

Highlights

Global exploration of phase behavior in frustrated Ising models using unsupervised learning techniques

Danilo Rodrigues de Assis Elias, Enzo Granato, Maurice de Koning

- We employ a machine learning approach based on unsupervised learning techniques to obtain a global picture of phase diagram from single data set.
- Resulting critical lines are in very good agreement with available exact results.
- Results indicate relation between structure of dimensionality-reduced latent space and fundamental physical system properties.

Global exploration of phase behavior in frustrated Ising models using unsupervised learning techniques

Danilo Rodrigues de Assis Elias^a, Enzo Granato^b, Maurice de Koning^{a,c}

^a*Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, UNICAMP, Campinas, 13083-859, São Paulo, Brazil*

^b*Laboratório Associado de Sensores e Materiais, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 12227-010, São Paulo, Brazil*

^c*Center for Computing in Engineering & Sciences, Universidade Estadual de Campinas, UNICAMP, Campinas, 13083-861, São Paulo, Brazil*

Abstract

We apply a set of machine-learning (ML) techniques for the global exploration of the phase diagrams of two frustrated 2D Ising models with competing interactions. Based on raw Monte Carlo spin configurations generated for random system parameters, we apply principal-component analysis (PCA) and auto-encoders to achieve dimensionality reduction, followed by clustering using the DBSCAN method and a support-vector machine classifier to construct the transition lines between the distinct phases in both models. The results are in very good agreement with available exact solutions, with the auto-encoders leading to quantitatively superior estimates, even for a data set containing only 1400 spin configurations. In addition, the results suggest the existence of a relationship between the structure of the optimized auto-encoder latent space and physical characteristics of both systems. This indicates that the employed approach can be useful in perceiving fundamental properties of physical systems in situations where *a priori* theoretical insight is unavailable.

Keywords: Machine learning, phase diagram, frustrated spin systems

1. Introduction

Over the past few years machine learning (ML) has revolutionized the way in which the behavior of complex systems is investigated, providing a data-driven approach that exploits the pattern-recognition powers of such techniques [1]. In particular, it has had a tremendous impact on the physical sciences, where ML methods have been applied to a wide variety of problems originating from areas as diverse as condensed-matter and statistical physics, particle physics, cosmology, quantum computing, chemistry and materials science [2].

Within the field of condensed-matter physics, a major purpose of the application of ML techniques has been to discover the phase behavior of different physical systems [3–22]. In this context, a variety of classical spin systems [3, 4, 6, 7, 12, 17, 18, 23–26] have played a particularly prominent role, displaying the promise of ML techniques in the discovery of complex phases of matter from raw sampling data.

In practice, many of these applications have focused on examining phase behavior as a function of a single parameter, often temperature, seeking to characterize transitions for given fixed values of possible other model parameters. Such an approach has shown to be successful, for instance, to quantify particular critical-temperature values in a number of such spin systems [3–5, 26].

A different kind of application of ML techniques is to employ their pattern-recognition capabilities to gain insight into a system's *global* phase behavior. In this case, instead of focusing on a single critical value for a restricted set of parameter values, the purpose now is to analyze a data set consisting of

configurations generated for the full system parameter space, with the goal of estimating entire critical lines and constructing a picture of the phase diagram as a whole. Such a capability is useful, for instance, when no *a priori* insight into the location of critical regions is available. Casert and co-workers [5] employed such a scheme for the active Ising model, constructing gas-liquid coexistence lines in the density-temperature parameter space. Their approach is based on a two-step procedure in which unsupervised learning techniques are first used to determine phase boundaries for a fixed value of the density, followed by a supervised learning step in which phase boundaries can be predicted for different densities. Acevedo *et al.* [6] reconstructed the critical line between the antiferromagnetic and paramagnetic phases of the two-dimensional frustrated antiferromagnetic Ising model on the honeycomb lattice using anomaly detection [27] based on convolutional autoencoders. For this purpose, they trained an autoencoder on temperature-dependent data for a particular fixed value of the frustration coupling constant and then used it to analyze data generated for other values to detect the transition between ordered and paramagnetic phases using anomaly detection.

In the present paper we consider a different global learning approach in which a single unsupervised learning procedure based on dimensionality reduction is applied to a data set containing samples generated for the entire parameter space. Moreover, these configurations correspond to random parameter values so as to obtain an unbiased data set, presuming no prior knowledge of the system. As an illustration, we construct the phase diagrams of two classical lattice spin systems involving several parameters, namely, the piled-up dominoes

(PUD) and zig-zag dominoes (ZZD) models [28]. These models were introduced in the 1970's as generalizations of the totally frustrated 2D Ising model [29] and incorporate effects of geometrical frustration by the existence of two different spin-spin coupling-parameter values distributed according to regular patterns on the 2D square lattice. The phase behavior of these models is nontrivial, depending on both temperature as well as the relative values of the two coupling constants. In particular, the PUD model displays three types of second-order phase transition: two of them occurring at finite temperature between a paramagnetic phase and either a ferromagnetic or striped antiferromagnetic phase, and a third featuring a transition with vanishing critical temperature.

We analyze the raw Monte Carlo (MC) spin-configuration data generated for randomly selected points in the models' parameter spaces by means of a three-stage data approach that consists of, (i) dimensionality reduction using principal-component analysis (PCA) and auto-encoders, (ii) clustering using the density-based spatial clustering of applications with noise (DBSCAN) algorithm, and (iii) classification using a support-vector machine (SVM). The obtained results enable us to construct the critical lines of the PUD and ZZD models and, given the availability of exact results for both systems [28], the quality of these estimates can be assessed.

The remainder of the manuscript has been organized as follows. In Sec. 2 we define the PUD and ZZD spin-model Hamiltonians and describe the geometric distribution of the two coupling-parameter values across the 2D square lattice. Subsequently, we discuss the details of the employed methodology in Sec. 3, describing the MC procedure employed to generate the set of spin configurations used in the analysis, as well as the ingredients of the ML approach used to process the data. The results are presented and discussed in Sec. 4 and we end with concluding remarks in Sec. 5.

2. Models

As for the standard 2D square Ising system, the PUD and ZZD models are defined by classical spins $s_i = \pm 1$ arranged on a square lattice with nearest-neighbor interactions. However, unlike the standard Ising model, the PUD and ZZD models are characterized by varying interaction-strength parameters. Specifically, the total energy of both models can be written as

$$H = - \sum_{\langle ij \rangle} J_{ij} s_i s_j, \quad (1)$$

where i and j label the spin sites, the notation $\langle ij \rangle$ implies a summation over nearest-neighbor spin pairs and J_{ij} is a spin-pair-dependent interaction-strength parameter with the dimensions of energy. For the PUD and ZZD models J_{ij} can assume only two values, J and J' , which can be either positive or negative. In addition, their distribution across the 2D square lattice is specified in a geometrically ordered pattern, as illustrated in Fig. 1. All horizontal spin pairs interact through the coupling parameter J . In contrast, the interaction parameter J' couples

