

Information-theoretic analysis of complex systems modeled by two dimensional pairwise Ising models

Alexandre L. M. Levada

Received: date / Accepted: date

Abstract Stochastic complex systems are composed by a large number of seemingly simple variables that exhibit non-linear interactions with each other, causing the emergence of complexity and non-deterministic dynamics in the edge between order and chaos. Hence, the evolution of these systems seems to be completely out of control, with unpredictable behaviors. In this paper, using information geometry as a mathematical approach to chaos and complexity, we investigate how information theory can be used to analyze the dynamics of pairwise Ising random fields along Markov Chain Monte Carlo simulations in which phase transitions are observed. Our experiments indicate that Fisher information regarding the inverse temperature parameter can bring important insights, since it signalizes changes in the global spatial dependence structure. Information-theoretic curves are built to show that, despite the random nature of the system, it is possible to identify an asymmetric pattern of evolution when the system moves towards different entropic states.

Keywords Ising model · Markov Chain Monte Carlo · Fisher information · Complex systems · Information geometry

1 Introduction

In the last decades, there has been a great interest in complex system science, whose goal is to understand how the interaction between different parts of a system give rise to collective behaviors [7, 23]. A feature of complex systems is related to the propagation of information through local rules in a way that the whole is controlled by a dynamics that cannot be fully understood by analyzing only its parts [30]. Besides, it has been observed that complexity emerges at

Alexandre L. M. Levada
Computing Department, Federal University of São Carlos
Tel.: +55-16-33518232
E-mail: alexandre.levada@ufscar.br

the edge between order and chaos [34]. Being able to capture patterns in a seemingly random behavior is the key for understanding the difference between chaos and randomness. The relevance of the study and characterization of complex systems dynamics resides in the fact that if we understand it, we have powerful insights regarding how to gain control over several natural phenomena and non-deterministic processes [25].

Several mathematical models for the study and characterization of complex systems can be found in the literature: cellular automata [15], whose deterministic rules cause the emergence of complexity through the definition of a transition function, complex networks [32], in which the connection of nodes cause the emergence of non-trivial topological features and random fields [10], which are particularly suitable for modeling non-deterministic phenomena where non-linear interactions between random variables lead to the emergence of long range correlations and phase transitions [33]. Many examples of the application of random field models arise in physics, especially in statistical mechanics and thermodynamics [19].

In complex systems dynamics, critical phenomena occur in critical states. In these states, the system is extremely susceptible to small perturbations. Phase transitions play an important role in complex systems. Usually, a phase transition is related to a change in the degree of order in the system and spontaneous symmetry breaking [8]. Being able to predict whether a complex system is approaching phase transitions is relevant to many process in physics, biology and economics. In this context, one question that is raised by many researchers is: can we predict if a system is moving towards a phase transition? Or, in other words, can we predict an imminent order/disorder phase transition using information flow dynamics? In this paper, we propose an information-theoretic approach for the analysis of pairwise Ising model dynamics using Fisher information regarding the inverse temperature parameter.

Our motivation for using information-theoretic measures to analyze complex systems dynamics is information geometry as a mathematical tool for the study of intrinsic geometric properties of parametric spaces associated with random variables [4,6]. The main goal is to build computational tools in order to predict the emergence of complex behavior in non-deterministic dynamical systems. At the beginning of the dynamics, the inverse temperature parameter is zero and we have a collection of independent Bernoulli random variables. However, when the inverse temperature increases, the underlying parametric space suffers a series of geometric deformations. Fisher information is capable of bringing insights about this complex process, since it is directly related to the metric tensor of the parametric space, being the most important ingredient to the computation of geodesic distances [1,17]. Statistical geometrodynamics have been proposed as a theoretical information geometric framework suitable to characterize chaotic dynamical behavior of arbitrary complex systems on curved statistical manifolds [14,11]. Both information geometry and statistical geometrodynamics have been explored in the derivation of measures of chaos and complexity in physical systems [12,13,17].

In summary, the main contributions of this paper are: 1) to be a proof of concept that statistical geometrodynamics and information geometry theories can be employed to perform complex systems analysis in practice, through a feasible computational application using the pairwise Ising model; 2) to propose an original equation to estimate the inverse temperature in the pairwise Ising model through maximum pseudo-likelihood estimation, avoiding the computation of the partition function in the joint Gibbs distribution; 3) to derive expressions for the observed Fisher information regarding the inverse temperature parameter in the pairwise Ising model; 4) to propose an information-theoretic curve to analyze the pairwise Ising model dynamics, characterizing the asymmetric behavior of the system when visiting different entropic states, which induces the emergence of the hysteresis phenomenon [26], often associated to systems that exhibit phase transitions. It has been verified the existence of attractors in the Ising model in terms of physical properties (i.e., magnetization) in regular lattices [5] and Cayley trees [20]. However, in this work, we derive a geometric based framework, build in terms of information-theoretic measures related to statistical divergences. Hysteresis has also been observed in Ising model dynamics in terms of physical properties [2,31], showing that information-theoretic analysis is capable of reproducing the same effects, but with a pure geometric interpretation.

The remaining of the paper is organized as follows: in Section 2, we define the pairwise Ising model and derive an original equation for the maximum pseudo-likelihood estimator of the inverse temperature parameter in this model. In Section 3, we discuss Fisher information, some properties and we derive expressions for the information-theoretic measures. Section 4 shows the computational simulations using Markov Chain Monte Carlo (MCMC) algorithms, the obtained information curves for complex systems modeled by pairwise Ising model dynamics. Finally, Section 5 presents the conclusions and final remarks.

2 The Ising model

Consider a 2D lattice composed by n sites, $S = \{s_1, s_2, \dots, s_n\}$, and a neighborhood system η_i that defines an adjacency relation between s_i and its Δ nearest neighbors. In this study, we assume a second-order neighborhood system, that is, $\Delta = 8$, defining the Moore neighborhood. In each site s_i we observe a random variable x_i that assume values in $\{-1, +1\}$, representing the spin of a particle [28]. A realization $X = (x_k)_{k \in S}$ is an assignment of a spin value to each s_i , for $i = 1, 2, \dots, n$. For any pair of adjacent sites $s_i, s_j \in S$ there is an interaction J_{ij} . Besides, a site s_i have an external magnetic field h_i interacting with it. The energy of the global configuration of spins X (Hamiltonian), is given by [16]:

$$U(X) = - \sum_{(i,j)} J_{ij} x_i x_j - \sum_i h_i x_i \quad (1)$$

In the absence of an external field and assuming that all nearest neighbors have the same interaction strength, we have:

$$U(X) = -J \sum_{(i,j)} x_i x_j \quad (2)$$

It is possible to assign a probability to the global configuration pattern X through the Boltzmann distribution:

$$P(X) = \frac{1}{Z} \exp \left\{ -\frac{1}{k_B T} U(X) \right\} \quad (3)$$

where k_B is the Boltzmann constant, T is the temperature and Z is the partition function, defined by the summation:

$$Z = \sum_X \exp \left\{ -\frac{1}{k_B T} U(X) \right\} \quad (4)$$

It is worth mentioning that the computation of the partition function requires a sum over all possible realizations of X , which is prohibitive for large fields. Making the inverse temperature parameter $\beta = 1/k_B T$ and considering a ferromagnetic model ($J = 1$), we can write:

$$P(X) = \frac{1}{Z} \exp \left\{ \beta \sum_{(i,j)} x_i x_j \right\} \quad (5)$$

