Spin-flip dynamics of the Curie-Weiss model: Loss of Gibbsianness with possibly broken symmetry

Christof Külske* and Arnaud Le Ny†

October 14, 2005

Abstract

We study the conditional probabilities of the Curie-Weiss Ising model in vanishing external field under a symmetric independent stochastic spin-flip dynamics and discuss their set of bad configurations (points of discontinuity). We exhibit a complete analysis of the transition between Gibbsian and non-Gibbsian behavior as a function of time, extending the results for the corresponding lattice model, where only partial answers can be obtained.

For initial inverse temperature $\beta \leq 1$, we prove that the time-evolved measure is always Gibbsian. For $1 < \beta \leq \frac{3}{2}$, the time-evolved measure loses its Gibbsian character at a sharp transition time. For $\beta > \frac{3}{2}$, we observe the new phenomenon of symmetry-breaking of bad configurations: The time-evolved measure loses its Gibbsian character at a sharp transition time, and bad configurations with non-zero spin-average appear. These bad configurations merge into a neutral configuration at a later transition time, while the measure stays non-Gibbs.

In our proof we give a detailed analysis of the phase-diagram of a Curie-Weiss random field Ising model with possibly non-symmetric random field distribution. This analysis requires a careful study of the minimizers of some rate-function in the framework of bifurcation analysis.

Key words: Ising and mean field models, disordered systems, Curie-Weiss random field Ising model, Gibbs and non-Gibbsian measures, tricritical point, extended phase diagram, bifurcation analysis and Catastrophe theory.

AMS 2000 subject classification: 60K35, 82C20, 82B26, 82B44, 58K35.

1 Introduction

Recent years have seen a variety of situations where non-Gibbsian lattice-spin measures appear from proper Gibbs measures that are subjected to natural transformations. One particularly interesting dynamical phenomenon is the loss (and possible recovery) of the Gibbs property of an initial Gibbs measure that evolves according to a stochastic spin-flip dynamics. This was discovered and analyzed by van Enter *et al.* in [4]. Related results for continuous spins that evolve diffusively starting from an initial Gibbs measure

*University of Groningen, Department of Mathematics and Computing Sciences, Blauwborgje 3, 9747 AC Groningen, The Netherlands. `E-mail :kuelske@math.rug.nl`, `http://www.math.rug.nl/~kuelske/`

†Université de Paris-Sud, Laboratoire de Mathématiques, Bâtiment 425, 91405 Orsay Cedex, France. `E-mail :arnaud.leny@math.u-psud.fr`, `http://www.math.u-psud.fr/~leny/`

were very recently obtained in [13]. Suppose a lattice spin system is in an equilibrium situation at initial inverse temperature β, described by a proper Gibbs measure in infinite volume. Suppose now the system undergoes a fast heating procedure, described by a high-temperature Glauber dynamics. A question arises: Is there a well-defined Hamiltonian for every time? Or, in the more catchy terms of [4]: Is there always a well-defined temperature? It turns out that there can be transition-phenomena of an 'in and out of Gibbsianness' as a function of time. These phenomena are already present when we consider an infinite-temperature time evolution, i.e. a spin-flip that is independent from site to site.

Let us highlight one particular phenomenon from [4], specializing to vanishing external magnetic field. Pick the plus state of a low temperature Ising model in zero external magnetic field as initial measure μ_β at time $t = 0$ and perform an unbiased spin-flip dynamics, independently over the sites, with rate 1. Denote the resulting probability measure at time t by $\mu_{\beta,t}$. In [4] in particular the following result is contained in Theorem 5.2.

Theorem 1.1 ([4]) *Assume that the initial inverse temperature β is large compared to the critical inverse temperature of the nearest neighbor Ising model. Then there exist $t_0(\beta) \leq t_1(\beta)$ such that*

1. *$\mu_{\beta,t}$ is a Gibbs measure for all $0 \leq t < t_0(\beta)$.*
2. *$\mu_{\beta,t}$ is* **not** *a Gibbs measure for all $t > t_1(\beta)$.*

There remain open questions in this picture: Is the transition between Gibbsianness and non-Gibbsianness sharp, i.e. $t_0(\beta) = t_1(\beta)$, as it was conjectured in [4]? Can one understand the trajectory of the interactions of the measure $\mu_{\beta,t}$ as time varies in more detail?

These and related questions seem to be difficult to answer for lattice systems, even for the infinite temperature dynamics. The purpose of the present paper is to provide detailed answers for the corresponding mean-field model. Moreover, in the course of our analysis we also find an interesting new mechanism of non-Gibbsianness that was not observed on the lattice. Recall that a discrete spin measure μ on the lattice is Gibbs iff (a version of) its conditional probabilities $\mu(\sigma_x| \cdot) : \sigma_{x^c} \mapsto \mu_{\beta,t}(\sigma_x|\sigma_{x^c})$ is continuous w.r.t. the product topology for any site x of the graph. This means that the influence of a perturbation of a conditioning σ_{x^c} outside of a large volume Λ will tend to zero, when Λ tends to $\mathbb{Z}^d$.

To investigate this property for the time-evolved measure $\mu_{\beta,t}$, we relate its conditional probabilities to expectations of a local observable w.r.t. a certain *quenched measure*, obtained by conditioning the measure at time $t = 0$ to be in a given spin configuration at time t. In this quenched measure the spin-configuration in the conditioning appears as an additional 'frozen' external magnetic field configuration. The investigation of such a quenched measure was performed in [4] as well as in [13] for the corresponding setup, and we will also perform it in this present paper for the mean-field model we are about to study.

The failure of the Gibbs-property for the time-evolved measure will occur if there is sensitive dependence of the quenched model on the external magnetic field configuration, at certain 'bad configurations'. For example, the checkerboard configuration, being the simplest example of a 'neutral configuration', serves as such a bad configuration in the proof of the second part of Theorem 1.1 in the large β and large t region. In turn, the absence of such bad configurations implies the Gibbs property.

To get a precise understanding of the quenched model is a difficult task for lattice models, where one usually cannot hope for exhaustive results. In the mean-field setup however we are able to compute the conditional probabilities (in infinite volume) in terms of the minimizers of the large deviation rate-function for the magnetization of the quenched model. To understand the structure of the minimizers is equivalent to understand the phase-diagram of a Curie-Weiss random field Ising model with possibly non-symmetric random-field distribution, as we shall see.

In this way, our analysis of the transitions between Gibbsian and non-Gibbsian behavior is reduced to the analytic problem of a bifurcation-analysis for this rate-function. We perform this analysis in the spirit of Catastrophe theory. As a result we discover an interesting new structure where bad configurations with non-zero spin-average appear in a spin-flip invariant model. This phenomenon happens in a regime of low enough temperatures, and this regime is strictly smaller than the regime of temperatures for which there is a phase transition for the initial measure. Moreover, it turns out that the model is explicitly solvable in the following strong sense: We can find a parametrization in terms of explicit functions of the boundary curve in (β, t)-space of the region for which $\mu_{\beta,t}$ is Gibbs. This curve consists of two different branches with different functional forms, depending whether they form the boundary to non-Gibbsianness with symmetry breaking, or non-Gibbsianness without symmetry breaking. This is a satisfying result, in view that even in a mean-field set-up one can not usually hope for explicit functional expressions for phase-boundaries.

We remark that we expect even more complicated phenomena to occur in the case of a non-vanishing external magnetic field for the initial measure. Indeed, [4] show that for an initial Ising model on the lattice at low temperature in a non-vanishing field, after a loss for a range of intermediate times, the Gibbs property can be restored for t large enough, whereas the recovery is obtained only for t infinite (i.e. at equilibrium) if the external field is vanishing. In contrast to that, [13] prove that there is a loss of the Gibbs property *independently* of the initial magnetic field in their model of continuous unbounded spins, and that non-Gibbsianness persists for all large enough times. In the present paper however, we restrict our attention to the case of a vanishing external field, where non-Gibbsianness also persists at low temperature and is recovered only at equilibrium, letting the detailed study the case of non-vanishing external fields in the mean-field setup for the companion paper [11].

Can non-Gibbsianness with symmetry breaking of bad configurations also happen for a lattice model? This is a challenging question. A natural candidate to establish this phenomenon is the corresponding Kac-Ising model under time-evolution and this will be the subject of future research.

1.1 The model

We start at time $t = 0$ with the Curie-Weiss Ising model in zero magnetic field at inverse temperature $\beta > 0$ whose finite-volume (N) Gibbs measures on spin-configurations $\sigma_{[1,N]} = (\sigma_i)_{i=1,\dots,N} \in \Omega_N := \{-1, +1\}^N$ are given by

$$\mu_{\beta,N}(\sigma_{[1,N]}) = \frac{\exp\left(\frac{\beta}{2N}\left(\sum_{i=1}^N \sigma_i\right)^2\right)}{Z_{\beta,N}} \tag{1}$$

where the normalization factor $Z_{\beta,N}$ is the standard partition function. This model shows a phase transition at the critical inverse temperature $\beta_c^{\rm CW} = 1$, distinguishing in the infinite-volume limit a region of uniqueness at high temperature ($\beta < 1$) from a region of coexistence of a '+'-like and a '-'-like phases at low temperature ($\beta \geq 1$).

Then we apply a stochastic spin-flip dynamics at rate 1, independently over the sites. Our object of interest in this paper will be the resulting image measure $\mu_{\beta,t,N}$ given by

$$\mu_{\beta,t,N}(\eta_{[1,N]}) := \sum_{\sigma_{[1,N]}\in\Omega_N} \mu_{\beta,N}(\sigma_{[1,N]}) \prod_{i=1}^{N} p_t(\sigma_i,\eta_i). \tag{2}$$

Here

$$p_t(\sigma_i,\eta_i) = \frac{e^{\eta_i\sigma_i h_t}}{2\cosh h_t}, \quad \text{with} \quad h_t = \frac{1}{2}\log\frac{1+e^{-2t}}{1-e^{-2t}}, \tag{3}$$

is the transition kernel from a spin-value σ_i at time $t=0$ to a spin value η_i at time $t>0$. The product form of the kernel in (2) is a special case of a Glauber dynamics for infinite temperature (cf. [4, 14, 15]).

It is well known that, for fixed N, the time-evolved measure $\mu_{\beta,t,N}$ tends to a spin-flip invariant product measure on $\{-1,+1\}^N$, with $t \uparrow \infty$. We shall however see that the large-N-behavior of the conditional probabilities of the time-evolved measure can be non-trivial, even when the corresponding measure is close to a product measure.

1.2 The notion of Gibbsianness for mean-field models

Before we state our main result in the mean-field set-up, we need to give a precise definition of Gibbsianness for mean-field models, that is based on the notion of continuity of conditional probabilities. For more information and discussion we refer the reader to [10],[7] and [12]. Observe that, due to exchangeability, the conditional probabilities of the time-evolved measure are automatically volume-dependent *functions of the empirical average* of the spins in the conditioning. This suggests to characterize a configuration by its magnetization and the following definition.

Definition 1.2 *We call $\alpha_0 \in (-1,1)$ a* **good configuration** *if and only if*

1. *The limit*

$$\gamma_{\beta,t}(\eta_1|\alpha) := \lim_{N\uparrow\infty} \mu_{\beta,t,N}(\eta_1|\eta_{[2,N]}) \qquad \textit{whenever} \ \lim_{N\uparrow\infty}\frac{1}{N}\sum_{i=2}^{N}\eta_i = \alpha \tag{4}$$

   exists for all α in a neighborhood of α_0.

2. *The function $\alpha \mapsto \gamma_{\beta,t}(\eta_1|\alpha)$ is continuous at $\alpha = \alpha_0$.*

We call the mean-field model $\mu_{\beta,t,N}$ **Gibbs** *iff every configuration is good.*

Note: For lattice systems a Gibbs measure in infinite volume is characterized by the fact that its conditional probabilities depend continuously w.r.t. product topology on the conditioning. Clearly, this notion would yield meaningless answers for mean field models. This is clear since the finite-volume Gibbs measures of mean field models converge to convex combinations of product measures, by de Finetti's theorem. Since non-trivial linear-combinations of product measures have conditional probabilities that have every point as a point of discontinuity (cf. [6]), non-Gibbsianness in this naive approach would always be trivially true for mean-field models in the phase transition region where the former combination is non-trivial. Thus, continuity of conditional probabilities for mean field models must be taken in the above sense where we demand continuity w.r.t. to the empirical average of the configuration in the conditioning. Note that the initial mean-field model (1) is of course a Gibbs measure in the sense of this definition since $\gamma_\beta(\eta_1|\alpha) := \gamma_{\beta,t=0}(\eta_1|\alpha) = \frac{e^{\beta\alpha\eta_1}}{2\cosh(\beta\alpha)}$ is a continuous function at every α.

In the case of the occurrence of bad configurations (i.e. that are not good in the sense of our definition) we can carry the analogue between mean-field and lattice to the notion of "almost sure Gibbsianness" (that is familiar on the lattice). For the mean-field model we look at the size of the set of the discontinuity points in the large volume-limit, with respect to the limiting measure on the time evolved measure at time t. If it gets measure zero, we have found the mean-field analogue of "almost sure Gibbsian" behavior (which will be studied in more details in [12]). An analysis of this sort was carried out in [10] for the decimation transformation of the Ising ferromagnet, and examples of joint measures in random systems including the random field Ising model, and in [7] for the fuzzy-Potts model. These examples very often show close analogy of the mechanisms of non-Gibbsianness between the lattice-model and their mean-field counterparts.

1.3 Main result

We are now ready to state our main result and describe the Gibbs vs. non-Gibbs character of the time-evolved measure depending on the temperature.

Theorem 1.3

1. *If $\beta^{-1} \geq 1$ the limiting conditional probabilities $\gamma_{\beta,t}$ are continuous functions of the empirical mean $\alpha \in (-1, 1)$, for all $t \geq 0$.*
2. *If $\frac{2}{3} \leq \beta^{-1} < 1$ there exists a sharp value $0 < t_0(\beta) < \infty$ such that:*
   - *For all $0 \leq t < t_0(\beta)$, the limiting conditional probabilities are continuous.*
   - *For all $t \geq t_0(\beta)$, the limiting conditional probabilities are discontinuous at the point $\alpha = 0$, and continuous at any $\alpha \neq 0$.*
3. *If $0 < \beta^{-1} < \frac{2}{3}$ then there exist sharp values $0 < t_0(\beta) < t_1(\beta) < \infty$ such that:*
   - *For all $0 \leq t < t_0(\beta)$, the limiting conditional probabilities are continuous.*
   - *For all $t_0(\beta) \leq t < t_1(\beta)$ there exists $\alpha_c = \alpha_c(\beta, t) \in]0, 1[$, such that the limiting conditional probabilities are discontinuous at the points α_c and $-\alpha_c$, and continuous else. For fixed β^{-1}, the function $t \mapsto \alpha_c(\beta, t)$ is decreasing and we have $\lim_{t\uparrow t_1} \alpha_c(\beta, t) = 0$.*
   - *For all $t \geq t_1(\beta)$, the limiting conditional probabilities are discontinuous at the point $\alpha = 0$, and continuous at any $\alpha \neq 0$.*

In this picture, we distinguish three regions with different qualitative behavior:

"Gibbs" Region: Gibbsianness for small times, and forever for $\beta^{-1} \geq 1$.

"non-Gibbs" Region: Non-Gibbsianness is due to neutral bad configurations.

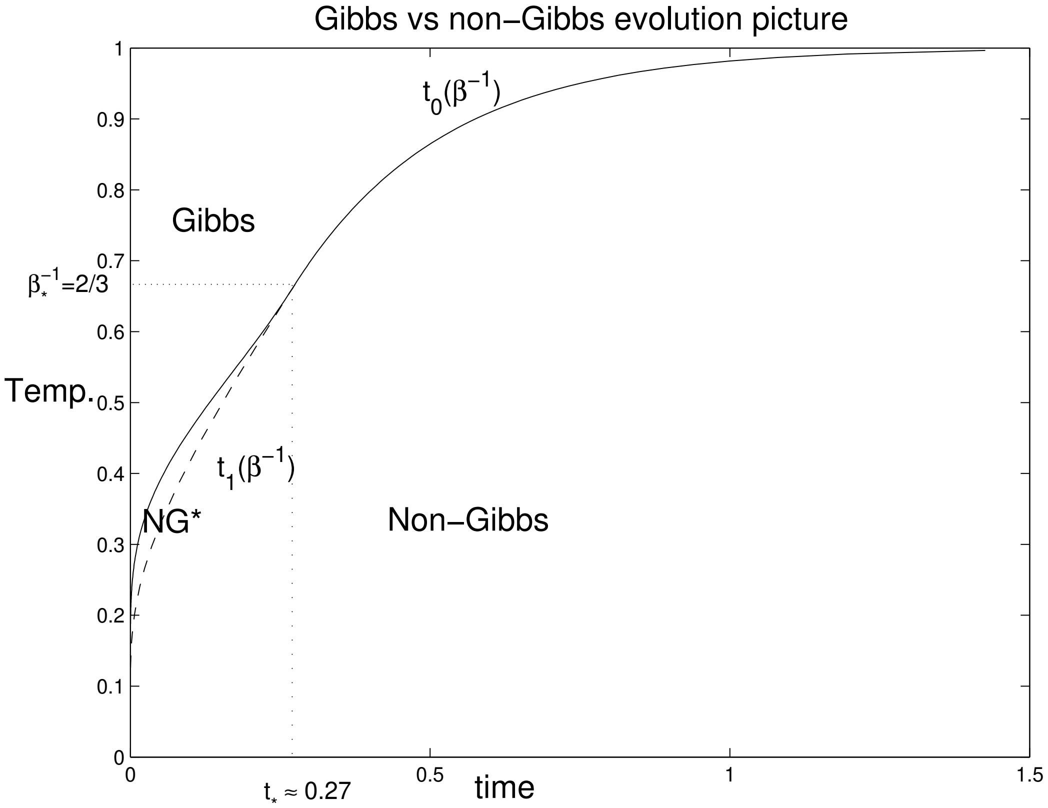

Figure 1: **Gibbs vs non-Gibbs time evolution picture**

"NG*" Region: Non-Gibbsianness due to bad configurations with broken symmetry.

