Large deviations of the finite-time magnetization of the Curie-Weiss random-field Ising model

Pierre Paga* and Reimer Kühn†

Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom

(Received 2 June 2016; revised manuscript received 16 January 2017; published 14 August 2017)

We study the large deviations of the magnetization at some finite time in the Curie-Weiss random field Ising model with parallel updating. While relaxation dynamics in an infinite-time horizon gives rise to unique dynamical trajectories [specified by initial conditions and governed by first-order dynamics of the form $m_{t+1} = f(m_t)$], we observe that the introduction of a finite-time horizon and the specification of terminal conditions can generate a host of metastable solutions obeying *second-order* dynamics. We show that these solutions are governed by a Newtonian-like dynamics in discrete time which permits solutions in terms of both the first-order relaxation ("forward") dynamics and the backward dynamics $m_{t+1} = f^{-1}(m_t)$. Our approach allows us to classify trajectories for a given final magnetization as stable or metastable according to the value of the rate function associated with them. We find that in analogy to the Freidlin-Wentzell description of the stochastic dynamics of escape from metastable states, the dominant trajectories may switch between the two types (forward and backward) of first-order dynamics. Additionally, we show how to compute rate functions when uncertainty in the quenched disorder is introduced.

DOI: 10.1103/PhysRevE.96.022126

I. INTRODUCTION

Disordered systems are characterized by large fluctuations in their physical observables induced by the presence of quenched randomness. This creates a situation where not only average physical quantities and the variances of fluctuations about these averages are of interest, but their entire distribution has physical relevance. In this context, the study of large deviations arises naturally [1,2], and their applicability is very general [3]. As a result, the topic has attracted much attention in recent years in a diversity of fields, including spin glasses [4–6], kinetically constrained models [7,8], random matrix theory [9–11], and epidemic spreading on networks [12,13].

The study of large deviations also has economic importance [14]: it is vital, for example, for an insurer to estimate the likelihood of situations in which large numbers of indemnities may have to be paid at once. From the regulator's perspective, understanding how often large-scale crises may occur in a given regulatory scenario is key to proper risk-management policy. But while attempts at studying rare events in semi-realistic credit risk models have been made [15], computing large-deviation functions has proven too difficult due in part to the presence of quenched disorder.

This difficulty motivates the study of simple models with quenched disorder for which a complete analysis of dynamical large deviation properties is possible and which could therefore be used as testbeds for approximation methods. Of such simple models, the Ising model is the prototypical example. Recently, the large deviations of the Ising model have received great attention, e.g., studies of the large deviations of the energy in the 1D Ising chain, or of the activity (the number of times a spin flips in a given trajectory), as in Refs. [16,17]. A natural extension of the Ising model to include quenched disorder is the random-field Ising model (RFIM). The large deviations of the equilibrium magnetization have long been well understood [18,19] and have proven a useful tool in understanding the phase diagram of the model. To the best of our knowledge, however, no studies of the large deviations of the finite-time magnetization in the RFIM have been made even in its simplest (Curie-Weiss) description. Yet it is of clear interest to have a simple but nontrivial model with disorder that admits closed-form solutions against which to test our intuition and approximate methods.

In this article, we thus treat the large deviations of the finite-time magnetization in the Curie-Weiss random field Ising model with parallel updating. Using generating functional methods, we derive equations of motion that include the value of the magnetization at a finite time as a constraint and find that it can be recast in a language reminiscent of Newtonian dynamics in discrete time. This language also arises naturally as a zero-noise limit of Langevin dynamics, where the magnitude of the noise scales like $N^{-1/2}$, with N the system size.

The remainder of this paper is organized as follows: we introduce the model and our main notations, then introduce a discrete time path integral formalism from which we derive saddle-point equations. We analyze these equations and find nontrivial fixed points (i.e., fixed points that are not equilibrium solutions) and find them to be elliptical. We then derive Newtonian-like dynamics for the system at the saddle point. We derive rate functions and compare them with explicit simulations for small numbers of time steps ($T = 50$) in the ferromagnetic parts of the phase diagram. Finally, we show how to incorporate uncertainty in the disorder and derive the associated rate function.

II. MODEL

We investigate the parallel dynamics of the Curie-Weiss RFIM, where the transition probabilities between configuration $\{\sigma_i\}_{i=1,N}$ and $\{\sigma_i'\}_{i=1,N}$ is given by

$$W(\{\sigma_i'\}|\{\sigma_i\}) = \prod_i \frac{e^{\beta\sigma_i'(Jm+h\theta_i)}}{2\cosh[\beta(Jm+h\theta_i)]}, \tag{1}$$

where $m = N^{-1}\sum_i \sigma_i$ is the magnetization of configuration $\{\sigma_i\}$, and the θ_i are random fields that take value in $\{\pm 1\}$. We

*pierre.paga@kcl.ac.uk
†reimer.kuehn@kcl.ac.uk

take $J = 1$ without loss of generality and assume the initial spins are independent and identically distributed: $P(\{\sigma_{i0}\}) = \prod_i p_0(\sigma_{i0})$. We parametrize the distribution of initial spins as $p_0(\sigma) = \frac{1+\sigma r_0}{2}$.

In most of our exposition, we will assume the random fields configuration to be fixed. We write $\langle \cdots \rangle$ the average with respect to the dynamics in Eq. (1), and by $\langle \cdots \rangle_\theta$ the empirical average with respect to the random fields:

$$\langle g(\theta)\rangle_\theta \equiv \frac{1}{N}\sum_i g(\theta_i) = \frac{N_+}{N} g(1) + \left(1 - \frac{N_+}{N}\right) g(-1), \tag{2}$$

where N_+ is the number of sites i such that $\theta_i = +1$. We will assume $\frac{N_+}{N} = \frac{1}{2} + O(\frac{1}{\sqrt{N}})$ in all subsequent numerical applications unless otherwise specified; i.e., we assume that the disorder configuration is *typical* for a symmetric random field configuration.

With such a setup, the probability of a sequence of configurations $\{\sigma_{it}\}$ is given by

$$P(\{\sigma_{it}\}) = \prod_{t=1}^{T} W(\{\sigma_{it}\}|\{\sigma_{i(t-1)}\}) \prod_i p_0(\sigma_{i0}). \tag{3}$$

At large times, the probability distribution of the magnetization is well-known as the equilibrium dynamics are governed by Peretto's pseudo-Hamiltonian ([20,21]),

$$H_\beta(\sigma) = -\frac{1}{\beta}\sum_i \log[2\cosh(\beta[m + h\theta_i])] - h\sum_i \sigma_i\theta_i, \tag{4}$$

which gives the rate function

$$I_\beta(m) = \lim_{N\to\infty} -\frac{1}{N}\log P\left(\frac{1}{N}\sum_i \sigma_i = m\right) \tag{5}$$

$$= \sup_x\{xm - \langle \log[\cosh(\beta h\theta + x)]\rangle_\theta\} - \langle \log[\cosh(\beta[h\theta + m])]\rangle_\theta + I_0. \tag{6}$$

Here, I_0 is a constant chosen such that $\min_m\{I_\beta(m)\}=0$. Such rate functions are plotted in Fig. 1.

But the understanding of the equilibrium distribution alone may not be sufficient in many problems and provides no way to estimate, e.g., the equilibration time. To remedy this, we compute the large deviation functions of the magnetization at finite times.

It is not harmless, in this setting, to assume a fixed *typical* random field configuration: if we assume a binomial distribution in the random fields (e.g., if the random fields at each site are independently distributed), a fluctuation of order $O(1)$ in the fraction of sites with $\theta_i = +1$ is exponentially suppressed to the same extent as the fluctuations in the magnetization whose probability we are trying to compute. In other words, in the absence of prior knowledge (e.g., in the absence of a large number of prior measurements) the observation of an anomalous magnetization is *equally* likely to be a rare result of the dynamics as it is to be the signature of an anomalous sample. We will show in our last section how to compute rate functions in this situation.

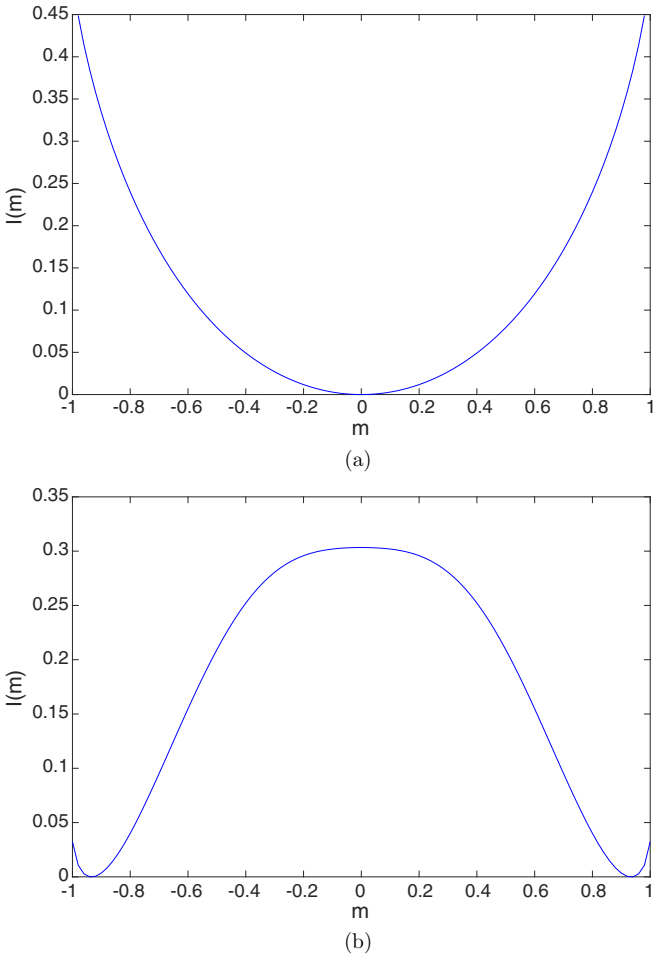


FIG. 1. Peretto rate functions in the paramagnetic ((a), $\beta = 0.67$, $h = 0.2$) and ferromagnetic ((b), $\beta = 2.5$, $h = 0.4$) parts of the phase diagram.