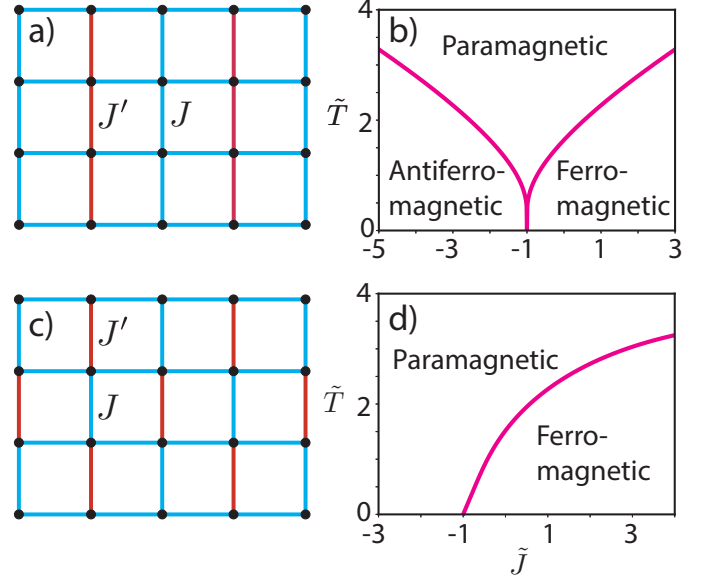


Figure 1: Definition of spin-interaction patterns in a), the piled-up dominoes (PUD) model and c), zig-zag dominoes (ZZD) model as described in Ref. 28. Black circles represent spin sites. Blue and red links represent interactions with strengths J and J' , respectively. Corresponding exact phase diagrams characterized by critical lines given by Eqs. (2) and (3) are depicted in b) and d), respectively.

only spin pairs that are *vertical* neighbors, but only a limited set of them, with the remainder being coupled by J . Specifically, for the PUD model, J' acts on all pairs of alternating vertical rows, as shown in Fig. 1 a), meaning that all spin pairs in even (odd) vertical rows interact through J' , whereas all pairs in the odd (even) vertical rows are coupled through J . In the ZZD model, on the other hand, the interaction between spin pairs from the vertical rows alternates between J and J' and such that neighboring vertical rows are “out of phase”, creating the zig-zag pattern depicted in Fig. 1 c). Given the structure of the Hamiltonian in Eq. (1) the phase behavior of both models can be characterized entirely in terms of the two adimensional parameters $\tilde{J} \equiv J'/J$ and $\tilde{T} \equiv k_B T/J$ [28], with T the absolute temperature and k_B Boltzmann's constant. Of course, both models reduce to the standard 2D square Ising model for $\tilde{J} = 1$. Furthermore, for $\tilde{J} = -1$, the models correspond to the fully frustrated Ising model [29]. Their phase behavior is known exactly [28], as displayed in Figs. 1 b) and d). Specifically, the PUD model features three distinct phases, paramagnetic, anti-ferromagnetic and ferromagnetic, separated by continuous phase transitions described by two critical lines that are solutions of the equations [28]

$$\sinh\left(\frac{2}{\tilde{T}}\right) \sinh\left(\frac{1+\tilde{J}}{\tilde{T}}\right) = \pm 1, \quad (2)$$

respectively, with the minus sign corresponding to the left branch. The ZZD model, on the other hand, is characterized by paramagnetic and ferromagnetic phases, separated by the critical manifold given by the solution of the equation [28]

$$2 \tanh\left(\frac{2}{\tilde{T}}\right) \tanh\left(\frac{1+\tilde{J}}{\tilde{T}}\right) = 1. \quad (3)$$

Interestingly, the disordered paramagnetic phase persists up to zero temperature for values of \tilde{J} below -1.

3. Methodology

3.1. Data Generation

The parameter space of the models is sampled randomly, employing uniform distributions for the parameters \tilde{J} and \tilde{T} within the intervals $\tilde{J} \in (-3, 1)$ and $\tilde{T} \in (0, 3)$ for the PUD model and $\tilde{J} \in (-3, 3)$ and $\tilde{T} \in (0, 4)$ for the ZZD system. To this end we fix $J = 1$ and sample J' and T according to the established intervals. Subsequently, for each randomly chosen pair (\tilde{J}, \tilde{T}) , we record a single representative spin configuration, generated as follows. After initializing the system in a random spin configuration it is subjected to a process in which it is cooled starting from a predefined high temperature, $T_0 = 5$, down to the sampled target temperature \tilde{T} . This particular value of T_0 has been chosen to assure that, for any parameter sample (\tilde{J}, \tilde{T}) , the generation process initializes in the same disordered paramagnetic state common to all Ising-like spin system, minimizing bias in the data set.

The cooling protocol is implemented using standard single-spin-flip Metropolis Monte Carlo (MC) simulations [30] in which the temperature T is reduced at a rate of 2×10^{-4} per MC sweep, which is defined as a set of N random spin-flip trials such that, on average, each of the N spins in the system is given the opportunity to alter its state. After the cooling stage is completed, the system evolves isothermally at the target temperature \tilde{T} for an additional 3×10^3 MC sweeps, after which the final configuration is recorded in the data set. Since only a single configuration is registered for each randomly sampled parameter pair (\tilde{J}, \tilde{T}) , all collected spin configurations are statistically independent.

3.2. Data-analysis

Our data-analysis strategy to estimate the critical manifolds is based on three elements. First, we subject the raw MC configurations to unsupervised-learning techniques with the aim of achieving dimensionality reduction of the data. In addition to having shown to be effective in capturing essential features of physical systems [7, 12], from a data-analysis standpoint it is useful for tackling difficulties associated with the high-dimensional nature of the raw data set [31]. Next, we process the reduced-space results using clustering algorithms [32] to identify distinct groups within the data. In some situations dimensionality reduction alone suffices to identify different groups within the data set, but in more complex scenarios, such as in this study, this is not the case. Either way, clustering techniques should be applied to the reduced space so that the identification of coherent groups is unbiased. After this clusterization we map the elements of the identified clusters to the phase diagram using the values of the parameters \tilde{J} and \tilde{T} associated with each data point and verify whether the distinct agglomerates are located in different regions, to be interpreted as distinct phases. Finally, we use the labels generated by the

clustering procedure to train a classifier and interpret the obtained decision thresholds in terms of the critical manifolds. Below we describe the details of each of the three elements, all of which have been implemented using the Scikit-learn platform [33].

3.2.1. Dimensionality reduction

We apply two different approaches for the dimensionality reduction step, using both principal-component analysis (PCA) [34, 35] as well as develop an auto-encoder [36]. PCA achieves dimensionality reduction by determining the eigenvectors, also known as principal components, of the covariance matrix of the raw MC data (with dimensions $N \times N$, with N the number of spins in the system). The principal components (PCs) are then ranked in order of decreasing eigenvalues. The first eigenvector (i.e., that with the largest eigenvalue) corresponds to the high-dimensional direction in the spin space that has the largest variance in the data set. Subsequently, the second eigenvector corresponds to the direction with the second-largest variance, and so on. The assumption then is that only a few PCs are sufficient to capture the essential information contained in the data set and, possibly, also allowing to group them. A restriction of the PCA approach, however, is that it is a fundamentally linear operation, excluding the possibility to detect non-linear relationships among the variables in the data set.

