Markov process ζ with generator L and invariant measure π on the graph Γ is defined as

$$\Phi(h) := -\langle h, Lh\rangle_{\pi} \tag{3.1.9}$$

for all $h \in l^2(\pi)$.

Remark 3.9 (a) In our setting ζ is assumed to have an invariant probability measure and we will use this to obtain a unique Dirichlet form.

(b) Φ has the alternative representation in terms of the conductance matrix C (see Remark 3.17)

$$\Phi(h) = \frac{1}{2} \sum_{(x,y)\in G^*} C_{xy} (h(x) - h(y))^2 \tag{3.1.10}$$

This can be seen by using equation (3.2.4) for the generator of ζ. Therefore the symmetry of C implies

$$\begin{aligned} \Phi(h) &= \sum_{x\in Y} h^2(x) \sum_{y\in\mathcal{N}_x} C_{xy} - \sum_{(x,y)\in G^*} h(x) C_{xy} h(y) \\ &= \frac{1}{2} \sum_{(x,y)\in G^*} C_{xy} (h(x) - h(y))^2 . \end{aligned} \tag{3.1.11}$$

(c) For an arbitrary subset H of G^*, the Dirichlet form *restricted to* H is defined by

$$\Phi_H(h) := \frac{1}{2} \sum_{(x,y)\in H} C_{xy} (h(x) - h(y))^2 . \tag{3.1.12}$$

The following variational representation of the capacity in terms of the Dirichlet form of ζ will turn out to be of fundamental importance. The reason is that it exhibits the monotonicity properties of the capacity. This "Dirichlet principle" can be found for example in the book of Liggett ([Lig85], p. 99, Theorem 6.1).

Theorem 3.10 *(Dirichlet principle) Let ζ be an irreducible Markov chain that is reversible with respect to the positive probability measure π. Let Φ be the associated Dirichlet form. We consider two disjoint subsets of Y, A and B. Let $\mathcal{H}_{A,B}$ denote the space of functions*

$$\mathcal{H}_{A,B} := \left\{ h \in l_2(\pi) \mid h_{|A} = 0 \text{ and } h_{|B} = 1 \right\}. \tag{3.1.13}$$

Then the equilibrium potential is the unique minimiser of Φ inside $\mathcal{H}_{A,B}$, i.e.

$$\Phi(h_{A,B}) = \inf_{h \in \mathcal{H}_{A,B}} \Phi(h). \tag{3.1.14}$$

Remark 3.11 Doyle [Doy89] gives an analogous variational principle in the non-reversible case. Consider the function space

$$\mathcal{G}_{A,B} \equiv \left\{ g \in l_2(\pi) \mid g_{|A \cup B} = 0 \right\}. \tag{3.1.15}$$

Then

$$\left\langle h^*_{A,B}, L h_{A,B} \right\rangle_\pi = \inf_{h \in \mathcal{H}_{A,B}} \sup_{g \in \mathcal{G}_{A,B}} \langle h - g, L(h+g) \rangle_\pi. \tag{3.1.16}$$

Here $h^*_{A,B}$ is the equilibrium potential for the reversed Markov chain ζ^* that has transition probability

$$p^*(x,y) := \frac{\pi(y)}{\pi(x)} p(y,x). \tag{3.1.17}$$

With the properties of the equilibrium potential follows

$$\left\langle h^*_{A,B}, L h_{A,B} \right\rangle_\pi = \Phi(h_{A,B}). \tag{3.1.18}$$

Unfortunately the variational representation (3.1.16) has not the same monotonicity properties as the Dirichlet principle.

The Dirichlet principle motivates the following

Definition 3.12 The *(Newtonian) capacity* of A and B with respect to ζ is defined as

$$\operatorname{cap}(A,B) := \Phi(h_{A,B}). \tag{3.1.19}$$

Remark 3.13 (a) Observe that the capacity is symmetric, since $h_{B,A} = 1 - h_{A,B}$ and $L1 = 0$. Due to the properties of the equilibrium potential and Definition 3.6 of the equilibrium measure, $e_{A,B}$, we can also write

$$\operatorname{cap}(A, B) = -\langle 1_A, e_{A,B}\rangle_\pi . \tag{3.1.20}$$

(b) In contrast to the equilibrium measure, the capacity of the embedded Markov chain of a continuous time process ζ is the same as for ζ. This is implied by the representations (3.0.7) and (3.0.8) for the generator and the invariant measure of the embedded chain.

(c) The representation (3.1.20) together with the identity (3.1.8) implies that

$$\operatorname{cap}(A, B) = \sum_{x \in B} \pi^{(d)}(x)\, P_x(\tau_A < \tau_B), \tag{3.1.21}$$

where $\pi^{(d)}$ is the reversible measure of the embedded Markov chain, see equation (3.0.8). Of course the embedded Markov chain of a Markov chain ζ is the chain itself.

In the special case $B = \{x\}$ we obtain therefore

$$P_x(\tau_A < \tau_x) = \frac{\operatorname{cap}(x, A)}{\pi^{(d)}(x)}. \tag{3.1.22}$$

The next proposition follows directly from Corollary 1.6 of [BEGK01]. It shows that the equilibrium potential can be approximated by capacities. Together with the Dirichlet principle this proposition will provide us a way to improve rough estimates on the capacity.

Proposition 3.14 *[BEGK01] For $A, B \subset Y$, disjoint, $x \notin A \cup B$ and ζ reversible, we obtain*

$$h_{A,B}(x) \le \frac{\operatorname{cap}(x, A)}{\operatorname{cap}(x, B)}. \tag{3.1.23}$$

Proof. Since $x \notin A \cup B$ we have $h_{A,B}(x) = P_x[\tau_A < \tau_B]$. If the process, started at a point x, wants to realise the event $\{\tau_A < \tau_B\}$, it may do so by going to A immediately and without returning to x again, or it may return to x without either going to A or B. Clearly, once the process returns to x it is in the same position as at the starting time, and we can use the strong Markov property. Formally:

$$\begin{aligned} P_x[\tau_A < \tau_B] &= P_x[\tau_A < \tau_{B\cup x}] + P_x[(\tau_x < \tau_{A\cup B}) \wedge (\tau_A < \tau_B)] \\ &= P_x[\tau_A < \tau_{B\cup x}] + P_x[\tau_x < \tau_{A\cup B}]\, P_x[\tau_A < \tau_B] . \end{aligned} \tag{3.1.24}$$

This is called a renewal equation. We can solve this equation for $P_x[\tau_A < \tau_B]$:

$$
\begin{aligned}
P_x[\tau_A < \tau_B] &= \frac{P_x[\tau_A < \tau_{B\cup x}]}{1 - P_x[\tau_x < \tau_{A\cup B}]} \\
&= \frac{P_x[\tau_A < \tau_{B\cup x}]}{P_x[\tau_{A\cup B} < \tau_x]}.
\end{aligned} \tag{3.1.25}
$$

By elementary monotonicity properties this representation yields the bound

$$
P_x[\tau_A < \tau_B] \leq \frac{P_x[\tau_A < \tau_x]}{P_x[\tau_B < \tau_x]} = \frac{\operatorname{cap}(x, A)}{\operatorname{cap}(x, B)}. \tag{3.1.26}
$$

□

3.2 Electrical networks

It will be convenient for the following to use the language of electrical networks. This subsection follows Doyle and Snell [DS84]. We introduce

Definition 3.15 Let $\Gamma = (Y, G)$ be a locally finite connected graph with edgeset G. We denote $G^* := \{(x, y) \in G \,|\, x \neq y\}$, i.e. we leave out all self-edges. Let A and B be subsets of Y and $C : G^* \to \mathbb{R}_{>0}$ a positive symmetric function, called the *conductance matrix* of Γ.

(a) Let $f : G^* \to \mathbb{R}$ be a function and define $\overline{f} : Y \to \mathbb{R}$ by $\overline{f}(x) := \sum_{y \in \mathcal{N}_x} f(x, y)$. f is called a *flow* from A to B and $\overline{f}(x)$ the *net flow out of* x, if

1. (anti-symmetry) $f(x, y) = -f(y, x)$,
2. (Kirchhoff's node law) $\overline{f}(x) = 0$ for all $x \in Y \setminus (A \cup B)$.

f is called *unit flow* if additionally $\sum_{x \in A} \overline{f}(x) = 1$.

(b) An *electrical network* is a weighted graph (Γ, C).

Remark 3.16 Given the values of a function $h : Y \to [0, 1]$, called *voltage*, on the sets A and B there exists a unique flow $i : G^* \to \mathbb{R}$ from A to B, called *current*, such that "Ohm's law"

$$
i(x, y) = C_{xy}(h(x) - h(y)) \tag{3.2.1}
$$

is valid. This follows from Proposition 3.1.2.

Proposition 3.17 *(i) Let $\zeta = (\zeta_t)_{t\in\mathbb{T}}$ be a reversible ergodic Markov chain on a locally finite graph $\Gamma = (Y, G)$. Put $G^* = \{(x,y) \in G \,|\, x \neq y\}$. Then ζ determines an electrical network (Γ, C) with conductance matrix $C : G^* \to \mathbb{R}_{>0}$ given by*

$$C_{xy} := \pi(x) L(x,y). \tag{3.2.2}$$

(ii) On the other hand a reversible ergodic Markov chain on Γ is determined by its invariant probability measure π and an arbitrary conductance matrix, $C : G^* \to \mathbb{R}_{>0}$ such that

$$\sup_{x\in Y} \frac{c(x)}{\pi(x)} < \infty, \qquad \textit{where } c(x) := \sum_{y\in\mathcal{N}_x} C(x,y). \tag{3.2.3}$$

Proof. ad (i). Suppose we are given the Markov chain ζ. Then the conductance matrix C given by (3.2.2) is indeed a symmetric function, because of the reversibility of ζ. Let $\mathcal{N} = \{\mathcal{N}_x\}_{x\in Y}$ be the corresponding neighbourhood system of G^*, i.e. $y \in \mathcal{N}_x$ iff $(x,y) \in G^*$. Then the generator of ζ can be written as

$$L(x,y) = \begin{cases} \frac{C_{xy}}{\pi(x)} & \text{for } y \in \mathcal{N}_x, \\ -\sum_{y\in\mathcal{N}_x} L(x,y) & \text{for } y = x, \\ 0 & \text{else.} \end{cases} \tag{3.2.4}$$

Hence the Dirichlet problem (3.1.4) is equivalent to

$$\begin{cases} \sum_{y\in\mathcal{N}_x} C_{xy}(h(y) - h(x)) = 0 & \text{for } x \in \Gamma \setminus (A \cup B), \\ h(x) = 1 & \text{for } x \in A, \\ h(x) = 0 & \text{for } x \in B. \end{cases} \tag{3.2.5}$$

Therefore the voltage is given by $h(x) = P_x(\tau_A < \tau_B)$ and i defined by (3.2.1) is a flow.

ad (ii). Given a conductance matrix C and a probability measure π that satisfies condition (3.2.3), we retrieve the transition matrix, p, of a reversible Markov chain ζ by setting $Z := \sup_{x\in Y} \frac{c(x)}{\pi(x)}$ and

$$p(x,y) := \begin{cases} \frac{1}{Z}\frac{C_{xy}}{\pi(x)} & \text{for } y \in \mathcal{N}_x, \\ 1 - \sum_{y\in\mathcal{N}_x} p(x,y) & \text{for } y = x, \\ 0 & \text{else.} \end{cases} \tag{3.2.6}$$

Obviously the reversible measure of ξ is indeed given by π. $\square$

Remark 3.18 The representation of the capacity via the Dirichlet principle (3.1.14) shows that $\mathrm{cap}\,(A,B)$ is the "effective conductance" of the electrical network (Γ, C) associated to ξ, when we apply a voltage 1 between A and B, i.e. we set the boundary conditions $h|_A = 1$ and $h|_B = 0$. Compare [DS84], section 3.5, page 63.

Equivalently to the Dirichlet principle (3.1.14) there exists a variational principle for the current, called *Thompson's principle*: Denote

$$\mathcal{F}_{A,B} := \{f : G \to \mathbb{R} \,|\, f \text{ unit flow from } A \text{ to } B\} \tag{3.2.7}$$

then

$$\frac{1}{\mathrm{cap}\,(A,B)} = \inf_{f \in \mathcal{F}_{A,B}} \frac{1}{2} \sum_{x,y \in Y} \frac{1}{C_{xy}} f\,(x,y)^2 \,. \tag{3.2.8}$$

The unique minimiser of this problem is the current i, that satisfies Ohm's law (3.2.1). For a proof see [DS84], p. 63. We will use this principle to obtain precise estimates of the capacity in the case of several relevant saddle points.

Example 3.19 In the case of a finite one-dimensional graph Γ we can calculate the equilibrium potential and the capacity of a network (Γ, C) directly. For $Y = \{0, 1, \ldots, N\}$ we denote $C_k = C\,(k-1, k)$ and obtain for $x \in Y$:

$$h_{0,N}\,(x) = \left(\sum_{k=1}^{x} \frac{1}{C_k} \right) / \left(\sum_{k=1}^{N} \frac{1}{C_k} \right) . \tag{3.2.9}$$

The capacity is given by

$$\begin{aligned} \mathrm{cap}\,(0,N) &= \sum_{k=1}^{N} C_k \left(h_{0,N}\,(k-1) - h_{0,N}\,(k) \right)^2 \\ &= \left(\sum_{k=1}^{N} C_k \frac{1}{C_k^2} \right) / \left(\sum_{k=1}^{N} \frac{1}{C_k} \right)^2 \\ &= 1/\left(\sum_{k=1}^{N} \frac{1}{C_k} \right) . \end{aligned} \tag{3.2.10}$$

3.3 Mean hitting time

Definition 3.20 We introduce the function $w_{A,B} : Y \to \mathbb{R}^+$ by setting

$$w_{A,B}\,(x) = \begin{cases} \mathbb{E}_x \tau_A 1_{\tau_A < \tau_B} & , \; x \notin A \cup B, \\ 0 & , \; x \in A \cup B. \end{cases} \tag{3.3.1}$$

If ζ is a continuous time process we consider again the embedded chain with transition probability given by $p(x,y) := \frac{r(x,y)}{r(x)}$ for $y \in \mathcal{N}_x$ and zero else. For a discrete time Markov process ξ we put $r(x) \equiv 1$ for all $x \in Y$. Then obtain for $w_{A,B}$ the following forward equation for $x \notin A \cup B$:

$$\begin{aligned} w_{A,B}(x) &= \mathbb{E}_x \tau_A 1_{\tau_A < \tau_B} \\ &= \frac{1}{r(x)} P_x [\tau_A < \tau_B] + \sum_{y \notin A \cup B} p(x,y)\, w_{A,B}(y) \\ &= \frac{1}{r(x)} h_{A,B}(x) + p w_{A,B}(x) . \end{aligned} \tag{3.3.2}$$

Therefore $w_{A,B}$ is a solution to the linear boundary problem

$$\begin{cases} -Lw &= h_{A,B} \quad \text{on } (A \cup B)^c , \\ w &= 0 \quad\quad \text{on } A \cup B. \end{cases} \tag{3.3.3}$$

Note that $-L$ is a positive operator. Proposition 3.4 implies that this problem has a unique solution.

Let D be a subset of Y. Define the Green function, $G_D : D \times D \to \mathbb{R}$, to be the kernel of the inverse operator of $-L$ on $l_2(D, \pi)$. The Green function contains all information about the law of the Markov process ζ.

We use the Definition 3.5(a) of the equilibrium potential to represent the Green function. Let C be another subset of Y, disjoint from D. Since $h_{C,D} = 0$ on D and $e_{C,D} = 0$ on $(C \cup D)^c$, we obtain

$$\begin{aligned} h_{C,D}(x) &= -G_{D^c} e_{C,D}(x) \\ &= -\sum_{y \in C} G_{D^c}(x,y)\, e_{C,D}(y) . \end{aligned} \tag{3.3.4}$$

We will use now the reversibility of ξ, that means $\pi(x)\, G_D(x,y) = \pi(y)\, G_D(y,x)$ and choose $C = \{y\}$. Then we obtain

$$\begin{aligned} G_{D^c}(x,y) &= -\frac{h_{y,D}(x)}{e_{y,D}(y)} = -\frac{\pi(y)\, h_{x,D}(y)}{\pi(x)\, e_{x,D}(x)} \\ &= \pi(y) \frac{h_{x,D}(y)}{\operatorname{cap}(x,D)}. \end{aligned} \tag{3.3.5}$$

This means, we can in principle determine the law of ζ completely, if we know the capacity and the equilibrium potential.

We summarise the results in the next Proposition that resembles Corollary 3.3 of [BEGK02]

Proposition 3.21 *[BEGK02] The Dirichlet Green function for any set $D \subset Y$ can be represented in terms of the equilibrium potential and capacities as*

$$G_{D^c}(x, y) = \pi(y) \frac{h_{x,D}(y)}{\mathrm{cap}(x, D)}. \tag{3.3.6}$$

The mean hitting time of $A \subset Y$ satisfies, for a starting point $x \notin A \cup B$,

$$\mathbb{E}_x \tau_A 1_{\tau_A < \tau_B} = \frac{1}{\mathrm{cap}(x, A \cup B)} \sum_{y \in (A \cup B)^c} \pi(y) h_{x, A \cup B}(y) h_{A,B}(y). \tag{3.3.7}$$

Especially for $B = \emptyset$ we obtain for all $x \notin A$

$$\mathbb{E}_x \tau_A = \frac{1}{\mathrm{cap}(x, A)} \sum_{y \in A^c} \pi(y) h_{x,A}(y). \tag{3.3.8}$$

4 Metastability

Let $\Lambda \subset \mathbb{R}^d$ be open and connected and consider a family of countable discrete sets, $(\Lambda_\epsilon)_{0<\epsilon<1} \subset \Lambda$. We assume that Λ_ϵ is equipped with a neighbourhood system $\mathcal{N}_\epsilon = \{\mathcal{N}_{\epsilon,x}\}_{x\in\Lambda_\epsilon}$, that makes it a connected set. Moreover, assume there exists $r > 0$, independent of ϵ, such that the number of neighbours is uniformly bounded by r, i.e. $|\mathcal{N}_{\epsilon,x}| \leq r$ for all $x \in \Lambda_\epsilon$. The associated r-uniformly locally finite graph is denoted by $\Gamma_\epsilon = (\Lambda_\epsilon, E_\epsilon)$.

Let $(\xi^\epsilon)_{0<\epsilon<1}$ be a family of ergodic time-homogeneous Markov chains on Γ_ϵ. Assume that ξ^ϵ is reversible with respect to the probability distribution π_ϵ. Let $p_\epsilon : \Lambda_\epsilon \times \Lambda_\epsilon \to [0,1]$ the transition probability of ξ^ϵ. Recall that we assume that ξ^ϵ only jumps between neighbours of Λ_ϵ, i.e. $p_\epsilon(x,y) = 0$ for all $y \notin \mathcal{N}_x \cup x$.

Since ξ^ϵ is reversible, p_ϵ can always be written in the form

$$p_\epsilon(x,y) = g_\epsilon(x,y) \min\left(1, \frac{\pi_\epsilon(y)}{\pi_\epsilon(x)}\right) \tag{4.0.1}$$

with a non negative symmetric function $g_\epsilon : \Lambda_\epsilon \times \Lambda_\epsilon \to \mathbb{R}_{\geq 0}$.

We assume

C1 the function g_ϵ is on compact sets uniformly bounded from below, i.e. for all $K \subset\subset \Lambda_\epsilon$ there exists a constant $c > 0$, independent of ϵ, such that $g_\epsilon(x,y) \geq c$ for all $x \in K$ and $y \in \mathcal{N}_{\epsilon,x}$.

This assures in particular that ξ^ϵ can jump between any two neighbours of Λ_ϵ and is not restricted to some connected subgraph.

Example 4.1 Consider $\Lambda_\epsilon \equiv \Lambda \cap \epsilon\mathbb{Z}^d$. Let x and y be neighbours, i.e. $\|x - y\| = \epsilon$.

(a) For $g_\epsilon(x,y) := \frac{1}{2d}$ we obtain, of course, the Metropolis algorithm.

(b) For

$$g_\epsilon(x,y) := \frac{1}{2d} \frac{\pi_\epsilon(x) \vee \pi_\epsilon(y)}{\pi_\epsilon(x) + \pi_\epsilon(y)} \geq \frac{1}{4d}, \tag{4.0.2}$$

we recover the heat bath dynamics, i.e.

$$p_\epsilon(x,y) = \frac{1}{2d} \frac{\pi_\epsilon(y)}{\pi_\epsilon(x) + \pi_\epsilon(y)}. \tag{4.0.3}$$

(c) For

$$g_\epsilon(x,y) := f_\epsilon(x,y)\sqrt{\frac{\pi_\epsilon(x)\wedge\pi_\epsilon(y)}{\pi_\epsilon(x)\vee\pi_\epsilon(y)}} \geq f_\epsilon(x,y), \tag{4.0.4}$$

where f is a non negative symmetric function that is uniformly bounded from below on compact subsets of $\Lambda_\epsilon \times \Lambda_\epsilon$, we recover the dynamics given by

$$p_\epsilon(x,y) = f_\epsilon(x,y)\sqrt{\frac{\pi_\epsilon(y)}{\pi_\epsilon(x)}}. \tag{4.0.5}$$

Definition 4.2 We define the *potential* of ξ^ϵ to be the function $F_\epsilon : \Lambda_\epsilon \to \mathbb{R}_{>0}$ with

$$F_\epsilon(x) := -\epsilon \ln \pi_\epsilon(x). \tag{4.0.6}$$

The interesting case for us occurs when F_ϵ has at least two local minima.

We assume that

F1 (F_ϵ) converges uniformly to a unique continuous function $F : \Lambda \to \mathbb{R}_{\geq 0}$, i.e. for all $\kappa > 0$ there exists $\epsilon_0 > 0$ such that for all $\epsilon < \epsilon_0$ we have

$$\sup_{x\in\Lambda_\epsilon} |F_\epsilon(x) - F(x)| < \kappa. \tag{4.0.7}$$

F2 The function F has compact lower level sets, i.e.

$$\{F \leq b\} \subset\subset \Lambda \quad \text{for all } b \geq 0. \tag{4.0.8}$$

The following definition of the so called relevant saddle points between two subsets A and B of Y will be essential for the dynamics of the Markov processes we consider.

Definition 4.3 Consider an arbitrary function $f : Y \to \mathbb{R}$ on a locally finite graph (Y, G). Let A and B be disjoint subsets of Λ_ϵ.

(a) A *path* γ is a finite sequence $(\gamma_1, \ldots, \gamma_k)$ of *communicating points*, i.e. $(\gamma_i, \gamma_{i+1}) \in G$ for $1 \leq i \leq k-1$. We write $x \in \gamma$ when γ visits the point x. We denote by $\mathcal{P}_{A,B}$ the set of paths starting in A and ending in B.

(b) The *communication height* between A and B is

$$\hat{f}(A,B) := \min_{\gamma\in\mathcal{P}_{A,B}} \max_{x\in\gamma} f(x). \tag{4.0.9}$$

Observe that the communication height depends, of course, on the edgeset G we have chosen.

(c) We introduce the level set

$$\mathcal{G}(A,B) := \left\{ z \in \Lambda_\epsilon \,|\, f(z) = \hat{f}(A,B) \right\}. \tag{4.0.10}$$

The *set of optimal path* is defined by

$$\mathcal{O}_{A,B} := \left\{ \gamma \in \mathcal{P}_{A,B} \,|\, \max_{x \in \gamma} f(x) = \hat{f}(A,B) \right\}. \tag{4.0.11}$$

A *gate* $G(A,B)$ is a minimal subset of $\mathcal{G}(A,B)$ with the property that all optimal paths intersect $G(A,B)$. That means for every $H \subsetneq G(A,B)$ there exists a path $\gamma \in \mathcal{O}_{A,B}$ such that $\gamma \cap H = \emptyset$. Note that $G(A,B)$ is in general not unique. The *set* $S(A,B)$ *of relevant saddle points* is the union over all gates $G(A,B)$.

The notion of communication height between two sets leads a decomposition of the state space into different valleys, described by the following

Definition 4.4 Let $A, B \subset\subset \Lambda$ be disjoint compact sets.

(a) We define the lower level set

$$W(A,B) := \left\{ x \in \Lambda \,|\, F(x) < \hat{F}(A,B) \right\}. \tag{4.0.12}$$

We assume that $A \subset W(A,B)$. We set $V_B(A)$, called the *valley of* A *with respect to* B, denotes the connected component of $W(A,B)$ containing A.

(b) Let $x \in \Lambda \setminus A$. Then we define the barrier between x and A by

$$B(x,A) := \hat{F}(x,A) - F(x). \tag{4.0.13}$$

$B(x,A)$ is the minimal height a path has to climb in order to connect x with A.

Analogously we define B_ϵ and $V_B^{(\epsilon)}(A)$ for F_ϵ.

4.1 Metastability

In the following, we will often be dealing with probabilities like $P_x[\tau_A < \tau_x]$ which we call *escape probability* from x to A.

Following Bovier, Eckhoff, Gayrard and Klein [BEGK02], we define metastability in the following way:

Definition 4.5 (metastability) Let $\mathscr{M}_\epsilon$ be a finite subset of Λ_ϵ such that the cardinality $|\mathscr{M}_\epsilon|$ is independent of ϵ. Let $\rho : (0,1) \to (0,1)$ be a monotone increasing function with $\lim_{\epsilon \downarrow 0} \rho(\epsilon) = 0$.

Then the family of Markov processes $(\xi^\epsilon)_{\epsilon \in (0,1)}$ is said to be *ρ-metastable* with respect to $(\mathscr{M}_\epsilon)$, if

$$\max_{m \in \mathscr{M}_\epsilon} P_m\left(\tau_{\mathscr{M}_\epsilon} < \tau_m\right) \leq \rho(\epsilon) \inf_{x \notin \mathscr{M}_\epsilon} P_x\left(\tau_{\mathscr{M}_\epsilon} < \tau_x\right). \tag{4.1.1}$$

The elements of $\mathscr{M}_\epsilon$ are called *ρ-metastable points* of (ξ^ϵ).

We say (ξ^ϵ) is metastable with respect to $(\mathscr{M}_\epsilon)$ if there exists a function ρ for which (ξ^ϵ) is ρ-metastable.

This definition suggest a decomposition of the state space into a finite collection of subsets. We define for each point $m \in \mathscr{M}_\epsilon$ the domain of attraction of m by

$$A(m, \mathscr{M}_\epsilon) := \left\{ x \in \Lambda_\epsilon \,|\, P_x\left(\tau_m = \tau_{\mathscr{M}_\epsilon}\right) \geq \max_{n \in \mathscr{M}_\epsilon} P_x\left(\tau_n = \tau_{\mathscr{M}_\epsilon}\right) \right\}. \tag{4.1.2}$$

It follows from Definition 4.5 of metastability that for all $m \in \mathscr{M}_\epsilon$

$$\lim_{\epsilon \downarrow 0} P_m\left(\tau_{\mathscr{M}_\epsilon \setminus m} < \tau_m\right) = 0. \tag{4.1.3}$$

Hence, if there exists a limiting Markov chain, it is reducible with at least $|\mathscr{M}_\epsilon|$ connected components.

In words, Definition 4.5 states the following: The infimum of the escape probabilities from any point $x \in \mathscr{M}_\epsilon^c$ to $\mathscr{M}_\epsilon$ is much bigger than the escape probability from a point, m, in $\mathscr{M}_\epsilon$ to another one. The function ρ in Definition 4.5 describes the factor, by which the escape probabilities between metastable points is smaller compared to the escape probability of any point with respect to the set of metastable points.

Therefore we have at least two different time scales: One that measures the time required for a typical excursion away from m that stays inside $A(m, \mathscr{M}_\epsilon)$ and another one on which we expect a changeover to $\mathscr{M}_\epsilon \setminus m$. This type of behaviour has been studied for a long time and is rigorously treated on the level of large deviations, in particular in the book of Freidlin and Wentzell [FW84].

The benefit of Definition 4.5 is that we only have to control hitting times of points or finite sets of points on the state space. In the analogues situation of a Diffusion in $\mathbb{R}^d$, one can deal with small balls around these points (see [BEGK04]).

Observe that Definition 4.5 does not determine a unique family $(\mathscr{M}_\epsilon)$ even for fixed ρ. Indeed, having isolated a very large set $\mathscr{M}_\epsilon$, in many cases one can find a subset $\mathcal{N}_\epsilon \subset \mathcal{M}_\epsilon$ such that the Markov chain also exhibits a metastable behaviour with respect to $\mathcal{N}_\epsilon$. We formulate this important property of Definition 4.5 in

Proposition 4.6 *Let I_ϵ be the set of all $i \in \mathscr{M}_\epsilon$ such that there exists c, independent of ϵ, and*

$$P_i\left(\tau_{\mathscr{M}_\epsilon} < \tau_i\right) \geq c \max_{m \in \mathscr{M}_\epsilon} P_m\left(\tau_{\mathscr{M}_\epsilon} < \tau_m\right). \tag{4.1.4}$$

Then we can construct a minimal set $J_\epsilon \subset I_\epsilon$ such that (ξ^ϵ) is metastable with respect to $\hat{\mathscr{M}}_\epsilon \equiv \mathscr{M}_\epsilon \setminus J_\epsilon$.

Proof. The definition of I_ϵ in (4.1.4) implies that there exist a monotone decreasing function, $r : (0,1) \to [0,1]$, with $\lim_{\epsilon \downarrow 0} r(\epsilon) = 0$ such that for all $m \in \mathscr{M}_\epsilon \setminus I_\epsilon$

$$P_m\left(\tau_{\mathscr{M}_\epsilon} < \tau_m\right) \leq r(\epsilon) \max_{x \in \mathscr{M}_\epsilon} P_x\left(\tau_{\mathscr{M}_\epsilon} < \tau_x\right). \tag{4.1.5}$$

So at first sight it might be possible just to leave out all elements of I_ϵ from $\mathscr{M}_\epsilon$ to get a new metastable set, but this is not possible if some or all relevant saddle points connect members of I_ϵ, i.e. there exists $i, j \in I_\epsilon$ and

$$\hat{F}_\epsilon\left(i, \mathscr{M}_\epsilon \setminus i\right) = \hat{F}_\epsilon\left(i, j\right). \tag{4.1.6}$$

In this case it may happen that by throwing away i and j there arises a valley of arbitrary depth that is not any more represented by an element of $\mathscr{M}_\epsilon \setminus I_\epsilon$.

We construct inductively the set J_ϵ by putting $J^{(0)} \equiv \emptyset$ and $J^{(n+1)} = J^{(n)} \cup \{j\}$ if there exists $j \in I_\epsilon \setminus J^{(n)}$ and $c > 0$, independent of ϵ, such that

$$P_j\left(\tau_{\mathscr{M}_\epsilon \setminus J^{(n)}} < \tau_j\right) \geq c \max_{m \in \mathscr{M}_\epsilon} P_m\left(\tau_{\mathscr{M}_\epsilon} < \tau_m\right). \tag{4.1.7}$$

Otherwise put $J_\epsilon \equiv J^{(n)}$. Without loss of generality we assume that for all $i \in J_\epsilon$

$$\pi_\epsilon(i) = \max_{x \in V_{\hat{\mathscr{M}}_\epsilon}(i)} \pi_\epsilon(x). \tag{4.1.8}$$

Let $i \in J_\epsilon$, then certainly it holds true that

$$\begin{aligned} P_i\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_i\right) &\leq P_i\left(\tau_{\mathscr{M}_\epsilon} < \tau_i\right) \\ &\leq \rho(\epsilon) \inf_{x \notin \mathscr{M}_\epsilon} P_x\left(\tau_{\mathscr{M}_\epsilon} < \tau_x\right). \end{aligned} \tag{4.1.9}$$

Moreover, for all $x \notin \mathscr{M}_\epsilon$

$$P_x\left(\tau_{\mathscr{M}_\epsilon} < \tau_x\right) \leq \sum_{i \in J_\epsilon} P_x\left(\tau_i < \tau_x\right) + P_x\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_x\right). \tag{4.1.10}$$

We denote $A_\epsilon(J_\epsilon) \equiv \cup_{i \in I_\epsilon} A(i, \mathscr{M}_\epsilon)$. Then for all $i \in J_\epsilon$ and $x \notin A_\epsilon(J_\epsilon)$ we know

$$P_x\left(\tau_i < \tau_x\right) \leq P_x\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_x\right). \tag{4.1.11}$$

Therefore we have shown up to now

$$\max_{i \in J_\epsilon} P_i\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_i\right) \leq 2\rho(\epsilon) \inf_{x \notin \mathscr{M}_\epsilon \cup A_\epsilon(J_\epsilon)} P_x\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_x\right). \tag{4.1.12}$$

To proceed we use that

$$P_x\left(\tau_A < \tau_x\right) = \frac{\operatorname{cap}(x, A)}{\pi_\epsilon(x)}. \tag{4.1.13}$$

Let $i \in J_\epsilon$, then certainly,

$$\hat{F}_\epsilon\left(x, \hat{\mathscr{M}}_\epsilon\right) = \hat{F}_\epsilon\left(i, \hat{\mathscr{M}}_\epsilon\right) \qquad \text{for all } x \in V_{\hat{\mathscr{M}}_\epsilon}(i). \tag{4.1.14}$$

On the other hand for $x \in A\left(i, \hat{\mathscr{M}}_\epsilon\right) \setminus V_{\hat{\mathscr{M}}_\epsilon}(i)$ we obtain

$$\hat{F}_\epsilon\left(x, \hat{\mathscr{M}}_\epsilon\right) = x. \tag{4.1.15}$$

Therefore Proposition 4.8 tells us that for all $i \in J_\epsilon$ and $x \in A\left(i, \hat{\mathscr{M}}_\epsilon\right)$ the condition (4.1.8) implies

$$P_i\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_i\right) \leq \inf_{x \in A(i, \mathscr{M}_\epsilon)} P_x\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_x\right). \tag{4.1.16}$$

Therefore

$$\begin{aligned} &\min_{i \in J_\epsilon} P_i\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_i\right) \\ &\leq \inf_{x \in A_\epsilon(J_\epsilon)} P_x\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_x\right). \end{aligned} \tag{4.1.17}$$

We have shown

$$\max_{i\in J_\epsilon} P_i\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_i\right) \leq 2\rho(\epsilon) \inf_{x\notin \mathscr{M}_\epsilon} P_x\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_x\right). \tag{4.1.18}$$

By construction of J_ϵ there exists a monotone decreasing function, $\hat{\rho} : (0,1) \to [0,1]$, with $\lim_{\epsilon\downarrow 0} \hat{\rho}(\epsilon) = 0$ such that

$$\max_{n\in \hat{\mathscr{M}}_\epsilon} P_n\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_n\right) \leq \hat{\rho}(\epsilon) \min_{i\in J_\epsilon} P_i\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_i\right). \tag{4.1.19}$$

Hence,

$$\max_{n\in \hat{\mathscr{M}}_\epsilon} P_n\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_n\right) \leq \hat{\rho}(\epsilon) \inf_{x\notin \hat{\mathscr{M}}_\epsilon} P_x\left(\tau_{\hat{\mathscr{M}}_\epsilon} < \tau_x\right) \tag{4.1.20}$$

holds true and we are done. $\square$

Remark 4.7 Observe that if I_ϵ contains more than one point, then the small eigenvalues of the generator $-L_\epsilon$ of ξ^ϵ depend on the structure of this set. We will give a non trivial example in the case of the Hopfield model in part III of this treatise.

4.2 priori bounds

In this section we will estimate capacities of ξ^ϵ on a subexponential scale and then use Proposition 3.14 to give an a priori bound on the equilibrium potential. We use the notions of the electrical network, $\left(\Gamma_\epsilon, C^{(\epsilon)}\right)$, associated to ξ^ϵ given in Definition 3.15, see Proposition 3.17. We consider only Markov chains, i.e. Markov processes with time set $\mathbb{T} = \mathbb{N}_0$ in this section. This corresponds to the following property of the generator: $\sum_{y\in\mathcal{N}_x} L(x,y) \leq 1$ for all $x \in \Lambda_\epsilon$. In the case of continuous time one can think this as an description of the embedded Markov chain.

The following proposition will play a key rôle in our treatment.

Proposition 4.8 *Let (ξ^ϵ) be a family of positive recurrent reversible Markov chain that satisfies the conditions at the beginning of this section, in particular **C1**. Let A and B be disjoint compact sets of Λ_ϵ such that $F_\epsilon(x) < \hat{F}_\epsilon(A,B)$ for all $x \in A$.*

Then there exist constants $c_1, c_2 > 0$ such that

$$c_1 \epsilon^d \leq \frac{\mathrm{cap}\,(A, B)}{\exp\left(-\hat{F}_\epsilon\,(A, B)\,/\epsilon\right)} \leq c_2 \epsilon^{-d}. \tag{4.2.1}$$

Proof. LOWER BOUND OF $\mathrm{cap}\,(A, B)$
The Dirichlet principle of Theorem 3.1.14 tells us

$$\begin{aligned} \mathrm{cap}\,(A, B) &= \inf_{h \in \mathcal{H}_{A,B}} \Phi\,(h) = \Phi\,(h_{A,B}) \\ &\geq \Phi_\gamma\,(h_{A,B}) \geq \inf_{h \in \mathcal{H}_{A,B}} \Phi_\gamma\,(h) \end{aligned} \tag{4.2.2}$$

for every subset $\gamma \subset \Lambda_\epsilon$ such that $\gamma \cap A$ and $\gamma \cap B$ are not empty. We choose now for γ an optimal path, i.e. $\gamma \in \mathcal{O}_{A,B}$. Identify γ with a graph with edges between nearest neighbours. By using the calculation in Remark 3.13(b), we obtain

$$\inf_{h \in \mathcal{H}_{A,B}} \Phi_\gamma\,(h) = 1/\left(\sum_{(x,y) \in \gamma} 1/C_{xy}^{(\epsilon)}\right). \tag{4.2.3}$$

Since γ is an optimal path we know $\gamma \in \left\{F_\epsilon \leq \hat{F}_\epsilon\,(A, B)\right\}$ which is a compact set, because of assumptions **F1** and **F2**. Assumption **C1** assures now the existence of a constant $c > 0$ such that we can estimate

$$C_{xy} \geq c \exp\left(-\hat{F}_\epsilon\,(A, B)\,/\epsilon\right) \qquad \text{for all } x, y \in \gamma. \tag{4.2.4}$$

Therefore

$$\begin{aligned} &\inf_{h \in \mathcal{H}_{A,B}} \Phi_\gamma\,(h) \\ &\geq \frac{c}{|\gamma|} \exp\left(-\hat{F}_\epsilon\,(A, B)\,/\epsilon\right) \\ &\geq \frac{c\epsilon^d}{\mathrm{vol}\left\{F \leq \hat{F}\,(A, B)\right\}} \exp\left(-\hat{F}_\epsilon\,(A, B)\,/\epsilon\right). \end{aligned} \tag{4.2.5}$$

Note that $\mathrm{vol}\left(\left\{F_\epsilon \leq \hat{F}_\epsilon\,(A, B)\right\}\right) < \infty$ follows from assumption **F2**.
UPPER BOUND OF CAP (x, B)

Denote by $V_\epsilon := V_B^{(\epsilon)}(A)$ the valley of A with respect to B (see Definition 4.4(a)). We choose a function h^+ with $h^+ = 0$ on V_ϵ and $h^+ = 1$ on V_ϵ^c. Then we obtain, since by reversibility $C_{xy}^{(\epsilon)} \leq \pi_\epsilon(x) \wedge \pi_\epsilon(y)$,

$$\begin{aligned}
\operatorname{cap}(A,B) &\leq \Phi\left(h^+\right) \\
&= \sum_{x\in\partial^- V_\epsilon}\sum_{y\in\partial^+ V_\epsilon} C_{xy}^{(\epsilon)} \\
&\leq r\left|\partial^- V_\epsilon\right| \exp\left(-\hat{F}_\epsilon(A,B)/\epsilon\right) \\
&\leq c\operatorname{vol}(V_B(A))\,\epsilon^{-d}\exp\left(-\hat{F}_\epsilon(A,B)/\epsilon\right). \qquad (4.2.6)
\end{aligned}$$

□

Remark 4.9 Let $x \in \Lambda_\epsilon$ and $D \subset \Lambda_\epsilon \setminus x$ such that $F_\epsilon(y) < \hat{F}_\epsilon(x,D)$ for all $y \in D$. Then we apply Proposition 4.8 with B containing only one point. With equation 3.1.22 we obtain that the escape probabilities are controlled on an exponential scale by the associated barriers:

$$P_x(\tau_D < \tau_x) \leq c_2\epsilon^{-d}\exp\left[-\frac{1}{\epsilon}B_\epsilon(x,D)\right] \qquad (4.2.7)$$

and

$$P_x(\tau_D < \tau_x) \geq c_1\epsilon^{d}\exp\left[-\frac{1}{\epsilon}B_\epsilon(x,D)\right]. \qquad (4.2.8)$$

This implies that $V_{\mathscr{M}_\epsilon\setminus m}(m) \subset A_\epsilon(m)$.

Example 4.10 (a) Assume that the set $\mathcal{M}$ of local minima of F consists of finitely many points. Denote $\kappa := \min_{m\in\mathcal{M}} B(m, \mathcal{M}\setminus m)$. Then we can find finite sets, $\mathscr{M}_\epsilon$, of local minima of F_ϵ such that $\mathscr{M}_\epsilon \to \mathcal{M}$ with respect to the Hausdorff distance of sets, $|\mathscr{M}_\epsilon| = |\mathcal{M}|$ and

$$\pi_\epsilon(m) = \max_{x\in V_{\mathscr{M}_\epsilon}(m)} \pi_\epsilon(x) \qquad (4.2.9)$$

for all $m \in \mathscr{M}_\epsilon$. Let $\rho(\epsilon) := \exp(-h/\epsilon)$ with $h < \kappa$. Proposition 4.8 shows that ξ^ϵ is then ρ-metastable with respect to $\mathscr{M}_\epsilon$, since for $m \in \mathscr{M}_\epsilon$

$$P_m\left(\tau_{\mathscr{M}_\epsilon\setminus m} < \tau_m\right) \;\leq\; c\epsilon^{-d}\exp\left(-\frac{1}{\epsilon}B_\epsilon(m,\mathscr{M}_\epsilon\setminus m)\right). \qquad (4.2.10)$$

Choose $\delta < \frac{1}{3}(\kappa - h)$. The uniform convergence of F_ϵ to F (assumption **F**) implies that there exists $\epsilon_0 > 0$ such that for all $\epsilon < \epsilon_0$ we have

$$|B_\epsilon(m, \mathscr{M}_\epsilon \setminus m) - B_\epsilon(m, \mathcal{M} \setminus m)| < \delta. \tag{4.2.11}$$

Hence

$$P_m\left(\tau_{\mathscr{M}_\epsilon \setminus m} < \tau_m\right) \leq c\epsilon^{-d} \exp\left(-\frac{\kappa - \delta}{\epsilon}\right). \tag{4.2.12}$$

Moreover, for $x \notin \mathscr{M}_\epsilon$ there exists ϵ_1 such that for all $\epsilon < \epsilon_1$ we have

$$\hat{F}(x, \mathscr{M}_\epsilon) - F_\epsilon(x) < \kappa - h - 2\delta \tag{4.2.13}$$

and therefore

$$P_x(\tau_{\mathcal{M}_\epsilon} < \tau_x) \quad \geq \quad c_1\epsilon^d \exp\left(\frac{\kappa - h - 2\delta}{\epsilon}\right). \tag{4.2.14}$$

Hence for all $\epsilon < \min(\epsilon_0, \epsilon_1)$ we obtain

$$\max_{m \in \mathscr{M}_\epsilon} P_m\left(\tau_{\mathscr{M}_\epsilon \setminus m} < \tau_m\right) \leq \rho(\epsilon) \inf_{x \notin \mathscr{M}_\epsilon} P_x(\tau_{\mathcal{M}_\epsilon} < \tau_x), \tag{4.2.15}$$

and therefore (ξ^ϵ) is ρ-metastable with respect to $\mathscr{M}_\epsilon$.

