

# Unsupervised learning of phase transitions: from principal component analysis to variational autoencoders

Sebastian J. Wetzel<sup>1</sup>

<sup>1</sup>*Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany*

We employ unsupervised machine learning techniques to learn latent parameters which best describe states of the two-dimensional Ising model and the three-dimensional XY model. These methods range from principal component analysis to artificial neural network based variational autoencoders. The states are sampled using a Monte-Carlo simulation above and below the critical temperature. We find that the predicted latent parameters correspond to the known order parameters. The latent representation of the states of the models in question are clustered, which makes it possible to identify **phases without prior knowledge of their existence or the underlying Hamiltonian**. Furthermore, we find that the reconstruction loss function can be used as a universal identifier for phase transitions.

## I. INTRODUCTION

Inferring macroscopic properties of physical systems from their microscopic description is an ongoing work in many disciplines of physics, like condensed matter, ultra cold atoms or quantum chromo dynamics. The most drastic changes in the macroscopic properties of a physical system occur at phase transitions, which often involve a symmetry breaking process. The theory of such phase transitions was formulated by Landau as a phenomenological model [1] and later devised from microscopic principles using the renormalization group [2, 3]. One can **identify phases by knowledge of an order parameter which is zero in the disordered phase and nonzero in the ordered phase**.

Whereas in many known models the order parameter can be determined by symmetry considerations of the underlying Hamiltonian, there are states of matter where such a **parameter can only be defined in a complicated non-local way** [4]. These systems include topological states like topological insulators, quantum spin hall states [5] or quantum spin liquids [6]. Therefore, we need to develop new methods to identify parameters capable of describing phase transitions.

Such methods might be borrowed from machine learning. Since the 1990s this field has undergone major changes with the development of more powerful computers and artificial neural networks. It has been shown that such neural networks can approximate every function under mild assumptions [7, 8]. They quickly found applications in image classification, speech recognition, natural language understanding and predicting from high-dimensional data. Furthermore, they began to outperform other algorithms on these tasks [9].

In the last years physicists started to employ machine learning techniques. Most of the tasks were tackled by supervised learning algorithms or with the help of reinforcement learning [10–22]. Supervised learning means one is given labeled training data from which the algorithm learns to assign labels to data points. After successful training it can then predict the labels of previously unseen data with high accuracy.

In addition to supervised learning, there are unsupervised learning algorithms which can find structure in unlabeled data. They can also classify data into clusters, which are however unlabelled. It is already possible to employ unsupervised learning techniques to reproduce Monte-Carlo-sampled states of the Ising model [23]. Phase transitions were found in an unsupervised manner using principal component analysis [24, 25]. We employ more powerful machine learning algorithms and transition to methods that can handle nonlinear data. A first nonlinear extension is kernel principal component analysis [26].

The first versions of autoencoders have been around for decades [27, 28] and were primarily used for dimensional reduction of data before feeding it to a machine learning algorithm. They are created from an encoding artificial neural network, which outputs a latent representation of the input data, and a decoding neural network that tries to accurately reconstruct the input data from its latent representation. Very shallow versions of autoencoders can reproduce the results of principal component analysis [29].

In 2013, variational autoencoders have been developed as one of the most successful unsupervised learning algorithms [30]. In contrast to traditional autoencoders, variational autoencoders impose restrictions on the distribution of latent variables. They have shown promising results in encoding and reconstructing data in the field of computer vision.

In this work we use unsupervised learning to determine phase transitions without any information about the microscopic theory or the order parameter. We transition from principal component analysis to variational autoencoders, and finally test how the latter handles different physical models. Our algorithms are able to find a low dimensional latent representation of the physical system which coincides with the correct order parameter. The decoder network reconstructs the encoded configuration from its latent representation. We find that the reconstruction is more accurate in the ordered phase, which suggests the use of the reconstruction error as a universal identifier for phase transitions.

Whereas for physicists this work is a promising way to find order parameters of systems where they are hard to identify, computer scientists and machine learning researchers might find an interpretation of the latent parameters.