Note that the summation over all bonds (i, j) for a given site s_i is precisely the difference between the number of neighboring sites having $x_j = x_i$, denoted by $U(x_i)$, and the number of neighboring sites having $x_j \neq x_i$, given by $\Delta - U(x_i)$. Hence, we can express $P(X)$ as:

$$P(X) = \frac{1}{Z} \exp \left\{ \beta \sum_i [U(x_i) - (\Delta - U(x_i))] \right\} = \frac{1}{Z} \exp \left\{ \beta \sum_i (2U(x_i) - \Delta) \right\} \quad (6)$$

Denoting $U'(x_i) = 2U(x_i) - \Delta$, we have:

$$P(X) = \frac{1}{Z} \exp \left\{ \beta \sum_i U'(x_i) \right\} \quad (7)$$

According to the Hammersley-Clifford theorem [21], which state the equivalence between Gibbs random fields (global models) and Markov random fields (local models), the pairwise Ising model can be defined by the set of local conditional density functions $p(x_i|\eta_i, \beta)$, for $i = 1, 2, \dots, n$, as:

$$p(x_i|\eta_i, \beta) = \frac{e^{\beta U'(x_i)}}{\sum_{k \in \{-1, +1\}} e^{\beta U'(k)}} = \frac{e^{\beta U'(x_i)}}{e^{\beta U'(-1)} + e^{\beta U'(+1)}} \quad (8)$$

The advantage of the local conditional density functions is that we do not have to deal with the partition function, making inference on the pairwise Ising model mathematically tractable and computationally viable. Note that when the inverse temperature parameter is zero, we have a sample of n independent Bernoulli random variables with $p = 0.5$.

2.1 Maximum pseudo-likelihood estimation of the inverse temperature parameter

An important problem in the proposed information-theoretic analysis of the dynamics of the Ising model is the estimation of the inverse temperature parameter of the system, given an outcome of the model. Maximum likelihood estimation is not tractable due to the partition function in the joint Gibbs distribution. An alternative is to use maximum pseudo-likelihood estimation by means of the local conditional density functions, which are combined to build the pseudo-likelihood function [9]. The pseudo-likelihood function for the Ising model is:

$$L(X; \beta) = \prod_{i=1}^n p(x_i | \eta_i, \beta) = \prod_{i=1}^n \frac{e^{\beta U'(x_i)}}{\sum_{k \in \{-1, +1\}} e^{\beta U'(k)}} \quad (9)$$

Usually, it is simpler to maximize the logarithm of the pseudo-likelihood function:

$$\begin{aligned} \ell(\beta) = \log L(X; \beta) &= \sum_{i=1}^n \log p(x_i | \eta_i, \beta) \\ &= \sum_{i=1}^n \left[\beta U'(x_i) - \log \sum_{k \in \{-1, +1\}} e^{\beta U'(k)} \right] \end{aligned} \quad (10)$$

Differentiating with respect to the inverse temperature parameter, we have:

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^n U'(x_i) - \sum_{i=1}^n \left[\frac{\sum_{k \in \{-1, +1\}} U'(k) e^{\beta U'(k)}}{\sum_{k \in \{-1, +1\}} e^{\beta U'(k)}} \right] = 0 \quad (11)$$

Note that from a statistical mechanics perspective, the solution for this equation provides the value of the inverse temperature for which the free energy (first term) is equal to the expected energy (second term). To solve the pseudo-likelihood equation, we propose an algorithm based on the identification of all different local contextual patterns that provide different contributions to the equation. Note that in a 2D Ising model defined on a second-order neighborhood system ($\Delta = 8$), there are only five patterns that offer different contributions, as illustrated by Figure 1 (the integer partitions of 8 into two categories). We now identify what are the contributions to the second term of

the pseudo-likelihood equation given by each one of the possible local configuration patterns. Note that for the five patterns, [8, 0], [7, 1], [6, 2], [5, 3], and [4, 4], we have:

+1	+1	+1	-1	-1	-1	[8, 0]
+1		+1	-1		-1	
+1	+1	+1	-1	-1	-1	
+1	+1	+1	-1	-1	-1	[7, 1]
+1		+1	-1		-1	
+1	+1	-1	-1	-1	+1	
+1	+1	+1	-1	-1	-1	[6, 2]
+1		+1	-1		-1	
+1	-1	-1	-1	+1	+1	
+1	+1	+1	-1	-1	-1	[5, 3]
+1		+1	-1		-1	
-1	-1	-1	+1	+1	+1	
+1	+1	+1	-1	-1	-1	[4, 4]
-1		+1	+1		-1	
-1	-1	-1	+1	+1	+1	

Fig. 1 Different local configurations patterns for the Ising model defined on a second-order neighborhood system. There are only five patterns that offer different contributions to the pseudo-likelihood equation for the Ising model on a 2D lattice with $\Delta = 8$.

$$\begin{aligned}
 K_{[8,0]} &= 8 \frac{(e^{8\beta} - e^{-8\beta})}{(e^{8\beta} + e^{-8\beta})} & K_{[7,1]} &= 6 \frac{(e^{6\beta} - e^{-6\beta})}{(e^{6\beta} + e^{-6\beta})} \\
 K_{[6,2]} &= 4 \frac{(e^{4\beta} - e^{-4\beta})}{(e^{4\beta} + e^{-4\beta})} & K_{[5,3]} &= 2 \frac{(e^{2\beta} - e^{-2\beta})}{(e^{2\beta} + e^{-2\beta})} \\
 K_{[4,4]} &= 0 \frac{(e^{0\beta} - e^{-0\beta})}{(e^{0\beta} + e^{-0\beta})} = 0
 \end{aligned} \tag{12}$$

Knowing that the hyperbolic tangent function is defined by:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{13}$$

and denoting by $N_{[i,j]}$ be the number of $[i, j]$ local configuration patterns along the random field, we obtain the maximum pseudo-likelihood estimator for the inverse temperature parameter is by solving the following transcendental equation:

$$8 \tanh(8\beta) \times N_{[8,0]} + 6 \tanh(6\beta) \times N_{[7,1]} + 4 \tanh(4\beta) \times N_{[6,2]} + 2 \tanh(2\beta) \times N_{[5,3]} = \sum_{i=1}^n U'(x_i) \quad (14)$$

In our simulations, we build a histogram of the pattern configurations at each iteration of the dynamics and numerically solve the equation in β to estimate the inverse temperature.