III. DYNAMICS OF THE MAGNETIZATION

We seek to compute the large deviation rate function of the magnetization m_T for some finite T.

A. Trajectory probability

Writing $\mathbf{m} = (m_t)_{t=0,\cdots,T}$, magnetization path probabilities are given by

$$P(\mathbf{m}) = \left\langle \prod_t \delta\left(m_t - \frac{1}{N}\sum_i \sigma_{it}\right)\right\rangle \tag{7}$$

$$= \sum_{\{\sigma_{it}\}}\prod_i p_0(\sigma_{i0}) \prod_{t=1}^{T}\left[\frac{e^{\beta(\sigma_{it}m_{t-1} + h\theta_i\sigma_{it})}}{2\cosh[\beta(m_{t-1} + h\theta_i)]}\right] \times \prod_{t=0}^{T}\delta\left(m_t - \frac{1}{N}\sum_i \sigma_{it}\right). \tag{8}$$

We insert the integral representation $\delta(x) = \int \frac{dy}{2\pi} e^{ixy}$ at every time step:

$$P(\mathbf{m}) = \sum_{\{\sigma_{it}\}} e^{N\beta \sum_{t=1}^{T} m_t m_{t-1}} \prod_i p_0(\sigma_{i0}) \times \prod_{t=1}^{T} \frac{e^{\beta h \theta_i \sigma_{it}}}{2\cosh[\beta(m_{t-1} + h\theta_i)]} \left[\frac{N}{2\pi}\right]^{T+1} \times \int d\hat{\mathbf{m}} \exp\left\{-i\hat{\mathbf{m}} \cdot \left(N\mathbf{m} - \sum_i \boldsymbol{\sigma}_i\right)\right\}, \quad (9)$$

where we write $\hat{\mathbf{m}} = (\hat{m}_t)_{t=0,\cdots,T}$ and $(\sigma)_i = (\sigma_{it})_{t=0,\cdots,T}$. This representation allows for the decoupling of the sum over microstates $\{\sigma_{it}\}$ with respect to the sites i, and thus $P(\mathbf{m})$ can be expressed as

$$P(\mathbf{m}) = \left[\frac{N}{2\pi}\right]^{T+1} \int d\hat{\mathbf{m}} \exp\left\{-iN\hat{\mathbf{m}} \cdot \mathbf{m} + N\beta \sum_{t=1}^{T} m_t m_{t-1} - \sum_i \sum_{t=1}^{T} \log\left[2\cosh\left(\beta[m_{t-1} + h\theta_i]\right)\right]\right\} \times \prod_i \sum_{\{\sigma_{it}\}} p_0(\sigma_{i0}) e^{i\hat{m}_0} \exp\{\sigma_t(\theta_i + i\hat{m}_t)\} \quad (10)$$

and can be expressed as a discrete time path integral,

$$P(\mathbf{m}) = \left[\frac{N}{2\pi}\right]^{T+1} \int d\hat{\mathbf{m}} \exp\left\{-iN\hat{\mathbf{m}} \cdot \mathbf{m} + \beta N \sum_{t=1}^{T} m_t m_{t-1} + N \sum_{t=0}^{T} \langle \log[Z_t(\theta)] \rangle_\theta\right\}, \quad (11)$$

with

$$Z_0(\theta) = \sqrt{1 - r_0^2} \frac{\cosh[\rho + i\hat{m}_0]}{\cosh[\beta(m_0 + h\theta)]}, \quad (12)$$

$$Z_t(\theta) = \frac{\cosh[i\hat{m}_t + \beta h\theta]}{\cosh[\beta(m_t + h\theta)]}, \quad 1 \leqslant t \leqslant T-1, \quad (13)$$

$$Z_T(\theta) = \cosh[i\hat{m}_T + \beta h\theta], \quad (14)$$

where $\rho = \tanh^{-1}(r_0)$.

B. Final magnetization

To find the marginal probability of the final magnetization m_T, we integrate over the rest of the trajectory. This gives

$$P(m_T) \propto \int \left[\prod_{t=0}^{T-1} dm_t\, d\hat{m}_t\right] d\hat{m}_T \exp\{-N\Omega(\mathbf{m},\hat{\mathbf{m}})\}, \quad (15)$$

with

$$\Omega(\mathbf{m},\hat{\mathbf{m}}) = \sum_{t=0}^{T} i\hat{m}_t m_t - \beta \sum_{t=1}^{T} m_{t-1} m_t - \sum_{t=0}^{T} \langle \log[Z_t(\theta)] \rangle_\theta. \quad (16)$$

The integral in Eq. (15) can at large N be evaluated by the saddle-point method. At the saddle point, we have the following conditions: at $t = 0$

$$m_0 = \tanh[\rho + i\hat{m}_0], \quad (17)$$

$$i\hat{m}_0 = \beta(m_1 - \langle \tanh[\beta(m_0 + \theta)] \rangle_\theta), \quad (18)$$

for $1 \leqslant t \leqslant T-1$:

$$m_t = \langle \tanh[i\hat{m}_t + \beta h\theta] \rangle_\theta, \quad (19)$$

$$i\hat{m}_t = \beta(m_{t-1} + m_{t+1} - \langle \tanh[\beta(m_t + \theta)] \rangle_\theta), \quad (20)$$

and at $t = T$,

$$m_T = \langle \tanh[i\hat{m}_T + \beta h\theta] \rangle_\theta. \quad (21)$$

Thus, the rate function for the final magnetization m_T reads

$$I(m) = -\lim_{N\to\infty} \frac{1}{N} \log P(m_T = m) = \Omega(\mathbf{m}^*, \hat{\mathbf{m}}^*)|_{m_T^* = m}, \quad (22)$$

where starred quantities denote saddle-point values.

C. Finite-time solutions

To simplify our expressions, we introduce the notations

$$f(x) = \langle \tanh[\beta(x + h\theta)] \rangle_\theta, \quad (23)$$

$$f_0(x) = \tanh[\rho + \beta x], \quad (24)$$

likewise,

$$F(x) = \langle \log[\cosh(\beta(x + h\theta))] \rangle_\theta, \quad (25)$$

$$F_0(x) = \tfrac{1}{2} \log\left(1 - r_0^2\right) + \log[\cosh(\rho + \beta x)], \quad (26)$$

finally,

$$\tilde{f}(x) = f(x) + f^{-1}(x), \quad (27)$$

$$\tilde{f}_0(x) = f_0^{-1}(x) + f(x). \quad (28)$$

Using these notations, the equations of motion, Eqs. (17)–(21), are rewritten as

$$m_0 = f_0(i\beta^{-1}\hat{m}_0), \quad (29)$$

$$i\hat{m}_0 = \beta(m_1 - f(m_0)), \quad (30)$$

$$m_t = f(i\beta^{-1}\hat{m}_t), \quad (31)$$

$$i\hat{m}_t = \beta(m_{t-1} + m_{t+1} - f(m_t)), \quad (32)$$

$$m_T = f(i\beta^{-1}\hat{m}_T). \quad (33)$$

We can insert the equations for the $i\hat{m}_t$ quantities in the equations for the m_t, and using the fact that both f and f_0 are invertible (see Appendix), we obtain

$$m_1 = \tilde{f}_0(m_0), \quad (34)$$

$$m_{t+1} + m_{t-1} = \tilde{f}(m_t), \quad (35)$$

$$i\hat{m}_T = f^{-1}(m_T), \quad (36)$$

and we can rewrite the Ω function at the saddle point in $\hat{\mathbf{m}}$ as a function of $\mathbf{m}$ only:

$$\begin{aligned}\omega(\mathbf{m}) &\equiv \Omega(\mathbf{m},\hat{\mathbf{m}}^*)\\ &= \sum_{t=0}^{T} i\hat{m}_t m_t - \beta\sum_{t=1}^{T} m_{t-1}m_t + F(m_0) - F_0(i\beta^{-1}\hat{m}_0)\\ &\quad + \sum_{t=1}^{T-1}[F(m_t) - F(i\beta^{-1}\hat{m}_t)] - F(i\beta^{-1}\hat{m}_T) \qquad (37)\\ &= \sum_{t=1}^{T}[\beta m_t(f^{-1}(m_t) - m_{t-1}) + F(m_t) - F(f^{-1}(m_t))]\\ &\quad + F(m_0) - F_0\big[f_0^{-1}(m_0)\big] + \beta m_0 f_0^{-1}(m_0) - F(m_T) \qquad (38)\\ &= \omega_0(m_0) + \omega(\mathbf{m}|m_0,m), \qquad (39)\end{aligned}$$

where

$$\omega_0(m_0) = \beta m_0 f_0^{-1}(m_0) - F_0\big[f_0^{-1}(m_0)\big] \qquad (40)$$

$$\begin{aligned}\omega(\mathbf{m}|m_0,m) = &\sum_{t=1}^{T}[\beta m_t(f^{-1}(m_t) - m_{t-1}) + F(m_t)\\ &- F(f^{-1}(m_t))] + F(m_0) - F(m_T). \qquad (41)\end{aligned}$$

In Eq. (38), ω_0 appears naturally as the influence of the distribution of the initial magnetization, while $\omega(\mathbf{m}|m,m_0)$ is the *conditional* contribution given the starting and end-points m_0 and m.