Auto-encoders, on the other hand, allow detection of such non-linearities. They are neural networks which take the elements of the data set as input and are trained to reproduce that input in the output layer. The structure is that of an hourglass, as shown schematically in Fig. 2. Starting from the input layer, with a number of neurons equal to the dimension of the data set, each subsequent hidden layer decreases its number of neurons until reaching the latent neuron block, which is the layer with the fewest neurons. This part of the structure is referred to as the encoder. The other part of the hourglass is known as the decoder, in which the number of neurons in the hidden layers increases again, in a manner symmetric to the encoder part, until reaching the original number of neurons in the output layer. The dimensionality reduction of the data is achieved by the encoder part, with the bottleneck layer spanning the so-called latent space, which represents a nonlinear distilled representation of each input sample.

Compared to PCA, which amounts to applying a straightforward linear transformation to the data, the creation of an auto-encoder is much more involved. In addition to specifying the number of layers in the hourglass structure and the number of neurons in each of them, including the minimum number of neurons of the latent space, it requires the definition of the neural network. This entails defining the connectivity between the layers, the weights of the connections between the neurons and the functional forms of the activation functions.

3.2.2. Clustering

Among the many available clustering techniques [32] we use the density-based spatial clustering of applications with noise (DBSCAN) algorithm [37]. This choice is motivated by a number of arguments. First, it does not require to pre-define the

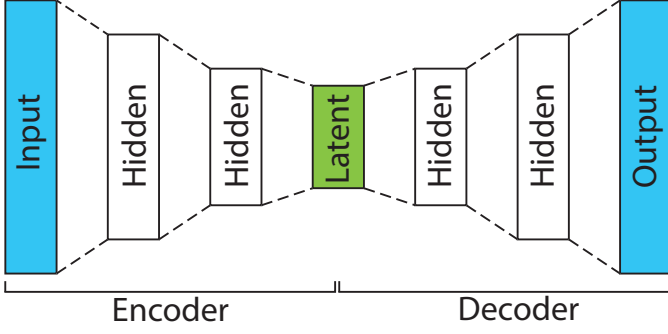


Figure 2: Hourglass structure of auto-encoder neural network. Each block represents a set of neurons. Number of neurons in input and output layers equals the dimension of the data set. Starting from the input layer, the subsequent hidden layers systematically reduce the number of neurons until reaching the bottleneck latent neuron block in which the number of neurons reaches a minimum. Subsequently, the intermediate, hidden layers, systematically reduce the number of neuron until reaching the latent neuron block in which the number of neurons reaches a minimum. Part between the input and the latent block is referred to as the encoder. Part between the latent block and output is referred to as the decoder.

number of clusters as input. Moreover, it is well-suited to handle noisy datasets and, as opposed to the vast majority of clustering methods, it can identify groups with arbitrary shape. The main principle of the DBSCAN approach is to group together data points that lie within a neighborhood of a specified radius ε . In particular, it searches for those data points that, within this neighborhood, have a specified minimum number of neighbors. Each point that satisfies this criterion is classified as a core point. If a data point does not satisfy this property but it is part of the neighborhood of a core point it is considered a border sample. Finally, if none of these conditions are met, the data point is considered noise.

3.2.3. Classification

Finally, after the clusterization step, we apply a classification approach to determine the boundaries between different clusters. To this end we use a support-vector machine (SVM) [39], which provides a robust method for determining nonlinear and fuzzy intersections between clusters.

4. Results and Discussion

The main results for the PUD model are based on a data set containing 1400 independent spin configurations on a 30×30 2D square lattice subject to periodic boundary conditions, each obtained for a single, uniformly sampled parameter pair (\tilde{J}, \tilde{T}) , as discussed in Sec. 3.1. Figure 3 a) displays the results obtained after a dimensionality reduction using PCA, with each data point representing one of the 1400 spin configurations as projected on the two-dimensional space spanned by the first two PCs. Next, to apply the DBSCAN clustering approach, we first need to select an appropriate value for the neighborhood radius parameter ε . To this end, we analyze the distribution of nearest-neighbor distances in PCA-reduced space depicted in Fig. 3 a). Figure 3b) plots the value of this nearest-neighbor distance for each data point, ranked from lowest to highest. It

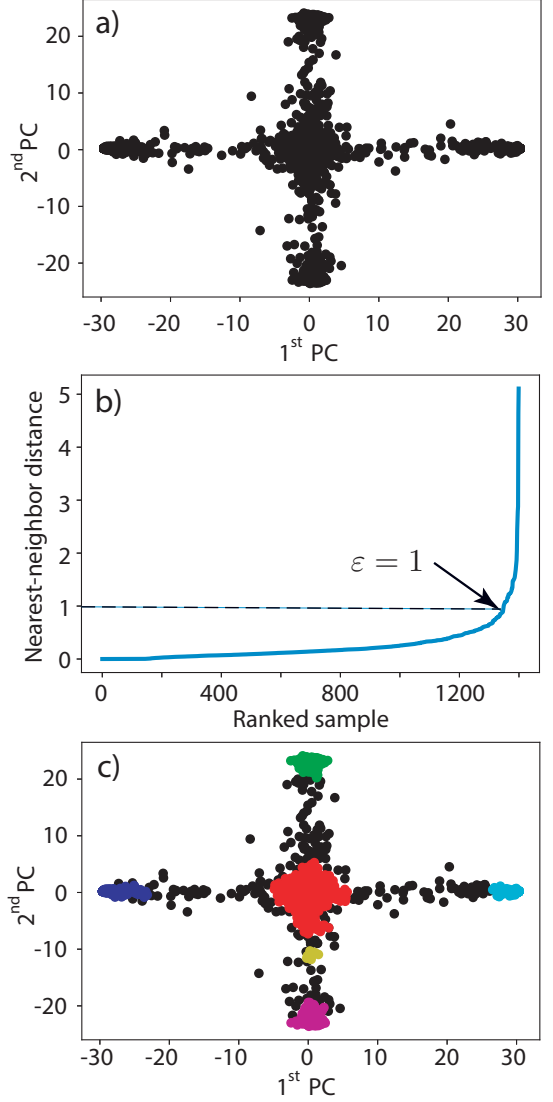


Figure 3: PCA dimensionality reduction of the MC data for the PUD model followed by clustering using DBSCAN. a) Projection of the data on the space spanned by the first 2 PCs. b) Nearest-neighbor distance for each data point in this space, ranked from the smallest to largest. Arrow indicates the location of maximum curvature, which has been shown to provide an appropriate value for the nearest-neighbor radius parameter ε for the DBSCAN clustering algorithm [38]. c) Results after clustering. Black circles represent data points classified as noise. Different colors correspond to data points attributed to distinct clusters.

has been shown [38] that the optimal value of ε corresponds to the distance at which the curvature of the distance versus rank curve is maximum, which in this case is $\varepsilon \approx 1$, as indicated by the arrow. Figure 3 c) then shows the results after clustering with DBSCAN, using a neighborhood radius $\varepsilon = 1$ and setting the minimum number of neighbors within this radius to be 10. Aside from the black circles, which are configurations that have been classified as noise, the data points painted with different colors belong to distinct clusters.