## II. MODELS

### A. Ising Model in 2d

The Ising model is one of the most-studied and well-understood models in physics. Whereas the one-dimensional Ising model does not possess a phase transition, the two-dimensional model does. The Hamiltonian of the Ising model on the square lattice with vanishing external magnetic  $h$  field reads

$$H = H(\mathbf{S}) = -J \sum_{\langle i, j \rangle_{nn}} s_i s_j , \quad (1)$$

with uniform interaction strength  $J$  and discrete spins  $s_i \in \{+1 = \uparrow, -1 = \downarrow\}$  on each site  $i = 1, \dots, N$ . The notation  $\langle i, j \rangle_{nn}$  indicates a summation over nearest neighbors. A spin configuration  $\mathbf{S} = (s_1, \dots, s_N)$  is a fixed assignment of a spin to each lattice site,  $\Lambda$  denotes the set of all possible configurations  $\mathbf{S}$ . We set the Boltzmann constant  $k_B = 1$  and the interaction strength  $J = 1$  for the ferromagnetic case and  $J = -1$  for the antiferromagnetic case. A spin configuration  $\mathbf{S}$  can be expressed in matrix form as

$$\underline{\mathbf{S}} \triangleq \begin{pmatrix} \uparrow & \downarrow & \uparrow & \dots & \uparrow \\ \downarrow & \uparrow & \uparrow & \dots & \uparrow \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \downarrow & \downarrow & \uparrow & \dots & \downarrow \end{pmatrix}_{L \times L} . \quad (2)$$

Lars Onsager solved the two dimensional Ising model in 1944 [31]. He showed that the critical temperature is  $T_c = 2/\ln(1 + \sqrt{2}) = 2.269$ .

For the purpose of this work, we assume a square lattice with length  $L = 28$  such that  $L \times L = N = 784$ , and periodic boundary conditions. We sample the Ising model using a Monte-Carlo algorithm [32] at temperatures  $T \in [0, 5]$  to generate 50000 samples in the ferromagnetic case and 10000 samples in the antiferromagnetic case. The Ising model obeys a discrete  $\mathbb{Z}_2$ -symmetry, which is spontaneously broken below  $T_c$ . The magnetization of a spin sample is defined as

$$M(\mathbf{S}) = \frac{1}{N} \sum_i s_i . \quad (3)$$

The partition function

$$Z = \sum_{\mathbf{S} \in \Lambda} \exp(-H(\mathbf{S})/T) \quad (4)$$

allows us to define the corresponding order parameter. It is the expectation value of the absolute value of the

magnetization at fixed temperature

$$\langle \|M(T)\| \rangle = \frac{1}{Z} \sum_{\mathbf{S} \in \Lambda} \|M(\mathbf{S})\| \exp(-H(\mathbf{S})/T) . \quad (5)$$

Similarly, with the help of the matrix  $A_{ij} = (-1)^{i+j}$ , we define the order parameter, as the expectation value of the staggered magnetization. The latter is calculated from an element-wise product with a matrix form of the spin configurations

$$M_{st} = M(\underline{\mathbf{S}} \odot \mathbf{A}) . \quad (6)$$

### B. XY Model in 3d

The Mermin-Wagner-Hohenberg theorem [33, 34] prohibits continuous phase transitions in  $d \leq 2$  dimensions at finite temperature when all interactions are sufficiently short-ranged. Hence, we choose the XY model in three dimensions as a model to probe the ability of a variational autoencoder to classify phases of models with continuous symmetries. The Hamiltonian of the XY model reads

$$H(\mathbf{S}) = -J \sum_{\langle i, j \rangle_{nn}} \mathbf{s}_i \cdot \mathbf{s}_j , \quad (7)$$

with spins on the one-sphere  $\mathbf{s}_i \in \mathbb{R}^2$ ,  $\|\mathbf{s}_i\| = 1$ . Employing  $J = 1$ , the transition temperature of this model is  $T_c = 2.2017$  [35]. Using a cubic lattice with  $L = 14$ , such that  $N = L^3 = 2744$ , we perform Monte-Carlo simulations to create 10000 independent sample spin configurations in the temperature range of  $T \in [0, 5]$ . The order parameter is defined analogously to the Ising model magnetization (5), but with the  $L^2$ -norm of a magnetization consisting of two components.

## III. METHODS

*Principal component analysis* [36] is an orthogonal linear transformation of the data to an ordered set of variables, sorted by their variance. The first variable, which has the largest variance, is called the first principal component, and so on. The linear function  $\langle \cdot, \mathbf{w} \rangle$ , which maps a collection of spin samples  $(\mathbf{S}_{(1)}, \dots, \mathbf{S}_{(n)})$  to its first principal component, is defined as

$$\arg \max_{\|\mathbf{w}\|=1} \left[ \sum_j ((\mathbf{S}_{(j)} - \mu) \cdot \mathbf{w})^2 \right] , \quad (8)$$

where  $\mu$  is the vector of mean values of each spin averaged over the whole dataset. Further principal components are obtained by subtracting the already calculated principal components and repeating (8).

*Kernel principal component analysis* [26] projects the data into a kernel space in which the principal component

analysis is then performed. In this work the nonlinearity is induced by a radial basis functions kernel.