(b) More generally, in the case $\rho(\epsilon) = \exp(k/\epsilon)$ (with $k > 0$) a metastable set $\mathcal{M}_k$ has the following property: In each valley of depth greater k exactly one of the deepest minima of this valley is in $\mathcal{M}_k$. In this case we have

$$B(m, n) > k \quad \forall m, n \in \mathcal{M}_k. \tag{4.2.16}$$

Moreover, for all other points $x \notin \mathcal{M}_k$ there has to be a point $m \in \mathcal{M}_k$ such that $B(x, m) < k$. Effectively we only have a condition for local minima of F that are outside of $\mathcal{M}_k$, namely

$$B(x, \mathcal{M}_k) < k \quad \forall x \in \mathcal{M} \setminus \mathcal{M}_k. \tag{4.2.17}$$

We have found a connection between Definition 4.5 of a metastable set and geometric properties of the function F_ϵ.

To prove the precise bounds on the capacity between minima, m and n, we need the following corollary, which will justify to restrict our attention to a neighbourhood of the set $S_\epsilon(m, n)$ of relevant saddle points.

Corollary 4.11 *(a) Let A and B be disjoint compact sets of Λ_ϵ such that $F_\epsilon(x) \le \hat{F}_\epsilon(A,B)$ for all $x \in A \cup B$. Then there exists $c > 0$ such that for $x \notin A \cup B$*

$$h_{A,B}(x) \le c\epsilon^{-2d} \exp\left\{-\frac{1}{\epsilon}\left(\hat{F}_\epsilon(x,A) - \hat{F}_\epsilon(x,B)\right)\right\}. \tag{4.2.18}$$

Proof. This follows from Proposition 3.14 combined with Proposition 4.8. □

4.3 Pathwise approach

A point $m \in \Lambda_\epsilon$ is called *metastable in the sense of Olivieri and Vares* iff the following two properties hold:

1. *"unpredictability of the tunneling time"*.
   Assume ξ^ϵ starts in m. Then $\tau_{\mathcal{M}_\epsilon \setminus m}$ is called *unpredictable* if it converges in distribution to an exponential random variable, i.e.
   $$\frac{\tau_{\mathcal{M}_\epsilon \setminus m}}{E_m\left(\tau_{\mathcal{M}_\epsilon \setminus m}\right)} \xrightarrow{\mathcal{D}} \mathcal{E} \quad \text{for } \epsilon \downarrow 0, \tag{4.3.1}$$
   where $\mathcal{E}$ is a unit mean exponential random variable.

2. *"thermalisation"*.
   Let $s, t \in \mathbb{N}_0$. We define the empirical average measure of ξ^ϵ between the times s and $s+t$ as
   $$\mu_{s,t} := \frac{1}{t} \sum_{k=s+1}^{s+t} \delta_{\xi^\epsilon_k}. \tag{4.3.2}$$
   Hence $\mu_{s,t}(B)$ is the fraction of time ξ^ϵ spends in $B \subset \Lambda_\epsilon$ between s and $s+t$. Let $V := V^{(\epsilon)}_{\mathscr{M}_\epsilon \setminus m}(m)$ be the valley of m with respect to $\mathscr{M}_\epsilon \setminus m$.

   Let ξ^ϵ again start at m. We say ξ^ϵ *thermalises at* m if there exists a deterministic time scale t_ϵ such that $\lim_{\epsilon\downarrow 0} t_\epsilon = \infty$, but $t_\epsilon = o\left(E_m\left(\tau_{\partial^+ V}\right)\right)$ and for every open set $B \subset \mathbb{R}^d$ containing m and every $\kappa > 0$
   $$\lim_{\epsilon \downarrow 0} P_m\left(\tau_{\partial^+ V} > t_\epsilon \text{ and } \sup_{s < \tau_{\partial^+ V} - t_\epsilon} \mu_{s,t_\epsilon}(B) > 1 - \kappa\right) = 1. \tag{4.3.3}$$

The next proposition gives a rough estimate for the distribution function of the hitting time of the boundary of a valley.

Proposition 4.12 *Let $V_\epsilon := V_n^{(\epsilon)}(m)$ be the valley of $m \in \mathcal{M}_\epsilon$ with respect to another point $n \in \mathcal{M}_\epsilon$. Then*

$$P_m\left(\tau_{\partial^+ V_\epsilon} < t\right) \le c\epsilon^{-d} \lfloor t \rfloor \exp\left(-\frac{1}{\epsilon} B_\epsilon(m,n)\right). \tag{4.3.4}$$

Proof. We observe similar to Olivieri and Vares (see [OV05] Proposition 4.7, p. 233) that for $x \in \partial^+ V_\epsilon$

$$\begin{aligned} P_m(\tau_x < t) &\le \sum_{k=1}^{\lfloor t \rfloor} P_m\left(\xi_k^\epsilon = x\right) \\ &= \sum_{k=1}^{\lfloor t \rfloor} \frac{1}{\pi_\epsilon(m)} P_{\pi_\epsilon}\left(\xi_0^\epsilon = m, \xi_k^\epsilon = x\right) \\ &\le \sum_{k=1}^{\lfloor t \rfloor} \frac{1}{\pi_\epsilon(m)} P_{\pi_\epsilon}\left(\xi_k^\epsilon = x\right) \\ &= \lfloor t \rfloor \frac{\pi_\epsilon(x)}{\pi_\epsilon(m)}. \end{aligned} \tag{4.3.5}$$

Therefore we obtain, since $\epsilon^d \mathrm{vol}(V_\epsilon)$ converges to the volume of compact set $V_n(m) \subset \Lambda$,

$$\begin{aligned} P_m\left(\tau_{\partial^+ V} < t\right) &\le \frac{\lfloor t \rfloor}{\pi_\epsilon(m)} \sum_{x \in \partial^+ V_\epsilon} \pi_\epsilon(x) \\ &\le \frac{\lfloor t \rfloor}{\pi_\epsilon(m)} \mathrm{vol}\left(V_n(m)\right) \epsilon^{-d} \exp\left(-\frac{1}{\epsilon} \hat{F}_\epsilon(m,n)\right). \end{aligned} \tag{4.3.6}$$

□

To prove the thermalisation of ξ^ϵ in the valley of a metastable point $m \in \mathcal{M}_\epsilon$ that contains no more minima, we need moreover the following

Lemma 4.13 *Let $V_\epsilon := V_{\mathcal{M}_\epsilon \setminus m}^{(\epsilon)}(m)$ be the valley of $m \in \mathcal{M}_\epsilon$ with respect to $\mathcal{M}_\epsilon \setminus m$. Denote $a_\epsilon := \exp(a/\epsilon)$, where $0 < a < B_\epsilon(m, \mathcal{M}_\epsilon \setminus m)$. Given*

$\kappa \in (0,1)$ *and an open set, B, containing m, there exists $c_\kappa > 0$ and ϵ_0 such that for all $\epsilon < \epsilon_0$ and all integer $t \geq a_\epsilon$*

$$\sup_{i \in V} P_i \left(\overline{\mu}_t (B) < 1 - \kappa\right) < \exp\left(-c_\kappa \frac{t}{a_\epsilon}\right). \tag{4.3.7}$$

Proof. It suffices to consider the case, where B is a small ball of radius $\rho > 0$ around m. Denote the *depth* of B by

$$F_\rho := \inf_{\|x-m\|=\rho} \left(F(x) - F(m)\right), \tag{4.3.8}$$

fix $b < a \wedge F_\rho$ and let $b_\epsilon := \exp(b/\epsilon)$. First we introduce

$$q_\epsilon := \sup_{i \in V} P_i \left(\overline{\tau}_m > \sqrt{b_\epsilon}\right) + P_m \left(\overline{\tau}_\rho < b_\epsilon\right) \tag{4.3.9}$$

with $\overline{\tau}_\rho := \min\left\{n \geq 1 \,|\, \overline{\xi}_k^\epsilon \notin B\right\}$. To see that $q_\epsilon \to 0$ in the limit of vanishing ϵ, we use the Chebyshev inequality to estimate the first summand and obtain

$$P_i \left(\overline{\tau}_m > \sqrt{b_\epsilon}\right) \leq \frac{1}{\sqrt{b_\epsilon}} E_i \left(\overline{\tau}_m\right). \tag{4.3.10}$$

Moreover, we obtain with Corollary 4.11

$$\sum_{y \in V_\epsilon \setminus i} \pi_\epsilon (y)\, \overline{h}_{i,m}(y) \leq k\epsilon^{-3d} \mathrm{vol}(V) \exp\left(-\frac{1}{\epsilon} \hat{F}_\epsilon \left(m, \mathcal{M}_\epsilon \setminus m\right)\right). \tag{4.3.11}$$

Therefore with Proposition 3.21 follows

$$E_i \left(\overline{\tau}_m\right) \leq c\epsilon^{-4d}. \tag{4.3.12}$$

With the same arguing as in Proposition 4.12, we obtain for the second summand of q_ϵ in (4.3.9):

$$P_m \left(\overline{\tau}_\rho < b_\epsilon\right) \quad \leq \quad \lfloor b_\epsilon \rfloor \exp\left(-\frac{1}{\epsilon} F_\rho\right). \tag{4.3.13}$$

Now we can proceed as Olivieri and Vares in the proof of their Lemma 4.11, p. 239 in [OV05]. That means we fix ϵ_0 such that $b_\epsilon / a_\epsilon \leq \frac{1}{2}\kappa$ and $\sqrt{b_\epsilon} \leq \frac{1}{4}\kappa$ as well as $q_\epsilon \leq \frac{1}{5}\kappa$ for all $\epsilon < \epsilon_0$.

If $\epsilon < \epsilon_0$ and $t \geq a_\epsilon$, due to$b_\epsilon / a_\epsilon \leq \frac{1}{2}\kappa$ we may write

$$P_i\left(\frac{1}{t}\sum_{k=1}^{t} 1\left(\overline{\xi}_k^\epsilon \in B\right) < 1-\kappa\right) \leq P_i\left(\frac{1}{k_\epsilon a_\epsilon}\sum_{k=1}^{k_\epsilon a_\epsilon} 1\left(\overline{\xi}_k^\epsilon \in B\right) < 1-\frac{\kappa}{2}\right) \tag{4.3.14}$$

where $k_\epsilon = \lfloor t/a_\epsilon \rfloor$.

For each $1 \leq k \leq k_\epsilon$ let us say that the time interval $[(k-1)\,a_\epsilon, ka_\epsilon)$ is good if the process $\overline{\xi}^\epsilon$ hits m before time $(k-1)\,a_\epsilon + \sqrt{a_\epsilon}$ and spends the rest of this time interval inside B. Otherwise, it is called bad. Let $Y_{\epsilon,k}$be the indicator function of the event $\{[(k-1)\,a_\epsilon, ka_\epsilon)$ is bad$\}$. Thus for any $i \in V_\epsilon$

$$\max_{k\in\{1,\ldots,k_\epsilon\}} P_i\left(Y_{\epsilon,k} = 1 \,\middle|\, Y_{\epsilon,1} = y_1, \ldots, Y_{\epsilon,k-1} = y_{k-1}\right) \leq q_\epsilon \tag{4.3.15}$$

for any choice of $y_1, \ldots, y_{k-1} \in \{0,1\}$. Since $q_\epsilon \leq \frac{1}{5}\kappa$, performing successive conditioning and applying (4.3.15) we obtain, for arbitrary $\lambda > 0$:

$$E_i\left(\exp\left(\lambda\sum_{k=1}^{k_\epsilon} Y_{\epsilon,k}\right)\right) \leq \left(1+\frac{\kappa}{5}\left(e^\lambda - 1\right)\right)^{k_\epsilon}. \tag{4.3.16}$$

Using (4.3.14) and (4.3.15) we see that

$$P_i\left(\overline{\mu}_t\left(B\right) < 1-\kappa\right) \leq P_i\left(\frac{1}{k_\epsilon}\sum_{k=1}^{k_\epsilon} Y_{\epsilon,k} \geq \frac{1}{4}\kappa\right) \leq e^{-k_\epsilon c_\kappa} \tag{4.3.17}$$

for all $\epsilon < \epsilon_0$, which implies the lemma. At the last inequality we have used the exponential Markov inequality and the preceding observation with $\lambda = \lambda\left(\kappa\right) > 0$ small enough such that

$$1+\frac{1}{5}\kappa\left(e^\lambda - 1\right) < e^{\frac{1}{4}\lambda\kappa}. \tag{4.3.18}$$

□

Now we can show

Theorem 4.14 *Let (ξ^ϵ) be a family of ergodic and reversible Markov chains. Let $\mathcal{M}_\epsilon$ be the set of local minima of F_ϵ. Assume that $\mathcal{M}_\epsilon$ is a finite set and $|\mathcal{M}_\epsilon|$ is independent of ϵ. Choose $m \in \mathcal{M}_\epsilon$ and let $V := V^{(\epsilon)}_{\mathcal{M}_\epsilon\backslash m}(m)$ be the valley of m with respect to $\mathcal{M}_\epsilon\backslash m$. Assume ξ^ϵ starts at m, then it thermalises at m.*

Proof. We proceed along the lines of the proof of Olivieri and Vares for thermalisation in the case of the Curie-Weiss model. We introduce the equilibrium measure *restricted to* V by setting for all $B \subset \Lambda_\epsilon$

$$\overline{\pi}_\epsilon(B) := \frac{\pi_\epsilon(B \cap V)}{\pi_\epsilon(V)}. \tag{4.3.19}$$

First we introduce of a restricted Markov chain, $\overline{\xi}^\epsilon$, that cannot leave the valley V. We determine $\overline{\xi}^\epsilon$ by defining its transition probability matrix

$$\overline{p}_\epsilon(x,y) := \begin{cases} p_\epsilon(x,y) & \text{if } x \in V, y \in \mathcal{N}_x \cap V \\ 1 - \sum_{z \in \mathcal{N}_x \cap V} p_\epsilon(x,z) & \text{if } y = x \in V \\ 0 & \text{else.} \end{cases} \tag{4.3.20}$$

The equilibrium measure for $\overline{\xi}^\epsilon$ is apparently $\overline{\pi}_\epsilon$. We introduce the following coupling between ξ^ϵ and $\overline{\xi}^\epsilon$: They both start in m and move together until for the first time ξ^ϵ jumps out of V. Recalling (4.3.20), at this step $\overline{\xi}^\epsilon$ remains at Σ and from then on they behave independently. Therefore $\overline{\tau}_\epsilon = \tau_\epsilon$ and $\overline{\xi}^\epsilon_t = \xi^\epsilon_t$ for all $t \le \tau_\epsilon$. Therefore the probability in (4.3.3) can be rewritten as

$$P_m\left(\sup_{s < \overline{\tau}_{\partial^+ V} - t_\epsilon} \overline{\mu}_{s,t_\epsilon}(B) > 1 - \kappa \text{ and } \overline{\tau}_{\partial^+ V} > t_\epsilon \right), \tag{4.3.21}$$

which is bounded from below by

$$1 - P_m(\overline{\tau}_{\partial^+ V} \le t_\epsilon) - P_m(G_\epsilon). \tag{4.3.22}$$

Here,

$$G_\epsilon := \bigcup_{l \in \{0,\dots,K_\epsilon\}} \left\{ \overline{\mu}_{lt_\epsilon,t_\epsilon}(B) > 1 - \frac{\kappa}{2} \text{ and } K_\epsilon \ge 1 \right\}, \tag{4.3.23}$$

where $K_\epsilon := \lfloor \overline{\tau}_{\partial^- V}/t_\epsilon \rfloor$.

Now, for every $k_\epsilon \in \mathbb{N}$ we obtain

$$\begin{aligned} P_m(G_\epsilon) &\le P_m(K_\epsilon \ge k_\epsilon) + k_\epsilon \sup_{i \in V} P_i\left(\overline{\mu}_{0,t_\epsilon}(B) > 1 - \frac{\kappa}{2}\right) \\ &\le P_m(\overline{\tau}_{\partial^+ V} \ge k_\epsilon t_\epsilon) + k_\epsilon \exp\left(-c_\kappa \sqrt{t_\epsilon}\right). \end{aligned} \tag{4.3.24}$$

For the last inequality we used Lemma 4.13. $\square$

5 Precise estimates for capacities and hitting times

5.1 Precise estimates of the capacity

We restrict ourselves now to the spaces $\Lambda_\epsilon \equiv \Lambda \cap \epsilon\mathbb{Z}^d$. Moreover, from now on we pose the following stronger assumption on the family of potentials (F_ϵ) that sharpens assumption **F1**, namely

sF1 We assume there exists functions $F_\epsilon : \Lambda \to \mathbb{R}_{>0}$ of class $C^3(\Lambda)$ such that

$$\pi_\epsilon(x) = e^{-F_\epsilon(x)/\epsilon} \qquad \forall x \in \Lambda_\epsilon. \tag{5.1.1}$$

and (F_ϵ) converges uniformly on Λ to a limiting function $F : \Lambda \to \mathbb{R}_{\geq 0}$ of class $C^3(\Lambda)$.

Let $\mathcal{M}$ be the set of local minima of F. A point s is called a *essential saddle point* if there exist minima $m, n \in \mathcal{M}$, such that $s \in S(m, n)$. The set of all essential saddle points will be denoted by $\mathcal{E}$. Analogously let $\mathcal{M}_\epsilon$ be the set of local minima of F_ϵ and $\mathcal{E}_\epsilon$ the set of essential saddle points of F_ϵ.

Remark 5.1 For all essential saddle points $s \in \mathcal{E}_\epsilon$ there exists $\hat{s} \in \Lambda$, such that $\nabla F_\epsilon(\hat{s}) = 0$ and $\|\hat{s} - s\|_2 < \epsilon$. Without restriction of generality for all $s \in \mathcal{E}_\epsilon$ we assume $\nabla F_\epsilon(s) = 0$, that is $s = \hat{s}$.

Definition 5.2 Let $f \in C^2(\mathbb{R}^d)$ be given. We call a critical point of f quadratic iff $\det \nabla^2 f \neq 0$. Otherwise it is called degenerate. A quadratic critical point of f, say x, is a k-saddle, iff $\nabla^2 f(x)$ has exactly k negative eigenvalues. We say, the function f is at x in k directions *unstable* and in $d-k$ directions *stable*.

To obtain precise estimates of the capacity and related quantities, we will now pose additional assumptions on the set of local minima, $\mathcal{M}_\epsilon$, of the family of potential $(F_\epsilon)_{\epsilon\in(0,1)}$.

We assume

S1 The functions F_ϵ and F have only finitely many critical points.

S2 All minima and all essential saddle points of F_ϵ and F are quadratic critical points. Moreover, $\nabla^2 F_\epsilon (x_\epsilon) \to \nabla^2 F(x)$ iff $x_\epsilon \to x$ for all $x_\epsilon \in \mathcal{M}_\epsilon \cup \mathcal{E}_\epsilon$.

S3 All metastable points and essential saddles are well in the interior of Λ, i.e. there exists $\kappa > 0$, such that for all $x \in \mathcal{M} \cup \mathcal{E}$ the distance to the boundary of Λ fulfills $\operatorname{dist}(x, \Lambda^c) > \kappa$.

Remark 5.3 (a) By enlarging the set Λ condition **S3** can be always satisfied.

(b) Condition **S2** implies that all essential saddle points are 1-saddles. In particular, it excludes situations prescribed in [MNOS04], Section 6.3. They give an example, where an unessential saddle point (with the same height) affects the prefactor of the capacity. This involves however essential saddle points s with $\det \nabla^2 F(s) = 0$.

We also need a to add another condition on the transition probability, p_ϵ, of the Markov chain ξ^ϵ, given in the form (4.0.1). We define $p_i(x) := p_\epsilon(x, x + \epsilon e_i)$ and $g_i^{(\epsilon)}(x) := g_\epsilon(x, x + \epsilon e_i)$. Since ξ^ϵ is a reversible Markov chain on a subset of the d-dimensional lattice with transitions only between nearest neighbours, all information are encoded in $(p_i(x) \mid x \in \Lambda_\epsilon, 1 \le i \le d)$.

C2 We assume $g_i^{(\epsilon)}$ is uniformly Lipschitz continuous on compact subsets of Λ_ϵ, i.e. for all $K \subset\subset \Lambda_\epsilon$ there exists a constant L independent of ϵ such that

$$\left| g_i^{(\epsilon)}(x) - g_i^{(\epsilon)}(y) \right| \le L \|x - y\|_2 \quad \text{for all } x, y \in K. \tag{5.1.2}$$

We will first consider the case of a unique relevant saddle point, called $s_\epsilon^*(m,n)$, between the metastable points $m, n \in \mathscr{M}_\epsilon$. As the treatment of the rough estimates indicates merely a neighbourhood of $s_\epsilon^*(m,n)$ contributes in leading order to the capacity between m and n.

We will use the parameter δ to measure the size of the neighbourhood of a relevant saddle point with vanishing gradient. We choose

$$\delta \equiv \delta(\epsilon) := \sqrt{k\epsilon |\ln \epsilon|} \tag{5.1.3}$$

and with $k \ge 4d$ constant. Whenever we use δ it will have this meaning.

The following lemma gives an approximation of the conductance matrix $C^{(\epsilon)}$ near a non degenerate critical point of F_ϵ.

Lemma 5.4 *Let s be a quadratic critical point of F_ϵ. Consider the ball $B_r(s)$ around s with radius $r = \mathcal{O}(\delta)$. Then for all $x \in B_r(s)$ we obtain*

$$C_i^{(\epsilon)}(x) = p_i(s)\,\pi_\epsilon(x)\,(1 + \mathcal{O}(\delta)). \tag{5.1.4}$$

Proof. By definition $C_i^{(\epsilon)}(x) = \pi_\epsilon(x)\,p_i(x)$. Since $x \in B_r(s)$ we have $\|x - s\|_2 = \mathcal{O}(\delta)$. Since $g_i^{(\epsilon)}$ is uniformly Lipschitz continuous and uniformly bounded by a constant from below on $B_r(s)$, we obtain

$$g_i^{(\epsilon)}(x) = g_i^{(s)}(s)\,(1 + \mathcal{O}(\delta)). \tag{5.1.5}$$

Since $F_\epsilon \in C^3(\Lambda)$ and $\nabla F_\epsilon(s) = 0$ we obtain for $x \in B_r(s)$ that

$$\frac{\pi_\epsilon(x + \epsilon e_i)}{\pi_\epsilon(x)} = 1 + \mathcal{O}(\delta). \tag{5.1.6}$$

Hence also $p_i(x) = p_i(s)\,(1 + \mathcal{O}(\delta))$ and the result follows. $\square$

As a lemma we show the continuous dependency of the capacity on (constant) boundary conditions.

Lemma 5.5 *Let $\Gamma = (Y, G)$ be a countable connected graph and $A, B \subset Y$ disjoint subsets. Let $a, b \in [0,1]$ with $a > b$. We define the function spaces*

$$\mathcal{H}_{A,B} := \{h \in l_2(\pi_\epsilon) \mid h|_A = 1 \text{ and } h|_B = 0\} \tag{5.1.7}$$

and

$$\widetilde{\mathcal{H}}_{A,B} := \{h \in l_2(\pi_\epsilon) \mid h|_A = a \text{ and } h|_B = b\}. \tag{5.1.8}$$

Define $\widetilde{\operatorname{cap}}(A,B) := \inf_{h \in \widetilde{\mathcal{H}}_{A,B}} \Phi(h)$. Then the minimiser $\tilde{h}_{A,B}$ is of the form $\tilde{h}_{A,B} = (a-b)\,h_{A,B} + b$.

Proof. $\tilde{h}_{A,B}$ fulfills the boundary value problem

$$\begin{cases} Lh = 0 & \text{on } Y \setminus (A \cup B), \\ h = a & \text{on } A, \\ h = b & \text{on } B. \end{cases} \tag{5.1.9}$$

Since this is a linear problem and $L1 = 0$, we are done. $\square$

Notation 5.6 For $v \in \mathbb{R}^d$ we define $v\cdot$ to be the diagonal matrix with entries $(v\cdot)_{ii} := v_i$.

Now we formulate the main theorem of this treatise. It gives a precise estimation of the capacity between two minima of F_ϵ. We formulate it here for the case of a unique relevant saddle point; for the case of several relevant saddles see Corollary 5.19.

Theorem 5.7 *Let ξ^ϵ be a reversible and ergodic Markov chain such that the assumptions at the beginning of this section are satisfied. Let $I, J \subset \mathcal{M}_\epsilon$ with $I \cap J = \emptyset$ and assume $s \equiv s^*_\epsilon(I, J)$ is the unique relevant saddle point between them. Then*

$$\begin{aligned} \operatorname{cap}(I,J) &= \left(\frac{2\pi}{\epsilon}\right)^{d/2-1} \frac{\hat{\lambda}_d}{\sqrt{|\det \nabla^2 F_\epsilon(s)|}} \exp\left(-F_\epsilon(s)/\epsilon\right) \times \\ &\quad \times \left(1 + \mathcal{O}\left(\sqrt{\epsilon}\,|\ln \epsilon|^{3/2}\right)\right), \end{aligned} \tag{5.1.10}$$

where $-\hat{\lambda}_d$ is the unique negative eigenvalue of the matrix given by

$$\left(p_i(s)\, \partial_i \partial_j F_\epsilon(s)\right). \tag{5.1.11}$$

To illustrate the general procedure we consider first the special case, where the orthonormal basis of eigenvectors $\{b_1, \ldots, b_d\}$ of $B \equiv \nabla^2 F_\epsilon(s)$ equals the canonical basis $\{e_1, \ldots, e_d\}$ of the lattice $\mathbb{Z}^d$, i.e. without loss of generality $b_i = e_i$ for all $i \in \{1, \ldots, d\}$. In this case the geometry of the lattice doesn't come into picture, because the process can take the direct way over the relevant saddle. This case can be treated in the same way as the problem for the function F in a continuous setting, compare [BEGK04]. Notice that in this case $\hat{\lambda}_d(\sigma) = p_d(s)\, \lambda_d$.

Without loss of generality we assume $s = 0$ and $\langle m, e_d \rangle < \langle n, e_d \rangle$.

The lower bound.

Denote

$$\delta_i := \left\lfloor \frac{1}{\epsilon} \frac{\delta}{\sqrt{(d-1)\,\lambda_i}} \right\rfloor \text{ for } 1 \le i < d \text{ and } \delta_d := \left\lfloor \frac{1}{\epsilon} \frac{1}{\sqrt{\lambda_d}} \delta \right\rfloor. \tag{5.1.12}$$

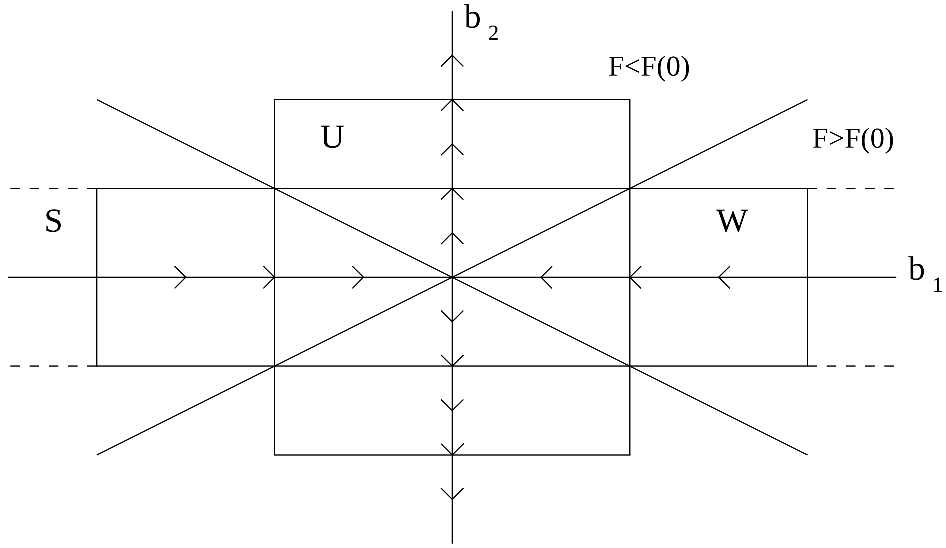

Figure 5.1.1: The different neighbourhoods of the saddle for the lower (U_δ) and upper (W_δ) bound.

We define index sets to designate the points of Λ_ϵ in a neighbourhood of zero:

$$R_\delta := \times_{i=1}^{d-1} \{-\delta_i, -\delta_i + 1, \ldots, \delta_i\} \tag{5.1.13}$$

and with a slight abuse of notation

$$2T_\delta := \{-2\delta_d, -2\delta_d + 1, \ldots, 2\delta_d - 1, 2\delta_d\}. \tag{5.1.14}$$

The associated neighbourhood of $s = 0$ is

$$U_\delta = \epsilon \left(R_\delta \times 2T_\delta\right) \tag{5.1.15}$$

We define the boundary toward m respectively n by

$$\partial_m U_\delta = \{(r, -2\delta_d) \mid r \in R_\delta\} \tag{5.1.16}$$

and $\partial_n U_\delta = \{(r, 2\delta_d) \mid r \in R_\delta\}$. U_δ is chosen in that way to secure that

$$F_\epsilon(x) < F_\epsilon(0) - \delta^2 \tag{5.1.17}$$

for $x \in \partial_m U_\delta \cup \partial_n U_\delta$.

We define for all $r \in R_\delta$ paths $\gamma_r : 2T_\delta \to U_\delta$ by

$$\gamma_r(t) := \epsilon t e_d + \epsilon \sum_{i=1}^{d-1} r_i e_i. \tag{5.1.18}$$

Let $(\gamma_r(2T_\delta), \gamma_r^*)$ be the one dimensional graph associated to γ_r with edges between nearest neighbours. Note that while all points of U_δ are hit by a path, only the edges parallel to e_d are included in these paths. To leave out some edges will only work in this case, because in general the process will use all edges inside a suitable defined neighbourhood of the relevant saddle.

We define the function spaces

$$\mathcal{H}_{U_\delta} := \{f : U_\delta \to [0,1] \mid f(z) = h_{n,m}(z) \text{ if } z \in \partial_m U_\delta \cup \partial_n U_\delta\} \tag{5.1.19}$$

and

$$\mathcal{H}_r := \{f : \gamma_r \to [0,1] \mid f(z) = h_{n,m}(z) \text{ if } z \in \{(r, -2\delta_d), (r, 2\delta_d)\}\}. \tag{5.1.20}$$

With the help of Lemma 5.5 and the representation (3.2.10) of the capacity of a one dimensional chain we obtain

$$\begin{aligned}
&\inf_{h \in \mathcal{H}_{n,m}} \Phi(h) = \Phi(h_{n,m}) \\
&\geq \Phi_{U_\delta}(h_{n,m}) \geq \inf_{h \in \mathcal{H}_{U_\delta}} \Phi_{U_\delta}(h) \\
&\geq \inf_{h \in \mathcal{H}_{U_\delta}} \sum_{r \in R_\delta} \Phi_{\gamma_r}(h) = \sum_{r \in R_\delta} \inf_{h \in \mathcal{H}_r} \Phi_{\gamma_r}(h) \\
&= \sum_{r \in R_\delta} \left(h_{n,m}(r, 2\delta_d) - h_{n,m}(r, -2\delta_d)\right)^2 \left(\frac{1}{2} \sum_{s \in \gamma_r^*} 1/C_s^{(\epsilon)}\right)^{-1}.
\end{aligned} \tag{5.1.21}$$

Now we use Corollary 4.11 and the inequality (5.1.17) to obtain a uniform bounds on the boundary. We obtain for $x \in \partial_m U_\delta$

$$\begin{aligned}
h_{n,m}(x) &\leq c\epsilon^{-2d} \exp\left(-\frac{1}{\epsilon}\left(\hat{F}_\epsilon(x,n) - \hat{F}_\epsilon(x,m)\right)\right) \\
&= c\epsilon^{-2d} e^{-\delta^2/\epsilon} = \mathcal{O}(\epsilon).
\end{aligned} \tag{5.1.22}$$

The last equation holds, since $\delta = \sqrt{k\epsilon |\ln \epsilon|}$ with $k \geq 3d$ large enough. For $x \in \partial_n U_\delta$ we obtain a uniform lower bound, namely

$$\begin{aligned}
h_{n,m}(x) &= 1 - h_{m,n}(x) \\
&= 1 - \mathcal{O}(\epsilon).
\end{aligned} \tag{5.1.23}$$

We denote by $\{\lambda_1, \ldots, \lambda_{d-1}, -\lambda_d\}$ the eigenvalues of the Hessian $\nabla^2 F_\epsilon(0)$. Since 0 is a 1-saddle of F_ϵ we can choose $\lambda_i > 0$ for all $1 \le i \le d$ and approximate F_ϵ inside U_δ by

$$F_\epsilon(x) = F_\epsilon(0) - \frac{1}{2}\lambda_d x_d^2 + \frac{1}{2}\sum_{i=1}^{d} \lambda_i x_i^2 + \mathcal{O}\left(\delta^3\right). \tag{5.1.24}$$

Therefore we conclude

$$\begin{aligned}
&\inf_{h\in\mathcal{H}_{n,m}} \Phi(h) \\
&\ge \sum_{r\in R_\delta} \left(\frac{1}{2}\sum_{s\in\gamma_r^*} 1/C_s^{(\epsilon)}\right)^{-1} (1+\mathcal{O}(\epsilon)) \\
&= \left(\sum_{r\in 2R_\delta} \exp\left(-\frac{1}{2}\epsilon\sum_{i=1}^{d-1}\lambda_i r_i^2\right)\right)\left(\frac{1}{2}\sum_{s\in\gamma_0^*} 1/C_s^{(\epsilon)}\right)^{-1} \times \\
&\quad \times \left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right).
\end{aligned} \tag{5.1.25}$$

The last equation uses Lemma 5.4. As we will see in the estimation of the upper bound this inequality is enough to match the associated upper bound up to multiplicative errors $(1+\mathcal{O}(\delta))$.

To evaluate these sums we use the quadratic approximation of F_ϵ inside of U_δ

Then we use Lemma 5.4 and obtain

$$C^{(\epsilon)}_{\gamma(t),\gamma(t+1)} = p_d(0)\, e^{-F_\epsilon(0)/\epsilon} e^{\frac{1}{2}\epsilon\lambda_d t^2}\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right). \tag{5.1.26}$$

The resulting Gaussian sums can be approximated by integrals (see Proposition A.1 in the Appendix). First we consider the sum over γ_0^* and obtain

$$\begin{aligned}
&\sum_{t=-2\delta_d}^{2\delta_d-1} \exp\left(-\frac{1}{2}\epsilon\lambda_d t^2\right) \\
&= \int_{-2\delta_d}^{2\delta_d-1} e^{-\frac{1}{2}\epsilon\lambda_d t^2}\, dt\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right) \\
&= \frac{2}{\sqrt{\epsilon\lambda_d}}\int_0^{2\sqrt{k|\ln\epsilon|}} e^{-\frac{1}{2}x^2}\, dx\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right) \\
&= \sqrt{\frac{2\pi}{\epsilon\lambda_d}}\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right).
\end{aligned} \tag{5.1.27}$$

The sum over $2R_\delta$ can be approximated by an $(d-1)$-dimensional Gaussian sum (see Proposition A.1 in the Appendix):

$$
\begin{aligned}
&\sum_{r\in R_\delta} \exp\left(-\frac{1}{2}\epsilon\sum_{i=1}^{d-1}\lambda_i r_i^2\right) \\
&= \int_{-\delta_{d-1}}^{\delta_{d-1}} \cdots \int_{-\delta_1}^{\delta_1} e^{-\frac{1}{2}\epsilon\sum_i \lambda_i r_i^2} dr_1 \dots dr_{d-1} \left(1+\mathcal{O}\left(\delta\right)\right) \\
&= 2^{d-1}\prod_{i=1}^{d-1}\int_0^{\delta_i} e^{-\frac{1}{2}\epsilon\lambda_i r_i^2} dr_i \left(1+\mathcal{O}\left(\delta\right)\right) \\
&= \prod_{i=1}^{d-1}\sqrt{\frac{2\pi}{\epsilon\lambda_i}}\left(1+\mathcal{O}\left(\delta\right)\right)
\end{aligned}
\tag{5.1.28}
$$

with the same transformation as before. Putting the pieces together we obtain

$$
\begin{aligned}
&\inf_{h\in\mathcal{H}_{n,m}} \Phi\left(h\right) \\
&= \left(\frac{2\pi}{\epsilon}\right)^{d/2-1} \frac{p_d\left(0\right)\lambda_d}{\sqrt{\lambda_1\lambda_2\dots\lambda_d}} e^{-F(0)/\epsilon}\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right).
\end{aligned}
\tag{5.1.29}
$$

The upper bound.

To prove an upper bound use δ_i from (5.1.12) and define

$$2R_\delta := \times_{i=1}^{d-1}\left\{-2\delta_i, -2\delta_i+1, \dots, 2\delta_i-1, 2\delta_i\right\} \tag{5.1.30}$$

and

$$T_\delta = \left\{-\delta_d, \dots, \delta_d\right\}. \tag{5.1.31}$$

Using these sets we put

$$W_\delta := \epsilon\left(2R_\delta \times T_\delta\right) \tag{5.1.32}$$

and

$$\partial_m W_\delta := \epsilon\left(2R_\delta \times \left\{-\delta_d\right\}\right) \text{ and } \partial_n W_\delta := \epsilon\left(2R_\delta \times \left\{\delta_d\right\}\right). \tag{5.1.33}$$

The remaining part of the inner boundary of W_δ is called the *central boundary* $\partial_c W_\delta$, i.e.

$$\partial_c W_\delta := \partial^- W_\delta \setminus \left(\partial_m W_\delta \cup \partial_n W_\delta\right). \tag{5.1.34}$$

The neighbourhood W_δ is chosen to secure that

$$F_\epsilon(x) > F_\epsilon(0) + \delta^2 \tag{5.1.35}$$

for all $x \in \partial_c W_\delta$.

We define $\tilde{D}_m$ as the connected component of

$$\left\{x \in \Lambda_\epsilon \,|\, F_\epsilon(x) \leq F_\epsilon(0) + \delta^2\right\} \tag{5.1.36}$$

that contains m. Define $D_m := \tilde{D}_m \setminus W_\delta$ and $D_n := \tilde{D}_m^c \setminus W_\delta$.

Now we choose a function h^+ to our convenience. We make the choice:

$$h^+|_{D_m} = 0, \qquad h^+|_{D_n} = 1. \tag{5.1.37}$$

By definition for all $z \in W_\delta$ there exist a unique $r \in 2R_\delta$ and $t \in T_\delta$ such that

$$z = \gamma_r(t) := \epsilon\left(\sum_{i=1}^{d-1} r_i e_i + t e_d\right). \tag{5.1.38}$$

Given this, we define on W_δ

$$h^+(\gamma_r(t)) \;:=\; \left(\sum_{k=-\delta_d}^{t-1} 1/C^{(\epsilon)}_{\gamma_0(k),\gamma_0(k+1)}\right) \Big/ \left(\frac{1}{2}\sum_{s\in\gamma_0^*} 1/C^{(\epsilon)}_s\right). \tag{5.1.39}$$

Observe that this does not depend on r.

We denote $\Sigma^- := \partial^+ D_m \setminus W_\delta^+$ and $\Sigma^+ := \partial^- D_m \setminus W_\delta^+$. Inserting h^+ into the Dirichlet form, we obtain

$$\begin{aligned}\Phi\left(h^+\right) &= \Phi_{W_\delta}\left(h^+\right) + \sum_{x\in\Sigma^-}\sum_{y\in\Sigma^+} C^{(\epsilon)}_{x,y} + \\ &\quad + \sum_{x\in\partial^- W_\delta}\sum_{y\in\partial^+ W_\delta} C^{(\epsilon)}_{x,y}\left(h^+(x) - h^+(y)\right)^2.\end{aligned} \tag{5.1.40}$$

Since we are in the case of discrete time, we have $C^{(\epsilon)}_{x,y} \leq \pi_\epsilon(x) \wedge \pi_\epsilon(y)$. Therefore

$$\sum_{x\in\Sigma^-}\sum_{y\in\Sigma^+} C^{(\epsilon)}_{x,y} \leq d\left|\Sigma^-\right| \exp\left(-\frac{1}{\epsilon}\left(F_\epsilon(s) + \delta^2\right)\right) \tag{5.1.41}$$

and

$$\sum_{x\in\partial^- W_\delta}\sum_{y\in\partial^+ W_\delta} C^{(\epsilon)}_{x,y}\left(h^+(x)-h^+(y)\right)^2 \le \left|\partial_c W_\delta\right| \exp\left(-\frac{1}{\epsilon}\left(F_\epsilon(s)+\delta^2\right)\right), \tag{5.1.42}$$

because $h^+=0$ on $\partial_m W_\delta$ and $h^+=1$ on $\partial_n W_\delta$.

With the help of Lemma 5.4 the first term can be estimated as

$$\begin{aligned}
\Phi_{W_\delta}\left(h^+\right) &= \left(\sum_{r\in 2R_\delta}\sum_{t=-\delta_d}^{\delta_d-1} C^{(\epsilon)}_{\gamma_r(t),\gamma_r(t+1)}\left(C^{(\epsilon)}_{\gamma_0(t),\gamma_0(t+1)}\right)^{-2}\right)\times \\
&\quad \times\left(\frac{1}{2}\sum_{s\in\gamma_0^*} 1/C^{(\epsilon)}_s\right)^{-2} \\
&= \left(\sum_{r\in 2R_\delta} e^{-\frac{1}{2}\epsilon\sum_{i=1}^{d-1}\lambda_i r_i^2}\right)\left(\frac{1}{2}\sum_{s\in\gamma_0^*} 1/C^{(\epsilon)}_s\right)^{-1}\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right) \\
&= \inf_{h\in\mathcal{H}_{A,B}} \Phi(h)\left(1+\mathcal{O}(\delta)\right).
\end{aligned} \tag{5.1.43}$$

Since $\delta=\sqrt{k\epsilon\left|\ln\epsilon\right|}$ with $k\ge 2d$, the quantities in (5.1.41) and (5.1.42) are by a factor ϵ^d smaller than the leading term.

Remark 5.8 Observe that, provided we have good a priori bounds, we only need one property of the conductance matrix $C^{(\epsilon)}$ to get matching upper and lower bounds (with multiplicative error tending to one), namely the existence of functions A_ϵ and B_ϵ such that

$$C^{(\epsilon)}_{\gamma_r(t),\gamma_r(t+1)} = A_\epsilon(t)\,B_\epsilon(r)\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right). \tag{5.1.44}$$

This means that we need approximately a separation of variables around the relevant saddle.

5.1.1 An associated inverse problem

To prove Theorem 5.7 in the general case, we will now formulate a corresponding inverse problem.

Definition 5.9 Let $\Gamma = (Y, G)$ be a locally finite graph with positive symmetric weights $C : G \to \mathbb{R}_{>0}$ on its edges. Let R be some index set and consider for $r \in R$ connected subgraphs $\eta_r = (Y_r, G_r)$ of Γ with positive symmetric weights $\tilde{C}_r : G_r \to \mathbb{R}_{>0}$. For convenience we put $\tilde{C}_r|_{G \setminus G_r} \equiv 0$. The family of $\left\{ \left(\eta_r, \tilde{C}_r \right) \mid r \in R \right\}$ is called a "partition" of (Γ, C), if

$$\sum_{r \in R} \tilde{C}_r (s) = C_s \text{ for all } s \in G. \tag{5.1.45}$$

Note that the Γ_r need not be disjoint.

There are of course very many ways of partitioning a given weighted graph, but as we will see in the next Proposition, given the equilibrium potential $h_{A,B}$ there exists particular useful partitions.

Let $\left(\eta_r, \tilde{C}_r \right)_{r \in R}$ be a partition of (Γ, C). We denote

$$\tilde{\Phi}_r (h) := \frac{1}{2} \sum_{s \in G_r} \tilde{C}_r (s) \left(h(s_2) - h(s_1) \right)^2, \tag{5.1.46}$$

the Dirichlet form on η_r.