As a second approach toward dimensionality reduction we develop an auto-encoder with the schematic structure shown in Fig. 2. The resulting optimized neural-network structure is

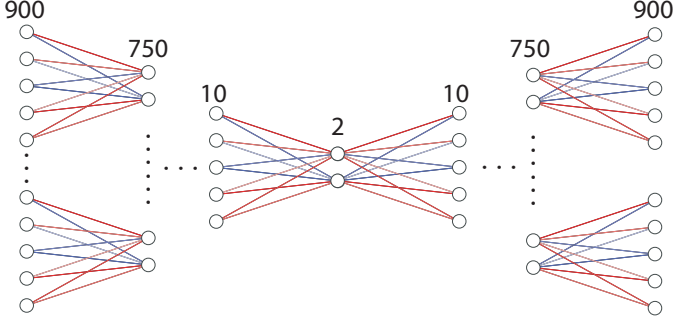


Figure 4: (Color online) Artificial neural network structure of auto-encoder. In addition to the input layer consisting of 900 neurons, the encoder portion consists of 8 hidden layers with, respectively, 750, 600, 450, 300, 150, 75, 30 and 10 neurons, before reaching the latent layer containing 2 neurons. The decoder portion is symmetric with respect to the latent layer. The neural network is linked such that subsequent layers are fully connected.

depicted in Fig. 4. In addition to the 900 neurons on the input layer, the encoder section features a succession of 8 hidden layers with, respectively, 750, 600, 450, 300, 150, 75, 30 and 10 neurons, before reaching the latent layer that consists of 2 neurons. The decoder part is symmetric with respect to the latent layer. The neural network between successive layers is fully connected and we employ hyperbolic tangents as activation functions [31]. The optimization of the auto-encoder neural network was implemented using the PyTorch package [40], employing its MSELoss function [41] as the loss measure and the Adam algorithm as the adaptive optimizer [41]. The optimization process is organized in three steps. First, the auto-encoder is pre-trained using 400 of the 1400 system configurations at a learning rate [41] of 5×10^{-4} for 1500 iterations (i.e., epochs[41]). Subsequently, the training procedure covers the entire data set for 4000 more iterations using the same learning rate. Finally, an additional 2000 iterations is carried out for the entire data set at a learning rate of 5×10^{-5} . The corresponding dimensionality-reduced representation of the data set is shown in Fig. 5 a), which depicts the outputs of the neurons L_1 and L_2 of the latent layer for the 1400 spin configurations. Subsequently, to determine the value of the radius parameter ϵ for DBSCAN, we determine the nearest-neighbor statistics of the data set in the same way as done in Fig. 3 b).

The results are shown in Fig. 5 b), which shows the nearest-neighbor distance for all data points in Fig. 5 a), ranked from lowest to largest. The fundamental difference between this profile and the one obtained for PCA is that, in this case, there are two local maxima in the curvature of the rank-distance plot, as shown by the arrows. This indicates that groups belonging to the latent space as represented in Fig. 5 a) have different characteristic densities. The existence of two density profiles can in fact be traced to the physical characteristics of the PUD model, as we will discuss later on, but under these conditions the DBSCAN clustering algorithm is known to be less effective. If one chooses ϵ according to the smaller of the two, i.e., focusing predominantly on clustering data points characteristic of high density regions, lower-density samples will not be recognized as belonging to any group at all. In contrast, if we select the

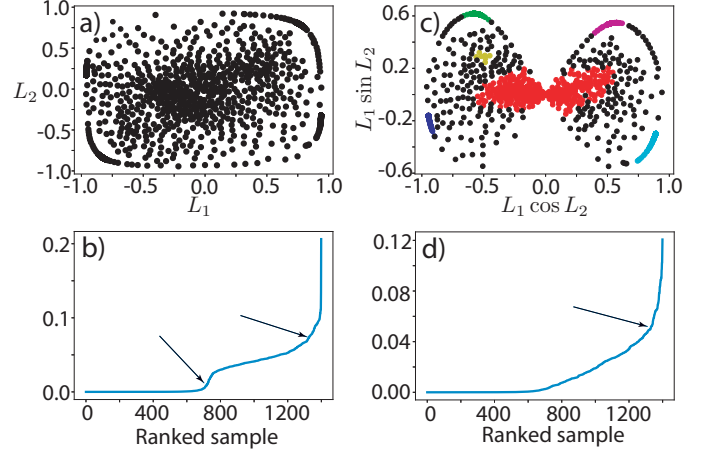


Figure 5: Auto-encoder dimensionality reduction and subsequent DBSCAN clusterization for the PUD data set. a) Data represented in the latent space as characterized by the output L_1 and L_2 of the corresponding 2 neurons. b) Cluster density analysis based on ranked nearest-neighbor distances for data points shown in a). c) Transformation of the latent space in terms of the transformed variables $L_1 \cos L_2$ and $L_1 \sin L_2$, respectively. Different colors represent data points in distinct clusters as obtained using DBSCAN. Black circles indicate configurations classified as noise. d) Cluster density analysis based on ranked nearest-neighbor distances for latent-space points shown in c).

larger of the two, i.e., choose cluster according to lower density regions, regions of higher density will incorporate data points that should not be included, giving rise to cluster overlap.

In this case, it has been shown that a transformation of coordinates in the latent space may be helpful to improve the identification of distinct clusters. Specifically, polar-coordinate-like interpretations have shown to be particularly useful on a number of occasions, including in the recognition of handwritten digits [42–44]. Following this approach, we transform the neuron outputs L_1 and L_2 of Fig. 5a) by interpreting L_1 as a generalized “radius” (allowing both positive and negative values) and L_2 as an “angle”, giving new Cartesian components defined as $L_1 \cos L_2$ and $L_1 \sin L_2$, respectively. Inverting the roles of L_1 and L_2 in this transformation does not alter the results. Figure 5 c) and d) display the corresponding structure of the transformed latent space and rank-distance curve, respectively, with the latter now having a single point of maximum curvature, allowing an effective application of the DBSCAN clustering algorithm. As in Fig. 3 c), the different colored data points in Fig. 5c) correspond to distinct clusters identified by DBSCAN, whereas the black circles correspond to configurations that are classified as noise.

Having clustered the data for both dimensionality-reduced representations, we now map the corresponding spin configurations onto the $\tilde{T} \times \tilde{J}$ parameter space of the PUD model, maintaining the color coding adopted in Figs. 3 c) and 5 c). The corresponding results are depicted in Figs 6 a) and b), which display the mappings produced by PCA and the auto-encoder, respectively. An immediate observation is that, in both cases, each non-noise color occupies only a restricted part of the parameter space, dividing it into three distinct regions. Figure 6c) shows representative spin arrangements from these