3 Information-theoretic measures

Information geometry provides the connection between geometry and statistics in a formal and elegant way. The key ingredient that unifies these two seemingly unrelated areas is the concept of Fisher information. In the following, we discuss what is Fisher information and how we can compute it in the pairwise Ising model.

3.1 Fisher information

Let $p(X; \theta)$ be the likelihood function (probability density as a function of θ), where θ is the parameter of the model. The Fisher information, which is related to the metric tensor of the parametric space, is defined by:

$$I(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \log p(X; \theta) \right)^2 \right] \quad (15)$$

It is known that Fisher information is the KL-divergence (relative entropy) in the limiting case. It has been shown that Fisher information has an important role in defining geodesic distances between different statistical models, since it defines the metric tensor of the parametric space [29].

3.2 The information equality

It is known that for several statistical models the information equality holds. For purposes of simplification we consider the uniparametric case, knowing that the extension to multiparametric models is quite natural. Let X be a random variable with a probability density function $p(X; \theta)$. Note that:

$$\frac{\partial^2}{\partial\theta^2} \log p(X; \theta) = \frac{\partial}{\partial\theta} \left[\frac{1}{p(X; \theta)} \frac{\partial}{\partial\theta} p(X; \theta) \right] \quad (16)$$

By the product rule, we have:

$$\frac{\partial}{\partial\theta} \left[\frac{1}{p(X; \theta)} \frac{\partial}{\partial\theta} p(X; \theta) \right] = -\frac{1}{p(X; \theta)^2} \left[\frac{\partial}{\partial\theta} p(X; \theta) \right]^2 + \frac{1}{p(X; \theta)} \frac{\partial^2}{\partial\theta^2} p(X; \theta) \quad (17)$$

which, by applying the expectation operator and rearranging the terms, leads to:

$$E \left[\left(\frac{\partial}{\partial\theta} \log p(X; \theta) \right)^2 \right] = -E \left[\frac{\partial^2}{\partial\theta^2} \log p(X; \theta) \right] + E \left[\frac{1}{p(X; \theta)} \frac{\partial^2}{\partial\theta^2} p(X; \theta) \right] \quad (18)$$

By the definition of expected value, the last term of the previous expression can be rewritten as:

$$E \left[\left(\frac{\partial}{\partial\theta} \log p(X; \theta) \right)^2 \right] = -E \left[\frac{\partial^2}{\partial\theta^2} \log p(X; \theta) \right] + \int \frac{\partial^2}{\partial\theta^2} p(X; \theta) dx \quad (19)$$

Under certain regularity conditions, it is possible to differentiate under the integral sign by interchanging differentiation and integration operators, which implies in:

$$\int \frac{\partial^2}{\partial\theta^2} p(X; \theta) dx = \frac{\partial^2}{\partial\theta^2} \int p(X; \theta) dx = \frac{\partial^2}{\partial\theta^2} 1 = 0 \quad (20)$$

leading to the information equality condition.

3.3 Fisher information and curvature of the relative entropy

Fisher information is also directly related to the relative entropy, or KL-divergence [18]. It has been shown that Fisher information is, in fact, the curvature of the relative entropy. Consider two distributions $p(x; \theta)$ and $p(x; \theta')$. Note that the KL-divergence is given by:

$$D_{KL}(\theta||\theta') = \int p(x; \theta) [\log p(x; \theta) - \log p(x; \theta')] \quad (21)$$

The second derivative of the relative entropy at the point $\theta' = \theta$ is:

$$\begin{aligned}
\left[\frac{\partial^2}{\partial \theta'^2} D_{KL}(\theta || \theta') \right]_{\theta'=\theta} &= - \frac{\partial}{\partial \theta'} \left[\int p(x; \theta) \frac{\partial}{\partial \theta'} \log p(x; \theta') dx \right]_{\theta'=\theta} \\
&= - \left[\int p(x; \theta) \frac{\partial^2}{\partial \theta_i'^2} \log p(x; \theta') dx \right]_{\theta'=\theta} \\
&= -E \left[\frac{\partial^2}{\partial \theta^2} \log p(x; \theta) \right] = I(\theta) \tag{22}
\end{aligned}$$

indicating that Fisher information encodes information about the curvature of the relative entropy. Another important connection between these two information-theoretic measures is that Fisher information is the KL-divergence in the limiting case, that is, when we have two nearby densities $p(x; \vec{\theta})$ and $p(x; \vec{\theta} + \Delta \vec{\theta})$. It has been shown that the KL-divergence between two infinitesimally close densities can be expressed by a quadratic form whose coefficients are given by the elements of the Fisher information matrix.

3.4 Fisher information as the KL-divergence in the limiting case

Another important connection between these two information-theoretic measures is that Fisher information is the KL-divergence in the limiting case, that is, when we have two nearby densities $p(x; \vec{\theta})$ and $p(x; \vec{\theta} + \Delta \vec{\theta})$. It has been shown that the KL-divergence between two infinitesimally close densities can be expressed by a quadratic form whose coefficients are given by the elements of the Fisher information matrix. First, recall that the symmetric KL-divergence between $p(x)$ and $q(x)$ is given by:

$$D_{KL}(p, q) = \int (p(x) - q(x)) \log \left(\frac{p(x)}{q(x)} \right) dx \tag{23}$$

Assuming we have a family of distributions parametrized by $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ and $\vec{\theta}' = (\theta_1 + \Delta \theta_1, \theta_2 + \Delta \theta_2, \dots, \theta_k + \Delta \theta_k)$ where $\Delta \theta_i$ is an infinitesimal displacement, then:

$$D_{KL}(p(x; \vec{\theta}), p(x; \vec{\theta}')) = \int [p(x; \vec{\theta}) - p(x; \vec{\theta}')] \log \left(\frac{p(x; \vec{\theta})}{p(x; \vec{\theta}')} \right) dx \tag{24}$$

By definition, let the variation $\Delta p(x; \vec{\theta})$ be:

$$\Delta p(x; \vec{\theta}) = p(x; \vec{\theta}) - p(x; \vec{\theta}') \tag{25}$$

which leads to:

$$D_{KL}(p(x; \vec{\theta}), p(x; \vec{\theta}')) = \int \frac{\Delta p(x; \vec{\theta})}{p(x; \vec{\theta})} \log \left(\frac{p(x; \vec{\theta})}{p(x; \vec{\theta}')} \right) p(x; \vec{\theta}) dx \tag{26}$$

Note that the argument of the logarithm can be expressed as:

$$\frac{p(x; \vec{\theta})}{p(x; \vec{\theta}')} = \frac{p(x; \vec{\theta})}{p(x; \vec{\theta}) - \Delta p(x; \vec{\theta})} = \frac{p(x; \vec{\theta}) + \Delta p(x; \vec{\theta})}{p(x; \vec{\theta})} = 1 + \frac{\Delta p(x; \vec{\theta})}{p(x; \vec{\theta})} \quad (27)$$