Proposition 5.10 *Assume $\left(\eta_r, \tilde{C}_r \right)$ is a partition of (Γ, C) that connects A and B, i.e.*

$$|\eta_r \cap A| = |\eta_r \cap B| = 1 \tag{5.1.47}$$

for all $r \in R$. Then

$$\operatorname{cap}(A, B) \geq \sum_{r \in R} \inf_{h \in \mathcal{H}_{A,B}} \tilde{\Phi}_r (h). \tag{5.1.48}$$

If $\tilde{C}_r$ satisfies additionally Kirchhoff's node law at each "node" $x \in \eta_r$ with voltage $h_{A,B}$, i.e. if

$$\sum_{y \in \mathcal{N}_x} \tilde{C}_r (x, y) \left(h_{A,B}(x) - h_{A,B}(y) \right) = 0. \tag{5.1.49}$$

then we obtain

$$\operatorname{cap}(A, B) = \sum_{r \in R} \inf_{h \in \mathcal{H}_{A,B}} \tilde{\Phi}_r (h). \tag{5.1.50}$$

Proof. Notice that due to condition (5.1.45) of Definition 5.9

$$\operatorname{cap}(A,B) \quad = \quad \inf_{h\in\mathcal{H}_{A,B}} \sum_{r\in R} \tilde{\Phi}_r(h)\,. \tag{5.1.51}$$

This implies the inequality (5.1.48), because we are taking the infimum over the larger class of functions

$$\tilde{\mathcal{H}}_{A,B} = \left\{h : \{Y_r\}_{r\in R} \to [0,1] \mid h|_A = 1,\, h|_B = 0\right\}. \tag{5.1.52}$$

To prove equation 5.1.50, denote by $\tilde{h}_r : Y_r \to [0,1]$ the minimiser of $\tilde{\Phi}_r$. The infimum and the sum in (5.1.51) can obviously be exchanged if $\tilde{h}_r = h_{A,B}|_{\eta_r}$. The variational problem (5.1.51) is equivalent to the linear problem (3.1.4) with generator $L : G \to [0,1]$ given by $L(x,y) := 1 - C_{xy}/\sum_{z\in\mathcal{N}_x} C_{xz}$. Thus we obtain

$$\begin{cases} \sum_{z\in\mathcal{N}_x} C_{xz}\left(h(z) - h(x)\right) = 0 & \text{for } x \in \Gamma \setminus (A\cup B)\,, \\ h(x) = 1 & \text{for } x \in A, \\ h(x) = 0 & \text{for } x \in B. \end{cases} \tag{5.1.53}$$

But this means, that the capacity of A and B is given by (5.1.50), iff the conductance matrices $\tilde{C}_r$ satisfies the Kirchhoff law for $h_{A,B}$, i.e.

$$\sum_{y\in\mathcal{N}_x} \tilde{C}_r(x,y)\left(h_{A,B}(y) - h_{A,B}(x)\right) = 0 \text{ for } x \in Y_r \setminus (A\cup B)\,. \tag{5.1.54}$$

□

Solution to the inverse flat problem

Let $p = (p_1, \ldots, p_d) \in \mathbb{R}^d_{>0}$ be given. We denote by $q \in \mathbb{R}^d_{>0}$ the vector with components $q_i := 1/\sqrt{p_i}$ and by Q the associated diagonal matrix with entries $Q_{ii} = q_i$.

We consider the electrical network that consists of the lattice $Y_q := \times_{i=1}^d (q_i\mathbb{Z})$ with edges between pairs of nearest neighbours and the constant conductance matrice C given by

$$C_i(x) \equiv C(x, x+q_ie_i) \equiv p_i. \tag{5.1.55}$$

For this special choice the equilibrium potential $h_{0,a} : Y_q \to [0,1]$ is of the simple form $h_{0,a}(x) = \frac{\langle a,x \rangle}{\|a\|^2}$ for any given direction $a \in \mathbb{R}^d_{>0}$.

We consider only those a such that $v \equiv Q^{-1}a \in \mathbb{Z}^d$ and

$$v_d \geq 1 \text{ and } \gcd(v_1, \ldots, v_d) = 1. \tag{5.1.56}$$

Under these assumptions we can construct explicitly a partition of (Y_q, p).

Definition 5.11 For simplicity we denote for any negative integer t the set $\{t, \ldots, 0\}$ by $\{0, \ldots, t\}$.

(a) Let $v = (v_1, \ldots, v_d) \in \mathbb{Z}^d$ with properties (5.1.56). Define the *element* $E \subset \Gamma$ *of size* v *and spacing* q by

$$E := \times_{i=1}^{d-1} \{0, q_i, 2q_i \ldots, v_i q_i\} \times \{q_d, 2q_d \ldots, v_d q_d\} \cup \{0\} \tag{5.1.57}$$

and identify E with the graph with edges between nearest neighbours $x, y \in E$.

(b) We want to define a family $(E_{r,t})$ for all $r, t \in \mathbb{Z}$. For $t \in \mathbb{Z}$ we define the translated set $E_{0,t}$ by

$$E_{0,t} := E + ta. \tag{5.1.58}$$

To define the elements $E_{r,0}$ we need to be more careful: Let H_a be the hyperplane orthogonal to a, that contains the origin. The elements $E_{r,0}$ should as good as possible start from the hyperplane H_a. Hence we put for $r \in \mathbb{Z}^{d-1}$

$$E_{r,0} := E_{0,t^*} + \sum_{i=1}^{d-1} r_i q_i e_i \tag{5.1.59}$$

with t^* such that the intersection of $E_{r,0}$ with H_a is non-empty. In the special case, where $x \in \mathbb{Z}^d \cap H_a$, there are possibly two elements $E_{0,t}$ and $E_{0,t+1}$ hitting that point. In this case we will choose the lower one.

We now define

$$E_{r,t} := E_{r,0} + ta. \tag{5.1.60}$$

(c) We define strings of elements by putting

$$Y_r := \bigcup_{t \in \mathbb{Z}} E_{r,t}. \tag{5.1.61}$$

Let η_r be the connected graph with vertexset Y_r and edges between nearest neighbours.

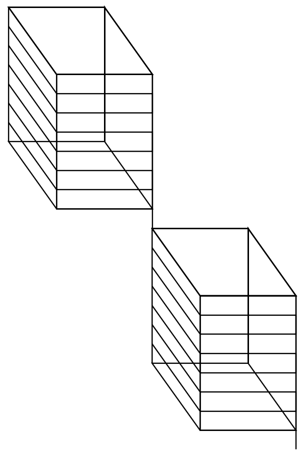

Figure 5.1.2: Two connected elements.

Remark 5.12 The construction of the family $(E_{r,t})$ implies that there exist shifts $s_{r,t} : Y \to Y$ such that $E_{r,t} = s_{r,t}(E)$.

Since the element $E_{t,r}$ is a translation of E and the weightfunctions C_i are constant on Λ_ϵ, it is enough to find a conductance matrix $\tilde{C}$ that satisfies the following equations: First the Kirchhoff equation

$$\sum_{y \in E} \tilde{C}(x,y)\left(h_{0,a}(y) - h_{0,a}(x)\right) = 0 \quad \text{for } x \in E \setminus \{0, a\} \tag{5.1.62}$$

and for all $i \in \{0, \ldots, d\}$ and $k \in \{0, \ldots, v_d\}$ the consistency condition:

$$\sum_{x : x_d = kq_d} \tilde{C}(x, x + q_i e_i) = p_i. \tag{5.1.63}$$

Observe that these conditions does not determine a unique conductance matrix.

We define the associated current, I, by Ohm's law, i.e.

$$I(x,y) := \tilde{C}(x,y)\left(h_{0,a}(y) - h_{0,a}(x)\right). \tag{5.1.64}$$

Then the two conditions (5.1.62) and (5.1.63) read respectively:

$$\sum_{y \in E} I(x,y) = 0 \tag{5.1.65}$$

$$\sum_{x : x_d = kq_d} I(x, x + q_i e_i) = \frac{v_i}{\|a\|^2}. \tag{5.1.66}$$

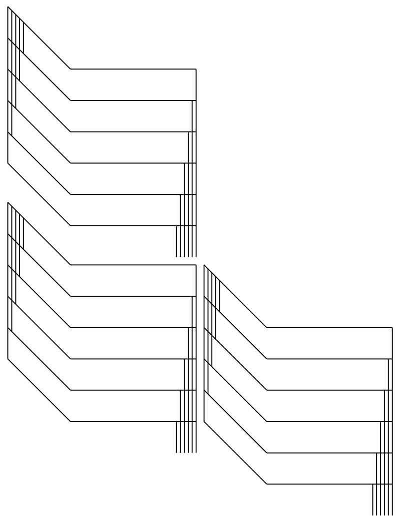

Figure 5.1.3: The picture represents the current flow on three selected elements in $\mathbb{Z}^3$, 6 lines symbolises one unit of the flow.

Proposition 5.13 *Let E be an element of size v and spacing q and $h_{0,a}(x) = \frac{\langle a,x\rangle}{\|a\|^2}$. Then a conductance matrix $\tilde{C}$ that satisfies conditions (5.1.62) and (5.1.63) is given by*

$$\tilde{C}(x, x + q_i e_i) := p_i \varphi_i(x) \tag{5.1.67}$$

for all $x, x + q_i e_i \in E$.

The function $\varphi : E \times \{1, \ldots, d\} \to [0,1]$ has the form

$$\varphi_i(x) := \begin{cases} 1 - x_d/a_d & \text{for } x = (0, 0, \ldots, 0, x_d), i = d, \\ 1/v_i & \text{for } x = (a_1, \ldots, a_{i-1}, x_i, 0, \ldots, 0, x_d), i < d, \\ x_d/a_d & \text{for } x = (a_1, \ldots, a_{d-1}, x_d), i = d, \\ 0 & \text{else}, \end{cases} \tag{5.1.68}$$

for all $(x, i) \in E \times \{1, \ldots, d\}$ such that $x + q_i e_i \in E$. Otherwise $\varphi_i(x) := 0$.

Proof. Insert $\tilde{C}$ into (5.1.62) and (5.1.63). □

Now we define flows for shifted elements.

Definition 5.14 Let $(E_{r,t}) = (s_{r,t}(E))$ be a family of translated elements and $\tilde{C} : E \times \{1, \ldots, d\} \to [0,1]$ as in Proposition 5.13.

(a) Define $\tilde{C}^{(r,t)} : E_{r,t} \times \{1, \ldots, d\} \rightarrow [0,1]$ by putting $\tilde{C}_i^{(r,t)}(x) := \tilde{C}_i\left(s_{r,t}^{-1}(x)\right)$.

(b) We define the capacity of an element by

$$\operatorname{cap}\left(E, \tilde{C}\right) := \inf_{h \in \mathcal{H}_{0,a}} \sum_{(x,y) \in E^*} \tilde{C}(x,y)(h(x) - h(y))^2, \tag{5.1.69}$$

where E^* is the edgeset of E. Analogously $\operatorname{cap}\left(E_{r,t}, \tilde{C}\right)$ is defined with the help of $s_{r,t}$.

(b) Let $n\left(E, \tilde{C}\right)$ be the average number of strings inside a unit volume on the hyperplane H_a (perpendicular to a).

Proposition 5.15 *Assume $C_i(x) \equiv p_i$, and $\tilde{C}_i(x) = p_i \varphi_i(x)$ as in Proposition 5.13. Then one element has the capacity*

$$\operatorname{cap}\left(E, \tilde{C}\right) = \frac{v_d}{\|a\|^2}, \tag{5.1.70}$$

and the average number of strings is

$$n\left(E, \tilde{C}\right) = \frac{\|a\|}{v_d \det Q}. \tag{5.1.71}$$

Proof. Using Definition (5.1.64) we obtain

$$\begin{aligned} \operatorname{cap}\left(E, \tilde{C}\right) &= \sum_{x,y \in E} \tilde{C}(x,y)(h_{0,a}(y) - h_{0,a}(x))^2 \\ &= \sum_{x,y \in E} I(x,y)(h_{0,a}(y) - h_{0,a}(x)). \end{aligned} \tag{5.1.72}$$

Notice that I is a flow in the sense of Doyle and Snell. Because of the conservation of energy principle (see [DS84], section 3.5, page 61) it follows

$$\operatorname{cap}\left(E, \tilde{C}\right) = I_0(h_{0,a}(a) - h_{0,a}(0)) = I_0,$$

where $I_0 \equiv \sum_{y \in E} I(0,y)$. Due to the geomtry of an element we obtain

$$\operatorname{cap}\left(E, \tilde{C}\right) = I(0, q_d e_d) = \frac{v_d}{\|a\|^2}. \tag{5.1.73}$$

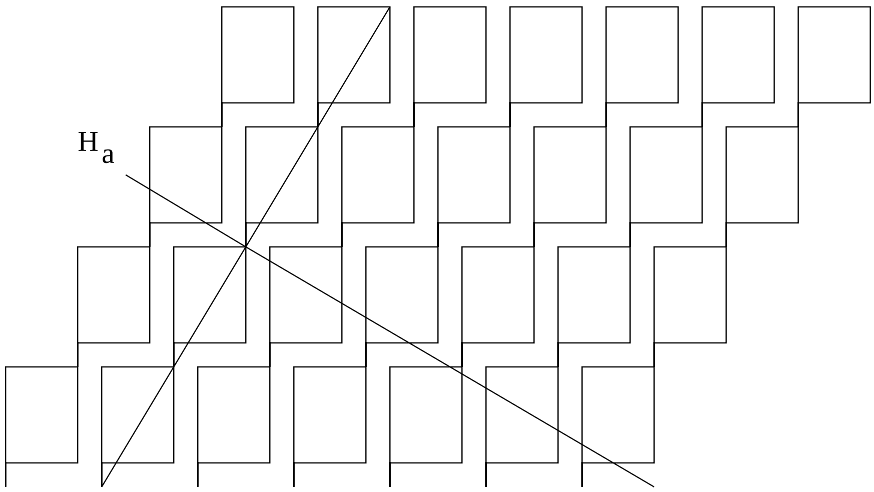

Figure 5.1.4: The picture represents merely every fourth string of elements.

Now we calculate the average number of strings $n\left(E,\tilde{C}\right)$. All edges are completely occupied by the elements, since $\sum_{r\in R}\varphi_i^{(r,t)}(x)=1$. A single element E uses the fraction $\varphi_i(x)$ of an edge $(x,x+q_ie_i)$, hence we obtain

$$\begin{aligned}
&\sum_{x\in E}\sum_{i=1}^{d}\varphi_i(x)\\
&=\varphi_d(0)+\sum_{k=1}^{v_d}\sum_{i=1}^{d-1}\sum_{n=0}^{v_i-1}\varphi_i(a_1,\ldots,a_{i-1},nq_i,0,\ldots,0,kq_d)+\\
&\qquad+\sum_{n=1}^{v_d-1}\varphi_d(0,\ldots,0,nq_d)+\sum_{n=1}^{v_d-1}\varphi_d(v_1,\ldots,v_{d-1},nq_d)\\
&=1+v_d\sum_{i=1}^{d-1}v_i\frac{1}{v_i}+\sum_{n=1}^{v_d-1}\left(1-\frac{n}{v_d}\right)+\sum_{n=1}^{v_d-1}\frac{n}{v_d}\\
&=1+(d-1)\,v_d+(v_d-1)=dv_d
\end{aligned}\tag{5.1.74}$$

edges. Notice that a half open cube in Λ_q contains d edges and has volume $\det Q$. Hence the effective volume of an element is $v_d\det Q$. Since the length of an element in direction a is $\|a\|$ we obtain

$$n\left(E,\tilde{C}\right)=\frac{\|a\|}{v_d\det Q}.\tag{5.1.75}$$

□

Remark 5.16 Since $n(E, \varphi)$ is the average number of strings inside a unite d-cube, we can rearrange the strings by small perturbations to have starting points inside the hyperplane H_a orthogonal to a on a cubic lattice with side length s such that $s^{(d-1)} = 1/n(E, \varphi)$.

5.1.2 General proof of Theorem 5.7

The case of a unique relevant saddle

Proof. Without loss of generality we assume $s^*(m, n) = 0$.

Let $\{\lambda_1, \ldots, \lambda_{d-1}, -\lambda_d\}$ be the eigenvalues of $\nabla^2 F_\epsilon(0)$ and $\{b_k\}_{1 \le k \le d}$ an orthonormal basis of eigenvectors, such that b_d belongs to the unique negative eigenvalue $-\lambda_d$ and $\langle m, b_d \rangle < 0 < \langle n, b_d \rangle$. If b_d coincides with a lattice direction, say e_d, the proof is simply is discrete version of the proof of Theorem 5.1 in [BEGK04]. But in the general case we have to use the partition of the last subsection.

Let $q \in \mathbb{R}^d$ the vector with components

$$q_i := \frac{1}{\sqrt{p_i(0)}} \tag{5.1.76}$$

and Q be the diagonal matrix with entries $Q_{ii} \equiv q_i$. We denote

$$\Lambda_q := \Lambda \cap (\times_{i=1}^d (q_i \mathbb{Z})). \tag{5.1.77}$$

During this proof we associate to a given function $f_\epsilon : \Lambda_\epsilon \to \mathbb{R}$, the transformed function $\overline{f}_\epsilon : \Lambda_q \to \mathbb{R}$ by defining

$$\overline{f_\epsilon} := f_\epsilon \circ \epsilon Q^{-1}. \tag{5.1.78}$$

We denote by $\left\{\hat{\lambda}_1, \ldots, \hat{\lambda}_{d-1}, -\hat{\lambda}_d\right\}$ the eigenvalues of

$$B_\epsilon := Q^{-1} \nabla^2 F_\epsilon(0)\, Q^{-1}. \tag{5.1.79}$$

Since all eigenvalues of Q are positive, we can choose $\hat{\lambda}_i > 0$ for all $i \in \{1, \ldots, d\}$. Let $\{w_1, \ldots, w_d\}$ be an orthonormal basis of eigenvectors of B_ϵ, such that w_d corresponds to the negative eigenvalue $-\hat{\lambda}_d$ and $\langle w_d, b_d \rangle > 0$.

We will see, that w_d points in the direction, in which the equilibrium potential rises.

Assume first $Q^{-1}w_d \in \mathbb{Q}^d$. Choose $a \parallel w_d$ with $\langle a, w_d \rangle > 0$ such that $v := Q^{-1}a \in \mathbb{Z}^d$ and $\gcd(v_1, \ldots, v_d) = 1$. Without loss of generality $v_d \geq 1$.

We transform the Dirichlet form by a substitution $y = \frac{1}{\epsilon}Qx$:

$$\begin{aligned}
\Phi(h) &= \sum_{x \in \Lambda_\epsilon} \sum_{i=1}^{d} C_{\epsilon,i}(x)\left(h(x+\epsilon e_i) - h(x)\right)^2 \\
&= \sum_{y \in \Lambda_\sigma} \sum_{i=1}^{d} \overline{C}_{\epsilon,i}(y)\left(\overline{h}(y+q_i e_i) - \overline{h}(y)\right)^2 \\
&=: \overline{\Phi}\left(\overline{h}\right).
\end{aligned} \tag{5.1.80}$$

We will use the parameter δ to measure the size of the neighbourhood the relevant saddle point. We choose

$$\delta \equiv \delta(\epsilon) := \sqrt{k\epsilon\,|\ln \epsilon|} \tag{5.1.81}$$

where $k \geq 3d$ constant.

The lower bound

We define the following neighbourhood of the saddle point:

$$U_\delta := \left\{ z \in \Lambda_\sigma \mid |\langle z, w_i \rangle| \leq \frac{\delta}{\epsilon\sqrt{\hat{\lambda}_i}}, |\langle z, w_d \rangle| \leq 2\frac{\delta}{\epsilon\sqrt{\hat{\lambda}_d}} \right\}. \tag{5.1.82}$$

We denote by $\partial_m U_\delta$ the face of the U_δ, that lies entirely in the valley $V_n(m)$ and analogously $\partial_n U_\delta$ the opposite face. We will use the space of functions

$$\mathcal{H}_{U_\delta} := \left\{ f : U_\delta \to [0,1] \mid f|_{\partial_n U_\delta \cup \partial_m U_\delta} = \overline{h}_{n,m} \right\}. \tag{5.1.83}$$

We obtain by cutting all edges outside the neighbourhood U_δ and then with the quadratic approximation (5.1.85)

$$\begin{aligned}
&\Phi(h_{n,m}) = \overline{\Phi}\left(\overline{h}_{n,m}\right) \\
&\geq \overline{\Phi}_{U_\delta}\left(\overline{h}_{n,m}\right) \geq \inf_{h \in \mathcal{H}_{U_\delta}} \overline{\Phi}_{U_\delta}(h).
\end{aligned} \tag{5.1.84}$$

Inside U_δ we can approximate $\overline{C}_{\epsilon,i}$ due to assumption **C1**, **C2** and **sF1** by

$$\overline{C}_{\epsilon,i}(y) = p_i(0)\exp\left(-\frac{1}{\epsilon}F_\epsilon(0) - \frac{\epsilon}{2}\langle y, B_\epsilon y\rangle\right)\left(1 + \mathcal{O}\left(\delta^3/\epsilon\right)\right). \tag{5.1.85}$$

Hence we have to investigate

$$k(\epsilon) \equiv \inf_{h\in\mathcal{H}_{U_\delta}} \sum_{y\in U_\delta} e^{-\frac{\epsilon}{2}\langle y, B_\epsilon y\rangle} \sum_{i=1}^{d} \left(\frac{h(y+q_i e_i) - h(y)}{q_i}\right)^2. \tag{5.1.86}$$

We abbreviate $f_\epsilon(y) := e^{-\frac{\epsilon}{2}\langle y, B_\epsilon y\rangle}$.

We use now a partition of $(U_\delta, p_i(0) f_\epsilon)$ with boundary sets $\partial_m U_\delta$ and $\partial_n U_\delta$ in the sense of Definition 5.9. This gives us in any case a lower bound as we noticed in (5.1.48). To obtain a good bound we choose the partition of the flat case and take as conductance matrix $\tilde{C}_{\epsilon,i}(x) := f_\epsilon(x)\,\tilde{c}_i(x)$. Here, $\tilde{c}_i(x) \equiv \varphi_i(x)\,p_i(0)$ and φ is given by 5.13. This gives us a good bound, because in the neighbourhood of the saddle point the potential f_ϵ is nearly flat.

Let $E \subset U_\delta$ be the elements of size v. We denote $\ell := n(E,\varphi)^{-1/(d-1)}$ and denote

$$\delta_i := \left\lfloor \frac{1}{\ell}\frac{\delta}{\sqrt{\lambda_i}} \right\rfloor \text{ for } 1 \le i < d \text{ and } \delta_d := \left\lfloor \frac{1}{\|a\|}\frac{\delta}{\sqrt{\hat{\lambda}_d}} \right\rfloor \tag{5.1.87}$$

as well as

$$R_\delta := \times_{i=1}^{d-1}\left\{-\delta_i + 1, -\delta_i + 2, \ldots, \delta_i - 2, \delta_i - 1\right\} \tag{5.1.88}$$

and

$$2T_\delta := \left\{-2\delta_d + 1, -2\delta_d + 2, \ldots, 2\delta_d - 2, 2\delta_d - 1\right\}. \tag{5.1.89}$$

We define $E_{r,t}$ like in Definition 5.11. The strings (η_r, G_r) with edges between nearest neighbours inside U_δ for r in a suitable neighbourhood of 0 are defined by

$$\eta_r := \bigcup_{t\in 2T_\delta} E_{r,t}. \tag{5.1.90}$$

Let $\check{\eta}_r := \{x \in \eta_r \mid \langle x, a\rangle = \min\}$ be the starting point and $\hat{\eta}_r := \{x \in \eta_r \mid \langle x, a\rangle = \max\}$ the endpoint of the rth string. Observe that

$\{\check{\eta}_r, \hat{\eta}_r\}_{r(\rho)} \subset U_\delta$. Furthermore define the sets $\hat{\eta} := \left\{\hat{\eta}_{r(\rho)} \,|\, \rho \in R_\delta\right\}$ and $\check{\eta} := \left\{\check{\eta}_{r(\rho)} \,|\, \rho \in R_\delta\right\}$. We denote the Dirichlet form of a single string by η_r

$$\overline{\Phi}_r(h) := \sum_{x\in\eta_r} f_\epsilon(x) \sum_{i=1}^{d} \tilde{c}_i(x) \left(h(x+q_i e_i) - h(x)\right)^2 \tag{5.1.91}$$

with φ defined in Proposition 5.13. We define the function space for a single string,

$$\mathcal{H}_r := \left\{h : \eta_r \to [0,1] \;\mid h(x) = h_{n,m}(x) \,\text{if}\, x \in \{\check{\eta}_r, \hat{\eta}_r\}\right\}. \tag{5.1.92}$$

Moreover, put

$$\check{h}_{n,m} := \sup\left\{\overline{h}_{n,m}(x) \;\mid x \in \check{\eta}\right\} \tag{5.1.93}$$

and $\hat{h}_{n,m} := \inf\left\{\overline{h}_{n,m}(x) \;\mid x \in \hat{\eta}\right\}$.

Proposition 5.10 yields

$$k(\epsilon) \geq \sum_{\rho\in R_\delta} \inf_{h\in\mathcal{H}_{r(\rho)}} \overline{\Phi}_{r(\rho)}(h).$$

Since we can calculate the capacity of a one dimensional chain, as in (3.2.10), we obtain with Lemma 5.5 that

$$\begin{aligned} k(\epsilon) &\geq \sum_{\rho\in R_\delta} \left(\overline{h}_{n,m}\left(\hat{\eta}_{r(\rho)}\right) - \overline{h}_{n,m}\left(\check{\eta}_{r(\rho)}\right)\right)^2 \times \\ &\qquad \times \left(\sum_{t\in 2T_\delta} \operatorname{cap}\left(E_{r,t}, f_\epsilon \tilde{c}\right)^{-1}\right)^{-1} \\ &\geq \operatorname{cap}(E, \tilde{c}) \left(\hat{h}_{n,m} - \check{h}_{n,m}\right)^2 \sum_{\rho\in R_\delta} \left(\sum_{t\in 2T_\delta} \max_{y\in E_{r(\rho),t}} f_\epsilon(y)^{-1}\right)^{-1} \end{aligned} \tag{5.1.94}$$

By construction of (η_r) we have for $y \in \check{\eta}$, using the definition of δ_i, (5.1.87),

$$\begin{aligned} \epsilon^2 \langle y, B_\epsilon y\rangle &= \epsilon^2 \sum_{i=1}^{d-1} \hat{\lambda}_i \langle y, w_i\rangle^2 - \epsilon^2 \hat{\lambda}_d \langle y, w_d\rangle^2 \\ &\leq \delta^2 - 4\delta^2 = -3\delta^2. \end{aligned} \tag{5.1.95}$$

Moreover, it holds

$$s^*(y,m) = y \text{ and } s^*(y,n) = 0.$$

Hence Proposition 4.11 implies for $y \in \check{\eta}$

$$\begin{aligned} h_{n,m}(y) &\le c\epsilon^{-2d} \exp\left(-\frac{1}{\epsilon}\left(\hat{F}_\epsilon(y,n) - \hat{F}_\epsilon(y,m)\right)\right) \\ &= c\epsilon^{-2d} e^{-3\delta^2/\epsilon}\left(1 + \mathcal{O}(\delta)\right) = \mathcal{O}(\epsilon). \end{aligned} \tag{5.1.96}$$

The last equation holds, since $\delta = \sqrt{k\epsilon\,|\ln\epsilon|}$ and $k \ge 3d$. For $y \in \hat{\eta}$ we obtain a uniform lower bound, namely

$$h_{n,m}(y) = 1 - h_{m,n}(y) = 1 + \mathcal{O}(\epsilon). \tag{5.1.97}$$

Altogether we obtain

$$\left(\hat{h}_{n,m} - \check{h}_{n,m}\right)^2 = 1 + \mathcal{O}(\epsilon). \tag{5.1.98}$$

Now we shift the strings η_r and rename them, such that $\tilde{E}_{\rho,0} = \tau\left(E_{r(\rho),t}\right)$ begins for all $\rho \in R_\delta$ at the point $\ell\sum_{i=1}^{d-1}\rho_i w_i$ in the hyperplane H_a orthogonal to a. The shifts τ can be chosen, such that their length is at most $\max\left\{\ell\sqrt{d}, \|a\|\right\}$. The starting points of elements in the ρth string can now be parametrised by

$$z_\rho(t) = ta + \ell\sum_{i=1}^{d-1}\rho_i w_i \tag{5.1.99}$$

for $t \in 2T_\delta$ and $\rho \in R_\delta$. Thus we have for $y \in E_{r(\rho),t}$

$$\left|\langle y, B_\epsilon y\rangle - \langle z_\rho(t), B_\epsilon z_\rho(t)\rangle\right| = \mathcal{O}(1)$$

Hence we obtain

$$\begin{aligned} k(\epsilon) &\ge \operatorname{cap}(E,\tilde{c}) \sum_{\rho\in R_\delta}\left(\sum_{t\in 2T_\delta} \exp\frac{\epsilon}{2}\langle z_\rho(t), B_\epsilon z_\rho(t)\rangle\right)^{-1} \times \\ &\quad \times\left(1 + \mathcal{O}(\epsilon)\right). \end{aligned} \tag{5.1.100}$$

By construction z_ρ lies parallel to a, and thus we can separate the sums in t and r from (5.1.100) and obtain:

$$\begin{aligned} k(\epsilon) &\ge \operatorname{cap}(E,\tilde{c})\left(\sum_{t\in 2T_\delta}\exp\left(-\frac{\epsilon}{2}\|a\|^2\hat{\lambda}_d t^2\right)\right)^{-1} \times \\ &\quad \times \sum_{\rho\in R_\delta}\exp\left(-\frac{\epsilon}{2}\ell^2\sum_{i=1}^{d-1}\hat{\lambda}_i\rho_i^2\right)\left(1 + \mathcal{O}(\epsilon)\right). \end{aligned} \tag{5.1.101}$$

We approximate the Gaussian sums of (5.1.100) with Gaussian integrals (see Appendix A). Hence we obtain

$$\sum_{t\in 2T_\delta} \exp\left(-\frac{\epsilon}{2}\|a\|^2 \hat{\lambda}_d t^2\right) = \frac{1}{\|a\|}\sqrt{\frac{2\pi}{\epsilon\hat{\lambda}_d}}\left(1+\mathcal{O}\left(\sqrt{\epsilon}\right)\right) \tag{5.1.102}$$

and

$$\sum_{\rho\in R_\delta} \exp\left(-\frac{\epsilon}{2}\ell^2\sum_{i=1}^{d-1}\hat{\lambda}_i\rho_i^2\right) = \frac{1}{\ell^{d-1}}\prod_{i=1}^{d-1}\sqrt{\frac{2\pi}{\epsilon\hat{\lambda}_i}}\left(1+\mathcal{O}\left(\sqrt{\epsilon}\right)\right). \tag{5.1.103}$$

The product can be evaluated by using, that $\{w_i\}_i$ is an orthonormal basis of eigenvectors of B:

$$\begin{aligned}\prod_{i=1}^{d-1}\hat{\lambda}_i &= \det\left(Q^{-1}\nabla^2 F_\epsilon(0)\,Q^{-1}\right)/\hat{\lambda}_d \\ &= (\det Q)^{-2}\det\nabla^2 F_\epsilon(0)/\hat{\lambda}_d.\end{aligned} \tag{5.1.104}$$

Inserting into (5.1.101), we obtain with Proposition 5.15:

$$\begin{aligned}k(\epsilon) &\geq \left(\frac{2\pi}{\epsilon}\right)^{d/2-1}\frac{\hat{\lambda}_d}{\sqrt{\det\nabla^2 F_\epsilon(0)}}\times \\ &\quad\times \operatorname{cap}(E,\tilde{c})\,n(E,\varphi)\,\|a\|\det Q\left(1+\mathcal{O}\left(\sqrt{\epsilon}\right)\right) \\ &= \left(\frac{2\pi}{\epsilon}\right)^{d/2-1}\frac{\hat{\lambda}_d}{\sqrt{|\det\nabla^2 F_\epsilon(0)|}}\left(1+\mathcal{O}\left(\sqrt{\epsilon}\right)\right).\end{aligned} \tag{5.1.105}$$

Observe that the eigenvalues of $B_\epsilon = Q^{-1}\nabla^2 F_\epsilon(0)\,Q^{-1}$ coincide with the eigenvalues of $Q^{-2}\left(\nabla^2 F_\epsilon(0)\right)$.

The upper bound.

We will directly use the transformed Dirichlet form $\overline{\Phi}$ of equation (5.1.80). We denote, using δ_i from (5.1.87),

$$2R_\delta := \times_{i=1}^{d-1}\{-2\delta_i, -2\delta_i+1, \ldots, 2\delta_i-1, 2\delta_i\} \tag{5.1.106}$$

and

$$T_\delta := \{-\delta_d, -\delta_d+1, \ldots, \delta_d-1, \delta_d\}. \tag{5.1.107}$$

Define now the neighbourhood $W_\delta \subset \Lambda_\sigma$ of the saddle point by

$$W_\delta := \left\{ z \in \Lambda_\sigma \mid |\langle z, w_i \rangle| \leq 2\frac{\delta}{\epsilon\sqrt{\hat{\lambda}_i}}, |\langle z, w_d \rangle| \leq \frac{\delta}{\epsilon\sqrt{\hat{\lambda}_d}} \right\} \tag{5.1.108}$$

and the slightly larger set

$$\hat{W}_\delta = \{ E_{r,t} \mid E_{r,t} \cap W_\delta \neq \emptyset \} . \tag{5.1.109}$$

The neighbourhood W_δ is chosen to secure that

$$\overline{F}_\epsilon(x) - \overline{F}_\epsilon(0) > \delta^2 \tag{5.1.110}$$

for $x \in \partial W_\delta \setminus (\partial_m W_\delta \cup \partial_n W_\delta)$.

We define $\tilde{D}_m$ as the connected component of

$$\left\{ x \in \Lambda_\epsilon \,|\, \overline{F}_\epsilon(x) \leq \overline{F}_\epsilon(0) + \delta^2 \right\} \tag{5.1.111}$$

that contains m. Define $D_m := \tilde{D}_m \setminus W_\delta$ and $D_n := \tilde{D}_m^c \setminus W_\delta$. To prove an upper bound we just choose a function h^+ to our convenience. We make the choice

$$h^+|_{D_m} = 0, \qquad h^+|_{D_n} = 1. \tag{5.1.112}$$

Up to now we didn't have to be very careful choosing h^+. But in a neighbourhood of the relevant saddle point of order $\mathcal{O}(\delta)$ we have to approximate the real equilibrium potential $h_{n,m}$ as good as possible. Surprisingly it suffices, to take h^+ constant on hyperplanes perpendicular to a. We take now a sum of resistances with value $1/\max_{\lambda\in[j,j+1)} \{\overline{\pi}_\epsilon(\lambda a)\}$ plus a term for the remainder.

We denote $f_\epsilon(x) \equiv \exp(-\epsilon \langle x, Bx \rangle /2)$ for $x \in \hat{W}_\delta$ and introduce as normalisation

$$N := \sum_{t=-\delta_d}^{\delta_d} \left(\max_{\lambda\in[t,t+1)} f_\epsilon(\lambda a) \right)^{-1} . \tag{5.1.113}$$

Denote the orthogonal projection onto the vector a with Pr_a, i.e. $\mathrm{Pr}_a = \langle a, \cdot \rangle \frac{a}{\|a\|^2}$. Denote $h_{0,a}(x) = \langle a, x \rangle / \|a\|^2$ and $\tilde{h}(x) = \left\lfloor \langle a, x \rangle / \|a\|^2 \right\rfloor$. For $x \in W_\delta$ we choose:

$$h^+(x) := \\ \frac{1}{N} \left(\frac{h_{0,a}(x) - \tilde{h}(x)}{f_\epsilon(Pr_a x)} + \sum_{j=-\delta_d}^{\tilde{h}(x)} \min_{\lambda\in[j,j+1)} \left(\frac{1}{f_\epsilon(\lambda a)} \right) \right) \tag{5.1.114}$$

We estimate now differences of h^+ between nearest neighbours: Let $i \in \{1, \dots, d\}$, then we obtain for $x \in W_\delta$ and if $\tilde{h}(x + q_i e_i) = \tilde{h}(x)$:

$$\begin{aligned}
&\left(h^+\left((x + q_i e_i)\right) - h^+(x)\right) N \\
&= \frac{h_{0,a}(x + q_i e_i) - \tilde{h}(y)}{f_\epsilon\left(Pr_a(x + \epsilon q_i e_i)\right)} - \frac{h_{0,a}(x) - \tilde{h}(y)}{f_\epsilon(Pr_a x)} \\
&= q_i \frac{a_i}{\|a\|^2} \frac{1}{f_\epsilon(Pr_a x)} \left(1 + \mathcal{O}\left(\epsilon^2\right)\right)
\end{aligned} \tag{5.1.115}$$

If on the other hand $\tilde{h}(x + q_i e_i) = \tilde{h}(x) + 1$ we obtain:

$$\begin{aligned}
&\left(h^+(x + q_i e_i) - h^+(x)\right) N \\
&= \min_{\lambda \in [k(x), k(x)+1)} \frac{1}{f_\epsilon(\lambda a)} + \\
&\quad + \left(\frac{h_{0,a}(x + q_i e_i) - \tilde{h}(x) - 1}{f_\epsilon\left(Pr_a(x + q_i e_i)\right)}\right) - \frac{h_{0,a}(x) - \tilde{h}(x)}{f_\epsilon(Pr_a x)} \\
&= q_i \frac{a_i}{\|a\|^2} \frac{1}{f_\epsilon(Pr_a x)} \left(1 + \mathcal{O}\left(\delta^3/\epsilon\right)\right).
\end{aligned} \tag{5.1.116}$$

Comparing (5.1.115) and (5.1.116), we see, that this hold independent of a possible jump of $\tilde{h}$.

We denote $\Sigma^- := \partial^+ D_m \setminus W_\delta^+$ and $\Sigma^+ := \partial^- D_m \setminus W_\delta^+$. Inserting h^+ into the Dirichlet form, we obtain

$$\begin{aligned}
\overline{\Phi}\left(h^+\right) &= \overline{\Phi}_{W_\delta}\left(h^+\right) + \sum_{x \in \Sigma^-} \sum_{y \in \Sigma^+} \overline{C}_\epsilon(x, y) + \\
&\quad + \sum_{x \in \partial^- W_\delta} \sum_{y \in \partial^+ W_\delta} \overline{C}_\epsilon(x, y) \left(h^+(x) - h^+(y)\right)^2.
\end{aligned} \tag{5.1.117}$$

Since we are in the case of discrete time, we have $\overline{C}_\epsilon(x, y) \leq \overline{\pi}_\epsilon(x) \wedge \overline{\pi}_\epsilon(y)$. Therefore

$$\sum_{x \in \Sigma^-} \sum_{y \in \Sigma^+} \overline{C}_\epsilon(x, y) \leq d \left|\Sigma^-\right| \exp\left(-\frac{1}{\epsilon}\left(F_\epsilon(0) + \delta^2\right)\right) \tag{5.1.118}$$

and

$$\begin{aligned}
&\sum_{x \in \partial^- W_\delta} \sum_{y \in \partial^+ W_\delta} \overline{C}_\epsilon(x, y) \left(h^+(x) - h^+(y)\right)^2 \\
&\leq \left|\partial^+ W_\delta\right| \exp\left(-\frac{1}{\epsilon}\left(F_\epsilon(0) + \delta^2\right)\right),
\end{aligned} \tag{5.1.119}$$

because $h^+ = 0$ on $\partial_m^- W_\delta$ and $h^+ = 1$ on $\partial_n^- W_\delta$ and the inequality (5.1.110). Since F has compact level sets and $\delta = \sqrt{k\epsilon \left|\ln \epsilon\right|}$, these term are negligible.

Now we estimate the first summand of (5.1.117). By a quadratic approximation inside W_δ we obtain

$$\overline{\Phi}_{W_\delta}\left(h^+\right) = K\left(\epsilon\right)\exp\left(-F_\epsilon\left(0\right)/\epsilon\right)\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right), \tag{5.1.120}$$

where $K\left(\epsilon\right)$ is defined by

$$K\left(\epsilon\right) := \sum_{x\in W_\delta} f_\epsilon\left(x\right)\sum_{i=1}^{d} p_i\left(0\right)\left(h^+\left(x+q_ie_i\right)-h^+\left(x\right)\right)^2. \tag{5.1.121}$$

With (5.1.116) we can estimate

$$K\left(\epsilon\right) \le \frac{\left(1+\mathcal{O}\left(\delta\right)\right)}{\left\|a\right\|^2 N^2}\sum_{x\in W_\delta} f_\epsilon\left(x\right)\left(\max_{\lambda\in[k(x),k(x)+1)}\left\{f_\epsilon\left(\lambda a\right)\right\}\right)^{-2}. \tag{5.1.122}$$

The crucial point is that the sum over $i \in \{1,\ldots,d\}$ vanishes. We use Proposition 5.13 and Definition 5.14 to bring the non-disjoint sets $E_{r,t}$ into the picture. They provide for every $i \in \{1,\ldots,d\}$ and $z \in W_\delta$:

$$\sum_{y\in E_{r,t}:y_d=z_d}\varphi_i^{(r,t)}\left(y\right) = \sum_{E_{s,u}\ni z}\varphi_i^{(s,u)}\left(z\right) = 1. \tag{5.1.123}$$

Therefore we can proceed like

$$\begin{aligned}
&\sum_{x\in W_\delta} f_\epsilon\left(x\right)\left(\max_{\lambda\in[k(x),k(x)+1)}\left\{f_\epsilon\left(\lambda a\right)\right\}\right)^{-2}\\
&\le \sum_{E_{r,t}\in\hat{W}_\delta}\sum_{x\in E_{r,t}} f_\epsilon\left(x\right)\sum_{i=1}^{d}\varphi_i^{(r,t)}\left(x\right)\frac{a_i^2}{\left\|a\right\|^2}\times\\
&\qquad\times\left(\max_{\lambda\in[t,t+1)}\left\{f_\epsilon\left(\lambda a\right)\right\}\right)^{-2}\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right)\\
&= \left\|a\right\|^2 \operatorname{cap}\left(E,\tilde{c}\right)\sum_{E_{r,t}\in\hat{W}_\delta}\max_{x\in E_{r,t}} f_\epsilon\left(x\right)\times\\
&\qquad\times\left(\max_{\lambda\in[t,t+1)}\left\{f_\epsilon\left(\lambda a\right)\right\}\right)^{-2}\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right).
\end{aligned} \tag{5.1.124}$$

The last equation holds, because Definition 5.14(b) and Proposition 5.13 provides

$$\frac{1}{\|a\|^4}\sum_{x\in E_{r,t}}\sum_{i=1}^{d}\varphi_i^{(r,t)}(x)\,a_i^2=\mathrm{cap}\,(E,\tilde{c})\,. \tag{5.1.125}$$

Inserting equation (5.1.124) into (5.1.122) provides with the help small shifts of element to get the parametrisation (5.1.99):

$$\begin{aligned}
&K(\epsilon)/\mathrm{cap}\,(E,\tilde{c})\\
&\leq \sum_{E_{r,t}\in\hat{W}_\delta}\left(\max_{x\in E_{r,t}} f_\epsilon(x)\right)\left(\max_{\lambda\in[t,t+1)} f_\epsilon(\lambda a)\right)^{-2}\times\\
&\quad\times\left(\sum_{t\in T_\delta}\left(\max_{\lambda\in[t,t+1)} f_\epsilon(\lambda a)\right)^{-1}\right)^{-2}\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right)\\
&=\sum_{\rho\in 2R_\delta}\exp\left(-\frac{\epsilon}{2}\ell^2\sum_{i=1}^{d-1}\hat{\lambda}_i\rho_i^2\right)\times\\
&\quad\times\left(\sum_{t\in T_\delta}\exp\left(-\frac{\epsilon}{2}\|a\|^2\hat{\lambda}_d t^2\right)\right)^{-1}\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right)\\
&\leq k(\epsilon)/\mathrm{cap}\,(E,\tilde{c})\left(1+\mathcal{O}\left(\delta^3/\epsilon\right)\right).
\end{aligned} \tag{5.1.126}$$

Therefore the upper bound coincides with the lower bound up to these error, and we are done. Since the expressions for the upper and lower bound of $\mathrm{cap}\,(m,n)$ agrees in this precision before an explicit evaluation of the sums in (5.1.126), it should be possible to get the same result for more general graphs.