In the next theorem we give the functional form of the two branches of the boundary curves of the Gibbs-region in time-temperature space. While the particular form of the functions is not important for the phenomenology of the Gibbs-non Gibbs transition, it is a worthwhile analytical fact that they are known.

Theorem 1.4

1. *If* $\frac{2}{3} \leq \beta^{-1} < 1$*, the threshold time is given by* $t_0(\beta) = -\frac{1}{4}\log(1-\beta^{-1})$
2. *If* $0 < \beta^{-1} < \frac{2}{3}$ *the boundary of the Gibbsian region has the parametrization, in terms of a parameter* $M > 0$*, of the form*

$$\left\{ \begin{pmatrix} t_0(\beta^{-1}) \\ \beta^{-1} \end{pmatrix}, 0 < \beta^{-1} < \frac{2}{3} \right\} = \left\{ \begin{pmatrix} t_4(M) \\ \beta_4^{-1}(M) \end{pmatrix}, 0 < M < \infty \right\} \tag{5}$$

with

$$\begin{aligned} t_4(M) =& -\frac{1}{4}\log\frac{2y^3 + M(1-y^2)^2}{2y + M(1-y^2)} \\ \beta_4^{-1}(M) =& \frac{y(2+My)(1+y^2)}{2M^2y^3 + 2(y+y^3) + M(1+3y^2+y^4)} \end{aligned} \tag{6}$$

where $y = y(M) = \tanh(M)$.

We do not have an explicit curve for line $t_1(\beta)$, see however the information given below.

In regions 1. and 2. of Theorem 1.3 there are no surprises with regard to the known results from the lattice case: In the high-temperature region 1. the measure stays Gibbs, for all t, in complete analogy to the situation on the lattice. In the intermediate range of temperature of case 2., the measure stays Gibbs for short times, in complete analogy with the lattice results of [14], and there is a sharp threshold-time that distinguishes between Gibbsian and non-Gibbsian behavior for the time-evolved measure, as it was conjectured for the behavior on the lattice. The only bad configuration for the empirical magnetization α (the point of discontinuity for the limiting conditional expectations) is the neutral value $\alpha = 0$. This is suggested by the analogy to the checkerboard configuration on the lattice. Notice that on the lattice any finite volume perturbation of a bad configuration produces another bad configuration, and so there are always infinitely many bad configuration on the lattice. This however does not change the magnetization in large volumes, and we expect just one point of discontinuity as a function of the magnetization in the mean-field model.

In region 3., i.e. for even lower temperatures and an intermediate range of time (see Figure 1), the more complicated phenomenon of *symmetry breaking in the set of bad configurations* is appearing: A non-empty set of bad configurations consisting of a pair $(\alpha^c, -\alpha^c)$ is born at a finite time that merges into the neutral configuration $\alpha = 0$ at a later time. We call the intermediate range of time *non-Gibbsianness with symmetry breaking*, sometimes abbreviated 'NG*', in the previous Figure 1 among others.

This new mechanism of non-Gibbsiannness is caused by the fact that the rate-function for the magnetization of the measure at time 0 conditioned on the evolved magnetization α at time t might have 1, 2, but also 3 minima. In region 2., at big enough times, a variation of the empirical magnetization α around zero then induces a phase transition between the two ordered $+$-like and $-$-like states that are related by symmetry. This is the mechanism of non-Gibbsianness known from [4].

More interestingly, in region 3., for intermediate times there is a transition between an ordered (say) $-$-like state and an approximately neutral '$\pm$-state', at a positive value of the empirical magnetization $\alpha = \alpha_c$. Both mechanisms collapse at a critical time where a particular singularity occurs for $\beta^{-1} = \frac{2}{3} := \beta_*^{-1}$ at $t = t_0(\frac{3}{2}) := t_* \approx 0.27$, and for $\alpha = 0$ where the quenched system is at a tricritical point.

In order to understand the symmetry breaking of bad configurations in detail we discuss the *extended phase diagram* of the quenched model, generalizing the known phase diagram from the symmetric case $\alpha = 0$ (see [1, 17, 2]) to non-symmetric distributions. We believe that this also has some interest in itself, beyond the study of non-Gibbsianness. Figure 1 then appears as the image of three of the lines of the extended phase diagram, by relating time in the original model to the coupling strength of the quenched fields in the quenched model via $t = \mathrm{atanh}(\exp(-2h_t))$. (Note that the inverse mapping of $t \mapsto h_t$ has the same functional form.)

An explicit expression for the trajectory $t \mapsto \gamma_{\beta,t}$ of the kernel in question in the "Gibbs" region is moreover given in Theorem 2.1.

To be able to give such the explicit form of Theorem 1.4 seems to be a lucky feature of the model. It is the result of a thorough discussion of the underlying bifurcation problem. For this we exploit the particular form of the rate-function, useful re-parametrization and linear dependence on two of the parameters.

The line separating the NG*-regime from the usual NG-regime is parametrized in terms of the dynamical field and a numerical solution of a mean-field type equation.

The paper is organized as follows. In Section 2 we derive explicit expressions for the limiting conditional probabilities of the time-evolved model in terms of the rate function of a mean-field model with possibly non-symmetric random field distribution. In Section 3 we discuss the bifurcation structure of this rate function and the phase diagram of the corresponding model. The proof of Theorems 1.3 and 1.4 follows immediately as a consequence of the structure of this extended phase diagram.

2 Infinite-volume limit of conditional probabilities

The existence of an infinite-volume limit for the conditional probabilities of the time-evolved model (cf. Definition 1.2) is not granted by abstract arguments, but must be deduced by an explicit analysis of the model. However, we will show that this limit exists and a kernel $\gamma_{\beta,t}(\eta_1|\alpha)$ is well-defined when a certain rate-function depending on the triple (β, t, α) has a unique minimizer. When this is the case we have moreover an explicit expression in terms of the minimizer of this rate-function. This also implies continuity, by the smooth dependence of the minimizer on small α-dependent perturbations of the rate-function. On the contrary, when this rate function has more than one minimizer, this dependence can be not smooth anymore and leads to different left and right infinite-volume limits[1] and thus to a configuration which is bad because the limit is not defined and cannot be extended by continuity arguments.

2.1 Continuity result - Explicit expression in terms of a rate-function

Theorem 2.1 *Denote by $\tilde{\eta}$ a random variable taking values ± 1 with average α, and write $\mathbb{E}_\alpha$ for its expectation. Put $\varepsilon = \varepsilon_t := h_t/\beta$ and suppose that the function*

$$\Phi_{\beta,\varepsilon,\alpha} : m \mapsto \Phi_{\beta,\varepsilon,\alpha}(m) = \frac{m^2}{2} - \frac{1}{\beta}\mathbb{E}_\alpha\Big(\log\cosh(\beta(m + \varepsilon\tilde{\eta}))\Big) \tag{7}$$

has a unique minimizer $m^*(\beta, \varepsilon, \alpha)$, for the fixed choice of $(\beta, \varepsilon, \alpha)$. Then the conditional probabilities of the time evolved measure (2) have a well-defined infinite volume limit $\gamma_{\beta,t}(\alpha)$ in the sense of (4) and we have the representation

$$\gamma_{\beta,t}(\eta_1|\alpha) = \frac{e^{\eta_1 g_{\beta,t}(\alpha)}}{2\cosh g_{\beta,t}(\alpha)}, \text{ with } g_{\beta,t}(\alpha) = \frac{1}{2}\log\frac{\cosh\Big(\beta\big(m^*(\beta,\varepsilon_t,\alpha)+\varepsilon_t\big)\Big)}{\cosh\Big(\beta\big(m^*(\beta,\varepsilon_t,\alpha)-\varepsilon_t\big)\Big)}. \tag{8}$$

The time- and α-dependent effective field $g_{\beta,t}(\alpha)$ captures all information about the interaction at time t of the spin η_1 with the empirical magnetization of the outside world α. It is not difficult to verify that the kernel (8) interpolates between the initial distribution and the final product measure. Indeed, we have[2] $\lim_{t\uparrow\infty}\gamma_{\beta,t}(\sigma_1|\alpha) = \frac{1}{2}$ and $\lim_{t\downarrow 0}\gamma_{\beta,t}(\sigma_1|\alpha) = \gamma_\beta(\sigma_1|\alpha)$.

Assuming Theorem 2.1, the further study of discontinuous behavior of the limiting conditional probabilities is then boiled down to an analytical problem which will be

[1] The left, resp. right, limit corresponds to approach α from below, resp. above, in (4).

[2] To show the second equality we have used that $\lim_{\varepsilon\uparrow\infty} m^*(\beta,\varepsilon,\alpha) = \alpha$. This follows from the mean-field equation in the next subsection.

treated in the next section. The result of the theorem itself relies on the following explicit expression of the conditional probabilities in finite volume in terms of a quenched model.

2.2 Proof: Finite-volume representation in terms of quenched model

For any fixed configuration of 'random fields' $\eta_{[1,N]} \in \{-1,+1\}^N$, let us introduce the following 'quenched' measure on the configurations $\sigma_{[1,N]} \in \{-1,+1\}^N$:

$$\mu_{\beta,\varepsilon,N}[\eta_{[1,N]}](\sigma_{[1,N]}) := \frac{\exp\Big(\frac{\beta}{2N}\big(\sum_{i=1}^N \sigma_i\big)^2 + \beta\varepsilon \sum_{i=1}^N \eta_i\sigma_i\Big)}{Z_{\beta,\varepsilon,N}[\eta_{[1,N]}]}. \tag{9}$$

This quenched model is a Curie-Weiss Ising model with additional random fields η that will be given by the spin-configuration at time t, constrained to have a fixed empirical distribution α. It will be important to understand non-symmetric distributions and this will be discussed in the next section.

The single-site conditional probabilities of the time-evolved model can then be expressed in terms of an expectation w.r.t a quenched model of the form (9) on N-1 spins, depending on $\eta_{[2,N]}$. We get the following explicit formula.

Proposition 2.2 *For each finite volume N we have the following representation of the conditional probabilities of the time-evolved model :*

$$\mu_{\beta,t,N}(\eta_1|\eta_{[2,N]}) = \frac{e^{\eta_1 g_{\beta,t,N}(\eta_{[2,N]})}}{2\cosh g_{\beta,t,N}(\eta_{[2,N]})}$$

where

$$g_{\beta,t,N}(\eta_{[2,N]}) = \frac{1}{2}\log \frac{\mu_{\beta_N,\varepsilon_N,N-1}[\eta_{[2,N]}]\Big[\cosh\Big(\beta_N\big(\frac{1}{N-1}\sum_{i=2}^N \sigma_i + \varepsilon_N\big)\Big)\Big]}{\mu_{\beta_N,\varepsilon_N,N-1}[\eta_{[2,N]}]\Big[\cosh\Big(\beta_N\big(\frac{1}{N-1}\sum_{i=2}^N \sigma_i - \varepsilon_N\big)\Big)\Big]}$$

and

$$\beta_N = \beta\cdot\frac{N-1}{N}, \quad \beta_N\varepsilon_N = h_t. \tag{10}$$

Remark: The dependence (10) of the parameters β_N etc. on N has no influence on the large-N limit.

Proof: The proof follows from the following rewriting by using the definition of the dynamics:

$$\frac{\mu_{\beta,t,N}(\eta_1=+|\eta_{[2,N]})}{\mu_{\beta,t,N}(\eta_1=-|\eta_{[2,N]})} = \frac{\sum_{\sigma_{[1,N]}}\exp\Big(\frac{\beta}{2N}\big(\sum_{i=1}^N\sigma_i\big)^2 + \beta\varepsilon\sigma_1 + \beta\varepsilon\sum_{i=2}^N\eta_i\sigma_i\Big)}{\sum_{\sigma_{[1,N]}}\exp\Big(\frac{\beta}{2N}\big(\sum_{i=1}^N\sigma_i\big)^2 - \beta\varepsilon\sigma_1 + \beta\varepsilon\sum_{i=2}^N\eta_i\sigma_i\Big)}.$$

Carrying out the σ_1-sums yields

$$\frac{\mu_{\beta,t,N}(\eta_1=+|\eta_{[2,N]})}{\mu_{\beta,t,N}(\eta_1=-|\eta_{[2,N]})} = \frac{\sum_{\sigma_{[2,N]}}\exp\Big(\frac{\beta}{2N}\big(\sum_{i=2}^N\sigma_i\big)^2 + \beta\varepsilon\sum_{i=2}^N\eta_i\sigma_i\Big)\cosh\Big(\beta\big(\frac{1}{N}\sum_{i=2}^N\sigma_i + \varepsilon\big)\Big)}{\sum_{\sigma_{[2,N]}}\exp\Big(\frac{\beta}{2N}\big(\sum_{i=2}^N\sigma_i\big)^2 + \beta\varepsilon\sum_{i=2}^N\eta_i\sigma_i\Big)\cosh\Big(\beta\big(\frac{1}{N}\sum_{i=2}^N\sigma_i - \varepsilon\big)\Big)}.$$

Now we recognize a quenched model on N-1 spins with frozen random fields $\eta_{[2,N]}$, both in numerator and denominator and this is enough to prove the representation in terms of a Curie-Weiss RFIM. $\square$

Once we have recognized this quenched model, we can reduce its large-N analysis to the determination of the minimizers of a corresponding rate-function. For any fixed configuration $\eta_{[1,N]}$ obeying the constraint $\frac{1}{N}\sum_{i=1}^N \eta_i = \alpha$, we use a Gaussian transformation for the empirical σ-sum (Hubbard-Stratonovitch transformation). The result is then the following:

Proposition 2.3 *Let the empirical spin sum $\frac{1}{N}\sum_{i=1}^N \sigma_i$ be distributed according to the quenched measure $\mu_{\beta,\varepsilon,N}[\eta_{[1,N]}]$ with random fields obeying the constraint $\frac{1}{N}\sum_{i=1}^N \eta_i = \alpha$. Denote by G an auxiliary independent Gaussian $\mathcal{N}(0,1)$ variable, and put*

$$X_N = \frac{1}{N}\sum_{i=1}^N \sigma_i + \frac{1}{\sqrt{\beta N}} G.$$

Then the distribution $\mathbb{P}_{\beta,\varepsilon,\alpha;N}$ of the variable X_N has the Lebesgue-density given by

$$\mathbb{P}_{\beta,\varepsilon,\alpha;N}[X_N \in (m, m+dm)] = \frac{e^{-\beta N \Phi_{\beta,\varepsilon,\alpha}(m)} dm}{\int e^{-\beta N \Phi_{\beta,\varepsilon,\alpha}(m)} dm}.$$

where the rate function $\Phi_{\beta,\varepsilon,\alpha}$ is given by (7).

The constraint makes the finite-volume distribution uniquely defined, by permutation invariance. Note that this is an identity for each finite N and in particular the function $\Phi_{\beta,\varepsilon,\alpha}(m)$ is the rate-function for X_N. So we have the convergence of $\frac{1}{N}\sum_{i=1}^N \sigma_i$ to the minimizer of Φ, in case that it is unique.

Proof: We have by definition of the quenched model

$$\begin{aligned}
&\mathbb{P}_{\beta,\varepsilon,\alpha;N}[X_N \in (m, m+dm)] \\
&= \sum_{\sigma_{[1,N]}} \frac{\exp\Big(\frac{\beta}{2N}\big(\sum_{i=1}^N \sigma_i\big)^2 + \beta\sum_{i=1}^N (\varepsilon\eta_i)\sigma_i\Big)}{Z_{\beta,\varepsilon,N}[\eta_{[1,N]}]} \sqrt{\frac{\beta N}{2\pi}} \exp\Big(-\frac{\beta N}{2}\Big(m - \frac{\sum_{i=1}^n \sigma_i}{N}\Big)^2\Big) dm \\
&= \frac{1}{\text{Norm}} \exp\Big(-\frac{\beta N}{2} m^2\Big) \sum_{\sigma_{[1,N]}} \prod_{i=1}^N \exp\big(\beta(m+\varepsilon\eta_i)\sigma_i\big) dm
\end{aligned}$$

with an m-independent normalization. Now use the constraint on $\eta_{[1,N]}$ which amounts to fixing the number of sites i with $\eta_i = +$ (resp. $= -$) and carry out the spin-sum. From that the formula follows easily. $\square$

Proof of Theorem 2.1: We put together the explicit finite-volume representation of Proposition 2.2 of the conditional probabilities with the characterization of Proposition 2.3 of the large-N-limit of the quenched model. Using the continuity of the rate-function in the parameters to get rid of the slight N-dependence (10) and the continuity of function under the expectation to ignore the influence of the Gaussian variable G, we obtain the proof. $\square$

3 Extended phase diagram of nonsymmetric random field model - Bifurcation analysis of the rate-function

We will now discuss the structure of the minima of the rate function (7) when we vary the parameters β^{-1} (temperature), ε (coupling strength of the random fields), α (asymmetry of the distribution). Recall that β^{-1} is the inverse temperature of the system at time $t = 0$ and $\epsilon = \epsilon_t = h_t/\beta$, cf. (3)).

The *extended phase diagram* is the decomposition of the parameter-space into regions where the structure of the rate-function does not change. By this we mean that there is no creation of minima (local bifurcations) and also no degeneracy of two global minima. While the potential function (7) depends on three parameters, the parameter α plays a special role to us here. In particular we are interested in values of the parameters $(\beta^{-1}, \varepsilon)$ for which there is a value of $\alpha = \alpha_c$ such that there is a degeneracy in the depth between two global minima. When we vary α around α_c for fixed (β^{-1}, ϵ), this degeneracy is lifted and the global minimizer will jump. Consequently α_c is a bad configuration (cf. Definition 1.2) for the time-evolved model at time t, by the representation of the limiting conditional probabilities given in Theorem 2.1. Since $t \mapsto h_t$ is a decreasing one-to-one map from $[0, \infty)$ to $[0, \infty)$, the knowledge of the phase diagram will directly indicate the 'Gibbs vs. non-Gibbs' picture during the time-evolution, and thus prove Theorem 1.3. The complete characterizations of the lines of this Figure 1 are easily deduced from the corresponding characterizations in the phase diagram of the forthcoming Theorem 3.1, using the relation $t = \operatorname{atanh}(\mathrm{e}^{-2\beta\varepsilon})$ (See Section 3.1).