Using a Taylor approximation for the logarithm, for small values of x we have:

$$\log(1 + x) \approx x \quad (28)$$

Thus, we can write the following approximation:

$$\log \left(1 + \frac{\Delta p(x; \vec{\theta})}{p(x; \vec{\theta})} \right) \approx \frac{\Delta p(x; \vec{\theta})}{p(x; \vec{\theta})} \quad (29)$$

which leads to:

$$D_{KL} \left(p(x; \vec{\theta}), p(x; \vec{\theta}') \right) = \int \left(\frac{\Delta p(x; \vec{\theta})}{p(x; \vec{\theta})} \right)^2 p(x; \vec{\theta}) dx \quad (30)$$

Since $\Delta p(x; \vec{\theta})$ is the arc length in the parametric space (manifold), we can express it in the tangent space as the dot product:

$$\Delta p(x; \vec{\theta}) \approx \sum_{i=1}^k \left[\frac{\partial}{\partial \theta_i} p(x; \vec{\theta}) \Delta \theta_i \right] \quad (31)$$

Note that the above equation is the dot product between the gradient (local tangent coordinates) and the displacement vector $\Delta \vec{\theta}$. Hence, we can write:

$$\begin{aligned} \frac{\Delta p(x; \vec{\theta})}{p(x; \vec{\theta})} &\approx \frac{1}{p(x; \vec{\theta})} \sum_{i=1}^k \frac{\partial}{\partial \theta_i} p(x; \vec{\theta}) \Delta \theta_i \\ &= \sum_{i=1}^k \frac{\partial}{\partial \theta_i} \log p(x; \vec{\theta}) \Delta \theta_i \end{aligned} \quad (32)$$

Then, we can express equation (30) as:

$$D_{KL} \left(p(x; \vec{\theta}), p(x; \vec{\theta}') \right) = \sum_{i=1}^k \sum_{j=1}^k g_{ij} \Delta \theta_i \Delta \theta_j \quad (33)$$

where

$$\begin{aligned}
g_{ij} &= \int \left(\frac{\partial}{\partial \theta_i} \log p(x; \vec{\theta}) \right) \left(\frac{\partial}{\partial \theta_j} \log p(x; \vec{\theta}) \right) p(x; \vec{\theta}) dx \\
&= E \left[\left(\frac{\partial}{\partial \theta_i} \log p(x; \vec{\theta}) \right) \left(\frac{\partial}{\partial \theta_j} \log p(x; \vec{\theta}) \right) \right] \\
&= I(\vec{\theta})_{ij}
\end{aligned} \tag{34}$$

In the matrix vector notation, we have:

$$D_{KL} \left(p(x; \vec{\theta}), p(x; \vec{\theta}') \right) = \Delta \vec{\theta}^T I(\vec{\theta}) \Delta \vec{\theta} \tag{35}$$

where $I(\vec{\theta})$ is the Fisher information matrix.

3.5 Observed Fisher information

In some cases, the computation of the expected values in Fisher information can be a challenging task, or even unfeasible. To overcome this problem, we can compute the observed Fisher information:

$$I_o^{(1)}(\theta) = \left(\frac{\partial}{\partial \theta} \log L(X; \theta) \right)^2 \tag{36}$$

which can be estimated by:

$$I_o^{(1)}(\theta) \approx \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial}{\partial \theta} \log p(x_i; \theta) \right]^2 \tag{37}$$

for $\theta = \hat{\theta}$ being the maximum likelihood estimator of the parameter. Similarly, using the information equality condition, we have:

$$I_o^{(2)}(\theta) = - \frac{\partial^2}{\partial \theta^2} \log L(X; \theta) \tag{38}$$

which can be estimated by:

$$I_o^{(2)}(\theta) \approx - \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log p(x_i; \theta) \tag{39}$$

Furthermore, it has been shown that equations (36) and (39) are consistent estimators to the expected Fisher information, that is, asymptotically they converge to $I(\theta)$.

3.6 Fisher information and entropy in the Ising model

In the pairwise Ising model, we cannot derive a closed-form expression for the expected Fisher information, due to the impossibility to analytically compute the expected values involving the derivatives of the logarithm of the pseudo-likelihood function. However, with the local conditional density functions, we can derive a closed-form expression for the observed Fisher information in this model to replace the expected information. The derivative of the local conditional density function with respect to the inverse temperature parameter is:

$$\frac{\partial}{\partial \beta} \log p(x_i | \eta_i, \beta) = U'(x_i) - \frac{\sum_k U'(k) e^{\beta U'(k)}}{\sum_k e^{\beta U'(k)}} = \frac{\sum_k [(U'(x_i) - U'(k)) e^{\beta U'(k)}]}{\sum_k e^{\beta U'(k)}} \quad (40)$$

Then, we have the following expression for $I_o^{(1)}(\beta)$, from now on denoted by Φ_β (PHI):

$$\Phi_\beta = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\sum_k \sum_l [(U'(x_i) - U'(k))(U'(x_i) - U'(l)) e^{\beta(U'(k)+U'(l))}]}{\sum_k \sum_l e^{\beta(U'(k)+U'(l))}} \right\} \quad (41)$$

Similarly, we can compute the negative of the second derivative of the local conditional density functions to derive $I_o^{(2)}(\beta)$, from now on, denoted by Ψ_β (PSI), as:

$$\Psi_\beta = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\sum_k \sum_l [(U'(k)(U'(k) - U'(l))) e^{\beta(U'(k)+U'(l))}]}{\sum_k \sum_l e^{\beta(U'(k)+U'(l))}} \right\} \quad (42)$$

It is straightforward to check that a local configuration pattern can provide different contributions to Φ_β and Ψ_β . Moreover, the condition for information equality can be summarized as:

$$(U'(x_i) - U'(k))(U'(x_i) - U'(l)) = U'(k)(U'(k) - U'(l)) \quad (43)$$

for $k, l \in \{-1, +1\}$. Note that, for the configuration pattern $[4, 4]$, the right hand side of equation (43) becomes zero, showing that it does not provide any contribution to the Fisher information regarding the inverse temperature parameter, since there is no prevalence of positive or negative spins in the neighborhood.

3.7 Entropy in the pairwise Ising model

To compute the entropy in the Ising model using the local conditional density functions, we recall that it can be defined as the expected value of self-information, which leads to:

$$H_\beta = -E[\log p(x_i|\eta_i, \beta)] = \frac{1}{n} \sum_{i=1}^n \left[\log \sum_k e^{\beta U'(k)} - \beta U'(x_i) \right] \quad (44)$$

Note that for $\beta = 0$, we have $H(\beta) = \log(2) = 0.69315$, which means that the two states are equiprobable. Recall that in this case, we have a sample of independent Bernoulli random variables with $\theta = 0.5$. The probability density function of a Bernoulli random variable is:

$$p(x; \theta) = \theta^x (1 - \theta)^{1-x} \quad (45)$$

The entropy is given by:

$$H(x) = -E[\log p(x; \theta)] = -E[x \log \theta + (1 - x) \log(1 - \theta)]$$

which, for $\theta = 0.5$, becomes:

$$H(x) = -E[\log(0.5)] = \log(2) \quad (46)$$

perfectly matching the value obtained by the pairwise Ising model.