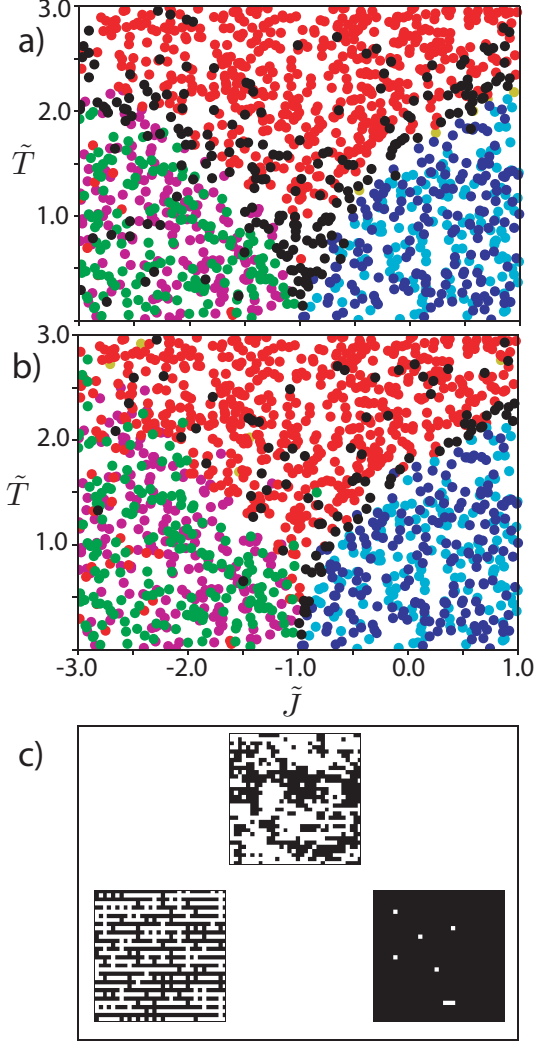


Figure 6: (Color online) Mapping of reduced-space clusters into the PUD parameter space. a) PCA clustered data. b) Auto-encoder clustered data. c) Representative spin configurations for the three phases with black/white squares representing spin up/down, respectively.

areas, clearly showing the distinct nature of the spin conformations in each of them. In this sense, each region represents a distinct spin phase, displaying a disordered paramagnetic phase, a ferromagnetic phase and a stripe-like anti-ferromagnetic phase. A further notable feature of the data mapping in Fig. 6a) and b) is that, while the paramagnetic region corresponds to a single cluster, the ferro and anti-ferromagnetic parts of the parameter space are occupied by two distinct clusters each. We will further discuss this point below.

Now that the various clusters in the dimensionality-reduced spaces of PCA and the auto-encoder in Fig. 6 have been associated with different types of spin configurations, we now analyze the data so as to determine the manifolds in the parameter space that separate the different phases. To this end, we define three different classes, corresponding to the three regions in the parameter space identified in Fig. 6 a) and b). In particular, the classes consist of, (i), the spin configurations from red cluster for the paramagnetic phase, (ii) the data from

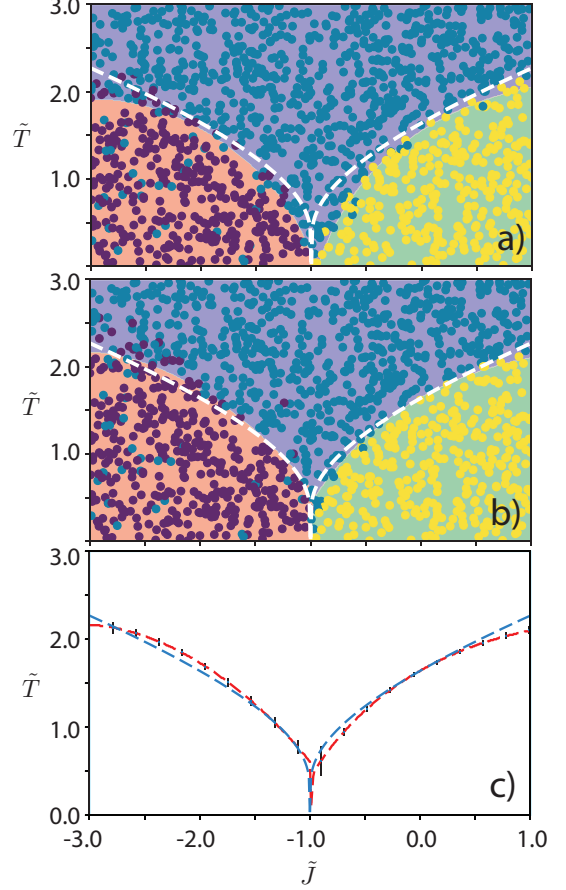


Figure 7: (Color online) Colored circles represent the spin configurations after DBSCAN clustering of the PCA and auto-encoder dimensionality reduction. Dark and light blue data points in Fig. 6a) and b) have been assigned to the yellow circles, the green and magenta to purple, and the red and black to blue. Different background colors represent regions defined by the SVM as belonging to a specific class. The boundaries between them represent the decision thresholds. White dashed lines represent analytical critical lines corresponding to the solutions of Eq. (2). a) PCA-based phase diagram. b) Polar-coordinate latent-space phase diagram from auto-encoder. c) SVM decision threshold (red line) including uncertainty estimates (vertical black bars) as estimated using the train-test split method [33] compared to analytical critical lines (blue line).

the light and dark blue clusters for the ferromagnetic phase and, (iii), the green and magenta clusters for the stripe-like anti-ferromagnetic phase. Based on these classes we employ a supervised-learning SVM multi-classification approach to establish the boundaries between these regions. The corresponding results for PCA and the auto-encoder are shown in Fig. 7 a) and b), respectively. The different background colors correspond to the different classes, with the boundaries between them representing the so-called decision thresholds that form the critical manifolds in the parameter space that separate the different phases. We can compare these results directly to the critical lines defined by Eq. (2), shown as the dashed white lines.

Overall, both PCA and the auto-encoder closely capture the qualitative features of the critical lines in the parameter space, including the existence of a critical point at zero temperature for $\tilde{J} = -1$. It is clear, however, that the phase-behavior description

provided by the auto-encoder is superior in terms of quantitative agreement, with the SVM decision thresholds closely overlapping the analytical results. This is further demonstrated in Fig. 7 c), which compares the SVM decision thresholds including confidence intervals as obtained using the train-test split method [33], to the exact results [28]. The agreement is encouraging, in particular because the data points were selected randomly and uniformly within a very broad range of the parameter space, without any bias toward known transition lines.

The main reason for the quantitative difference between PCA and the auto-encoder is the fact that the first 2 PCs cover only 34% of the variance in the data, with the remaining 66% being diluted over the other 897 eigenvectors. The auto-encoder, on the other hand, provides a more comprehensive data reduction scheme. Although it does not involve the concept of explained variance, as for PCA, the fidelity obtained in training the auto-encoder can be used as a measure for the accuracy in the recovery of the input image by the encoder. In this particular case, the auto-encoder can be pushed to reproduce input images with a fidelity superior to 90%. As discussed previously for a variety of other classical spin systems [4, 6], the superiority of the auto-encoder approach is not surprising given that PCA is intrinsically limited to linear transformations of the data.

Even so, the PCA dimensionality reduction, followed by DBSCAN clustering still provides insightful information regarding the physical characteristics of the system. In particular, it has been shown to be useful in the identification of order parameters characterizing phase transitions [4, 45]. This can be appreciated in the cluster structure depicted in Fig. 3 c), in which vertically and horizontally opposite clusters belong to the same striped anti-ferromagnetic and ferromagnetic phases, respectively, whereas the centralized red agglomerate corresponds to the paramagnetic phase. In this sense, the first PC clearly distinguishes between the two symmetric ferromagnetic states (i.e., spin up and spin down) and the paramagnetic phase, whereas the second PC does so to differentiate the (spin-up and spin-down) striped anti-ferromagnetic phases from the paramagnetic phase.