For the actual computations it is convenient to make a change of variables

$$E = \beta\varepsilon, \quad M = \beta m.$$

We write here E instead of h_t in order to indicate that it will be considered as an independent variable. We write for the rate-function

$$\Psi_{\beta,E,\alpha}(M) := \beta\Phi_{\beta,\varepsilon,\alpha}(m) = \frac{M^2}{2\beta} - \mathbb{E}_\alpha\big[\log\cosh\big(M + E\tilde{\eta}\big)\big]. \tag{11}$$

Note that $M \mapsto \Psi_{\beta,E,\alpha=0}(M)$ is an even function, and furthermore is remains unchanged under the joint substitution of (α, M) by $(-\alpha, -M)$. Therefore it suffices to consider the case $\alpha \in [0, 1]$.

Using the language of dynamical systems (see [9]) we will encounter two mechanisms of creation of new minima, a *fold bifurcation* and a *pitchfork bifurcation*[3]. This has partially been done for the case $\alpha = 0$ [17, 1, 2] and our analysis will in particular incorporate these results. As we will see however, there are interesting new phenomena arising when we look at general, possibly non-zero values of α. Note that our analysis will be simplified by the linearity of $\Psi_{\beta,E,\alpha}(M)$ in the parameters β^{-1} and α, as we will see below. To state this theorem, we write $\bar{\Psi}_{\beta,\epsilon,\alpha} = \Psi_{\beta,\beta\epsilon,\alpha}$ and get the following Figure 2 describing the different regions of the extended phase diagram :

[3]In dynamical systems one would study the ordinary differential equation $\frac{d}{dt}M = -\partial_M \Psi_{\beta,E,\alpha}(M)$ with time t, while we are interested here only in equilibrium properties.

Region R_{40}: Uniqueness for all α; no phase transition and no degeneracy.

Region R_{01}: Phase transition for $\alpha = 0$; coexistence of two phases, the "+"/"-"-like phases. uniqueness with possibly metastability for $\alpha \neq 0$.

Region R_{12}: Phase transition with metastability for $\alpha = 0$; coexistence of "+"/"-"-like phases, and a "$\pm$"-metastable phase. Uniqueness for $\alpha \neq 0$.

Region R_{23}: Metastability without phase transition; unique neutral "$\pm$"-like phase with two "+"/"-"-like metastable phases when $\alpha = 0$. Coexistence for $\alpha_c \neq 0$.

Region R_{34}: Uniqueness with metastability for $\alpha = 0$, degeneracy at some $\alpha \neq 0$ and phase coexistence for $\alpha_c \neq 0$.

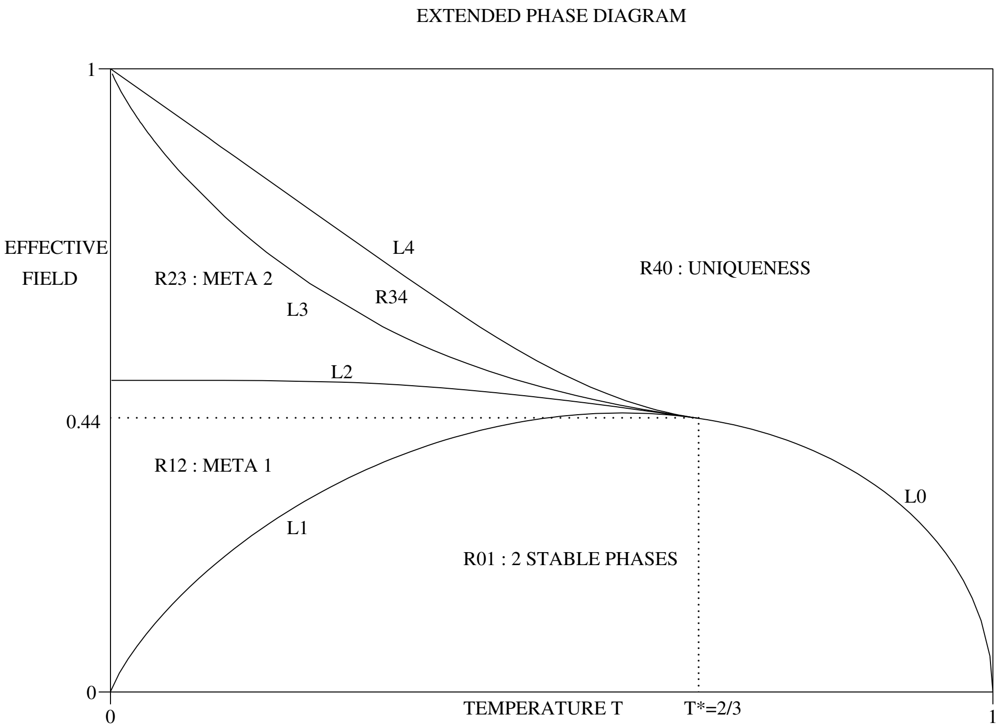

Figure 2: **Extended phase diagram of quenched CW-RFIM**

Theorem 3.1 *1. If* $\beta^{-1} \geq 1$, $\bar{\Psi}_{\beta,\varepsilon,\alpha}$ *has a unique minimizer.*

2. If $\frac{2}{3} \leq \beta^{-1} < 1$, *there exists a sharp value* $0 \leq \varepsilon_0(\beta^{-1}) \leq \varepsilon_* = \frac{2}{3} \cdot \operatorname{atanh}\left(\sqrt{\frac{1}{3}}\right)$ *s.t.*

- *For* $\varepsilon < \varepsilon_0$, $\bar{\Psi}_{\beta,\varepsilon,0}$ *has precisely two symmetric global minimizers* $\pm M^*$ *and* $M = 0$ *is a local maximizer, but* $\bar{\Psi}_{\beta,\varepsilon,\alpha}$ *has a unique global minimizer and no maximizer for all* $\alpha \neq 0$.
- *For* $\varepsilon \geq \varepsilon_0$, $\bar{\Psi}_{\beta,\varepsilon,\alpha}$ *has a unique minimizer for all* $\alpha \in [-1, +1]$.

3. If $0 < \beta^{-1} < \frac{2}{3}$*, there exist sharp values*

$$0 < \varepsilon_1(\beta^{-1}) < \varepsilon_2(\beta^{-1}) < \varepsilon_3(\beta^{-1}) < \varepsilon_4(\beta^{-1}) \text{ s.t.} \tag{12}$$

- *For* $\varepsilon \leq \varepsilon_1$*,* $\bar{\Psi}_{\beta,\varepsilon,0}$ *has precisely two symmetric global minimizers* $\pm M^*$*,* $M = 0$ *is a local maximizer, but* $\bar{\Psi}_{\beta,\varepsilon,\alpha}$ *has a unique global minimizer and no maximizer for* $\alpha \neq 0$*.*
- *For* $\varepsilon_1 < \varepsilon < \varepsilon_2$*,* $\bar{\Psi}_{\beta,\varepsilon,0}$ *has precisely two symmetric global minimizers* $\pm M^*$*,* $M = 0$ *is a local minimizer and there are two symmetric local maximizers* $\pm\hat{M}$ *with* $0 < \hat{M} < M^*$*. All minimizers become global for* $\varepsilon = \varepsilon_2$ *and* $\bar{\Psi}_{\beta,\varepsilon,\alpha}$ *has a unique global minimizer for all* $\alpha \neq 0$*.*
- *For* $\varepsilon_2 < \varepsilon < \varepsilon_3$*,* $\bar{\Psi}_{\beta,\varepsilon,0}$ *has two symmetric local minimizers* $\pm M^*$*,* $M = 0$ *is the unique global minimizer and there are two symmetric local maximizers* $\pm\hat{M}$ *with* $0 < \hat{M} < M^*$*.*
- *For* $\varepsilon_3 \leq \varepsilon$*,* $\bar{\Psi}_{\beta,\varepsilon,0}$ *has a unique global minimizer* $M = 0$ *and no other local minimizer.*
- *For* $\varepsilon_2 < \varepsilon < \varepsilon_4$ *there exists* $\alpha_c \equiv \alpha_c(\beta^{-1}, \varepsilon) > 0$ *s.t* $\bar{\Psi}_{\beta,\varepsilon,\alpha_c}$ *takes its global minimum at precisely two global minimizers.*
- *For* $\varepsilon \geq \varepsilon_4$*,* $\bar{\Psi}_{\beta,\varepsilon,\alpha}$ *has a unique global minimizer, for all* α*.*

The collection of the functions ε_i describes the extended phase diagram (Figure 2). The regions of the $(\beta^{-1}, \varepsilon)$-plane where the organization of the extrema of the rate-function changes, i.e. where our Gibbs property is different for the time-evolved measure, are separated by the lines $L_i, i = 0, 1, 2, 3, 4$ whose equations are given by $\varepsilon_i : D_i \longmapsto [0, 1]$; $\beta^{-1} \longrightarrow \varepsilon_i(\beta^{-1})$, where $D_0 = [\frac{2}{3}, 1], D_1 = D_2 = D_3 = D_4 = [0, \frac{2}{3}]$. we obtain explicitly L_0 and L_1, and we get a parametric characterization of the lines L_3, L_3 and L_4. Since we have an extra explicit form of the curve describing the line L_4, we prefer to state this form in the theorem.

Theorem 3.2 (Characterization of the lines L_i) *:*

1. *There exists a* tricritical point $(\beta_*^{-1}, \varepsilon_*)$ *with* $\beta_*^{-1} = \frac{2}{3}$ *where all the lines* L_i *meet.*
2. L_0 *and* L_1 *are defined explicitly for* $\beta^{-1} \in [0, 1]$ *by the same expression for* $i = 0, 1$*:*

$$\varepsilon_i(\beta^{-1}) = \beta^{-1} \cdot \operatorname{atanh}\big(\sqrt{1 - \beta^{-1}}\big), \ \ \forall \beta^{-1} \in \mathrm{D_i}.$$

3. L_2 *and* L_3 *can be parametrized by* $E > E_*$ *in the* $(\beta^{-1}, \varepsilon)$*-plane, for* $i = 1, 2$*,*

$$\begin{cases} \beta_i^{-1}(E) & = & G_i(M_i(E), E) \\ \\ \epsilon_i(E) & = & \beta_i^{-1}(E) \cdot E \end{cases} \tag{13}$$

where G_i *is explicitly known,* $M_i = M_i(E) \neq 0$ *is given implicitly by* $F_i(M, E) = 0$*, with* F_i *explicitly known.*

4. *The line* L_4 *is explicitly parametrized by* $M > 0$ *in the sense that*

$$\left\{ \begin{pmatrix} \beta^{-1} \\ \varepsilon_4(\beta^{-1}) \end{pmatrix}, 0 < \beta^{-1} < \frac{2}{3} \right\} = \left\{ \beta_4^{-1}(M) \cdot \begin{pmatrix} 1 \\ E_4(M) \end{pmatrix}, 0 < M < \infty \right\} \tag{14}$$

with

$$E_4(M) = \text{atanh}\Big(\sqrt{\frac{2y^3 + M(1-y^2)^2}{2y + M(1-y^2)}}\Big) \tag{15}$$

where $y = y(M) = \tanh(M)$*, and* $\beta_4^{-1}(M)$ *was defined in (6).*

The explicit expressions of the G_i's entering in the parametrizations of the lines L_i for $i = 2, 3$ are given in the proof (Equations (43), (48)). Inequality (12) means that the line L_2 is above L_1, the line L_3 is above L_2 and the line L_4 is above L_3. This fact will follow also from the derivation of (13) and the general discussion from Catastrophe theory. The line L_0 could also be recovered for $E < E_*$ via a variational formula also used to describe L_4 (Equation (24)), and L_4 could also be described in an analogous way as L_2 and L_3. More detailed and explicit local descriptions near the axis and near the tricritical point can be derived from an expansion of the implicit characterizations F_i of the magnetization, also for L_4, and this should be detailed in [11].

Remark: The well-known phase-diagram of the Curie-Weiss random field Ising model with symmetric distribution $\alpha = 0$ was already described by Salinas and Wreszinski [17]. They obtained L_0 analytically and L_2 numerically. There is uniqueness of the Gibbs measure above $L_0 \cup L_2$, and there are two ferromagnetic Gibbs measures related by spin-flip below $L_0 \cup L_2$. We also refer to the information given in [1, 2]. Nevertheless, non-Gibbsianness can occur in *both regions*, as we shall see. So, a more complete analysis is required and we need to know the nature of the extrema as a function of β^{-1}, ε and α. In particular we rediscover the previous result of [17]. For the proof of Theorem 1.3 on non-Gibbsianness with or without non-Gibbsianness, only the lines L_0, L_2, and L_4 are relevant. However, the study of the lines L_1, L_3 is valuable to help the understanding of the bifurcation structure of the rate-function. Moreover it is of independent interest to study the metastable minima of the non-symmetric quenched model, also independently of the question of non-Gibbsianness.

3.1 Bifurcation-structure of the rate function

The structure of the extrema of the rate-function is best understood when we fix $E = \beta\varepsilon$ and vary the temperature β^{-1}, so that we are moving along lines of fixed slope E in the phase diagram. There are two fundamentally different cases: $E \le E_*$ and $E > E_*$ where $E_* = \beta_*\varepsilon_*$. This is due to the fact that the fourth derivative of the rate-function for $\alpha = 0$ vanishes at $M = 0$ if and only if $E = E_*$.

(A) Let us assume first $\alpha = 0$. Take $E \le E_*$ and start from the region above the line L_0 (high-temperature region). Then the rate function has a unique minimizer. When we lower β^{-1} we will eventually cross the line L_0 from the outside. At L_0 we encounter a *pitchfork bifurcation*: The unique minimizer $M = 0$ becomes a local maximizer and two symmetric minimizer are emerging from it (Figure 6 in Section 4). This does not happen when $\alpha \neq 0$ (Figure 7).

(B) Let us now fix $E > E_*$. Put again $\alpha = 0$ first. We start from the region below L_0-L_1 (i.e. below L_1 when $E > E_*$) where there are two global minima and no maximum (cf. the proof), and increase the temperature β^{-1} such that we move along lines of constant E: Then we first see a pitchfork-bifurcation at the line L_1 that produces a new minimum at $M = 0$ and two local maxima. When we further increase β^{-1} this minimum gets deeper and deeper w.r.t. to the depth of the non-zero minimum (Figure 8 and 9). Next, when we reach the line L_2 the depth of the minima becomes equal (Figure 10). When we increase the temperature β^{-1} the zero-minima become deeper. The non-zero minima will finally vanish at the line L_3 where they collide with the neighboring local maxima (Figure 13). The mechanism of this is a *fold bifurcation*, to be discussed below.

When we vary α around 0, the degeneracy of the global minima for the non-zero minimizer is lifted in the open region below L_0-L_2. The global minimizer jumps from a positive value to a negative value when we pass $\alpha = 0$ from positive values of α to negative values of α. This discontinuity implies that $\alpha = 0$ is a bad configuration for the time-evolved measure, and this does not happen for other α's (Figures 13 and 14).

Let us suppose we are in the open region between L_2 and L_3 (Figure 14). Then, by increasing α from zero, the minimum of the positive minimizer will become deeper w.r.t. the (perturbed) minimum around zero. At the critical value $\alpha_c(\beta^{-1}, \varepsilon)$ the depth of the two minima becomes equal. The depth of the negative minimizer will decrease and play no role for us. This mechanism creates a bad configuration for the time-evolved measure at a non-zero value of α.

This sort of degeneracy for the specifically chosen $\alpha = \alpha_c(\beta^{-1}, \varepsilon)$ is also possible in the whole open region between L_2 and L_4 (Figure 15), even though there is a unique minimizer between L_3 and L_4 for $\alpha = 0$. To understand this, we remark that increasing α can create a new minimum. Indeed, suppose we are in particular on the line L_3. Then there is a fold bifurcation driven by the parameter α at $\alpha = 0$, for fixed $(\beta^{-1}, \varepsilon)$, that creates a new local minimizer for α positive (Figure 14). This creation of a new minimizer will also occur in the region between L_3 and L_4, but at strictly positive values of α. Further increasing the α will then create equal depth of these minima at $\alpha = \alpha_c(\beta^{-1}, \varepsilon)$. Finally, suppose we are looking at the rate function for $\alpha = \alpha_c(\beta^{-1}, \varepsilon)$, so that we have two equal-depth global minima, and increase β^{-1} along the line of constant E. Then the two equal-depth minima will eventually collapse, by means of a pitchfork bifurcation, and this defines the line L_4.