Solutions to Eq. (35) can be parametrized via any choice of two points on the trajectory, although not all choices lead to physical trajectories (i.e., $|m_t| < 1$ for all t). With the additional constraint of Eq. (34) and since m_T is fixed, we expect there to be only a finite number of possible solutions. Equation (36), meanwhile, has no effect on the trajectories themselves, it only intervenes in the computation of Ω.

Moreover, we notice that Eq. (35), which can be considered the proper "equation of motion" of the system, include the "unconstrained" dynamics as solutions: the unconstrained (or average) equations of motions being

$$m_{t+1} = f(m_t). \qquad (42)$$

Moreover, Eq. (35) is time-reversal invariant, and thus the time-reversed trajectories

$$m_{t+1} = f^{-1}(m_t) \qquad (43)$$

are also valid solutions when far from $t = 0$.

IV. EFFECTIVE DYNAMICS

A. Quasi-Newtonian dynamics

Given Eqs. (34) and (35), the dynamics of the system can be recast into a form reminiscent of Newtonian dynamics:

$$\Delta_t^2 m \equiv (m_{t+1} - m_t) - (m_t - m_{t-1}) = k(m_t), \qquad (44)$$

where we recognize in the left-hand side a discrete time second derivative and write

$$k(x) = \tilde{f}(x) - 2x. \qquad (45)$$

In Figs. 2 and 3 we plot phase portraits of trajectories corresponding to solutions of Eq. (35). We compute the trajectories by specifying m_0 and m_1, taken from a uniform grid in $[-1,1]^2$, and computing the values of the subsequent magnetization according to Eq. (35). A subset of these solutions represent solutions of the full system Eqs. (34) and (35) for suitably chosen m_T and r_0. Additionally, we plot the relaxation dynamics map $m_{t+1} = f(m_t)$ and the diagonal $m_{t+1} = m_t$ to highlight the corresponding equilibrium fixed points. Since we are dealing with discrete dynamics, it is natural to plot the "momentum" $m_{t+1} - m_t$ as a function of the average position $\frac{m_t+m_{t+1}}{2}$ rather than m_t or m_{t+1} to avoid a tilting of the phase portraits. The potential wells appear clearly, with up to four potential wells for $\beta = 2.5, h = 0.485$ in the regime where a ferromagnetic phase coexists with a metastable paramagnetic phase.

B. Energy conservation

Since Eq. (44) is reminiscent of Newtonian dynamics, we expect it to follow a discrete form of energy conservation. Indeed, multiplying both sides by $(m_{t+1} - m_{t-1})/2$, we have

$$\frac{(m_{t+1} - m_t)^2 - (m_t - m_{t-1})^2}{2} = \frac{1}{2}(m_{t+1} - m_{t-1})k(m_t). \qquad (46)$$

We now sum between two times t_1 and t_2:

$$\begin{aligned}&\frac{1}{2}\left(\big[m_{t_2+1} - m_{t_2}\big]^2 - \big[m_{t_1} - m_{t_1-1}\big]^2\right)\\ &\quad = \sum_{t_1}^{t_2}\frac{1}{2}(m_{t+1} - m_{t-1})k(m_t) \qquad (47)\\ &= V_{t_1} - V_{t_2}, \qquad (48)\end{aligned}$$

with

$$V_t = -\sum_{\tau=1}^{t}\frac{1}{2}(m_{\tau+1} - m_{\tau-1})k(m_\tau). \qquad (49)$$

Thus, we have

$$E \equiv \frac{1}{2}\big[m_{t_2+1} - m_{t_2}\big]^2 + V_{t_2} = \frac{1}{2}\big[m_{t_1} - m_{t_1-1}\big]^2 + V_{t_1}, \qquad (50)$$

which shows conservation of energy in discrete time, V playing the role of a path-dependent potential. Notice that if the increments $m_{\tau+1} - m_\tau$ are small throughout the trajectory (up to t), then

$$\begin{aligned}V_t &\simeq -\int_{(m_1+m_0)/2}^{(m_t+m_{t+1})/2} k(m)dm\\ &= V((m_t + m_{t+1})/2) - V((m_1 + m_0)/2). \qquad (51)\end{aligned}$$

Thus, in the low-speed regime the dynamics are Newtonian-like with a purely position-dependent potential. Up to a constant,

$$V(x) = x(x - f^{-1}(x)) + \beta^{-1}(F(f^{-1}(x)) - F(x)). \qquad (52)$$

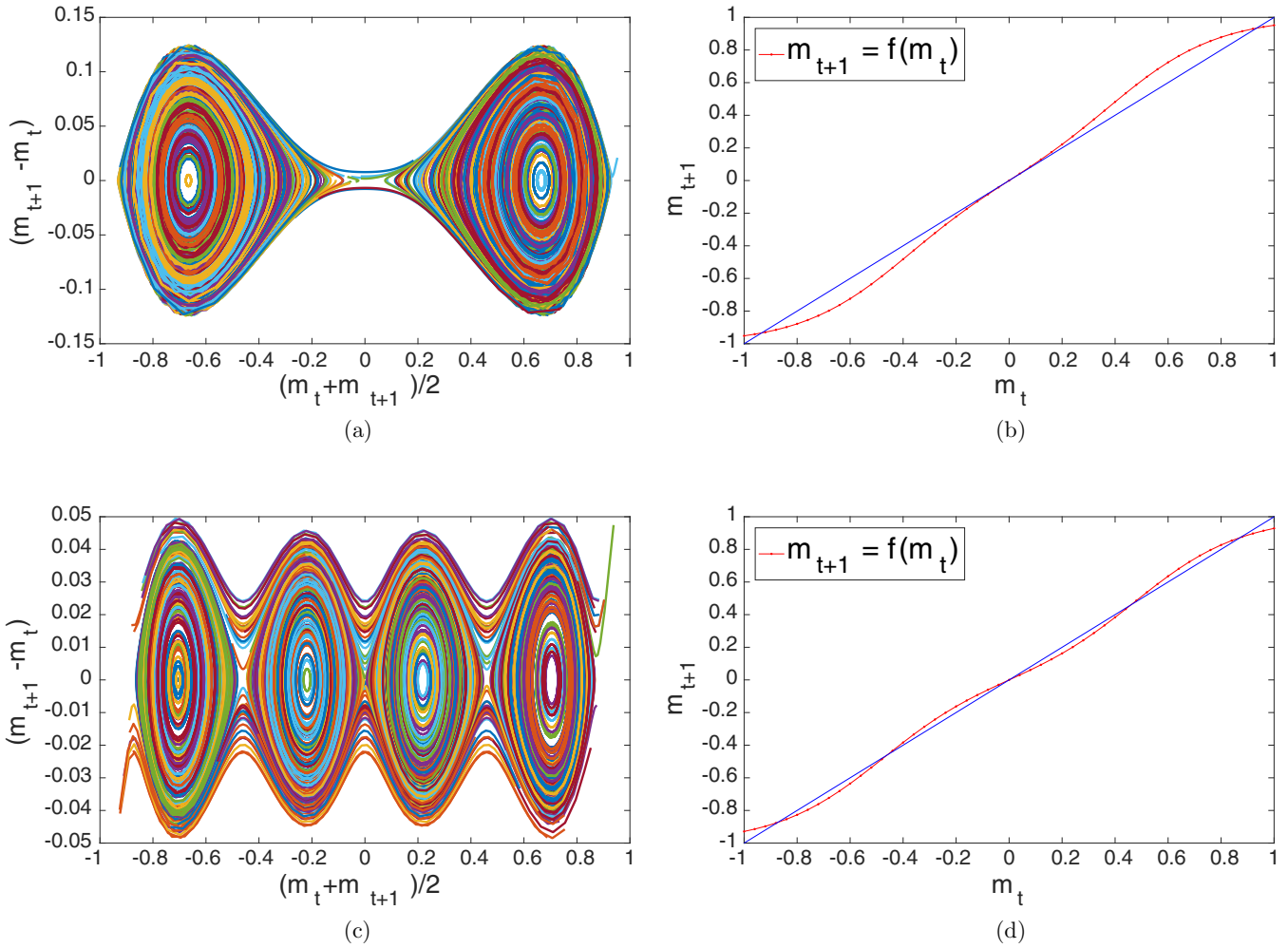


FIG. 2. Phase portraits of trajectories described by Eq. (35) [(a) and (c)] and maps of the relaxation dynamics $m_{t+1} = f(m_t)$ together with the diagonal $m_{t+1} = m_t$ [(b) and (d)] for the parameter settings: (a, b) $\beta = 2.5, h = 0.4$ (ferromagnetic phase), (c, d) $\beta = 2.5, h = 0.485$ (ferromagnetic phase with metastable state at $m = 0$).

We plot the position-dependent potential and the path-dependent potential in Fig. 4. We obtain the trajectory by solving Eqs. (34) and (35) with m_T fixed in a range of values in $[-1,1]$ using a numerical nonlinear equation solver. The path-dependent potential for a trajectory up to m_t is plotted as a function of $m = (m_t + m_{t+1})/2$, following the convention used in Eq. (51). We see that the position-dependent potential is a good approximation for the path-dependent potential.

C. Influence of initial conditions

Since the dynamics is second order, two boundary conditions are required to specify a solution. Notice that Eq. (34), which we can rewrite as

$$m_1 - m_0 = \tilde{f}_0(m_0) - m_0, \tag{53}$$

provides an initial condition in the form of an initial velocity field.

We can therefore compute and plot the initial total energy of a trajectory as a function of m_0 in the continuum approximation of Eq. (51), and compare it with the position-dependent potential. We do so in Fig. 5.

If $E > V_c$, where V_c is the level of the rightmost potential energy peak, the trajectory is unbounded and will eventually leave the domain $[-1,1]$, which is forbidden. But because the time horizon is finite, some trajectories do not have time to leave the domain in T time steps, and so we see some trajectories with initial energy $E > V_c$. As T increases, however, the range of allowed values of E above V_c becomes narrower. If $E < V_c$, on the other hand, the trajectory is bounded by the potential energy peaks. At large T, these trajectories will oscillate around the minimum of a potential well.