*Traditional neural network-based autoencoders consist of two artificial neural networks stacked on top of each other.* The encoder network is responsible for encoding the input data into some latent variables. The decoder network is used to decode these parameters in order to return an accurate recreation of the input data, shown in Fig. 1. The parameters of this algorithm are trained by performing gradient descent updates in order to minimize the reconstruction loss (reconstruction error) between input data and output data.

*Variational autoencoders* are a modern version of autoencoders which impose additional constraints on the encoded representations, see latent variables in Fig. 1. These constraints transform the autoencoder to an algorithm that learns a latent variable model for its input data. Whereas the neural networks of traditional autoencoders learn an arbitrary function to encode and decode the input data, variational autoencoders learn the parameters of a probability distribution modeling the data. After learning the probability distribution, one can sample parameters from it and then let the encoder network generate samples closely resembling the training data.

To achieve this, variational autoencoders employ the assumption that one can sample the input data from a unit Gaussian distribution of latent parameters. The weights of the model are trained by simultaneously optimizing two loss functions, a reconstruction loss and the Kullback-Leibler divergence between the learned latent distribution and a prior unit Gaussian.

In this work we use autoencoders and variational autoencoders [37] with one fully connected hidden layer in the encoder as well as one fully connected hidden layer in the decoder, each consisting of 256 neurons. The number of latent variables is chosen to match the model from which we sample the input data. The activation functions of the intermediate layers are rectified linear units. The activation function of the final layer is a *sigmoid* in order to predict probabilities of spin  $\uparrow$  or  $\downarrow$  in the Ising model, or *tanh* for predicting continuous values of spin components in the XY model. We do not employ any  $L^1$ ,  $L^2$  or Dropout regularization. However, we tune the relative weight of the two loss functions of the variational autoencoder to fit the problem at hand. The Kullback-Leibler divergence of the variational autoencoder can be regarded as regularization of the traditional autoencoder. In our autoencoder the reconstruction loss is the cross-entropy loss between the input and output probability of discrete spins, as in the Ising model. The reconstruction loss is the mean-squared-error between the input and the output data of continuous spin variables in the XY model.

To understand why a variational autoencoder can be a suitable choice for the task of classifying phases, we recall what happens during training. The weights of the autoencoder learn two things: on the one hand, they learn to encode the similarities of all samples to allow for an

efficient reconstruction. On the other hand, they learn a latent distribution of the parameters which encode the most information possible to distinguish between different input samples.

Let us translate these considerations to the physics of phase transitions. If all the training samples are in the unordered phase, the autoencoder learns the common structure of all samples. The autoencoder fails to learn any random entropy fluctuations, which are averaged out over all data points. However, in the ordered phase there exists a common order in samples belonging into the same phase. This common order translates to a nonzero latent parameter, which encodes correlations on each input sample. It turns out that in our cases this parameter is the order parameter corresponding to the broken symmetry. It is not necessary to find a perfect linear transformation between the order parameter and the latent parameter as it is the case in Fig. 3. A one-to-one correspondence is sufficient, such that one is able to define a function that maps these parameters onto each other and captures all discontinuities of the derivatives of the order parameter.

We point out similarities between principal component analysis and autoencoders. Although both methods seem very different, they both share common characteristics. Principal component analysis is a dimensionality reduction method which finds the linear projections of the data that maximizes the variance. Reconstructing the input data from its principal components minimizes the mean squared reconstruction error. Although the principal components do not need to follow a Gaussian distribution, principal components have the highest mutual agreement with the data if it emerges from a Gaussian prior. Moreover, a single layer autoencoder with linear activation functions closely resembles principal component analysis [29]. principal component analysis is much easier to apply and in general uses less parameters than autoencoders. However, it scales very badly to a large dataset. Autoencoders based on convolutional layers can reduce the number of parameters. In extreme cases this number can be even less than the parameters of principal component analysis. Furthermore, such autoencoders can promote locality of features in the data.