3.2 General facts about the rate-function

In order to discuss the phase diagram we will need to consider the first four derivatives w.r.t. M. If we denote ∂_V^k the k^{th} derivative with respect to any variable V, they read for $V = M$,

$$\partial_M \Psi_{\beta,E,\alpha}(M) = \frac{M}{\beta} - \mathbb{E}_\alpha\big[\tanh\big(M + E\tilde{\eta}\big)\big]$$

$$\partial_M^2 \Psi_{\beta,E,\alpha}(M) = \beta^{-1} - 1 + \mathbb{E}_\alpha\big[\tanh^2\big(M + E\tilde{\eta}\big)\big] = \beta^{-1} - \mathbb{E}_\alpha\big[\cosh^{-2}\big(M + E\tilde{\eta}\big)\big]$$

and further

$$\partial_M^3 \Psi_{\beta,E,\alpha}(M) = 2\mathbb{E}_\alpha\Big[\Big(\tanh(1 - \tanh^2)\Big)\big(M + E\tilde{\eta}\big)\Big]$$

and

$$\partial_M^4 \Psi_{\beta,E,\alpha}(M) = 2\mathbb{E}_\alpha\Big[\Big(1 - 4\tanh^2 + 3\tanh^4\Big)\big(M + E\tilde{\eta}\big)\Big]$$
$$= 2\mathbb{E}_\alpha\Big[\Big(3\cosh^{-4} - 2\cosh^{-2}\Big)\big(M + E\tilde{\eta}\big)\Big].$$

Any minimizer of the rate function must satisfy the so called mean-field equation

$$\frac{M}{\beta} = \mathbb{E}_\alpha\big[\tanh\big(M + E\tilde{\eta}\big)\big]. \tag{16}$$

For any real function $f(M,E)$ we use the notations

$$\bar{f}(M,E) = \frac{1}{2}(f(M+E)+f(M-E)) \text{ and } \Delta f(M,E) = \frac{1}{2}(f(M+E)-f(M-E)) \tag{17}$$

for the symmetric average and the discrete derivative around M. In this spirit we write e.g.

$$\partial_M \Psi_{\beta,E,\alpha}(M) = \frac{M}{\beta} - \overline{\tanh}\big(M,E\big) - \alpha\Delta\tanh\big(M,E\big).$$

Let us add a few general observations. Note that $\Psi_{\beta,E,\alpha}(M) \sim \frac{M^2}{2\beta}$ as $|M| \uparrow \infty$ and since it is continuous and bounded below its infimum is attained at one of its local minima.

Lemma 3.3 *For any choice of the parameters (β,E,α), the function $M \mapsto \Psi_{\beta,E,\alpha}(M)$ has at most three local minima. For $\beta^{-1} \geq 1$, it has no maximum and its minimum is the only local extremum.*

Proof: To prove the desired bound on the number of local minima it suffices to show that the number of zeroes of $M \mapsto \partial_M^2 \Psi_{\beta,E,\alpha}(M)$ (the number of the *inflection points* of the rate-function) is bounded by 4, since the function is bounded below. To do this we prove that, for any (α,E), the number of local extrema of the function $M \mapsto \mathbb{E}_\alpha\Big(\cosh^{-2}\big(M + E\tilde{\eta}\big)\Big)$ is bounded by 3. This function is non-negative and tends to zero with $|M| \uparrow \infty$. So we see by a look at the expression for the second derivative of the rate-function that this implies indeed that, for any choice of β the number of the zeroes of its second derivative is bounded by 4. Now, taking one more M-derivative, we may rewrite

$$\partial_M \mathbb{E}_\alpha\Big(\cosh^{-2}\big(M + E\tilde{\eta}\big)\Big) = -\partial_M^3 \Psi_{\beta,E,\alpha}(M) = -2\mathbb{E}_\alpha\Big(\tanh(1-\tanh^2)\big(M + E\tilde{\eta}\big)\Big)$$
$$= -\frac{4z}{(z+1)^3(z+c)^3}\Big((1+\alpha)(z-1)(z+c)^3 + (1-\alpha)c(z-c)(z+1)^3\Big)$$

where we have put $z = e^{2(M+E)} > 0$, $c = e^{4E} > 1$. Let us look at the number of zeros of this expression, for fixed α, c. Clearly, the polynomial $z \mapsto (1+\alpha)(z-1)(z+c)^3 + (1-\alpha)c(z-c)(z+1)^3$ has at most 4 real zeros, for any value of (α,c). However, we observe that there will be always one negative zero z_-. This is clear since the polynomial turns to $+\infty$ for $z \downarrow -\infty$ and its value at $z=0$ is negative. But since this negative zero z_- does not correspond to a real value of M we are done : the rate-function has at most three minima. When $\beta^{-1} > 1$ or $\epsilon > 1$, the absolute value derivative of the rate-function is always strictly bigger than one, and one deduces easily that $M=0$ is the only extremum, and is a global minimum. Similar adapted arguments also hold for $\beta^{-1} = 1$. $\square$

3.3 Catastrophe manifold

To understand the structure of the minima of the *potential function* $M \mapsto \Psi_{\beta,E,\alpha}(M)$ when the three parameters (β, E, α) are varied is an elementary but not completely trivial analytical problem.

We find it useful to introduce to this end some systematic basic notions from Catastrophe theory going back to Thom and Zeeman (See [16]), to the level that we need them. What is the benefit of Catastrophe theory in a probabilistic paper? On the one hand, Catastrophe theory allows for the classifications of the possible singularities of local bifurcations of potential functions, as provided by Thom's theorem. It makes rigorous that it suffices to look at polynomial approximations for the potential around critical points to understand locally the nature of bifurcations, that is the changes of the structure of the extrema of the potential. In fact, in our present case we will recognize the so-called *butterfly-singularity* which appears for a four-dimensional family, while we view it here from a three-dimensional sub-space. Indeed, our family has only one parameter α that distinguishes M from $-M$, while in the full butterfly-unfolding one has two such "odd" parameters. As a consequence, not all symmetries are lifted in our present problem.

However, more than just locally identifying the nature of the singularity around a critical point, we manage to provide a fairly explicit analysis over the whole manifold. This is possible by carefully exploiting the specific form of our potential. Taking an elementary but useful message from Catastrophe theory we use M as a coordinate and do not try to solve the appearing equations for M directly. This will be useful throughout the analysis, and allows us in particular to find the explicit parametric form of boundary curve between Gibbsian- and non-Gibbsian parts of the phase diagram. Here the magnetization (or state-variable) M appears as parameter of the curve.

Let us now proceed with some basic definitions. The *Catastrophe manifold* X is the set of points in the product space of state space $\mathbb{R}$ and parameter space (also called the space of controls) $C = \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}$ satisfying the mean field equation

$$X = \left\{(M, \beta^{-1}, E, \alpha) \in \mathbb{R} \times C | \partial_M \Psi_{\beta,E,\alpha}(M) = 0\right\}.$$

It is a smooth sub-manifold of $\mathbb{R}^4$ which can be seen by inverting the mean-field equation for α. Indeed, this defines a function

$$\alpha_{\mathrm{MF}}(M, \beta^{-1}, E) := \frac{\frac{M}{\beta} - \overline{\tanh}\big(M, E\big)}{\Delta \tanh\big(M, E\big)} \tag{18}$$

which is well-defined since $E \neq 0$ implies that the denominator does not vanish. Hence X can be written as a graph of the smooth map α_{MF} with the coordinates (M, β^{-1}, E).

The points where this manifold folds "under", when we view the direction of the state variable M is the vertical direction, are points of the vertical tangent plane, and these are given by the vanishing of the curvature of the potential,

$$\partial_M^2 \Psi_{\beta,E,\alpha}(M) = 0.$$

The projection to the parameter space of this space, is called the *bifurcation set* B, i.e.

$$B := \left\{(\beta^{-1}, E, \alpha) | \exists M \in \mathbb{R}, \partial_M \Psi_{\beta,E,\alpha}(M) = 0, \partial_M^2 \Psi_{\beta,E,\alpha}(M) = 0\right\}.$$

It is our aim to describe first this bifurcation set. For our analysis of non-Gibbsianness we are of course interested in the points of global bifurcations where two minima acquire equal depths. However, to understand these, we first discuss local bifurcations, i.e. the production resp. merging of minima.

In the next step we show how the analysis of the local bifurcations (production of new minima) of the full three-parameter problem can be reduced to a bifurcation problem of a potential function of one variable and the remaining one-dimensional control variable E. Our point is that we can do this globally, that is over the whole manifold, and not only near a critical point. To do that we are now going to exploit the linear nature of the potential as a function of the parameters β^{-1} and α. We have the following result for this first step.

Proposition 3.4 *Define the functions*

$$\beta_{12}^{-1}(M,E) := \frac{1+\tanh(M+E)\tanh(M-E)}{1+M(\tanh(M+E)+\tanh(M-E))} \tag{19}$$

and

$$\alpha_{12}(M,E) := \frac{M\Big(1-\overline{\tanh^2}\big(M,E\big)\Big)-\overline{\tanh}\big(M,E\big)}{\Delta\tanh\big(M,E\big)+M\Delta\tanh^2\big(M,E\big)} \tag{20}$$

Then the bifurcation set has the parametric representation

$$B=\Big\{\big(\alpha_{12}(M,E),\beta_{12}^{-1}(M,E),E\big),E>0,M\in\mathbb{R}\Big\} \tag{21}$$

Remark: (21) gives a parametric representation of the bifurcation set B. The geometry of this set, consisting of a two-dimensional surface with self-intersections, is best understood when we look at intersections of B with planes of fixed E. Performing parametric plots for various fixed values of E, with running variable M we see a qualitative change of the curves in (β^{-1},α)-space at $E=E^*$.

In going from values $E<E_*$ we see a *cusp-like shape* (Figure 3) that unfolds into a *pentagram-shaped curved shape* for $E>E_*$ (Figure 4). We like to call it pentagram-shaped although the leftmost corner in Figure 4 lies within the region that is bounded by the vertical axis and the line-segments extending from the righmost corners.

This singularity is well known from the so-called *butterfly unfolding* (compare [16], from p 178). We note that $\alpha_{12}(M,E)=-\alpha_{12}(-M,E)$ and $\beta_{12}^{-1}(M,E)=\beta_{12}^{-1}(-M,E)$, so the curve is symmetric about the axis $\alpha=0$. The critical value $E=E^*$ is explained by the fact that $\partial_M^4\Psi_{\beta^{-1},\alpha=0,E}(M=0)$ changes signs at $E=E^*$.

We just mention that, allowing for another parameter $H=\beta h$, appearing as a magnetic field, we expect that this symmetry would be lifted, leading to the full unfolding of the butterfly-singularity, for the four-parameter family (α,β^{-1},E,H). We will analyze this and the relation to non-Gibbsianness more precisely in a follow-up paper ([11]).

Proof of the Proposition: We look at the defining equations of the bifurcation manifold B

$$\begin{cases} \frac{M}{\beta}=\overline{\tanh}\big(M,E\big)+\alpha\Delta\tanh\big(M,E\big) \\ \\ \frac{1}{\beta}=1-\overline{\tanh^2}\big(M,E\big)-\alpha\Delta\big(\tanh^2\big)\big(M,E\big). \end{cases}$$

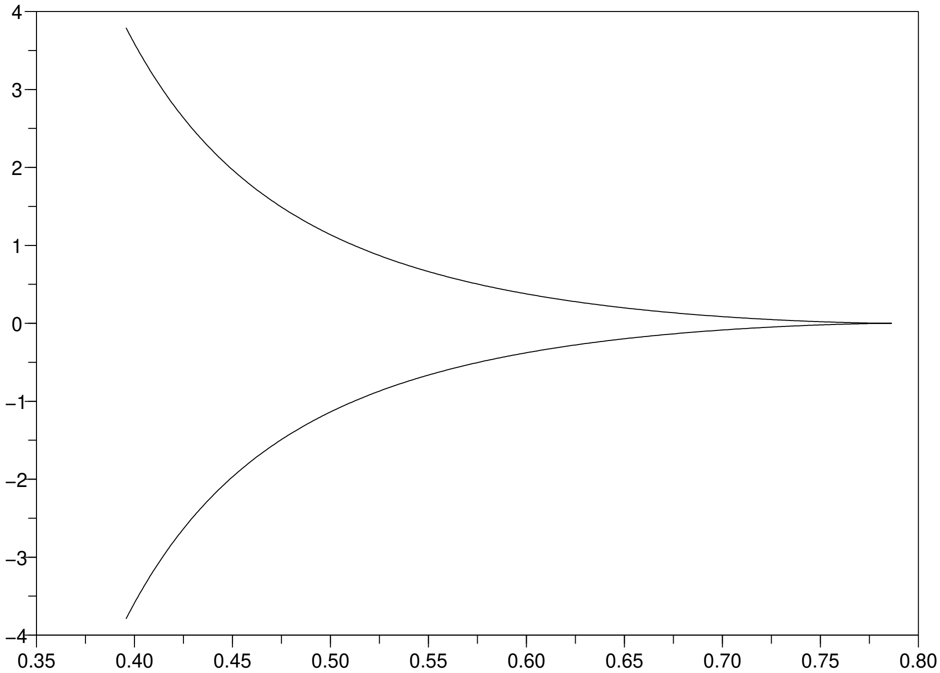

Figure 3: **Bifurcation set: α vs. β^{-1} at fixed $E = 0.6 < E^*$ - Cusp bifurcation**

Fix M and $E > 0$ and consider this as a linear system for β^{-1} and α. Then the proposition follows from the following statement: For all $M \in \mathbb{R}$, $E > 0$, the above linear system in (β^{-1}, α) has always a unique solution of the given form $(\beta_{12}^{-1}(M,E), \alpha_{12}(M,E))$. Now, to show this we note that the determinant of the corresponding coefficient matrix becomes

$$\Big(1 + 2M\overline{\tanh}\big(M,E\big)\Big)\Delta\tanh\big(M,E\big)$$

where we used that

$$\Delta\big(\tanh^2\big)\big(M,E\big) = 2\overline{\tanh}\big(M,E\big)\Delta\tanh\big(M,E\big).$$

We claim first that it is always strictly positive, for $E > 0$. Indeed, $E > 0$ implies that $\Delta\tanh\big(M,E\big) > 0$. Furthermore $M\overline{\tanh}\big(M,E\big) \geq 0$ is always true. Now, eliminating α we obtain the expression

$$\beta_{12}^{-1}(M,E) = \frac{1 - \overline{\tanh^2}\big(M,E\big) + 2\Big(\overline{\tanh}\big(M,E\big)\Big)^2}{1 + 2M\overline{\tanh}\big(M,E\big)}.$$

It is then a simple computation to rewrite the function $\beta_{12}^{-1}(M,E)$ in the form of the Lemma. The form of $\alpha_{12}(M,E)$ follows in the same way. $\square$

The point $\big(E_*, \alpha = 0, \beta^{-1} = \frac{2}{3}\big)$ is characterized by the fact that the first four derivatives of the potential at $M = 0$ vanish. Correspondingly we have just observed

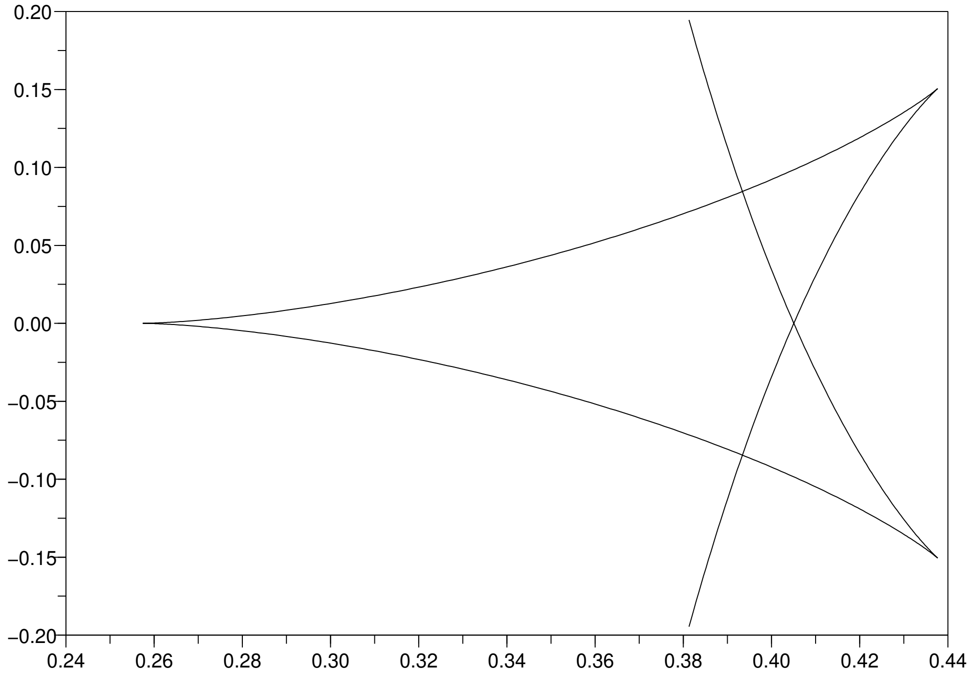

Figure 4: **Pentagram-shaped bifurcation : α vs. β^{-1} at fixed $E = 1.3 > E^*$**

the unfolding of a singularity that resembles the unfolding in the 3-parameter toy-model with potential $\frac{M^6}{6} + \frac{a_4}{4}M^4 + \frac{a_2}{2}M^2 + a_1 M$ with parameters a_1, a_2, a_4 at $a_1 = a_2 = a_4 = 0$. (Compare with the *butterfly catastrophe* in [16], from p 178).

Let us discuss the number of minima of the potential function $\Psi_{\beta^{-1},\alpha,E}(M)$ for parameter values outside of B. This structure is well-known for the polynomial model. For $E < E^*$, the potential has a single minimum outside the cusp-shaped region in Figure 5. Within the cusp, the potential has two minima which are created by a fold bifurcation by crossing a branch of the cusp from the outside. This singularity is the well-known as the *cusp-catastrophe*, see [16] p. 174.

For $E > E^*$, the situation is more complicated and well-described for instance in figure 9.12 in [16]. To the right of the pentagram-shaped curve the potential has a unique minimum. In the two triangle-shaped regions the potential has two minima. For the upper triangle one of these minima is created by a fold crossing the upper boundary from above; the other one is created by crossing the right boundary from the left. In the four-corned region the potential has three minima. Finally, in the region to the left, the potential has two minima.

Since the number of minima is a constant for parameter values in the complement of the bifurcation set, we know the number of minima of the potential for any parameter values by extension from the known shape of the potential locally around the critical point, when we know that there are no other types of singularities appearing. That is, we want to prove now that over the whole phase diagram, and not only locally around

$(E_*, \alpha = 0, \beta^{-1} = \frac{2}{3})$, this structure persists.