V. RATE FUNCTIONS

A. Dominant trajectories

Once we have obtained solutions to the equations of motion, Eqs. (34) and (35), we can compute the associated value of ω. We plot such values in Fig. 6. Due to the existence of metastable trajectories, for each m there can be several possible trajectories ending in $m_T = m$ and thus several possible values of $I(m_T)$, as can be seen in Fig. 6(a). The physically relevant rate function is obtained by taking the minimum value of $I(m_T)$ at any given m_T.

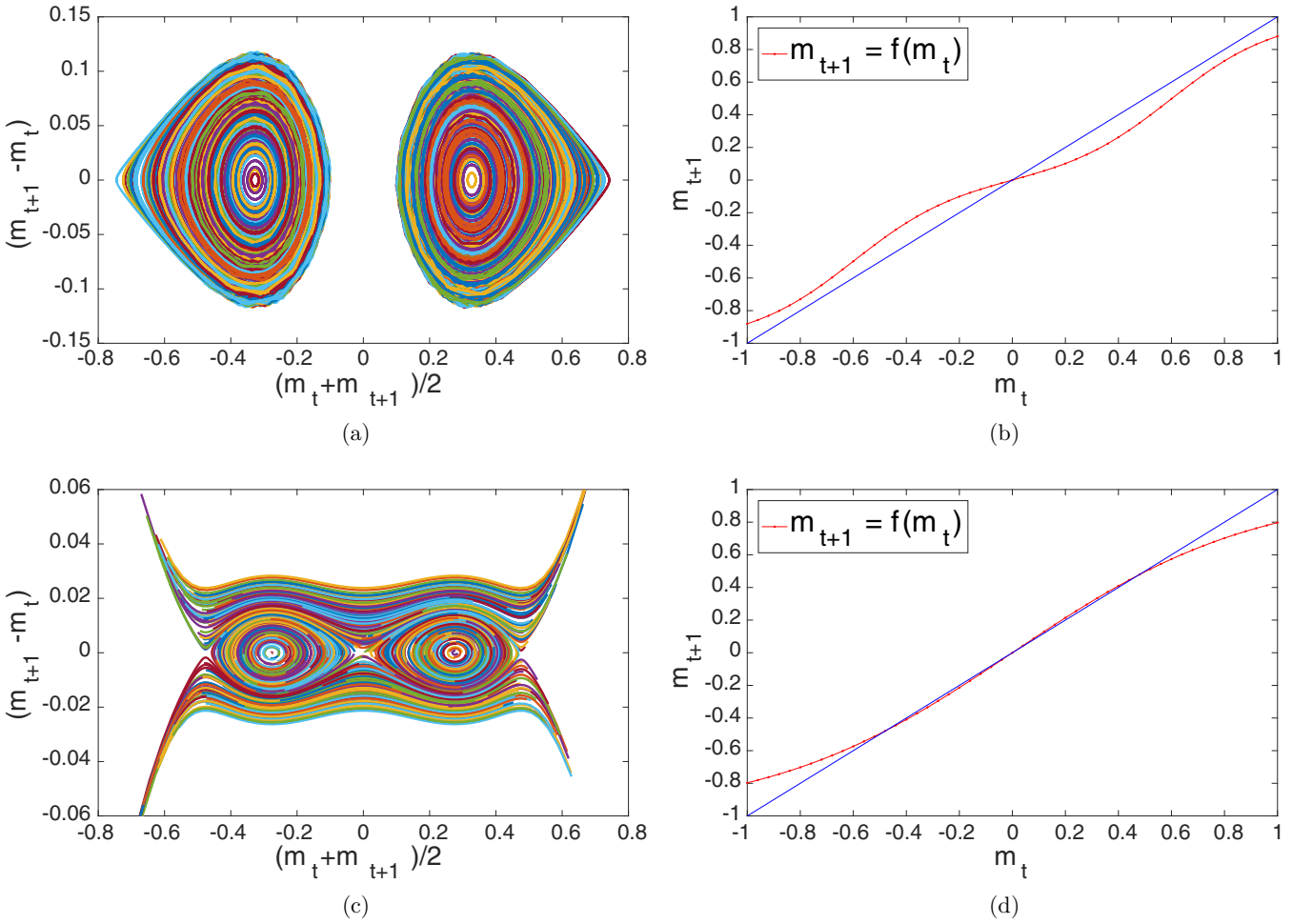


FIG. 3. Phase portraits of trajectories described by Eq. (35) [(a) and (c)] and maps of the relaxation dynamics $m_{t+1} = f(m_t)$ together with the diagonal $m_{t+1} = m_t$ [(b) and (d)] for the parameter settings: (a, b) $\beta = 2.5, h = 0.6$ (paramagnetic phase) (c, d) $\beta = 1.1, h = 0.1$ (ferromagnetic phase near a second-order phase transition).

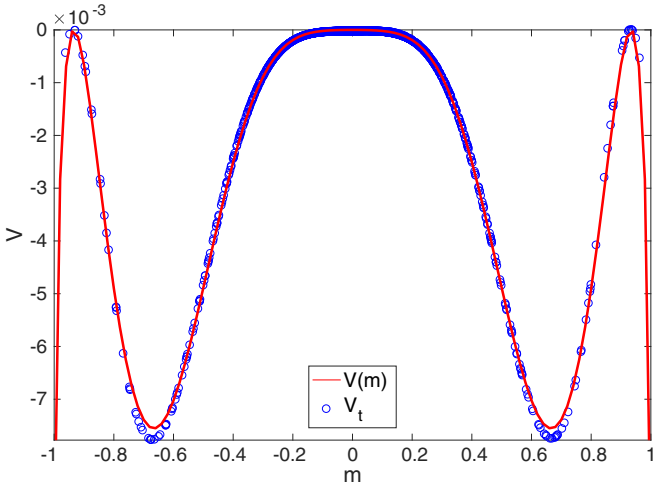


FIG. 4. Position-dependent potential (solid line) and path-dependent potential (circles) with $T = 20$ for the magnetic regime at $\beta = 2.5, h = 0.4$, and $r_0 = 0$.

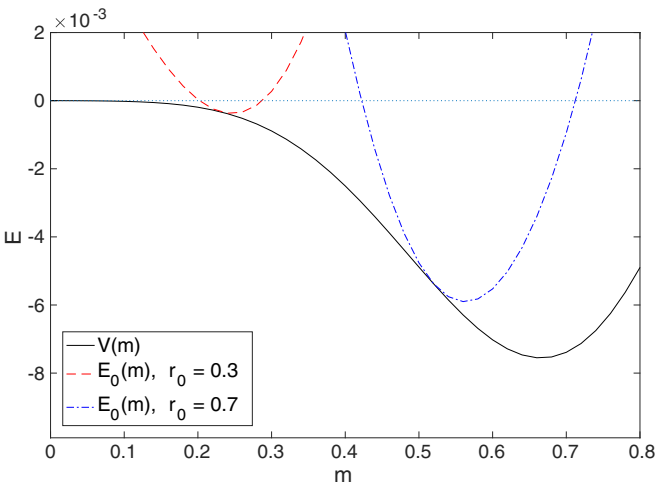


FIG. 5. Position-dependent potential (solid line) and initial total energy for $r_0 = 0.3$ (dashed) and $r_0 = 0.7$ (dash-dotted) for the ferromagnetic regime at $\beta = 2.5$ and $h = 0.4$. The level of the highest potential peak is fixed at 0.

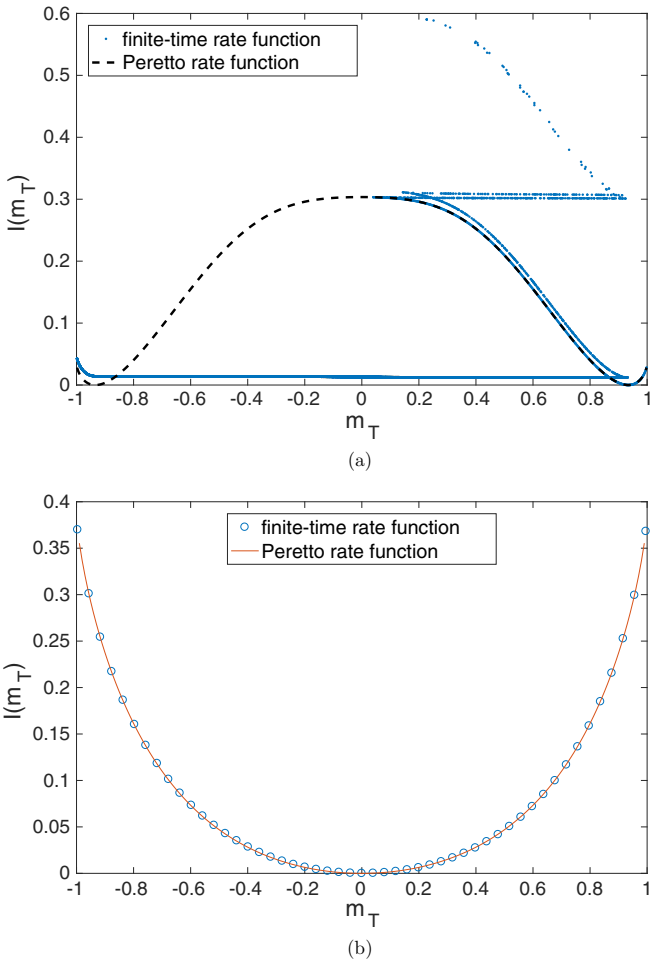


FIG. 6. (a) Peretto rate function (dashed) and finite-time rate function ($T = 50$, dots) for the ferromagnetic phase at $\beta = 2.5$, $h = 0.4$, $r_0 = 0.3$. (b) Peretto rate function (solid line) and finite-time rate function ($T = 20$, circles) for the paramagnetic phase at $\beta = 1$, $h = 0.4$, $r_0 = 0.3$.

