How non-Gibbsianness helps a metastable Morita minimizer to provide a stable free energy

Christof Külske[†]

April 15, 2004

Abstract: We analyze a simple approximation scheme based on the Morita-approach for the example of the mean field random field Ising model where it is claimed to be exact in some of the physics literature. We show that the approximation scheme is flawed, but it provides a set of equations whose metastable solutions surprisingly yield the correct solution of the model. We explain how the same equations appear in a different way as rigorous consistency equations. We clarify the relation between the validity of their solutions and the almost surely discontinuous behavior of the single-site conditional probabilities.

AMS 2000 Subject Classification. 60G60, 82B20, 82B44.
Keywords and phrases. Disordered systems, Morita approach, non-Gibbsian measures, mean field models, random field model.

[†] EURANDOM, P.O.Box 513 5600 MB Eindhoven, The Netherlands and
Institut für Mathematik, Sekr. MA 7-4, TU Berlin, Strasse des 17. Juni 136, D-10623 Berlin, Germany

1 Introduction

The *Morita-approach* or *equilibrium ensemble approach* to systems with quenched disorder goes back to [18]. A fair and clear review from a theoretical physicist's point of view containing a quick outline of the theory and various interesting recent applications is given by Kühn in [14] (see also [13]).

The central idea in the Morita-approach is: Look at the *joint measure* governing the distribution of the quenched degrees of freedom and the dynamical variables, rather than directly trying to describe the *quenched measure* for the dynamical variables for fixed realization of the disorder. Ideally one would then like to write this joint measure as a formal equilibrium average over the *joint variables* in terms of a *joint Hamiltonian*. This joint Hamiltonian would then be the sum of the original one and another "disorder-Hamiltonian" depending only on the quenched degrees of freedom. The resulting model possesses full spatial symmetries and one might hope that it is amenable to techniques known from systems without disorder.

Mathematically there are problems with this idea. In finite volume this "disorder-Hamiltonian" can in principle be chosen in such a way that the resulting joint distribution coincides with that of the true model. For lattice systems in the infinite volume this is however a serious problem. In fact, for many models an absolutely summable joint Hamiltonian does not exist, and the joint measures in the infinite volume are non-Gibbsian measures. The appearance of a non-Gibbsian joint distribution was first discovered in the example of the Grising model in [5] and studied in a general context in [9, 10]. See also the discussion in [6, 15, 16]. A well-understood example for this are in particular the joint measures of the random field Ising model [3] in more than 3 dimensions at low temperature and small disorder. They provide an illuminating example of strong non-Gibbsian pathologies. In fact, their conditional probabilities are shown to be discontinuous functions of the conditionings, for a set of conditionings with (joint) measure one [10]. This means that the measure is not even "almost surely Gibbs" (in the sense of [17]). This pathology even causes the usual Gibbs variational principle to fail [12]. Close analogies to this behavior on the lattice can be already found in the corresponding mean-field model. Here the corresponding functions describing the conditional expectations can be explicitly computed [11]. For more on the analogies between non-Gibbsian measures on the lattice and discontinuous behavior of conditional probabilities in mean-field models see [7] and [11].

The motivation for using the Morita-approach from the point of view of theoretical physics is however to leave these conceptual problems aside and take it as a source for approximation schemes [14, 15]. Such schemes can be obtained by taking certain simplified trial disorder-Hamiltonians that are chosen e.g. demanding that a finite number of moments of the distribution of the disorder variables coincide with that of the true distribution. Then one would like to solve the resulting Morita approximant model and hope that relevant features of the solution are the same as that of the true model.

It might seem hopeless to justify such approximations in general for non-trivial lattice models. It is therefore valuable to fully understand at least simple toy models that can be explicitly treated. This is want we want to do here. We will give here a complete discussion of the quick naive "solution" of the mean-field random field Ising model, based on a very simple approximant joint measure containing only one parameter [23]. This so-called solution is fairly old and the computations are trivial, but the justification of the resulting equations is subtle. So it is worth to reconsider it from a rigorous point of view and straighten out some wrong claims in the literature, answering a question of Kühn. In particular we take issue with the statement of Kühn [14], describing the work of [23]. He writes: "it reproduces the exact solution at the cost of introducing a single 'chemical potential' to fix the average value of the random field, which creates a term in the modified Hamiltonian that introduces no non-locality into the model over and above that already contained in the definition of the Curie-Weiss limit. This is a remarkable result in the light of concerns raised about the appearance of non-Gibbsian measures within the equilibrium ensemble approach and the identification of the RFIM as providing a realization of a kind of 'worse-case scenario' in the non-Gibbsian world..."

We will indeed see that even for this simple model the situation is subtle and the validity of the solution is fundamentally related to the analogue of "non-Gibbsianness" in the mean-field context. To see this, we will start in Section 2 by reviewing the quick "solution", following [14]. This provides us with two equations for two parameters, the magnetisation and the Morita chemical potential. These equation have in fact solutions for which the magnetisation-variable takes the known value of the spontaneous magnetization. However, we note that this solution corresponds to a wrong (metastable) saddle-point approximation for the approximant measure and therefore the naive derivation given above is flawed. Moreover, we will prove that in the low temperature regime it is even strictly impossible to choose a chemical potential such that the magnetic fields become symmetric. In brief, the theory based on the Morita approximant measure with just one chemical potential fails.

How can we understand then that the two equations derived by a wrong line of argument yield the correct value of the magnetization? Is this just accidental? We will see in Section 3 that the same two equations come up in a different way as consistency equations for the conditional probabilities of the true joint measures of the model without approximations. Here however the fixed Morita-chemical potential is replaced by a random variable. It is in this context that we will finally understand that the validity of these equations and the almost sure discontinuity of the conditional expectations are consequences of each other.

2 Invalidity of single-site Morita approximation approach for the Curie Weiss Random Field Ising Model

We consider the mean-field random field Ising model. It is defined in terms of the following formula for the quenched Gibbs expectation for fixed choice of the random fields.