3.3.1 Exact form of L_4 from one-dimensional bifurcation problem

Additionally to fix E, we fix now the parameter β^{-1}. We want to characterize the points α in the bifurcation set over the (β^{-1}, E)-plane. We will see that there are either $0, 1, 2, 3$ or 4 points of such α, and no more, for any fixed (β^{-1}, E), as the plots of the fixed E-slices of B given in Figure 3 and 4 suggest.

We have from the proposition

$$\Big\{\alpha \in \mathbb{R} | (\alpha, \beta^{-1}, E) \in B\Big\} = \Big\{\alpha_{\mathrm{MF}}(M, \beta^{-1}, E), \exists M \in \mathbb{R} : \beta^{-1} = \beta_{12}^{-1}(M, E)\Big\} \tag{22}$$

Where α_{MF} is given by (18). In this way we see that by varying the independent parameter β^{-1} we change the number of solutions for M depending on the form of the function $\beta_{12}^{-1}(M, E)$ and the number of its local extrema. The corresponding M gives us a solution $\alpha = \alpha_{12}(M, E)$ on the bifurcation set B. In particular $\beta^{-1} \geq \sup_{M \in \mathbb{R}} \beta_{12}^{-1}(M, E)$ implies that there are no solutions for α, and hence the potential $\Psi_{\beta^{-1}, E, \alpha}$ has a unique minimizer, for any α. We note that the map of M to α is not injective at degenerate points. This is responsible for the occurrence of three values of α in the bifurcation set at fixed values of (β^{-1}, E) that correspond to the common point of the two triangles in Figure 4.

To understand the number of solutions for M in (22) we can obtain for different E, we need to proceed now with the bifurcation-analysis of $M \mapsto \beta_{12}^{-1}(M, E)$, viewed as (the inverse of) an auxiliary potential function of the one-dimensional parameter E.

Remark that one can rewrite

$$\beta_{12}^{-1}(M, E) = (1 - \tanh^2(E)) \cdot \frac{1 + \tanh^2(M)}{1 - \tanh^2(M)\tanh^2(E) + 2M\tanh(M)(1 - \tanh^2(E))} \tag{23}$$

This representation follows from simple computation using the formula for $\tanh(x + y)$. Let us denote for $E \leq E_*$

$$\beta_0^{-1}(E) := \sup_{M \in \mathbb{R}} \beta_{12}^{-1}(M, E) \tag{24}$$

and for $E \geq E_*$

$$\beta_4^{-1}(E) := \sup_{M \in \mathbb{R}} \beta_{12}^{-1}(M, E). \tag{25}$$

In Figure 4 the temperature $\beta_4^{-1}(E)$ appears as the projection of the right corners of the pentagram to the β^{-1}-axes. We shall see in this proof that these functions indeed define the main critical line L_0-L_4 of the phase diagram. The importance of this line lies in the fact that its image to the time-temperature space is in fact the boundary curve of the Gibbsian region given in Figure 1, but we have chosen different names to distinguish qualitative different behaviors, and thus different expressions.

One finds for $E < E_*$

$$\beta_0^{-1}(E) = 1 - \tanh^2(E)$$

(see also the next section). This functional form for the maximum over M is not true anymore for $E > E_*$, as we will see below. Instead we define for $E \geq E_*$ the function

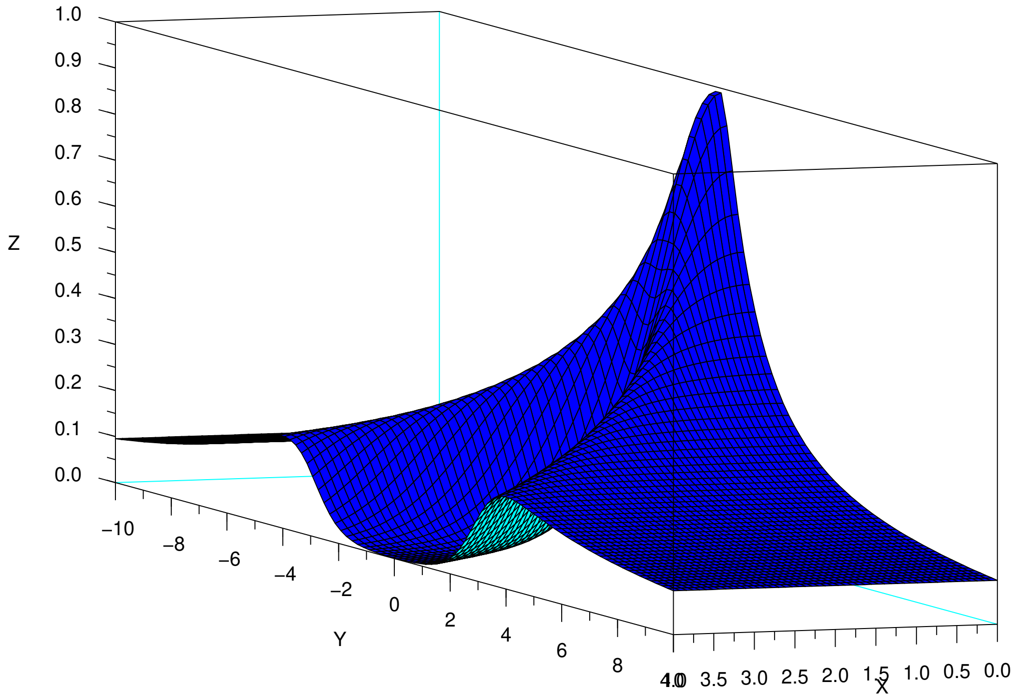

Figure 5: **Pitchfork-bifurcation driven by E at $E = E_*$**: Plot of the auxiliary potential function $\beta_{12}^{-1}(X = E, Y = M, Z = \beta^{-1})$.

$\beta_1^{-1}(E) := \beta_{12}^{-1}(M = 0, E) = 1 - \tanh^2(E)$ by this very same functional form as β_0^{-1}. This is again corresponding to the line L_1 in the phase diagram presented earlier. This temperature appears as the leftmost corner in Figure 4.

To perform actual computations it is more convenient to consider

$$\beta_{12}(M, E) := \frac{1}{\beta_{12}^{-1}(M, E)}$$

and to view this as a potential function with state space M and control parameter E.

We denote by Y the corresponding *Catastrophe set*

$$Y = \{(M, E) \in \mathbb{R} \times \mathbb{R}_+ | \partial_M \beta_{12}(M, E) = 0\}$$

that is the set of critical points (M, E), with M taking values in the state space $\mathbb{R}$ and E taking values in the control space $\mathbb{R}_+$.

What is remarkable is that we can find an explicit representation of $\bar{Y}$ as a smooth curve in terms of elementary functions. The reason for this is that taking the inverse of (23) gives us

$$\beta_{12}(M, E) = \beta_1(E)\Big(U_0(M) - (\tanh^2 E)U_1(M)\Big) \tag{26}$$

with the functions

$$U_0(M) = \frac{1 + 2M \tanh M}{1 + \tanh^2 M} \text{ and } U_1(M) = \frac{\tanh^2 M + 2M \tanh M}{1 + \tanh^2 M}.$$

The following proposition describes the global structure of $\beta_{12}(M, E)$.

Proposition 3.5 *The function $M \mapsto \beta_{12}(M, E)$ undergoes a pitchfork bifurcation where two minima are created by increasing E at the point E_*. More precisely, the Catastrophe set has the form $Y = \bar{Y} \cup \Big(\{0\} \times \mathbb{R}_+\Big)$ where*

$$\bar{Y} = \Big\{\big(M, E_4(M)\big), M \in \mathbb{R}\Big\} \tag{27}$$

is the graph of the function

$$E_4(M) = \frac{1}{2} \log \frac{1 + y_4(M)}{1 - y_4(M)} \tag{28}$$

where $y_4(M) \equiv \sqrt{y_4^2(M)} \geq 0$ is given by

$$y_4^2(M) := \frac{U_0'(M)}{U_1'(M)} = \frac{2 \tanh^3 M + M(1 - \tanh^2 M)^2}{2 \tanh M + M(1 - \tanh^2 M)} \tag{29}$$

for $M \neq 0$, and $y_4^2(M = 0) = 0$.

The bifurcation set of (26), set of E's for which there exists an M such that

$$\partial_M \beta_{12}(M, E) = \partial_M^2 \beta_{12}(M, E) = 0, \tag{30}$$

consists of the single point $E = E_*$.

Remark: $\bar{Y} = 0$ describes a line of global maximizers of $M \mapsto \beta_{12}(M, E)$. Varying the parameter E, we thus have the following picture:

- If $E \leq E_*$ the function $M \mapsto \beta_{12}(M, E)$ has its unique local minimum at the unique minimizer $M = 0$, and no other local extrema.
- If $E > E_*$ the minimizer of the function $M \mapsto \beta_{12}(M, E)$ are given by a pair $\pm M_4(E)$, with $M_4(E) > 0$. $M = 0$ is a local maximizer and there are no other local extrema.

Proof: It is easy to show that a pitchfork bifurcation occurs at $E = E_*$. In fact, for this suffices to show that, at $E = E_*$ and $M = 0$ the derivatives satisfy $\partial_M^2 \beta_{12} = 0$ $\partial_E \partial_M^2 \beta_{12} < 0$ and $\partial_M^4 \beta_{12} > 0$. It is easy to check this, and we skip details here.

More interesting is the global analysis. It implies in particular that there are no other bifurcations appearing and the local picture around $E = E_*$ and $M = 0$ extends to the whole space. Note that $\partial_M \beta_{12}(M, E) = 0$ is equivalent to

$$U_0'(M) = (\tanh^2 E) U_1'(M) \tag{31}$$

and also that the parameter $E > 0$ appears only in terms of multiplication by a function, so that $y^2 = \tanh^2 E$ acts like a re-parametrized linear control variable. It is easy to check that this equation for M always has the solution $M = 0$. Next, we find all the other solutions (after dividing out $M = 0$) by putting

$$\tanh^2 E = y_4^2(M) \tag{32}$$

To show that there are no further bifurcation points other than $E = E_*$ realize that the equations $\partial_M \beta_{12}(M,E) = \partial_M^2 \beta_{12}(M,E) = 0$ imply that $\partial_M y_4^2(M) = 0$. But, given the explicit form (29) of this function of one variable and no parameter, it is an elementary task to show that this implies that $M = 0$. However, from this follows by the second equation $\partial_M^2 \beta_{12}(M = 0, E) = 0$ that in fact $E = E_*$ is the only point in the bifurcation set. This finishes our proof. □

As an immediate and important consequence of this we get the desired exact form for the line L_4.

Theorem 3.6 *The line L_4 has the parametric representation of the form*

$$\left\{(\beta_4^{-1}(E), E), E \geq E_*\right\} = \left\{\Big(\frac{1 - y_4^2(M)}{U_0(M) - y_4^2(M) U_1(M)}, \frac{1}{2}\log\frac{1 + y_4(M)}{1 - y_4(M)}\Big), M > 0\right\}. \tag{33}$$

Remark: Note M is really the inverse temperature times the magnetization that the system acquires at the melting of the minima at $\alpha_c(\beta, \varepsilon)$.

As a corollary we obtain from here the simple form of L_4 that was given Theorem 3.2 by a simple computation using the explicit expressions.

Proof: Recall that $\beta_4^{-1}(E)$ is the maximum over M of $\beta_{12}^{-1}(M,E)$ for $E \geq E_*$. But from the previous proposition follows that a *local* maximum is achieved along the curves provided by $\bar{Y}$ in (M,E)-space, for $E \geq E_*$. Since there are no other bifurcations appearing, and the functions goes to zero with $|M| \uparrow \infty$ these two symmetric local maxima must automatically be global maxima.

So we may parametrize the curve in (β^{-1}, E)-space in terms of the corresponding M, the value of β^{-1} taken at $\beta_{12}^{-1}(M, E_4(M))$. From this the representation follows by substitution of the explicit expressions. □

Having understood the form and bifurcation of $\beta_{12}(M,E)$ it is now immediate to go back to the full bifurcation problem and understand the sets of solutions of M for the equation appearing on the r.h.s. of (22). We fix this result in the form of the following corollary.

Corollary 3.7 *Consider the equation $\beta^{-1} = \beta_{12}^{-1}(M,E)$.*

- *Fix $E \leq E_*$.*
  - *For $\beta^{-1} \geq \beta_0^{-1}(E)$, it has no solutions M.*
  - *For $\beta^{-1} = \beta_0^{-1}(E)$, it has precisely one solution $M = 0$.*

- *For* $\beta^{-1} < \beta_0^{-1}(E)$*, it has precisely a pair of solutions* $\pm M_0(E) \neq 0$*.*

- *Fix* $E > E_*$*.*
  - *For* $\beta^{-1} \geq \beta_4^{-1}(E)$*, it has no solutions* M*.*
  - *For* $\beta^{-1} = \beta_4^{-1}(E)$*, it has precisely the pair of solutions* $\pm M_4(E)$*.*
  - *For* $\beta_1^{-1} < \beta^{-1} < \beta_4^{-1}(E)$*, it has two pairs of solutions* $\pm M_0(E), \pm M_4(E)$*, where* $0 < M_0(E) < M_4(E)$*.*
  - *For* $\beta^{-1} = \beta_1^{-1}(E)$*, it has the solutions* $M = 0$ *and* $\pm M_0(E) \neq 0$*.*
  - *For* $\beta^{-1} < \beta_1^{-1}(E)$*, it has the pair of solutions* $\pm M_0(E) \neq 0$*.*

We have understood that the set of α's that lie in the bifurcation set above a point (β^{-1}, E), is given by applying the map α_{12} to the solutions of the above equation. However, this map is not always injective, indeed the curve $M \mapsto (\alpha_{12}(M,E), \beta_{12}^{-1}(M,E))$ has double-points for $E > E_*$ (we see from the pentagram-shaped Figure 4 that there are three double-points). Indeed, denote by $M_3(E)$, for $E > E_*$, the positive solution of $\alpha_{12}(M,E) = 0$. Then we have $\beta_3^{-1}(E) = \beta_{12}^{-1}(M_3(E), E)$. This is the point of fold-bifurcation at $\alpha = 0$. However, this point has no role in the discussion of non-Gibbsianness.

3.3.2 Global bifurcations: Degeneracy of minima

Let us now finally turn to the discussion of the set of α's such that $M \mapsto \Phi_{\beta,\varepsilon,\alpha}(M)$ has two different global minimizers. We recall that the temperature $\beta_2^{-1}(E)$ is the unique temperature of equal depth of all three minima of the potential at $\alpha = 0$. Therefore it is clear that $\beta_0^{-1}(E) \leq \beta_2^{-1}(E) \leq \beta_3^{-1}(E)$. Exact characterizations of the latter two temperatures will be given using different techniques next section.

Proposition 3.8 *For any* $\beta^{-1} \in (\beta_2^{-1}(E), \beta_4^{-1}(E))$ *there is a unique* $\alpha_c(\beta^{-1}, E) > 0$ *such that there are two different global minima of equal depth. For any other* β^{-1} *the global minimum is unique for any* $\alpha \neq 0$*.*

Proof of the Proposition: Throughout the discussion $E > E_*$ is fixed. Now, for $\beta = \beta_2(E)$, we have equal depth for the three minimima of Φ at $\alpha = 0$. By increasing α, the corresponding minimizers will move along smooth curves, until they vanish. The unicity of the value of α_c will now follow from the following lemma.

Lemma 3.9 *Suppose that, for fixed* (β, E)*, there are two different smooth curves* $\alpha \mapsto M_i(\beta, \alpha)$*,* $i = 1, 2$*,* α *running in the same interval* $\alpha \in I$*, of local extreme points, where* $M_1(\beta, \alpha) < M_2(\beta, \alpha)$*.*

Then the function $\alpha \mapsto \Psi_{\beta,E,\alpha}(M_2(\beta,\alpha)) - \Psi_{\beta,E,\alpha}(M_1(\beta,\alpha))$ *is decreasing in* α*. In particular, there is at most one value of* α *for which* $\Psi_{\beta,E,\alpha}(M_1(\beta,\alpha)) = \Psi_{\beta,E,\alpha}(M_2(\beta,\alpha))$*.*

Proof of the Lemma: Looking at the evolution of the values of the rate-function at both curves of extreme points with the parameter α we have

$$\frac{d}{d\alpha} \Psi_{\beta,E,\alpha}(M_i(\beta,\alpha)) = -\frac{1}{2} \log \frac{\cosh(M_i(\beta,\alpha) + E)}{\cosh(M_i(\beta,\alpha) - E)}. \tag{34}$$

This is clear since the M-derivative of the rate-function vanishes at the minima. But it is elementary to see that the function $M \mapsto -\frac{1}{2}\log\frac{\cosh(M+E)}{\cosh(M-E)}$ is decreasing with M (and odd in M), for any fixed $E > 0$. This implies the claim. $\square$

We start at some point $\beta^{-1} \in (\beta_2^{-1}(E), \beta_3(E)^{-1})$ and $\alpha = 0$. Here precisely three local minimizers $\tilde{M}_0 = 0$ and $\tilde{M}_1 = -\tilde{M}_{-1}$ with $\tilde{M}_1 > 0$ exist. Furthermore we have $\Phi_{\beta^{-1},E,\alpha=0}(\tilde{M}_0) < \Phi_{\beta^{-1},E,\alpha=0}(\tilde{M}_1)$. Indeed, recall that $\beta_2(E)$ is the boundary curve for ferromagnetic order in the phase-diagram of the well-known symmetric ($\alpha = 0$) Curie-Weiss random field Ising model. Now, increasing α these minimizers move along curves $\tilde{M}_i(\alpha)$. The lemma implies that $\Phi_{\beta^{-1},E,\alpha}(\tilde{M}_0(\alpha)) - \Phi_{\beta^{-1},E,\alpha}(\tilde{M}_{-1}(\alpha))$ will decrease, and $\Phi_{\beta^{-1},E,\alpha}(\tilde{M}_1(\alpha)) - \Phi_{\beta^{-1},E,\alpha}(\tilde{M}_0(\alpha))$ will decrease, along these curves. Hence $\tilde{M}_{-1}(\alpha)$ will play no role for the discussion of the global minimum because the zero-type minimum is already deeper. It is well-known from the geometry of the butterfly bifurcation and the form of $\Phi_{\beta^{-1},E,\alpha}$ in the pentagram-shaped region that the two left minima are vanishing by fold-bifurcations when increasing α. The point of α when the curve $M_{-1}(\alpha)$ ends and $M_{-1}(\alpha)$ vanishes by a fold bifurcation will thus play no role. When the point $M_0(\alpha)$ vanishes by a fold bifurcation, at the corresponding point α^* we must have $\Phi_{\beta^{-1},E,\alpha^*}(\tilde{M}_1(\alpha^*)) - \Phi_{\beta^{-1},E,\alpha^*}(\tilde{M}_0(\alpha^*)) < 0$.