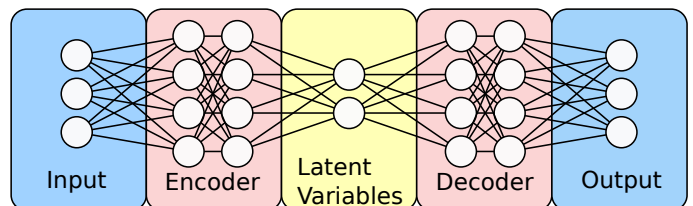


FIG. 1: Neural network architecture

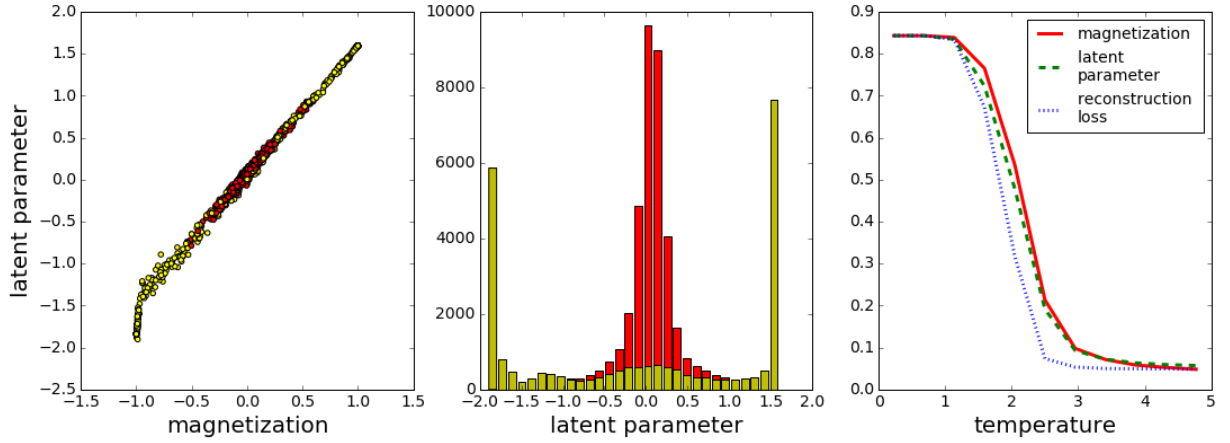


FIG. 2: Ferromagnetic Ising model. Left: correlation between latent parameter and magnetization for each spin sample. Red dots indicate points in the unordered phase, while yellow dots indicate points in the ordered phase. Middle: histogram of occurrences of latent parameters. Red bars correspond to data of the unordered phase, yellow bars correspond to the ordered phase. Right: for each absolute value of magnetization, absolute value of latent parameter and cross-entropy reconstruction loss: average at fixed temperature. The reconstruction loss is mapped on the  $T = 0$  and  $T = 5$  value of the magnetization, the latent parameter is rescaled to the magnetization at  $T = 0$ .

## IV. RESULTS

### A. Ising Model

The four different algorithms can be applied to the Ising model to determine the role of the first principal components or the latent parameters. Fig. 3 shows a clear correlation between these parameters and the magnetization for all four methods. However, the traditional autoencoder is inaccurate; this fact leads us to enhancing traditional autoencoders to variational autoencoders. The principal component methods show the most accurate results, slightly better than the variational autoencoder. This is to be expected, since the former are modeled by fewer parameters.

In the following results section, we concentrate on the variational autoencoder as the most advanced algorithm for unsupervised learning.

To begin with, we choose the number of latent parameters in the variational autoencoder to be one. After training for 50 epochs and a saturation of the training loss, we visualize the results in Fig. 2. On the left, we see a close linear correlation between the latent parameter and the magnetization. In the middle we see a histogram of encoded spin configurations into their latent parameter. The model learned to classify the configurations into three clusters. Having identified the latent parameter to be a close approximation to the magnetization  $M(S)$  allows us to interpret the properties of the clusters. The right and left clusters in the middle image correspond to an average magnetization of  $M(S) \approx \pm 1$ , while the middle cluster corresponds to the magnetization  $M(S) \approx 0$ . Employing a different viewpoint, from Fig. 2 we conclude that the parameter which holds the most information on how to distinguish Ising spin samples is the order pa-

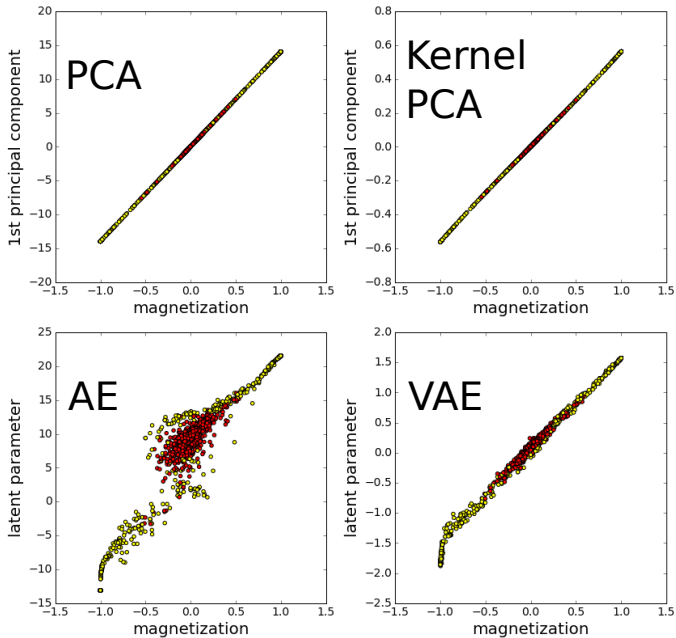


FIG. 3: Ferromagnetic Ising Model. Principal components and latent representations versus magnetization for different algorithms. PCA - principal component analysis, Kernel PCA - kernel principal component analysis, AE - autoencoder, VAE - variational autoencoder.