Non rational directions.

To prove the case $z\equiv Q^{-1}w_d\notin\mathbb{Q}^d$, we first observe that z is an element of the one-dimensional eigenspace associated to the negative eigenvalue, λ_d, of $A:=Q^{-1}\nabla^2F_\epsilon(0)$. Consider an increasing sequence $z_n\in\mathbb{Q}^d$ such that $\|z_n-z\|_2<1$ and $\lim_{n\to\infty}z_n=z$. Choose $r\in\mathbb{R}$, such that $U_\delta\cup W_\delta\subset B_r(0)$, the $\|\cdot\|_1$-ball in Λ_q. Let $D_{v,w}:\mathbb{R}^d\to\mathbb{R}^d$ be the rotation from v to w, such that $D_{v,w}(x)=x$ for $x\in\mathbb{R}^d\setminus\mathrm{span}\,(v,w)$.

We define $g_n \in C^\infty\left(\mathbb{R}^d\right)$ with the following properties: g_n is bijective and

$$g_n(x) = \begin{cases} D_{z_n,z}(x) & \text{for } x \in B_r(0), \\ x & \text{for } x \notin B_{r+1}(0). \end{cases} \tag{5.1.127}$$

Consider now the sequence of functions $F_{\epsilon,n} := F_\epsilon \circ g_n$. Then

$$Q^{-1}\nabla^2 F_{\epsilon,n}(0) = D^T_{z_n,z} A D_{z_n,z} \tag{5.1.128}$$

has an eigenvector $D^T_{z_n,z} z = z_n \in \mathbb{Q}^d$ associated to λ_d. By construction $F_{\epsilon,n} \to F_\epsilon$ uniformly.

We denote by $\pi_{\epsilon,n}$ the probability measure given by

$$\pi_{\epsilon,n}(x) = \frac{1}{Z_{\epsilon,n}} \exp\left(-\frac{1}{\epsilon} F_{\epsilon,n}(x)\right) \tag{5.1.129}$$

with normalisation $Z = \sum \exp\left(-\frac{1}{\epsilon} F_{\epsilon,n}(x)\right)$. We define a Markov process $\xi^{\epsilon,n}$ by putting

$$p_{\epsilon,n}(x,y) := g_\epsilon(x,y) \min\left(1, \frac{\pi_{\epsilon,n}(y)}{\pi_{\epsilon,n}(x)}\right). \tag{5.1.130}$$

Apparently $\xi^{\epsilon,n}$ is reversible with respect to $\pi_{\epsilon,n}$. Moreover, $p_{\epsilon,n} \to p_\epsilon$ in the operator-norm associated to $l_2(\pi_\epsilon)$. Since a Markov chain is uniquely determined by its transition matrix and the sequence $(L_{\epsilon,n})$ is uniformly tight, we obtain (see e.g. Theorem 15.5 on p. 127 in Billingsley [Bil68]) that $\xi^{\epsilon,n} \to \xi^\epsilon$ in $D([0,\infty),\Lambda)$.

Therefore the stochastic representation of $h_{A,B}$ of Proposition 3.4 yields $h^n_{A,B} \to h_{A,B}$ pointwise. We estimate

$$\begin{aligned} &\left\|e^n_{A,B} - e_{A,B}\right\|_{\pi_\epsilon} \\ &= \left\|L_{\epsilon,n} h^n_{A,B} - L_\epsilon h_{A,B}\right\|_{\pi_\epsilon} \\ &\le \left\|L_{\epsilon,n}\left(h^n_{A,B} - h_{A,B}\right)\right\|_{\pi_\epsilon} + \left\|\left(L_{\epsilon,n} - L_\epsilon\right) h_{A,B}\right\|_{\pi_\epsilon} \\ &\le \left\|L_{\epsilon,n}\right\|_\infty \left\|\left(h^n_{A,B} - h_{A,B}\right)\right\|_{\pi_\epsilon} + \left\|L_{\epsilon,n} - L_\epsilon\right\|_\infty \left\|h_{A,B}\right\|_{\pi_\epsilon}. \end{aligned} \tag{5.1.131}$$

Here we used again the operator norm

$$\|L\|_\infty := \sup_{h \in l^2(\pi_\epsilon)} \frac{\|Lh\|_{\pi_e}}{\|h\|_{\pi_e}}. \tag{5.1.132}$$

Therefore $e^n_{A,B} \to e_{A,B}$ in $l^2(\pi_\epsilon)$ and the capacity of ξ^ϵ is also the limit of the capacities of the approximating Markov processes ξ^ϵ_n. $\square$

Remark 5.17 (Rectangular lattices) Consider the case of a rectangular lattice, i.e. $\Lambda_{\epsilon r} := \Lambda \cap \left(\times_{i=1}^{d} (\epsilon r_i \mathbb{Z})\right)$. This problem can be reduced to the one treated before: Let $\tau \in \mathbb{R}^{d\times d}$ be the diagonal matrix with entries $\tau_{ii} := r_i$ and put $\check{f} := f \circ \tau$ for every function $f : \Lambda_{\epsilon r} \to \mathbb{R}$. Then the Dirichlet form can be transformed by a substitution $y := \tau^{-1} x$ as follows:

$$\begin{aligned}\Phi_{\Gamma_s}(h) &= \sum_{x\in\Lambda_{\epsilon r}} \sum_{i=1}^{d} C_i(x)\left(h(x+\epsilon r_i e_i) - h(x)\right)^2 \\ &= \sum_{y\in\Lambda_{\epsilon}} \sum_{i=1}^{d} \check{C}_i(y)\left(\check{h}(y+\epsilon e_i) - \check{h}(y)\right)^2 \qquad (5.1.133)\end{aligned}$$

We approximate $\check{C}_i$ in the neighbourhood of a relevant saddle s through

$$\begin{aligned}\check{C}_i(y) &= p_i \check{\pi}_\epsilon(y)\left(1+\mathcal{O}(\delta)\right) \\ &= p_i e^{F_\epsilon(s)/\epsilon} e^{\langle y, \tau B \tau y\rangle}\left(1+\mathcal{O}(\delta)\right) \qquad (5.1.134)\end{aligned}$$

with $B := \nabla^2 F_\epsilon(s)$. Now Theorem 5.7 yields that

$$\begin{aligned}\operatorname{cap}(m,n) &= \left(\frac{2\pi}{\epsilon}\right)^{d/2-1} \frac{|\lambda_r|}{\sqrt{|\det \nabla^2 F_\epsilon(s)|}} \times \\ &\quad \times \exp\left(-F_\epsilon(s)/\epsilon\right)\left(1+\mathcal{O}(\delta)\right), \qquad (5.1.135)\end{aligned}$$

where λ_r is the unique negative eigenvalue of $\tau^2 p \cdot (\nabla^2 F_\epsilon(s))$.

Several relevant saddles

Now we treat the case of finitely many relevant saddle points, i.e.

$$S_\epsilon(m, \mathcal{M}\setminus m) = \{s_i \,|\, i \in J\}, \qquad (5.1.136)$$

where the cardinality $|J|$ does not depend on ϵ. We show that the transition over each saddle point can be considered separately.

In the following definition we use that we have only quadratic essential saddle points.

Definition 5.18 Let A and $B \subset \Lambda$ be disjoint and compact. Assume $|S(A,B)| \geq 2$. We call the relevant saddle points in $S(A,B)$ *serial* if every optimal path $\gamma \in \mathcal{O}(A,B)$ visits all of them. The other extreme are parallel saddle points: We call a set of relevant saddle points *parallel* if there is no optimal path that visits two of them.

Corollary 5.19 *Let ξ^ϵ be a family of Markov chains that satisfies the assumptions of Theorem 5.7.*

(a) Then we obtain for parallel relevant saddle points

$$\operatorname{cap}(m, \mathcal{M} \setminus m) = \left(\frac{2\pi}{\epsilon}\right)^{d/2-1} \sum_{i \in J} \frac{\hat{\lambda}_d^{(i)}}{\sqrt{|\det \nabla^2 F_\epsilon(s_i)|}} e^{-\frac{1}{\epsilon}\hat{F}_\epsilon(m, \mathcal{M}_\epsilon \setminus m)} \times \times (1 + \mathcal{O}\left(\sqrt{\epsilon}\,|\ln \epsilon|^{3/2}\right)). \tag{5.1.137}$$

(b) For serial relevant saddle points we obtain

$$\begin{aligned}&\operatorname{cap}(m, \mathcal{M} \setminus m)\\ &= \left(\frac{2\pi}{\epsilon}\right)^{d/2-1} \left[\sum_{i \in J} \frac{\sqrt{|\det \nabla^2 F_\epsilon(s_i)|}}{\hat{\lambda}_d^{(i)}}\right]^{-1} e^{-\frac{1}{\epsilon}\hat{F}_\epsilon(m, \mathcal{M}_\epsilon \setminus m)} \times\\ &\quad \times (1 + \mathcal{O}\left(\sqrt{\epsilon}\,|\ln \epsilon|^{3/2}\right)).\end{aligned} \tag{5.1.138}$$

Here, $-\hat{\lambda}_d^{(i)}$ is the unique negative eigenvalue of $(p_j(s_i)\,\partial_j\partial_k F_\epsilon(s_i))_{j,k}$.

Remark 5.20 Observe, that case (b) can only occur, if the potential F_ϵ has local minima, that does not belong to $\mathcal{M}_\epsilon$.

In the general case we have a graph structure between the relevant saddle points. This can, as the cases of parallel and serial saddles, be treated like an electrical network, where we want to calculate the effective conductance, given the conductance of all edges.

Proof. The proof of Theorem 5.7 shows that under our assumptions the prefactor of the capacity is determined by a neighbourhood of the relevant saddle points of radius $\delta \equiv \sqrt{k\epsilon\,|\ln \epsilon|}$ with $k > 0$ constant. Denote by $A_i := \bigcup_{j \in J \setminus i} B_{\sqrt{\epsilon}|\ln \epsilon|}(s_i)$ the union of balls with radius $\sqrt{\epsilon}\,|\ln \epsilon|$ around $S_\epsilon(m, \mathcal{M}_\epsilon \setminus m) \setminus s_i$ for $i \in J$.

ad (a). There exists disjoint optimal paths $\gamma_i \in \mathcal{O}(m, \mathcal{M}_\epsilon \setminus m)$, such that $s_i \in \gamma_j$ iff $i = j$ and therefore the a priori bounds are valid and we can choose neighbourhoods $U_i := U_\delta^{(i)}$ of s_i, such that

$$\begin{aligned}
\operatorname{cap}(m, \mathcal{M}_\epsilon \setminus m) &= \sum_{i \in J} \Phi_{U_i}\left(h_{m,\mathcal{M}_\epsilon \setminus m}\right)(1 + \mathcal{O}(\delta)) \\
&= \sum_{i \in J} \inf_{h_i} \Phi_{U_i}(h_i)(1 + \mathcal{O}(\delta)) \\
&= \pi(m) \sum_{i \in J} P_m\left(\tau_{\mathcal{M}_\epsilon \setminus m} < \tau_{m \cup A_i}\right)(1 + \mathcal{O}(\delta)) . \qquad (5.1.139)
\end{aligned}$$

The explicit form follows with Theorem 5.7. We can also apply the method of the upper bound: then the neighbourhoods in the separatrix can be chosen separately.

ad (b). Denote $n := |J|$. We choose an optimal path $\gamma \in \mathcal{O}(m, \mathcal{M}_\epsilon \setminus m)$. By definition γ visits all relevant saddle points between m and $\mathcal{M}_\epsilon \setminus m$. We arrange them as (s_i) according to their appearance in γ. Now we define $x_0 = m$ and let x_i be the first minimum γ visits between s_i and s_{i+1} for $1 \le i \le n-1$. Moreover, let x_n be the first minimum γ visits in $\mathcal{M}_\epsilon \setminus m$. Denote by

$$\mathcal{F}_{i,j} := \{f : G_\epsilon \to \mathbb{R} \mid f \text{ unit flow from } x_i \text{ to } x_j\}, \qquad (5.1.140)$$

then it follows with Thompson's principle (3.2.8):

$$\begin{aligned}
\frac{1}{\operatorname{cap}(m, \mathcal{M} \setminus m)} &= \inf_{f \in \mathcal{F}_{0,n}} \sum_{x,y \in \Lambda_\epsilon} \frac{1}{C_{xy}} f_{xy}^2 \\
&= \inf_{f \in \mathcal{F}_{0,n}} \sum_{i=1}^{n} \sum_{x,y \in U_i} \frac{1}{C_{xy}} f_{xy}^2 \left(1 + \mathcal{O}\left(\delta^3/\epsilon\right)\right) \\
&= \sum_{i=1}^{n} \inf_{f_i \in \mathcal{F}_{i-1,i}} \sum_{x,y \in U_i} \frac{1}{C_{xy}} (f_i)_{xy}^2 \left(1 + \mathcal{O}\left(\delta^3/\epsilon\right)\right) \\
&= \sum_{i=1}^{n} \frac{1}{\operatorname{cap}(x_{i-1}, x_i)} \left(1 + \mathcal{O}\left(\delta^3/\epsilon\right)\right) . \qquad (5.1.141)
\end{aligned}$$

The explicit form follows again with Theorem 5.7. $\square$

5.2 Eyring-Kramers formula

We will use now Proposition 3.21 to compute mean hitting times. Starting from a minimum $m \in \mathcal{M}_\epsilon$, the first quantity we are interested in is the expected time ξ^ϵ needs to change over to $\mathcal{M}_\epsilon \setminus m$.

To get explicit formula we introduce another assumption on F_ϵ, namely

F3 The function F_ϵ has exponentially tight level sets, i.e. there exists $c_a > 0$ independent of ϵ and at most polynomial in a such that

$$\sum_{x \in \Lambda_\epsilon : F_\epsilon(x) \geq a} \exp\left(-\frac{1}{\epsilon} F_\epsilon(x)\right) \leq c_a \epsilon^{-d} \exp\left(-\frac{a}{\epsilon}\right). \tag{5.2.1}$$

We need this assumption to estimate integral by the Laplace method, see Appendix, Proposition B.2.

The main theorem in this section is

Theorem 5.21 *(Eyring-Kramers formula) Let $\mathcal{M}_\epsilon$ be the set of local minima of F_ϵ. Let $m \in \mathcal{M}_\epsilon$ and $I \subset \mathcal{M}_\epsilon \setminus m$ such that for all $n \in \mathcal{M}_\epsilon \setminus (I \cup m)$ the barriers satisfies*

$$B_\epsilon(m, n) > B_\epsilon(n, I) \tag{5.2.2}$$

or

$$B_\epsilon(n, m) < B_\epsilon(n, I). \tag{5.2.3}$$

Then

$$\begin{aligned} \mathbb{E}_m(\tau_I) &= \epsilon^{-d/2} \frac{(2\pi)^{d/2}}{\operatorname{cap}(m, I)} \sum_{n \in V_I(m)} \frac{1}{\sqrt{\det(\nabla^2 F(n))}} e^{-F(n)/\epsilon} \times \\ & \quad \times (1 + \mathcal{O}\left(\sqrt{\epsilon} \left|\ln \epsilon\right|^{3/2}\right)) \end{aligned} \tag{5.2.4}$$

The sum is meant to reach all $n \in \mathcal{M}_\epsilon \setminus I$ and in particular includes always $n = m$.

Proof. Proposition 3.21 yields in our setting

$$\mathbb{E}_m\left(\tau_{\mathcal{M}_\epsilon \setminus m}\right) = \frac{1}{\operatorname{cap}(m, \mathcal{M}_\epsilon \setminus m)} \sum_{y \notin \mathcal{M}_\epsilon \setminus m} \pi_\epsilon(y)\, h_{m, \mathcal{M}_\epsilon \setminus m}(y). \tag{5.2.5}$$

The a priori estimates on the equilibrium potential $h_{m,\mathcal{M}_\epsilon\backslash m}$, see Corollary 4.11, are qualitatively of the same form as in the continuous case, see [BEGK04], Corollary 4.8, p. 414. Moreover, Proposition B.2 of the appendix reveals that also the Laplace asymptotics are, up to a factor ϵ^{-d}, the same as in the continuous case. Hence the proof is identical to the one of Theorem 6.2, p. 420 in [BEGK04]. Observe that the range of the sum in ([BEGK04]) is

$$n : \hat{F}_\epsilon (m,n) < \hat{F} (n,I) . \tag{5.2.6}$$

This is indeed the same as ours, since if n satisfies (5.2.6) then $m \in V_I(n)$ and hence also $n \in V_I(m)$. □

5.3 The global picture

In this section we summarise the results of [BEGK01] and apply our more precise estimates of the capacity. For the results on admissible transitions, we need the following stronger assumptions:

T1 Given any two minima $m, n \in \mathscr{M}_\epsilon$ the set of relevant saddle points $S_\epsilon (m,n)$ contains a unique element $s^* (m,n)$.

T2 F_ϵ can be represented as $F_\epsilon = F_{0,\epsilon} + \epsilon F_{1,\epsilon}$, where $F_{1,\epsilon}$ is Lipschitz and $F_{0,\epsilon}$ is twice Lipschitz, i.e. for $i \in \{1,2\}$

$$|F_{i,\epsilon} (x) - F_{i,\epsilon} (y)| \leq C \|x-y\|_1 \tag{5.3.1}$$

and moreover

$$\|\nabla_\epsilon F_{0,\epsilon} (x) - \nabla_\epsilon F_{0,\epsilon} (y)\|_\infty \leq C \|x-y\|_1 , \tag{5.3.2}$$

where $\|x\|_\infty = \max_{1\leq i\leq d} |x_i|$ is the maximum norm in $\mathbb{R}^d$.

Notation 5.22 In case assumption **T1** holds and $s \equiv s^* (m,n)$, we denote the valley $V_n^{(\epsilon)} (m)$ also by $V_s (m)$.

Assumption **S2** yields that all essential saddle points are quadratic. Therefore V_s consists of two components, that we denote by $V_s^\pm$ with the understanding, that

$$\inf_{x\in V_s^+} F_\epsilon (x) < \inf_{x\in V_s^-} F_\epsilon (x) \tag{5.3.3}$$

holds.

Under assumption of uniqueness of the relevant saddle points the structure of the landscape F_ϵ is encoded in a tree structure, that we define on the set $\mathcal{M}_\epsilon \cup \mathcal{E}_\epsilon$. Define for any essential saddle $s \in \mathcal{E}_\epsilon$ the two "children"

$$a_s^\pm = \begin{cases} \arg\max \{F_\epsilon(x) \mid x \in \mathcal{E}_\epsilon \cap V_s^\pm\} & \text{for } \mathcal{E}_\epsilon \cap V_s^\pm \neq \emptyset, \\ \mathcal{M}_\epsilon \cap V_s^\pm & \text{else.} \end{cases} \tag{5.3.4}$$

Note, that the set $\mathcal{M}_\epsilon \cap V_s^\pm$ consists of a single point, if $\mathcal{E}_\epsilon \cap V_s^\pm = \emptyset$. Now draw a link from any essential saddle to the two points $a_s^\pm$. This produces a connected tree, $\mathcal{T}_\epsilon$, with underlying set $\mathcal{E}_\epsilon \cup \mathcal{M}_\epsilon$ having the property, that all leaves are local minima, while all other points are essential saddle points.

An alternative way to construct this tree is by starting from below: From each local minimum draw a link to the lowest essential saddles connecting it to other minima. Then from each saddle point, that was reached before, draw a line to the lowest saddle point above it, that connects it to further minima. Continue until all minima are connected. Since we have assumed that there is always a unique relevant saddle point between two minima, both procedures give a unique answer. Denote by $\mathcal{T}_{s,x}$ the branch of $\mathcal{T}_\epsilon$ emanating from s, that contains x and by $\mathcal{T}_s$ the union of the two branches emanating from s.

The tree $\mathcal{T}_\epsilon$ induces a natural hierarchical distance between two points in $\mathcal{E}_\epsilon \cup \mathcal{M}_\epsilon$, given by the length of the shortest path on $\mathcal{T}_\epsilon$ needed to join them. This distance encodes the all information on the time scales of "exits" from valleys. What is missing, is how the process descends into a neighbouring valley after such an exit. It turns out, that all we need to know in addition, is which minimum the process visits first after crossing a saddle point. In general, the process has the option to visit various minima first with certain probabilities. We will here only refer to the case, where F_ϵ is such, that there is always one minimum that is visited first with overwhelming probability. This situation is discussed in [BG99] and they showed, that under condition **T** one can construct a certain deterministic dynamical system, which selects in every valley, $V_s(x)$ a unique minimum, that is first visited after entering the valley through the saddle point s. To make this more precisely, we introduce the event

$$T_\epsilon(x,y) := \{\tau_y \leq \tau(V_s(x)^c \cap \mathcal{M}_\epsilon) \text{ and } \xi_0^\epsilon = x\}, \tag{5.3.5}$$

where $x, y \in \mathcal{M}$, $s = s^*(x,y)$ and $V_s(x)^c \equiv \Lambda_\epsilon \setminus V_s(x)$. In words y is the first minimum outside the valley $V_s(x)$, that the Markov process ξ^ϵ is visiting.

Bovier and Gayrard showed by using large deviation estimates on the path space (look [BEGK01], Prop 4.3, p. 125)

Proposition 5.23 *Let $m, n \in \mathcal{M}_\epsilon$ and $s \equiv s^*(m,n)$ their unique relevant saddle. Assume* **T2** *and that the probability for ξ^ϵ when started in m to reach a δ-neighbourhood of the boundary of Λ in finite time T is exponentially small. Then there exists a unique minimum $x \in V_s(n)$ and $\alpha > 0$, such that*

$$\mathbb{P}_m(T_\epsilon(m,x)) \geq 1 - e^{-\epsilon^{-\alpha}}. \tag{5.3.6}$$

This proposition motivates the following

Definition 5.24 A pair of minima $(m,n) \subset \mathcal{M}_\epsilon$ is called connected, if

1. m is the deepest minimum in the valley $V_s(m)$ for $s = s^*(m,n)$,
2. n is the unique minimum in $V_s(n)$, such that $\mathbb{P}_m(T_\epsilon(m,n)) \geq 1 - e^{-K_\epsilon/\epsilon}$.

In this case the event $T_\epsilon(x,y)$, defined by (5.3.5) is called an admissible transition. Note, that the number of points connected to a special $m \in \mathcal{M}_\epsilon$ is of course greater or equal to one and can be arbitrary large.

As [BEGK01] pointed out, the rough estimate of Corollary 4.11 shows that each transition can be decomposed into a sequence of admissible transitions. The time scale for the transition is determined by the first admissible transition, because this involves the relevant saddle point between the starting point and the end point.

Another result we take from [BEGK01], Prop 5.5, p. 139, is

Proposition 5.25 *Let $s \in \mathcal{E}_\epsilon$ and $m \in V_s$ the deepest minimum of $V_s(m)$. Then for $\beta < \alpha$,*

$$E_m(\tau_m \,|\, \tau_m < \tau(V_s(m)^c \cap \mathcal{M}_\epsilon)) = \frac{\pi_\epsilon(V_s(m))}{\pi_\epsilon(m)}\left(1 + \mathcal{O}\left(e^{-\epsilon^{-\beta}}\right)\right). \tag{5.3.7}$$

This shows, that the expected recurrence time at m without leaving the valley $V_s(m)$ is up to exponentially small errors equal to the same time of the restricted Markov chain $\tilde{\xi}_\epsilon$ with state space $V \equiv V_s(m) \subset \Lambda_\epsilon$ and transition probabilities

$$\tilde{p}_\epsilon(x,y) := \begin{cases} p_\epsilon(x,y) & \text{if } x \in V, y \in \mathcal{N}_x \cap V \\ 1 - \sum_{z \in \mathcal{N}_x \cap V} p_\epsilon(x,z) & \text{if } y = x \in V \\ 0 & \text{else.} \end{cases} \tag{5.3.8}$$

Let $\tilde{\pi}_\epsilon$ be the invariant measure of $\tilde{\xi}_\epsilon$, then obviously

$$\tilde{\pi}_\epsilon (B) = \pi_\epsilon (B \cap V) / \pi_\epsilon (V) . \tag{5.3.9}$$

Using the ergodic Theorem (see see [HLL03], Proposition 3.3.1, p. 44), we obtain

$$\tilde{E}_m \tau_m = \frac{1}{\tilde{\pi}_\epsilon (m)} = \frac{\pi_\epsilon (V)}{\pi_\epsilon (m)}. \tag{5.3.10}$$

We have seen that ξ^ϵ will choose with overwhelming probability the way over the relevant saddle point $s^* (m, n)$ to change over to another minimum $n \in \mathcal{M}_\epsilon$. There is of course some probability, that it will take a completely different way. We will compute the mean hitting time τ_n for ξ^ϵ starting in m and conditioned that (m, n) is a pair of connected minima.

Theorem 5.26 *Let $m, n \in \mathcal{M}_\epsilon$ and $B_\delta (m)$ the ball with radius δ around m. Assume that (m, n) are connected minima and there is a unique relevant saddle point $s = s^* (m, n)$ between m and n. Assume, there exists $c > 0$ small, such that for $\delta > 0$ small enough*

$$F_\epsilon (x) \geq F_\epsilon (m) + c\delta^2 \textit{ for all } x \in V_s (m) \setminus B_\delta (d) . \tag{5.3.11}$$

Then $P_m (T_\epsilon (m, n))$ converges for $\epsilon \downarrow 0$ exponentially fast to one and the Eyring-Kramers formula is valid, i.e.

$$\begin{aligned} &\mathbb{E}_m (\tau_n \,|\, T_\epsilon (m, n)) \\ &= \frac{2\pi}{\epsilon \hat{\lambda}_d (q)} \frac{\sqrt{|\det \nabla^2 F_\epsilon (s)|}}{\sqrt{\det \nabla^2 F_\epsilon (m)}} \exp \left(\frac{1}{\epsilon} B_\epsilon (m, n) \right) \left(1 + \mathcal{O} \left(\frac{\delta^3}{\epsilon} \right) \right) \end{aligned} \tag{5.3.12}$$

Proof. Suppose ξ^ϵ starts in $x \in \Lambda_\epsilon$. Let $I \subset \Lambda_\epsilon$ and $y \notin I \cup x$. We will derive another renewal equation by splitting the events that ξ^ϵ returns to x or goes directly to y :

$$\begin{aligned} &\mathbb{E}_x (\tau_y \,|\, \tau_y < \tau_I) \\ &= \mathbb{P}_x (\tau_{I\cup y} < \tau_x) \, \mathbb{E}_x (\tau_y \,|\, \tau_y < \tau_{I\cup x}) \\ &\quad + \mathbb{P}_x (\tau_x < \tau_{I\cup y}) \left(\mathbb{E}_x (\tau_x \,|\, \tau_x < \tau_{I\cup y}) + \mathbb{E}_x (\tau_y \,|\, \tau_y < \tau_I) \right) . \end{aligned} \tag{5.3.13}$$

Therefore

$$\begin{aligned} \mathbb{E}_x (\tau_y \,|\, \tau_y < \tau_I) &= \frac{\mathbb{E}_x (\tau_x \,|\, \tau_x < \tau_{I\cup y})}{\mathbb{P}_x (\tau_{I\cup y} < \tau_x)} \mathbb{P}_x (\tau_x < \tau_{I\cup y}) + \\ &\quad + \mathbb{E}_x (\tau_y \,|\, \tau_y < \tau_{I\cup x}) . \end{aligned} \tag{5.3.14}$$

We will use this equation now for $x = m$, $y = n$ and $I = (V_s(m)^c \cap \mathcal{M}_\epsilon) \setminus n$. Bovier et al. proved (compare their proof of Theorem 5.1, page 137 in [BEGK01]), that in this case equation 5.3.14 can be estimated as

$$\mathbb{E}_m(\tau_n \,|\, \tau_n < \tau_I) = \frac{\mathbb{E}_m(\tau_m \,|\, \tau_m < \tau_{I\cup n})}{\mathbb{P}_m(\tau_{I\cup n} < \tau_m)} \left(1 + \mathcal{O}\left(e^{-\epsilon^\alpha}\right)\right). \tag{5.3.15}$$

With the help of Proposition 5.25 and the definition of the capacity, we obtain

$$\mathbb{E}_m(\tau_n \,|\, \tau_n < \tau_I) = \frac{\pi_\epsilon(V_s(m))}{\operatorname{cap}(m, I\cup n)} \left(1 + \mathcal{O}\left(\epsilon^{-\kappa} e^{-K_\epsilon/\epsilon}\right)\right). \tag{5.3.16}$$

Since we have assumed condition (5.3.11), we can now directly apply the Laplace method, see Appendix, Proposition B.1, and obtain

$$\sum_{y\in V_s(m)} e^{-F_\epsilon(y)} \le \left(\det \nabla^2 F_\epsilon(m)\right)^{-1/2} \left(\frac{2\pi}{\epsilon}\right)^{d/2} e^{-F_\epsilon(m)/\epsilon} \left(1 + \mathcal{O}\left(\sqrt{\epsilon}\right)\right). \tag{5.3.17}$$

The capacity $\operatorname{cap}(m, I\cup n)$ can be estimate with Theorem 5.7, because $s^*(m, I\cup n) = s^*(m,n)$. Inserting these results into the formula for the conditioned mean hitting time (5.2.5) yields the Eyring-Kramers formula for the lattice

$$\begin{aligned}&\mathbb{E}_m(\tau_n \,|\, T_\epsilon(m,n)) \\ &= \frac{2\pi}{\epsilon} \frac{1}{\hat{\lambda}_d(q)} \frac{\sqrt{|\det \nabla^2 F_\epsilon(s)|}}{\sqrt{|\det \nabla^2 F_\epsilon(m)|}} e^{B_\epsilon(m,n)/\epsilon} \left(1 + \mathcal{O}\left(\frac{\delta^3}{\epsilon}\right)\right).\end{aligned} \tag{5.3.18}$$

$\square$

5.4 Discrete approximation of SDE

Let $\Lambda \subset \mathbb{R}^d$ be an open connected set. Let $F \in C^3(\Lambda)$ with exponentially tight level sets, i.e.

$$\int_{x\in\Lambda: F(x)\ge a} e^{-F(x)/\epsilon} dx \le c_a e^{-a/\epsilon}. \tag{5.4.1}$$

In the following, we will construct the generator of a Markov process with continuous time, that provides a discrete version of the stochastic differential equation

$$dX_t = \nabla F(X_t)\, dt + \sqrt{2\epsilon} dB_t. \tag{5.4.2}$$

Denote $\Lambda_h := \Lambda \cap h\mathbb{Z}^d$ and let $\Gamma_h = (\Lambda_h, G_h)$ be the graph with edges between nearest neighbours. We choose h small enough so that Λ_h is a connected graph. We define on Γ_h:

$$\begin{aligned}
\nabla_h f(x) &:= \frac{1}{h\sqrt{2}} \left(f(y) - f(x) \right)_{y \in \mathcal{N}_x}, & (5.4.3)\\
\mathrm{div}_h Z(x) &:= \frac{1}{h\sqrt{2}} \sum_{y \in \mathcal{N}_x} \left(Z(x,y) - Z(y,x) \right), & (5.4.4)\\
\Delta_h f(x) &:= \frac{1}{h^2} \sum_{y \in \mathcal{N}_x} \left(f(y) - f(x) \right). & (5.4.5)
\end{aligned}$$

Note that with this definitions and the scalar products

$$\langle f, g \rangle := \sum_{x \in \Lambda_h} f(x)\, g(x) \tag{5.4.6}$$

on $l_2(\Lambda_h)$ and

$$\langle Y, Z \rangle = \sum_{x \in \Lambda_h} \sum_{y \in \mathcal{N}_x} Y(x,y)\, Z(x,y) \tag{5.4.7}$$

on $l_2(G_h)$, the following relations are valid

$$\begin{aligned}
\langle \nabla_h f, Z \rangle &= - \langle f, \mathrm{div}_h Z \rangle, & (5.4.8)\\
\langle \nabla_h f, \nabla_h f \rangle &= - \langle f, \Delta_h f \rangle. & (5.4.9)
\end{aligned}$$

Now consider the generator of the diffusion process X_t^ϵ

$$\mathscr{L} = \epsilon e^{F/\epsilon} \mathrm{div} \left(e^{-F/\epsilon} \nabla \right). \tag{5.4.10}$$

It's discrete analogue on Λ_h is

$$\mathcal{L}_h = \epsilon e^{F/\epsilon} \mathrm{div}_h \left(e^{-F/\epsilon} \nabla_h \right). \tag{5.4.11}$$

Therefore

$$\begin{aligned}
\mathcal{L}_h f(x) &= \frac{\epsilon}{2h^2} e^{F(x)/\epsilon} \sum_{y \in \mathcal{N}_x} \left(e^{-F(x)/\epsilon} + e^{-F(y)/\epsilon} \right) \left(f(y) - f(x) \right)\\
&= \frac{\epsilon}{2h^2} \sum_{y \in \mathcal{N}_x} \left(1 + e^{(F(x) - F(y))/\epsilon} \right) \left(f(y) - f(x) \right). & (5.4.12)
\end{aligned}$$

This is the generator of a continuous time Markov process ζ_h with rates

$$r_h(x,y) = \frac{\epsilon}{2h^2}\left(1 + e^{(F(x)-F(y))/\epsilon}\right) \qquad \text{for } y \in \mathcal{N}_x, \tag{5.4.13}$$

$r_\epsilon(x,x) = -\sum_{y\in\mathcal{N}_x} r_\epsilon(x,y)$ and zero else. Therefore ζ_h is stable and conservative. The embedded Markov chain, ξ^h, has a transition matrix

$$p_\epsilon(x,y) := \frac{r_\epsilon(x,y)}{|r_\epsilon(x,x)|} = \frac{1 + e^{(F(x)-F(y))/\epsilon}}{\sum_{z\in\mathcal{N}_x}\left(1 + e^{(F(x)-F(z))/\epsilon}\right)}. \tag{5.4.14}$$

ξ^ϵ is irreducible and has reversible probability measure, ν_ϵ, given by

$$\nu_\epsilon(x) = \frac{\sum_{y\in\mathcal{N}_x}\left(1 + e^{(F(x)-F(y))/\epsilon}\right)}{4d\sum_{z\in\Lambda_h} e^{-F(z)/\epsilon}} e^{-F(x)/\epsilon}. \tag{5.4.15}$$

Hence ξ^ϵ is positive recurrent. Therefore ζ^ϵ is nonexplosive.

The invariant probability measure of ζ_h is

$$\pi_h(x) = \frac{1}{Z_h}\exp\left(-F(x)/\epsilon\right) \tag{5.4.16}$$

with normalisation factor $Z_h := \sum_{x\in\Lambda_h}\exp\left(-F(x)/\epsilon\right)$.

To show a convergence result of ζ_h, we look at it as a process on the Skorohod space $D([0,\infty),\Lambda)$.

Theorem 5.27 *The Markov processes ζ_h given by $\mathcal{L}_h$ converges in $D([0,\infty),\Lambda)$ for $h\to 0$ to the diffusion X^ϵ on Λ with generator $\mathscr{L}$.*

Proof. First we show, that

$$\mathcal{L}_h f(x) \to \mathscr{L} f(x) \tag{5.4.17}$$

for every $f \in C_b^2(\Lambda)$. Consider the following calculation

$$\begin{aligned}
&\frac{1}{2h^2}\sum_{y\in\mathcal{N}_x}\left(e^{-F(x)/\epsilon}+e^{-F(y)/\epsilon}\right)(f(y)-f(x))\\
&=\frac{1}{2h^2}\sum_{i=1}^{d}e^{-F(x+he_i)/\epsilon}(f(x+he_i)-f(x))-\\
&\qquad -e^{-F(x)/\epsilon}(f(x)-f(x-he_i))+\\
&\qquad +e^{-F(x)/\epsilon}(f(x+he_i)-f(x))-\\
&\qquad -e^{-F(x-he_i)/\epsilon}(f(x)-f(x-he_i))\\
&=\frac{1}{2h}\sum_{i=1}^{d}e^{-F(x+he_i)/\epsilon}\partial_i f(x+he_i)-e^{-F(x)/\epsilon}\partial_i f(x)\\
&\quad +e^{-F(x)/\epsilon}\partial_i f(x)-e^{-F(x-he_i)/\epsilon}\partial_i f(x-he_i)+R(h)\\
&\to \operatorname{div}\left(e^{-F(x)/\epsilon}\nabla f(x)\right)\qquad \text{for } h\downarrow 0.
\end{aligned}\tag{5.4.18}$$

The correction term R is defined by

$$\begin{aligned}
&R(h)\\
&=\frac{1}{2h}\sum_{i=1}^{d}e^{-F(x)/\epsilon}(\partial_i f(x)-\partial_i f(x+(h-\xi_1)e_i))+\\
&\quad +e^{-F(x+he_i)/\epsilon}(\partial_i f(x+(h-\xi_1)e_i)-\partial_i f(x+he_i))+\\
&\quad +e^{-F(x)/\epsilon}(\partial_i f(x)-\partial_i f(x-(h-\xi_2)e_i))+\\
&\quad +e^{-F(x-he_i)/\epsilon}(\partial_i f(x-he_i)-\partial_i f(x-(h-\xi_2)e_i))\\
&\to 0\qquad \text{for } h\downarrow 0.
\end{aligned}\tag{5.4.19}$$

Here the mean value theorem yields $\xi_1, \xi_2 \in (0, h)$, i.e. small real numbers, going to 0 for $h \downarrow 0$.

We still have to show the tightness of $(P_h)_{h\in(0,1)}$, the laws of $(\zeta_h)_{h\in(0,1)}$ in $D([0,\infty), \Lambda)$. To do this, we introduce its modulus of continuity, w_{ζ_h} by

$$w_{\zeta_h}(\delta) := \sup_{|s-t|<\delta}\|\zeta_h(s)-\zeta_h(t)\|. \tag{5.4.20}$$

Now we use Theorem 15.5 on p. 127 of Billingsley [Bil68], that says

Proposition 5.28 *Suppose that for each positive η, there exists an $a \in \mathbb{R}$ such that*

$$P_h\left(|\zeta_h(0)| > a\right) \leq \eta, \qquad \forall h < 1. \tag{5.4.21}$$

Suppose further that, for each positive κ and η, there exist a $\delta \in (0,1)$, and an $h_0 \in (0,1)$, such that

$$P_h\left(w_{\zeta_h}(\delta) \geq \kappa\right) \leq \eta, \qquad \forall h \leq h_0. \tag{5.4.22}$$

Then $(P_h)_{h\in(0,1)}$ is tight, and, if P is the weak limit of a subsequence $(P_{h'})$, then $P(C) = 1$.

To verify the conditions of Proposition 5.28 first notice, that the first condition is satisfied if the processes ζ_h are started in single points x_h, such that $\lim_{h\downarrow 0} x_h = x \in \Lambda$. To show the second condition we assume $\kappa < \epsilon$ and denote by σ_1 the time of the first jump of ζ_h. Hence σ_1 is a random variable with exponential distribution and parameter $r_\epsilon(x) := \sum_{y\in\mathcal{N}_x} r_\epsilon(x,y)$, where x denotes the starting point of ζ_h. We denote the transition probability of ζ_h by p_h, that means

$$p_h(t,x,y) = P_x\left(\zeta_h(t) = y\right). \tag{5.4.23}$$

We obtain for a fixed starting point $x \in \Lambda_h$:

$$\begin{aligned}
&P_x\left(\sup_{s<t<s+\delta} \|\zeta_h(t) - \zeta_h(s)\| \geq \kappa\right)\\
&\leq \sup_s \sum_{y\in\Lambda_h} P_x\left(\zeta_h(s) = y\right) P_y\left(\sigma_1 < \delta\right)\\
&= \sup_s \sum_{y\in\Lambda_h} p_h(s,x,y)\left(1 - e^{-r_\epsilon(y)\delta}\right).
\end{aligned} \tag{5.4.24}$$

Denote by $A(x) := \{y \in \Lambda_h \,|\, \exists z \sim y,\ \text{s.t.}\ \max\{F(y), F(z)\} < F(x)\}$. Since F has exponentially tight level sets (5.4.1) $A(x)$ is a compact set. Moreover we obtain

$$p_h(s,x,y) \leq e^{-(F(y)-F(x))/\epsilon} \quad \text{for } y \notin A(x), \tag{5.4.25}$$

since the process has to climb onto the level $F(y)$.

ζ_h is an irreducible time-continuous Markov process. Hence a fixed starting point x and $t > 0$, $p_h(t,x,y) > 0$ for all $y \in \Lambda_h$. Therefore the h dependence of p_h has the form

$$p_h(t,x,y) = h^d g(t,x,y)\left(1 + o(1)\right) \tag{5.4.26}$$

to allow $\sum_{y\in\Lambda_h} p_h(t,x,y)=1$. This yields up to multiplicative errors $(1+o(1))$ in h :

$$
\begin{aligned}
&P_x\left(\sup_{s<t<s+\delta}\|\zeta_h(t)-\zeta_h(s)\|\geq\kappa\right)\\
&\leq\delta\sum_{y\in A(x)}h^d r_\epsilon(y)+\sum_{y\in A(x)^c}h^d e^{-(F(y)-F(x))/\epsilon}r_\epsilon(y)\\
&=\delta\sum_{y\in A(x)}\sum_{z\in\mathcal{N}_y}h^d\left(1+e^{(F(y)-F(z))/\epsilon}\right)+\\
&\qquad+e^{F(x)/\epsilon}\sum_{y\in A(x)^c}\sum_{z\in\mathcal{N}_y}h^d\left(e^{-F(y)/\epsilon}+e^{-F(z)/\epsilon}\right)\\
&\leq\delta\left(2de^{F(x)/\epsilon}h^d|A(x)|+4dc_{F(x)}\right),
\end{aligned}
\tag{5.4.27}
$$

where we have used again (5.4.1). □

Now we consider the case $h=\epsilon$.

Corollary 5.29 *(of Theorem 5.7 and Theorem 5.21) Assume $F\in C^3(\Lambda)$ has exponentielly tight level sets and satisfies the conditions **S1-S3**. Let ζ_ϵ be the continuous time Markov process with statespace Λ_ϵ and generator $\mathcal{L}_\epsilon$ given by (5.4.11). Let $\mathcal{M}$ be the set of local mimima of F. Let $I,J\subset\mathcal{M}_\epsilon$ with $I\cap J=\emptyset$ and assume $s\equiv s^*_\epsilon(I,J)$ is the unique relevant saddle point between them. Then the capacity of ζ_ϵ is given by*

$$
\begin{aligned}
\operatorname{cap}(I,J) &= \left(\frac{2\pi}{\epsilon}\right)^{d/2-1}\frac{\lambda_d}{\sqrt{|\det\nabla^2F(s)|}}\frac{\exp(-F(s)/\epsilon)}{\sum_{x\in\Lambda_\epsilon}e^{-F(x)/\epsilon}}\times\\
&\quad\times\left(1+\mathcal{O}\left(\sqrt{\epsilon}\,|\ln\epsilon|^{3/2}\right)\right),
\end{aligned}
\tag{5.4.28}
$$

where $-\lambda_d$ is the unique negative eigenvalue of $\nabla^2F(s)$.

The expected hitting times between local minima are given by Theorem 5.21.