Quenched measure:

$$\mu_{\beta,\varepsilon,h_0,N}[\eta_{[1,N]}](\sigma_{[1,N]}) := \frac{\exp\Big(\frac{\beta}{2N}\big(\sum_{i=1}^N \sigma_i\big)^2 + \beta\sum_{i=1}^N(\varepsilon\eta_i + h_0)\sigma_i\Big)}{Z_{\beta,\varepsilon,h_0,N}[\eta_{[1,N]}]} \tag{1}$$

Here the spins ("dynamical variables") take values $\sigma_i = \pm 1$ and the random fields take values $\eta_i = \pm 1$ with equal probability. We denote their distribution by $\mathbb{P}$. We stress that the partition function appearing in the denominator depends on the realization of the random fields $\eta_{[1,N]}$ describing the disorder. We allow from the beginning also an external magnetic field h_0, but we are mainly interested in the case $h_0 \downarrow 0$.

What one understands by the "solution of the model" is the characterization of the behavior of this measure on the σ's for a large set of η's, having asymptotically $\mathbb{P}$-measure one. This has been done in great detail [22, 1, 8], and so in this model there is no need for any approximation based on the Morita-approach in order to solve the model. Most basically, we know the phase structure in zero external field, for any choice of the parameters β, ε. We recall that for large β and small ε in zero external magnetic field h_0 the model exhibits a spontaneous magnetization whose value m is a solution of the equation

$$m = \frac{1}{2}\Big(\tanh\beta(m+\varepsilon) + \tanh\beta(m-\varepsilon)\Big) \tag{2}$$

We also know finer properties of the quenched distribution above, like its dependence on the volume label N, for fixed realisation of the random fields. This can be asymptotically described by the in the "metastates formalism" [8], a notion due to Newman and Stein [19, 20]. For general background on this notion in the theory of disordered systems see [21, 2].

Knowing the correct solution, our point in this note will be however to put the Morita approximation scheme outlined above to the test. Now, in the Morita-approach one looks at the joint measures on the product space of the spin variables σ and the disorder variables η. They are simply composed from the quenched measures and the a priori uniform distribution of the random fields by the following obvious formula.

True joint measure:

$$K_{\beta,\varepsilon,h_0,N}(\sigma_{[1,N]}, \eta_{[1,N]}) = \frac{1}{2^N}\mu_{\beta,\varepsilon,h_0,N}[\eta_{[1,N]}](\sigma_{[1,N]}) \tag{3}$$

The approximant measure we want to consider is obtained by putting a single-site disorder potential $\lambda\sum_{i=1}^N \eta_i$ with just one free parameter λ that has the meaning of a chemical potential governing the mean value of the random fields.

Morita-approximant measure:

$$\hat{K}_{\lambda;\beta,\varepsilon,h_0,N}(\sigma_{[1,N]}, \eta_{[1,N]}) = \frac{\exp\Big(\frac{\beta}{2N}\big(\sum_{i=1}^N \sigma_i\big)^2 + \beta\sum_{i=1}^N(\varepsilon\eta_i + h_0)\sigma_i + \lambda\sum_{i=1}^N \eta_i\Big)}{Z_{\lambda;\beta,\varepsilon,h_0,N}} \tag{4}$$

We stress that the partition function does *not* depend on $\eta_{[1,N]}$, in contrast to (1). The Hamiltonian of this measure contains no non-local couplings of the random fields.

Then the idea of the naive Morita approximation-approach is as follows: 1) For any fixed λ, compute the large-N limit of distribution of this model. 2) Choose $\lambda = \lambda(\beta, \varepsilon, h_0)$ such that the expectation of the random fields coincides with the true joint measures, i.e. it vanishes, $\lim_N \int \hat{K}_{\lambda;\beta,\varepsilon,h_0,N}(d\eta_1)\eta_1 = 0$. More precisely the value of λ will depend on N, but it will have a well-defined limit as $N \uparrow \infty$. 3) Then, the distribution of the Morita approximant measure taken with this value of the bias of the random fields λ, should be close to the true joint measure. E.g. we should have that the distribution of the spin average $\frac{1}{N}\sum_{i=1}^N \sigma_i$ has the same infinite volume limit in the true joint measure and in the Morita approximant measure.

Let us write down the following precise formulation in order to have a well-defined starting point of discussion.

Single-Site approximation conjecture: Let β, ε, h_0 be fixed. Then the conjecture is that there is a value $\lambda(\beta, \varepsilon, h_0)$ such that $\lim_N \hat{K}_{\lambda;\beta,\varepsilon,h_0,N}(d\eta_1)\eta_1 = 0$ and that for this value we have that

$$\lim_{N\uparrow\infty} K_{\beta,\varepsilon,h_0,N}\Bigl(\frac{1}{N}\sum_{i=1}^{N}\sigma_i \in \cdot\Bigr) = \lim_{N\uparrow\infty} \hat{K}_{\lambda(\beta,\varepsilon,h_0);\beta,\varepsilon,h_0,N}\Bigl(\frac{1}{N}\sum_{i=1}^{N}\sigma_i \in \cdot\Bigr) \tag{5}$$

Is this conjecture true? How does this relate to the proved a.s. discontinuity of the conditional expectations of the true joint measures? Let us review the quick derivation of the solution of the model based on this conjecture (we follow here [14]).

Naive (problematic!) derivation: Look at the partition function of the Morita approximant measure, putting $h_0 = 0$ from the beginning, and use a simple Gaussian identity (Hubbard-Stratonovitch transformation) to write

$$Z_{\lambda;\beta,\varepsilon,N} = 2^N \int \frac{dm}{\sqrt{2\pi/(\beta N)}} \exp\Bigl(-\beta N \Phi_{\lambda;\beta,\varepsilon}(m)\Bigr) \tag{6}$$

The function appearing in the exponent is N-independent and is given below in (24) by putting $h_0 = 0$. Using the Laplace method to compute the integral we must have $\frac{\partial}{\partial m}\Phi_{\lambda;\beta,\varepsilon}(m) = 0$. This is an equation for the minimizer m of the form

$$m = \frac{\sum_{k=\pm 1}\sinh\bigl(\beta(m+\varepsilon k)\bigr)e^{\lambda k}}{\sum_{k=\pm 1}\cosh\bigl(\beta(m+\varepsilon k)\bigr)e^{\lambda k}} \tag{7}$$

The parameter λ is fixed such that the mean of the magnetic field sum divided by N vanishes, i.e. $\lim_{N\uparrow\infty}\frac{\partial}{\partial\lambda}\log Z_{\lambda;\beta,\varepsilon,N} = 0$. This requires at the minimizer m that $\frac{\partial}{\partial\lambda}\Phi_{\lambda;\beta,\varepsilon}(m) = 0$. This requires that

$$e^{-2\lambda} = \frac{\cosh(\beta(m+\varepsilon))}{\cosh(\beta(m-\varepsilon))} \tag{8}$$

The equation (8) shows that m and λ are in one-to-one correspondence to each other. From (7) and (8) follows the well known (and correct) mean field equation (2). So it seems that the Morita approximation approach becomes exact in this case and we are done.