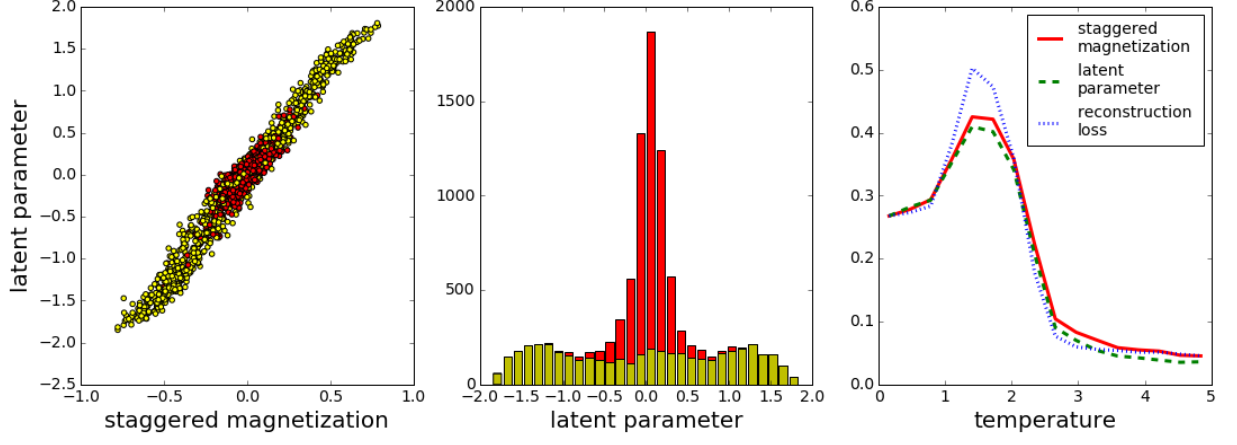


FIG. 4: Antiferromagnetic Ising Model. Left: correlation between latent parameter and staggered magnetization for each spin sample. Red dots indicate points in the unordered phase, while yellow dots indicate points in the ordered phase. Middle: histogram of occurrences of latent parameters. Red bars correspond to data of the unordered phase, yellow bars correspond to the ordered phase. Right: for each absolute value of staggered magnetization, absolute value of latent parameter and cross-entropy reconstruction loss: average at fixed temperature. The reconstruction loss is mapped on the  $T = 0$  and  $T = 5$  value of the staggered magnetization, the latent parameter is rescaled to the magnetization at  $T = 0$ .

parameter. On the right panel, the average of the magnetization, the latent parameter and the reconstruction loss are shown as a function of the temperature. A sudden change in the magnetization at  $T_c \approx 2.269$  defines the phase transition between paramagnetism and ferromagnetism. Even without knowing this order parameter, we can now use the results of the autoencoder to infer the position of the phase transition. As an approximate order parameter, the average absolute value of latent parameter also shows a steep change at  $T_c$ . The averaged reconstruction loss also changes drastically at  $T_c$  during a phase transition. While the latent parameter is different for each physical model, the reconstruction loss can be used as a universal parameter to identify phase tran-

sitions. To summarize, without any knowledge of the Ising model and its order parameter, but sample configurations, we can find a good estimation for its order parameter and the occurrence of a phase transition.

It is a priori not clear how to determine the number of latent neurons in the creation of the neural network of the autoencoder. Due to the lack of theoretical groundwork, we find the optimal number by experimenting. If we expand the number of latent dimensions by one, see Fig. 5, the results of our analysis only change slightly. The second parameter contains a lot less information compared to the first, since it stays very close to zero. Hence, for the Ising model, one parameter is sufficient to store most

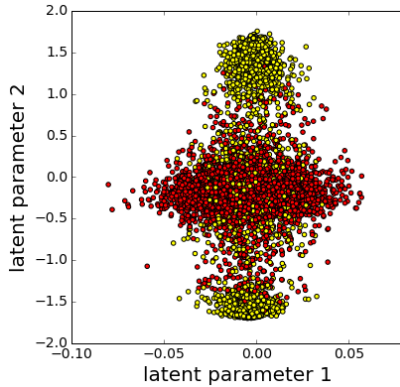


FIG. 5: Ferromagnetic Ising Model. Visualization of data in a two dimensional latent space. Red dots indicate points in the unordered phase, while yellow dots correspond to the ordered phase. The axis for parameter 1 has a smaller range than the axis for parameter 2.