Proof. Let ξ^ϵ be the embedded Markov chain of ζ_ϵ, whose transition probability is given by (5.4.14). We compare it with the Metropolis Markov chain of π_ϵ, given in (5.4.16). The Metropolis algorithm has a transition matrix p^M_ϵ on the Λ_ϵ with

$$
p^M_\epsilon(x,y)=\frac{1}{2d}e^{-[F(y)-F(x)]^+/\epsilon}\qquad\text{for } y\in\mathcal{N}_x \tag{5.4.29}
$$

and $p_\epsilon^M(x,x) := 1 - \sum_{y\in\mathcal{N}_x} p_\epsilon^M(x,y) \geq 0$. The connection to the process χ_ϵ is given by

$$p_\epsilon(x,y) = g_\epsilon(x,y)\, p_\epsilon^M(x,y), \tag{5.4.30}$$

where g_ϵ is the symmetric function defined by

$$\begin{aligned} g_\epsilon(x,y) &= 2de^{[F(x)-F(y)]^+/\epsilon} \frac{1+e^{(F(x)-F(y))/\epsilon}}{\sum_{y\in\mathcal{N}_x}\left(1+e^{(F(x)-F(y))/\epsilon}\right)} \\ &= \frac{2d\left(1+e^{|F(y)-F(x)|/\epsilon}\right)}{\sum_{y\in\mathcal{N}_x}\left(1+e^{(F(x)-F(y))/\epsilon}\right)}. \end{aligned} \tag{5.4.31}$$

The function g_i defined by $g_i(x) := g_\epsilon(x, x+\epsilon e_i)$ is on $K \subset\subset \Lambda$ bounded from below for ϵ small enough by

$$\min_{x\in K} \frac{1+e^{|\partial_i F(x)|}}{1+2\max_{j=1}^d \left(e^{|\partial_j F(x)|}\right)}. \tag{5.4.32}$$

Hence, condition **C1** is satisfied. Moreover, condition **C2** is satisfied, since g_i is Lipschitz continuous in a neighbourhood of a critical point $s \in \Lambda_\epsilon$. We obtain

$$\begin{aligned} g_i(s) &= 2d\frac{1+e^{|F(s+\epsilon e_i)-F(s)|/\epsilon}}{\sum_{z\in\mathcal{N}_s}\left(1+e^{(F(s)-F(z))/\epsilon}\right)} \\ &= 1+\mathcal{O}(\epsilon), \end{aligned} \tag{5.4.33}$$

since

$$F(s+\epsilon e_i) = F(s) + \frac{1}{2}\epsilon^2 \partial_i^2 F(s). \tag{5.4.34}$$

Similarly the reversible probability measure ν_ϵ of ξ^ϵ given by

$$\nu_\epsilon(x) = \frac{\sum_{y\in\mathcal{N}_x}\left(1+e^{(F(x)-F(y))/\epsilon}\right)}{4d\sum_{z\in\Lambda_h} e^{-F(z)/\epsilon}} e^{-F(x)/\epsilon} \tag{5.4.35}$$

satisfies

$$\nu_\epsilon(s) = \frac{e^{-F(s)/\epsilon}}{\sum_{z\in\Lambda_h} e^{-F(z)/\epsilon}}\left(1+\mathcal{O}(\epsilon)\right). \tag{5.4.36}$$

Applying Theorem 5.7 yields formula (5.4.28) for the capacity $\mathrm{cap}(m,n)$. Apparently the conditions of Theorem 5.21 are also satisfied. $\square$

Part III

Metastability in the Hopfield model

6 The Hopfield model

6.1 The Hopfield Hamiltonian

Let N be a natural number and consider the vertexset

$$\Lambda := \{1, \ldots, N\} . \tag{6.1.1}$$

Virtually all objects we introduce will depend on N, so we will hide this dependence in some cases. We call $\mathcal{S}_N := \{-1, 1\}^N$ the *set of spin configurations.* Let $\{\xi^1, \ldots, \xi^M\}$ be fixed spin configurations. We consider the *Hopfield Hamiltonian* $H_N : \mathcal{S}_N \to \mathbb{R}_{\leq 0}$ given by

$$H_N(\sigma) = -\frac{1}{2N} \sum_{\mu=1}^{M} \langle \xi^\mu, \sigma \rangle^2 . \tag{6.1.2}$$

Observe that several sites $i \in \Lambda$ are subject to the same *force*

$$\frac{\partial H_N}{\partial \sigma_i} = -\frac{1}{N} \sum_{\mu=1}^{M} \xi_i^\mu \langle \xi^\mu, \sigma \rangle . \tag{6.1.3}$$

Therefore we can change to a reduced representation of the Hopfield model, in which the independent degrees of freedom are $d := 2^M$ mean field variables. This transformation was first used by Grensing and Kühn in [GK86].

Let $\{b_1, \ldots, b_d\}$ be a fixed enumeration of all vectors in $\{-1, 1\}^M$. Any choice of M patterns can then be regarded as a map

$$\xi : i \mapsto \xi_i \equiv \left(\xi_i^1, \xi_i^2, \ldots, \xi_i^M\right) \tag{6.1.4}$$

that associates to each site $i \in \Lambda$ one of the vectors b_k. Hence the map ξ determines a partition of Λ into sets Λ_k given by

$$\Lambda_k := \{i \in \Lambda \mid \xi_i = b_k\} . \tag{6.1.5}$$

We restrict now the choices of patterns such that each Λ_k is non empty.

Denote the number of sites in Λ_k by

$$\ell_k := |\Lambda_k| , \tag{6.1.6}$$

therefore $\sum_{k=1}^{d} \ell_k = N$. Note that of course ℓ_k depends on N and ξ although this is not indicated.

Denote by $\mathcal{L}_N := \times_{k=1}^{d} \left(\frac{2}{\ell_k} \mathbb{Z} \right)$ the rectangular lattice with spacings $2/\ell_k$. We define the set of *mean field configurations* to be

$$\mathcal{X}_N := [-1,1]^d \cap \mathcal{L}_N \tag{6.1.7}$$

and the map $X_N : \mathcal{S}_N \to \mathcal{X}_N$ by setting

$$X_{N,k}(\sigma) := \frac{1}{\ell_k} \sum_{i \in \Lambda_k} \sigma_i. \tag{6.1.8}$$

X_N determines a partition of the spin configuration space $\mathcal{S}_N$ into ξ dependent subsets $\mathcal{S}_N(x) := X_N^{-1}(x)$, indexed by $x \in \mathcal{X}_N$. We say that X_N *lumps together* the sites in each Λ_k. Notice, that X_N maps the space $\mathcal{S}_N$ of asymptotically infinite dimension to a subset of $[-1,1]^d$ and therefore mean field configurations are much better to handle. Using the partition $\{\Lambda_k\}$ of Λ, we obtain

$$\begin{aligned} \langle \xi^\mu, \sigma \rangle &= \sum_{k=1}^{d} \sum_{i \in \Lambda_k} \xi_i^\mu \sigma_i \\ &= \sum_{k=1}^{d} b_k^\mu \ell_k X_{N,k}(\sigma). \end{aligned} \tag{6.1.9}$$

Let L denote the diagonal matrix with entries $L_{kk} := \ell_k$. We denote by P the orthogonal projection of $\mathbb{R}^d$ onto the subspace spanned by the vectors $\{b^1, \ldots, b^M\}$, i.e.

$$P_{jk} := \frac{1}{d} \sum_{\mu=1}^{M} b_j^\mu b_k^\mu. \tag{6.1.10}$$

Then we may write

$$\begin{aligned} H_N(\sigma) &= -\frac{1}{2N} \sum_{\mu=1}^{M} \langle b^\mu, L X_N(\sigma) \rangle^2 \\ &= -\frac{d}{2N} |P L X_N(\sigma)|^2. \end{aligned} \tag{6.1.11}$$

Remark 6.1 We denote in this whole chapter the euclidean norm in any $\mathbb{R}^n$ by $|\cdot|$.

For any $\beta \in \mathbb{R}_{\geq 0}$ we define the Gibbs measure $\pi \equiv \pi_{N,\beta}$ on the finite set Λ by setting:

$$\pi(\sigma) := \frac{1}{Z_{N,\beta}} e^{-\beta H_N(\sigma)}. \tag{6.1.12}$$

Here, the *partition function* $Z_{N,\beta} := \sum_{\sigma \in \mathcal{S}} e^{-\beta H_N(\sigma)}$ is a normalising factor.

6.2 Dynamics

To model the dynamics, we construct a reversible Markov chain

$$\sigma_{N,\beta} = \{\sigma_{N,\beta}(t)\}_{t \in \mathbb{N}_0} \tag{6.2.1}$$

on $\mathcal{S}_N$. The kind of stochastic dynamics we use is called *Glauber* dynamics, because in each time step only a single spin flip occurs. We denote by σ^i the configuration with spins

$$\left(\sigma^i\right)_j = \begin{cases} \sigma_j & \text{for } j \neq i, \\ -\sigma_j & \text{for } j = i. \end{cases} \tag{6.2.2}$$

In order to use the lumping procedure induced by X_N defined in (6.1.8), we choose transition probabilities $w_N \equiv w_{N,\beta}$ of the form

$$w_N(\sigma, \tau) := \begin{cases} \frac{1}{N} c_N\left(X_N(\sigma), X_N(\tau)\right) \min\left(1, \frac{\pi(\tau)}{\pi(\sigma)}\right), & \|\tau - \sigma\|_1 = 2, \\ 1 - \sum_{i=1}^N w_N\left(\sigma, \sigma^i\right), & \tau = \sigma, \\ 0, & \text{else}, \end{cases} \tag{6.2.3}$$

where $c_N : \mathcal{X}_N \times \mathcal{X}_N \to \mathbb{R}_{\geq 0}$ is a symmetric function. Therefore $\sigma_{\beta,N}$ is reversible.

We define $c_{N,k}(x) := c_N\left(x, x + \frac{2}{\ell_k} e_k\right)$ for all $1 \leq k \leq d$ and assume

D there exists $c > 0$, independent of N, such that

$$c_{N,k}(x) \geq c \tag{6.2.4}$$

for all $x \in \Lambda_N$ and $1 \leq k \leq d$. Moreover, we assume $c_{N,k}$ is Lipschitz continuous, more precisely there exists $L > 0$, independent of N, such that

$$|c_{N,k}(x) - c_{N,k}(y)| \leq L\,|x - y| \quad \text{for all } x, y \in \mathcal{X}_N. \tag{6.2.5}$$

To lift $\sigma_{N,\beta}$ onto the space of mean field configurations, we define a linear transformation, A_N, which maps functions on $\mathcal{S}_N$ to functions on $\mathcal{X}_N$ by

$$(A_N f)(x) := \sum_{\sigma \in \mathcal{S}_N(x)} f(\sigma). \tag{6.2.6}$$

Proposition 6.2 *The chain $\zeta_{N,\beta} \equiv \{\zeta_{N,\beta}(t)\}_{t\in\mathbb{N}_0}$ on the set $\mathcal{X}_N$ of mean field configurations defined by $\zeta_{N,\beta}(t) := X_N(\sigma_{N,\beta}(t))$ is again a Markov chain and has transition matrix, $p_N \equiv p_{N,\beta}$, given by*

$$\begin{aligned} &p_N(x,y) \\ &= \begin{cases} \frac{\ell_k}{2N} c_N(x,y)\left((1-x_k)\wedge(1+y_k)\frac{\varrho(y)}{\varrho(x)}\right), & y = x + \frac{2}{\ell_k}e_k, \\ \frac{\ell_k}{2N} c_N(x,y)\left((1+x_k)\wedge(1-y_k)\frac{\varrho(y)}{\varrho(x)}\right), & y = x - \frac{2}{\ell_k}e_k, \\ 1 - \sum_{y\in\mathcal{N}_x} p_N(x,y), & y = x, \\ 0 & \textit{else.} \end{cases} \end{aligned} \tag{6.2.7}$$

$\zeta_{N,\beta}$ is reversible with respect to the new Gibbs measure $\varrho \equiv \varrho_{N,\beta}$ that is determined by the Hamiltonian

$$\mathcal{H}_{N,\beta}(x) = -\frac{d}{2N}|PLx|^2 - \frac{1}{\beta}\ln|\mathcal{S}_N(x)|. \tag{6.2.8}$$

Remark 6.3 Observe that $p_N(x,y) > 0$ for all nearest neighbours x, y. To show this, we assume without loss of generality that $y = x + \frac{2}{\ell_k}e_k$. Therefore $x_k = y_k - \frac{2}{\ell_k} \le 1 - \frac{2}{\ell_k}$ and hence $(1-x_k) \ge \frac{2}{\ell_k} > 0$. Analogously, $1 + y_k \ge \frac{2}{\ell_k}$ holds true.

Proof. The Gibbs measure on $\mathcal{X}_N$ is defined by

$$\varrho(x) = (A_N \pi)(x). \tag{6.2.9}$$

Since π depends not on all information of σ but only on $X_N(\sigma)$ as we showed in (6.1.11), we can write with a slight abuse of notation

$$\varrho(x) = |\mathcal{S}_N(x)|\,\pi(x). \tag{6.2.10}$$

ϱ is the Gibbs distribution for the mean field Hamiltonian $\mathcal{H}_{N,\beta}$, given by

$$\mathcal{H}_{N,\beta}(x) = -\frac{d}{2N}|PLx|^2 - \frac{1}{\beta}\ln|\mathcal{S}_N(x)|. \tag{6.2.11}$$

The matrix p_N is defined by the equation $p_N(A_N f) = A_N(w_N f)$. Inserting $x \in \mathcal{X}_N$ and putting $f = 1_{\mathcal{S}_N(y)}$ provides the form

$$p_N(x, y) = \frac{1}{|\mathcal{S}_N(x)|} \sum_{\sigma \in \mathcal{S}_N(x)} \sum_{\tau \in \mathcal{S}_N(y)} w_N(\sigma, \tau). \tag{6.2.12}$$

The Markov property holds, iff the probability to go from $\mathcal{S}_N(x)$ to $\mathcal{S}_N(y)$ does not depend on the starting point, i.e.

$$\sum_{\tau \in \mathcal{S}_N(y)} w_N(\sigma, \tau) = \sum_{\tau \in \mathcal{S}_N(y)} w_N(\sigma', \tau) \tag{6.2.13}$$

for all $\sigma, \sigma' \in \mathcal{S}_N(x)$. To prove this, we show that the left hand side does not depend on σ. We denote the canonical basis of $\mathbb{R}^d$ by $\{e_1, \ldots, e_d\}$ and assume $y = x + \frac{2}{\ell_k} e_k$. If the mean field configuration should increase in Λ_k, then the flipped spin has to be a minus-spin. Hence

$$\begin{aligned} &\sum_{\tau \in \mathcal{S}_N\left(x + \frac{2}{\ell_k} e_k\right)} w_N(\sigma, \tau) \\ &= \sum_{i \in \Lambda_k} w_N\left(\sigma, \sigma^i\right) \delta_{\sigma_i, -1} \\ &= \frac{\ell_k}{2N} (1 - x_k)\, c_N(x, y) \left(1 \wedge \frac{|\mathcal{S}_N(x)|\, \varrho(y)}{|\mathcal{S}_N(y)|\, \varrho(x)}\right). \end{aligned} \tag{6.2.14}$$

We used again that the Gibbs measure $\pi(\sigma)$ depends only on $X(\sigma)$, i.e. $\pi(\sigma) = \frac{\varrho(X_N(\sigma))}{|\mathcal{S}(X_N(\sigma))|}$ and the number of minus spins in Λ_k is $\frac{1}{2}\ell_k(1 - x_k)$. For $y = x - \frac{2}{\ell_k} e_k$ we can derive analogously

$$\begin{aligned} &\sum_{\tau \in \mathcal{S}_N\left(x - \frac{2}{\ell_k} e_k\right)} w_N(\sigma, \tau) \\ &= \frac{\ell_k}{2N} (1 + x_k)\, c_N(x, y) \left(1 \wedge \frac{|\mathcal{S}_N(x)|\, \varrho(y)}{|\mathcal{S}_N(y)|\, \varrho(x)}\right). \end{aligned} \tag{6.2.15}$$

Since these expressions does not depend on which σ in $\mathcal{S}_N(x)$ we have chosen, condition (6.2.13) is satisfied and we obtain

$$p_N(x, y) = \sum_{\tau \in \mathcal{S}(y)} w_N(\sigma, \tau) \quad \text{for any } \sigma \in \mathcal{S}_N(x). \tag{6.2.16}$$

To simplify expression (6.2.14) we use

$$\begin{aligned}
\frac{|\mathcal{S}_N(x)|}{\left|\mathcal{S}_N\left(x+\frac{2}{\ell_k}e_k\right)\right|} &= \binom{\ell_k}{\frac{1}{2}(\ell_k+x_k\ell_k)} \Big/ \binom{\ell_k}{\frac{1}{2}(\ell_k+x_k\ell_k+2)} \\
&= \frac{\ell_k+x_k\ell_k+2}{\ell_k-x_k\ell_k} = \frac{1+\left(x_k+\frac{2}{\ell_k}\right)}{1-x_k}. \qquad (6.2.17)
\end{aligned}$$

Plugging this into equations (6.2.14) and (6.2.15) we obtain the form given in Proposition 6.2.

The reversibility of p_N with respect to ϱ follows directly from equation (6.2.12):

$$\begin{aligned}
\varrho(x)\, p_N(x,y) &= \pi(x) \sum_{\sigma\in\mathcal{S}_N(x)} \sum_{\tau\in\mathcal{S}_N(y)} w_N(\sigma,\tau) \\
&= \sum_{\sigma\in\mathcal{S}_N(x)} \sum_{\tau\in\mathcal{S}_N(y)} \pi(\sigma)\, w_N(\sigma,\tau) \\
&= \sum_{\sigma\in\mathcal{S}_N(x)} \sum_{\tau\in\mathcal{S}_N(y)} \pi(\tau)\, w_N(\tau,\sigma) \\
&= \varrho(y)\, p_N(y,x). \qquad (6.2.18)
\end{aligned}$$

We have used here again $\pi(x)$ to denote $\pi(\sigma)$ for any $\sigma\in\mathcal{S}_N(x)$. $\square$

Example 6.4 (a) As a particular example we consider the Metropolis sample for the Gibbs distribution π

$$w_N(\sigma,\tau) := \begin{cases} \frac{1}{N}\left(1\wedge\frac{\pi(\sigma^i)}{\pi(\sigma)}\right), & \tau=\sigma^i, \\ 1-\sum_{i=1}^N w_N(\sigma,\sigma^i), & \tau=\sigma, \\ 0 & \text{else.} \end{cases} \qquad (6.2.19)$$

In this case, the transition matrix p_N of the Markov chain $\zeta_{N,\beta}$ on the mean field configurations has of course the form

$$p_N(x,y) = \begin{cases} \frac{\ell_d}{2N}\left((1-x_k)\wedge(1+y_k)\frac{\varrho(y)}{\varrho(x)}\right), & y=x+\frac{2}{\ell_k}e_k, \\ \frac{\ell_d}{2N}\left((1+x_k)\wedge(1-y_k)\frac{\varrho(y)}{\varrho(x)}\right), & y=x-\frac{2}{\ell_k}e_k, \\ 1-\sum_{y\in\mathcal{N}_x} p_N(x,y), & y=x, \\ 0 & \text{else.} \end{cases} \qquad (6.2.20)$$

(b) Another interesting dynamics use the "magnetic field" $h : \mathcal{S}_N \to \mathbb{R}^N$ defined by

$$h_i(\sigma) := \frac{1}{N}\sum_{j(\neq i)} J_{ij}\sigma_j = \frac{1}{N}\sum_{\mu=1}^{M}\langle \xi^\mu, \sigma\rangle \xi_i^\mu - \frac{M}{N}\sigma_i. \tag{6.2.21}$$

Like Biroli and Monasson, see [BM98], we define a transition matrix, v_N, on the spin space $\mathcal{S}_N$ by

$$v_N(\sigma,\tau) := \begin{cases} \frac{1}{2N}\left(1-\sigma_i \tanh\left(\beta h_i(\sigma)\right)\right), & \tau = \sigma^i, \\ 1 - \sum_{i=1}^N v_N\left(\sigma, \sigma^i\right), & \tau = \sigma, \\ 0 & \text{else.} \end{cases} . \tag{6.2.22}$$

To prove that v_N is reversible with respect to π of (6.1.12), we observe

$$h_i\left(\sigma^i\right) = h_i(\sigma) \tag{6.2.23}$$

and

$$\begin{aligned} H_N\left(\sigma^i\right) &= -\frac{1}{2N}\sum_{\mu=1}^{M}\left\langle \xi^\mu, \sigma^i\right\rangle^2 \\ &= -\frac{1}{2N}\sum_{\mu=1}^{M}\left(\langle \xi^\mu, \sigma\rangle - 2\xi_i^\mu\sigma_i\right)^2 \\ &= H_N(\sigma) + \frac{2}{N}\sum_{\mu=1}^{M}\langle \xi^\mu, \sigma\rangle \xi_i^\mu \sigma_i - \frac{2M}{N} \\ &= H_N(\sigma) + 2h_i(\sigma)\sigma_i. \end{aligned} \tag{6.2.24}$$

by the definition in (6.2.21). We use $1+\tanh(x) = \frac{2e^x}{e^x+e^{-x}}$ and abbreviate $a_i \equiv \beta h_i(\sigma)\sigma_i$. Then we obtain

$$\pi_{N,\beta}(\sigma)\, v_N\left(\sigma, \sigma^i\right) = \frac{1}{NZ_{N,\beta}}\exp\left(-\beta H_N(\sigma)\right)\frac{e^{-a_i}}{e^{a_i}+e^{-a_i}}. \tag{6.2.25}$$

and

$$\pi_{N,\beta}\left(\sigma^i\right) v_N\left(\sigma^i, \sigma\right) = \frac{1}{NZ_{N,\beta}}\exp\left(-\beta H_N(\sigma) - 2a_i\right)\frac{e^{a_i}}{e^{a_i}+e^{-a_i}}. \tag{6.2.26}$$

Hence we have proved equality.

Analogously to proof of Proposition 6.2 we construct a Markov chain $\zeta_{N,\beta}$ on the space of mean field configurations $\mathcal{X}_N$, that is reversible with respect to ϱ. We only have to check condition (6.2.13) for v_N. Assume $y = x + \frac{2}{\ell_k} e_k$, then for $\sigma \in \mathcal{S}_N(x)$

$$
\begin{aligned}
&\sum_{\tau \in \mathcal{S}_N(y)} v_N(\sigma, \tau) \\
&= \sum_{i \in \Lambda_k} v_N\left(\sigma, \sigma^i\right) \delta_{\sigma_i, -1} \\
&= \frac{\ell_k}{4N}(1 - x_k)\left(1 + \tanh \beta \frac{d}{N}\left((PLx)_k + \frac{M}{d}\right)\right). \qquad (6.2.27)
\end{aligned}
$$

We have used the definition of h in (6.2.21). The last expression is indeed independent of $\sigma \in \mathcal{S}_N(x)$. This works analogously for $y = x - \frac{2}{\ell_k} e_k$. Therefore the transition matrix, q_N, of $\zeta_{N,\beta}$ is given by

$$
\begin{aligned}
&q_N\left(x, x + \frac{2}{\ell_k} e_k\right) \\
&= \frac{\ell_k}{4N}(1 - x_k)\left(1 + \tanh \beta \frac{d}{N}\left((PLx)_k + \frac{M}{d}\right)\right) \qquad (6.2.28)
\end{aligned}
$$

and

$$
\begin{aligned}
&q_N\left(x, x - \frac{2}{\ell_k} e_k\right) \\
&= \frac{\ell_k}{4N}(1 + x_k)\left(1 - \tanh \beta \frac{d}{N}\left((PLx)_k - \frac{M}{d}\right)\right) \qquad (6.2.29)
\end{aligned}
$$

and the usual conditions $q_N(x, x) = 1 - \sum_{y \sim x} q_N(x, y)$ and $q_N(x, y) = 0$ if x and y are not equal or nearest neighbours.

6.3 Random patterns

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We choose $\{\xi_i^\mu\}_{1 \le \mu \le M, i \in \mathbb{N}}$ as a family of mutually independent random variables that attain the values 1 and -1 with equal probability $\frac{1}{2}$. We continue to use the same letters for the objects we have defined. athough the most of them are of course now random variables.

For example, $(\ell_k)_{1\leq k\leq d}$ is a random vector with a multinomial distribution with parameters N and $\frac{1}{d}$. Its components are correlated random variables with mean value $\frac{N}{d}$ and covariance

$$\mathrm{Cov}\left(\ell_k,\ell_j\right)=\frac{N}{d^2}\left(d\delta_{jk}-1\right). \tag{6.3.1}$$

In order to discuss the N dependence of $\mathcal{H}_N$, let us now change to normalised variables by writing ℓ_k in the form

$$\ell_k:=\frac{N}{d}\left(1+\frac{1}{\sqrt{N}}\lambda_k\right), \tag{6.3.2}$$

where λ_k are centered random variables and have covariance

$$\mathrm{Cov}\left(\lambda_k,\lambda_j\right)=d\delta_{jk}-1. \tag{6.3.3}$$

The range of λ_k is the set $\frac{1}{\sqrt{N}}\left\{-N,-N+d,\ldots,(d-1)\,N\right\}\subset\mathbb{R}$. Certainly λ depends on N, although this is not indicated.

Define the Cramér entropy function $I:[-1,1]\to\mathbb{R}$ by

$$I\left(x\right)=\begin{cases}\frac{1}{2}\left(\left(1+x\right)\ln\left(1+x\right)+\left(1-x\right)\ln\left(1-x\right)\right), & x\in\left(-1,1\right),\\ \ln 2, & x\in\left\{-1,1\right\}.\end{cases} \tag{6.3.4}$$

and denote $\Lambda:=\mathrm{diag}\left(\lambda_k\right)$, the diagonal $d\times d$ - matrix with entries $\Lambda_{kk}=\lambda_k$.

Definition 6.5 In this definition we stress the dependence of λ on N and ξ.

To work on a common probability space we define

$$\lambda_N[\xi]:=\lambda_N\left[\left(\xi_1,\ldots,\xi_N\right)\right] \tag{6.3.5}$$

for all $\xi\in\left(\{-1,1\}^M\right)^{\mathbb{N}}$. This can, of course, be done analogously for all quantities that depend on N and ξ.

For each N we define

$$\Xi_N:=\left\{\xi\in\left(\{-1,1\}^M\right)^{\mathbb{N}}\;\middle|\;|\lambda_N[\xi]|<2\sqrt{d\log N}\right\}. \tag{6.3.6}$$

Moreover, denote

$$\Xi:=\liminf_{N\to\infty}\Xi_N, \tag{6.3.7}$$

i.e. Ξ is the space of all $\xi=(\xi_i^\mu)_{1\leq\mu\leq M,i\in\mathbb{N}}$ such that there exists $N_0[\xi]$ and for all $N\geq N_0[\xi]$

$$|\lambda_N[\xi]|\leq 2\sqrt{d\log N}. \tag{6.3.8}$$

Remark 6.6 Observe that for $\xi \in \Xi$ and $N \geq N_0[\xi]$ indeed all subsets $\Lambda_k[\xi]$ are non empty.

Part (a) of the next proposition resembles Lemma 2.2 of [Gen96], while part (b) is due to [KP89] (see equation (2.6) on p. 909).

Proposition 6.7 *(a) Ξ is a set of full measure, i.e. $\mathbb{P}(\Xi) = 1$.*

(b) For $\xi \in \Xi$ and $N \geq N_0[\xi]$, the Hamiltonian $\mathcal{H}_{N,\beta}$ can be written as

$$\mathcal{H}_{\beta,N}(x) = N f_{\beta,\lambda}(x) - \frac{N}{\beta} \ln 2 + \mathcal{O}(\ln N), \tag{6.3.9}$$

where $f_{\beta,\lambda} : [-1,1]^d \to \mathbb{R}$ is the function

$$f_{\beta,\lambda}(x) := \frac{1}{\beta d} \sum_{k=1}^{d} \left(1 + \frac{1}{\sqrt{N}} \lambda_k\right) I(x_k) - \frac{1}{2d} \left| P\left(\mathbb{1} + \frac{1}{\sqrt{N}} \Lambda\right) x\right|^2. \tag{6.3.10}$$

Proof. ad (a). $\frac{\sqrt{N}}{d}\lambda$ is the partial sum of the N centered i.i.d. random variables $\left(1(\xi_i = b_k) - \frac{1}{d}\right)_{k \in \{1,\ldots,d\}}$ with values in $[-1,1]^d$. Therefore the statement follows from the Law of Iterated Logarithm for partial sums of $\mathbb{R}^k$-valued random variables, whose proof can be found more generally for Banach spaces in [LT91], Theorem 8.2 on p. 197.

ad (b). With the help of Stirling's formula

$$\log(n!) = n \log n - n + \log 2\pi n + \mathcal{O}(1/12n) \tag{6.3.11}$$

we can approximate for $a > 0$ and $-1 < b < 1$:

$$\begin{aligned} \binom{a}{\frac{1}{2}a(1+b)} &= a \ln 2 - aI(b) - \frac{1}{2} \ln\left(\frac{\pi}{2} a\left(1 - b^2\right)\right) + \\ &\quad + \mathcal{O}\left(a\left(1 - b^2\right)\right)^{-1}. \end{aligned} \tag{6.3.12}$$

Therefore

$$\begin{aligned} \ln |\mathcal{S}_N(x)| &= \sum_{k=1}^{d} \binom{\ell_k}{\frac{1}{2}\ell_k(1 + x_k)} \\ &= N \ln 2 - \frac{N}{d} \sum_{k=1}^{d} \left(1 + \frac{1}{\sqrt{N}} \lambda_k\right) I(x_k) + \\ &\quad + \mathcal{O}(\ln N). \end{aligned} \tag{6.3.13}$$

The last estimation holds, since for $\xi \in \Xi$ and $N \geq N_0[\xi]$ we obtain

$$\begin{aligned} \left| \ln \ell_k - \ln \frac{N}{d} \right| &= \left| \ln \left(1 + \frac{1}{\sqrt{N}} \lambda_k \right) \right| \\ &\leq \left| \ln \left(1 - 2 \frac{\sqrt{d \ln N}}{\sqrt{N}} \right) \right| . \end{aligned} \tag{6.3.14}$$

This last expression converges to zero for $N \to \infty$. □

Remark 6.8 Note that the function $f_{\beta,\lambda}$ depends only over terms $\frac{\lambda_k}{\sqrt{N}}$ on λ and N. In particular, $f_\beta := f_{\beta,0}$ depends neither on ξ nor N (except of course if M would depend on N).

Proposition 6.9 *For $\xi \in \Xi$ and $N \geq N_0[\xi]$, the sequence of functions $f_{\beta,\lambda}$ converges for $N \to \infty$ uniformly to the deterministic function f_β, i.e.*

$$\|f_{\beta,\lambda} - f_\beta\|_\infty \leq 3 \frac{1+\beta}{\beta} \sqrt{d \frac{\ln N}{N}} . \tag{6.3.15}$$

Proof. This is exactly the meaning of Proposition 2.3 in [KP89], p. 912 with $\lambda_{KP} := \frac{1}{\sqrt{N}} \lambda$, $\delta_{KP} := 2\sqrt{d \ln N}/\sqrt{N}$, $U_{N,\delta} := \Xi_N$ and $\eta_{KP} := 0$. To be clear we indexed the quantities Koch and Piasko use with a KP. □

We introduce the matrix A_N that will be crucial to control the random deviation of the minima and 1-saddles of $f_{\beta,\lambda}$ compared to the deterministic ones of f_β.

Definition 6.10 (a) Denote by $\mathcal{A}_M$ the $M\,(M-1)\,/2$ dimensional vector space of symmetric $M \times M$ matrices with vanishing diagonal.

(b) Define $A_N \in \mathcal{A}_M$ by setting

$$A_N^{\mu,\nu} := \frac{1}{d} \langle b^\mu, \Lambda b^\nu \rangle \tag{6.3.16}$$

for all $\mu, \nu \in \{1, \ldots, M\}$.

We prove some properties of A_N in the next

Proposition 6.11 *(a) $\{A_N^{\mu,\nu}, \mu < \nu\}$ are uncorrelated random variables with expectation zero and variance one on $(\Omega, \mathcal{F}, \mathbb{P})$. Alternatively A_N can be written for $\mu \neq \nu$ in the form*

$$A_N^{\mu,\nu} = \frac{1}{\sqrt{N}} \langle \xi^\mu, \xi^\nu \rangle . \tag{6.3.17}$$

(b) For all $\xi \in \Xi$ and $N \geq N_0[\xi]$, we obtain for all $x \in \mathbb{R}^M$

$$|A_N x| \leq 2\sqrt{p \ln N}\, |x| . \tag{6.3.18}$$

(c) There exists $(\gamma_n^{\mu,\alpha})_{1 \leq \mu < \alpha \leq p; n \in \mathbb{N}}$ i.i.d. one dimensional standard normal distributed random variables on a common probability space with ξ such that

$$|A_N^{\mu,\alpha} - g_N^{\mu,\alpha}| = \mathcal{O}\left(\frac{\log N}{\sqrt{N}}\right) \tag{6.3.19}$$

almost surely, where

$$g_N^{\mu,\alpha} = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \gamma_n^{\mu,\alpha} \tag{6.3.20}$$

for $\mu < \alpha$ and $g_N \in \mathscr{A}_M$.

Remark 6.12 The matrices $(g_N)_{N \in \mathbb{N}}$ can be understood as a random walk in $\mathcal{A}_M$ with time parameter $N \in \mathbb{N}$ that starts in zero and has i.i.d. Gaussian increments. For any N the $M(M-1)/2$ independent components of g_N are one dimensional standard Gaussians.

Proof. ad (a). We have $A_N^{\mu,\mu} = \frac{1}{d}\mathrm{tr}(\Lambda) = 0$. For $\alpha \neq \mu$ and $\overline{\alpha} \neq \overline{\mu}$ we obtain with (6.3.3)

$$\mathbb{E}[A_N^{\alpha,\mu}] = \frac{1}{d} \sum_k b_k^\alpha b_k^\mu \mathbb{E}[\lambda_k] = 0 \tag{6.3.21}$$

and

$$\begin{aligned} \mathbb{E}\left[A_N^{\alpha,\mu} A_N^{\overline{\alpha},\overline{\mu}}\right] &= \frac{1}{d^2} \sum_{j,k} b_j^\alpha b_j^\mu \mathbb{E}[\lambda_j \lambda_k]\, b_k^{\overline{\alpha}} b_k^{\overline{\mu}} \\ &= \frac{1}{d} \left\langle b^{\alpha\mu}, b^{\overline{\alpha\mu}} \right\rangle - \frac{1}{d^2} \langle b^\alpha, b^\mu \rangle \left\langle b^{\overline{\alpha}}, b^{\overline{\mu}} \right\rangle \\ &= \delta_{\{\alpha,\mu\},\{\overline{\alpha},\overline{\mu}\}}. \end{aligned} \tag{6.3.22}$$

In other words the $\{A_N^{\alpha,\mu}, \alpha < \mu\}$ are uncorrelated random variables with expectation zero and variance one.

To prove the alternative representation for $\mu \neq \nu$, notice that

$$
\begin{aligned}
&\frac{1}{\sqrt{N}} \sum_{i \in \Lambda} \xi_i^\mu \xi_i^\nu \\
&= \frac{1}{\sqrt{N}} \sum_{k=1}^{d} \sum_{i \in \Lambda_k} \xi_i^\mu \xi_i^\nu \\
&= \frac{1}{\sqrt{N}} \sum_{k=1}^{d} b_k^\mu b_k^\nu \ell_k = \frac{1}{d} \left\langle b^\mu, \Lambda b^\nu \right\rangle ,
\end{aligned}
\tag{6.3.23}
$$

because of the orthogonality of b^μ and b^ν.

ad (b) This is Corollary 2.4 in [Gen96], p. 250, except that we have relaxed iterated logarithm to a logarithm.

ad (c) This property is adopted from Külske ([Kül97], p. 1286). It follows from a strong invariance principle for partial sum processes for $\mathbb{R}^k$-valued independent random variables, whose proof can be found in [Rio93], Cor. 4 on p. 1712. $\square$

7 Properties of the effective energy

In the following unlabelled sums with Latin index have range $\{1,\ldots,d\}$ and Greek index means range $\{1,\ldots,M\}$. We will always assume $\beta > 1$, which means we are in the low temperature regime. Denote by $m^* \equiv m^*(\beta)$ the unique positive solution of the transcendental equation

$$m = \tanh(\beta m). \tag{7.0.1}$$

We use now that $f_{\beta,\lambda}$ is a C^∞-function from $(-1,1)^d$ to $\mathbb{R}$. Since $\mathbb{E}\lambda = 0$ and λ fulfills a law of large numbers for $N \to \infty$, we first discuss the deterministic function f_β. Some of the proofs of the following statements are postponed to section 9.

7.1 Critical points of f_β

Using $I'(y) = \operatorname{artanh}(y)$ we obtain

$$\frac{d}{dy_k} f_\beta(y) = \frac{1}{\beta d}\operatorname{artanh}(y_k) - \frac{1}{d}(Py)_k. \tag{7.1.1}$$

The zeros of this functions are the solutions of the mean field equation

$$\tanh\left[\beta (Py)_k\right] = y_k. \tag{7.1.2}$$

In other words we are searching for fixed points of the mapping

$$y \mapsto \operatorname{Tanh}(\beta P y), \quad y \in [-1,1]^d \tag{7.1.3}$$

where $\operatorname{Tanh}(y) := (\tanh y_1, \tanh y_2, \ldots, \tanh y_d)$.

An important result in Koch and Piasko [KP89] describes the so-called "symmetric solutions of order n" of this equation for $n > 0$ (the case $n = 0$ corresponds to the trivial solution $y = 0$).

A symmetric solution of order n can be obtained by making the Ansatz $Py = a_n v^{(n)}$ and

$$v^{(n)} := \textstyle\sum_\alpha c_\alpha b^\alpha, \quad c_\alpha \in \{-1,0,1\}, \quad |c|^2 = n. \tag{7.1.4}$$

This Ansatz leads to the following equation for a_n :

$$a_n = 2^{-n+1} \sum_{0 \le m < n/2} \binom{n}{m} \frac{n-2m}{n} \tanh\left[(n-2m)\beta a_n\right]. \tag{7.1.5}$$

For $\beta \leq 1$, it is easy to see that equation (7.1.2) admits only the trivial solution, and that f_β takes its minimum value for $y = 0$. This minimum turns into a local maximum as β is increased past its critical value $\beta = 1$, and the remaining $3^M - 1$ symmetric solutions bifurcate away from the origin.

Definition 7.1 Let $f \in C^2\left(\mathbb{R}^d\right)$ be given. We call a critical point of f quadratic iff $\det \nabla^2 f \neq 0$. Otherwise it is called degenerate. A quadratic critical point of f, say x, is a k-saddle, iff $\nabla^2 f(x)$ has exactly k negative eigenvalues. We say, the function f is at x in k directions *unstable* and in $d-k$ directions *stable*. The $d \times d$ matrix $\nabla^2 f(x)$ is said to have *signature* k.

Theorem 7.2 *(Koch, Piasko) (a) Let $\beta > 1$ and $n \in \mathbb{N}$ be given, then equation (7.1.5) has a unique positive solution $a_n \equiv a_n(\beta)$. Furthermore, if $v^{(n)}$ satisfies (7.1.4) and $y^{(n)} \in \mathbb{R}^d$ is defined by*

$$y_k^{(n)} := \tanh\left[\beta a_n v^{(n)}\right], \quad 1 \leq k \leq d\ , \tag{7.1.6}$$

then the function f_β has a critical point at $y^{(n)}$.

(b) Let $1 < \beta < 1 + \left(9d + 500M^8\right)^{-1}$ and $y \in \mathbb{R}^d$. If f_β has a critical point at y, then y is a symmetric solution of some order $n \leq p$. In particular,

if y is a local minimum of f_β, then y is a symmetric solution of order 1, and if y is a 1-saddle of f_β, then y is a symmetric solution of order 2.

Proof. ad (a). This is Theorem 1.3, p. 907 in [KP89].

ad (b). The first part is Theorem 1.4 (i), p. 908 in [KP89] and we only have to show the second. Define the map

$$\Omega_\beta : [-1,1]^d \to P\left([-1,1]^d\right) \tag{7.1.7}$$

by

$$\Omega_\beta(x) := P\mathrm{Tanh}(\beta P x), \quad x \in [-1,1]^d. \tag{7.1.8}$$

Denote by P_1 and P_3 the orthogonal projections on $\mathbb{R}^d$ onto the subspaces $\mathrm{span}\{v_n\}$ and $\mathrm{span}\{b^\mu \mid \langle v_n, b^\mu\rangle = 0\}$, respectively and let $P_2 := P - P_1 - P_3$. It has been shown in [KP89], that the linearisation of Ω_β at the point $z^{(n)} := a_n(\beta)\, v_n$ has the following spectral representation

$$D\Omega_\beta\left(z^{(n)}\right) = \left(s_n + (n-1)\, r_n\right) P_1 + \left(s_n - r_n\right) P_2 + s_n P_3. \tag{7.1.9}$$

Here, s_n and r_n are given by the equations

$$s_n = \beta - \beta\frac{1}{d}\sum_k \tanh^2\left(\beta z_k^{(n)}\right) \tag{7.1.10}$$

$$r_n = -\beta\frac{1}{d}c_\mu c_\nu \sum_k \tanh^2\left(\beta z_k^{(n)}\right) b_k^\mu b_k^\nu, \tag{7.1.11}$$

where $\mu \neq \nu$ are arbitrary numbers between one and M such that $c_\mu c_\nu \neq 0$. Since $s_n + (n-1)\, r_n < 1$, $D\Omega_\beta$ can have an eigenvalue greater one only if $s_n - r_n > 1$.

As Koch and Piasko[KP89] pointed out (compare formula (3.5) on p. 917), we obtain

$$\nabla^2 f_\beta\left(\mathrm{Tanh}\left(\beta z^{(n)}\right)\right) = \frac{1}{d\beta}\left(\mathbb{1} - \beta P \mathrm{Tanh}'\left(\beta z^{(n)}\right)\right)\left(\mathrm{Tanh}'\left(\beta z^{(n)}\right)\cdot\right)^{-1}, \tag{7.1.12}$$

where $\mathrm{Tanh}'\left(\beta z^{(n)}\right)\cdot$ denotes the diagonal matrix with entries given by the vector. Since it is a positive definite matrix the signature of the matrix

$$\nabla^2 f_\beta\left(\mathrm{Tanh}\left(\beta z^{(n)}\right)\right) \tag{7.1.13}$$

coincides with the signature of $\left(\mathbb{1} - \beta P \mathrm{Tanh}'\left(\beta z^{(n)}\right)\right)$.

If $s_n - r_n < 1$, we know that all eigenvalues of $\nabla^2 g_\beta$ are positive and z is a minimum and for $s_n - r_n > 1$, we obtain at least $\dim\left(P_2\mathbb{R}^d\right)$ negative eigenvalues. Therefore only the points $y^{(2)}$ can be 1-saddles.

For $n = 2$, we obtain $s_2 = \beta\left(1 - \frac{1}{2}m^{*2}\right)$ and $r_2 = -\frac{1}{2}m^{*2}$. Hence, $s_2 - r_2 = \beta > 1$. Let β_s denote the unique solution of the equation

$$\beta = \frac{2}{2 - m^*\left(\beta\right)^2}. \tag{7.1.14}$$

Then the eigenvalue s_n gets bigger than 1 at β_s. Therefore $y^{(2)}$ is a 1-saddle of f_β only in the temperature interval $(1, \beta_s)$. □

Corollary 7.3 *Let $1 < \beta < 1 + \left(9d + 500M^8\right)^{-1}$. We define a vertex-set*

$$V := \{-M, \ldots, M\} \setminus \{0\} \tag{7.1.15}$$

and an edgeset

$$E := \{\{\mu, \nu\} \in V \times V \mid \mu \notin \{\nu, -\nu\}\}. \tag{7.1.16}$$

We denote $\overline{m}_{\pm\mu} := \pm m^*(\beta)\, b^\mu$ *and* $\overline{s}_{\mu,\pm\nu} := \frac{1}{2} m^*(b^\mu \pm b^\nu)$. *Then*

$$\overline{\mathcal{M}}_N := \{\overline{m}_\mu \,|\, \mu \in V\} \tag{7.1.17}$$

is the set of local minima of f_β *and*

$$\overline{S}_N := \{\overline{s}_{\mu\nu} \,|\, \{\mu,\nu\} \in E\} \tag{7.1.18}$$

is the set of 1*-saddles of* f_β.

In the following we will use another result of Koch and Piasko [KP89], p. 919, namely

Proposition 7.4 *Let* $I = \{1, 2, \ldots, p\}$ *and for every subset* $J \subset I$ *define*

$$b_k^J := \prod\nolimits_{\mu \in J} b_k^\mu, \; 1 \leq k \leq d, \tag{7.1.19}$$

where the value of an empty product is defined to be 1. *Then the set* $\left\{b^J : J \subset I\right\}$ *is an orthogonal basis for* $\mathbb{R}^d$.