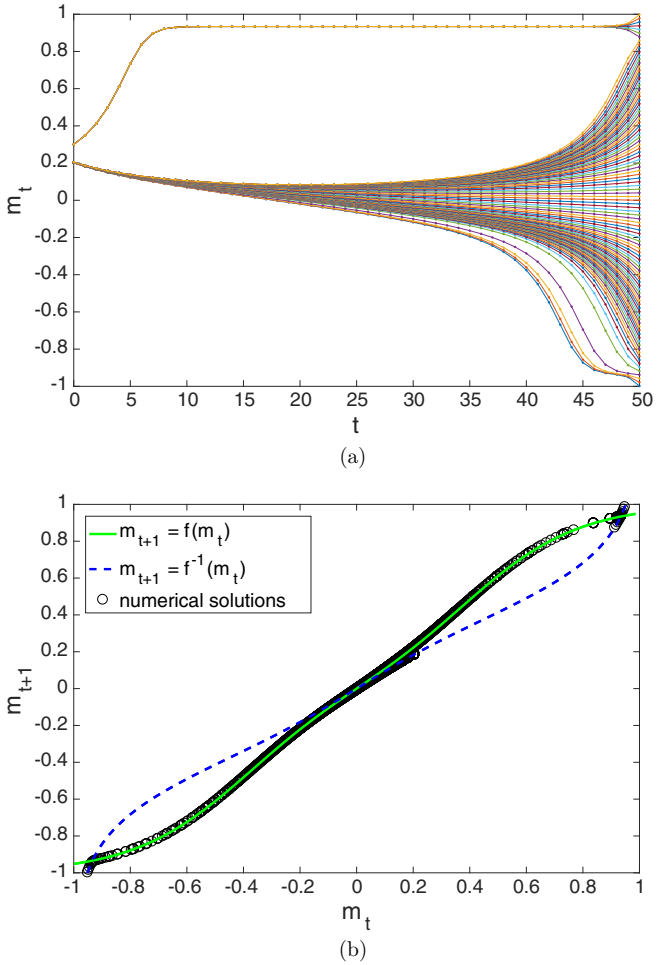


FIG. 7. (a) Dominant trajectories as a function of time for $\beta = 2.5$, $h = 0.4$, $r_0 = 0.3$. (b) dominant trajectories (circles) and limit curves $m_{t+1} = f(m_t)$ (solid line) and $m_{t+1} = f^{-1}(m_t)$ (dashed line) for $\beta = 2.5$, $h = 0.4$.

A natural step is then to characterize the solutions which give rise to these lowest values. We plot these trajectories in Fig. 7(a).

The time-dependent magnetization of the Curie-Weiss model obeys the first-order dynamics $m_{t+1} = f(m_t)$. We know from Eq. (41) that the contribution of this dynamics to the ω function vanish except for initial terms, and we thus expect the dominant trajectories to follow this dynamics. In practice, the situation is more complex: as can be seen in Fig. 7(b) where we plot m_{t+1} as a function of m_t for the ω-minimizing trajectories, the dominant trajectories seem to follow both the forward dynamics $m_{t+1} = f(m_t)$ and backward dynamics $m_{t+1} = f^{-1}(m_t)$.

From Fig. 7, we see several cases:

(i) if m_T is close to the stable point of the forward dynamics $m_{\rm eq} \simeq 0.94$, the trajectories initially follow the relaxation dynamics $m_{t+1} = f(m_t)$ then veer away from $m_{\rm eq}$ using backward dynamics.

(ii) if m_T is sufficiently far away from $m_{\rm eq}$, the trajectories instead initially follow the backward dynamics $m_{t+1} = f^{-1}(m_t)$, converging toward $m = 0$, then veer away from $m = 0$ using forward dynamics.

(iii) if $m_T < -m_{\rm eq}$, the trajectories initially follow backward dynamics toward $m = 0$, then away from $m = 0$ and toward $m = -m_{\rm eq}$ using forward dynamics, then away from $m = -m_{\rm eq}$ using backward dynamics again.

The first type of trajectories start near $m_0 = r_0 = 0.3$, while the second and third start near $m_0 = 0.2$. Both starting points can be seen from Fig. 5 as the points where the initial energy E_0 is approximately equal to the maximum value of the potential energy peak at $m = 0$. This remains true for different values of β and h.

This can be understood from the form of the ω function in Eq. (41): if a trajectory $\mathbf{m}$ satisfies $m_{t+1} = f(m_t)$, the sum in Eq. (41) becomes telescoping, leading to $\omega(\mathbf{m}|m_0,m) = 0$ for such a trajectory. A similar thing happens if $m_{t+1} = f^{-1}(m_t)$, but with a nonzero value of $\omega(\mathbf{m}|m_0,m)$ given by

$$\begin{aligned}&\omega_{\rm backward}(m_0,m)\\ &\quad = F[f^{-1}(m_0)] + F(m_0) - \beta m_0 f^{-1}(m_0)\\ &\qquad - \{F[f^{-1}(m)] + F(m) - \beta m f^{-1}(m)\}. \end{aligned} \tag{54}$$

Thus, the backward and forward dynamics seem to minimize the cost in terms of ω: partially in the case of backward dynamics, and fully in the case of forward dynamics. Note that while ω_{backward} is not obviously positive and indeed can be negative for some pairs (m_0,m_T), it is positive whenever m_T is reached from m_0 via backward dynamics.

The necessity of backward dynamics stems from the condition that the endpoint of a trajectory be m_T, which in general cannot be satisfied without switching: looking at the first group of trajectories in Fig. 7(a), trajectories which follow forward dynamics consistently reach a stable fixed-point in about 10 time steps. Therefore, trajectories that switch dynamics at least once are favored.

The location of the switch depends on the length of the time horizon: the longer the time horizon, the closer the switching location is to the top of the potential energy peak. This can be understood by considering a very large time horizon T: the trajectories must reach $m_T = m$ in T time steps, which means they must either oscillate around a potential energy minimum (and thus switch dynamics repeatedly, which incurs a large ω cost) or spend a large amount of time in transit. The second option is only possible if the trajectory spends a large amount of time reaching close to a potential energy peak, while having an initial energy very close to the maximum value of the potential energy. Thus, the initial energy of the trajectory must be close to the value of the potential energy peak, and the switching must happen close to the location of the peak. Note that in cases where there are several potential energy peaks of equal height, this means there are two possibilities for the initial magnetization, as in Fig. 5. We see these two possibilities clearly in Fig. 7(a).

Incidentally, this explains the large flat section of the rate function in Fig. 6(a): as we see in Fig. 7(a), the initial conditions are concentrated around the two values $m_0 = 0.2$ and $m_0 = 0.3$, with the largely flat section (as seen in Fig. 10, there is a slight decrease at the crossing of $m = 0$) arising from those trajectories which start near $m = 0.2$. These trajectories start at $m_0 \simeq 0.2$ at a cost $\omega_0(m_0)$, progress to the $m = 0$ peak through backward dynamics at a cost $\omega_{\text{backward}}(m_0,0)$, then proceed through forward dynamics (which does not incur a cost in ω) to their endpoint. The only change in the value of ω occurs for those trajectories which end at $m_T < -m_{\text{eq}}$, as these trajectories must switch trajectories once more from forward to backward at $m = -m_{\text{eq}}$.

B. Quasi-Peretto trajectories

In addition to the dominant trajectories, we find there are a class of trajectories solutions of the equations of motion, Eqs. (17)–(21) whose associated rate function are very close to the Peretto rate function even when the dominant trajectories do not. We call them quasi-Peretto trajectories, and find that they correspond to the first case in the discussion above. We plot them below in Fig. 8 for $\beta = 2.5, h = 0.4, r_0 = 0.3$.

C. Comparisons with simulations

We run $N_s = 10^8$ simulations with $N = 10^5$ and compare the results with analytic predictions in Fig. 9. Since we find no difference between the finite-time rate function and the Peretto result for $\beta < 1$, $T = 20$, we focus on the ferromagnetic part of the phase diagram and run simulations for $\beta = 2.5, h = 0.4$, where the finite-time rate function and the Peretto result are highly dissimilar even as late as $T = 150$. The value of the rate function is, for a large range of values of m_T, of the order of magnitude of N^{-1}. We therefore include first-order corrections to the analytically derived rate function as discussed in Appendix C (note that in order for these first-order corrections to be exact, we must have $\frac{N_+}{N} = \frac{1}{2} + o(N^{-1})$ instead of $\frac{N_+}{N} = \frac{1}{2} + O(N^{-1/2})$). We find excellent agreement. The associated Peretto rate function is plotted in Fig. 8(a) and is several orders of magnitude larger than the finite-time result.

D. Connection with Langevin dynamics and Freidlin-Wentzell theory

The behavior we discuss for dominant trajectories is common in Freidlin-Wentzell theory with conservative potential, where it is well-known that the trajectories realizing the minimum of the action are those that follow both forward and backward dynamics ([22–24]). Indeed, the above discussion is a natural adaptation of Freidlin-Wentzell theory in discrete time (a general discussion of the extension of Freidlin-Wentzell theory to Markov chains can be found in Ref. [25]): starting with Eq. (3), we see that the magnetization obtained at each time step t can be decomposed into the "deterministic" magnetization that would be obtained in the limit of infinite system size, $m_t^*(m_{t-1})$ and a random error term δm_t, dependent on m_{t-1}, which if N is large is Gaussian distributed via the central limit theorem, with a variance scaling like N.