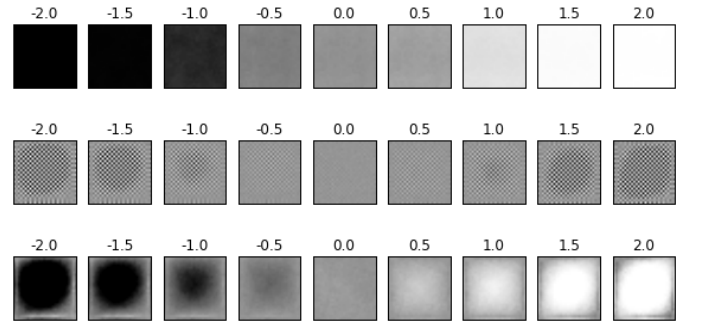


FIG. 6: Reconstruction of images, each consisting of  $28 \times 28$  pixels, from the latent parameter. The brightness indicates the probability of the spin to be up (white:  $p(\uparrow) = 1$ , black:  $p(\downarrow) = 1$ ). The first row is a reconstruction of sample configurations from the ferromagnetic Ising model. The second row corresponds to the antiferromagnetic Ising model. The third row is the prediction from the AF latent parameter, where each second spin is multiplied by  $-1$ , to show that the second row indeed predicts an antiferromagnetic state.



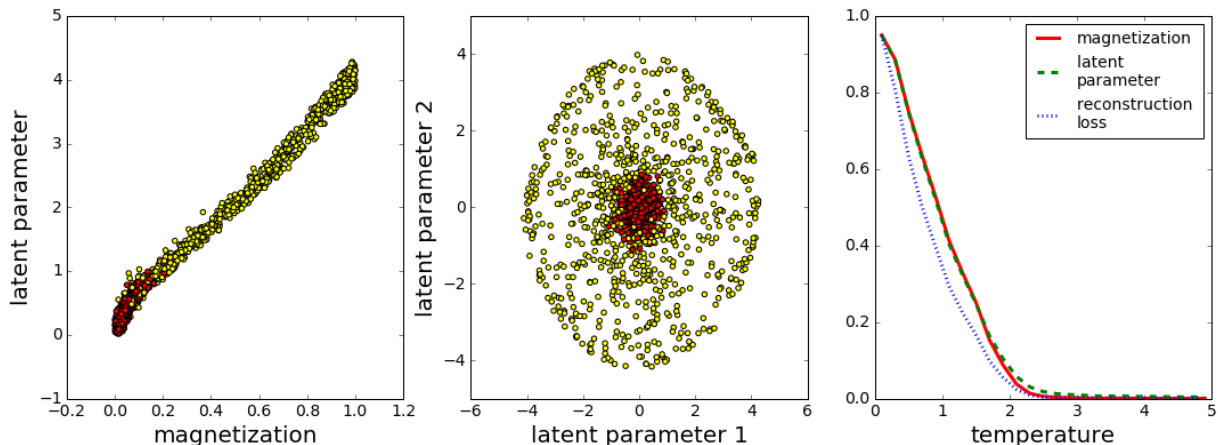


FIG. 7: XY Model. Left: correlation between  $L^2$ -norm of latent parameter vector and  $L^2$ -norm of magnetization for each spin sample. Red dots indicate points in the unordered phase, while yellow dots indicate points in the ordered phase. Middle: representation in two-dimensional latent space. Right: for each  $L^2$ -norm of the magnetization,  $L^2$ -norm of latent parameter and average of the square root of the mean squared error reconstruction loss: average at fixed temperature. The reconstruction loss is mapped on the  $T = 0$  and  $T = 5$  value of the magnetization, the latent parameter is rescaled to the magnetization at  $T = 0$ .

of the information of the latent representation.

While the ferromagnetic Ising model serves as an ideal starting ground, in the next step we are interested in models where different sites in the samples contribute in a different manner to the order parameter. We do this in order to show that our model is even sensitive to structure on the smallest scales. For the magnetization in the ferromagnetic Ising model, all spins contribute with the same weight. In contrast, in the antiferromagnetic Ising model, neighboring spins contribute with opposite weight to the order parameter (6).

Again the variational autoencoder manages to capture the traditional order parameter. The staggered magnetization is strongly correlated with the latent parameter, see Fig. 4. The three clusters in the latent representation make it possible to interpret different phases. Furthermore, we see that all three averaged quantities - the magnetization, the latent parameter and the reconstruction loss - can serve as indicators of a phase transition.