We introduce some abbreviations: denote

$$\gamma_1 := \frac{1}{\beta(1 - m^{*2})}, \tag{7.1.20}$$

$$\gamma_2 := \frac{1}{\beta}, \tag{7.1.21}$$

$$\gamma_3 := \frac{1}{2}\left(\gamma_1 + \gamma_2 - 1 + \sqrt{1 + (\gamma_1 - \gamma_2)^2}\right), \tag{7.1.22}$$

$$\gamma_4 := \frac{1}{2}\left(\gamma_1 + \gamma_2 - 1 - \sqrt{1 + (\gamma_1 - \gamma_2)^2}\right). \tag{7.1.23}$$

For $J \subset I \equiv \{1, \ldots, M\}$ define $u^J \in \mathbb{R}^d$ by

$$u^J = b^J (1 + b^\mu b^\nu) \tag{7.1.24}$$

and

$$\tilde{u}^J := b^J (1 - b^\mu b^\nu) \tag{7.1.25}$$

and the mixtures

$$v^\alpha := \frac{1}{\gamma_1 - \gamma_3} u^\alpha + \frac{1}{\gamma_2 - \gamma_3} \tilde{u}^\alpha \tag{7.1.26}$$

and

$$\tilde{v}^\alpha := \frac{1}{\gamma_1 - \gamma_4} u^\alpha + \frac{1}{\gamma_2 - \gamma_4} \tilde{u}^\alpha. \tag{7.1.27}$$

Now we can formulate

Proposition 7.5 *(a) The points $\overline{m}_\mu$ are minima for all $\beta > 1$ and the Hessian $d\nabla^2 f_\beta(\overline{m}_\mu)$ has eigenvectors b^α with eigenvalue $\gamma_1 - 1$ for $1 \le \alpha \le M$ and eigenvectors b^J with eigenvalue γ_1, where $J \subset \{1, \dots, M\}$ such that $|J| \ne 1$.*

(b) The points $\overline{s}_{\mu\nu}$ are 1-saddles for $1 < \beta < \beta_s$, where β_s is the unique solution of the equation $\beta = \frac{2}{2-m^*(\beta)^2}$. The corresponding eigenvalues of the Hessian $d\nabla^2 f_\beta(\overline{s}_{\mu\nu})$ are

eigenvalue	*multiplicity*	*eigenvector*
γ_1	$\frac{1}{2}d - M + 1$	u^J for $\|J\| \ne 1$
γ_2	$\frac{1}{2}d - M + 1$	$\tilde{u}^J$ for $\|J\|, \|J \setminus \{\mu, \nu\}\| \ne 1$
γ_3	$M - 2$	v^α for $\alpha \notin \{\mu, \nu\}$
γ_4	$M - 2$	$\tilde{v}^\alpha$ for $\alpha \notin \{\mu, \nu\}$
$\gamma_1 - 1$	1	$b^\mu + b^\nu$
$\gamma_2 - 1$	1	$b^\mu - b^\nu$

Proof. We use the representation (7.1.12) of $\nabla^2 f_\beta$ at a symmetric solution.

For the symmetric solution of order 1, we have simply

$$\mathrm{Tanh}'\left(\beta z^{(1)}\right) \cdot = \left(1 - m^{*2}\right) \mathbb{1}. \tag{7.1.28}$$

Therefore $\left\{b^J\right\}_{J \subset \{1,\dots,M\}}$ is a basis of eigenvectors for $\left(P\mathrm{Tanh}'\left(\beta z^{(1)}\right) \cdot\right)$ with eigenvalues $\lambda_J = 1 - m^{*2}$ if $|J| = 1$ and $\lambda_J = 0$ if $|J| \ne 1$. This leads to part (a).

ad (b). We consider without restriction of generality $z^{(2)} = \frac{1}{2}m^*\left(b^1 + b^2\right)$. The matrix $\left(\mathrm{Tanh}'\left(\beta z^{(2)}\right) \cdot\right)$ has the representation

$$\left(\mathrm{Tanh}'\left(\beta z^{(2)}\right) \cdot\right) = \left(1 - \frac{1}{2}m^{*2}\right) \mathbb{1} - \frac{1}{2}m^*\left(b^{\{1,2\}} \cdot\right). \tag{7.1.29}$$

Hence, we have a connection between pairs of vectors (u, v) like $(1, b^{12})$ and (b^{13}, b^{23}) that are related by $v = ub^{\{1,2\}}$, as well as $u = vb^{\{1,2\}}$.

We define

$$a^\alpha := b^\alpha b^{\{1,2\}} + \frac{m^{*2}}{2 - m^{*2}} b^\alpha. \tag{7.1.30}$$

The representation (7.1.29) yields a basis of eigenvectors of $\left(P\mathrm{Tanh}'\left(\beta z^{(2)}\right) \cdot\right)$, namely $(b^1 - b^2)$ with eigenvalue 1, $(b^1 + b^2)$ with eigenvalue $(1 - m^{*2})$, as well as $(M-2)$ eigenvectors b^α with eigenvalue $\left(1-\frac{1}{2}m^{*2}\right)$ and a^α with eigenvalue 0, for all $\alpha \in \{3,\ldots,M\}$. Moreover, there are $(d-2M+2)$ eigenvectors of the form b^J where $|J|, |J\setminus\{1,2\}| \neq 1$. All of these have eigenvalue 0.

Due to equation (7.1.29) the matrix $\left(\mathrm{Tanh}'\left(\beta z^{(2)}\right)\cdot\right)$ has for this basis of eigenvectors a block diagonal representation, namely two single valued entries, 1 and $(1-m^{*2})$, associated to b^1-b^2 and b^1+b^2. Then, for $\alpha \in \{3,\ldots,M\}$, associated to (b^α, a^α) there are blocks of the form

$$\begin{pmatrix} \frac{2-2m^{*2}+m^{*4}}{2-m^{*2}} & -\frac{1}{2}m^{*2} \\ -\frac{2m^{*2}\left(1-m^{*2}\right)}{\left(2-m^{*2}\right)^2} & \frac{2\left(1-m^{*2}\right)}{2-m^{*2}} \end{pmatrix}, \tag{7.1.31}$$

followed by $\left(\frac{1}{2}d-M+1\right)$ blocks of form

$$\begin{pmatrix} 1-\frac{1}{2}m^{*2} & -\frac{1}{2}m^{*2} \\ -\frac{1}{2}m^{*2} & 1-\frac{1}{2}m^{*2} \end{pmatrix}, \tag{7.1.32}$$

which are associated to pairs $\left(b^J, b^J b^{\{1,2\}}\right)$, where $|J|, |J\setminus\{1,2\}| \neq 1$. Diagonalising the inverted blocks multiplied from the left with the associated 2×2 blocks of the diagonal matrix $\frac{1}{\beta}\left(1-\beta P \mathrm{Tanh}'\left(\beta z^{(2)}\right)\right)$ leads to the statement of the proposition. $\square$

7.2 Precise critical points and barrier

Theorem 7.6 *Denote* $\gamma_1 \equiv \frac{1}{\beta(1-m^{*2})}$, $a_1 \equiv \frac{m^*}{\gamma_1 - 1}$ *and* $a_2 \equiv \frac{1}{2}\frac{m^*\beta}{1-\beta\left(1-\frac{1}{2}m^{*2}\right)}$. *Then for all* $\xi \in \Xi$ *and* $N \geq N_0[\xi]$, *the function* $f_{\beta,\lambda}$ *has* $2M$ *deepest minima, namely*

$$m_{\pm\mu} = \pm m^* \left(b^\mu + \frac{a_1}{\sqrt{N}} \sum_\alpha A_N^{\mu\alpha} b^\alpha\right) + \mathcal{O}\left(\frac{\ln N}{N}\right). \tag{7.2.1}$$

For $1 < \beta < 1+(9d+500M^8)^{-1}$ *it has exactly* $\left(\binom{2M}{2} - M\right)$ *1-saddles, namely*

$$\begin{aligned} s_{\mu,\pm\nu} &:= \overline{s}_{\mu,\pm\nu}\left(1 \mp 2a_1a_2\frac{1}{\sqrt{N}}A_N^{\mu\nu}\right) + \\ &\quad + a_2\frac{1}{\sqrt{N}}\left(\mathbb{1} - \overline{S}^2_{\mu,\pm\nu}\right)\sum_\alpha \left(A_N^{\mu,\alpha} \pm A_N^{\nu,\alpha}\right) b^\alpha + \\ &\quad + \mathcal{O}\left(\frac{\ln N}{N}\right) \end{aligned} \tag{7.2.2}$$

for $\mu \neq \nu \in \{1,\ldots,M\}$ *and* $s_{-\mu,\pm\nu} := -s_{\mu,\mp\nu}$. *Here,* $\overline{S}_{\mu,\pm\nu}$ *denotes the diagonal matrix with entries given by the vector* $\overline{s}_{\mu,\pm\nu}$.

The proof of this theorem will be given in section 9. Very similar results as in Theorem 7.6 are already known, compare for the precise location of the minimising order parameters e.g. [Gen96], Theorem. 1.1, p. 246.

In the next proposition we give an explicit estimate of the random heights of these minima and 1-saddles. We define the following constants:

$$k_0 := a_1\left(\frac{1}{2\beta}\ln\frac{1+m^*}{1-m^*} - \frac{1}{2}m^*(\gamma_1+2)\right), \tag{7.2.3}$$

$$k_1 := a_1\left(m^*\gamma_1 - \frac{1}{4\beta}\ln\frac{1+m^*}{1-m^*}\right) - \frac{1}{2\beta}I(m^*), \tag{7.2.4}$$

$$k_2 := \left(1-m^{*2}\right)a_2\left(\frac{1}{4}m^*(\gamma_1+2) - \frac{1}{4\beta}\ln\frac{1+m^*}{1-m^*}\right), \tag{7.2.5}$$

$$k_3 := m^*a_1a_2\left(\frac{1}{4}m^*(\gamma_1+2) + \frac{1}{4\beta}\ln\frac{1+m^*}{1-m^*}\right). \tag{7.2.6}$$

Here, we have used the constants γ_1, a_1, a_2 as in Theorem 7.6. Observe that $k_0 \in \left(-\frac{1}{2}, \frac{3}{4}\right)$ and $k_1 \in (0,1)$ and k_2, k_3 have a singularity at β_s, which is the unique solution of $\beta = \frac{2}{2-m^*(\beta)^2}$.

We denote the free energy of the Curie-Weiss model by

$$f_{CW}(\beta) := \frac{1}{2}m^{*2} - \frac{1}{\beta}I(m^*). \tag{7.2.7}$$

Proposition 7.7 *For all* $\xi \in \Xi$ *and* $N \geq N_0[\xi]$, *the explicit representation of* $f_{\beta,\lambda}$ *at the minima and the saddle points is given by*

$$\begin{aligned} &f_{\beta,\lambda}(\pm m_\mu) \\ &= -f_{CW}(\beta) + \frac{k_0}{N}\left(A_N^2\right)^{\mu,\mu} + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right)^3 \end{aligned} \tag{7.2.8}$$

and

$$
\begin{aligned}
& f_{\beta,\lambda}\left(s_{\mu\nu}\right) \\
& = -\frac{1}{2} f_{CW}\left(\beta\right) - \frac{k_1}{\sqrt{N}} A_N^{\mu,\nu} + \\
& \quad - \frac{k_2}{N} \sum_{\alpha} \left(A_N^{\alpha\mu} + A_N^{\alpha\nu}\right)^2 - \frac{k_3}{N} \left(A_N^{\mu\nu}\right)^2 + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right)^3 .
\end{aligned} \tag{7.2.9}
$$

To obtain $f_{\beta,\lambda}\left(s_{\mu,-\nu}\right)$ we have to substitute $A_N^{\alpha\nu}$ by $-A_N^{\alpha\nu}$ for all $\alpha \in 1, \ldots, M$ in equation (7.2.9).

Remark 7.8 Let g_N be a random walk in $\mathscr{A}$, the space of symmetric $M \times M$ matrices with vanishing diagonal as introduced in Proposition 6.11(b). Since we can approximate A_N by g_N we see that the height of the minima of $\mathcal{H}_N$ varies only of order $\mathcal{O}\left(1\right)$ times a chi-square (with M degrees of freedom) distributed random variable. The height of the saddles fluctuates of order $\mathcal{O}\left(\sqrt{N}\right)$ times a normal random variable plus terms of higher order.

8 Structure of the ground states

8.1 Eyring-Kramers formula

In this subsection we put together the ingredients to arrive at an Eyring-Kramers formula for the Hopfield model. We consider the (random) set of deepest minima of $f_{\beta,\lambda}$ by

$$\mathcal{M}_N := \{m_\mu \,|\, \mu \in V\}\,. \tag{8.1.1}$$

We will use the following Theorem of Lidskii (compare Kato [Kat76], Theorem. 6.10, p. 126).

Proposition 8.1 *[Lidskii, 1950] Let A and B be symmetric $d \times d$-matrices and $C = B - A$. Denote respectively by α_k, β_k and γ_k, $k \in \{1, \ldots, d\}$ the repeated eigenvalues of A, B and C. Then the d-dimensional numerical vector $(\beta_1 - \alpha_1, \ldots, \beta_d - \alpha_d)$ lies in the convex hull of the vectors obtained from $(\gamma_1, \ldots, \gamma_d)$ by all possible permutations of its elements.*

Moreover, we need the following

Definition 8.2 Consider $\delta \in (0, \frac{1}{2}$. Denote $\overline{E} := \left\{\{\mu, \nu\} \in \overline{V} \times \overline{V} \,|\, \mu \neq \nu\right\}$. We define a random set of "good" numbers

$$J_\delta := \left\{ n \in \mathbb{N} \,|\, \min_{a \neq b \in \overline{E}} \left(A_n^a - A_n^b\right) \geq n^{-\frac{1}{2}+\delta} \right\}. \tag{8.1.2}$$

As we have seen in Proposition 6.11 (c) we can think of the components of A_N (up to the symmetry) to be independent random walk, hence a number is not good if two of them come to close together. The next lemma shows that almost surely the most $n \in \mathbb{N}$ are "good".

Lemma 8.3 *We define*

$$Z'_\delta := \left\{ \xi \in \Xi \,\Big|\, \lim_{N \uparrow \infty} \frac{1}{N} \sum_{n=1}^{N} 1\,(n \in J_\delta) = 1 \right\}. \tag{8.1.3}$$

Then $\mathbb{P}\left(Z'_\delta\right) = 1$.

Proof. We define $\eta^{ab} = \left(\eta_i^{ab}\right)_{i\in\mathbb{N}}$ by $\eta_i^{ab} := \frac{1}{\sqrt{N}}\left(\xi_i^{a_1}\xi_i^{a_2} - \xi_i^{b_1}\xi_i^{b_2}\right)$ for $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$. Then η^{ab} is a sequence of centered i.i.d. random variables with finite variance and $S_n^{ab} = \frac{1}{\sqrt{N}}\sum_{i=1}^n \eta_i^{ab}$ their normalised partial sum and $S_n^{ab} = A_n^a - A_n^b$. Apparently $\left\{S_n^{ab}\right\}_{a,b\in\overline{E}}$ are identically distributed; let S_n be another random variable with the same distribution. Then

$$\mathbb{P}\left(\min_{a,b\in\overline{E}} S_n^{ab} \geq n^{-\frac{1}{2}+\delta}\right) \leq \binom{M}{2}\mathbb{P}\left(S_n \geq n^{-\frac{1}{2}+\delta}\right). \tag{8.1.4}$$

Therefore we can use Lemma 3, in [Kül97], p. 1279. □

We want to control the expected time $\zeta_{N,\beta}$ needs to get from one minimum in $\mathcal{M}_N$ to another one. Since $\zeta_{N,\beta}$ is for each realisation of the patterns a reversible Markov chain on a (compact) subset of a lattice we can apply the Eyring-Kramers formula in the form proved in Theorem 5.21 of part II.

We incorporate the notions of Definition 4.3 of part II. In particular, we denote the communication height between two subsets $I, J \subset \mathcal{M}_N$ by $\hat{f}_{\beta,\lambda}(I, J)$. The associated set of relevant saddle points is named $S_N(I, J)$. Recall the notion of valley of Definition 4.4 of part II. Finally the barrier between $m \in \mathcal{M}_N \setminus I$ and I is defined as

$$b_N(m, I) := \beta(\hat{f}_{\beta,\lambda}(m, I) - f_{\beta,\lambda}(m). \tag{8.1.5}$$

Theorem 8.4 *We assume* $1 < \beta < 1 + \left(9d + 500M^8\right)^{-1}$. *Choose* $\delta \in (0, \frac{1}{2})$ *and assume* $\xi \in Z'_\delta$ *and* $N \geq N_0[\xi]$, *as well as* $N \in J_\delta$. *Let* I *and* J *be disjoint subsets of* $\mathcal{M}_N$. *Assume* $c_{N,k} \equiv 1$, *i.e. we consider (Glauber) Metropolis dynamics for the original Hopfield Markov chain.*

If $s \in S_N(I, J)$ *is a relevant saddle point between* I *and* J *we obtain*

$$\begin{aligned}\operatorname{cap}(I, J) &= k_4|S_N(I, J)|N^{(d-2)/2}\varrho_{N,\beta}(s)\times \\ &\quad\times\left(1 + \mathcal{O}\left(\sqrt{\ln^3 N}/\sqrt{N}\right)\right),\end{aligned} \tag{8.1.6}$$

where

$$k_4 := \frac{\sqrt{\beta - 1}\left(1 - m^{*2}\right)^{d/4}\left(2\pi\beta d\right)^{d/2}}{\pi\sqrt{1 - \beta\left(1 - m^{*2}\right)}\left(1 - \beta\left(1 - \frac{1}{2}m^{*2}\right)\right)^{(M-2)/2}}. \tag{8.1.7}$$

Starting in $m \in \mathcal{M}_N \backslash I$ the expected (quenched) hitting time of J satisfies

$$\begin{aligned} \mathrm{E}_m(\tau_I) &= \frac{k_5 N}{|S_N(m,J)|} \sum_{n \in V_J(m)} \exp(N b_N(n,J)) \times \\ &\quad \times \left(1 + \mathcal{O}\left(\sqrt{\ln^3 N}/\sqrt{N}\right)\right), \end{aligned} \tag{8.1.8}$$

where

$$k_5 := \frac{\pi\left(1-m^{*2}\right)^{d/4}}{\sqrt{\beta-1}\sqrt{1-\beta\left(1-\frac{1}{2}m^{*2}\right)}} \left(\frac{1-\beta\left(1-\frac{1}{2}m^{*2}\right)}{1-\beta\left(1-m^{*2}\right)}\right)^{(M-1)/2}. \tag{8.1.9}$$

Remark 8.5 (a) If we do not specialise to the Metropolis dynamics, we have to multiply k_4 by

$$\sqrt{\frac{\beta d|\gamma|}{\beta-1}}, \tag{8.1.10}$$

where γ is the unique negative eigenvalue of the matrix

$$\left(a_i \left(\nabla^2 f_\beta\right)_{ij}\right) \tag{8.1.11}$$

with

$$a_k := \begin{cases} (1-m^*)\, c_{N,k}(s) & \text{for } k \in U_{\mu\nu}, \\ c_{N,k}(s) & \text{for } k \notin U_{\mu\nu}. \end{cases} \tag{8.1.12}$$

Here, $U_{\mu\nu} := \{k \in \{1,\ldots,d\} \mid b_k^\mu = b_k^\nu\}$.

k_5 has to be divided by the same quantity.

Whenever $c_{N,k}(x)$ depends only on $\pi(x)$ and x_k, this will yield , up to a constant factor, again the result (8.1.6).

(b) The validity of this theorem could possibly be extended to $\beta \in (1, \beta_s)$, where β_s is the unique solution of the equation $\beta = \frac{2}{2-m^*(\beta)^2}$. Outside this interval the points $\{\pm s^{\mu,\pm\nu}\}$ are no longer candidates for the relevant saddle points and therefore there has to be others, which however are unknown up to now.

Proof. We choose $\epsilon := \frac{1}{N}$ and $F_\epsilon := \beta f_{\beta,\lambda}$ as well as $F := \beta f_\beta$. Proposition 6.9 shows that condition **sF1** is satisfied and **F2** holds since the statespace is relatively compact and f_β is continuous. Since

$$\beta \in \left(1, 1 + \left(9d + 500M^8\right)^{-1}\right), \tag{8.1.13}$$

Theorem 7.2 implies that there are only finitely many candidates for essential saddle points of $f_{\beta,\lambda}$. Therefore conditions **S1**-**S3** of section 5 of part II are satisfied. Assumption **D** assures that conditions **D1** and **D2** of chapter 5 of part II are satisfied.

Hence we can apply Theorem 5.7 to estimate the capacity. This yields for $\operatorname{cap}(I,J)$ up to multiplicative errors $\left(1+\mathcal{O}\left(\sqrt{\ln^3 N}/\sqrt{N}\right)\right)$ the value

$$(2\pi\beta N)^{(d/2-1)}\frac{|\gamma|}{\sqrt{|\det\nabla^2 f_{\beta,\lambda}(s)|}}\varrho_{N,\beta}(s), \tag{8.1.14}$$

where γ is the unique negative eigenvalue of $L^{-2}p_N(s)\cdot\nabla^2 f_{\beta,\lambda}(s)$.

Moreover, Proposition 7.7 shows that the fluctuations of the minima are small compared to the fluctuations of the 1-saddles. Since we assumed that $N\in J_\delta$ we see that the additional condition of Theorem 5.21 is satisfied. This yields for $\mathrm{E}_m\left(\tau_{\mathcal{M}_N\setminus m}\right)$ up to multiplicative errors $\left(1+\mathcal{O}\left(\sqrt{\ln^3 N}/\sqrt{N}\right)\right)$ the value

$$\frac{\pi}{2\beta N|\gamma|}\frac{\sqrt{|\det\nabla^2 f_{\beta,\lambda}(s)|}}{\sqrt{\det\nabla^2 f_{\beta,\lambda}(m)}}\sum_{n\in V_J(m)}\exp\left(Nb_N(n,J)\right). \tag{8.1.15}$$

Now we show that we can estimate the prefactor explicitly. From Proposition 6.2 we obtain

$$\begin{aligned}&p_{N,k}(s)\\&=\frac{\ell_k}{2N}c_{N,k}(s)\left((1-s_k)\wedge\left(1+s_k+\frac{2}{\ell_k}\right)\frac{\varrho\left(s+\frac{2}{\ell_k}e_k\right)}{\varrho(s)}\right).\end{aligned} \tag{8.1.16}$$

Since $c_{N,k}$ is Lipschitz, this shows that also $p_{N,k}$ is Lipschitz continuous. The representation (6.3.2) of ℓ yields:

$$\begin{aligned}\ell_k &= \frac{N}{d}\left(1+\frac{1}{\sqrt{N}}\lambda_k\right)\\ &= \frac{N}{d}\left(1+\mathcal{O}\left(\frac{\sqrt{\ln N}}{\sqrt{N}}\right)\right).\end{aligned} \tag{8.1.17}$$

Theorem 7.6 yields $s=\overline{s}\left(1+\mathcal{O}\left(\ln N/\sqrt{N}\right)\right)$. Therefore

$$(1-s_k)\wedge\left(1+s_k+\frac{2}{\ell_k}\right)=(1-|\overline{s}_k|)\left(1+\mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right)\right). \tag{8.1.18}$$

Since $f_{\beta,\lambda} \in C^{\infty}(\mathbb{R})$ and $\nabla f_{\beta,\lambda}(s) = 0$ we obtain

$$f_{\beta,\lambda}\left(s + \frac{2}{\ell_k} e_k\right) = f_{\beta,\lambda}(s) + \mathcal{O}\left(N^{-2}\right). \tag{8.1.19}$$

Hence,

$$\varrho\left(s + \frac{2}{\ell_k} e_k\right) = \varrho(s)\left(1 + \mathcal{O}\left(N^{-1}\right)\right). \tag{8.1.20}$$

Altogether we obtain

$$p_k(s) = \frac{1}{2d} a_k \left(1 + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right)\right) \tag{8.1.21}$$

where $a \in \mathbb{R}^d$ is defined by 8.1.12.

With the formula (9.0.2) for the Hessian of $f_{\beta,\lambda}$ we obtain, since $\xi \in \Xi$ and $N \geq N_0[\xi]$

$$\nabla^2 f_{\beta,\lambda}(s) = \nabla^2 f_{\beta}(s)\left(1 + \mathcal{O}\left(\frac{\sqrt{\ln N}}{\sqrt{N}}\right)\right). \tag{8.1.22}$$

Now we apply the Theorem of Lidskii (Proposition 8.1). Therefore the deviation of the eigenvalues of $\nabla^2 f_{\beta,\lambda}(s)$ compared with the eigenvalues of $\nabla^2 f_{\beta}(\overline{s}_{\mu\nu})$ are of order $\mathcal{O}\left(\ln N/\sqrt{N}\right)$. In the same way we can relate the eigenvalues of $\nabla^2 f_{\beta,\lambda}(m_\mu)$ to $\nabla^2 f_{\beta}(\overline{m}_\mu)$ and of $L^{-2} p_N(s) \cdot \nabla^2 f_{\beta,\lambda}(s)$ to $\frac{d}{2N^2}\left(a_i\left(\nabla^2 f_{\beta}(\overline{s})\right)_{ij}\right)$.

In the case of the Metropolis algorithm (6.2.20), we have $c_{N,k}(x) \equiv 1$. The only (normed) eigenvector with negative eigenvalue of $\nabla^2 f_{\beta}(\overline{s})$ is $v := \frac{1}{\sqrt{2d}}\left(b^{\mu} - b^{\nu}\right)$ and hence $v_k = 0$ for $k \in U_{\mu\nu}$ (compare Proposition 7.5). Therefore using (8.1.12) the unique negative eigenvalue of $\frac{d}{2N^2}\left(a_i\left(\nabla^2 f_{\beta}(\overline{s})\right)_{ij}\right)$ is $\left(-\frac{(\beta-1)}{2N^2\beta}\right)$ and the associated eigenvector is v.

If $c_{N,k}(x)$ depends only on $\varrho(x)$ and x_k then a_k is constant for $k \notin U_{\mu\nu}$. Therefore, up to a constant factor, the result of Theorem 8.4 holds also in these cases.

The remaining part of the prefactor in (8.1.15) can be approximated using the estimate (8.1.22) and the Theorem of Lidskii by

$$\frac{\left|\det \nabla^2 f_{\beta,\lambda}(s)\right|}{\det \nabla^2 f_{\beta,\lambda}(m)} = \frac{\left|\det \nabla^2 f_{\beta,\lambda}(s)\right|}{\det \nabla^2 f_{\beta,\lambda}(m)}\left(1 + \mathcal{O}\left(\frac{\sqrt{\ln N}}{\sqrt{N}}\right)\right). \tag{8.1.23}$$

This does not depend any more on μ or ν and we obtain from Proposition 7.5 the following explicit form: Denote $c \equiv \frac{\gamma_3\gamma_4}{\gamma_1\gamma_2} = 1-\beta\left(1-\frac{1}{2}m^{*2}\right)$, then

$$\begin{aligned}
&\frac{\left|\det \nabla^2 f_\beta\left(\overline{s}_{\mu,\nu}\right)\right|}{\det \nabla^2 f_\beta\left(\overline{m}_\mu\right)} \\
&= \frac{\left(\gamma_1\gamma_2\right)^{d/2-M+1}\left(\gamma_3\gamma_4\right)^{M-2}\left(\gamma_1-1\right)\left(1-\gamma_2\right)}{\gamma_1^{d-M}\left(\gamma_1-1\right)^M} \\
&= \left(\beta\gamma_1\right)^{-d/2}\left(\frac{c\gamma_1}{\gamma_1-1}\right)^{M-1}\frac{1-\gamma_2}{c\gamma_2} \\
&= \frac{\left(\beta-1\right)\left(1-m^{*2}\right)^{d/2}}{1-\beta\left(1-\frac{1}{2}m^{*2}\right)}\left(\frac{1-\beta\left(1-\frac{1}{2}m^{*2}\right)}{1-\beta\left(1-m^{*2}\right)}\right)^{M-1}. \qquad (8.1.24)
\end{aligned}$$

The prefactor for the capacity can be estimated analogously. $\square$

Remark 8.6 If $\zeta_{N,\beta}$ has the transition probability matrix q_N defined by (6.2.28), we receive

$$g_k\left(x\right) = \frac{\ell_k}{4N}\left(1-x_k\right)\left(1+\tanh\frac{\beta d}{N}\left(\left(PLx\right)_k+\frac{M}{d}\right)\right). \qquad (8.1.25)$$

For $x = \overline{s}_{\mu\nu} + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right)$ this equals (since $\xi \in \Xi$)

$$\frac{d}{N}\left(PLx\right)_k = \left(P\left(1+\frac{1}{\sqrt{N}}\Lambda\right)x\right)_k = \overline{s}_{\mu\nu} + \mathcal{O}\left(\delta\right). \qquad (8.1.26)$$

Therefore $c_{N,k}\left(x\right) \equiv \frac{1}{2}$. Hence the expected value of the hitting time is 2 times the value of (8.4).

Proposition 8.7 *Denote by* $v := \overline{m}_\mu - \overline{s}_{\mu\nu}$. *Then for all* $0<t<1$

$$\nabla f_\beta\left(\overline{s}_{\mu\nu}+tv\right) \upuparrows v, \qquad (8.1.27)$$

i.e. the gradient of f_β *points along the connecting line between* $\overline{s}_{\mu\nu}$ *and* $\overline{m}_\mu$ *towards* $\overline{m}_\mu$.

Proof. We directly compute the gradient of f_β using the definition of m^*. We obtain

$$
\begin{aligned}
&\partial_k f_\beta\left(\overline{s}_{\mu\nu}+tv\right)\\
&=\frac{1}{\beta}\operatorname{artanh}\left(\overline{s}_{\mu\nu}+tv\right)-P\left(\overline{s}_{\mu\nu}+tv\right)_k\\
&=\left(\frac{1}{\beta m^*}\operatorname{artanh}\left(m^*\left(1-t\right)\right)-\left(1-t\right)\right)v.
\end{aligned}
\tag{8.1.28}
$$

The last identity follows, since

$$
\overline{s}_{\mu\nu,k}+tv_k=\begin{cases} m^* b_k^\mu, & k\in U\\ m^*\left(1-t\right)b_k^\mu, & k\notin U,\end{cases}
\tag{8.1.29}
$$

where $U:=\{k\in\{1,\dots,d\}\mid b_k^\mu=b_k^\nu\}$. Moreover, again with the definition of m^* follows for all $t\in(0,1)$

$$
a\left(t\right)\equiv\frac{1}{\beta m^*}\operatorname{artanh}\left(m^*\left(1-t\right)\right)-\left(1-t\right)>0.
\tag{8.1.30}
$$

Hence

$$
\nabla f_\beta\left(\overline{s}_{\mu\nu}+tv\right)=a\left(t\right)v.
\tag{8.1.31}
$$

$\square$

8.2 Random graphs

In this section we estimate the communication height between two minima of $f_{\beta,\lambda}$. Theorem 7.7 gives an explicit expression for the height of the 1-saddles $(s_{\mu,\nu})$. We see that for large N their order statistic is given by the order statistic of the standard Gaussians $(g_N^{\mu\nu})$ constructed in Proposition 6.11.

To give deterministic bounds for the communication height we consider the undirected weighted graph (V,E,g) with V,E as introduced in Corollary 7.3. The weights are defined by

$$
g\left(\mu,\nu\right):=\operatorname{sign}(\mu)\operatorname{sign}(\nu)g_N^{|\mu||\nu|}\quad\text{for }\mu,\nu\in E,
\tag{8.2.1}
$$

reflecting the structure of the heights of the 1-saddles, $f_{\beta,\lambda}\left(s_{\mu\nu}\right)$.

Let $\mathcal{T}_{\min}$ a minimal spanning tree of (V, E, g). We define the unique set

$$\overline{\mathcal{T}}_{\min} := \{t \in E \,|\, t \in \mathcal{T}_{\min} \text{ or } -t \in \mathcal{T}_{\min}\} . \tag{8.2.2}$$

Apparently set of essential saddle points is then

$$\mathcal{E}_\epsilon = \{s_{\mu\nu} \,|\, \mu, \nu \in \overline{\mathcal{T}}\}. \tag{8.2.3}$$

Taking into account the symmetry of the weights $g(\mu, \nu)$ it is enough to consider a simpler graph to estimate the communication heights. We identify μ and $-\mu$ and associate the weight $|g(\mu, \nu)|$ to the edge $\{\mu, \nu\}$. Then we get the vertex set

$$\overline{V} = \{1, \ldots, M\} \tag{8.2.4}$$

and the edgeset

$$\overline{E} = \left\{\{\mu, \nu\} \in \overline{V} \times \overline{V} \,|\, \mu \neq \nu\right\} . \tag{8.2.5}$$

Now let us introduce some notions of random graph theory, see e.g. [Bol01]. Let $p \in (0, 1)$ be given. $\mathscr{G}(n, p)$ is the set of all graphs G with n vertices such that each possible edge has independent probability p to be in G. In other words, if G is a graph with m edges, then

$$P_p(G) = p^m (1 - p)^{\binom{n}{2} - m} . \tag{8.2.6}$$

We write P_p and E_p to emphasise that the probability and expectation are taken in $\mathscr{G}(n, p)$.

We will use the following theorem proved by Erdös and Rényi in 1959. It gives a threshold value for the probability $p(n)$ such that asymptotically for $n \to \infty$ almost all graphs out of $\mathscr{G}(n, p(n))$ are connected.

Theorem 8.8 *(Erdös and Rényi) Let $c \in \mathbb{R}$ be fixed and*

$$p(n) := \frac{1}{n} (\log n + c + o(1)) . \tag{8.2.7}$$

Let $G \in \mathscr{G}(n, p(n))$ be a random graph. Then

$$P_{p(n)} (G \textit{ is connected}) \to \exp\left(-e^{-c}\right) \tag{8.2.8}$$

for $n \to \infty$.

The proof of this theorem can be found in Bollobas [Bol01], Theorem 7.3, p. 164.

We consider again the weighted graph $\left(\overline{V}, \overline{E}, g\right)$. By regarding only the edges with height bigger or equal to a given number x_M, we obtain a random graph. This graph is an element of $\mathscr{G}(M, p(M))$ with $p(M) := P(|g| \geq x_M)$.

All edges associated to essential saddle points are included in the maximal spanning tree of $\left(\overline{V}, \overline{E}, |g_N|\right)$. Hence we are searching for the minimal $p(M)$ such that asymptotically a.s. all the graphs in $\mathscr{G}(M, p(M))$ are connected.

Theorem 8.9 *Let $\xi \in \Xi$ and $N \geq N_0[\xi]$ and assume $1 < \beta < 1 + (9d + 500M^8)^{-1}$. Then asymptotically almost surely (for $M \to \infty$), the communication height between two elements of $\mathcal{M}_N$, say m and n, can be estimated by*

$$\hat{f}_{\beta,\lambda}(m, n) \leq \frac{1}{2} f_{CW}(\beta) - \frac{k_1}{\sqrt{N}} \sqrt{2 \ln M}. \tag{8.2.9}$$

Proof. Proposition 7.7 yields an estimate of $f_{\beta,\lambda}(s_{\{}\mu\nu\})$, which involves the standard Gaussian random variable $|g_N^{\mu,\nu}|$ for $\{\mu, \nu\} \in \overline{E}$. Theorem 8.8 implies that a random graph, $G_{M,p(M)}$ with edge probability $p(M) := P(|g| \geq x_M)$ is almost surely for $M \to \infty$ a connected graph, if

$$p(M) = \frac{1}{M}(\ln M + c(M)), \tag{8.2.10}$$

where $c(M) \to \infty$ for $M \to \infty$. Since we have for $x > 0$ the bound (compare [Fel66], p. 175)

$$P(|g| \geq x) \geq \left(\frac{1}{x} - \frac{1}{x^3}\right) \sqrt{\frac{2}{\pi}} \exp\left(-\frac{1}{2} x^2\right), \tag{8.2.11}$$

we obtain for $M > 20$ that

$$x_M = \sqrt{\ln \frac{2M^2}{\pi} - 3 \ln_2 \frac{2M^2}{\pi}} + o(1). \tag{8.2.12}$$

satisfies condition (8.2.10). Hence almost surely for $M \uparrow \infty$ every essential saddle point satisfies the inequality (8.2.9).

We prove now that this result holds also for the original graph (V, E, g_N). Let G be a subgraph of (V, E, g_N) that leads to a connected subgraph $\overline{G}$ of $(\overline{V}, \overline{E}, |g|)$. By definition $G = G_1 \cup G_2$ such that $\{m, n\} \in G_1$ iff $\{-m, -n\} \in$

G_2. Now every edge that does not belong to the maximal spanning tree of $(\overline{V}, \overline{E}, |g|)$ has equal probability to connect either two vertices out of the same subgraph or a vertex of G_1 with one of G_2. Hence the probability that G_1 and G_2 are not connected is $1 - \left(\frac{1}{2}\right)^{n\left(\frac{1}{2}\ln n - 1\right)}$, which converges to zero exponentielly fast. $\square$

8.3 Low lying eigenvalues of the generator

In this section we consider the generator of the Markov chain $\zeta_{N,\beta}$ that is defined by $L_{N,\beta} := p_{N,\beta} - 1$. We abbreviate $L \equiv L_{N,\beta}$. Due to the reversibility of $\zeta_{N,\beta}$, L is a negative operator in $\ell_2(\varrho)$, i.e. it is symmetric and has only negative eigenvalues. By 'low lying' eigenvalues of L we mean eigenvalues with small absolute value.

Let $D \subset \mathcal{X}_N$. We say that $\lambda \in \mathbb{C}$ is an eigenvalue of the Dirichlet operator L^D if the equation

$$\begin{cases} Lf(x) = \lambda f(x), & x \in D^c, \\ f(x) = 0, & x \in D \end{cases} \tag{8.3.1}$$

has a non-zero solution $f_{D,\lambda}$. The solution $f_{D,\lambda}$ is called eigenfunction of L^D. Let λ_D denote the smallest eigenvalue of L^D.

We assume again $1 < \beta < 1 + (9d + 500M^8)^{-1}$. The validity of the statements in this section could possibly be extended to $\beta \in (1, \beta_s)$, where β_s is the unique solution of the equation $\beta = \frac{2}{2 - m^*(\beta)^2}$.

Since $\mathcal{M}_N$ is the complete set of local minimal of $\mathcal{H}_N$ on $\mathcal{X}_{N,\beta}$ and has constant cardinality $|\mathcal{M}_N| = 2M$, we know that $\zeta_{N,\beta}$ behaves metastable with respect to $\mathcal{M}_N$ (in the sense of Definition 4.5 of part II, see Example 4.10). It is already known that therefore $-L$ has $2M$ eigenvalues that are exponentially small in N, and all other eigenvalues are at most polynomially small in N, see e.g. [BEGK02], Theorem 1.3, p. 222.

There exists a classical bound for the low lying eigenvalues of the Generator of a diffusion process proved by Donsker and Varadhan in 1976 [DV76]. By analogue arguments it can be proved (see [BEGK02], Lemma 4.2, p. 236)

Proposition 8.10 *For every nonempty subset $J \subset \mathcal{M}_N$ we have*

$$\lambda_J \geq \left(\sup_{x \notin J} E_x(\tau_J) \right)^{-1}. \tag{8.3.2}$$

We will show that the low lying eigenvalues of L have indeed a similar structure.

We define the *entrance time of $\zeta_{N,\beta}$ into the set A* as

$$\sigma_A := \min\{t \geq 0 \,|\, \zeta_{N,\beta}(t) \in A\}. \tag{8.3.3}$$

Observe that σ_A differs from the hitting time τ_A since it takes the value 0 if $\zeta_{N,\beta}(0) \in A$.

As we have seen in Proposition 3.4 of part II the *equilibrium potential* of $\zeta_{N,\beta}$ with respect to disjoint subsets $A, B \subset \mathcal{X}_N$ satisfies

$$h_{A,B}(x) = P_x(\sigma_A < \sigma_B) \quad \text{for all } x \in \mathcal{X}_N. \tag{8.3.4}$$

We use the abbreviation

$$h_\mu := h_{m_\mu, \mathcal{M}_N \setminus m_\mu} \quad \text{for all } \mu \in V. \tag{8.3.5}$$

Now we can state the crucial proposition that allows us to control the low lying eigenvalues of L :

Proposition 8.11 *([Bovier, Gayrard, Klein]) Assume that $\zeta_{N,\beta}$ is κ-metastable with respect to $\mathcal{M}_N$. Let λ be one of the $2M$ smallest eigenvalues of $-L$, then there exists an eigenvalue γ of the $2M \times 2M$-matrix $\mathcal{K} \equiv \mathcal{K}_{N,\beta}$ whose elements are given by*

$$\mathcal{K}_{\mu\nu} = -\frac{\langle h_\mu, L h_\nu \rangle_\varrho}{\|h_\mu\|_\varrho \|h_\nu\|_\varrho} \tag{8.3.6}$$

such that $\lambda = \gamma(1 + \mathcal{O}(\kappa))$. We call $\mathcal{K}$ the capacity matrix of $\zeta_{N,\beta}$.

Proof. The proof can be found in [Bov04] Theorem 5.1, p. 36. Compare also Section 4 of [BGK05]. □

Remark 8.12 (a) To motivate the name, recall that the capacity between m_μ and $\mathcal{M}_N \setminus m_\mu$ is given by $\text{cap}(m_\mu, \mathcal{M}_N \setminus m_\mu) = -\langle h_\mu, L h_\mu \rangle_\varrho$.

(b) The row sum of denominators of $\mathcal{K}$ is zero, i.e.

$$\sum_{\nu \in V} \langle h_\mu, L h_\nu \rangle_\varrho = 0. \tag{8.3.7}$$

This follows, since L is a linear operator, for all $x \in \mathcal{X}_N$

$$\sum_{\nu \in V} h_\nu(x) = \sum_{\nu \in V} P_x\left(\tau_{m_\nu} \le \tau_{\mathcal{M}_N}\right) = 1 \tag{8.3.8}$$

and $L1 = 0$.

Due to the symmetry under total spin flip and the unusually small fluctuations of the heights of the minima in $\mathcal{M}_N$ we cannot directly use the results of [BEGK02] or [BGK05], but we can apply similar methods. Let $\{\lambda_0, \ldots, \lambda_{2M-1}\}$ with $0 = \lambda_0 \le \ldots \le \lambda_{2M-1}$ be the smallest eigenvalues of the generator $-L_{N,\beta}$ of the Hopfield model.

Proposition 8.13 *We assume $1 < \beta < 1 + \left(9d + 500M^8\right)^{-1}$. Choose $\delta \in \left(0, \frac{1}{2}\right)$ and assume $\xi \in Z'_\delta$ and $N \ge N_0[\xi]$, as well as $N \in J_\delta$. We define $\{\mu, \nu\} \in E$ to be such that*

$$A_N^{\mu\nu} = -\max_{a \in \overline{E}} \left|A_N^a\right|. \tag{8.3.9}$$

Then the two largest eigenvalues of $\mathcal{K}$ are equal, i.e.

$$\lambda_{2M-1} = \lambda_{2M-2} \tag{8.3.10}$$

and there exists a constant c such that their value can be estimated as

$$\left(E_{m_\mu}\left(\tau_{\mathcal{M}_N \setminus m_\mu}\right)^{-1} + E_{m_\mu}\left(\tau_{\mathcal{M}_N \setminus m_\nu}\right)^{-1}\right)\left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right). \tag{8.3.11}$$

All other eigenvalues of $\mathcal{K}$ satisfy

$$\lambda \le 2Me^{-cN^{-\delta}}\lambda_{2M-1}. \tag{8.3.12}$$

Remark 8.14 This yields together with Theorem 8.4 an explicit formula for λ_{2M-1} with multiplicative errors $\left(1 + \mathcal{O}\left(\sqrt{\ln^3 N}/\sqrt{N}\right)\right)$.

