# Probing Hidden Spin Order with Interpretable Machine Learning

Jonas Greitemann,[1] Ke Liu,[1,*] and Lode Pollet[1]

[1] *Arnold Sommerfeld Center for Theoretical Physics, University of Munich, Theresienstrasse 37, 80333 Munich, Germany*

(Dated: October 30, 2019)

Machine learning shows promise for improving our understanding of many-body problems. Tackling an unsolved problem, or classifying intricate phases, remains however a daunting task. Building on a recently introduced interpretable supervised learning scheme, we introduce a generic protocol to probe and identify nonlinear orientational spin order. We extract the tensorial order parameter up to rank 6. Moreover, we find that our approach yields reliable results already for a modest amount of training data and without knowledge of the exact transition temperature. Our approach may prove useful for identifying novel spin order and ruling out spurious spin liquid candidates.

The statistical learning of phases is nowadays an active field of research [1–15]. Despite the enormous recent progress, most studies focused on benchmarking state-of-the-art algorithms against well-understood (regimes of) physical models thereby typically identifying a local and linear order parameter. However, intriguing systems often feature complicated orders which are commonly composed of nonlinear combinations of elementary fields.

Frustrated classical Heisenberg spin systems are canonical examples of this kind [16]. These systems accommodate various symmetry-broken spin-nematic phases [17–28] and topological spin liquid phases [29–34]. The former phases develop exotic multipolar orders, typically breaking the $O(3)$ symmetry. These orders are usually referred to as hidden order, as they display no (staggered) magnetic momentum and elude the measurement techniques of conventional magnetic order. This feature not only makes their detection more subtle, it also results in confusion with spin liquids whose identification often relies on ruling out symmetry-broken order. Indeed, there have been a number of instances where a state proposed as a spin liquid was later found to possess a hidden quadrupolar or octupolar order [35–40]. To make matters worse, there are in fact myriads of such nematic orders admitted by the subgroups of $O(3)$, while their occurrence is in general hard to foresee or to exclude simply on the basis of the Hamiltonian.

In this Letter, we present a generic protocol to probe classical $O(3)$-breaking orientational spin order, requiring no *a priori* knowledge other than the possible existence of a phase transition. The building blocks of our approach are support vector machines (SVMs) [41] and an interpretable scheme, which was recently introduced by Ponte and Melko [42], upon which we construct a versatile kernel integrating both interpretability and universality. We apply the algorithm to explicitly detect a wide variety of exotic spin orders, including the tetrahedral, dodecahedral, octahedral and icosahedral orders, which go beyond the conventional Heisenberg and Néel order, and the unconventional quadrupolar order.

*Model and samples.*—There are various models that are known to accommodate a hidden quadrupolar or octupolar order, and in principle could be used to generate the training data. However, these only cover a small set of possible hidden nematic orders. To validate the generality of our approach, we instead generate the testing data samples from an effective gauge theory which allows one to simulate all possible $O(3)$-breaking orders [43]. We note that the definition of the associated order parameter depends only on the symmetry group and its representation and does not require any specification of the microscopic model nor the lattice geometry.

The gauge theory is defined by the Hamiltonian [44]

$$H = \sum_{\langle i,j\rangle} \sum_{\alpha\beta\gamma} \mathbb{J}^{\alpha\beta} \mathbf{S}_i^{\alpha} \cdot U_{ij}^{\beta\gamma} \mathbf{S}_j^{\gamma} \tag{1}$$

on a cubic lattice. At each lattice site $i$, there are three $O(3)$ spins, $\mathbf{S}_i^{\alpha} = (S_{i,x}^{\alpha}, S_{i,y}^{\alpha}, S_{i,z}^{\alpha})$, distinguished by a 'color' index $\alpha \in \{\mathrm{l}, \mathrm{m}, \mathrm{n}\}$, which locally form orthogonal triads to represent general spin orientations. Additionally, at each bond $\langle i, j\rangle$, there is a matrix gauge field, $U_{ij}$, which takes the values from a three-dimensional point group, $U_{ij} \in G$, $G \subset O(3)$. The coupling $\mathbb{J}$ is in general color-dependent, whereas it suffices to work in its isotropic limit $\mathbb{J}^{\alpha\beta} = -J\delta_{\alpha\beta}$ for our purpose. The sign of $J$ is not critical as it may be absorbed by a gauge transformation.