Kühn writes appropriately: *This result [23] - simple and reassuring as it is - must be regarded as remarkable in the light of concerns raised about the appearance of non-Gibbsian measures within the equilibrium ensemble approach [5, 9, 6] and the identification of the RFIM as providing a realization of a system exhibiting almost surely non-Gibbsian joint measures [9, 11].*

Indeed, we note that the "derivation" is flawed because of the following fact.

Worrisome fact why this derivation is wrong: Suppose that $\beta > 1$, $\varepsilon > 0$ and $\lambda > 0$ are fixed. Then the minimum of the function $m \mapsto \Phi_{\lambda;\beta,\varepsilon}(m)$ is attained at a unique positive value $m^*(\lambda)$ (as we will see below). Therefore there cannot be a pair $(m^*(\lambda), \lambda)$ satisfying (8). So the solution (m, λ) obtained by (7), (8) corresponds to a wrong value for the free energy.

The remaining Morita mystery: Why does the wrong minimizer give the correct equation for the magnetization?

It is the purpose of this note to clarify the situation. We will be even more general and more careful here and allow for a possibly non-zero external magnetic field h_0. This we do in order to investigate whether taking the limit $h_0 \downarrow 0$ only in the end will help us to solve the problem of this approach.

We can readily solve the model for any choice of the parameters of inverse temperature β, strength of random fields ε, external field h_0 and Morita chemical potential λ. As usual in mean field models there is convergence to (linear combinations) of product measures over the sites i. Indeed, any limit measure must be a mixture of product measures. This is clear by de Finetti's theorem since the limit of exchangeable measures inherits the property of exchangeability.

Now, solving our simple model is almost trivial when we note that by summing over the η first we obtain a resulting effective Curie-Weiss Ising model with a new effective homogenous magnetic field acting on the σ's. The computations are simple and will be given below for the sake of completeness. Before we do so let us however state the most important consequence of this in the present context.

Theorem 2.1 (Impossibility of single-site approximation of true joint measure) *Assume that $\beta > 1$ and $\varepsilon > 0$ are fixed. Then*

$$\begin{aligned}&\Big\{h_0 \in \mathbb{R}, \exists \lambda \in \mathbb{R} : \lim_{N\uparrow\infty} \hat{K}_{\lambda;\beta,\varepsilon,h_0,N}(d\eta) = \mathbb{P}(d\eta)\Big\}\\ &= \mathbb{R}\backslash\Big[-a(\beta,\varepsilon), +a(\beta,\varepsilon)\Big]\end{aligned} \tag{9}$$

where $a(\beta, \varepsilon)$ is strictly bigger than zero. Here the symbol lim *denotes a weak limit.*

In words the theorem states that the set of external homogenous magnetic fields h_0 for which there exists a "compensating" Morita-field λ that reproduces the neutral i.i.d. distribution for the random fields is bounded away from zero for any $\beta > 1$. This means that the approximation scheme must necessarily fail in the relevant low temperature regime: First of all, in zero external field h_0 it is impossible to produce asymptotically symmetric i.i.d. random fields by an appropriate choice of λ. This result however, might not be too surprising. But the theorem says more: Even choosing h_0 strictly positive and letting it tend to zero afterwards won't help us.

Having said this, it is interesting to investigate the set of parameters for which the distribution of random fields becomes neutral i.i.d. in more detail. Let us make the following definition.

Neutral Set: Fix the inverse temperature $\beta > 0$ and $\varepsilon > 0$. We call the parameter set

$$\mathcal{R}(\beta,\varepsilon) := \Big\{(h_0, \lambda) \in \mathbb{R}\times\mathbb{R}, \lim_{N\uparrow\infty} \hat{K}_{\lambda;\beta,\varepsilon,h_0,N}(d\eta) = \mathbb{P}(d\eta)\Big\} \tag{10}$$

the neutral set. Obviously $\mathcal{R}(\beta, \varepsilon) = -\mathcal{R}(\beta, \varepsilon)$ by the symmetry of the model.

Then we have the following theorem.

Theorem 2.2 (Structure of neutral set) *Assume that $\beta > 0$ and $\varepsilon > 0$ are fixed. Then the set $\mathcal{R}(\beta, \varepsilon)$ is the union of two semi-infinite curves, related to each other by reflection at the origin. They are connected if and only if $\beta \leq 1$.*

More precisely these curves are of the following form.

There is a continuous increasing function $h_0 \mapsto l_{\beta,\varepsilon}(h_0)$ that is defined an open interval of the form $(a(\beta,\varepsilon),\infty)$ and takes positive values. The left endpoint of the interval satisfies $a(\beta,\varepsilon)\begin{cases} > 0 \text{ for } \beta > 1 \\ = 0 \text{ for } \beta \le 1 \end{cases}$.