Fig. 6 demonstrates the reconstruction from the latent parameter. In the first row we see the reconstruction from samples of the ferromagnetic Ising model, the latent parameter encodes the whole spin order in the ordered phase. Reconstructions from the antiferromagnetic Ising model are shown in the second and third row. Since the reconstructions clearly show an antiferromagnetic phase, we infer that the autoencoder encodes the spin samples even to the most microscopic level.

### B. XY Model

In the XY model we examine the capabilities of a variational autoencoder to encode models with continuous symmetries. In models like the Ising model, where dis-

crete symmetries are present, the autoencoder only needs to learn a discrete set, which is often finite, of possible representations of the symmetry broken phase. If a continuous symmetry is broken, there are infinitely many possibilities of how the ordered phase can be realized. Hence, in this section we test the ability of the autoencoder to embed all these different realizations into latent variables.

The variational autoencoder handles this model equally well as the Ising model. We find that two latent parameters model the phase transition best. The latent representation in the middle of Fig. 7 shows the distribution of various states around a central cluster. The radial symmetry in this distribution leads to the assumption that a sensible order parameter is constructed from the  $L^2$ -norm of the latent parameter vector. In Fig. 7, one sees the correlation between the magnetization and the absolute value of the latent parameter vector. Averaging the samples for the same temperature hints to the facts that the latent parameter and the reconstruction loss can serve as an indicator for the phase transition.

## V. CONCLUSION

We have shown that it is possible to observe phase transitions using unsupervised learning. We compared different unsupervised learning algorithms ranging from principal component analysis to variational autoencoders and thereby motivated the need for the upgrade of the traditional autoencoder to a variational autoencoder. The weights and latent parameters of the variational autoencoder approach are able to store information about microscopic and macroscopic properties of the underlying systems. The most distinguished latent parameters co-

incide with the known order parameters. Furthermore, we have established the reconstruction loss as a new universal indicator for phase transitions. We have expanded the toolbox of unsupervised learning algorithms in physics by powerful methods, most notably the variational autoencoder, which can handle nonlinear features in the data and scale very well to huge datasets. Using these techniques, we expect to predict unseen phases or uncover unknown order parameters, e.g. in quantum spin liquids. We hope to develop deep convolutional autoencoders which have a reduced number of parameters compared to fully connected autoencoders and can also

promote locality in feature selection. Furthermore, since there exists a connection between deep neural networks and renormalization group [38], it may be helpful to employ deep convolutional autoencoders to further expose this connection.

*Acknowledgments* We would like to thank Timo Milbich, Björn Ommer, Michael Scherer, Manuel Scherzer and Christof Wetterich for useful discussions. We thank Shirin Nkongolo for proofreading the manuscript. S.W. acknowledges support by the Heidelberg Graduate School of Fundamental Physics.