This model has a gauge symmetry $G$ defined by the gauge transformations $\mathbf{S}_i^{\alpha} \to \Lambda_i^{\alpha'\alpha} \mathbf{S}_i^{\alpha}$, $U_{ij}^{\alpha\beta} \to \Lambda_i^{\alpha'\alpha} U_{ij}^{\alpha\beta} \Lambda_j^{\beta'\beta}, \forall\, \Lambda_i \in G$, and a global $O(3)$ symmetry defined by global rotations $\mathbf{S}_i^{\alpha} \to \mathbf{S}_i^{\alpha} \Omega^{\alpha\alpha'}$, $\forall\, \Omega \in O(3)$. As the gauge symmetry cannot break spontaneously [45], the gauge model Eq. (1) exhibits a phase transition between a featureless disordered phase and a $G$-ordered phase. Except for a few trivial instances, this $G$-ordered phase is in general defined by highly non-linear combinations of the ordinary fields $\{\mathbf{S}^{\alpha}\}$, whereas those individually exhibit disorder. This exactly mimics spin states with a hidden order $G$. For instance, by choosing $G = D_{\infty h}$ or $G = D_3$, Eq. (1) will respectively recover the quadrupolar and triatic order with the same order parameter tensor as their realization on a triangular [37] or Kagome lattice [35, 36].

Thus, we can use the gauge model Eq. (1) to gen-

erate spin configurations representative of an arbitrary $G$-ordered phase, and use them as training data to justify our algorithm. Specifically, we perform classical Metropolis Monte Carlo simulations on Eq. (1) for a given gauge symmetry $G$ at an ensemble of temperatures above and below a phase transition, and sample the spin configuration $\{S^\alpha_{i,a}\}$. These spin configurations serve then as input data to the SVM. Unless otherwise specified, we typically perform simulations on lattices with volume $V = 16^3$ and train the machine with about $0.4 \sim 4 \times 10^5$ samples.

*Local spin cluster.*—Before proceeding with the details of the algorithm, it is useful to introduce the notion of a local spin cluster which can dramatically reduce the computational cost. Let $\mathbf{x}$ denote the full spin configuration as a vector, $\mathbf{x} = \{x_\mu\} = \{S^\alpha_{i,a}\}$, where the collective indices $(i, \alpha, a)$ are re-arranged into a single index $\mu$. The size of $\mathbf{x}$ depends linearly on the volume of the system, $\dim(\mathbf{x}) \propto V$, resulting in an empirical $\mathcal{O}(V^{2.2})$ complexity of the underlying quadratic programming problem [46]. As a large number of samples is typically generated, this scaling makes the application to large systems with massive samples prohibitively expensive.

However, as a local order parameter can be defined by a finite number of local fields, aiming at the detection of a hidden local order, we can lattice-average configuration vectors $\mathbf{x}$ or functions thereof whenever needed, up to a local cluster containing a finite number of spins. The choice of this spin cluster can often be guided by knowledge of the Hamiltonian and the lattice. In case of the gauge model Eq. (1), the three spins forming a local triad, $\mathbf{S}^\alpha_i$, are sufficient to serve as such a spin cluster. For general models it may involve multiple sites, but will still only lead to a constant complexity regarding to the lattice size.

*SVM for tensorial order parameters.*—The machine-learning problem is formulated as a standard supervised binary classification [41]. First, we classify a set of configurations, $\{\mathbf{x}^{(k)}\}$, which have been sampled from Monte Carlo simulations at temperatures $T^{(k)}$ into a disordered and an ordered class, and correspondingly assign to each sample a binary label, $y^{(k)} = \pm 1$. This is done by comparing the simulation temperature $T^{(k)}$ to a discriminatory temperature $T_{\text{disc}}$ which does not need to coincide with the critical temperature $T_c$ since the SVM is designed to be robust against misclassified data. Once the training data are collected and classified we rely on the Sequential Minimal Optimization algorithm [46] to solve the quadratic programming problem that underlies the SVM. As a result, we gain access to the decision function that classifies new test samples $\mathbf{x}$ into either class, $y$. The decision function is formally defined as

$$d(\mathbf{x}) = \sum_k \lambda_k y_k K(\mathbf{x}^{(k)}, \mathbf{x}), \qquad y = \operatorname{sgn}(d(\mathbf{x})). \tag{2}$$