Proof. Recall that

$$E_{m_\mu}\left(\tau_{\mathcal{M}_N \setminus m_\mu}\right) = \frac{\varrho\left(h_\mu\right)}{\operatorname{cap}\left(m_\mu, \mathcal{M}_N \setminus m_\mu\right)}. \tag{8.3.13}$$

First we investigate the quantities $\|h_\mu\|_\varrho$. We can approximate, as in the proof of Theorem 5.21 of part II,

$$\|h_\mu\|_\varrho^2 = k_5 N^{d/2} \varrho\left(m_\mu\right)\left(1 + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right)\right), \tag{8.3.14}$$

where

$$\begin{aligned} k_5 : &= (2\pi)^{d/2} / \sqrt{\det \nabla^2 f_\beta \left(\overline{m}_\mu\right)} \\ &= (2\pi d\beta)^{d/2} \frac{\left(1 - m^{*2}\right)^{d/2}}{\left(1 - \beta\left(1 - m^{*2}\right)\right)^{M/2}}. \end{aligned} \tag{8.3.15}$$

Therefore Theorem 7.7 implies

$$\begin{aligned} \frac{\|h_\mu\|_\varrho}{\|h_\nu\|_\varrho} &= \exp\left\{k_0\beta\left(\left(g_N^2\right)^{\mu\mu} - \left(g_N^2\right)^{\nu\nu}\right)\right\} \times \\ &\quad \times \left(1 + \mathcal{O}\left(\frac{\ln^3 N}{\sqrt{N}}\right)\right). \end{aligned} \tag{8.3.16}$$

Since $g_N \in \mathcal{A}_M$ and

$$\left\{g_N^{\mu\nu}\right\} \sim \mathcal{N}_{0,1} \quad \forall \{\mu, \nu\} \in E, \tag{8.3.17}$$

the quotient $\frac{\|h_\mu\|_\varrho}{\|h_\nu\|_\varrho}$ is of order $\mathcal{O}(1)$ for $N \to \infty$ (and M finite).

Due to the assumptions $\xi \in Z'_\delta$ and $N \notin J_\delta$, we have

$$A_N^{\mu\nu} \geq A_N^{ab} + N^{-\frac{1}{2}+\delta} \quad \text{for all } 1 \leq a < b \leq M. \tag{8.3.18}$$

In the following we use a modification of the argument for the proof of Proposition 7.12 and Theorem 7.13 in [Bov04]. We denote

$$G_{\mu\nu} := \{\mu, \nu\}^2 \cup \{-\mu, -\nu\}^2. \tag{8.3.19}$$

Now, we investigate the matrix $\hat{\mathcal{K}}$ given by

$$\hat{\mathcal{K}}_{xy} := \begin{cases} \mathcal{K}_{xy}, & \{x, y\} \in G_{\mu\nu} \\ 0, & \text{else.} \end{cases} \tag{8.3.20}$$

Thereafter we show that the capacity matrix $\mathcal{K}$ is a perturbation of $\hat{\mathcal{K}}$.

We claim that the non-zero part of $\hat{\mathcal{K}}$ has the structure

$$\left(\hat{\mathcal{K}}_{xy}\right)_{\{x,y\} \in \{\mu,\nu\}^2} = A\mathcal{K}_{\mu\mu}\left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right), \tag{8.3.21}$$

where we have denoted

$$A := \begin{pmatrix} 1 & -a \\ -a & a^2 \end{pmatrix} \qquad \text{with } a := \frac{\|h_\mu\|_\varrho}{\|h_\nu\|_\varrho}. \tag{8.3.22}$$

Due to the spin-flip symmetry

$$\left(\hat{\mathcal{K}}_{xy}\right)_{\{x,y\}\in\{-\mu,-\nu\}^2} = \left(\hat{\mathcal{K}}_{xy}\right)_{\{x,y\}\in\{\mu,\nu\}^2}. \tag{8.3.23}$$

Equality (8.3.21) holds true, because the property (8.3.9) of $\{\mu,\nu\}$ yields the following identity between sets of relevant saddle points

$$S_N\left(m_\mu, \mathcal{M}_N \setminus m_\mu\right) \;=\; S_N\left(m_\nu, \mathcal{M}_N \setminus m_\nu\right) = \{s_{\mu\nu}\}. \tag{8.3.24}$$

Moreover, $V_{\mathcal{M}_N\setminus\{m_\mu,m_\nu\}}\left(\{m_\mu,m_\nu\}\right)$ contains no other minima.

We distinguish three cases:

1. Assume $x \in A\left(m_\mu\right) \cup A\left(m_\nu\right)$ and $f_{\beta,\lambda}\left(x\right) \le f_{\beta,\lambda}\left(s_{\mu\nu}\right) + \frac{\delta}{\sqrt{N}}$. Then we obtain
$$P_x\left(\tau_{\{m_\mu,m_\nu\}} \le \tau_{\mathcal{M}_N}\right) = 1 - \mathcal{O}\left(e^{-cN^\delta}\right). \tag{8.3.25}$$
Therefore
$$\begin{aligned} h_\mu\left(x\right) \;&=\; P_x\left(\tau_{m_\mu} < \tau_{m_\nu} \,|\, \tau_{\{m_\mu,m_\nu\}} \le \tau_{\mathcal{M}_N}\right) \times \\ &\quad \times \left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right) \\ &=\; 1 - P_x\left(\tau_{m_\nu} < \tau_{m_\mu} \,|\, \tau_{\{m_\mu,m_\nu\}} \le \tau_{\mathcal{M}_N}\right) \times \\ &\quad \times \left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right) \\ &=\; \left(1 - h_\nu\left(x\right)\right)\left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right). \end{aligned} \tag{8.3.26}$$

2. Assume $x \notin A\left(m_\mu\right) \cup A\left(m_\nu\right)$ and $f_{\beta,\lambda}\left(x\right) \le f_{\beta,\lambda}\left(s_{\mu\nu}\right) + \frac{\delta}{\sqrt{N}}$. Then we obtain
$$h_\mu\left(x\right) = \mathcal{O}\left(e^{-cN^\delta}\right) = h_\nu\left(x\right). \tag{8.3.27}$$

3. For x such that $f_{\beta,\lambda}\left(x\right) > f_{\beta,\lambda}\left(s_{\mu\nu}\right) + \frac{\delta}{\sqrt{N}}$ we obtain
$$\varrho\left(x\right) < \exp\left(-\beta N f_{\beta,\lambda}\left(s_{\mu\nu}\right) - \delta\sqrt{N}\right). \tag{8.3.28}$$

Since $L1 = 0$ we conclude

$$\left\langle h_\nu, L h_\mu\right\rangle_\varrho = -\left\langle h_\mu, L h_\mu\right\rangle_\varrho \left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right) \tag{8.3.29}$$

and hence $\mathcal{K}_{\mu\nu} = -a\mathcal{K}_{\mu\mu}\left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right)$. Moreover,

$$\langle h_\nu, Lh_\nu \rangle_\varrho = \langle h_\mu, Lh_\mu \rangle_\varrho \left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right) \tag{8.3.30}$$

and hence $\mathcal{K}_{\nu\nu} = a\mathcal{K}_{\mu\mu}\left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right)$ and we have proved the representation (8.3.21).

We can say even more, since with Remark8.12 $\sum_{\nu \in V} \langle h_\mu, Lh_\nu \rangle = 0$ and therefore in particular

$$\mathcal{K}_{\alpha\beta} = \mathcal{K}_{\mu\mu}\mathcal{O}\left(\left(e^{-cN^\delta}\right)\right) \quad \forall \{\alpha, \beta\} \in \{\pm\mu, \pm\nu\}^2 \setminus G_{\mu\nu}. \tag{8.3.31}$$

The eigenvalues of A of (8.3.22) are 0 and $(1 + a^2)$. Therefore the largest eigenvalue of $\hat{\mathcal{K}}$ is

$$\hat{\lambda} = \mathcal{K}_{\mu\mu}\left(1 + \frac{\|h_\mu\|_\varrho^2}{\|h_\nu\|_\varrho^2}\right)\left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right) \tag{8.3.32}$$

and it has multiplicity two.

Now, we claim that $\mathcal{K}$ is a perturbation of $\hat{\mathcal{K}}$. For this purpose we write

$$\mathcal{K} = \hat{\mathcal{K}} + \check{\mathcal{K}} \tag{8.3.33}$$

To justify the claim we estimate the norm of $\check{\mathcal{K}}$. We take as matrix norm the Euclidean norm in $\mathbb{R}^{4M^2}$. We observe

$$\mathcal{K}_{xx} = \frac{\operatorname{cap}(m_x, \mathcal{M}_N \setminus m_x)}{\|h_x\|_\varrho^2}. \tag{8.3.34}$$

Therefore with Theorem 8.4 we obtain for all $N \geq N_0[\xi]$

$$\mathcal{K}_{\mu\mu} \geq e^{cN^\delta} \max_{x \in V \setminus \{\pm\mu, \pm\nu\}} \mathcal{K}_{xx}. \tag{8.3.35}$$

For $x \neq y$ we obtain like in [Bov04] by the Cauchy-Schwarz inequality that $\mathcal{K}_{xy}^2 \leq \mathcal{K}_{xx}\mathcal{K}_{yy}$. This and the estimate in (8.3.31) implies

$$\left\|\check{\mathcal{K}}\right\| \leq 2Me^{-cN^\delta} \max(\mathcal{K}_{\mu\mu}, \mathcal{K}_{\nu\nu}). \tag{8.3.36}$$

With the result (8.3.36) follows that the biggest eigenvalue of $\mathcal{K}$ and $\hat{K}$ coincide up to multiplicative errors $\left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right)$. $\square$

We recall that we have introduced the minimal spanning tree $\mathcal{T}_{\min}$ of the weighted graph (V, E, g) in the last section.

We enumerate the edges of this tree $\mathcal{T}_{\min}$ by $(t_1, \ldots, t_{2M-1})$ such that $g(t_{2M-1}) \leq g(t_{2M-2}) \leq \ldots \leq g(t_1) < 0$. Notice that (up to the order and sometimes choice of equally weighted edges) the construction with Kruskal's algorithm starts with t_{2M-1} and adds along our enumeration edges to the spanning tree until it ends with t_1. Let $I_{\mathcal{T}} \subset \{1, \ldots, 2M-1\}$ denote the set of indices such that $g(t_i) < g(t_{i-1})$.

Theorem 8.15 *We assume* $1 < \beta < 1 + \left(9d + 500M^8\right)^{-1}$. *Choose* $\delta \in (0, \frac{1}{2})$ *and assume* $\xi \in Z'_\delta$ *and* $N \geq N_0[\xi]$, *as well as* $N \in J_\delta$.

Then there exists an increasing sequence $(\mathscr{M}_{i+1} \,|\, i \in I_{\mathcal{T}})$ *of metastable sets of* $\zeta_{N,\beta}$. *Let* $\{\lambda_0, \ldots, \lambda_{2M-1}\}$ *with* $0 = \lambda_0 \leq \ldots \leq \lambda_{2M-1}$ *be the smallest eigenvalues of the generator* $-L_{N,\beta}$ *of the transformed Markov chain* $\zeta_{N,\beta}$. *We define*

$$S_i^* := \arg \min_{\{m,n\} \in \mathscr{M}_i \times \mathscr{M}_i} \left(\hat{f}_{\beta,\lambda}(m,n)\right). \tag{8.3.37}$$

Denote for $m \in \mathscr{M}_i$

$$\gamma_{m,i} := \left(E_m\left(\tau_{\mathscr{M}_i \setminus m}\right)\right)^{-1}. \tag{8.3.38}$$

We distinguish three cases:

- *Assume* $S_i^* = \{\{m,n\}, \{-m,-n\}\}$, *then*

$$\lambda_{i-1} = \lambda_i = (\gamma_{m,i} + \gamma_{n,i})\left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right). \tag{8.3.39}$$

- *Assume* $S_i^* = \{\{m,n\}, \{-m,n\}\}$, *then*

$$\lambda_i = (2\gamma_{m,i} + \gamma_{n,i})\left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right) \tag{8.3.40}$$

  and

$$\lambda_{i-1} = \gamma_{m,i}\left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right). \tag{8.3.41}$$

- *Assume* $S_i^* = \{m,n\}$, *then*

$$\lambda_i = (\gamma_{m,i} + \gamma_{n,i})\left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right). \tag{8.3.42}$$

Remark 8.16 Combined with Theorem 8.4 this implies explicit estimates provided we know $(\mathcal{T}_i)_{1 \leq i < 2M}$.

Proof. For the proof we reduce step by step the cardinality of the set $\mathscr{M}_{2M} \equiv \mathcal{M}_N$ that is by definition described by the vertex-set of $\mathcal{T}_{\min}$.

We use now Proposition 4.6 of part II to find inductively smaller and smaller metastable sets $\mathscr{M}_i$ of the Markov chain $\zeta_{N,\beta}$. We define $\mathcal{K}^{(2M)} \equiv \mathcal{K}$ and $\mathcal{K}^{(i)}$ to be the capacity matrix of $\mathscr{M}_i$.

Assume we have already constructed $\mathscr{M}_i$ and the associated capacity matrix is $\mathcal{K}^{(i)}$. As in Definition 4.4 of part II, we define the valley $V_n(m)$ as the connected component of the set $\left\{x \in \mathcal{X}_{N,\beta} \,|\, f_{\beta,\lambda}(x) < \hat{f}_{\beta,\lambda}(m,n)\right\}$ that contains m.

Assume $\{m,n\} \in S_i^*$, then from Definition 8.3.37 follows that the only element of $\mathscr{M}_i$ that is contained in $V_n(m)$ is m itself. Analogously $V_m(n)$ contains only n. Therefore we can conclude that

$$s_{\alpha\beta} \in S_N(m,n)\,, \tag{8.3.43}$$

where $\{\alpha, \beta\}$ are minimiser of

$$\min_{\substack{\mu\,:\, m_\mu \in V_n(m) \cap \mathcal{M}_N \\ \nu\,:\, m_\nu \in V_m(n) \cap \mathcal{M}_N}} \langle \xi^\mu, \xi^\nu \rangle\,. \tag{8.3.44}$$

- If $S_i^* = \{\{m,n\}, \{-m,-n\}\}$ (this has to be the case for $\mathscr{M}_{2M}$), where $\varrho(m) \le \varrho(n)$, we put $\mathscr{M}_{i-2} := \mathscr{M}_i \setminus \{m,-m\}$. Observe that in this case there is no metastable set of $\zeta_{N,\beta}$ with $(i-1)$ elements.

- If $S_i^* = \{m,n\}$, where $\varrho(m) \le \varrho(n)$, we have tie edges and we put $\mathscr{M}_{i-1} = \mathscr{M}_i \setminus m$. Observe that in particular n may be equal to $-m$ in this case.

- And third, if $S_i^* = \{\{m,n\}, \{-m,n\}\}$, we distinguish two cases. If $\varrho(m) \le \varrho(n)$, we put again $\mathscr{M}_{i-2} := \mathscr{M}_i \setminus \{m,-m\}$. If on the other hand $\varrho(n) \le \varrho(m)$, we put $\mathscr{M}_{i-2} := \mathscr{M}_i \setminus \{m,n\}$.

Hence $\mathscr{M}_i$ contains exactly i points. Assume $\mathscr{M}_i = \{m_1, \dots, m_i\}$ and denote $h_{i,x} := h_{m_x, \mathcal{M}_i \setminus m_x}$. Then

$$\mathcal{K}_{xy}^{(i)} = \frac{\langle h_{i,x}, L h_{i,y} \rangle}{\|h_{i,x}\|_\varrho \|h_{i,y}\|_\varrho}. \tag{8.3.45}$$

We abbreviate $c \equiv \frac{\|h_{i,1}\|_\varrho}{\|h_{i,2}\|_\varrho}$. We show now that analogously to the proof of Proposition 8.13 we can treat $\mathcal{K}^{(i)}$ as a perturbation of a simpler matrix. This matrix depends now on the structure of $\mathcal{T}_i$. In all three cases the estimate of the smallness of this perturbation is shown exactly as in the proof of Proposition 8.13.

- Assume $S_i^* = \{\{m_1, m_2, \}, \{-m_1, -m_2\}\}$, possibly by renumbering the elements of $\mathscr{M}_i$. Then we can describe $\mathcal{K}^{(i)}$ as a perturbation of the matrix $\mathcal{K}^{(V)}$ given by

$$\mathcal{K}^{(V)}_{xy} := \begin{cases} \mathcal{K}_{xy}, & \{x,y\} \in \{1,2\}^2 \cup \{-1,-2\}^2 \\ 0, & \text{else.} \end{cases} \tag{8.3.46}$$

  Moreover, this matrix has the following structure:

$$\left(\mathcal{K}^{(U)}_{xy}\right)_{\{x,y\}\in\{1,2\}^2} = \mathcal{K}_{11} A \left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right), \tag{8.3.47}$$

  where we denoted

$$A := \begin{pmatrix} 1 & -c \\ -c & c^2 \end{pmatrix}. \tag{8.3.48}$$

  Due to the spin-flip symmetry

$$\left(\mathcal{K}^{(V)}_{xy}\right)_{\{x,y\}\in\{-1,-2\}^2} = \left(\mathcal{K}^{(V)}_{xy}\right)_{\{x,y\}\in\{1,2\}^2}. \tag{8.3.49}$$

  The eigenvalues of A are $\{0, (1+c^2)\}$ and therefore the largest eigenvalue of $\mathcal{K}^{(V)}$ has multiplicity two and is, up to multiplicative errors $\left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right)$, equal to

$$\operatorname{cap}(m_1, m_2) \left(\frac{1}{\|h_{i,1}\|_\varrho^2} + \frac{1}{\|h_{i,2}\|_\varrho^2}\right). \tag{8.3.50}$$

- Assume $S_i^* = \{m_1, m_2\}$, again by renumbering the elements of $\mathcal{M}_i$. Then we can describe $\mathcal{K}_i$ as a perturbation of the matrix $\mathcal{K}^{(U)}$ given by

$$\mathcal{K}^{(U)}_{xy} := \begin{cases} \mathcal{K}_{xy}, & \{x,y\} \in \{1,2\}^2 \\ 0, & \text{else.} \end{cases} \tag{8.3.51}$$

  Moreover, this matrix has the following structure:

$$\left(\mathcal{K}^{(U)}_{xy}\right)_{\{x,y\}\in\{1,2\}^2} = \mathcal{K}_{11} A \left(1 + \mathcal{O}\left(e^{-cN^\delta}\right)\right) \tag{8.3.52}$$

with A defined in (8.3.48). Therefore the unique largest eigenvalue of $\mathcal{K}^{(U)}$ is, up to multiplicative errors $\left(1+\mathcal{O}\left(e^{-cN^\delta}\right)\right)$, equal to

$$\operatorname{cap}(m_1, m_2)\left(\frac{1}{\|h_{i,1}\|_\varrho^2}+\frac{1}{\|h_{i,2}\|_\varrho^2}\right). \tag{8.3.53}$$

Observe that (only) in this case the set of relevant saddle points between m_1 and m_2 contains two elements.

- The last possible case is $S_i^* = \{\{m_1, m_2\}, \{-m_1, m_2\}\}$. Hence we can describe $\mathcal{K}_i$ as a perturbation of the matrix $\mathcal{K}^{(W)}$ given by

$$\mathcal{K}_{xy}^{(W)} := \begin{cases} \mathcal{K}_{xy}, & \{x,y\} \in \{1,2\}^2 \cup \{-1,2\}^2 \\ 0, & \text{else.} \end{cases} \tag{8.3.54}$$

  Moreover this matrix has the following structure:

$$\left(\mathcal{K}_{xy}^{(W)}\right)_{\{x,y\}\in\{1,2,-1\}^2} = \mathcal{K}_{11} C\left(1+\mathcal{O}\left(e^{-cN^\delta}\right)\right), \tag{8.3.55}$$

  where we have, due to the spin flip symmetry,

$$C := \begin{pmatrix} 1 & -c & 0 \\ -c & 2c^2 & -c \\ 0 & -c & 1 \end{pmatrix}. \tag{8.3.56}$$

  The eigenvalues of C are $\{0, 1, 1+2c^2\}$ and therefore the two largest eigenvalues of $\mathcal{K}^{(W)}$ are, up to multiplicative errors $\left(1+\mathcal{O}\left(e^{-cN^\delta}\right)\right)$, equal to

$$\operatorname{cap}(m_1, m_2)\left(\frac{1}{\|h_{i,1}\|_\varrho^2}+\frac{2}{\|h_{i,2}\|_\varrho^2}\right) \tag{8.3.57}$$

  and

$$\frac{\operatorname{cap}(m_1, m_2)}{\|h_{i,1}\|_\varrho^2}. \tag{8.3.58}$$

This proves the theorem. □

9 Some proofs

From now on we leave out the N-dependence of A_N and write $A \equiv A_N$. We will frequently need the first and second derivative of $f_{\beta,\lambda}$. We use $I'(y) = \operatorname{artanh}(y)$ and $I''(y) = \frac{1}{1-y^2}$. Thus

$$\frac{\partial}{\partial y_k} f_{\beta,\lambda}(y) = \left(1 + \frac{1}{\sqrt{N}}\lambda_k\right) \times \\ \times \left(\frac{1}{\beta}\operatorname{artanh}(y_k) - \sum_j P_{kj}\left(1 + \frac{1}{\sqrt{N}}\lambda_j\right) y_j\right) \tag{9.0.1}$$

and

$$\frac{\partial^2}{\partial y_j \partial y_k} f_{\beta,\lambda}(y) = \left(1 + \frac{1}{\sqrt{N}}\lambda_j\right) \times \\ \times \left(\frac{1}{\beta\left(1 - y_j^2\right)}\delta_{jk} - P_{jk}\left(1 + \frac{1}{\sqrt{N}}\lambda_k\right)\right). \tag{9.0.2}$$

9.1 Precise location of critical points

This subsection contains the proof of Theorem 7.6.

We consider $\xi \in \Xi$ and $N \geq N_0[\xi]$. Due to the uniform convergence of $f_{\beta,\lambda}$ to f_β (proved in Proposition 6.9) the cluster points of a sequence $\left(m^{(N)}\right)$ of global minima of $f_{\beta,\lambda}$ has to be contained in the set $\mathcal{M}$ of global minima of f_β. Therefore we can divide a given sequence into subsequences that converge to a global minimum of f_β. We show that if $m^{(N)}$ converges for $N \to \infty$ to $\overline{m}_{\pm\mu}$ then it is unique and has the form $m_{\pm\mu}$ given in Theorem 7.6.

Note that this follows already from the general theorem of Bovier and Gayrard [BG98], Theorem 6.2, p. 40, since here their $\beta_c(2) = 1$.

Assume that $\overline{y} \in P\mathbb{R}^d$ is a critical point of f_β. Now we perform the Ansatz $y := \overline{y} + \frac{1}{\sqrt{N}}\kappa$ to find a critical point of $f_{\beta,\lambda}$. Here $\kappa \equiv \kappa_N$ is an arbitrary random variable such that $|\kappa| = o\left(\sqrt{N}\right)$.

y is a critical point of $f_{\beta,\lambda}$ iff

$$\frac{d}{dy_k} f_{\beta,\lambda}(y) \overset{!}{=} 0. \tag{9.1.1}$$

Using (9.0.1) this is equivalent to

$$\frac{1}{\beta}\text{Artanh}\left(\overline{y}+\frac{1}{\sqrt{N}}\kappa\right) \quad = \quad P\left(\mathbb{1}+\frac{1}{\sqrt{N}}\Lambda\right)\left(\overline{y}+\frac{1}{\sqrt{N}}\kappa\right). \quad (9.1.2)$$

We were allowed to cancel the common factor $\left(1+\frac{1}{\sqrt{N}}\lambda_k\right)$, since $\xi \in \Xi$ and $N \geq N_0[\xi]$. We use now a Taylor expansion for arctanh $\left(\overline{y}+\frac{1}{\sqrt{N}}\kappa\right)$ and arrive at

$$\begin{aligned}&\frac{1}{\beta}\text{artanh}\left(\overline{y}_k\right)+\frac{1}{\beta\left(1-\overline{y}_k^2\right)}\frac{1}{\sqrt{N}}\kappa_k+\mathcal{O}\left(\frac{1}{N}\left|\kappa\right|^2\right)\\ &=\overline{y}_k+\frac{1}{\sqrt{N}}\sum_{\alpha}\left\langle b^\alpha,\Lambda y\right\rangle b_k^\alpha+\frac{1}{\sqrt{N}}P\left(\mathbb{1}+\frac{1}{\sqrt{N}}\Lambda\right)\kappa_k. \quad (9.1.3)\end{aligned}$$

Using that $\overline{y}$ is a critical point of f_β leads us to

$$\kappa_k=\beta\left(1-\overline{y}_k^2\right)\sum_{\alpha}\left(r_\alpha+\left\langle b^\alpha,\left(\mathbb{1}+\frac{1}{\sqrt{N}}\Lambda\right)\kappa\right\rangle\right)b_k^\alpha+\mathcal{O}\left(\frac{1}{N}\left|\kappa\right|^2\right), \quad (9.1.4)$$

where we denoted $r_\alpha := \langle b^\alpha, \Lambda y\rangle$. Now we multiply this equation with

$$b_k^\sigma\left(1+\frac{1}{\sqrt{N}}\lambda_k\right) \quad (9.1.5)$$

and sum over all $k \in \{1,\ldots,d\}$. This yields the matrix equation for $t \in \mathbb{R}^M$ with $t_\alpha := \left\langle b^\alpha, \left(\mathbb{1}+\frac{1}{\sqrt{N}}\Lambda\right)\kappa\right\rangle$. Moreover, t is of the same order as κ and therefore we obtain

$$\begin{cases}\kappa_k &= \beta\left(1-\overline{y}_k^2\right)\sum_\alpha\left(r_\alpha+t_\alpha\right)b_k^\alpha+\mathcal{O}\left(\frac{1}{N}\left|t\right|^2\right),\\ t &= G\left(r+t\right)+\mathcal{O}\left(\frac{1}{N}\left|t\right|^2\right),\end{cases} \quad (9.1.6)$$

where G is the $M \times M$ matrix with

$$G^{\sigma\alpha}=\beta\left\langle b^\sigma,\left(\mathbb{1}-\left(\overline{y}\cdot\right)^2\right)\left(\mathbb{1}+\frac{1}{\sqrt{N}}\Lambda\right)b^\alpha\right\rangle. \quad (9.1.7)$$

Here, $\overline{y}\cdot$ denotes the diagonal matrix with entries $\overline{y}$. To evaluate this further, we have to use specific information about the point $\overline{y}$.

Minima

Fix a $\mu \in \{1, \dots, M\}$. We will calculate the precise location of the minima, hence $\overline{y} = m^* b^\mu$. Denote again $\gamma_1 := \frac{1}{\beta(1-m^{*2})}$. We find $r = m^* A^\mu$ with $A^\mu \equiv (A^{\mu,1}, \dots, A^{\mu,p})$ and

$$G = \beta \left(1 - m^{*2}\right) \left(\mathbb{1} + \frac{1}{\sqrt{N}} A\right). \tag{9.1.8}$$

Therefore equation (9.1.6) equals

$$\begin{aligned} \gamma_1 t &= \left(\mathbb{1} + \frac{1}{\sqrt{N}} A\right)(t + r) + \mathcal{O}\left(\frac{1}{N}|t|^2\right) \\ \Leftrightarrow \left(\mathbb{1} - \frac{\hat{a}}{\sqrt{N}} A\right) t &= \hat{a}\left(\mathbb{1} + \frac{1}{\sqrt{N}} A\right) r + \mathcal{O}\left(\frac{1}{N}|t|^2\right), \end{aligned} \tag{9.1.9}$$

where $\hat{a} \equiv \frac{1}{\gamma_1 - 1}$. The matrix $\left(\mathbb{1} - \frac{\hat{a}}{\sqrt{N}} A\right)$ is invertible, since $\xi \in \Xi$ and $N \geq N_0[\xi]$, and therefore

$$\begin{aligned} \Leftrightarrow t &= \hat{a}\left(\mathbb{1} - \frac{\hat{a}}{\sqrt{N}} A\right)^{-1}\left(\mathbb{1} + \frac{1}{\sqrt{N}} A\right) r + \mathcal{O}\left(\frac{1}{N}|t|^2\right) \\ &= \hat{a} \sum_{n=0}^{\infty} \left(\frac{\hat{a}}{\sqrt{N}} A\right)^n \left(\mathbb{1} + \frac{1}{\sqrt{N}} A\right) r + \mathcal{O}\left(\frac{1}{N}|t|^2\right) \\ &= \hat{a} m^* A^\mu + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right). \end{aligned} \tag{9.1.10}$$

The last equation uses $r_\alpha = m^* A^{\mu,\alpha} = \mathcal{O}\left(\sqrt{\ln N}\right)$ for $\xi \in \Xi$ and $N \geq N_0[\xi]$. Inserting this in equation (9.1.6) gives us directly the correction κ for the minima m_μ :

$$\kappa_k = \frac{m^*}{\gamma_1}(1 + \hat{a}) \sum_\alpha A^{\mu,\alpha} b_k^\alpha + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right) \tag{9.1.11}$$

or equivalently

$$\kappa_k = \frac{m^*}{\gamma_1 - 1} \sum_\alpha A_N^{\mu,\alpha} b_k^\alpha + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right). \tag{9.1.12}$$

Saddle points

Let $\xi \in \Xi$ and $N \geq N_0[\xi]\,[\xi]$. To show that the essential saddle points of $f_{\beta,\lambda}$ are small deviations of the essential saddles of f_β, we use again the uniform convergence of Proposition 6.9: Let $m^{(N)}, n^{(N)} \in \mathcal{M}_N$. Let γ be an optimal path of f_β. Then γ can be uniformly approximated by paths γ_N in $\mathcal{X}_N$ and therefore by definition the communication height $\hat{f}_{\beta,\lambda}\left(m^{(N)}, n^{(N)}\right)$ converges to $\hat{f}_\beta\,(m,n)$. Let $S_N\,(m,n)$ be the set of all relevant saddle points between m and n in $\mathcal{X}_N$. Then each sequence in $S_N\left(m^{(N)}, n^{(N)}\right)$ has a subsequence that converges to an element of $S\,(m,n)$, the relevant saddles in $[-1,1]^d$. Now we show that $\left|S^N\left(m^{(N)}, n^{(N)}\right)\right| = |S\,(m,n)|$.

We use the abbreviation $\overline{s} := \overline{s}_{\mu,\nu}$.

Lemma 9.1 *We define the diagonal matrix S with elements $S_{kk} := \overline{s}_k$. Then the following properties hold:*

1. $S^2\overline{s} = m^{*2}\overline{s}$.
2. $\left\langle b^\alpha, S^2 b^\beta\right\rangle = \frac{1}{2}m^{*2}\left(\delta_{\alpha,\beta} \pm \delta_{\{\alpha,\beta\},\{\mu,\nu\}}\right)$.

Proof. ad 1. We have $\overline{s}_k = m^* b_k^\mu 1_{b_k^\mu = b_k^\nu}$. Therefore

$$\begin{aligned} S^2\overline{s}_k &= (\overline{s}_k)^3 = m^{*3} b_k^\mu 1_{e_k^\mu = e_k^\nu} \\ &= m^{*2}\overline{s}_k. \end{aligned} \tag{9.1.13}$$

ad 2. This follows directly from

$$S_k^2 = \overline{s}_k^2 = \frac{1}{2}m^{*2}\left(1 \pm b_k^\mu b_k^\nu\right). \tag{9.1.14}$$

□

We want to use again equation (9.1.6), now with $\overline{y} := \overline{s}_{\mu,\nu}$. Let B be the $M \times M$ matrix with $B^{\sigma\alpha} = A^{\sigma\alpha} - m^{*2}\frac{1}{d}\left\langle b^{\sigma\alpha}, \Lambda b^{\mu\nu}\right\rangle$. We receive $r_\alpha = \frac{1}{2}m^*\left(V^{\alpha\mu} + V^{\alpha\nu}\right)$ and with the help of Lemma 9.1 for $\sigma \neq \mu, \nu$

$$G^{\sigma\alpha} = \left(1 - \frac{1}{2}m^{*2}\right)\delta_{\sigma\alpha} + B^{\sigma\alpha}. \tag{9.1.15}$$

Thus equation (9.1.6) takes for $\sigma \neq \mu, \nu$ now the shape

$$t_\sigma = \beta\left(\left(1 - \frac{1}{2}m^{*2}\right)\mathbb{1} + \frac{1}{\sqrt{N}}B\right)(r+t)_\sigma. \tag{9.1.16}$$

With $c := \frac{\beta}{1-\beta\left(1-\frac{1}{2}m^{*2}\right)}$ this is equivalent to

$$\left(\mathbb{1} - \frac{c}{\sqrt{N}}B\right) t_\sigma = c\left(\left(1 - \frac{1}{2}m^{*2}\right)\mathbb{1} + \frac{1}{\sqrt{N}}B\right) r_\sigma. \tag{9.1.17}$$

The matrix $\left(\mathbb{1} - \frac{c}{\sqrt{N}}B\right)$ is invertible for $\xi \in \Xi$ and $N \geq N_0[\xi]$, and thus

$$\begin{aligned} t_\sigma &= c\left(\mathbb{1} - \frac{c}{\sqrt{N}}B\right)^{-1}\left(\left(1 - \frac{1}{2}m^{*2}\right)\mathbb{1} + \frac{1}{\sqrt{N}}B\right) r_\sigma \\ &= c\sum_{n=0}^{\infty}\left(\frac{c}{\sqrt{N}}B\right)^n\left(\left(1 - \frac{1}{2}m^{*2}\right)\mathbb{1} + \frac{1}{\sqrt{N}}B\right) r_\sigma \\ &= c\left(1 - \frac{1}{2}m^{*2}\right) r_\sigma + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right). \end{aligned} \tag{9.1.18}$$

We have used $B^{\sigma\alpha} = \mathcal{O}\left(\sqrt{\ln N}\right)$ and $r_\alpha = \mathcal{O}\left(\sqrt{\ln N}\right)$ from Proposition 6.7.

For $\sigma = \mu$ we get

$$\begin{aligned} t_\mu &= \beta\left(\left(1 - \frac{1}{2}m^{*2}\right)\mathbb{1} + \frac{1}{\sqrt{N}}B\right)(r+t)_\mu - \\ &\quad -\frac{1}{2}m^{*2}\beta\left(r_\nu + t_\nu\right) \end{aligned} \tag{9.1.19}$$

and analogues for $\sigma = \nu$

$$\begin{aligned} t_\nu &= \beta\left(\left(1 - \frac{1}{2}m^{*2}\right)\mathbb{1} + \frac{1}{\sqrt{N}}B\right)(r+t)_\nu - \\ &\quad -\frac{1}{2}m^{*2}\beta\left(r_\mu + t_\mu\right). \end{aligned} \tag{9.1.20}$$

Since $r_\mu = r_\nu$, we obtain

$$t_\mu - t_\nu = \beta\left(\mathbb{1} + \frac{1}{\sqrt{N}}B\right)(t_\mu - t_\nu) = 0, \tag{9.1.21}$$

because $\beta\left(1 + \frac{1}{\sqrt{N}}A\right) \neq 1$ for N large. Thus we can deduce

$$t_\mu = \beta\left(\left(1 - m^{*2}\right)\mathbb{1} + \frac{1}{\sqrt{N}}B\right)(r+t)_\mu \tag{9.1.22}$$

and, analogues to the derivation of (9.1.18), we obtain

$$t_\mu = \frac{1}{\gamma_1 - 1} r_\mu + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right). \tag{9.1.23}$$

With Equation (9.1.6) and $c\left(1 - \frac{1}{2}m^{*2}\right) + 1 = \frac{c}{\beta}$ we conclude

$$\begin{aligned}
\kappa_k &= \beta\left(1 - \overline{s}_k^2\right) \sum_\alpha \left(t_\alpha + r_\alpha\right) b_k^\alpha + \mathcal{O}\left(\frac{1}{N}\|\epsilon\|^2\right) \\
&= \left(1 - \overline{s}_k^2\right) \sum_\alpha \left(c 1_{\{\mu,\nu\}^c}(\alpha) + \frac{\beta\gamma_1}{\gamma_1 - 1} 1_{\{\mu,\nu\}}(\alpha)\right) r_\alpha b_k^\alpha + \\
&\quad + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right).
\end{aligned} \tag{9.1.24}$$

We will use that $\beta\frac{\gamma_1}{\gamma_1-1} - c = -\frac{1}{2}m^{*2}\frac{\hat{a}c}{1-m^{*2}}$ and $S^2\overline{s} = m^{*2}\overline{s}$ to derive from (9.1.24) another representation of ϵ:

$$\begin{aligned}
\epsilon_k &= c\left[\left(1 - \overline{s}_k^2\right) \sum_\alpha r_\alpha b_k^\alpha - m^* \hat{a} r_\mu \overline{s}_k\right] + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right) \\
&= \frac{m^* c}{2}\left(\left(1 - \overline{s}_k^2\right) \sum_\alpha \left(A_N^{\mu,\alpha} + A_N^{\nu,\alpha}\right) b_k^\alpha - 2m^* \hat{a} A_N^{\mu,\nu} \overline{s}_k\right) + \\
&\quad + \mathcal{O}\left(\frac{\ln N}{\sqrt{N}}\right).
\end{aligned} \tag{9.1.25}$$

9.2 Precise height of the minima and 1-saddles

We prove now Proposition 7.7 about the precise height of the minima $\{m_\mu\}_{1\le\mu\le M}$ and the 1-saddles $\{s_{\mu,\nu}\}_{1\le\mu,\nu\le M}$ between them.

Proof. We use the Taylor expansion of the logarithm to estimate the Cramér entropy term, defined in (6.3.4), and obtain for $v = \mathcal{O}\left(\sqrt{\ln N}\right)$:

$$\begin{aligned}
& I\left(u + \frac{1}{\sqrt{N}} v\right) \\
&= \frac{1}{2}\ln\left(1 - u^2\right) + \frac{1}{2}\left(u + \frac{1}{\sqrt{N}} v\right) \ln\frac{1+u}{1-u} + \\
&\quad + \frac{1}{2N}\frac{1}{1-u^2} v^2 + \mathcal{O}\left(\frac{\ln N}{N}\right)^{3/2}.
\end{aligned} \tag{9.2.1}$$

Now, we evaluate the function $f_{\beta,\lambda}$ defined in (6.3.10) at a point $z := x + \frac{1}{\sqrt{N}} y \in B_r(x)$ with $r = k\frac{\ln N}{\sqrt{N}}$:

$$\begin{aligned}
& f_{\beta,\lambda}(z) \\
&= \frac{1}{2\beta d}\sum_{k=1}^{d}\left(1+\frac{1}{\sqrt{N}}\lambda_k\right)\times \\
&\times\left(\ln\left(1-x_k^2\right)+z_k\ln\frac{1+x_k}{1-x_k}+\frac{1}{N}\frac{1}{1-x_k^2}y_k^2\right)- \\
&-\frac{1}{2d}\left|P\left(\mathbb{1}+\frac{1}{\sqrt{N}}\Lambda\right)z\right|^2+\mathcal{O}\left(\frac{\ln N}{N}\right)^{3/2}. \qquad (9.2.2)
\end{aligned}$$

9.2.1 Minima.

We denote $a \equiv \frac{m^*}{\gamma_1 - 1}$. First we consider the minimum m^μ, in other words we put $x \equiv m^* b^\mu$ and $y \equiv a\sum_\alpha A^{\mu\alpha} b^\alpha$ and use equation (9.2.2). In the following we use trace$(\Lambda) = 0$ and

$$\ln\frac{1+m^*b_k^\mu}{1-m^*b_k^\mu} = b_k^\mu \ln\frac{1+m^*}{1-m^*}. \qquad (9.2.3)$$

Hence equation (9.2.2) simplifies to

$$\begin{aligned}
& f_{\beta,\lambda}(m^\mu) \\
&= \frac{1}{2\beta}\ln\left(1-m^{*2}\right)+\frac{\gamma_1}{2dN}\langle y,y\rangle + \\
&+\frac{1}{2\beta}\ln\frac{1+m^*}{1-m^*}\mathrm{op}_{\mu\mu}-\frac{1}{2}\sum_\alpha\left(\mathrm{op}_{\mu\alpha}\right)^2+\mathcal{O}\left(\frac{\ln N}{N}\right)^{3/2}. \qquad (9.2.4)
\end{aligned}$$

Here, we used the overlap parameter at the minimum $\mathrm{op}_{\mu\alpha} := \frac{1}{N}\langle b^\alpha, Lm^\mu\rangle$. We obtain

$$\begin{aligned}
\mathrm{op}_{\mu\alpha} &= \frac{1}{d}\left\langle b^\alpha, \left(\mathbb{1}+\frac{1}{\sqrt{N}}\Lambda\right)\left(m^*b^\mu+\frac{a}{\sqrt{N}}\sum_\beta A^{\mu\beta}b^\beta\right)\right\rangle \\
&= m^*\delta_{\mu\alpha}+\frac{1}{\sqrt{N}}\left(m^*+a\right)A^{\mu\alpha}+\frac{a}{N}\left(A^2\right)^{\mu\alpha}. \qquad (9.2.5)
\end{aligned}$$

Therefore

$$\mathrm{op}_{\mu\mu} = m^* + \frac{a}{N}\left(A^2\right)^{\mu\mu} \qquad (9.2.6)$$

and

$$\sum_{\alpha} \left(\mathrm{op}_{\mu\alpha}\right)^2 - m^{*2}$$
$$= \frac{1}{N}\left[\left(m^* + a\right)^2 + 2m^*a\right]\left(A^2\right)^{\mu\mu} + \mathcal{O}\left(\frac{\ln N}{N}\right)^{3/2}. \qquad (9.2.7)$$

To compute (9.2.4), we need moreover

$$\langle y, y\rangle = a^2 d\left(V^2\right)^{\mu\mu}. \qquad (9.2.8)$$

Altogether this leads us to

$$f_{\beta,\lambda}\left(m^{\mu}\right)$$
$$= \left(\frac{1}{\beta} I\left(m^*\right) - \frac{1}{2} m^{*2}\right) +$$
$$+\frac{1}{N}\left(\frac{a}{2\beta}\ln\frac{1+m^*}{1-m^*} - \frac{1}{2} m^*\left(3a + m^*\right)\right)\left(A^2\right)^{\mu\mu} + \mathcal{O}\left(\frac{\ln N}{N}\right)^{3/2} \qquad (9.2.9)$$

Now $f : \mathcal{A} \to \mathbb{R}^+$ with $f(a) = (a^2)^{\mu\nu}$ is Lipschitz-continuous with respect to the matrix norm $\|\cdot\|_2$ defined by

$$\|a\|_2^2 := \max_{\mu} \sum_{\nu} \left(a^{\mu\nu}\right)^2. \qquad (9.2.10)$$

To see this consider

$$\begin{aligned} \left|\left(b^2\right)^{\mu\nu} - \left(a^2\right)^{\mu\nu}\right| &= \left|\sum_{\alpha} b^{\mu\alpha} b^{\nu\alpha} - a^{\mu\alpha} a^{\nu\alpha}\right| \\ &\le \left|\max_{\alpha}\left\{a^{\mu\alpha}, b^{\mu\alpha}\right\}\right| \left|\sum_{\alpha} b^{\nu\alpha} - a^{\nu\alpha}\right| \\ &\le \left(\|a\|_s^2 + \|b-a\|_s^2\right)\|b-a\|_s^2. \qquad (9.2.11) \end{aligned}$$

Therefore with the law of the iterated logarithm and the strong approximation property of Proposition 6.11 (a) and (b), we can replace $(A^2)^{\mu\mu}$ by $(g_N^2)^{\mu\mu}$.

Saddle Points.