- 
- [1] LD Landau. Zur Theorie der Phasenumwandlungen II. *Phys. Z. Sowjetunion*, 11:26–35, 1937.
  - [2] L. P. Kadanoff. Scaling laws for Ising models near  $T(c)$ . *Physics*, 2:263–272, 1966.
  - [3] Kenneth G. Wilson. The renormalization group: Critical phenomena and the kondo problem. *Reviews of Modern Physics*, 47(4):773–840, oct 1975. doi:10.1103/revmodphys.47.773.
  - [4] Xiao gang Wen. *Quantum Field Theory of Many-Body Systems*. Oxford University Press, 2004.
  - [5] C. L. Kane and E. J. Mele. Z<sub>2</sub>topological order and the quantum spin hall effect. *Physical Review Letters*, 95(14), sep 2005. doi:10.1103/physrevlett.95.146802.
  - [6] P.W. Anderson. Resonating valence bonds: A new kind of insulator? *Materials Research Bulletin*, 8(2):153–160, feb 1973. doi:10.1016/0025-5408(73)90167-0.
  - [7] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4):303–314, dec 1989. doi:10.1007/bf02551274.
  - [8] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, jan 1991. doi:10.1016/0893-6080(91)90009-t.
  - [9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
  - [10] Stefano Curtarolo, Dane Morgan, Kristin Persson, John Rodgers, and Gerbrand Ceder. Predicting crystal structures with data mining of quantum calculations. *Physical Review Letters*, 91(13), sep 2003. doi:10.1103/physrevlett.91.135503.
  - [11] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical Review Letters*, 108(5), jan 2012. doi:10.1103/physrevlett.108.058301.
  - [12] Zhenwei Li, James R. Kermode, and Alessandro De Vita. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Physical Review Letters*, 114(9), mar 2015. doi:10.1103/physrevlett.114.096405.
  - [13] Erin LeDell, Prabhat, Dmitry Yu. Zubarev, Brian Austin, and William A. Lester. Classification of nodal pockets in many-electron wave functions via machine learning. *Journal of Mathematical Chemistry*, 50(7): 2043–2050, may 2012. doi:10.1007/s10910-012-0019-5.
  - [14] G. Pilińska, J. E. Gubernatis, and T. Lookman. Structure classification and melting temperature prediction in octet AB solids via machine learning. *Physical Review B*, 91(21), jun 2015. doi:10.1103/physrevb.91.214302.
  - [15] Yousef Saad, Da Gao, Thanh Ngo, Scotty Bobbitt, James R. Chelikowsky, and Wanda Andreoni. Data mining for materials: Computational experiments with ABcompounds. *Physical Review B*, 85(10), mar 2012. doi:10.1103/physrevb.85.104104.
  - [16] O. S. Ovchinnikov, S. Jesse, P. Bincacchit, S. Trolier-McKinstry, and S. V. Kalinin. Disorder identification in hysteresis data: Recognition analysis of the random-bond-random-field ising model. *Physical Review Letters*, 103(15), oct 2009. doi:10.1103/physrevlett.103.157203.
  - [17] Louis-François Arsenault, Alejandro Lopez-Bezanilla, O. Anatole von Lilienfeld, and Andrew J. Millis. Machine learning for many-body physics: The case of the anderson impurity model. *Physical Review B*, 90(15), oct 2014. doi:10.1103/physrevb.90.155136.
  - [18] John C. Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke. Finding density functionals with machine learning. *Physical Review Letters*, 108(25), jun 2012. doi:10.1103/physrevlett.108.253002.
  - [19] Geoffroy Hautier, Christopher C. Fischer, Anubhav Jain, Tim Mueller, and Gerbrand Ceder. Finding nature’s missing ternary oxide compounds using machine learning and density functional theory. *Chemistry of Materials*, 22(12):3762–3767, jun 2010. doi:10.1021/cm100795d.
  - [20] J. Carrasquilla and R. G. Melko. Machine learning phases of matter. *ArXiv e-prints*, May 2016.
  - [21] G. Kasieczka, T. Plehn, M. Russell, and T. Schell. Deep-learning top taggers or the end of qcd? *ArXiv e-prints*, January 2017.
  - [22] G. Carleo and M. Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, February 2017.
  - [23] Giacomo Torlai and Roger G. Melko. Learning thermodynamics with boltzmann machines. *Physical Review B*, 94(16), oct 2016. doi:10.1103/physrevb.94.165134.
  - [24] Lei Wang. Discovering phase transitions with unsupervised learning. *Physical Review B*, 94(19), nov 2016. doi:10.1103/PhysRevB.94.195105.
  - [25] Evert P. L. van Nieuwenburg, Ye-Hua Liu, and Sebastian D. Huber. Learning phase transitions by confusion. *Nature Physics*, feb 2017. doi:10.1038/nphys4037.
  - [26] Bernhard Scholkopf, Alexander Smola, and Klaus-Robert

- Müller. Kernel principal component analysis. In *Advances in kernel methods*, pages 327–352. MIT Press, 1999.
- [27] Hervé Bouillard and Yves Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics*, 59(4):291–294, 1988.
- [28] Geoffrey E. Hinton and Richard S. Zemel. Autoencoders, minimum description length and helmholtz free energy. *NIPS*, 1993.
- [29] Pierre Baldi and Kurt Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks*, 2(1):53–58, jan 1989. doi:10.1016/0893-6080(89)90014-2.
- [30] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *ICML*, December 2013.
- [31] Lars Onsager. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Physical Review*, 65(3-4):117–149, feb 1944. doi:10.1103/physrev.65.117.
- [32] Nicholas Metropolis and S. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341, sep 1949. doi:10.1080/01621459.1949.10483310.
- [33] N. D. Mermin and H. Wagner. Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic heisenberg models. *Physical Review Letters*, 17(22):1133–1136, nov 1966. doi:10.1103/physrevlett.17.1133.
- [34] P. C. Hohenberg. Existence of long-range order in one and two dimensions. *Physical Review*, 158(2):383–386, jun 1967. doi:10.1103/physrev.158.383.
- [35] Aloysius P. Gottlob and Martin Hasenbusch. Critical behaviour of the 3d XY-model: a monte carlo study. *Physica A: Statistical Mechanics and its Applications*, 201(4): 593–613, dec 1993. doi:10.1016/0378-4371(93)90131-m.
- [36] Karl Pearson. LIII.on lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572, nov 1901. doi:10.1080/14786440109462720.
- [37] Francois Chollet. Building autoencoders in keras, May 2014. URL <https://blog.keras.io/building-autoencoders-in-keras.html>.
- [38] P. Mehta and D. J. Schwab. An exact mapping between the variational renormalization group and deep learning. *ArXiv e-prints*, October 2014.