4 Computational simulations

In this study, we consider two different scenarios for the information-theoretic analysis of the Ising model dynamics, denoted by: 1) known inverse temperature; 2) unknown inverse temperature. In the first case, we assume that we have direct access to the real inverse temperature parameter in the computation of entropy and Fisher information. In other words, there is no uncertainty in the definition of this important parameter. On the other hand, in the blind analysis, we observe the system's outcomes and the unknown inverse temperature must be estimated from data using maximum pseudo-likelihood. To simulate the pairwise Ising model dynamics, we employ the Metropolis-Hastings algorithm, a popular Markov Chain Monte Carlo method used to generate outputs of random fields, given the parameter values [27, 22]. In every simulation, the initial condition is composed by a $n \times n$ lattice of Bernoulli(1/2) samples, which is the degenerate distribution for the pairwise Ising model when the inverse temperature is zero. During the first half of the simulation, after the k -th iteration, a positive small increment of $\Delta\beta$ is added to the inverse temperature until we reach a maximum value β_{MAX} . In the second half of the simulation, a

small negative increment $-\Delta\beta$ is added to the inverse temperature after each iteration, until we reach zero once again. In all experiments, a full cycle takes 600 iterations, with $n = 512$ and $\beta_{MAX} = 1.5$. Figure 2 shows some samples of the random field during the evolution of the system.

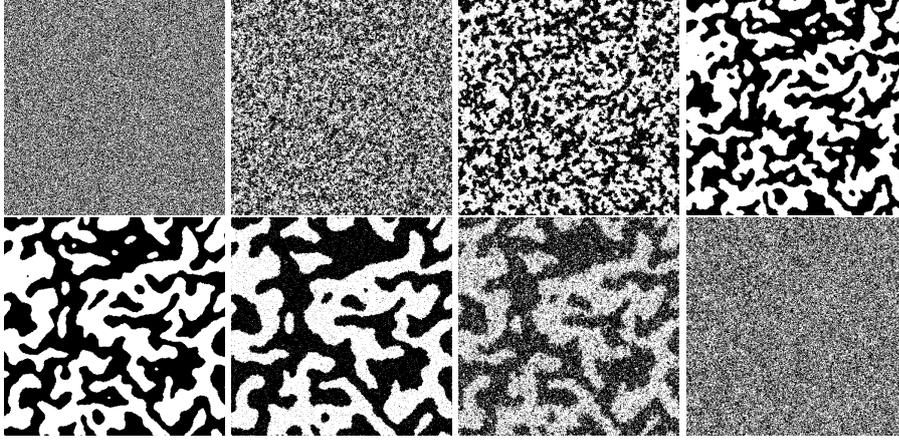


Fig. 2 Ising model dynamics along a Markov Chain Monte Carlo (MCMC) simulation. Evolution of the random field as the inverse temperature parameter β is first increased from zero to β_{MAX} and then decreased to zero again.

4.1 The information curves of the Ising model

The proposed information curve for the analysis of the pairwise Ising model dynamics is build by a parametric curve defined as the sequence of vectors $\vec{p}_\beta = (\Phi_\beta, \Psi_\beta, H_\beta)$ in the positive (first) octant of the three dimensional space (R^3), obtained by varying the inverse temperature parameter from an initial value A to a final value B and back. Let $\mathbf{X}^{(\beta_1)}, \mathbf{X}^{(\beta_2)}, \dots, \mathbf{X}^{(\beta_n)}$ be a sequence outcomes from the pairwise Ising model produced by a MCMC simulation for which $A = 0 = < \beta_1 < \beta_2 < \dots < \beta_n = B$. Then,

$$\vec{F}_A^B(\beta) = (\Phi_\beta, \Psi_\beta, H_\beta) \quad \beta = A, \dots, B \quad (47)$$

defines the information curve of the system from A to B . Therefore, we can define one cycle of the curve W as the concatenation of the information curves $\vec{F}_A^B(\beta)$ and $\vec{F}_B^A(\beta)$. In the following, we will see through several computational experiments, that the characteristic shape of this information curve for the Ising model dynamics, does not suffer significant changes even for large variations in the initial conditions.

4.2 Analysis of random field dynamics: known inverse temperature

In the first set of experiments, we consider that we have access to the real inverse temperature parameter, which is not a reasonable hypothesis in the analysis of real world systems. Figure 3 shows two different plots: 1) the variation of $\Phi(\beta)$ (blue curve) and $\Psi(\beta)$ (red curve) along the MCMC simulation; and 2) the variation of $H(\beta)$ (entropy) along the MCMC simulation.

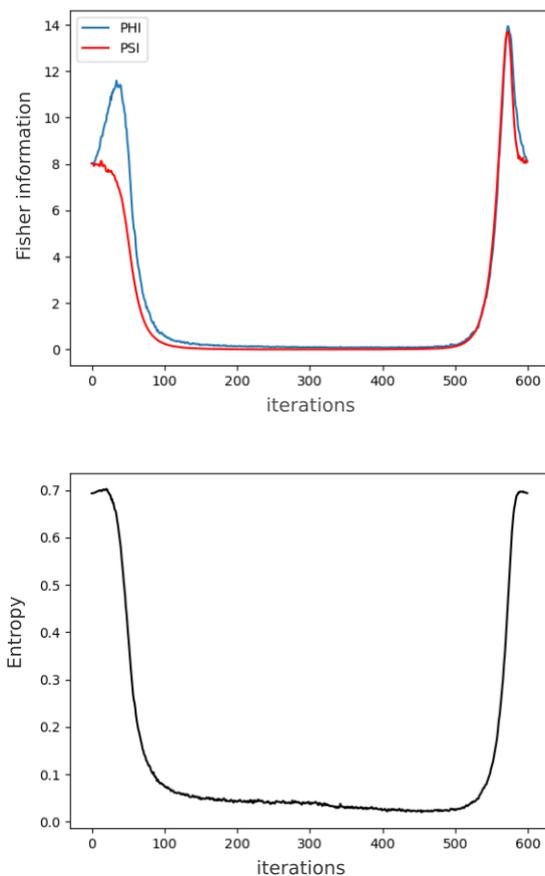


Fig. 3 Inverse temperature, Fisher information and entropy along Ising model dynamics field dynamics. In the Ising model dynamics, for some values of the inverse temperature parameter, the information equality does not hold.