Without loss of generality we consider the case $\overline{s} = \frac{1}{2}m^*(b^\mu + b^\nu)$. Denote $v_\alpha := A^{\mu,\alpha} + A^{\nu,\alpha}$ as an abbreviation. We will use here $a \equiv \frac{m^*}{\gamma_1 - 1}$ and $k \equiv \frac{1}{2}\frac{\beta m^*}{1-\beta\left(1-\frac{1}{2}m^{*2}\right)}$. We start with equation (9.2.2) putting $x \equiv \overline{s}$ with perturbation $y \equiv k\left(1-S^2\right)\sum_\alpha v_\alpha b^\alpha - akv_\mu \overline{s}$ from Proposition 7.6. We will use

$$\ln\left(1-(\overline{s}_k)^2\right) = \frac{1}{m^{*2}}\overline{s}_k^2 \ln\left(1-m^{*2}\right) \tag{9.2.12}$$

and

$$\ln\frac{1+\overline{s}_k}{1-\overline{s}_k} = \frac{1}{m^*}\overline{s}_k \ln\frac{1+m^*}{1-m^*}. \tag{9.2.13}$$

Then we obtain

$$\begin{aligned}
&f_{\beta,\lambda}\left(s_{\mu,\nu}\right)\\
&= \frac{1}{2\beta m^{*2} d}\ln\left(1-m^{*2}\right)\left\langle \overline{s}, \left(\mathbb{1} + \frac{1}{\sqrt{N}}\Lambda\right)\overline{s}\right\rangle + \\
&\quad + \frac{1}{4\beta}\ln\frac{1+m^*}{1-m^*}\left(\mathrm{op}_\mu + \mathrm{op}_\nu\right) + \\
&\quad + \frac{1}{N}\frac{1}{\beta d}\left\langle y, \left(\mathbb{1}-S^2\right)^{-1} y\right\rangle - \frac{1}{2}\sum_\alpha\left(\mathrm{op}_\alpha\right)^2 + \mathcal{O}\left(\frac{\ln N}{N}\right)^{3/2}
\end{aligned} \tag{9.2.14}$$

We used here the overlap parameter $\mathrm{op}_\alpha := \frac{1}{N}\left\langle b^\alpha, Ls_{\mu,\nu}\right\rangle$. We obtain

$$\begin{aligned}
\mathrm{op}_\alpha &= \frac{1}{d}\left\langle b^\alpha, \left(\mathbb{1} + \frac{1}{\sqrt{N}}\Lambda\right)\left(\overline{s} + \frac{1}{\sqrt{N}}y\right)\right\rangle \\
&= \frac{1}{2}m^*\left(1 - \frac{1}{\sqrt{N}}akv_\mu\right)\left(\delta_{\alpha\mu} + \delta_{\alpha\nu} + \frac{1}{\sqrt{N}}v_\alpha\right) + \\
&\quad + \frac{k}{\sqrt{N}}\left[v_\alpha - \frac{1}{2}m^{*2}\left(v_\alpha + v_\mu\left(\delta_{\alpha,\nu} + \delta_{\alpha,\mu}\right)\right)\right] + \\
&\quad + \left(1 - \frac{1}{2}m^{*2}\right)\frac{k}{N}\sum_\beta v_\beta A^{\beta\alpha} - \\
&\quad - \frac{m^{*2}k}{2dN}\sum_\beta v_\beta\left\langle b^{\alpha\beta}, \Lambda b^{\mu\nu}\right\rangle.
\end{aligned} \tag{9.2.15}$$

This is valid, because

$$\begin{aligned}&\left\langle b^\alpha, \left(\mathbb{1} + \frac{1}{\sqrt{N}}\Lambda\right)\overline{s}\right\rangle\\&= \frac{1}{2}m^* d\left(\delta_{\alpha\mu} + \delta_{\alpha\nu} + \frac{1}{\sqrt{N}}v_\alpha\right)\end{aligned} \tag{9.2.16}$$

and, with the help of Lemma 9.1,

$$\begin{aligned}&\left\langle b^\alpha, \left(\mathbb{1} + \frac{1}{\sqrt{N}}\Lambda\right)\left(\mathbb{1} - S^2\right)b^\beta\right\rangle\\&= d\left(\delta_{\alpha,\beta} + \frac{1}{\sqrt{N}}A^{\alpha\beta}\right) - \frac{1}{2}m^{*2}d\left(\delta_{\alpha\beta} + \delta_{\{\alpha,\beta\},\{\mu,\nu\}}\right)\\&\quad -\frac{m^{*2}}{2\sqrt{N}}\left(dA^{\alpha\beta} + \left\langle b^{\alpha\beta}, \Lambda b^{\mu\nu}\right\rangle\right).\end{aligned} \tag{9.2.17}$$

We obtain from (9.2.15) using $v_\mu = v_\nu$

$$\begin{aligned}&\sum_\alpha \left(\mathrm{op}_\alpha\right)^2\\&= \frac{1}{4}m^{*2}\left(2\left(1 - \frac{2}{\sqrt{N}}akv_\mu\right) + \frac{1}{\sqrt{N}}4v_\mu\right) +\\&\quad + \frac{mk}{\sqrt{N}}\left(\left(1 - \frac{1}{2}m^{*2}\right)2v_\mu - \frac{1}{2}m^{*2}2v_\mu\right) +\\&\quad + \frac{m^{*2}}{4N}\left(2a^2k^2v_\mu^2 - 8akv_\mu^2 + \sum_\alpha v_\alpha\right) -\\&\quad - \frac{m^*k^2a}{N}v_\mu\left(\left(1 - \frac{1}{2}m^{*2}\right)2v_\mu - m^{*2}v_\mu\right) +\\&\quad + \frac{m^*k}{N}\left(\left(1 - \frac{1}{2}m^{*2}\right)\sum_\alpha v_\alpha^2 - m^{*2}v_\mu^2\right) +\\&\quad + \frac{k^2}{N}\left(\left(1 - \frac{1}{2}m^{*2}\right)^2\sum_\alpha v_\alpha^2 + \frac{1}{2}m^{*4}v_\mu^2 - 2m^{*2}\left(1 - \frac{1}{2}m^{*2}\right)v_\mu^2\right) +\\&\quad + \frac{m^*k}{N}\left(\left(1 - \frac{1}{2}m^{*2}\right)\sum_\alpha v_\alpha^2 - \frac{1}{2}m^{*2}\sum_\alpha v_\alpha^2\right).\end{aligned} \tag{9.2.18}$$

This can be simplified to

$$\begin{aligned}
&\sum_\alpha (\mathrm{op}_\alpha)^2 \\
&= \frac{1}{2}m^{*2} + \frac{m^* a\gamma_1}{\sqrt{N}} v_\mu + \\
&\quad + \frac{k}{N}\left(\frac{k}{\beta^2} + m^*\left(1-m^{*2}\right)\right)\sum_\alpha v_\alpha^2 + \\
&\quad + \frac{m^{*2}ak}{N}\left(\frac{1}{2}\gamma_1^2 ak - \frac{2k\gamma_1}{m^*\beta} - 1\right) v_\mu^2 + \mathcal{O}\left(\frac{\ln N}{N}\right)^{3/2}. \qquad (9.2.19)
\end{aligned}$$

As other ingredients we need

$$\begin{aligned}
\mathrm{op}_\mu + \mathrm{op}_\nu &= m^*\left(1 - \frac{1}{\sqrt{N}} ak v_\mu\right)\left(1 + \frac{1}{\sqrt{N}} v_\mu\right) + \\
&\quad + \frac{2k}{\sqrt{N}}\left(1-m^{*2}\right) v_\mu + \\
&\quad + \frac{k}{N}\sum_\beta v_\beta^2 - \frac{m^{*2}k}{N}\sum_\beta v_\beta^2 \\
&= m^* + \frac{2k}{\sqrt{N}}\left(1 - m^{*2} + \frac{1}{2}m^* a + \frac{m^*}{2k}\right) v_\mu \\
&\quad + \frac{m^* ak}{N} v_\mu^2 + \frac{k}{N}\left(1-m^{*2}\right)\sum_\alpha v_\alpha^2 \qquad (9.2.20)
\end{aligned}$$

and

$$\begin{aligned}
&\left\langle y, \left(\mathbb{1} - S^2\right)^{-1} y\right\rangle \\
&= k^2 \left\langle \sum_\alpha v_\alpha \left(\mathbb{1} - S^2\right) b^\alpha - a v_\mu \overline{s}, \sum_\beta v_\beta b^\beta - \frac{a}{1-m^{*2}} v_\mu \overline{s} \right\rangle \\
&= k^2 \sum_{\alpha,\beta} v_\beta v_\alpha \left\langle \left(\mathbb{1} - S^2\right) b^\alpha, b^\beta \right\rangle + k^2 \frac{a^2}{1-m^{*2}} v_\mu^2 \left\langle \overline{s}, \overline{s}\right\rangle - \\
&\quad - ak^2 \sum_\alpha v_\alpha v_\mu \left(\frac{1}{1-m^{*2}}\left\langle \left(\mathbb{1} - S^2\right) b^\alpha, \overline{s}\right\rangle + \left\langle b^\alpha, \overline{s}\right\rangle\right). \qquad (9.2.21)
\end{aligned}$$

With Lemma 9.1 we obtain

$$\begin{aligned}
&\left\langle y, \left(\mathbb{1} - S^2\right)^{-1} y \right\rangle \\
&= \left(1 - \frac{1}{2} m^{*2}\right) k^2 d \sum_\alpha v_\alpha^2 - m^{*2} k^2 d v_\mu^2 + \frac{1}{2} m^{*2} k^2 \frac{a^2}{1 - m^{*2}} d v_\mu^2 - \\
&\quad - 2 m^* k^2 a d v_\mu^2 \\
&= \left(1 - \frac{1}{2} m^{*2}\right) k^2 d \sum_\alpha v_\alpha^2 - \frac{1}{4} \left(2k + \beta a\right) a \gamma_1 m^{*2} d v_\mu^2. \qquad (9.2.22)
\end{aligned}$$

Putting this together leads to

$$\begin{aligned}
&f_{\beta,\lambda}\left(\overline{s} + \epsilon\right) \\
&= \frac{1}{2}\left(\frac{1}{\beta} I\left(m^*\right) - \frac{1}{2} m^{*2}\right) + \\
&\quad + \frac{1}{\sqrt{N}} A_N^{\mu\nu} \left(\frac{1}{2\beta} I\left(m^*\right) + \frac{a}{4\beta} \ln \frac{1 + m^*}{1 - m^*} - \frac{1}{2} m^* \gamma_1 a\right) + \\
&\quad + \frac{k}{N} \left(1 - m^{*2}\right) \left(\frac{1}{4\beta} \ln \frac{1 + m^*}{1 - m^*} - \frac{1}{4} m^* \left(\gamma_1 + 2\right)\right) \sum_\alpha v_\alpha^2 - \\
&\quad - \frac{m^* a k}{N} \left(\frac{1}{4\beta} \ln \frac{1 + m^*}{1 - m^*} + \frac{1}{4} m^* \left(\gamma_1 + 2\right)\right) v_\mu^2 \\
&\quad + \mathcal{O}\left(\frac{\ln N}{N}\right)^{3/2}. \qquad (9.2.23)
\end{aligned}$$

By the inequality (9.2.11) and the strong approximation property of Proposition 6.11, we can asymptotically replace $A^{\mu\nu}$ by $g_N^{\mu\nu}$. So we are done. □

Appendix A

Approximation of Gaussian sums via integrals

Let H be a positive definite $d \times d$ matrix. Then we can evaluate the associated Gaussian integral by

$$\int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\epsilon \langle x, Hx\rangle\right) dx = \frac{(2\pi)^{d/2}}{\sqrt{\det H}} \epsilon^{-d/2}. \tag{A.0.1}$$

We show now that Gaussian sums can be approximated by these integrals in a very precise way.

Proposition A.1 *Let $(H_\epsilon)_{\epsilon \in (0,1)}$ be a family of positive definite $d \times d$-matrices. We assume there exists $\kappa > 0$, independent of ϵ, such that*

$$\|H_\epsilon x\| \geq \kappa \|x\| \quad \text{for all } x \in \mathbb{R}^d. \tag{A.0.2}$$

Then the related Gaussian sum can be approximated by a Gaussian integral, that means we have

$$\sum_{k \in \mathbb{Z}^d} \exp\left(-\frac{1}{2}\epsilon \langle k, H_\epsilon k\rangle\right) \;=\; \frac{(2\pi)^{d/2}}{\sqrt{\det H_\epsilon}} \epsilon^{-d/2} \left(1 + \mathcal{O}\left(\sqrt{\epsilon}\right)\right). \tag{A.0.3}$$

Proof. Denote

$$s_d := \sum_{x \in \mathbb{Z}^d} \phi_\epsilon(x) \tag{A.0.4}$$

We prove the result by induction over the dimension n.

1. Fixing the induction at $n = 1$: We obtain by approximating the Gaussian integral via step functions from below and above using monotonicity

$$\sum_{k=1}^{\infty} \phi_\epsilon(k) \leq \int_0^\infty \phi_\epsilon(x)\, dx \leq \sum_{k=0}^{\infty} \phi_\epsilon(k). \tag{A.0.5}$$

Thus we have

$$s_1 - 1 \leq \int_{\mathbb{R}} e^{-\frac{1}{2}\epsilon H_\epsilon x^2} dx \leq s_1 + 1. \tag{A.0.6}$$

Since we have assumed that the spectrum of H_ϵ is uniformly bounded from below (A.0.2), we obtain

$$s_1 = \sqrt{\frac{2\pi}{\epsilon H_\epsilon}} \left(1 + \mathcal{O}\left(\sqrt{\epsilon}\right)\right). \tag{A.0.7}$$

2. Induction step $\{1, \ldots, n\} \to n+1$: By approximating again the Gaussian integral via step functions from below and above and using monotonicity, we obtain

$$\sum_{k \in \mathbb{N}^{n+1}} \phi_\epsilon(k) \leq \int_{\mathbb{R}_+^{n+1}} \phi_\epsilon(x)\, dx \leq \sum_{k \in \mathbb{N}_0^{n+1}} \phi_\epsilon(k). \tag{A.0.8}$$

Let $K \subset \{1, \ldots, n\}$ and define $A_{n,K} := \{x \in \mathbb{Z}^n \mid x_k = 0 \text{ for } k \in K\}$. Denote for $H_\epsilon \in \mathbb{R}^{n \times n}$ by $H_\epsilon^{(K)}$ the $(n - |K|) \times (n - |K|)$-matrix that arises by dropping the jth row and column of H_ϵ for all $j \in K$. We use now the fact, that the projection of a normal density on $\mathbb{R}^{n+1}$ onto a k-dimensional subspace is again a normal density, i.e.

$$\begin{aligned} &\sum_{x \in A_{n,K}} \exp\left(-\frac{1}{2}\epsilon \langle x, H_\epsilon x\rangle\right) \\ &= \sum_{x \in \mathbb{Z}^{n-|K|}} \exp\left(-\frac{1}{2}\epsilon \left\langle x, H_\epsilon^{(K)} x\right\rangle\right). \end{aligned} \tag{A.0.9}$$

Since we know from the induction hypothesis $s_k = \mathcal{O}\left(\epsilon^{-k/2}\right)$ for $k \leq n$, we obtain with the inclusion-exclusion principle

$$\begin{aligned} s_{n+1} &= \int_{\mathbb{R}^{n+1}} e^{-\frac{1}{2}\epsilon\langle x, H_\epsilon x\rangle} + \mathcal{O}\left(\epsilon^{-n/2}\right) \\ &= \frac{(2\pi)^{(n+1)/2}}{\sqrt{\det H_\epsilon}} \epsilon^{-(n+1)/2} \left(1 + \mathcal{O}\left(\sqrt{\epsilon}\right)\right). \end{aligned} \tag{A.0.10}$$

Hence, we have shown the proposition. $\square$

Appendix B

Estimation of sums by the Laplace-method

Let $a \in \mathbb{R}^d$ and $b > 0$. The Gaussian integral

$$I(a,b) := \int_0^\infty \exp\left(-ax - \frac{1}{2}\epsilon b x^2\right) dx \tag{B.0.1}$$

can be evaluated by a quadratic completion

$$I(a,b) = \sqrt{\frac{2\pi}{\epsilon b}} \exp\left(\frac{a^2}{2\epsilon b}\right)\left(1 - \mathscr{N}_{0,1}\left(\frac{a}{\sqrt{\epsilon b}}\right)\right). \tag{B.0.2}$$

Here $\mathscr{N}_{0,1}$ denotes the standard normal distribution function.

Now, we distinguish two different asymptotic behaviours. For notational convenience we leave out the dependence of a and b on ϵ.

(a) Assume there exists a constant $\beta > 0$ such that $a/\sqrt{\epsilon b} = \mathcal{O}\left(\epsilon^\beta\right)$. Then we obtain

$$I(a,b) = \sqrt{\frac{\pi}{2}} \frac{1}{\sqrt{\epsilon b}} \left(1 + \mathcal{O}\left(\epsilon^\beta\right)\right). \tag{B.0.3}$$

This holds since

$$\begin{aligned} \mathscr{N}_{0,1}\left(\frac{a}{\sqrt{\epsilon b}}\right) &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^{a/\sqrt{\epsilon b}} e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{2} + \mathcal{O}\left(\epsilon^\beta\right). \end{aligned} \tag{B.0.4}$$

(b) Assume there exists a constant $\gamma \in \left(0, \frac{1}{2}\right]$ such that $\sqrt{\epsilon b}/a = \mathcal{O}\left(\epsilon^{\gamma}\right)$. Then we obtain

$$I\left(a,b\right) = \frac{1}{a}\left(1+\mathcal{O}\left(\epsilon^{2\gamma}\right)\right). \tag{B.0.5}$$

This holds since

$$\begin{aligned} 1-\mathscr{N}_{0,1}\left(\frac{a}{\sqrt{\epsilon b}}\right) &= \frac{1}{\sqrt{2\pi}}\int_{a/\sqrt{\epsilon b}}^{\infty} e^{-\frac{1}{2}x^2}dx \\ &= \frac{1}{a}\sqrt{\frac{\epsilon b}{2\pi}}\exp\left(\frac{a^2}{2\epsilon b}\right)\left(1+\mathcal{O}\left(\epsilon^{2\gamma}\right)\right). \end{aligned} \tag{B.0.6}$$

For a proof of this see e.g. [Fe], p. 175.

We will show now that exponential sums have a very similar behaviour.

Proposition B.1 *Let Λ be an open interval that contains 0. We consider a family $\left(f_\epsilon\right)_{\epsilon\in(0,1)}$ with $f_\epsilon \in C^3\left(\Lambda,\mathbb{R}\right)$. Define the one sided lattice $\mathcal{N}_\epsilon := \Lambda \cap \epsilon\mathbb{N}_0$ and assume f_ϵ has exponentially tight level sets on $\mathcal{N}_\epsilon$, i.e.*

$$\sum_{x\in\mathcal{N}_\epsilon : f_\epsilon(x)\geq a} \exp\left(-f_\epsilon\left(x\right)/\epsilon\right) \leq c_a e^{a/\epsilon}. \tag{B.0.7}$$

We distinguish two cases:

(a) Assume there exists constants $c,\beta>0$ such that $\lim_{\epsilon\downarrow 0} f''_\epsilon\left(0\right) \geq c$ and $f'_\epsilon\left(0\right) = \mathcal{O}\left(\epsilon^{\frac{1}{2}+\beta}\right)$. Moreover, assume there exists $a>0$ small, such that for all $\delta>0$ small enough

$$f_\epsilon\left(x\right) \geq f_\epsilon\left(0\right) + a\delta^2 \text{ for all } x \geq \delta. \tag{B.0.8}$$

Then we obtain

$$\begin{aligned} &\sum_{x\in\mathcal{N}_\epsilon} \exp\left(-f_\epsilon\left(x\right)/\epsilon\right) \\ &= \frac{\sqrt{\pi}}{\sqrt{2\epsilon f''_\epsilon\left(0\right)}}\exp\left(-f_\epsilon\left(0\right)/\epsilon\right)\left(1+\mathcal{O}\left(\epsilon^{\overline{\beta}}\left|\ln\epsilon\right|^3\right)\right), \end{aligned} \tag{B.0.9}$$

where $\overline{\beta} := \frac{1}{2}\wedge\beta$.

(b) Assume there exists $\gamma \in \left(0, \frac{1}{2}\right]$ and $c > 0$, independent of ϵ, such that $\lim_{\epsilon \downarrow 0} \left(f_\epsilon' (0) / \epsilon^{\frac{1}{2} - \gamma} \right) \geq c$. Assume there exists $a > 0$ small, such that for all δ small enough

$$f_\epsilon (x) \geq f_\epsilon (0) + a\delta \quad \text{for all } x \geq \delta. \tag{B.0.10}$$

Then we obtain

$$\begin{aligned} &\sum_{t \in \mathcal{N}_\epsilon} \exp \left(-f_\epsilon (\epsilon t) / \epsilon \right) \\ &= \frac{1}{\left(1 - e^{-f_\epsilon'(0)}\right)} \exp \left(-f_\epsilon (0) / \epsilon \right) \left(1 + \mathcal{O} \left(\epsilon^\gamma \right) \right). \end{aligned} \tag{B.0.11}$$

In the case $\lim_{\epsilon \downarrow 0} f_\epsilon' (0) \geq c$, (i.e. $\gamma = \frac{1}{2}$) we get the more precise estimate

$$\begin{aligned} &\sum_{x \in \mathcal{N}_\epsilon} \exp \left(-f_\epsilon (x) / \epsilon \right) \\ &= \left(\frac{1}{1 - e^{-f_\epsilon'(0)}} - \frac{1}{2} \epsilon f_\epsilon'' (0) \frac{e^{-f_\epsilon'(0)} \left(1 + e^{-f_\epsilon'(0)} \right)}{\left(1 - e^{-f_\epsilon'(0)}\right)^3} \right) \times \\ &\quad \times \exp \left(-f_\epsilon (0) / \epsilon \right) \left(1 + \mathcal{O} \left(\epsilon^{3/2} \right) \right). \end{aligned} \tag{B.0.12}$$

Proof. ad (a). We choose $\delta \equiv \delta (\epsilon) := \sqrt{k\epsilon \left| \ln \epsilon \right|}$ with k constant. The sum (B.0.29) can be written as

$$\sum_{x \in \mathcal{N}_\epsilon} e^{-f_\epsilon(x)/\epsilon} = e^{-f_\epsilon(0)/\epsilon} \Big(\sum_{x < \delta} e^{-(f_\epsilon(x) - f_\epsilon(0))/\epsilon} + \sum_{x \geq \delta} e^{-(f_\epsilon(x) - f_\epsilon(0))/\epsilon} \Big). \tag{B.0.13}$$

The sums on the right hand side contains, of course, also only $x \in \mathcal{N}_\epsilon$. With the help of (B.0.8) and the exponentially small level sets of f_ϵ (assumption **F3**), the second sum of (B.0.13) is bounded by $c_{a\delta^2} \epsilon^{-d} e^{-a\delta^2/\epsilon} = c_\epsilon \epsilon^{ka-d} < \sqrt{\epsilon}$ for k large enough. As we will see this summand is negligible.

We abbreviate $a \equiv f_\epsilon' (0)$ and $b = f_\epsilon'' (0)$. Then we approximate f_ϵ by a Taylor series of second order around 0:

$$f_\epsilon (\epsilon t) - f_\epsilon (0) = \epsilon a t + \frac{1}{2} \epsilon^2 b t^2 + \mathcal{O} \left((\epsilon t)^3 \right). \tag{B.0.14}$$

Inserting this into the first sum of (B.0.13), we obtain

$$\begin{aligned}
&\sum_{t\in\mathbb{N}_0, t<\delta/\epsilon} \exp\left(-f_\epsilon\left(\epsilon t\right) + f_\epsilon\left(0\right)\right)/\epsilon \\
&= \sum_{t<\delta/\epsilon} \exp\left(-at - \frac{1}{2}\epsilon b t^2 + \mathcal{O}\left(\epsilon^2 t^3\right)\right) \\
&= \sum_{t<\delta/\epsilon} \exp\left(-at - \frac{1}{2}\epsilon b t^2\right)\left(1 + \mathcal{O}\left(\delta^3/\epsilon\right)\right). \qquad \text{(B.0.15)}
\end{aligned}$$

Notice that remainder of the sum satisfies

$$\begin{aligned}
&\sum_{t\geq\delta/\epsilon} \exp\left(-at - \frac{1}{2}\epsilon b t^2\right) \\
&= \sum_{t=0}^{\infty} \exp\left(-a\left(t + \lceil\delta/\epsilon\rceil\right) - \frac{1}{2}\epsilon b\left(t + \lceil\delta/\epsilon\rceil\right)^2\right) \\
&\leq \epsilon^{\frac{1}{2}ck} \sum_{t=0}^{\infty} \exp\left(-at - \frac{1}{2}\epsilon b t^2\right), \qquad \text{(B.0.16)}
\end{aligned}$$

which is negligible compared to the last sum for $k > \frac{1}{c}$. Therefore

$$\begin{aligned}
&\sum_{t<\lfloor\delta/\epsilon\rfloor} \exp\left(-f_\epsilon\left(\epsilon t\right) + f_\epsilon\left(0\right)\right)/\epsilon \\
&= \sum_{t=0}^{\infty}\left(\exp\left(-at - \frac{1}{2}\epsilon b t^2\right)\right)\left(1 + \mathcal{O}\left(\delta^3/\epsilon\right)\right) \qquad \text{(B.0.17)}
\end{aligned}$$

We approximate now this sum by an integral. Due to the monotonicity of $\left(-at - \frac{1}{2}\epsilon b t^2\right)$ on $(0,\infty)$ we have

$$\begin{aligned}
&\sum_{t=0}^{\infty}\left(\exp\left(-at - \frac{1}{2}\epsilon b t^2\right)\right) \\
&\leq \int_0^\infty \exp\left(-ax - \frac{1}{2}\epsilon b x^2\right) dx \\
&\leq \sum_{t=1}^{\infty}\left(\exp\left(-at - \frac{1}{2}\epsilon b t^2\right)\right). \qquad \text{(B.0.18)}
\end{aligned}$$

Therefore

$$\sum_{t=0}^{\infty} \exp\left(-at - \frac{1}{2}\epsilon b t^2\right) = I(a,b) + \mathcal{O}(1). \tag{B.0.19}$$

with the Gaussian integral I defined in (B.0.1).

Since $a = \mathcal{O}\left(\epsilon^{\beta}\sqrt{\epsilon b}\right)$ we obtain as in (B.0.3)

$$I(a_\epsilon, b) = \sqrt{\frac{\pi}{2}}\frac{1}{\sqrt{\epsilon b}}\left(1 + \mathcal{O}\left(\epsilon^{\beta}\right)\right). \tag{B.0.20}$$

Altogether we obtain for ν small enough

$$\sum_{t<\delta/\epsilon} \exp\left(-f_\epsilon(\epsilon t) + f_\epsilon(0)\right)/\epsilon = \sqrt{\frac{\pi}{2}}\frac{1}{\sqrt{\epsilon f_\epsilon''(0)}}\left(1 + \mathcal{O}\left(\epsilon^{\overline{\beta}}\left|\ln\epsilon\right|^3\right)\right) \tag{B.0.21}$$

with $\overline{\beta} = \min\left\{\frac{1}{2}, \beta\right\}$.

ad (b). The sum (B.0.29) can be written as

$$\sum_{x\in\mathcal{N}_\epsilon} e^{-f_\epsilon(x)/\epsilon} = e^{-f_\epsilon(0)/\epsilon}\left(\sum_{t<\kappa/\epsilon} e^{-(f_\epsilon(\epsilon t)-f_\epsilon(0))/\epsilon} + \sum_{t\geq\kappa/\epsilon} e^{-(f_\epsilon(\epsilon t)-f_\epsilon(0))/\epsilon}\right). \tag{B.0.22}$$

With the help of (B.0.10) and the exponentially small level sets of f_ϵ (assumption **F3**), the second sum of (B.0.22) is bounded by $c_\kappa e^{-c\kappa/\epsilon}$. We choose $\kappa \equiv \kappa_\epsilon = \epsilon^{1-\alpha}$ with $\alpha > 0$ small, hence this summand is exponentially small.

We abbreviate $a \equiv f_\epsilon'(0)$. Then we approximate f_ϵ by a Taylor series of second order around 0:

$$f_\epsilon(\epsilon t) - f_\epsilon(0) = \epsilon a t + \mathcal{O}\left((\epsilon t)^2\right). \tag{B.0.23}$$

Inserting this into the first sum of (B.0.22), we obtain

$$\begin{aligned}&\sum_{t<\kappa/\epsilon} \exp\left(-f_\epsilon(\epsilon t) + f_\epsilon(0)\right)/\epsilon \\ &= \sum_{t<\kappa/\epsilon} \exp(-at)\left(1 + \mathcal{O}\left(\epsilon^{1-2\alpha}\right)\right).\end{aligned} \tag{B.0.24}$$

Notice that remainder of the sum satisfies

$$\begin{aligned}&\sum_{t\geq\kappa/\epsilon} \exp(-at) \\ &= \exp\left(-\left\lceil\epsilon^{-\alpha}\right\rceil a\right)\sum_{t=0}^{\infty}\exp(-at),\end{aligned} \tag{B.0.25}$$

which is asymptotically exponentially smaller than this last sum for $\alpha > \frac{1}{2} - \gamma$. Therefore

$$\begin{aligned}
&\sum_{t<\kappa/\epsilon} \exp\left(-f_\epsilon\left(\epsilon t\right) + f_\epsilon\left(0\right)\right)/\epsilon \\
&= \sum_{t=0}^{\infty} \left(\exp\left(-a\right)\right)^t \left(1 + \mathcal{O}\left(\epsilon^{1-2\alpha}\right)\right) \\
&= \frac{1}{1-e^{-a}} \left(1 + \mathcal{O}\left(\epsilon^{1-2\alpha}\right)\right).
\end{aligned} \tag{B.0.26}$$

Hence we obtain for $\alpha = \frac{1}{2}\left(1-\gamma\right)$ the desired estimate.

In the case $\gamma = \frac{1}{2}$ we obtain

$$\begin{aligned}
&\sum_{t=0}^{\lfloor \kappa/\epsilon \rfloor} \exp\left(-f_\epsilon\left(\epsilon t\right) + f_\epsilon\left(0\right)\right)/\epsilon \\
&= \sum_{t=0}^{\infty} \exp\left(-f'_\epsilon\left(0\right)\right)^t \left(1 - \frac{1}{2}\epsilon f''_\epsilon\left(0\right) t^2\right) \left(1 + \mathcal{O}\left(\epsilon^{2-4\alpha}\right)\right) \\
&= \left(\frac{1}{1-e^{-f'_\epsilon(0)}} - \frac{1}{2}\epsilon f''_\epsilon\left(0\right) \frac{e^{-f'_\epsilon(0)}\left(1+e^{-f'_\epsilon(0)}\right)}{\left(1-e^{-f'_\epsilon(0)}\right)^3}\right) \left(1 + \mathcal{O}\left(\epsilon^{2-4\alpha}\right)\right)
\end{aligned} \tag{B.0.27}$$

The last step follows from

$$\sum_{t=0}^{\infty} t^2 e^{at} = \frac{d^2}{da^2} \sum_{t=0}^{\infty} e^{at}. \tag{B.0.28}$$

Hence we obtain for $\alpha = \frac{1}{8}$ the assertion.
□

Now we want to estimate sums in $\mathbb{Z}^d$ of the form

$$\sum_{x\in\mathbb{Z}^d} \exp\left(f_\epsilon\left(\epsilon x\right)/\epsilon\right) \qquad \text{as N} \to \infty. \tag{B.0.29}$$

For $v = \{v_1, \ldots, v_d\} \in \mathbb{R}^d$, we introduce as usual the norm $\|\cdot\|_\infty$ by

$$\|v\|_\infty := \max_{i\in\{1,\ldots,d\}} |v_i|. \tag{B.0.30}$$

Proposition B.2 *Consider a family* $(f_\epsilon)_{\epsilon\in(0,1)}$ *with* $f_\epsilon \in C^3\left(\mathbb{R}^d\right)$. *We assume* f_ϵ *has exponentially small level sets, i.e.*

$$\sum_{x\in\epsilon\mathbb{Z}^d : f_\epsilon(x)\geq a} \exp\left(-f_\epsilon(x)/\epsilon\right) \leq c_a e^{a/\epsilon}. \tag{B.0.31}$$

Assume that (f_ϵ) *converges uniformly on* Λ *to a function* $f \in C^3(\Lambda)$. *We assume that* f *has only finitely many critical points. Moreover, we assume that* f *and* f_ϵ *have a unique global minimum at* 0 *and* $\nabla^2 f(0)$ *and* $(\nabla^2 f_\epsilon(0))$ *are positive definite matrices such that*

$$\lim_{\epsilon\downarrow 0} \nabla^2 f_\epsilon(0) = \nabla^2 f(0). \tag{B.0.32}$$

Then we obtain

$$\begin{aligned}\sum_{x\in\epsilon\mathbb{Z}^d} \exp\left(-f_\epsilon(x)/\epsilon\right) &= \epsilon^{-d}\int_{\mathbb{R}^d} \exp\left(-f_\epsilon(x)/\epsilon\right) dx \times \\ &\times \left(1+\mathcal{O}\left(\sqrt{\epsilon}\left|\ln\epsilon\right|^{3/2}\right)\right).\end{aligned}$$

In particular

$$\begin{aligned}\sum_{x\in\epsilon\mathbb{Z}^d} \exp\left(-f_\epsilon(x)/\epsilon\right) &= \epsilon^{-d/2}\frac{(2\pi)^{d/2}}{\sqrt{\det\nabla^2 f_\epsilon(0)}}\exp\left(-f_\epsilon(0)/\epsilon\right) \times \\ &= \times\left(1+\mathcal{O}\left(\sqrt{\epsilon}\left|\ln\epsilon\right|^{3/2}\right)\right).\end{aligned} \tag{B.0.33}$$

Proof. Let $\delta := \sqrt{k\epsilon\left|\ln\epsilon\right|}$ with $k>0$ constant. The sum (B.0.29) can be written as

$$\sum_{x\in\epsilon\mathbb{Z}^d} e^{-f_\epsilon(x)/\epsilon} = e^{-f_\epsilon(0)/\epsilon}\Big(\sum_{\|x\|_\infty<\delta} e^{-(f_\epsilon(x)-f_\epsilon(0))/\epsilon} + \sum_{\|x\|_\infty\geq\delta} e^{-(f_\epsilon(x)-f_\epsilon(0))/\epsilon}\Big). \tag{B.0.34}$$

The sums on the right hand side contains, of course, also only $x\in\epsilon\mathbb{Z}^d$.

Since 0 is the unique global minimum of f_ϵ and $\nabla^2 f_\epsilon(0)$ is positive definite, there exists $a>0$ such that for all $\delta>0$ small enough we have

$$f_\epsilon(x) \geq f_\epsilon(0) + a\delta^2 \text{ for all } \|x\|_\infty \geq \delta. \tag{B.0.35}$$

With the help of (B.0.8) and the exponentially small level sets of f_ϵ (B.0.31), the second sum of (B.0.34) is bounded by $c_{a\delta^2}e^{-a\delta^2/\epsilon}$. Inserting $\delta = \sqrt{k\epsilon\left|\ln\epsilon\right|}$

we obtain $c_\epsilon \epsilon^{ak} < \sqrt{\epsilon}$ for k large enough. As we will see, this summand can be neglected.

We denote $H_\epsilon \equiv \nabla^2 f_\epsilon(0)$. To estimate the first sum in (B.0.34), we use the second order Taylor series

$$f_\epsilon(\epsilon k) - f_\epsilon(0) = \frac{1}{2}\epsilon^2 \langle k, H_\epsilon k\rangle + \mathcal{O}\left(\|\epsilon k\|_\infty^3\right). \tag{B.0.36}$$

Inserting this yields

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}^d:\|k\|_\infty<\delta/\epsilon} \exp\left[-\left(f_\epsilon(\epsilon k) - f_\epsilon(0)\right)/\epsilon\right] \\
&= \sum_{k:\|k\|_\infty<\delta/\epsilon} \exp\left(-\frac{1}{2}\epsilon\langle k, H_\epsilon k\rangle + \mathcal{O}\left(\delta^3/\epsilon\right)\right) \\
&= \sum_{k:\|k\|_\infty<\delta/\epsilon} \exp\left(-\frac{1}{2}\epsilon\langle k, H_\epsilon k\rangle\right)\left(1+\mathcal{O}\left(\sqrt{\epsilon}\,|\ln\epsilon|^{3/2}\right)\right) \\
&= \sum_{k\in\mathbb{Z}^d} \exp\left(-\frac{1}{2}\epsilon\langle k, H_\epsilon k\rangle\right)\left(1+\mathcal{O}\left(\sqrt{\epsilon}\,|\ln\epsilon|^{3/2}\right)\right).
\end{aligned} \tag{B.0.37}$$

To obtain the last equality notice that

$$\begin{aligned}
&\sum_{k:\|k\|_\infty\geq\delta/\epsilon} \exp\left(-\frac{1}{2}\epsilon\langle k, H_\epsilon k\rangle\right) \\
&\leq \exp\left(-\frac{1}{2}\epsilon^{-2\nu}\lambda\right)\sum_{k\in\mathbb{Z}^d} \exp\left(-\frac{1}{2}\epsilon\langle k, H_\epsilon k\rangle\right),
\end{aligned} \tag{B.0.38}$$

where λ denotes the smallest eigenvalue of H_ϵ.

Since $\nabla^2 f(0)$ and $(\nabla^2 f_\epsilon(0))$ are positive definite matrices and (B.0.32), we can apply Proposition A.1 and obtain

$$\sum_{x\in\epsilon\mathbb{Z}^d} e^{f(x)/\epsilon} = \epsilon^{-d/2}\frac{(2\pi)^{d/2}}{\sqrt{|\det H_\epsilon|}}\exp\left(f_\epsilon(0)/\epsilon\right)\times \tag{B.0.39}$$

$$\times\left(1+\mathcal{O}\left(\sqrt{\epsilon}\,|\ln\epsilon|^{3/2}\right)\right). \tag{B.0.40}$$

□

Bibliography

[AGS85] D.J. Amit, G. Gutfreund, and H. Sompolinsky. “Spin-Glass Models of Neural Networks”. *Phys. Rev. A*, **32**:1007–1018, 1985.

[AGS87] D.J. Amit, G. Gutfreund, and H. Sompolinsky. “Statistical Mechanics of Neural Networks near Saturation”. *Ann. Phys.*, **173**:30–67, 1987.

[BdHN06] A. Bovier, F. den Hollander, and F. Nardi. “Sharp asymptotics for Kawasaki dynamics on a finite box with open boundary conditions”. *Probab. Theor. Rel. Fields.*, **135**:265–310, 2006.

[BEGK01] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. “Metastability in Stochastic Dynamics of Disordered Mean-Field Models”. *Probab. Theor. Rel. Fields*, **119**:99–161, 2001.

[BEGK02] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. “Metastability and Low Lying Spectra in Reversible Markov Chains”. *Commun. Math. Phys.*, **228**:219–255, 2002.

[BEGK04] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. “Metastability in Reversible Diffusion Processes I. Sharp Estimates for Capacities and Exit Times”. *J. Eur. Math. Soc.*, **6**:399–424, 2004.

[BF05] A. Bovier and A. Faggionato. Spectral analysis of sinai’s walk for smal eigenvalues, 2005.

[BG94] A. Bovier and V. Gayrard. “Rigorous Results on the Hopfield Model of Neural Networks”. *Resenhas do IME-USP*, pages 161–173, 1994.

[BG98] A. Bovier and V. Gayrard. “Hopfield Models as Generalized Random Mean Field Models”. In A. Bovier and P. Picco, editors, *“Mathematical Aspects of Spin Glasses and Neural Networks”*. Birkhäuser, Boston, 1998.

[BG99] A. Bovier and V. Gayrard. “Sample Path Large Deviations for a Class of Markov Chains related to Disordered Mean Field Models”. Technical report, WIAS-preprint, 1999.

[BGK05] A. Bovier, V. Gayrard, and M. Klein. “Metastability in Reversible Diffusion Processes II. Precise Asymptotics for small Eigenvalues”. *J. Eur. Math. Soc.*, **7**:69–99, 2005.

[Bil68] P. Billingsley. *“Convergence of Probability Measures”*. John Wiley & Sons, New York, 1968.

[BM98] G. Biroli and R. Monasson. “Relationship between Long Time Scales and the Static Free-Energy in the Hopfield Model”. *J. Phys. A*, **31**:L391–L396, 1998.

[Bol01] B. Bollobas. *“Random Graphs”*. Cambrigde University Press, second edition, 2001.

[Bov04] A. Bovier. “Metastability and Ageing in Stochastic Dynamics”. In A. Maas, S. Martinez, and J. San Martin, editors, *“Dynamics and Randomness II, Nonlinear Phenomena and Complex Systems”*, volume **10**. Kluwer Acad. Publishers, Nederlands, 2004.

[Bré99] P. Brémaud. *“Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues”*. Springer, New York, 1999.

[Dav82] E.B. Davies. “Metastable States of Symmetric Markov Semigroups. I.”. *Proc. Lond. Math. Soc. III*, **45**:133–150, 1982.

[Doy89] P. Doyle. “Energy for Markov Chains”. http://math.dartmouth.edu/~doyle/docs/energy/energy.pdf, 1989.

[DS84] P.G. Doyle and J.L. Snell. *“Random Walks and Electrical Networks”*, volume 22 of *Carus Mathematical Monographs.* Mathematical Association of America, Washington DC, 1984.

[DV76] M.D. Donsker and S.R.S. Varadhan. "On the principal eigenvalue of second-order elliptic differential operators". *Comm. Pure Appl. Math*, **29**:595–621, 1976.

[Fel66] W. Feller. *"An Introduction to Probability Theory and its Applications vol.1"*. J. Wiley, New York, 1966.

[FP77] A.L. Figotin and L.A. Pastur. "Exactly Soluble Model of a Spin Glass". *Sov. J. Low Temp. Phys.*, **3**(6):378–383, 1977.

[FW84] M.I. Freidlin and A.D. Wentzell. *"Random Perturbations of Dynamical Systems"*. Springer, New York, 1984.

[Gen96] B. Genz. "An Almost Sure Central Limit Theorem for the Overlap Parameters in the Hopfield Model". *Stochastic Processes and their Applications*, **62**:243–262, 1996.

[GK86] D. Grensing and R. Kühn. "Random-Site Spin-Glass Models". *J. Phys. A*, **19**:L1153, 1986.

[GS98] B. Gaveau and L.S. Schulmann. "Theory of Nonequilibrium First-order Phase Transitions for Stochastic Dynamics". *J. Math. Phys.*, **39**(3):1517–1533, 1998.

[HLL03] O. Hernández-Lerma and J. B. Lasserre. *"Markov Chains and Invariant Probabilities"*. Birkhäuser, Basel, 2003.

[Hop82] J.J. Hopfield. "Neural Networks and Physical Systems with Emergent Collective Computational Abilities". *Proc. Nat. Acad. Sci*, **79**:2554–2558, 1982.

[Kat76] T. Kato. *"Perturbation Theory for Linear Operators"*. Springer, Berlin, second edition, 1976.

[KP89] H. Koch and J. Piasko. "Some Rigorous Results on the Hopfield Neural Network Model". *J. Stat. Phys.*, **55**(5/6):903–928, 1989.

[Kül97] C. Külske. "Metastates in Disordered Mean Field Models". *J. Stat. Phys.*, **88**(5/6):1257–1293, 1997.

[Lig85] T.M. Liggett. *"Interacting Particle Systems"*. Springer, Berlin, 1985.

[LT91] M. Ledoux and M. Talagrand. *"Probability in Banach Spaces"*. Springer, 1991.

[Löw98] M. Löwe. "On the Storage Capacity of Hopfield Models with Weakly Correlated Patterns". *Ann. Appl. Probab.*, **8**:1216–1250, 1998.

[MNOS04] F. Manzo, F.R. Nardi, E. Olivieri, and E. Scoppola. "On the Essential Features of Metastability: Tunnelling Time and Critical Configurations". *J. Stat. Phys.*, **115**(1/2):591–642, 2004.

[MS97] V.A. Malyshev and F.M. Spieksma. "Dynamics in Binary Neural Networks with a Finite Number of Patterns". *MPEJ*, **3**(1):1–36, 1997.

[OV05] E. Olivieri and M.E. Vares. *"Large Deviations and Metastability"*. Cambrigde University Press, 2005.

[Rio93] E. Rio. "Strong Approximation for Set-indexed Partial-sum Processes, via KMT Constructions II". *Ann. Prob.*, **21**(3):1706–1727, 1993.