More precisely, from Eq. (3) the magnetization at time t is given by

$$m_t = \frac{1}{N} 2(b_+(m_{t-1}) + b_-(m_{t-1})) - 1, \tag{55}$$

where $b_\pm$ are independent binomial random variables with $\frac{N}{2}$ trials and success probability

$$p_\pm(m_{t-1}) = \frac{e^{\beta(m_{t-1}\pm h)}}{2\cosh(\beta(m_{t-1} \pm h))}. \tag{56}$$

In the limit of large N these binomial distribution are approximately Gaussian, and thus m_t has variance $\sigma = N^{-1/2}\sigma(m_{t-1})$, with

$$\sigma^2(m_{t-1}) = 2(p_+(m_{t-1})(1 - p_+(m_{t-1})) + p_-(m_{t-1})(1 - p_-(m_{t-1}))). \tag{57}$$

Thus,

$$m_t \simeq f(m_{t-1}) + N^{-1/2}\sigma(m_{t-1})x_t, \tag{58}$$

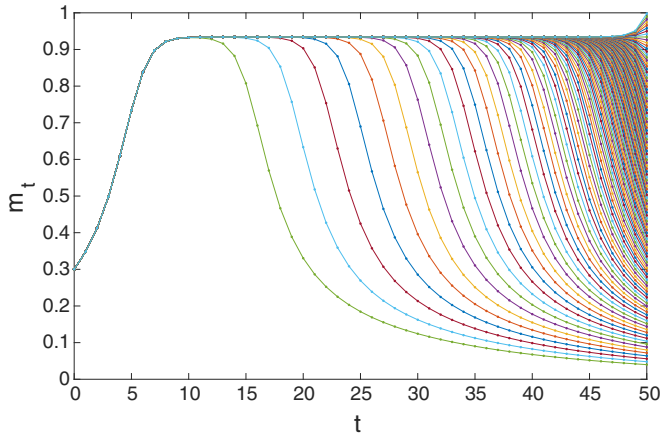
where $x_t \sim \mathcal{N}(0,1)$. We can rewrite this as

$$m_t - m_{t-1} \simeq f(m_{t-1}) - m_{t-1} + N^{-1/2}\sigma(m_{t-1})x_t, \tag{59}$$

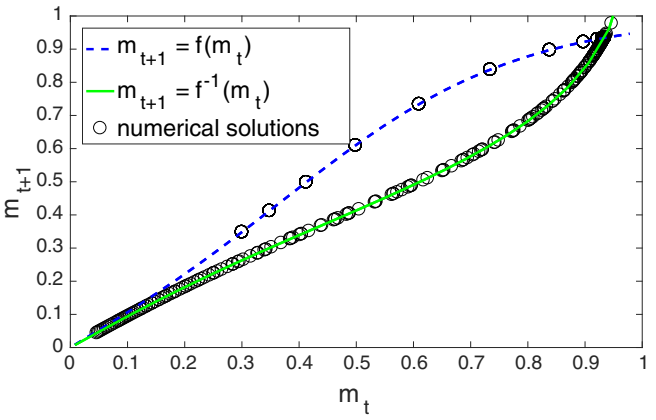
i.e., the dynamics follow a discrete analog of overdamped Langevin dynamics in an inhomogeneous medium.

VI. ATYPICAL DISORDER

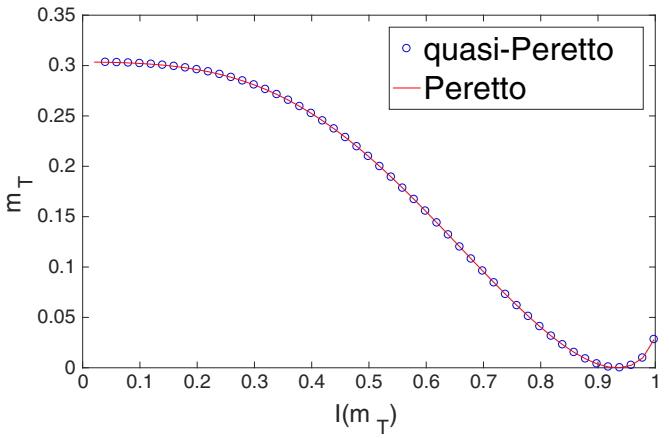
We have so far worked with a configuration of random fields with a fixed fraction p (which we have taken equal to $\frac{1}{2}$) of



(a)



(b)



(c)

FIG. 8. Quasi-Peretto solutions for $\beta = 2.5$, $h = 0.4$, $r_0 = 0.3$ (a) as a function of time. (b) in the (m_t, m_{t+1}) plane (circles) along with limit curves $m_{t+1} = f(m_t)$ (dashed line) and $m_{t+1} = f^{-1}(m_t)$ (solid line) for $\beta = 2.5, h = 0.4$. (c) Rate function associated with the quasi-Peretto trajectories (circles) and Peretto rate function (solid line).

spins having random field $\theta = +1$. To model situations where the magnetization is measured repeatedly from a fixed sample,

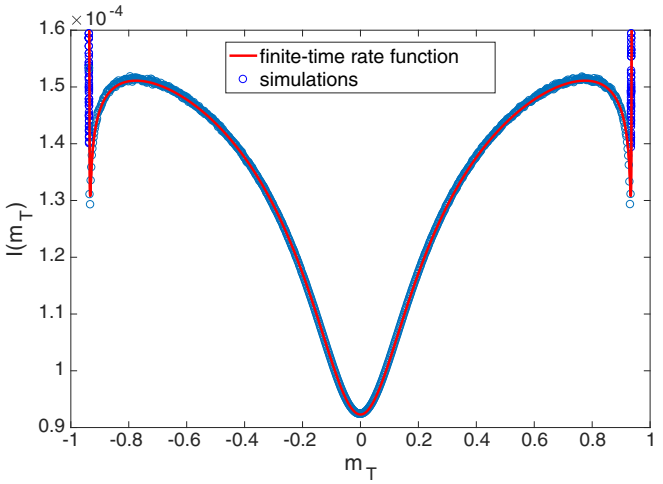


FIG. 9. Finite-time ($T = 50$) theoretical (solid red line) and empirical (circles) rate functions for $\beta = 2.5$, $h = 0.4$, $r_0 = 0.0$.

this is appropriate as there is no uncertainty in the sample: faced with a distribution different from the one predicted, we would simply have to adjust the value of p.

If we do not have access to repeated measurement on a fixed sample however, we must take into account the uncertainty of the random fields configuration. We can account for situations where the configuration is not known exactly, but its distribution is known, with the same formalism. Consider the case where the number of spins with $\theta_i = +1$ follows a binomial distribution with parameter p. Then,

$$P(m_T) \propto \sum_n \binom{N}{n} p^n (1-p)^{N-n} \int \left[\prod_{t=0}^{T-1} dm_t \, d\hat{m}_t \right] d\hat{m}_T \times \exp\left\{ -N \sum_{t=0}^{T} i\hat{m}_t m_t - N\beta \sum_{t=1}^{T} m_{t-1} m_t + \sum_{t=0}^{T} [n \log(Z_t(+1)) + (N-n)\log(Z_t(-1))] \right\}, \tag{60}$$

where $P(m_T)$ is now the probability density function of the final magnetization m_T *prior* to any measurements on the sample, i.e., knowing only the distribution of configurations. Alternatively, it is the probability distribution function of the final magnetization when the sample is changed at every measurement.

Performing the sum over n creates an annealed version of the previous problem, where we must replace $\langle \log Z_t(\theta)\rangle_\theta$ by $\log \langle Z_t(\theta)\rangle_\theta$. This creates an additional coupling between the variables $i\hat{m}_t$ and m_t which obscures the Newtonian structure of the problem discussed previously. Instead, we can maintain the Newtonian structure by performing the same δ-trick as previously and writing

$$1 = \frac{N}{2\pi} \int dq\, d\hat{q}\, e^{-i\hat{q}(Nq-n)}. \tag{61}$$

Thus,

$$\begin{aligned} P(m_T) \propto \int \left[\prod_{t=0}^{T-1} dm_t\, d\hat{m}_t\right] d\hat{m}_T\, dq\, d\hat{q} \\ \times \exp\left\{-N\sum_{t=0}^{T} i\hat{m}_t m_t - N\beta \sum_{t=1}^{T} m_{t-1}m_t \right. \\ + N\sum_{t=0}^{T}[q\log(Z_t(+1)) + (1-q)\log(Z_t(-1))] \\ - iN\hat{q}(q-p) + N\log[(1-p)e^{-i\hat{q}p} \\ \left. + pe^{i\hat{q}(1-p)}]\right\}. \end{aligned} \tag{62}$$

The reasoning then proceeds as before with two changes:

(1) the functions F and f are replaced with $F(\cdot|q)$ and $f(\cdot|q)$, where

$$\begin{aligned} F(m|q) = q\log(2\cosh[\beta(m+h)]) \\ + (1-q)\log(2\cosh[\beta(m-h)]), \end{aligned} \tag{63}$$

$$f(m|q) = q\tanh[\beta(m+h)] + (1-q)\tanh[\beta(m-h)], \tag{64}$$

(2) the original saddle-point equations (with f and F replaced by their $F(\cdot|q)$ and $f(\cdot|q)$ equivalents) are supplemented with the saddle-point equations for the q and $\hat{q}$ variables:

$$\begin{aligned} i\hat{q} &= \partial_q \sum_{t=0}^{T}[q\log(Z_t(+1)) - (1-q)\log(Z_t(-1))] \\ &= \sum_{t=0}^{T}[\log(Z_t(+1)) - \log(Z_t(-1))], \end{aligned} \tag{65}$$

$$q = \frac{pe^{i\hat{q}}}{pe^{i\hat{q}} + (1-p)}. \tag{66}$$

Equation (66) can be interpreted as giving the best estimate of the *actual* configuration of the sample given only a single measurement of the magnetization with result m_T.