Define

$$\mathcal{R}^+(\beta,\varepsilon) := \Big\{(h_0, -l_{\beta,\varepsilon}(h_0))\Big| h_0 \in (a(\beta,\varepsilon),\infty)\Big\} \tag{11}$$

Then, the neutral set has the form

$$\mathcal{R}(\beta,\varepsilon) = \begin{cases} \mathcal{R}^+(\beta,\varepsilon) \cup \Big(-\mathcal{R}^+(\beta,\varepsilon)\Big) & \text{for } \beta > 1 \\ \mathcal{R}^+(\beta,\varepsilon) \cup \Big(-\mathcal{R}^+(\beta,\varepsilon)\Big) \cup \Big\{(0,0)\Big\} & \text{for } \beta \le 1 \end{cases} \tag{12}$$

Hence, for $\beta > 1$ the neutral set is disconnected. For $\beta \le 1$ we have moreover $l^*_{\beta,\varepsilon}(0+) = 0$, and hence the neutral set is connected.

This result on the neutral set is a consequence of the solution of the Morita approximant for any choice of the parameters. We will now describe the behavior of the Morita approximant for general choice of the parameters. Then we will derive as a conclusion the explicit condition for the neutral set.

We need some definitions. Define the effective magnetic field-like parameter

$$\hat{h} = h_0 + \bar{h}_{\beta,\varepsilon}(\lambda) \tag{13}$$

with the function

$$\bar{h}_{\beta,\varepsilon}(\lambda) := \frac{1}{2\beta}\log\frac{\cosh(\lambda+\beta\varepsilon)}{\cosh(\lambda-\beta\varepsilon)} \tag{14}$$

Define the joint single-site measures depending on the parameter set, and on an additional (magnetization-like) parameter $m \in \mathbb{R}$.

$$\begin{aligned}\pi_{\lambda;\beta,\varepsilon,h_0}[m](\sigma_i,\eta_i) &:= \frac{\exp\Big(\beta\big((m+\varepsilon\eta_i+h_0)\sigma_i+\lambda\eta_i\big)\Big)}{2\sum_{k=\pm1}\cosh\big(\beta(m+\varepsilon k+h_0)\big)e^{\lambda k}} \\ &= \frac{\exp\big(\beta(m+\hat{h})\sigma_i\big)}{2\cosh\beta(m+\hat{h})}\frac{\exp\big((\beta\varepsilon\sigma_i+\lambda)\eta_i\big)}{2\cosh(\beta\varepsilon\sigma_i+\lambda)}\end{aligned} \tag{15}$$

Here we found it convenient to express the joint distribution on the r.h.s. appearing under the i-product in the form $\text{Prob}(\sigma_i,\eta_i) = \text{Prob}(\sigma_i)\text{Prob}(\eta_i|\sigma_i)$. In this way the marginal on the σ's can be readily read off. We see that the role of the parameter $\hat{h}$ is to provide an "effective magnetic field" acting on the spins.

Theorem 2.3 (Solution of Morita approximant) *Assume that the parameters $\beta,\varepsilon \in (0,\infty)$ and $\lambda, h_0 \in (-\infty,\infty)$ are fixed.*

(i): *Assume at first that $\hat{h} \ne 0$. Then we have the weak convergence*

$$\lim_{N\uparrow\infty}\hat{K}_{\lambda;\beta,\varepsilon,h_0,N}(\sigma_{[1,N_0]},\eta_{[1,N_0]}) = \prod_{i=1}^{N_0}\pi_{\lambda;\beta,\varepsilon,h_0}\big[m^{CW}(\beta,\hat{h})\big](\sigma_i,\eta_i) \tag{16}$$

Here we have denoted by $m^{CW}(\beta,h)$ the solution of $m=\tanh(\beta(m+h))$ that has the sign of h, for $h\neq 0$.

(ii): *For $\hat{h}=0$ we have the weak convergence to the symmetric linear combination of product measures*

$$\begin{aligned}&\lim_{N\uparrow\infty}\hat{K}_{\lambda;\beta,\varepsilon,h_0,N}(\sigma_{[1,N_0]},\eta_{[1,N_0]})\\&=\frac{1}{2}\prod_{i=1}^{N_0}\pi_{\lambda;\beta,\varepsilon,h_0}\big[m^{CW}(\beta,0+)\big](\sigma_i,\eta_i)+\frac{1}{2}\prod_{i=1}^{N_0}\pi_{\lambda;\beta,\varepsilon,h_0}\big[m^{CW}(\beta,0-)\big](\sigma_i,\eta_i)\end{aligned}\tag{17}$$

Remark: Of course $m^{\mathrm{CW}}(\beta,h)$ is the magnetization of an ordinary Curie Weiss Ising model in an external field.

Proof: Let us write the Morita approximant joint measure as a marginal on the σ's times the conditional measure of the random fields given the σ's, that is

$$\begin{aligned}&\hat{K}_{\lambda;\beta,\varepsilon,h_0,N}(\sigma_{[1,N]},\eta_{[1,N]})\\&=\frac{\exp\Big(\frac{\beta}{2N}\big(\sum_{i=1}^{N}\sigma_i\big)^2+\beta h_0\sum_{i=1}^{N}\sigma_i+\sum_{i=1}^{N}\log\cosh(\beta\varepsilon\sigma_i+\lambda)\Big)}{\mathrm{Norm}\,.}\prod_{i=1}^{N}\frac{e^{(\beta\varepsilon\sigma_i+\lambda)\eta_i}}{2\cosh(\beta\varepsilon\sigma_i+\lambda)}\end{aligned}\tag{18}$$

This shows us that the marginal distribution on the σ's is given by an ordinary ordered mean field Ising model of the form $\propto\exp\Big(\frac{\beta}{2N}\big(\sum_{i=1}^{N}\sigma_i\big)^2+\beta\hat{h}\sum_{i=1}^{N}\sigma_i\Big)/\mathrm{Norm}\,.$ with the effective field $\hat{h}$. From here the limit statements are obvious by the known convergence results of the Curie Weiss Ising model to the corresponding (linear combination of) product measures. □

Now, from the explicit solution we may derive explicit information on the neutral set. In order to do so, note at first the elementary properties

$$\bar{h}_{\beta,\varepsilon}(\lambda)\quad=\quad\begin{cases}\downarrow-\varepsilon, & \text{for }\lambda\downarrow-\infty\\0, & \text{for }\lambda=0\\\uparrow\varepsilon & \text{for}\lambda\uparrow\infty\,,\end{cases}$$

and it is a monotonically increasing in λ and odd. It maps $\mathbb{R}$ to the interval $(-\varepsilon,\varepsilon)$.