Fig. 8 shows the results obtained for the other frustrated spin system, defined by the ZZD model, which is characterized by a different geometric patterns of the coupling parameters J and J' . The dimensionality reduction is obtained using an auto-encoder with the same two-neuron bottleneck architecture applied for the PUD model. Fig. 8 a) displays the resulting latent-space structure after applying the same polar-coordinate-like transformation used for the PUD model and using DBSCAN clusterization. It displays the same butterfly-like shape as for the PUD, but in this case the number of distinct clusters is reduced by two. Specifically, the two mirror-symmetric clusters on the lower part of the butterfly for the PUD model have disappeared. When mapping the data points onto the (\tilde{J}, \tilde{T}) parameter space, as depicted in Fig. 8 b), it is clear this reduction is related to the fact that the ZZD model is characterized by two instead of three phases, displaying a ferromagnetic phase for positive values of \tilde{J} and low \tilde{T} and a disordered paramagnetic phase for negative values of \tilde{J} or high temperatures. After grouping the clusters belonging to the same region of the phase

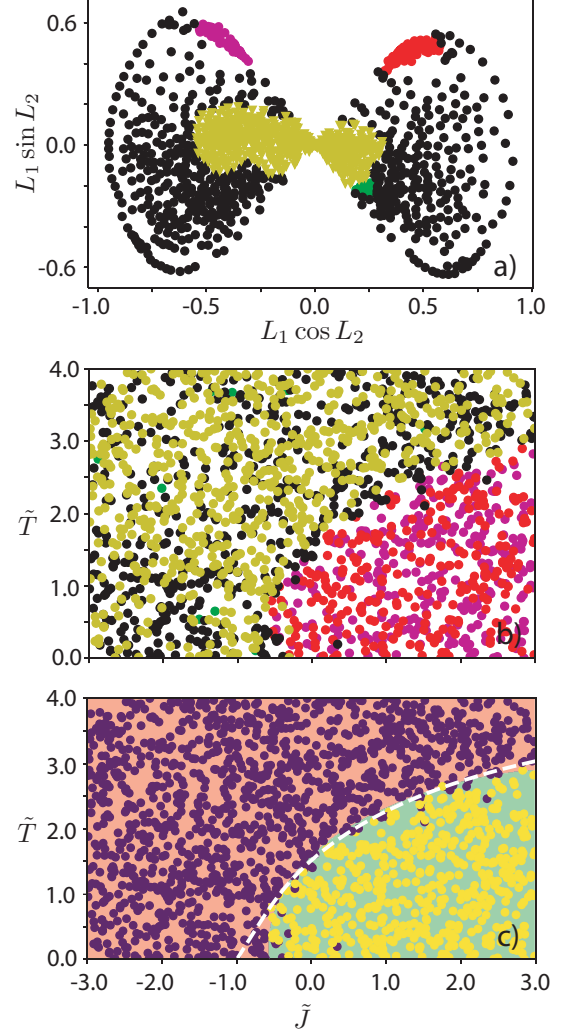


Figure 8: (Color online) Auto-encoder results for ZZD model. a) Hyperspherical latent space after DBSCAN clustering, with different clusters represented by different colored circles. b) Spin configurations plotted in the (\tilde{J}, \tilde{T}) parameter space, maintaining the colors from a). c) Corresponding ZZD phase diagram as obtained by SVM classification. Different background colors represent regions defined by the SVM as belonging to a specific class. The boundaries between them represent the decision thresholds. White dashed line represents analytical critical line from Eq. (3).

diagram into distinct classes and training an SVM classifier we obtain the critical line separating both phases as the boundary separating both background colors, as shown in Fig. 8 c). As for the PUD model, the agreement with the exact result described by Eq. (3), shown as the white dashed line, is excellent.

An interesting question is to what extent there is a relation between the structural characteristics of the optimized auto-encoder and the fundamental physical characteristics of the two spin models. In this context, we note that for both spin models the best results are obtained using a bottleneck layer formed by 2 neurons, obtaining a reproduction fidelity $\gtrsim 90\%$. If, for instance, only a single neuron is used, the optimized fidelity does not exceed $\sim 50\%$, whereas a bottleneck layer containing three neurons does not significantly improve the fidelity. The fact that the optimal latent-space dimension is two for both systems,

even though the number of distinct phases for both models is different (3 for the PUD versus 2 for the ZZD), suggests that the optimal dimension of the latent space may be related to the number of independent parameters required to characterize the phase diagram. A further indication for the connection between the structure of the auto-encoder latent space and the physical characteristics of the system can be seen in Figs. 5 c) and 8 a). The butterfly-shaped hyper-spherical latent spaces are mirror-symmetric with respect to the vertical axis passing through the origin, with the symmetry-related cluster pairs belonging to the same regions in the phase diagram. The symmetry brought out by the clustered auto-encoder results thus provides insight into the physical system under consideration, revealing that opposite clusters contain configurations that are spin reversals of each other. In other words, the vertical axes in Figs. 5 c) and 8 a) correspond to the spin-reversal symmetry underlying the system Hamiltonians. This “discovery” of a fundamental system symmetry through the optimization of the auto-encoder is a further example that the global ML phase exploration employed here can be useful in perceiving fundamental properties of physical systems when prior theoretical insight is unavailable.

In a similar context, the existence of two density profiles in the Cartesian auto-encoder latent space for the PUD model, as shown in Fig. 5 b), is in fact related to the nature of its ferro and antiferromagnetic phases of the PUD model. Theoretical analysis [28] indicates that the entropy of its anti-ferromagnetic phase is much greater, meaning that the variability among different samples is much larger compared to that for the ferromagnetic configurations. This is the reason underlying the existence of the two density profiles for the ranked sample-sample distances in Fig. 5 b). Whereas the clusters belonging to the lower ε (i.e., higher density) correspond to ferromagnetic configurations, the ones for the larger ε value represent anti-ferromagnetic spin conformations. Even though the data does not allow quantification of the configurational entropy for both phases, it does provide qualitative indications that it is larger for the striped anti-ferromagnetic phase, without previous knowledge of the system.

It is interesting to contrast the present fully-connected auto-encoder approach to that employed for the two-dimensional frustrated antiferromagnetic Ising model as reported in Ref. 6. In the latter, a convolutional auto-encoder (CAE) was used to determine the critical line between the antiferromagnetic and paramagnetic phases. The advantage of CAEs over the fully connected auto encoder utilized in the present work is that they are characterized by a much smaller number of parameters and feature the ability to recognize spatial correlations by the application of filters or local transformations on the images, allowing, for instance, to construct the phase diagram of the Bose-Hubbard model [46]. Even so, in addition to the fact that the implementation of CAEs is generally substantially more involved, the CAEs in Ref. 6 do not have a low-dimensional latent space, such that its detection of phase transitions is not based on dimensionality reduction, but rather on anomaly detection. In this sense, both auto-encoder techniques approach the construction of phase diagrams in very different manners and may be considered as complementary to each other. Particularly useful fea-

tures of the present fully-connected auto-encoder approach are its interpretability in terms of the latent space features and its efficiency, giving accurate estimates for the critical lines based on only 1400 MC spin configurations.