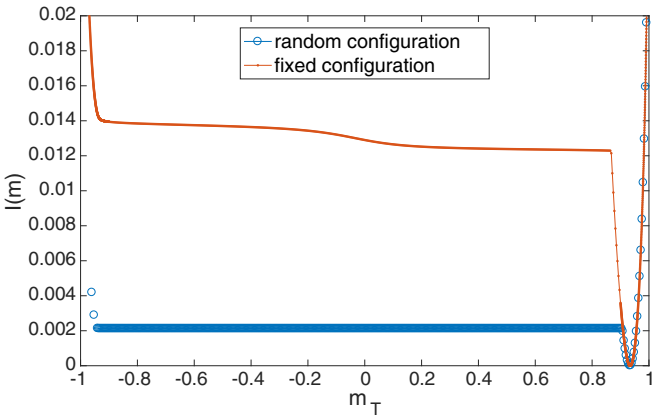


FIG. 10. Rate functions for $\beta = 2.5$, $h = 0.4$, $r_0 = 0.3$, and $T = 50$ with atypical disorder (circles) and without (red dotted line).

We expect the presence of atypical disorder to decrease the value of the rate function away from the equilibrium magnetization (which does not change), and this is confirmed by numerical results in Fig. 10.

VII. CONCLUSION

In this article, we have studied the large deviations of the magnetization at finite times in a simple model with quenched randomness. The advantage of such simplicity is that a number of findings—some quite unexpected—could be studied in great analytic detail. Among the unexpected results, we mention in particular the emergence of a multiplicity of metastable trajectories, and of second-order conservative dynamics: we find that while the dynamics of the RFIM are first order in time, the dynamics of the relevant trajectories for large deviations are second-order and follow simple equations of motion reminiscent of Newton's second law. In particular, they obey a form of energy conservation. Moreover, we find that the most likely trajectories with prescribed final magnetization switch between two types of first-order dynamics: the relaxation dynamics of the Curie-Weiss RFIM, $m_{t+1} = f(m_t)$, and their time-reversed counterpart $m_{t+1} = f^{-1}(m_t)$. The order and time of switching is dependent on the location of temperature and field strength parameters in the RFIM phase diagram, and the initial and terminal conditions. We study these equations numerically, obtain the relevant trajectories and compute the associated rate functions. We observe that the problem allows a multiplicity of (meta-)stable solutions, in contrast to simple relaxation dynamics. We find very good agreement with simulations. Extension of the formalism to a situation where the quenched-disorder configuration is uncertain is straightforward and largely preserves these features.

This mapping of the dynamics of the magnetization to that of a particle obeying some form of second-order dynamics in a multi-well potential is well-known in the study of large deviations of, e.g., first-passage times, as in Freidlin-Wentzell theory. The switching of dynamics between forward and backwards has been observed in other systems as well ([22,24,26]). To the best of our knowledge, it has not been studied in the RFIM before, despite its shedding light on many surprising and counter-intuitive aspects of large deviations.

In particular, it would be of interest to see if the switching phenomena appears in irreversible models such as studied in Ref. [15], where the time-reversed dynamics $m_{t+1} = f^{-1}(m_t)$ cannot exist, or in the study of observables which are not local in time (e.g., average magnetization or average activity).

ACKNOWLEDGMENTS

This work has been supported by the People Programme (Marie Curie Actions) of the European Unions Seventh Framework Programme Grant No. FP7/2007-2013/ and under REA Grant Agreement No. 290038(C.F.).

APPENDIX A: EQUILIBRIUM RATE FUNCTIONS

1. RFIM Hamiltonian

We can derive the equilibrium distribution assuming a Hamiltonian,

$$\mathcal{H}(\boldsymbol{\sigma}) = -N\frac{m^2(\boldsymbol{\sigma})}{2} - h\sum_i \sigma_i\theta_i, \tag{A1}$$

as appears in Ref. [19].

The induced distribution on the magnetization m is given by

$$P(m) = \left\langle \delta\left(m - \frac{1}{N}\sum_i \sigma_i\right)\right\rangle \tag{A2}$$

$$= \frac{1}{Z}\sum_{\{\sigma_i\}} e^{\beta(\frac{1}{2N}(\sum_i \sigma_i)^2 + \sum_i \sigma_i\theta_i)}\delta\left(m - \frac{1}{N}\sum_i \sigma_i\right) \tag{A3}$$

$$= \frac{1}{Z}\sum_{\{\sigma_i\}} \exp\left\{\beta\left(N\frac{m^2}{2} + \sum_i \sigma_i\theta_i\right)\right\} \times \int d\hat{m}\, e^{-i\hat{m}(Nm - \sum_i \sigma_i)} \tag{A4}$$

$$= \frac{1}{Z}\int d\hat{m}\, \exp\left\{-Ni\hat{m}m + N\beta\frac{m^2}{2} + N\langle \log 2\cosh[\beta\theta + i\hat{m}]\rangle_\theta\right\}. \tag{A5}$$

Taking the saddle-point yields equations for real $i\hat{m}$, which we write μ, and we obtain

$$I^0(m) = \sup_\mu\{\mu m - \langle \log\cosh[\mu + \beta\theta]\rangle_\theta\} - \beta\frac{m^2}{2}, \tag{A6}$$

which would be the rate function, except we are missing the partition function Z. However, Z is a constant (relative to m), hence it can be recovered by noticing that the minimum of the rate function is 0; hence,

$$I(m) = I^0(m) - \min_x\{I^0(x)\}. \tag{A7}$$

2. Equilibrium distribution

The previous approach, however, has the shortcoming of assuming a Hamiltonian, whereas we are more interested in obtaining the Hamiltonian from the dynamics. This has been done by Peretto (1984). We assume parallel dynamics with one-step transition probabilities given by

$$W(\boldsymbol{\sigma}'|\boldsymbol{\sigma}) = \prod_i \frac{\exp\left\{\sigma_i'\left[\beta h\theta_i + \frac{1}{N}\sum_j \beta\sigma_j\right]\right\}}{2\cosh\left[\beta\left(h\theta_i + \frac{1}{N}\sum_j \sigma_j\right)\right]}. \tag{A8}$$

This transition probability satisfies detailed balance with an equilibrium distribution $p(\boldsymbol{\sigma})$:

$$\frac{p(\boldsymbol{\sigma})}{p(\boldsymbol{\sigma}')} = \frac{W(\boldsymbol{\sigma}|\boldsymbol{\sigma}')}{W(\boldsymbol{\sigma}'|\boldsymbol{\sigma})} = \prod_i \frac{\frac{\exp\left\{\sigma_i\left[\beta h\theta_i + \frac{1}{N}\sum_j \beta\sigma_j'\right]\right\}}{2\cosh\left[\beta\left(h\theta_i + \frac{1}{N}\sum_j \sigma_j'\right)\right]}}{\frac{\exp\left\{\sigma_i'\left[\beta h\theta_i + \frac{1}{N}\sum_j \beta\sigma_j\right]\right\}}{2\cosh\left[\beta\left(h\theta_i + \frac{1}{N}\sum_j \sigma_j\right)\right]}} \tag{A9}$$

$$= \prod_i \frac{\cosh\left[\beta\left(h\theta_i + \frac{1}{N}\sum_j \sigma_j\right)\right]\exp\{\sigma_i\beta h\theta_i\}}{\cosh\left[\beta\left(h\theta_i + \frac{1}{N}\sum_j \sigma_j'\right)\right]\exp\{\sigma_i'\beta h\theta_i\}}, \tag{A10}$$

giving naturally

$$p(\boldsymbol{\sigma}) = \frac{1}{Z}\exp\left\{\sum_i \log\left[\cosh\left(\beta[h\theta_i + m]\right)\right] + \sigma_i\beta h\theta_i\right\}. \tag{A11}$$

We then work out the distribution on m induced by this distribution on $\boldsymbol{\sigma}$ in the usual way:

$$P(m) = \left\langle \delta\left(m - \frac{1}{N}\sum_i \sigma_i\right)\right\rangle \tag{A12}$$

$$= \frac{1}{Z}\sum_{\{\sigma_i\}} \exp\left\{\sum_i \log\left[\cosh\left(\beta(h\theta_i + m)\right)\right] + \sigma_i\beta h\theta_i\right\} \times \delta\left(m - \frac{1}{N}\sum_i \sigma_i\right) \tag{A13}$$

$$= \frac{1}{Z}\sum_{\{\sigma_i\}} \exp\left\{\sum_i \log\left[\cosh\left(\beta(h\theta_i + m)\right)\right] + \sigma_i\beta h\theta_i\right\} \times \int \frac{d\hat{m}}{2\pi/N} e^{-i\hat{m}(Nm - \sum_i \sigma_i)} \tag{A14}$$

$$= \frac{1}{Z}\int d\hat{m}\, \frac{N}{2\pi}\exp N\{-i\hat{m}m + \langle \log[\cosh(\beta[h\theta + m])]\rangle_\theta + \langle \log[\cosh(\beta h\theta + i\hat{m})]\rangle_\theta\}, \tag{A15}$$

and we see that the end result is

$$I(m) = \sup_x\{xm - \langle \log[\cosh(\beta h\theta + x)]\rangle_\theta\} - \langle \log[\cosh(\beta[h\theta + m])]\rangle_\theta + \frac{1}{N}\log[Z], \tag{A16}$$

and we see again that the log Z constant only serves to ensure that $\inf_m\{I(m)\} = 0$.

APPENDIX B: INVERSES

We make extensive use of the inverses $f^{-1}(x)$ and $f_0^{-1}(x)$, and we derive them here.