Looking at the graph of $\Phi(\beta)$, the the first local maximum point seems to be related to a rapid decrease of the system's entropy. In fact, this point signalizes that a complex phase-transition is emerging. Figure 4 shows the

estimated information curves $\vec{F}_A^B(\beta)$ for $\beta = 0, \dots, 1.5$ (the blue curve) and $\vec{F}_B^A(\beta)$ for $\beta = 1.5, \dots, 0$ (the red curve) regarding the random field dynamics. Note that the model can be considered as a three dimensional parametric curve for the underlying complex dynamical system, since our simulations confirmed that for large variations in the initial conditions, the trajectory in the information space, which defines the shape of our trajectory, remains the same. In this context, these curves are similar to an attractor of a dynamical system.

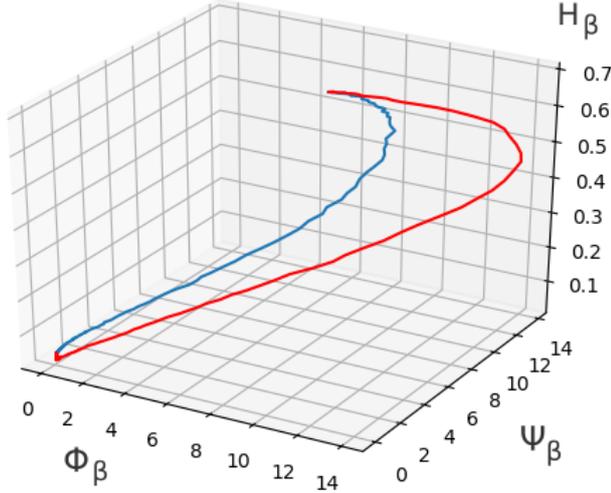


Fig. 4 Information curve of the random field dynamics knowing the true inverse temperature parameter. The 3D information curve was built by varying the inverse temperature parameter β from $\beta_{MIN} = 0$ (state A) to $\beta_{MAX} = 1.5$ (state B) and back. The results show that moving along different entropic states causes the emergence of a natural orientation in terms of information.

When the inverse temperature is zero, we are at the top of the blue curve and as we increase its value, we move along down to the state with minimum entropy. There is a natural orientation in the information curves along the evolution of the random field. In order to have a different visualization, Figure 5 illustrates that the use both forms of Fisher information defines a 2D curve for the Ising model dynamics.

4.3 Analysis of random field dynamics: unknown inverse temperature

In our second set of experiments, we assume a more realistic hypothesis: the inverse temperature parameter is not known a priori, that is, we have to estimate β for each one of the iterations of the MCMC simulation. The main

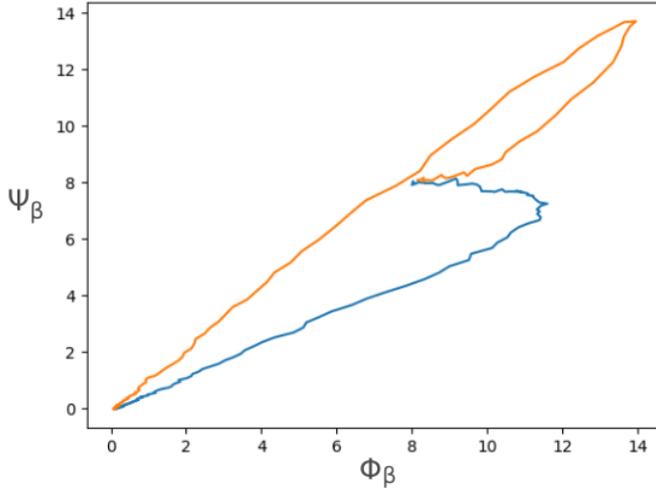


Fig. 5 2D information curve built by combining Fisher information forms $\Phi(\beta)$ and $\Psi(\beta)$. Note that the blue line refers to increasing inverse temperature and the red line refers to decreasing inverse temperature. The curve shows that the dynamics is highly asymmetrical in terms of variations in the spatial dependence structure.

difference with this scenario is that we have some degree of uncertainty regarding the inverse temperature parameter, which controls the spatial dependence structure of the random field. Figure 6 shows four different plots: 1) the variation of $\Phi(\beta)$ (blue curve) and $\Psi(\beta)$ (red curve) along the MCMC simulation; and 2) the variation of $H(\beta)$ (entropy) along the MCMC simulation.

Note that the maximum pseudo-likelihood estimator of the inverse temperature parameter, in average, underestimates the real value for large parameter values. However, the Fisher information and entropy curves does not suffer significant changes. It is worth mentioning that the two peaks in the Fisher information curves indicate that the system is undergoing phase-transitions. Figure 7 shows the 3D information curve and Figure 8 shows the 2D curve.

The analysis of the curves shows that from an information-theoretic perspective, the behavior of the system when it moves towards lower entropy states (β increases) is significantly different from the behavior of the system when it moves towards higher entropy states (β decreases). This effect is known as the hysteresis phenomenon [3] and basically suggests that the system has non-local memory [24] and undergoes a series of irreversible changes. In this sense, our model can be considered mathematical model of hysteresis. In summary, when temperature is infinite ($\beta = 0$) entropy reaches a maximum value ($\log 2$) and the information equality prevails. From an information-theoretic perspective, a reduction in temperature (increase in β) causes a series of changes in the parametric space, since its metric tensor (Fisher information)

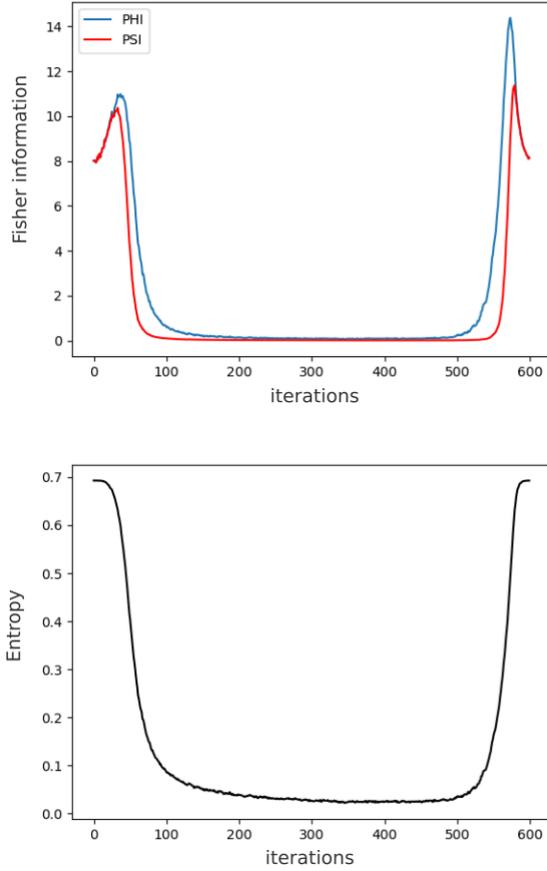


Fig. 6 Estimated inverse temperature, Fisher information and entropy along random field dynamics. The estimated inverse temperature has an upper bound, making the behavior of Fisher information quite different than in the case where the true inverse temperature is known.

is changed in an apparently asymmetric non-reversible way, inducing the emergence of a natural orientation of evolution.