A final point concerns the influence of the system size on the global estimate of phase diagrams. In applications of ML techniques used to locate a single critical point, such a size-dependence analysis has been instrumental in achieving quantitative accuracy [3, 4]. In the context of classical spin systems, for instance, such an analysis involves generating configurations in the vicinity of a critical point for different linear lattice dimensions L , after which pertinent outcomes are analyzed as a function of some power of L . For instance, using PCA for the standard 2D ferromagnetic Ising model, its critical temperature can be determined by plotting the temperature for which the quantified second principal component reaches its maximum as a function of L^{-1} , followed by extrapolation to the limit of infinite lattice dimension [4].

In the present global ML approach, however, in which one uses a single data set consisting of configurations sampled from the entire parameter space of the system, local scaling properties become blurred. First, the dimensionality reduction encodes the characteristics of the phase diagram as a whole, incorporating a continuum of critical temperatures instead of a particular single value. Furthermore, in a similar fashion, the final estimates for the critical lines are obtained using a SVM classifier, whose result corresponds to the minimum of a classification error that is global in nature rather than reflecting behavior in the vicinity of a single critical point. This is illustrated in Fig. 9, which displays the PCA results for the PUD model for lattice dimensions $L = 8, 16$ and 64 . The size effect on the distribution of the data points in the clustered latent space is evident. Whereas the results display appreciable dispersion for $L = 8$, the projections onto the first two principal components become progressively sharper as the system size grows, increasing the distinguishability among the ferromagnetic, anti-ferromagnetic and paramagnetic phases. On the other hand, even though the SVM critical lines for $L = 64$ are visibly better than those obtained for $L = 8$ in an overall sense, there is no manifestly visible size-scaling behavior. This is shown in the inset of Fig. 9 f), which plots the SVM critical temperature as a function of $1/L$ for the case of the standard 2D Ising model (i.e., $\tilde{J} = 1$). Even though the estimates are within 7% of the exact value for all system sizes, there is no discernible L^{-1} dependence.

In this sense, the present global ML approach can be considered to be complementary to the schemes that have been utilized to quantify the parameters characterizing a particular critical point [3, 4]. Whereas the present approach allows one to obtain a good first picture of the overall phase diagram when no *a priori* insight is available, the latter can subsequently be used to systematically obtain better accuracy for the critical parameters by targeting the generation of additional data in the vicinity of the initial transition estimate and using techniques as described in Refs. 3 and 4. This is also consistent with the findings of Théveniaut and Alet for the many-body localization transition in the Heisenberg spin 1/2 chain [25], who reported that uncer-

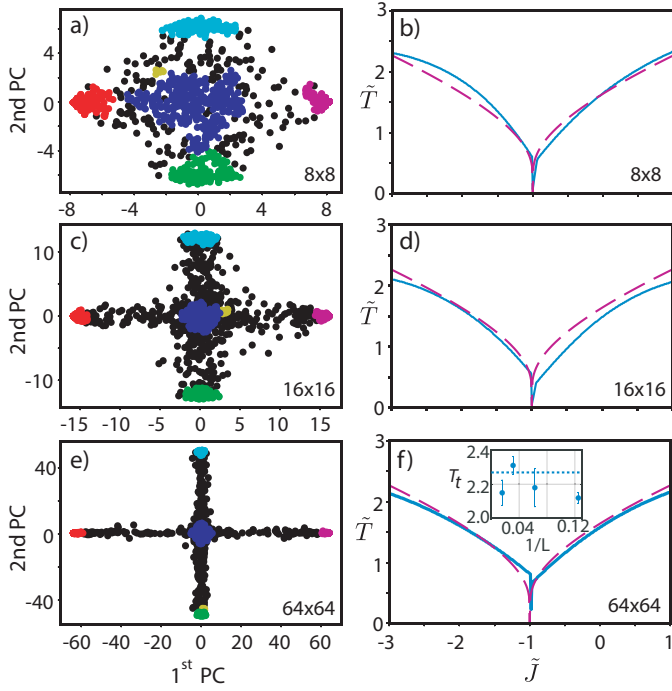


Figure 9: (Color online) System size dependence of latent space and corresponding SVM estimates for the critical lines for PUD model for 8x8 16x16 and 64x64 lattice sizes. Inset in f) displays SVM critical temperature estimate as a function of L^{-1} for $\tilde{J} = 1$, with horizontal dashed line indicating exact value $T_c = 2 / \ln(1 + \sqrt{2})$.

tainties in the estimates for critical parameters extracted from neural-network structures tend to be larger than those obtained from conventional approaches.

5. Conclusions

In summary, we consider a ML approach for obtaining a global prediction of the phase diagrams for two frustrated Ising models, which are characterized by the presence of critical lines rather than a single critical point. Using raw MC spin configurations generated for random parameter values from the entire parameter space, we first apply unsupervised learning techniques such as PCA and auto-encoders to achieve dimensionality reduction, followed by clustering using the DBSCAN method and a classification step using a SVM approach. The resulting estimates for the critical lines of both systems are in excellent agreement with available exact results, even though the data points were selected randomly and uniformly within an ample region of the parameter space, without any predisposition toward the critical lines.

In both cases, the auto-encoder dimensionality-reduction approach is found to provide better quantitative results as compared to those based on PCA. The main origin of this difference is that the optimized auto-encoder gives an image fidelity $\gtrsim 90\%$, with two neurons in the bottleneck layer. In contrast, the corresponding two-dimensional latent space for PCA captures only 34% of variance in the data. A further observation is that the latent-space characteristics of the optimized auto-

encoders appear to relate to fundamental physical characteristics of both considered spin models. In addition to recognizing the spin-up/spin-down symmetry, in both cases the best results are obtained using a bottleneck layer formed by 2 neurons, indicating that the optimal latent-space dimension is linked to the number of independent parameters required to characterize the phase diagram. Although this relation between the optimal auto-encoder structure and the physical properties requires further study, it suggests that the present ML approach may be helpful in perceiving fundamental physical properties when, for instance, *a priori* theoretical insight is unavailable.

Finally, although the approach provides good estimates for the phase diagram as a whole, it is not naturally suited to be used in finite-size scaling procedures that are often employed to accurately quantify critical parameters. This is due to the fact that the predictions are based on data sets generated for the entire parameter space rather than in the vicinity of a particular critical point. In addition, the estimates for the transition lines are obtained using a classification procedure that seeks to minimize a global classification error across the phase diagram as a whole such that local critical properties become obscured by global optimization criteria. In this sense, the present approach is complementary schemes that are utilized to quantify critical parameters by focusing on data generated in the vicinity of a critical point.

CRedit authorship contribution statement

D.R.A.E: Conceptualization, execution of computations, writing. E.G: Conceptualization, review and editing. M.K.: Conceptualization, writing and supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge support from CNPq, Fapesp grants no. 2016/23891-6 (MK) and 2018/19586-9 (EG), as well as the Center for Computing in Engineering & Sciences - Fapesp/Cepid no. 2013/08293-7 (MK). We gratefully acknowledge Rodolfo Paula Leite, Levy Boccato, Romis Ribeiro de Faissol Attux and Gabriel Reis Garcia.