The inverse of $f_0(x) = \tanh(\rho + \beta x)$ presents no difficulty:

$$f_0^{-1}(x) = \beta^{-1}(\tanh^{-1}(x) - \rho). \tag{B1}$$

The inverse of $f(x) = \langle \tanh[\beta(x + h\theta)]\rangle_\theta$ is a bit more complicated:

$$f(x) = \langle \tanh[\beta(x + h\theta)]\rangle_\theta \tag{B2}$$

$$= \tfrac{1}{2}(\tanh[\beta(x + h)] + \tanh[\beta(x - h)]), \tag{B3}$$

making use of the addition formula $\tanh(x + y) = \frac{\tanh(x)+\tanh(y)}{1+\tanh(x)\tanh(y)}$,

$$f(x) = \frac{1}{2}\left(\frac{\tanh[\beta x] + \tanh[\beta h]}{1 + \tanh[\beta x]\tanh[\beta h]} + \frac{\tanh[\beta x] - \tanh[\beta h]}{1 - \tanh[\beta x]\tanh[\beta h]}\right), \tag{B4}$$

we write $t_y = \tanh(\beta y)$ for simplicity:

$$f(x) = \frac{1}{2}\left(\frac{t_x + t_h}{1 + t_x t_h} + \frac{t_x - t_h}{1 - t_x t_h}\right) \tag{B5}$$

$$= \frac{1}{2}\frac{(t_x + t_h)(1 - t_x t_h) + (t_x - t_h)(1 + t_x t_h)}{1 - t_x^2 t_h^2} \tag{B6}$$

$$= \frac{1}{2}\frac{(t_x + t_h) - t_x^2 t_h - t_x t_h^2 + t_x - t_h + t_x^2 t_h - t_x t_h^2}{1 - t_x^2 t_h^2} \tag{B7}$$

$$= \frac{t_x\left(1 - t_h^2\right)}{1 - t_x^2 t_h^2}. \tag{B8}$$

Thus,

$$t_x^2 t_h^2 f(x) + t_x\left(1 - t_h^2\right) - f(x) = 0, \tag{B9}$$

and

$$t_x = -\frac{\left(1 - t_h^2\right)}{2t_h^2 f(x)} \pm \sqrt{\left(\frac{1 - t_h^2}{2t_h^2 f(x)}\right)^2 + \frac{1}{t_h^2}}. \tag{B10}$$

For the choice of branch, we consider $\beta h \ll 1$:

$$t_x = -\frac{1}{2(\beta h)^2 f(x)} \pm \sqrt{\left(\frac{1}{2(\beta h)^2 f(x)}\right)^2 + \frac{1}{(\beta h)^2}} \tag{B11}$$

$$= \frac{1}{(\beta h)^2}\left(-\frac{1}{2f(x)} \pm \frac{1}{2|f(x)|}\right), \tag{B12}$$

and since the result must be finite, we must have

$$-\frac{1}{2f(x)} \pm \frac{1}{2|f(x)|} = 0, \tag{B13}$$

i.e., we select the $+$ branch if $f(x) > 0$ and the $-$ branch if $f(x) < 0$:

$$f^{-1}(x) = \beta^{-1}\tanh^{-1}\left(-\frac{\left(1 - t_h^2\right)}{2t_h^2 x} + \operatorname{sign}(x)\sqrt{\left(\frac{1 - t_h^2}{2t_h^2 x}\right)^2 + \frac{1}{t_h^2}}\right) \tag{B14}$$

$$= \beta^{-1}\operatorname{sign}(x)\tanh^{-1}\left(\frac{2\cosh(\beta h)^2|x|}{1 + \sqrt{1 + (x\sinh(2\beta h))^2}}\right). \tag{B15}$$

APPENDIX C: FIRST-ORDER CORRECTIONS

From the path integral formulation, we can rewrite Eq. (15) as

$$P(m_T) \propto \int\left[\prod_{t=0}^{T-1} d\delta m_t\, d\delta\hat{m}_t\right] d\delta\hat{m}_T \times \exp\{-N\Omega(\mathbf{m}^* + \delta\mathbf{m}, \hat{\mathbf{m}}^* + \delta\hat{\mathbf{m}})\}, \tag{C1}$$

where the starred quantities represent the saddle-point values and $\delta m_T = 0$. Expanding to second order, we have

$$P(m_T) \propto \int\left[\prod_{t=0}^{T-1} d\delta m_t\, d\delta\hat{m}_t\right] d\delta\hat{m}_T \times \exp\{-N\Omega(\mathbf{m}^*, \hat{\mathbf{m}}^*, m, \hat{m}^*) \tag{C2}$$

$$-\frac{N}{2}(\delta\mathbf{m}\,\delta\hat{\mathbf{m}})\partial^2\Omega(\mathbf{m}^*, \hat{\mathbf{m}}^*)\begin{pmatrix}\delta\mathbf{m}\\ \delta\hat{\mathbf{m}}\end{pmatrix}\Big\}, \tag{C3}$$

i.e., we have a Gaussian colored noise on top of the Newtonian path $(\mathbf{m}^*, \hat{\mathbf{m}}^*)$. This noise scales like $N^{-1/2}$, hence taking the $N \to \infty$ limit is equivalent to taking a zero-noise limit of the noisy (Langevin) dynamics.

This suggests an obvious correction to the numerical results: the $\partial^2\Omega$ matrix has size $2T + 1$ (as m_T is a parameter rather than an integration variable); hence, for low values of T it can be diagonalized easily. The first-order corrections can be computed once the relevant saddle-point trajectories have been obtained:

$$\frac{\delta I(m)}{N} = \frac{1}{2}\sum_{i=1}^{2T+1}\log|\lambda_i|, \tag{C4}$$

where the λ_i are the eigenvalues of the $\partial^2\Omega$ matrix, which for a trajectory $\mathbf{m}$ is given by if $t = 0$,

$$\partial^2_{m_0 m_{t'}}\Omega = \delta_{t',0}\beta f'(m_0) - \beta\delta_{t',1}, \tag{C5}$$

$$\partial^2_{\hat{m}_0\hat{m}_{t'}}\Omega = \delta_{t',0}\beta^{-1} f_0'\left(f_0^{-1}(m_0)\right), \tag{C6}$$

$$\partial^2_{m_0\hat{m}_{t'}}\Omega = i\delta_{t',0}, \tag{C7}$$

if $t \geqslant 1$,

$$\partial^2_{m_t m_{t'}}\Omega = \delta_{t,t'}\beta f'(m_t) - \beta(\delta_{t',t+1} + \delta_{t',t-1}), \tag{C8}$$

$$\partial^2_{\hat{m}_t\hat{m}_{t'}}\Omega = \delta_{t,t'}\beta^{-1} f'(f^{-1}(m_t)), \tag{C9}$$

$$\partial^2_{m_t\hat{m}_{t'}}\Omega = i\delta_{t,t'}. \tag{C10}$$

We note that any constant (in m) factors can be obtained simply by computing the probability distribution $P(m) = \exp\{-N(I(m) + \delta I)\}$ and requiring normalization.

[1] H. Touchette, Phys. Rep. **478**, 1 (2009).
[2] G. Biroli, J.-P. Bouchaud, and M. Potters, J. Stat. Mech.: Theory Exp. (2007) P07019.
[3] J. L. Lebowitz and H. Spohn, J. Stat. Phys. **95**, 333 (1999).
[4] F. Morone, G. Parisi, and F. Ricci-Tersenghi, Phys. Rev. B **89**, 214202 (2014).
[5] G. Parisi and T. Rizzo, Phys. Rev. Lett. **101**, 117205 (2008).
[6] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, J. Phys. A: Math. Theoret. **42**, 075007 (2009).
[7] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland, Phys. Rev. Lett. **98**, 195702 (2007).
[8] T. Bodineau and C. Toninelli, Commun. Math. Phys. **311**, 357 (2012).
[9] C. De Bacco, A. Guggiola, R. Kühn, and P. Paga, arXiv:1506.08436 (2015).
[10] P. Vivo, S. N. Majumdar, and O. Bohigas, J. Phys. A: Math. Theoret. **40**, 4317 (2007).
[11] E. Katzav and I. Perez Castillo, Phys. Rev. E **82**, 040104 (2010).
[12] A. Greven and F. den Hollander, Ann. Probab. **22**, 1381 (1994).
[13] F. Altarelli, A. Braunstein, L. Dall'Asta, and R. Zecchina, Phys. Rev. E **87**, 062115 (2013).
[14] C. Klüppelberg and T. Mikosch, J. Appl. Probabil. **34**, 293 (1997).
[15] P. Paga and R. Kühn, J. Stat. Mech.: Theory Exp. (2015) P03008.
[16] R. L. Jack and P. Sollich, Prog. Theor. Phys. Suppl. **184**, 304 (2010).
[17] E. S. Loscar, A. S. Mey, and J. P. Garrahan, J. Stat. Mech.: Theory Exp. (2011) P12011.
[18] J. A. de Matos and J. F. Perez, J. Stat. Phys. **62**, 587 (1991).
[19] M. Löwe, R. Meiners, and F. Torres, J. Phys. A: Math. Theoret. **46**, 125004 (2013).
[20] P. Peretto, Biol. Cybernet. **50**, 51 (1984).
[21] N. Skantzos and A. Coolen, J. Phys. A: Math. Gen. **33**, 1841 (2000).
[22] A. Ventsel' and M. I. Freidlin, Russian Math. Surveys **25**, 1 (1970).
[23] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Vol. 260 (Springer Science & Business Media, Berlin, 2012).
[24] F. Bouchet and J. Reygner, J. Ann. Henri Poincaré **17**, 3499 (2016).
[25] E. Olivieri and M. E. Vares, *Large Deviations and Metastability*, Vol. 100 (Cambridge University Press, Cambridge, 2005).
[26] R. Nicole and P. Sollich (unpublished).