Hence there must be a point $\alpha_c(\beta^{-1}, E)$ when the two right-most minima have equal depth. This point must be unique because the lemma excludes double crossings.

For the other regions of the phase diagram we use the same arguments. We see in that way that for any other β^{-1} the global minimum is unique for $\alpha \neq 0$. Indeed, consider $\beta^{-1} < \beta_2^{-1}(E)$, $E > E_*$. Then the plus-type minimum is already deeper than the zero-type minimum and become even deeper with increasing α, and this excludes the degeneracy of the global minimum.

3.4 Behavior along the horizontal and vertical axes

The structure on the horizontal axis $\varepsilon = 0$ is trivial: The quenched model is nothing but the usual Curie-Weiss Ising model without disorder, for which a phase transition occurs for $\beta^{-1} < 1$ (when the rate-function has a global minimum at $M = \pm M^* \neq 0$ and one local maximum at $M = 0$), and uniqueness for $\beta^{-1} \geq 1$ (when the rate-function has a unique local and global minimum at $M = 0$), see e.g. [3].

When β goes to infinity, i.e. near the vertical axis, the original rate-function Φ converges , for all α, to

$$\Phi_{\infty,\epsilon,\alpha}(m) = \frac{m^2}{2} - \mathbb{E}_\alpha[|m + \epsilon\tilde{\eta}|]. \tag{35}$$

We consider the case $\alpha = 0$, the other cases will be included because L_4 meets L_3 at the vertical axis. The rate-function becomes

$$\Phi_{\infty,\epsilon,\alpha=0} = \frac{m^2 - (|m+\epsilon| + |m-\epsilon|)}{2}.$$

For $\epsilon \geq 1$, this gives $\Phi(m) = m^2/2 - \epsilon$; it is shaped like a parabola and has a unique local extremum, the global minimum $m = 0$. When $\epsilon \in [0, 1]$, it has a piecewise definition and three different analytic expressions : $\Phi(m) = m^2/2 + m$ for $m \le -\epsilon$, $\Phi(m) = m^2/2 - \epsilon$ for $m \in [-\epsilon, +\epsilon]$ or $\Phi(m) = m^2/2 - m$ for $m \ge \epsilon$. In the first case, there is a local minimum for $m = -1$, with $\Phi(-1) = -1/2$, and a local maximum for $m = -\epsilon$, with $\Phi(-\epsilon) = \epsilon^2/2 - \epsilon$. In the second case, there is a local minimum for $m = 0$ with $\Phi(0) = -\epsilon$ and a local maximum for $m = +\epsilon$, with $\Phi(\epsilon) = \epsilon^2/2 - \epsilon$. In the third case, there is local minimum for $m = +1$, with $\Phi(1) = -1/2$. The comparison of the different values taken by Φ at these extrema depending on ϵ gives the picture predicted by the theorem at zero temperature : for $\epsilon \in [0, \frac{1}{2}]$, there are two global minima for $m = \pm 1$, two global maxima at $m = \pm\epsilon$ and one local minimum at $m = 0$, whereas for $\epsilon \in [\frac{1}{2}, 1]$, $m = 0$ is a global minimum, $m = \pm 1$ are local minima and $m = \pm\epsilon$ stays global minima[4].

3.5 The line L_0-L_1: pitchfork-bifurcation

The region below the line L_0-L_1 (two local minima and global maxima) is characterized by the requirement that $\Psi_{\beta,E,\alpha=0}(M)$ has a local maximum at $M = 0$. This gives

$$0 > \partial_M^2 \Psi_{\beta,E,\alpha=0}(M = 0) = \beta^{-1} - 1 + \tanh^2\big(E\big).$$

From this follows that the bounded region below the curve L_0-L_1 is

$$\big\{(\beta, E), \partial_M^2 \Psi_{\beta,E,\alpha=0}(M = 0) < 0\big\} = \big\{(\beta, E), \beta < \cosh^2(E)\big\}.$$

and the curve L_0-L_1 itself is given by $\partial_M^2 \Psi_{\beta,E,\alpha=0}(M = 0) = 0$ which reads, in complete agreement the expressions given previoulsy,

$$\beta^{-1} = 1 - \tanh^2(E) \tag{36}$$

or, equivalently $\beta = \cosh^2(E)$ (Compare [17, 2]). This gives the equation $\epsilon_0 : \beta^{-1} \mapsto \epsilon_0(\beta^{-1})$ with

$$\epsilon_0(\beta^{-1}) = \beta^{-1}\ \mathrm{atanh}\big(\sqrt{1 - \beta^{-1}}\big) \tag{37}$$

and the same expression for ε_1.

Let us start from the high temperature region, where there is a unique minimizer, and consider $\varepsilon \le \varepsilon_*$. Crossing the line L_0 from the outside (increasing β) we encounter a *pitchfork bifurcation*. (See [9, 16], the toy model for a pitchfork bifurcation is given by the potential $M \mapsto AM^2 + M^4$, where the parameter A varies around $A = 0$. For $A > 0$ there is one minimum at $M = 0$, for $A < 0$ there are two symmetric minima with $M^2 = -A/2$.)

More generally, a pitchfork bifurcation occurs at a one-dimensional parameter $p = p_0$ whenever

$$\begin{cases} \partial_M \Phi_{p_0}(M_0) = 0 \\ \partial_M^2 \Phi_{p_0}(M_0) = 0 \\ \partial_M^3 \Phi_{p_0}(M_0) = 0 \\ \partial_p \partial_M \Phi_{p_0}(M_0) = 0 \\ \partial_M^4 \Phi_{p_0}(M_0) \ne 0 \\ \partial_p \partial_M^2 \Phi_{p_0}(M_0) \ne 0 \end{cases} \tag{38}$$

[4]At $\epsilon = 0$, we get the picture corresponding to the area above L_0 because the limit $\beta^{-1} \to 0$ has been taken first.

The first equation follows since M_0 is a local minimizer or maximizer.

The second arises because minimizer and maximizer of quadratic type are merging and hence the second derivative is zero.

The third says follows since we have an extremum at M_0 and not a saddle.

The fourth equality says that the switching on the parameter does not create a tilting between the emerging new extrema, and both have equal depth to lowest order.

The next inequality says that M_0 is definitely an extremum, i.e. it is a maximizer (in the case > 0) or a minimizer (in the case < 0).

The last inequality says that there really is a definite change in sign of the curvature of Φ when the parameter p is perturbed. In the case that $\partial_M^4 \Phi_{p_0}(M_0)$ and $\partial_p \partial_M^2 \Phi_{p_0}(M_0)$ have the opposite signs the creation of 3 extrema from 1 extremum is in the direction of increasing p.

Let us in particular fix E and vary the parameter $p = \beta^{-1}$ along those lines. In the high temperature region, $M = 0$ is the only extremum and thus the unique candidate for this bifurcation. We have $\partial_{\beta^{-1}} \partial_M^2 \Psi_{\beta,E,\alpha}(M) = 1 > 0$, and let us now discuss the behavior of the fourth derivative. The so-called *tricritical point* is given by the requirement that also the fourth derivative is vanishing, $\partial_M^4 \Psi_{\beta,E,\alpha=0}(M = 0) = 0$. This means

$$0 = 1 - 4\tanh^2(E) + 3\tanh^4(E)$$

which has the solution

$$\tanh^2(E_*) = \frac{1}{3}$$

and thus

$$E_* = \operatorname{atanh}\Big(\frac{\sqrt{3}}{3}\Big) \approx 0.6585. \tag{39}$$

The other solution $\tanh^2(E) = 1$ amounts to the vertical slope $E = \infty$. Together with (36), the characteristic equation of the line L_0-L_1, this gives

$$\beta_*^{-1} = \frac{2}{3} \ \text{ and } \ \varepsilon_* = \frac{2}{3}\operatorname{atanh}\Big(\frac{\sqrt{3}}{3}\Big) \approx 0.44 \tag{40}$$

for the temperature and the effective field at the tricritical point and we have

$$\big\{(\beta, E), \partial_M^4 \Psi_{\beta,E,\alpha=0}(M = 0) > 0\big\} = \big\{(\beta, E), E < E_*\big\}. \tag{41}$$

So, the bifurcation at L_0 corresponds to the creation of two minima and a maximum from a minimum (Figure 7 in Section 4). Since $\partial_M^4 \Phi_{p_0}(M_0) > 0$ and $\partial_{\beta^{-1}} \partial_M^2 \Phi_{p_0}(M_0) > 0$ have the same signs along L_0, the creation of two minima and a maximum from a minimum is in the direction of decreasing temperature β^{-1}. On the other hand, since $\partial_M^4 \Phi_{p_0}(M_0) < 0$ and $\partial_{\beta^{-1}} \partial_M^2 \Phi_{p_0}(M_0) > 0$ have the opposite signs along L_1, the creation of two new maxima and a minimum from a maximum is in the direction of increasing temperature β^{-1} (Figure 9 and 10).

Since we know from the above Catastrophe theory discussion about degeneracies that only one degeneracy can occur in the considered region, no bifurcation is possible for the newly created local extremum below L_0-L_1. Moreover, from the same discussion we learn that such phenomenon cannot occur for other $\alpha \neq 0$, for which Ψ still have a unique global minimum below L_0-L_1 (Figure 8 and 13).

3.6 The line L_2: degeneracy of non-zero minima and neutral minimum

When $E > E*$, there could have at most two non-negative local minima and it is indeed the case when we cross L_1. When β^{-1} increases, no other local extremum can merge but another type of degeneracy can occur, depending on the depth of the local minima. The region R_{12} between L_1 and L_2 is now characterized by the presence of two symmetric absolute minima at M and $-M$ and a metastable minimum at 0 (Figure 11), together with two local maxima between 0 and the other minima.

We first prove that we indeed get a line for L_2 by stating the next lemma which says that degeneracies of the depths of the minima of the 'same type' can only occur for one specific value of β^{-1}, which also provides the sharpness of the transition times of Theorem 1.3.

Lemma 3.10 *Suppose that, for fixed (β, E, α), there are two different smooth curves $\beta^{-1} \mapsto M_i(\beta, E, \alpha)$, $i = 1, 2$, β^{-1} running in the same interval $\beta^{-1} \in I$, of local extreme points, where $M_1^2(\beta, \alpha) < M_2^2(\beta, \alpha)$. Then the function $\beta^{-1} \mapsto \Psi_{\beta,E,\alpha}(M_2(\beta, E, \alpha)) - \Psi_{\beta,E,\alpha}(M_1(\beta, E, \alpha))$ is decreasing in β^{-1}. In particular, for fixed $E > E_*$, there is at most one value of β^{-1} for which $\Psi_{\beta,E,\alpha}(M_1(\beta, E, \alpha)) = \Psi_{\beta,E,\alpha}(M_2(\beta, E, \alpha))$*

Proof: The proof is completely analogous to Lemma 3.9 on the variation w.r.t. α. Looking at the evolution of value of the rate-function at both curves of extreme points with the parameter β^{-1} we have

$$\frac{d}{d\beta^{-1}}\Psi_{\beta,E,\alpha}(M_i(\beta, E, \alpha)) = \frac{M_i^2(\beta, E, \alpha)}{2}$$

This implies the claim. $\square$

We will apply this to the present case of $\alpha = 0$ when we fix E and move along those lines by varying β^{-1}, crossing the different lines in the phase diagram. Coming from the region below L_0-L_1 we first see a pitchfork-bifurcation at the line L_1 that produces a minimum at $M = 0$. This minimum value does not depend on the temperature but gets deeper and deeper w.r.t. to the depth of the non-zero minimum (whose depth is decreasing with temperature) and when we reach the line L_2 there is in particular degeneracy (The depth of the non-zero minimum become equal to the depth at zero). When we further increase the temperature β^{-1} the non-zero minimum will finally vanish at the line L_3 (Figure 12). The mechanism of this is a *fold bifurcation*, to be discussed below. In particular we see from the lemma that the critical values of $\beta^{-1}(E)$ where these two phenomena are happening are unique. Moreover, the lowest depth of the non-zero minimum being reached for $\alpha = 0$, this former is the lowest and first value (when β^{-1} increases) where such an equality is possible. Above L_2, we shall see that this occur for bigger and bigger values of α until a critical value characterizing L_4.

So we get the following two equations valid at L_2 for (M, E, β). We always have for $\alpha = 0$ the mean field equation (16) at M and at the line L_2 we have the equation demanding equality of the two values of the potential at 0 and $M > 0$,

$$\Psi_{\beta,E,\alpha=0}(M = 0) = \Psi_{\beta,E,\alpha=0}(M).$$

This is reexpressed as

$$\begin{cases} \frac{M}{\beta} = \overline{\tanh}(M, E) \\ \\ \frac{M^2}{2\beta} - \overline{\log(\cosh)}(M, E) = -\log(\cosh)(E). \end{cases}$$

Dividing the two equations eliminates the linear variable β^{-1} and implies the following implicit relation of (M, E) characterizing the solutions $F_2(M, E) = 0$ with

$$F_2(M, E) = M\ \overline{\tanh}(M, E) - 2\ \overline{\log\cosh}(M, E) + 2\log\cosh(E). \tag{42}$$

Note that $M = 0$ is always a solution, for any $E \in \mathbb{R}_+$. Looking however for $E \in (E_*, \infty)$, we can numerically find a non-zero solution $M_2 = M_2(E)$. We also emphasize that there can only be one value of $M \neq 0$ for fixed $E > E_*$. This is clear, since we have proved before that the rate-function can have at most 3 local minima, for any choice of the parameters.

Having solved this (M, E)-relation, the curve L_2 is obtained immediately by solving the mean-field equation for β^{-1}. This gives us the parametric representation

$$E \mapsto \beta_2^{-1}(E) = \frac{\overline{\tanh}(M_2(E), E)}{M_2(E)}. \tag{43}$$

We obtain from this a parametric plot in the usual $(\beta^{-1}, \varepsilon)$-plane by putting $\varepsilon_2(E) := \beta_2^{-1}(E) \cdot E$.

Let us show that the line has a representation as a function $\varepsilon_2 = \varepsilon_2(\beta^{-1})$. To show this we show that $\frac{d}{dE}\beta^{-1}(E) < 0$. To see this we take the derivative of the second equation w.r.t. E and note that the terms involving the M-derivative vanish, due to the mean-field equation. We get

$$\frac{d}{dE}\beta^{-1}(E)\frac{M(E)^2}{2} = \Delta\tanh(M(E), E) - \tanh(E).$$

The proof is now finished by showing that the r.h.s. is strictly less than zero, for all $M > 0$ and $E > 0$. Note that we can not use convexity in order to conclude this since $M - E$ might be negative. Instead introduce the variables $y = e^{2M}$, $b = e^{2E}$ and write

$$\begin{aligned} &\Delta\tanh(M(E), E) - \tanh(E) = \frac{1}{2}\left(\frac{yb-1}{yb+1} - \frac{yb^{-1}-1}{yb^{-1}+1}\right) - \frac{b-1}{b+1} \\ &= -\frac{(b-1)(y-1)^2}{(yb+1)(yb^{-1}+1)(b+1)} < 0 \end{aligned}$$

and then we really have a function $\varepsilon_2(\beta^{-1})$.

Unfortunately, it does not seem to be possible to get an explicit expression for this function. We shall nevertheless manage to give explicitly its local behavior around the tricritical point, while we have used numerical solutions $M_2(E)$ of the mean-field type equation $F_2(M, E) = 0$ to draw the line L_2 on the extended phase diagram (Figure 2 in Section 3.1).

3.7 The line L_3: Fold bifurcation

The region between the lines L_2 and L_3 is characterized by a global minimum at $M = 0$ and two symmetric local minima at nonzero values $\pm M^*$, $M^* > 0$ (Figure 15) for $\alpha = 0$, non-symmetric for $\alpha \neq 0$ (Figure 16). As already said, we do not really need to characterize L_3 for the purpose of non-Gibbsianness, but determining the whole non-Gibbs region becomes easier once we have focused first on $\alpha = 0$. In this case, varying (decreasing) β^{-1} along the lines of fixed $E > E_*$ a *fold bifurcation* occurs at the line L_3 : a minimum and a maximum are created from a saddle point by varying the parameter p into one direction. It occurs at a one-dimensional parameter $p = p_0$ whenever

$$\begin{cases} \partial_M \Phi_{p_0}(M_0) = 0 \\ \partial_M^2 \Phi_{p_0}(M_0) = 0 \\ \partial_p \partial_M \Phi_{p_0}(M_0) \neq 0 \\ \partial_M^3 \Phi_{p_0}(M_0) \neq 0. \end{cases} \tag{44}$$

The first equation just says that M_0 is a local minimizer. The second arises because minimizer and maximizer of quadratic type are merging and creating a saddle point with $p \to p_0$ and hence the second derivative is zero. The two extrema occur for $p > p_0$ if $\partial_p \partial_M \Phi_{p_0}(M_0)$ and $\partial_M^3 \Phi_{p_0}(M_0)$ have opposite sign and are not present for $p < p_0$.