Finally, to show how the shape of the proposed information-theoretic curve does not suffer significantly changes along different pairwise Ising model dynamics, we performed three complete cycles in which the inverse temperature is gradually increased up to the maximum value $\beta_{MAX} = 1.5$ and then slowly reduced to zero once again. Each cycle is composed by 600 iterations. Figures 9 and 10 show the obtained 3D and 2D information curves, where the first cycle is represented by the blue curve, the second cycle is represented by the red curve and the third cycle is represented by the green curve. Note that despite small perturbations, the characteristic shape of the curves is preserved in all

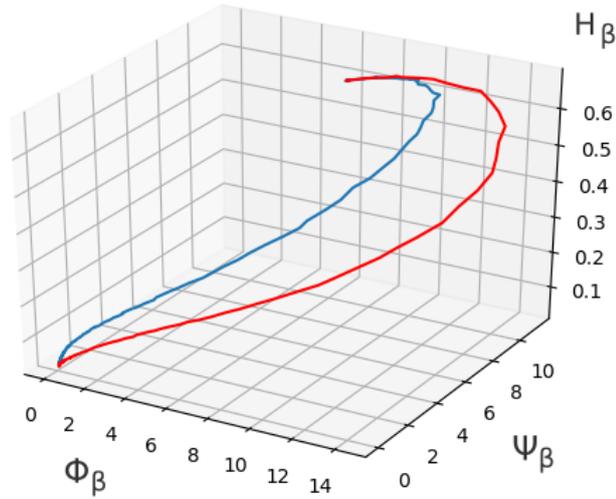


Fig. 7 Information curve of the random field dynamics. The 3D curve was built by varying the inverse temperature parameter β from $\beta_{MIN} = 0$ (state A) to β_{MAX} (state B) and back. The results show that moving along different entropic states causes the emergence of a natural orientation in terms of information.

cycles, indicating that Fisher information regarding the inverse temperature parameter is a robust measure in the analysis of the global system's behavior along the pairwise Ising model dynamics. Besides, it is clear the distinction between the two regimes of the system: the point denoted by A in Figure 10 denotes total lack of spatial dependence structure, that is, we have several independent Bernoulli random variables and a very noisy configuration in which entropy is maximum, while in B we have a high degree of spatial dependence, large clusters and a very smooth, low entropy configuration.

5 Conclusions

The Ising model is a mathematical structure that express complex behavior. In this work, we addressed the problem of analyzing the dynamics of a non-deterministic complex system modeled by pairwise Ising random field models by means of an information-geometric perspective. Investigations on the relation between entropy and Fisher information, led us to the definition of the information curve of the system. In summary, the information curve models the dynamics of random fields in a geometrical perspective, allowing us to predict whether the system is approaching a phase transition: large values of $\Phi(\beta)$ and $\Psi(\beta)$ indicate that the system's dynamics is abruptly changing. Computational experiments with different initial conditions and parameter settings lead to the same pattern in the information space. We have observed that, when the random field moves along different entropic states, its parametric space is

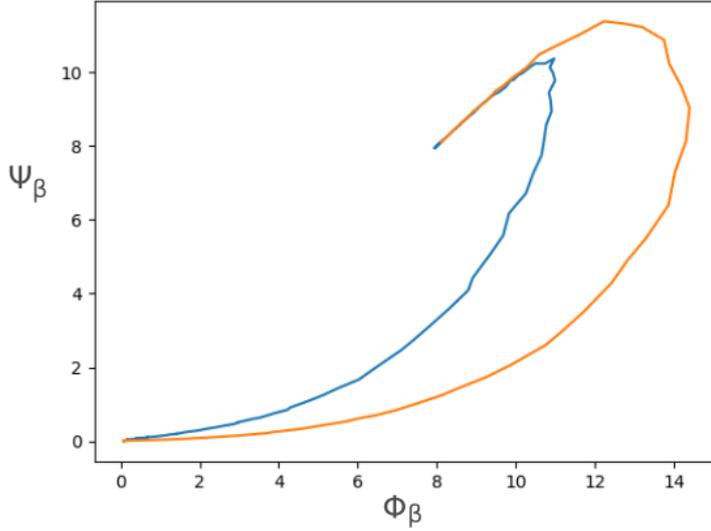


Fig. 8 2D information curve built by combining both forms of Fisher information. Note that when the inverse temperature is unknown, the shape of the curve is changed. However, in both cases, the phenomenon of hysteresis is observed.

actually being deformed by changes that happen in Fisher information. Our findings show that when temperature deviates from infinity, entropy decreases as a cause of variations in Fisher information. Moreover, moving towards lower entropy states is different from moving towards higher entropy states, since the information curves are not the same. This asymmetry induces a natural orientation to the information curve indicating that during the phase transitions our system has memory and experience irreversible changes. In this scenario, the proposed information-theoretic model can be considered a mathematical model of hysteresis in which the natural orientation is pointed by an arrow of time. One feature of the proposed model is that it does not rely on physical interpretations: we believe the proposed information-theoretic framework can be adapted to study the evolution of complex systems exhibiting local interactions along time, such as deterministic cellular automata models. Besides, with the proposed maximum pseudo-likelihood equation it is possible to estimate the unknown inverse temperature parameter in a computationally efficient way. Future works may include the study of other techniques for the estimation of the inverse temperature parameter, simulations of the Ising model dynamics on general non-regular graphs and the application of this information-theoretic model in other discrete random field models, such as the q -state Potts models, in which each cell assume one of q different states. Finally, we intend to study Gaussian-Markov random field model dynamics, where we have an infinite number of states for each cell.

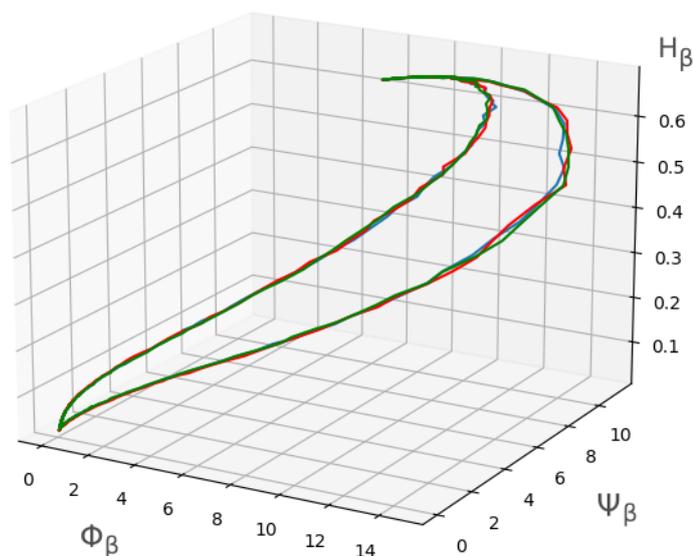


Fig. 9 3D Information curves of the random field dynamics. The three curves (blue, red, green) from three different cycles of the pairwise Ising dynamics show that the characteristic shape is invariant to changes in the initial conditions.