Theorem 2.4 (Explicit description of neutral set) *Assume that $\beta>0$ and $\varepsilon>0$ are fixed. The decomposition (12) holds with*

$$\begin{aligned}\mathcal{R}^{+}(\beta,\varepsilon)=\Big\{(h_0,-l)\Big|\,&0<\bar{h}_{\beta,\varepsilon}(l)<h_0,\\&m^{CW}\Big(\beta,h_0-\bar{h}_{\beta,\varepsilon}(l)\Big)=\frac{\sinh(2l)}{\sinh(2\beta\varepsilon)}\Big\}\end{aligned}\tag{19}$$

This set can be written as a graph in the form (11) with a continuous increasing $l_{\beta,\varepsilon}(h_0)$ that maps the interval $(a(\beta,\varepsilon),\infty)$ onto the interval $(\bar{h}_{\beta,\varepsilon}(a(\beta,\varepsilon)),\beta\varepsilon)$ where $a(\beta,\varepsilon)$ is uniquely given by $\bar{h}_{\beta,\varepsilon}\big(a(\beta,\varepsilon)\big)=\frac{1}{2}\sinh^{-1}\Big(\sinh(2\beta\varepsilon)\,m^{CW}\big(\beta,0+\big)\Big)$.

Remark: Note that the above expression for $a(\beta,\varepsilon)$ implies that $a(\beta,\varepsilon)=0$ if and only if the spontaneous magnetization $m^{\mathrm{CW}}\big(\beta,0+\big)$ vanishes, i.e. $\beta\leq 1$.

Proof: Suppose that $\beta>1$. Then, in order to have convergence to a symmetric product measure on the random fields we must have that the parameters $\lambda;\beta,\varepsilon,h_0$ are such that $\hat h\neq 0$. (Indeed, for $\hat h=0$ the distribution on the spins converges to a symmetric mixture of two different product measures. But from this it is obvious that also the random field distribution will be a mixture between two different product measures.) This shows that $(0,0)\not\in\mathcal{R}(\beta,\varepsilon)$ in that case.

Suppose however $\beta\leq 1$. Then $(h_0,\lambda)=(0,0)$ implies $\hat h=0$ which implies that the distribution of the σ's is a symmetric product measure. But this implies that the distribution of the random fields will be a symmetric product measure so that $(0,0)\not\in\mathcal{R}(\beta,\varepsilon)$ in that case.

So, we are left with the case $\hat h\neq 0$. We can treat the cases $\beta>1$ and $\beta\leq 1$ on a unified basis. Now, conditional on the value of σ the η_i have an expectation value of $\tanh(\lambda+\beta\varepsilon\sigma_i)$. We use the simple identity

$$\begin{aligned}&\tanh(\lambda+\beta\varepsilon\sigma_i)=\frac{B(1-L^2)\sigma_i+L(1-B^2)}{1-B^2L^2}\quad\text{where}\\&L=\tanh\lambda,\quad B=\tanh\beta\varepsilon\end{aligned}\tag{20}$$

for $\sigma_i=\pm 1$. So the distribution on the random fields η_i converges weakly to a product measure with individual expectation value

$$\lim_{N\uparrow\infty}\int\hat K_{\lambda;\beta,\varepsilon,h_0,N}(d\eta_1)\eta_1=\frac{B(1-L^2)m^{\mathrm{CW}}(\beta,\hat h)+L(1-B^2)}{1-B^2L^2}\tag{21}$$

Put $l=-\lambda$ and use $\tanh(l)/(1-\tanh^2(l))=\sinh(2l)$. So, in order to have the desired convergence to the symmetric product measure we must have

$$m^{\mathrm{CW}}(\beta,h_0-\bar h_{\beta,\varepsilon}(l))=\frac{\sinh(2l)}{\sinh(2\beta\varepsilon)}\tag{22}$$

This equation can only hold if $h_0-\bar h_{\beta,\varepsilon}(l)$ and l have the same sign. By symmetry we can assume that $l>0$. But this implies that $h_0>0$ (since $\bar h_{\beta,\varepsilon}(l)>0$.)

So, it suffices to look for all pairs (l,h_0) with $l>0$ that satisfy the consistency equations (22). The small trick we are using now is to fix the l and ask for h_0 rather than doing it the opposite way. Fixing l we see that the l.h.s. runs monotonically through the open interval $(m^{\mathrm{CW}}(\beta,0+),1)$ when h_0 runs in the "allowed range" $(\bar h_{\beta,\varepsilon}(l),\infty)$.

So, the set of $l>0$ such that there exists a solution h_0 is determined by the condition $(m^{\mathrm{CW}}(\beta,0+),1)\ni\frac{\sinh(2l)}{\sinh(2\beta\varepsilon)}$. Equivalently, this is the open interval $l\in(a(\beta,\varepsilon),\beta\varepsilon)$. Moreover the map to h_0 is continuous and monotone by known properties of the function $m^{\mathrm{CW}}(\beta,h)$. So it can be inverted and this yields the claim. $\square$

So what has happened in the naive (but wrong) derivation of the mean-field equations (7) and (8)? In order to see this let us write down a representation of the finite-N approximant measures. As a result of a Gaussian transformation on the level of measures we get the following formula.

Proposition 2.5 *In finite volume N we have the identity*

$$\hat{K}_{\lambda;\beta,\varepsilon,h_0,N}(\sigma_{[1,N]},\eta_{[1,N]}) = \frac{\int dm \exp\bigl(-\beta N \hat{\Phi}_{\beta,\varepsilon,h_0}(m)\bigr)}{\int d\tilde{m} \exp\bigl(-\beta N \hat{\Phi}_{\lambda;\beta,\varepsilon,h_0}(\tilde{m})\bigr)} \prod_{i=1}^{N} \pi_{\lambda;\beta,\varepsilon,h_0}[m](\sigma_i,\eta_i) \tag{23}$$

Here

$$\begin{aligned}\hat{\Phi}_{\lambda;\beta,\varepsilon,h_0}(m) &= \frac{m^2}{2} - \frac{1}{\beta}\log\sum_{k=\pm 1}\cosh\bigl(\beta(m+\varepsilon k+h_0)\bigr)e^{\lambda k}\\ &= \frac{m^2}{2} - \frac{1}{\beta}\log\cosh\bigl(\beta(m+\hat{h})\bigr) + Const(\beta,\varepsilon)\end{aligned} \tag{24}$$

where $Const(\beta,\varepsilon)$ does not depend on m.