This does not happen on L_2 for $\alpha \neq 0$, where there also are two non-negative local minima but the depths (which are monotone in α) are not equal. We can however check that crossing L_2, this degeneracy does not hold anymore for $\alpha = 0$ (Figure 13), leaving the door open for other candidates α for equal depth, that indeed exists and increases with the temperature along the line of constant E, from $\alpha = 0$ to α_c (see next subsection the characterization of L_4). Let us come back to $\alpha = 0$ to describe L_3. We have

$$\partial_{\beta^{-1}} \partial_M \Psi_{\beta,E,\alpha}(M) = M \neq 0. \tag{45}$$

The line L_3 is characterized by the mean field equation at M and furthermore

$$\partial_M^2 \Psi_{\beta,E,\alpha=0}(M) = 0.$$

These two equations are reexpressed as

$$\begin{cases} \frac{M}{\beta} = \overline{\tanh}(M, E) \\ \\ \frac{1}{\beta} = 1 - \overline{\tanh^2}(M, E). \end{cases} \tag{46}$$

Again, dividing the two equations eliminates the linear variable β^{-1} and implies an implicit (M, E)-relation of the form $F_3(M, E) = 0$ with

$$F_3(M, E) = M\left(\overline{1 - \tanh^2}(M, E)\right) - \overline{\tanh}(M, E). \tag{47}$$

We follow the same logic as for the line L_2. Note that $M = 0$ is always a solution, for any E. For $E \in (E_*, \infty)$ there is a unique solution $M = M_3(E)$ and we can find it numerically. As before, this gives us the parametric representation

$$E \mapsto \beta_3^{-1}(E) = \frac{\overline{\tanh}(M_3(E), E)}{M_3(E)} \tag{48}$$

or equivalently

$$\beta_3^{-1}(E) = 1 - \overline{\tanh^2}(M_3(E), E)$$

and $\varepsilon_3(E) := \beta_3^{-1}(E).E$.

Let us now vary the parameter α around $\alpha = 0$ at the line L_3. We have

$$\partial_\alpha \partial_M \Psi_{\beta,E,\alpha}(M) = -\Delta \tanh\big(M, E\big) < 0$$

if and only if $M > 0$. But this means that we have a fold bifurcation w.r.t. α, too. The creation of a minimum/maximum-pair is in the direction of increasing α, since the third derivative w.r.t. M of the rate-function has a positive sign. This will allow the extension of the non-Gibbs region to a wider range of temperature for fixed E, namely in the "NG*" region R_{34} between L_3 and L_4.

Let us show that the line has a representation as a function $\varepsilon = \varepsilon_3(\beta^{-1})$ i.e. that $\frac{d}{dE}\beta^{-1}(E) < 0$. To see this we take the derivative of the mean field equation w.r.t. E

$$0 = \beta^{-1}(E)M(E) - \overline{\tanh}(M(E), E).$$

This gives

$$0 = \Big(\frac{d}{dE}\beta^{-1}(E)\Big)M(E) + \frac{1}{2}\big(\tanh^2(M(E)+E) - \tanh^2(M(E)-E)\big)$$

where we have used that the second derivative of the rate-function w.r.t. M vanishes. The difference on the right is always positive, for positive $M(E)$, and this implies the claim that we do have a function. It seems again difficult to find an explicit expression for this function, but we shall also study some local behaviors explicitly.

3.8 The line L_4: pitchfork bifurcation for non-zero α

The line L_4 has already been characterized during the general Catastrophe theory discussion and the explicit expression we got enables us to draw it on the phase diagram. Nevertheless, it could also be derived in the same way as the previous derivation for the lines L_3 and L_3, and we indicate here how one should proceed.

For $E > E_*$, β_4^{-1} corresponds the higher value of β for which the creation of two minima of equal depth by increasing α is possible, and can thus be characterized by the existence of a solution $(\alpha, M, \beta) \in (0,1) \times (\mathbb{R}_*^+)^2$ of the system of equations

$$\begin{cases} 0 = \partial_M \Psi_{\beta,E,\alpha}(M) \\ 0 = \partial_M^2 \Psi_{\beta,E,\alpha}(M) \\ 0 = \partial_M^3 \Psi_{\beta,E,\alpha}(M). \end{cases} \tag{49}$$

To understand the system (49), note the following: the open region R_{34} in the $(\beta^{-1}, \varepsilon)$-plane between the lines L_3 and L_4 is defined by the property that there exists an $\alpha = \alpha_c(\beta^{-1}, E)$ strictly between 0 and 1 such that there are precisely two local minimizers $0 < M_1(\beta^{-1}, E) < M_2(\beta^{-1}, E) < 1$ and for those the value of the rate-function coincides (Figure 16). Let us denote the local maximizer between $M_1(\beta^{-1}, E)$

and $M_2(\beta^{-1},E)$ by $\hat{M}(\beta^{-1},E)$. When we approach the line L_4 from within the region R_{34} we encounter a pitchfork bifurcation again (i.e. a collapse of these extrema). It means these two minimizers of $\Psi_{\beta,E,\alpha=\alpha_c(\beta,E)}(M)$ come closer and collapse with each other and the maximizer at the line L_4, that is $M_1(\beta^{-1},\varepsilon)=M_2(\beta^{-1},\varepsilon)=\hat{M}(\beta^{-1},\varepsilon)$ for $(\beta^{-1},\varepsilon)\in L_4$. This however implies that the second derivative of the rate-function w.r.t. M at $\hat{M}(\beta^{-1},\varepsilon)$ vanishes on the line L_4 (this follows since the second derivative of the rate-function w.r.t. M is non-negative at $M_1(\beta^{-1},\varepsilon)$ and non-positive at $\hat{M}(\beta^{-1},\varepsilon)$ within the region R_{34}. So, by continuity, it must vanish on L_4). Since $\hat{M}(\beta^{-1},\varepsilon)$ is a local minimizer on the 'collapse line' L_4 this implies that also the third derivative of the rate-function w.r.t. M vanishes on L_4. Beyond L_4, no creation of saddle is possible whatever α is (in $[-1,1]$) and thus no bifurcation occurs.

Varying β^{-1} along the curve of fixed E, it is possible to eliminate α from the system (49) and get an implicit characterization of the non-zero minima on the line L_4. One gets afterwards a parametrization of L_4 in term of the dynamical field $h_t=E$, and it is also possible to check that L_4 has indeed a representation as a function $\varepsilon_4=\varepsilon_4(\beta^{-1})$.

Nevertheless, the general discussion of the previous section provides a more explicit expression of this line, that we have used to draw numerically the phase diagram, because it does not involve any numerical computation.

3.9 Proof of Theorems 1.3 and 1.4

Let us summarize the picture. We are dealing with a quenched system depending on the parameters β^{-1} (initial temperature), ε (time, $\equiv$ coupling strength of the random fields) and α (conditioning of empirical average of spins at time t, $\equiv$ tilting of the random field distribution). Equivalently we are looking at the parameter space spanned by (β^{-1},E,α).

To get the time evolution picture (Figure 1) from the extended phase diagram (Figure 2), and thus Theorems 1.3 and 1.4 from Theorems (3.1) and (3.2), we use the smooth relation (3) between time t and $h_t=E_t$. The regions of the extended phase diagram with qualitatively different phase structure are mapped into the regions of Gibbsianness, of non-Gibbsianness or of non-Gibbsianness with symmetry breaking depending whether the singularities induce different left and right limit of the conditional probabilities or not. Theorem 1.3 is then easily obtained using the smooth one-to-one decreasing relation between t and E.

To prove Theorem 1.4 one also uses this relation together with the logarithmic expression of the inverse hyperbolic tangent to get explicitly $t_0(\beta^{-1})=-\frac{1}{4}\log(1-\beta^{-1})$ for $\beta^{-1}\geq\frac{2}{3}$, and the same expression for $t_1(\beta^{-1})$ when $\beta^{-1}<\frac{2}{3}$. For $0<\beta^{-1}<\frac{2}{3}$, we use $t_4(M)=\mathrm{atanh}(\mathrm{E_4(M)})$ to parametrize (t_0,β_4^{-1}) by $M>0$ via $y_4(M)$ through the expressions given by Theorem(3.2). We may also write $t=-\frac{1}{2}\log\tanh(h_t)=-\frac{1}{2}\log\tanh(\beta\varepsilon_t)\equiv-\frac{1}{2}\log\tanh(E_t)$ for the function giving time in terms of the dynamical field. Thus $E=\infty$ corresponds to $t=0$.

Recall that the rate-function of the system can have $3,2$ or 1 minima, that is the system can have $3,2$ or 1 different phases which can be characterized as $+$-like, $-$-like, $\pm$-like, or "high-temperature" phase in case of uniqueness. In the later case, Theorem 2.1 proves Gibbsianness at high temperature $\beta^{-1}\geq 1$ where Theorem 3.1 tells that there is uniqueness. It tells that it is also the case when $\beta^{-1}\leq 1$ in the region R_{40}, that is for short times. In the other regions, there is non uniqueness, and this leads to non-Gibbsianness in the mean-field sense because approaching a prescribed magnetization α from below or above does not yield the same conditional probabilities, no limit can be defined and thus some conditional probabilities are discontinuous. A discontinuity in α, for fixed (β^{-1}, E) corresponds thus to a point of coexistence of two phases; depending of the type of metastability detected via Theorem 3.1, we could have as discontinuity $\alpha = 0$ (regions R_{12}) or $\alpha \neq 0$ (region R_{23} and R_{34}).

For $\alpha = 0$ there is a $\pm$-symmetry of the rate-function in M. This implies that the $+$- and $-$- phase are automatically degenerate and this degeneracy is lifted by varying α around zero. In the region R_{01}, they are the only stable phases and $\alpha = 0$ is a point of discontinuity because the right and left limits when α approach 0 are different : when it is approached from below, the degeneracy of the $-$-phase is lifted and the $+$-phase is eventually selected, whereas the contrary happens from above. In the region R_{12}, the stable phases are still the $+$- and $-$-phase, but the $\pm$-phase arises as a metastable phase. Due to its unstable character, this phase does not modify the discontinuity for $\alpha = 0$ described above, and does not lead to any other point of discontinuity. In R_{23}, the role are exchanged for $\alpha = 0$: the $\pm$-phase is stable and the $+$- and $-$ phases are metastable. We may then achieve $(+, \pm)$-degeneracy by turning on an $\alpha = \alpha_c(\beta, E) > 0$ which stabilizes $\pm$, gives it a minimum of equal depth with the $+$ one and creates a discontinuity. The converse happens at $-\alpha_c(\beta, E) < 0$ and we have discontinuity on our mean-field sense for $\alpha_c \neq 0$ and $-\alpha_c$. In R_{34}, there is no metastability at $\alpha = 0$ anymore and the neutral phase is the only one. Nevertheless, tuning on $\alpha > 0$ gives rise to a metastable $+$-phase which becomes stable at $\alpha = \alpha_c(\beta, E) \neq 0$, for which there again discontinuity. Moreover, for certain values of α there can be also degeneracies of two metastable minima that are energetically above the absolute minimum. This degeneracy will not give rise to a discontinuity, and is not interesting for our purposes. There is no metastability beyond L_4, and thus the "high temperature" region R_{40} is the region of Gibbsianness of the time-evolved measures. This occurs for short times at low temperature and at every time for high temperature $\beta \geq 1$.

Finally let us prove monotonicity of the critical α in the NG*-region as a function of time.

Proposition 3.11 *For fixed $\beta^{-1} > 0$, the bad configuration $E \mapsto \alpha_c(\beta^{-1}, E)$ is an increasing function in E (and hence a decreasing function of the time).*

Proof: Look at the system

$$\begin{cases} \Psi_{\beta,E,\alpha}(M_1) = \Psi_{\beta,E,\alpha}(M_2) \\ \partial_M \Psi_{\beta,E,\alpha}(M_1) = 0 \\ \partial_M \Psi_{\beta,E,\alpha}(M_2) = 0. \end{cases}$$

We show that $\frac{d\alpha}{dE}\Big|_{\beta^{-1}} \geq 0$. Taking the derivative of the first equation w.r.t. E and using the second and the third equation we get

$$\begin{aligned} &- \mathbb{E}_\alpha \tanh\big(M_1 + E\tilde{\eta} + H\big)\tilde{\eta} - \Delta \log\cosh\big(M_1 + H, E\big)\frac{d\alpha}{dE} \\ &= -\mathbb{E}_\alpha \tanh\big(M_2 + E\tilde{\eta} + H\big)\tilde{\eta} - \Delta \log\cosh\big(M_2 + H, E\big)\frac{d\alpha}{dE} \end{aligned}$$

Note that $M \mapsto \Delta \log\cosh\big(M,E\big)$ is an increasing function. So the claim follows if we show that $M \mapsto \mathbb{E}_\alpha \tanh\big(M + E\tilde\eta\big)\tilde\eta$ is a decreasing function.

But note that

$$\begin{aligned}&\partial_M \mathbb{E}_\alpha \tanh\big(M+E\tilde\eta\big)\tilde\eta\\ &= -\frac{1+\alpha}{2}\tanh^2\big(M+E\big) + \frac{1-\alpha}{2}\tanh^2\big(M-E\big) \le 0\end{aligned}$$

for $M, E \ge 0$ and $\alpha \ge 0$.

□

Acknowledgments : We thank the Laboratoire de mathématiques d'Orsay (CK) and Eurandom (ALN, CK) for invitations, crucial in the achievement of this work. We thank A. van Enter, R. Fernández, E. Orlandi, M. Vares and V. Zagrebnov for encouraging discussions, H. Broer (Groningen) for advice on Catastrophe theory, and we are also grateful to L.M. Le Ny (Irisa, Rennes) for the help in scilab and matlab to draw the pictures.

References

[1] Amaro de Matos, J.M.G., Patrick and A.E., Zagrebnov, V. (1992) Random infinite-volume Gibbs states for the Curie-Weiss random field Ising model, *J. Statist. Phys.* **66**, 139–164.

[2] Cassandro, M., Orlandi, E., Picco, P. and Vares, M.E. (2005) One-dimensional random field Kac's model: localization of the phases, *Elec. J. Probab.*, **10**, 786–864, 2005

[3] Ellis, R.S. (1985) *Entropy, Large Deviations and Statistical Mechanics*, Springer-Verlag, New York etc.

[4] van Enter, A.C.D., Fernández, R., den Hollander, F. and Redig, F. (2002) Possible loss and recovery of Gibbsianness during the stochastic evolution of Gibbs measures, *Comm. Math. Phys.* **226**, 101–130.

[5] van Enter, A.C.D., Fernández, R. and Sokal, A.D. (1993) Regularity properties of position-space renormalization group transformations: Scope and limitations of Gibbsian theory, *J. Statist. Phys.* **72**, 879–1167.

[6] van Enter, A.C.D., Lörinczi, J. (1996), Robustness of the non-Gibbsian property: some examples. *J. Phys. A* **29**,t 2465–2473.

[7] Häggström, O. and Külske, C. (2004) Gibbs property of the fuzzy Potts model on trees and in mean-field. In *Proceedings of the workshop "Gibbs vs. non-Gibbs in statistical mechanics and related fields" (Eurandom, December 2003)*, Markov proc. Relat. Fields **10**, no 2, 2004.

[8] Georgii, H.-O. (1988) *Gibbs Measures and Phase Transitions*, de Gruyter, New York.

[9] Guckenheimer, J. and Holmes, P., (1990) Nonlinear oscillations, dynamical systems, and bifurcations of vector fields, *Applied Mathematical Sciences*, **42**, Springer, New York.

[10] Külske, C. (2003) Analogues of non-Gibbsianness in joint measures of disordered mean field models, *J. Statist. Phys.* **112**, 1101–1130.

[11] Külske, C. and Le Ny, A. (2005) Gibbs properties of mean-field models under time evolution II: Case of general external fields $h \neq 0$. In preparation.

[12] Külske, C. and Le Ny, A. (2005) 'DLR-like' description of mean-field models. In preparation.

[13] Külske, C. and Redig, F. (2005), Loss without recovery of Gibbsianness during diffusion of continuous spins, to be published in Probab. Theor. Relat. Fields. DOI: 10.1007/s00440-005-0469-9

[14] Le Ny, A. and Redig, F. (2002), Short times conservation of Gibbsianness under local stochastic evolutions, *J. Statist. Phys.* **109** nos 5/6, 1073–1090.

[15] Liggett, T.M. (1985) *Interacting Particle Systems*. Springer-Verlag, New York, 1985.

[16] Poston, T. and Stewart, I. (1978), Catastrophe Theory and its Applications, Surveys and reference works in mathematics, Pitman, London, San Francisco, Melbourne.

[17] Salinas, S.R. and Wreszinski, W.F. (1985), On the mean-field Ising model in a random external field, *J. Statist. Phys.*, **41**, nos 1/2, 299–313.