Acknowledgements This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

1. Abdel-All, N.H., Abd-Ellah, H., Moustafa, H.: Information geometry and statistical manifold. *Chaos, Solitons & Fractals* **15**(1), 161 – 172 (2003)
2. Ümit Akıncı, Karakoyun, G.: Hysteresis behavior of the binary Ising model. *Physica B: Condensed Matter* **521**, 365 – 370 (2017)
3. Al-Bender, F., Symens, W.: Dynamic characterization of hysteresis elements in mechanical systems.i. theoretical analysis. *Chaos* **15**, 013105 (2005)
4. Amari, S., Nagaoka, H.: *Methods of Information Geometry*. American Mathematical Society (2000)
5. Ananikian, N.S., Dallakian, S.K., Izmailian, N.S., Oganessyan, K.A.: Strange attractors in an antiferromagnetic Ising model. *Fractals* **05**(01), 175–185 (1997)
6. Arwini, K., Dodson, C.T.J.: *Information Geometru: Near Randomness and Near Independence*. Springer (2008)
7. Bar-yam, Y.: *Dynamics of complex systems*. Taylor & Francis (1997)
8. Barnett, L., Lizier, J.T., Harré, M., Seth, A.K., Bossomaier, T.: Information flow in a kinetic Ising model peaks in the disordered phase. *Phys. Rev. Lett.* **111**, 177203 (2013)
9. Besag, J.: Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)* **36**(2), 192–236 (1974)
10. Bulinski, A., Spodarev, E.: Introduction to random fields. In: E. Spodarev (ed.) *Stochastic Geometry, Spatial Statistics and Random Fields: Asymptotic Methods*, pp. 277–335. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
11. Cafaro, C.: Works on an information geometrodynamical approach to chaos. *Chaos, Solitons & Fractals* **41**(2), 886–891 (2009)

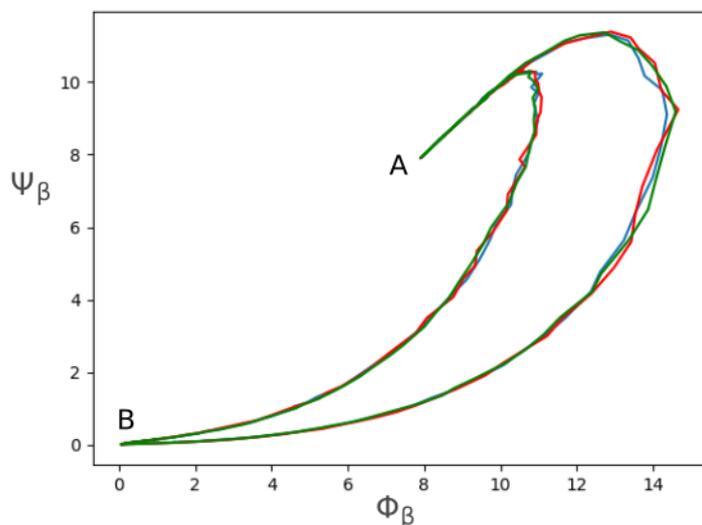


Fig. 10 2D Information curves of the random field dynamics. The characteristic shape of the information curves is invariant to changes in the initial conditions. Note also that in the information-theoretic space, the two regimes of the system are clearly identified by two distinct points A and B.

12. Cafaro, C., Ali, S.: Can chaotic quantum energy levels statistics be characterized using information geometry and inference methods? *Physica A: Statistical Mechanics and its Applications* **387**(27), 6876–6894 (2008)
13. Cafaro, C., Giffin, A., Ali, S., Kim, D.H.: Reexamination of an information geometric construction of entropic indicators of complexity. *Applied Mathematics and Computation* **217**(7), 2944–2951 (2010)
14. Caticha, A.: Towards a statistical geometrodynamics. *Lecture Notes in Physics* **633**, 103–116 (2004)
15. Chopard, B., Droz, M.: *Cellular Automata Modeling of Physical Systems*. Cambridge University Press (2005)
16. Cibra, B.A.: An introduction to the ising model. *The American Mathematical Monthly* **94**(10), 937–959 (1987)
17. Felice, D., Cafaro, C., Mancini, S.: Information geometric methods for complexity. *Chaos* **28**, 032101 (2018)
18. Frieden, B.R.: *Science from Fisher Information*. Cambridge University Press (2004)
19. Friedli, S., Velenik, Y.: *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction*. Cambridge University Press (2017)
20. Ganikhodjaev, N., Mohamed Nawi, A.: Attractors on ising model with restricted interactions on cayley tree of order 2. *AIP Conference Proceedings* **1522**(1), 1300–1305 (2013)
21. Hammersley, J.M., Clifford, P.: Markov field on finite graphs and lattices (preprint) (1971). URL www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf. Unpublished
22. Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**(1), 97–109 (1970)
23. Holovatch, Y., Kenna, R., Thurner, S.: Complex systems: physics beyond physics. *European Journal of Physics* **38**(2), 023002 (2017)

24. Laine, E.M., Breuer, H.P., Piilo, J., Li, C.F., Guo, G.C.: Nonlocal memory effects in the dynamics of open quantum systems. *Physical Review Letters* **108**(21), 210402 (2012)
25. Lima, L., Oliveira, S.: Two-dimensional stochastic dynamics as model for time evolution of the financial market. *Chaos, Solitons & Fractals* **136**, 109792 (2020)
26. Mayergoyz, I.D.: *Mathematical Models of Hysteresis and Their Applications*. Academic Press (2003)
27. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**(6), 1087–1092 (1953)
28. Montalva-Medel, M., Rica, S., Urbina, F.: Phase space classification of an ising cellular automaton: The q2r model. *Chaos, Solitons & Fractals* **133**, 109618 (2020)
29. Pinele, J., Strapasson, J.E., Costa, S.I.R.: The fisher–rao distance between multivariate normal distributions: Special cases, bounds and applications. *Entropy* **22**(4), 404 (2020)
30. Roli, A.: An introduction to complex system science. Tech. rep., University of Bologna, Department of Computer Science and Engineering (2015). URL <https://core.ac.uk/download/pdf/31046012.pdf>
31. Shukla, P.: Hysteresis in the ising model with glauber dynamics. *Phys. Rev. E* **97**, 062127 (2018)
32. Strogatz, S.: Exploring complex networks. *Nature* **410**, 268–276 (2001)
33. Willsky, A., Sandell, N.: The stochastic analysis of dynamic systems moving through random fields. *IEEE Transactions on Automatic Control* **27**(4), 830–838 (1982)
34. Šarlošič, J., Bocko, J., Surovec, R.: Deterministic chaos. *Procedia Engineering* **96**, 458–466 (2014)