Remark: The second equality for $\hat{\Phi}$ can be seen e.g. by reexpressing the first cosh as a sum over a spin $s = \pm 1$ and exchanging the s and k-sums.
Proof: We use a Gaussian transition kernel from the σ-variables to an auxiliary real valued variable m given by $T(dm|\sigma_{[1,N]}) = \exp\Bigl(-\frac{\beta N}{2}\Bigl(m - \frac{\sum_{i=1}^{n}\sigma_i}{N}\Bigr)^2\Bigr)dm/\text{Norm}$.. We define a "big joint measure" on the spins, the random fields and also the auxiliary magnetization-like continuous variable by the formula

$$\hat{M}_{\lambda;\beta,\varepsilon,h_0,N}(dm,\sigma_{[1,N]},\eta_{[1,N]}) := \hat{K}_{\lambda;\beta,\varepsilon,h_0,N}(\sigma_{[1,N]},\eta_{[1,N]})T(dm|\sigma_{[1,N]}) \tag{25}$$

We see that m concentrates very nicely around the value of the empirical average of the true spins in this measure. Then the non-normalized density of this "big joint measure" is given by $\exp\Bigl(-\frac{\beta N}{2}m^2 + \beta\sum_i\bigl((m+\varepsilon\eta_i+h_0)\sigma_i+\lambda\eta_i\bigr)\Bigr)$. Use this to express the "big joint measure" in the form of a marginal on the m times a conditional measure on the (σ,η) given the m. From here it is simple to get the desired formula. □

So, conditional on a value of m, the pairs (σ_i,η_i) are independent. We have then for their conditional mean values

$$\begin{aligned}\sum_{\sigma_1=\pm}\pi_{\lambda;\beta,\varepsilon,h_0}[m](\sigma_1)\sigma_1 &= \frac{\sum_{k=\pm 1}\sinh\bigl(\beta(m+\varepsilon k+h_0)\bigr)e^{\lambda k}}{\sum_{k=\pm 1}\cosh\bigl(\beta(m+\varepsilon k+h_0)\bigr)e^{\lambda k}}\\ \sum_{\eta_1=\pm}\pi_{\lambda;\beta,\varepsilon,h_0}[m](\eta_1)\eta_1 &= \frac{\sum_{k=\pm 1}k\cosh\bigl(\beta(m+\varepsilon k+h_0)\bigr)e^{\lambda k}}{\sum_{k=\pm 1}\cosh\bigl(\beta(m+\varepsilon k+h_0)\bigr)e^{\lambda k}}\end{aligned} \tag{26}$$

We remark that, with this notation, we have that (the version for general h_0 of) the saddle point equation (7) is equivalent to the consistency equation for the magnetization written as

$$m = \sum_{\sigma_1=\pm}\pi_{\lambda;\beta,\varepsilon,h_0}[m](\sigma_1)\sigma_1 \tag{27}$$

The (version for general h_0 of) the neutrality equation (8) is written as is equivalent to

$$0 = \sum_{\eta_1=\pm}\pi_{\lambda;\beta,\varepsilon,h_0}[m](\eta_1)\eta_1 \tag{28}$$

Now, the large-N limit of the model is obtained by looking at the absolute minimizer of the function $m \mapsto \hat{\Phi}(m)$. But note that the representation for $\hat{\Phi}(m)$ given in the second line shows that is has the double-well form of the corresponding function in an Ising model in the external field $\hat{h}$. It is an elementary property of this function that its absolute minimizer has the same sign as $\hat{h}$. But this shows that the relation (8) can not be true for the absolute minimizer. Instead the solution of (7), (8) corresponds to the second local minimum which is not the absolute minimum but the metastable minimum.

3 Validity of consistency equations and almost sure discontinuity of conditional expectations

So how can we understand the fact that the correct solution of the model is obtained by solving equations (27) and (28) although the solution corresponds to the wrong saddle point? The solution to this puzzle is due to the fact that the naive equations have rigorous counterparts in the following sense. The equations we are going to state now appear as consistency equations for the conditional probabilities of the *true* joint measures.

Proposition 3.1 (Consistency equations for true joint measure) *There is a function $\lambda_N(\eta_{[2,N]})$, depending on the parameters β, ε, h_0, which is invariant under permutation of $(\eta_i)_{i=2,\dots,N}$ such that we have*

$$\begin{aligned}&\sum_{\sigma_1} K_{\beta,\varepsilon,h_0,N}(\sigma_1)\sigma_1\\ &= \sum_{\sigma_{[2,N]},\eta_{[2,N]}} K_{\beta,\varepsilon,h_0,N}(\sigma_{[2,N]},\eta_{[2,N]})\left(\sum_{\sigma_1=\pm}\pi_{\lambda_N(\eta_{[2,N]});\beta,\varepsilon,h_0}\Big[\frac{1}{N}\sum_{i=2}^N\sigma_i\Big](\sigma_1)\sigma_1\right)\end{aligned} \tag{29}$$

$$0 = \sum_{\sigma_{[2,N]},\eta_{[2,N]}} K_{\beta,\varepsilon,h_0,N}(\sigma_{[2,N]},\eta_{[2,N]})\left(\sum_{\eta_1=\pm 1}\pi_{\lambda_N(\eta_{[2,N]});\beta,\varepsilon,h_0}\Big[\frac{1}{N}\sum_{i=2}^N\sigma_i\Big](\eta_1)\eta_1\right) \tag{30}$$

Proof of the proposition: The proof is based on the following lemma.