4 Pictures

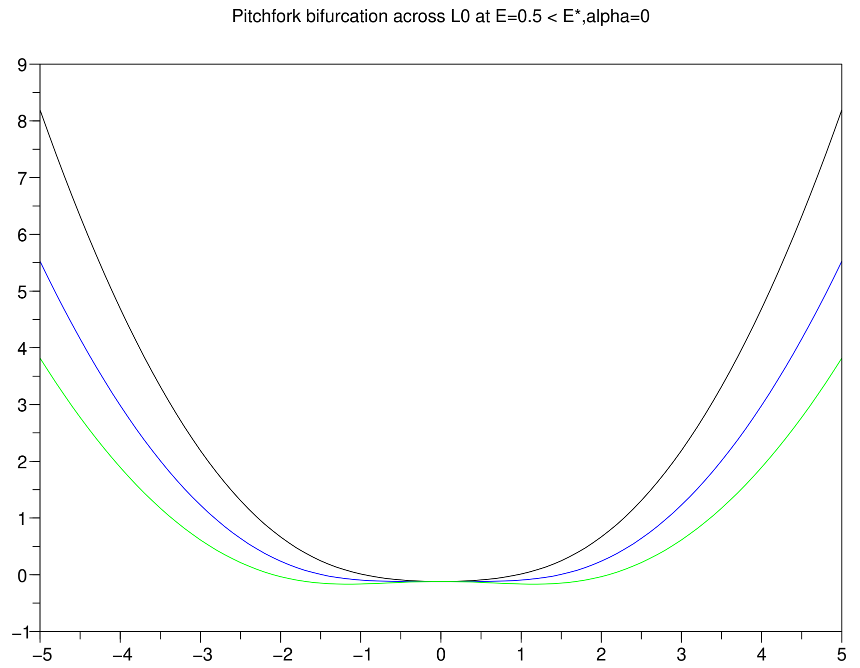

Figure 6: **Across L_0 at $\alpha = 0$**

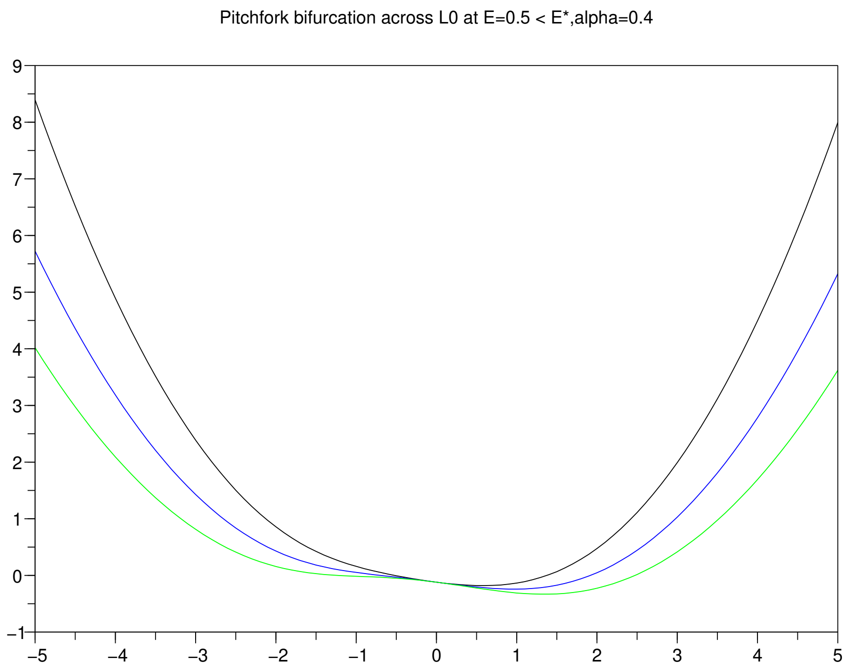

Figure 7: **Across** L_0 **at** $\alpha \neq 0$

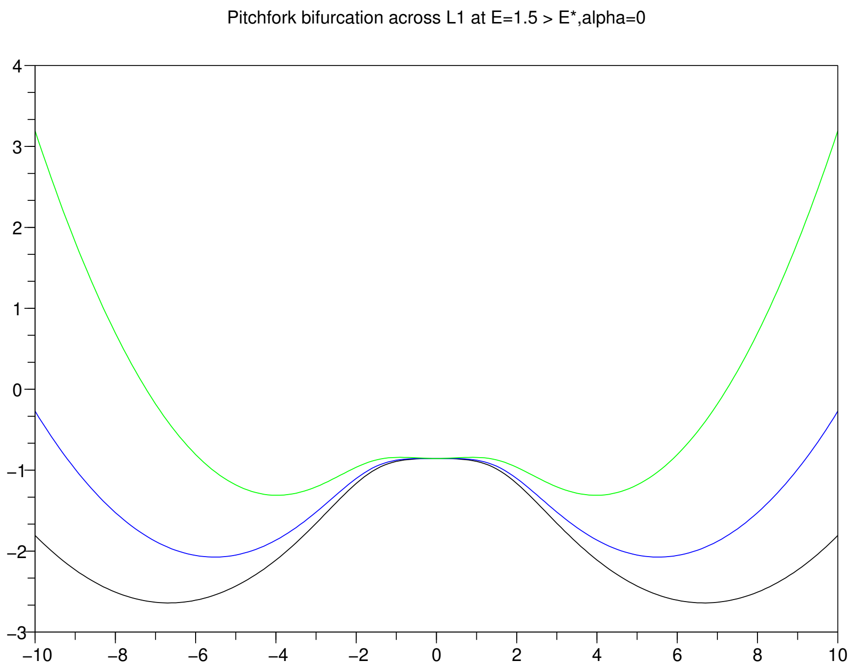

Figure 8: **Across** L_1 **at** $\alpha = 0$

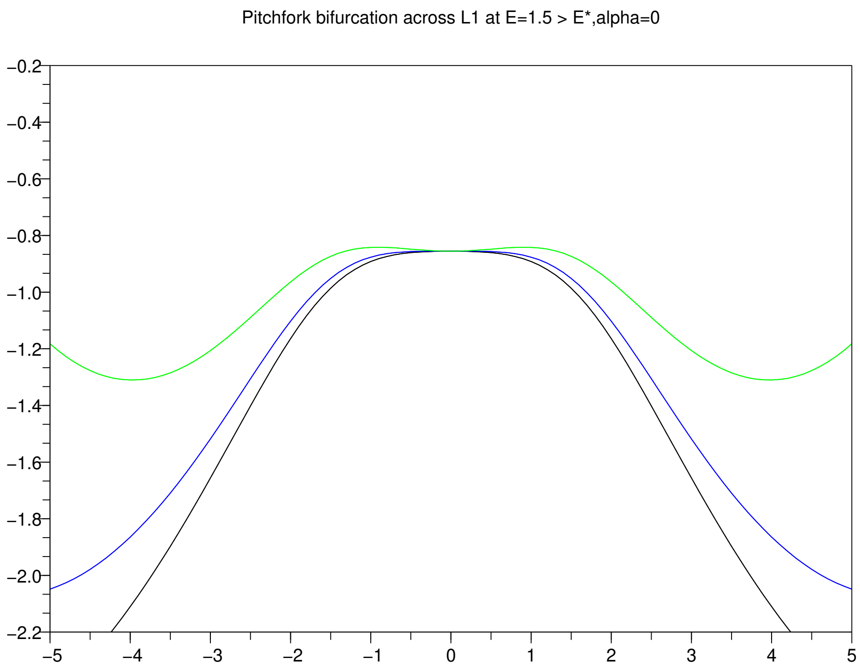

Figure 9: **Across** L_1 **at** $\alpha = 0$ **Zoomed**

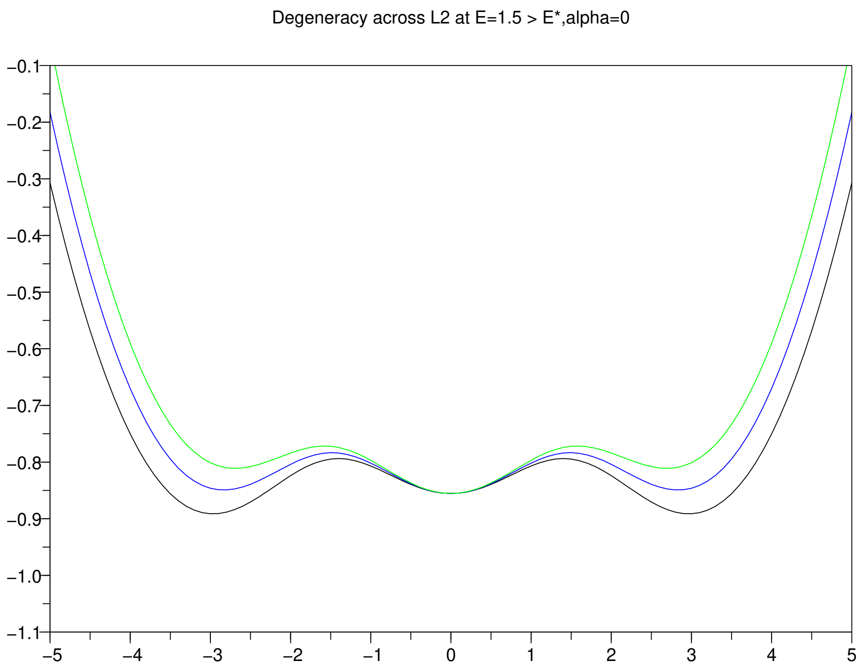

Figure 10: **Across** L_2 **at** $\alpha = 0$

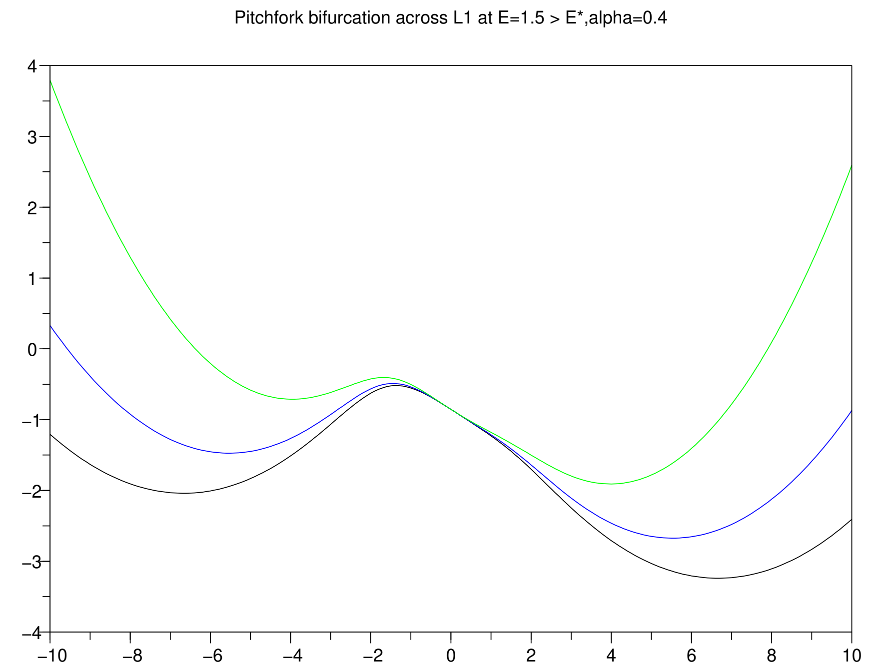

Figure 11: **Across** L_1 **at** $\alpha \neq 0$

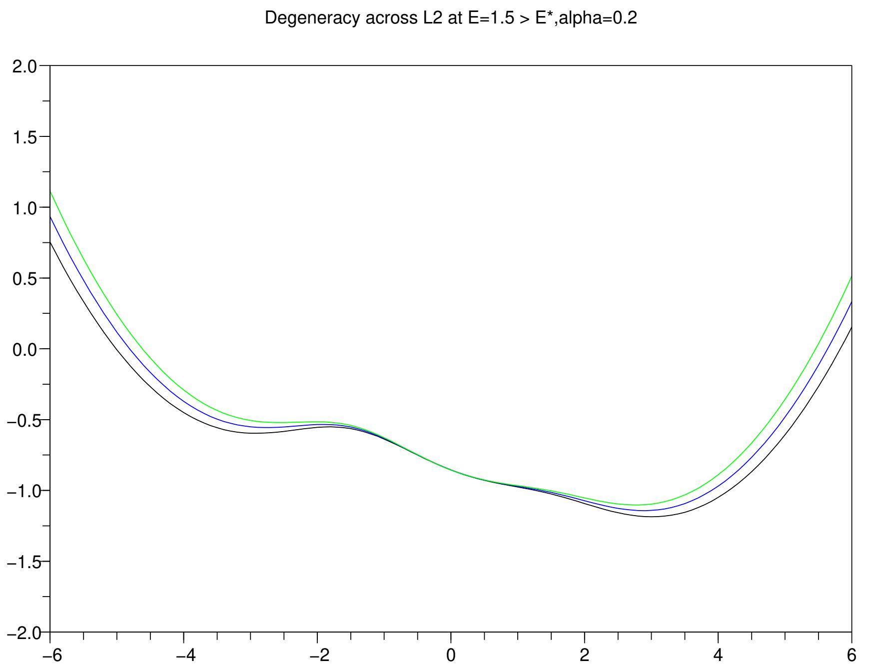

Figure 12: **Across** L_2 **at** $\alpha \neq 0$

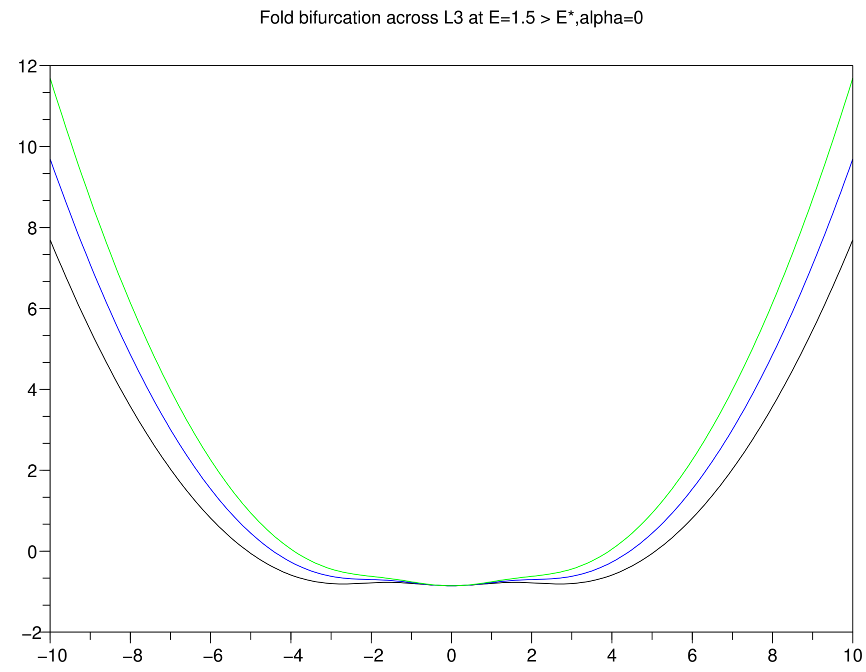

Figure 13: **Across L_3 at $\alpha = 0$**

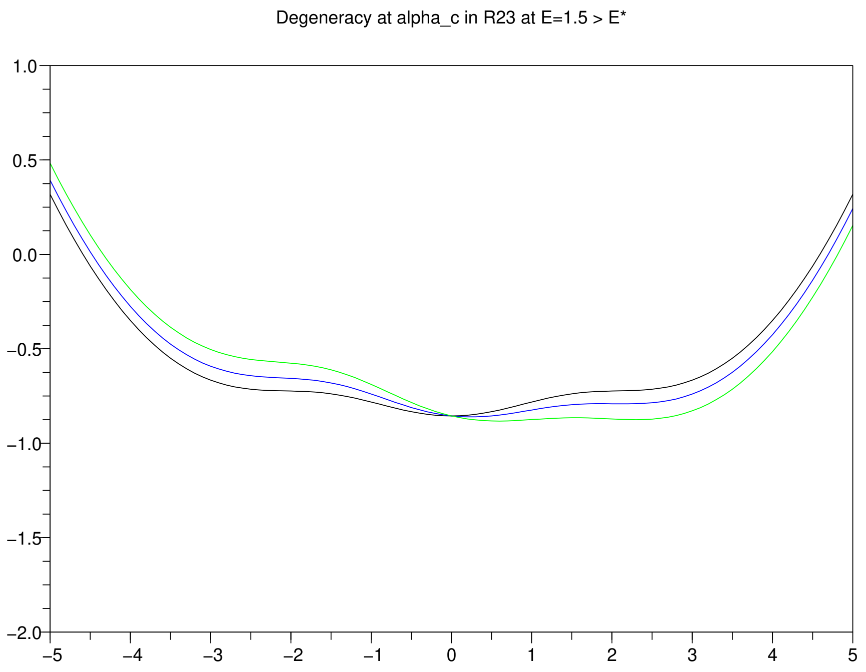

Figure 14: **Degeneracy at $\alpha \neq 0$ in R_{23}**

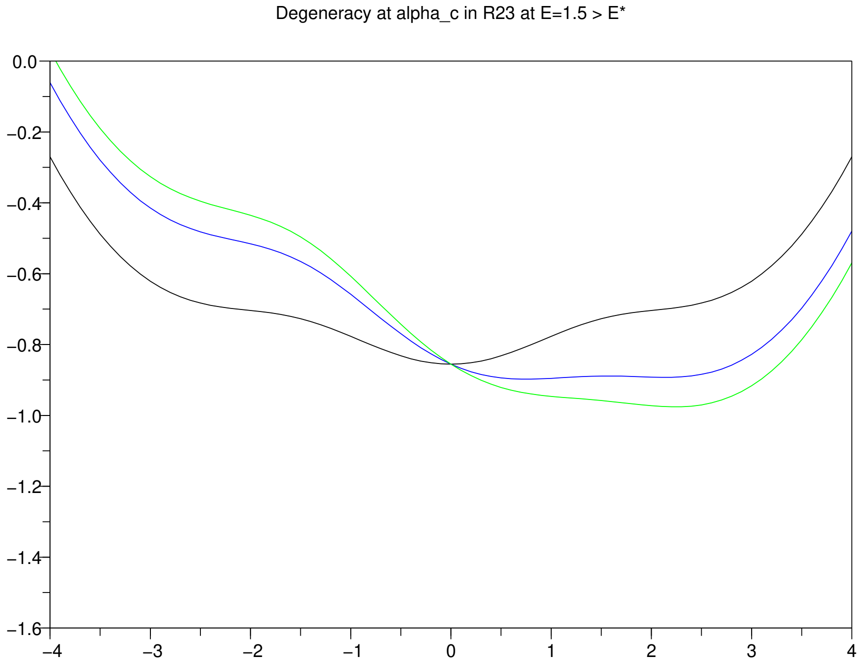

Figure 15: **Degeneracy at $\alpha \neq 0$ in R_{34}**