References

- [1] C. M. Bishop, Pattern Recognition and Machine Learning, Information Science and Statistics, Springer, 2006.
- [2] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. 91 (2019) 045002.
- [3] J. Carrasquilla, R. G. Melko, Machine learning phases of matter, Nat. Phys. 13 (2017) 431.

- [4] W. Hu, R. R. P. Singh, R. T. Scalettar, Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination, *Phys. Rev. E* 95 (2017) 062122.
- [5] C. Casert, T. Vieijra, J. Nys, J. Ryckebusch, Interpretable machine learning for inferring the phase boundaries in a nonequilibrium system, *Phys. Rev. E* 99 (2019) 023304.
- [6] S. Acevedo, M. Arlego, C. A. Lamas, Phase diagram study of a two-dimensional frustrated antiferromagnet via unsupervised machine learning, *Phys. Rev. B* 103 (2021) 134422.
- [7] L. Wang, Discovering phase transitions with unsupervised learning, *Phys. Rev. B* 94 (2016) 195105.
- [8] K. Ch'ng, J. Carrasquilla, R. G. Melko, E. Khatami, Machine learning phases of strongly correlated fermions, *Phys. Rev. X* 7 (2017).
- [9] P. Ponte, R. G. Melko, Kernel methods for interpretable machine learning of order parameters, *Phys. Rev. B* 96 (2017).
- [10] E. P. L. Nieuwenburg, Y.-H. Liu, S. D. Huber, Learning phase transitions by confusion, *Nat. Phys.* 13 (2017).
- [11] D. L. Deng, X. Li, S. Sarma, Machine learning topological states, *Phys. Rev. B* 96 (2017).
- [12] S. J. Wetzel, Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders, *Phys. Rev. E* 96 (2017) 022140.
- [13] S. J. Wetzel, M. Scherzer, Machine learning of explicit order parameters: from the ising model to su(2) lattice gauge theory, *Phys. Rev. B* 96 (2017).
- [14] K. Ch'ng, N. Vazquez, E. Khatami, Unsupervised machine learning account of magnetic transitions in the hubbard model, *Phys. Rev. E* 97 (2018).
- [15] Y.-H. Liu, E. P. L. van Nieuwenburg, Discriminative cooperative networks for detecting phase transitions, *Phys. Rev. Lett.* 120 (2018) 176401.
- [16] E. Nieuwenburg, E. Bairey, G. Refael, Learning phase transitions from dynamics, *Phys. Rev. B* 98 (2018).
- [17] D. Kim, D.-H. Kim, Smallest neural network to learn the ising criticality, *Phys. Rev. E* 98 (2018) 022138.
- [18] K. Mills, I. Tamblyn, Deep neural networks for direct, featureless learning through observation: The case of two-dimensional spin models, *Phys. Rev. E* 97 (2018) 032119.
- [19] J. Venderley, V. Khemani, E.-A. Kim, Machine learning out-of-equilibrium phases of matter, *Phys. Rev. Lett.* 120 (2018) 257204.
- [20] M. Ceriotti, Unsupervised machine learning in atomistic simulations, between predictions and understanding, *J. Chem. Phys.* 150 (2019) 150901.
- [21] J. F. Rodriguez-Nieva, M. S. Scheurer, Identifying topological order through unsupervised machine learning, *Nat. Phys.* 15 (2019) 790–795.
- [22] R. Freitas, E. J. Reed, Uncovering the effects of interface-induced ordering of liquid on crystal growth using machine learning, *Nat. Commun.* 11 (2020) 3260.
- [23] C. Wang, H. Zhai, Machine learning of frustrated classical spin models. i. principal component analysis, *Phys. Rev. B* 96 (2017) 144432.
- [24] C. Wang, H. Zhai, Machine learning of frustrated classical spin models (ii): Kernel principal component analysis, *Front Phys-beijing* 13 (2018) 130507.
- [25] H. Théveniaut, F. Alet, Neural network setups for a precise detection of the many-body localization transition: Finite-size scaling and limitations, *Phys. Rev. B* 100 (2019) 224202.
- [26] C. Alexandrou, A. Athenodorou, C. Chrysostomou, S. Paul, The critical temperature of the 2d-ising model through deep learning autoencoders, *Eur. Phys. J. B* 93 (2020) 226.
- [27] H. D. Purnomo, T. Wahyono, Y. Heryadi, Machine learning applications for anomaly detection, IGI Global, Hershey, PA, USA, 2019, pp. 49–83.
- [28] G. André, R. Bidaux, J.-P. Carton, R. Conte, L. de Seze, Frustration in periodic systems : exact results for some 2d ising models, *J. Phys. France* 40 (1979) 479–488.
- [29] J. Villain, Spin glass with non-random interactions, *J. Phys. C: Solid State Phys.* 10 (1977) 1717–1734.
- [30] M. E. J. Newman, G. T. Barkema, Monte Carlo Methods in Statistical Physics, Oxford University Press, 1999.
- [31] K. P. Murphy, Machine Learning: A Probabilistic Perspective, Adaptive Computation and Machine Learning series, MIT Press, 2012.
- [32] V. Estivill-Castro, Why so many clustering algorithms: A position paper, *SIGKDD Explor.* 4 (2002) 65–75.
- [33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [34] K. Pearson, Liii. on lines and planes of closest fit to systems of points in space, *Phil. Mag. Series 6* 2 (1901) 559–572.
- [35] I. T. Jolliffe, Principal Component Analysis, Springer Series in Statistics, Springer New York, 2006.
- [36] M. A. Kramer, Nonlinear principal component analysis using autoassociative neural networks, *AIChE J.* 37 (1991) 233–243.
- [37] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, in: E. Simoudis, J. Han, U. Fayyad (Eds.), Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, AAAI, Palo Alto, 1996, p. 226.
- [38] N. Rahmah, I. S. Sitanggang, Determination of optimal epsilon (eps) value on dbscan algorithm to clustering data on peatland hotspots in sumatra, *IOP Conf. Ser.: Earth Environ. Sci.* 31 (2016) 012012.
- [39] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (1995) 273–297.
- [40] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035.
- [41] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, Adaptive Computation and Machine Learning series, MIT Press, 2016.
- [42] T. R. Davidson, L. Falorsi, N. D. Cao, T. Kipf, J. M. Tomczak, Hyperspherical variational auto-encoders., in: Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 2018, Monterey, California, USA, August 6-10, 2018, 2018, pp. 856–865.
- [43] M. R. Joshi, Y. S. Patil, Analysis of change in coordinate system on clustering, in: 2016 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC), 2016, pp. 1–7.
- [44] Y. S. Patil, M. R. Joshi, Clustering with polar coordinates system: Exploring possibilities, in: Smart Intelligent Computing and Applications, Springer Singapore, Singapore, 2019, pp. 553–560.
- [45] H. Kiwata, Deriving the order parameters of a spin-glass model using principal component analysis, *Phys. Rev. E* 99 (2019) 063304.
- [46] K. Kottmann, P. Huembeli, M. Lewenstein, A. Acín, Unsupervised phase discovery with deep anomaly detection, *Phys. Rev. Lett.* 125 (2020) 170603.