Lemma 3.2 (Representation of conditional probability of true joint measure) *The single-site conditional probabilities can be written in the form*

$$K_{\beta,\varepsilon,h_0,N}(\sigma_1,\eta_1|\sigma_{[2,N]},\eta_{[2,N]}) = \pi_{\lambda_N(\eta_{[2,N]});\beta,\varepsilon,h_0}\Big[\frac{1}{N}\sum_{i=2}^N\sigma_i\Big](\sigma_1,\eta_1) \tag{31}$$

where

$$\lambda_N(\eta_{[2,N]}) = \frac{1}{2}\log\frac{Z_{\beta,\varepsilon,h_0,N}[\eta_1=-,\eta_{[2,N]}]}{Z_{\beta,\varepsilon,h_0,N}[\eta_1=+,\eta_{[2,N]}]} \tag{32}$$

Proof of the lemma: By a simple computation we have for the single-site distribution

$$\begin{aligned}
&K_{\beta,\varepsilon,h_0,N}(\sigma_1,\eta_1|\sigma_{[2,N]},\eta_{[2,N]})\\
&=\frac{1}{\text{Norm}}\exp\Big(\beta\big(\frac{1}{N}\sum_{i=2}^N\sigma_i+\varepsilon\eta_1+h_0\big)\sigma_1+\frac{1}{2}\log\frac{Z_{\beta,\varepsilon,h_0,N}[\eta_1=-,\eta_{[2,N]}]}{Z_{\beta,\varepsilon,h_0,N}[\eta_1=+,\eta_{[2,N]}]}\times\eta_1\Big)
\end{aligned}\tag{33}$$

and this shows the claim. □

Continuing with the proof of the proposition we use the formula for the conditional probabilities writing

$$\begin{aligned}
&K_{\beta,\varepsilon,h_0,N}(\sigma_1,\eta_1)\\
&=\sum_{\sigma_{[2,N]},\eta_{[2,N]}}K_{\beta,\varepsilon,h_0,N}(\sigma_{[2,N]},\eta_{[2,N]})\quad\pi_{\lambda_N(\eta_{[2,N]});\beta,\varepsilon,h_0}\Big[\frac{1}{N}\sum_{i=2}^N\sigma_i\Big](\sigma_1,\eta_1)
\end{aligned}\tag{34}$$

But this equation gives the equation for the magnetization (29) by summing over σ_1. Using the symmetry of the distribution of η_1 we get (30). □

Let us now summarize what we know by the rigorous solution of the random field model about the limiting distribution of the pair of random quantities entering the single-site kernel π. In words, the distribution becomes sharp in the case of non-zero external field. It becomes sharp but double valued in the case of vanishing external field. In view of the last lemma this statement is a different way of saying that there is a jump of the conditional probabilities when the empirical random field sum of the conditioning is infinitesimally perturbed around its typical value 0. Now, the rigorous statement is as follows.

Theorem 3.3 (Convergence of true joint measures)
(i) *Suppose that $h_0>0$. Then we have the weak limit*

$$\lim_{N\uparrow\infty}K_{\beta,\varepsilon,h_0,N}\left(\frac{1}{N}\sum_{i=2}^N\sigma_i\in\cdot\quad,\lambda_N(\eta_{[2,N]})\in\cdot\right)\to\delta_{m^*(h_0)}\times\delta_{\lambda^*(h_0)}$$

Here $(m^*(h_0),\lambda^*(h_0))$ is a solution of the consistency equations (27) and (28).

(ii) *Suppose that $h_0=0$. Then*

$$\lim_{N\uparrow\infty}K_{\beta,\varepsilon,h_0=0,N}\left(\frac{1}{N}\sum_{i=2}^N\sigma_i\in\cdot\quad,\lambda_N(\eta_{[2,N]})\in\cdot\,\middle|\sum_{i=1}^N\eta_i>0\right)\to\delta_{m^*}\times\delta_{\lambda^*}$$

where (m^*,λ^*) is the unique solution of the consistency equations (27) and (28) with $m^*>0$ (and, as a consequence $\lambda^*<0$).

As a consequence we have

$$\lim_{N\uparrow\infty}K_{\beta,\varepsilon,h_0=0,N}\left(\frac{1}{N}\sum_{i=2}^N\sigma_i\in\cdot\quad,\lambda_N(\eta_{[2,N]})\in\cdot\right)\to\frac{1}{2}\delta_{m^*}\times\delta_{\lambda^*}+\frac{1}{2}\delta_{-m^*}\times\delta_{-\lambda^*}$$

Remark: We see that the system chooses the particular value of $\lambda_N(\eta_{[2,N]})$ (that has the opposite sign of the magnetisation) itself!

Proof: We only sketch the proof. We rewrite the quotient of partition functions appearing in the definition of $\lambda_N(\eta_{[2,N]})$ in the form

$$\lambda_N(\eta_{[2,N]}) = \frac{1}{2}\log \mu_{\beta,\varepsilon,h_0,N}[\eta_1 = +, \eta_{[2,N]}]\Big(\exp\big(-2\beta\varepsilon\sigma_1\big)\Big) \tag{35}$$

From here Theorem 3.3 follows from the work done for the quenched model in [8, 11]. Let us focus here only on the interesting case of vanishing external magnetic field $h_0 = 0$. In this case it was shown that, under the condition of positive sum of the random fields the empirical average of the spins concentrates sharply around the positive magnetisation m^* (positive solution of (2)) w.r.t. to the quenched Gibbs probability. (This is true for "typical values" of the random field sum, that is for $N^{\frac{1}{2}-\delta} \leq \sum_{i=1}^N \eta_i \leq N^{\frac{1}{2}+\delta}$, and these values get all mass w.r.t. $\mathbb{P}$ in the large-N limit). At the same time the quenched Gibbs probability $\mu_{\beta,\varepsilon,h_0,N}[\eta_1 = +, \eta_{[2,N]}](\sigma_1 = +)$ aquires a sharp value that is related in a simple way to m^*. From (35) this gives the value of λ^*. □

Not assuming the knowledge of the solution of the quenched model we can reverse the argument in the following way in order to solve the model. Look at the consistency equations for the true joint measure (29),(30). Take $h_0 > 0$. Then it is very plausible without much a priori knowledge that the distribution of the pair $\Big(\frac{1}{N}\sum_{i=2}^N \sigma_i, \lambda_N(\eta_{[2,N]})\Big)$ under the true joint measure should converge to a Dirac measure $\delta_{m,\lambda}$. (This is in particular clear, if we assume that λ_N has the form (35) and assume that the quenched magnetization becomes sharp for typical realization of the random fields in a positive homogeneous external field.) But this means that the outer integrals in the rigorous consistency equations become sharp. So, the limiting value (m, λ) must necessarily satisfy the naive consistency equations (27) and (28). These equations can then be solved and afterwards we let the external magnetic field h_0 tend to zero from above to discover the known solutions for the model.

Let us finally see that, in the case of $h_0 = 0$ the validity of the naive equations implies that *there must be* discontinuous behavior of the conditional expectations as a function of the average of the random fields appearing in the conditioning. Indeed, suppose that $\lambda_N(\eta_{[2,N]})$ were a continuous function of $\frac{1}{N}\sum_{i=2}^N \eta_i$. Then, by the law of large numbers it would have to be constant in the large-N limit. But by reasons of symmetry this constant would have to be zero in the case of $h_0 = 0$. But this is in contradiction to the non-trivial solution of the naive equations (7) and (8). To summarize the last line of argument in catchy terms: Non-Gibbsianness is necessary to help the metastable solution to provide the right answer.

We remark that the purpose of this note is not to attack the Morita approach in general as a valuable heuristic method in theoretical physics to predict the behavior of disordered systems when a rigorous analysis is not available or not yet available.

As pointed out to us by Reimer Kühn, one could also argue that the second of the naive equations (8), which demands that m and λ at the physical fixed point must have opposite sign, renders the region of integration for the partition function (6) which includes the other fixed point as *unphysical* and so to be excluded from the domain of integration. This line of reasoning would render the ”naive” argument correct and this is not the first occasion in physics where such things happen. While there seems no direct rigorous justification for this procedure we have shown that one is able to understand the validity of the naive equations by viewing the parameter λ properly as a stochastic quantity. This might give hope that results obtained by approximation schemes based on the Morita approach provide correct answers also in more complicated situations where a rigorous analysis is lacking. A better understanding of this would pose a fascinating challenge.

Acknowledgments:

The author thanks Aernout van Enter and Reimer Kühn for discussions at the conference ”Gibbs vs. non-Gibbs” (EURANDOM, december 2003) which led to this paper and valuable comments on an earlier draft.

References

[1] Amaro de Matos, J.M.G., Patrick, A.E., Zagrebnov,V.A. (1992), Random Infinite-Volume Gibbs States for the Curie-Weiss Random Field Ising Model., *J. Statist. Phys.* **66**, 139–164.

[2] Bovier, A. (2001) Statistical Mechanics of Disordered Systems, *MaPhySto lecture notes* **10**, available at http://www.maphysto.dk/

[3] Bricmont J., Kupiainen A. (1988) Phase transition in the $3d$ random field Ising model. *Comm. Math. Phys.* **142**, 539–572.

[4] 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.

[5] van Enter, A.C.D., Maes, C., Schonmann, R.H. and Shlosman, S. (2000) The Griffiths singularity random field, *On Dobrushin's way. From probability to statistical mechanics* (R. Minlos, Yu. Suhov and S. Shlosman, eds) pp 59–70, AMS.

[6] van Enter, A.C.D., Maes C. and Külske C. (2000) Comment on: [15], *Phys. Rev. Lett.* **84**, 6134.

[7] Häggström, O. and Külske, C. (2004) Gibbs properties of the fuzzy Potts model on trees and in mean field, to appear in *Markov Proc. Relat. Fields*

[8] Külske, C. (1997) Metastates in Disordered Mean-Field Models: Random Field and Hopfield Models, *J. Statist. Phys.* **88** 5/6, 1257–1293 (1997)

[9] Külske, C. (1999) (Non-) Gibbsianness and phase transitions in random lattice spin models, *Markov Proc. Relat. Fields* **5**, 357–383.

[10] Külske, C. (2001) Weakly Gibbsian representations for joint measures of quenched lattice spin models, *Probab. Th. Relat. Fields* **119**, 1–30.

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

[12] Külske, C., Le Ny, A., Redig, F. (2004) Relative entropy and variational properties of generalized Gibbsian measures, to appear in *Ann. Probab.*

[13] Kühn, R (1996) Equilibrium Ensemble Approach to Disordered Systems I: General Theory, Exact Results', *Z. Phys.* **100**, 231–242.

[14] Kühn, R (2004) Gibbs vs. Non-Gibbs in the Equilibrium Ensemble Approach to Disordered Systems, preprint, to appear in *Markov Proc. Relat. Fields*

[15] Kühn, R. (1994) Critical behavior of the randomly spin diluted 2D Ising model: A grand ensemble approach, *Phys. Rev. Lett.* **73**, 2268–2271.

[16] Kühn, R. and Mazzeo, G. reply to : [6], *Phys. Rev. Lett.* **84**, 6135 (2000)

[17] Maes, C., Redig, F, Van Moffaert, A. (1999) Almost Gibbsian versus weakly Gibbsian measures. *Stochastic Process. Appl.* **79**, 1–15.

[18] Morita, T. (1964) Statistical Mechanics of quenched solid solutions with application to magnetically dilute alloys, *J. Math. Phys.* **5**:1402–1405.

[19] Newman, C. M., Stein, D. (1996) Non-mean-field behavior of realistic spin glasses, *Phys. Rev. Lett.* **76**, 515–518.

[20] Newman, C. M., Stein, D. (1996) Spatial inhomogeneity and thermodynamic chaos, *Phys. Rev. Lett.* **76**, 4821–4824.

[21] Newman, C. M. (1997) *Topics in disordered systems*, Birkhäuser, Basel.

[22] Salinas, S.R., Wreszinski, W.F. (1985) On the Mean-Field Ising Model in a Random External Field, *J. Statist. Phys.* **41**, 299–313.

[23] Serva, M. and Paladin, G. (1993), Gibbs thermodynamic potentials for disordered systems, *Phys. Rev. Lett.* **70**, 105–108.