Here $K(\mathbf{x}', \mathbf{x})$ is a kernel function which (implicitly) maps the raw data to an auxiliary space where the data are separable by a hyperplane. The $\lambda_k$ are essentially Lagrange multipliers and can be understood as the weight of a training sample $\mathbf{x}^{(k)}$ entering in the definition of the hyperplane (samples with $\lambda \neq 0$ are referred to as support vectors), and are learnt during training.

It was established in Ref. [42] that the converged decision function can be identified with the order parameter given that its general form is suitably constrained by the choice of the kernel. Thus, rather than just using its sign for binary classification, one can measure $d(\mathbf{x})$ as an observable in subsequent Monte Carlo simulations, and it can be analyzed to extract a corresponding analytical expression. As was demonstrated for various Ising-type models, a quadratic kernel function $K(\mathbf{x}', \mathbf{x}) = (\mathbf{x}' \cdot \mathbf{x})^2$ has a direct relation to a local order parameter: With this kernel the decision function Eq. (2) can be expressed as $d(\mathbf{x}) = \sum_{\mu\nu} C_{\mu\nu} x_\mu x_\nu$, where $C_{\mu\nu} = \sum_k \lambda_k y_k x^{(k)}_\mu x^{(k)}_\nu$ denotes a coefficient matrix. This coefficient matrix is extracted from the SVM and encodes the expression of the (squared) Ising magnetization and similar quantities. Still, the remarkable interpretability comes at the price of a limited universality of this kernel, as it may only apply to instances of linear order parameters.

However, we can amend the quadratic kernel to be universally applicable for general $O(3)$-breaking orientational orders, without losing the interpretability. This is done by observing that, mathematically, the orientation of an arbitrary order $G$ can be expressed in terms of tensors, while it is highly non-trivial to identify the explicit form of the tensor [47, 48]. Therefore, we can define a mapping,

$$\mathbf{x} = \{S^\alpha_{i,a}\} \quad \mapsto \quad \boldsymbol{\phi}(\mathbf{x}) = \{\langle S^{\alpha_1}_{a_1} \dots S^{\alpha_n}_{a_n} \rangle_{cl}\}, \tag{3}$$

which maps the raw configuration $\mathbf{x}$ to a monomial configuration $\boldsymbol{\phi}(\mathbf{x})$, where $\alpha_1, \dots, \alpha_n \in \{\mathrm{l}, \mathrm{m}, \mathrm{n}\}$ and $a_1, \dots, a_n \in \{x, y, z\}$. Note that the mapping is performed locally on spin clusters and the lattice average, $\langle \dots \rangle_{cl}$, is taken over the monomials. The problem then lies in finding the explicit coordinates $c_{\boldsymbol{\alpha}}$ of the order parameter, $\mathbb{O}^{(G)}$, in a space spanned by rank-$n$ tensors $\mathbf{S}^{\alpha_1} \otimes \dots \otimes \mathbf{S}^{\alpha_n}$:

$$\mathbb{O}^{(G)} = \sum_{\boldsymbol{\alpha}} c_{\boldsymbol{\alpha}} \mathbf{S}^{\alpha_1} \otimes \mathbf{S}^{\alpha_2} \otimes \dots \otimes \mathbf{S}^{\alpha_n}. \tag{4}$$

Naively, $\boldsymbol{\phi}(\mathbf{x})$ contains $r^n$ elements for a spin cluster with $r$ ordinary field components $S^\alpha_a$ (here $r = 9$). However, because of redundant permutations, we can effectively work with a reduced number of elements given by the multiset coefficient $\left(\!\binom{r}{n}\!\right) \coloneqq \frac{(r+n-1)!}{n!(r-1)!}$. This decreases the computational cost significantly, especially for high-rank tensors. By contrast, the interpretation of the decision function is generally easier in the redundant form. Furthermore, perhaps counterintuitively, it is more efficient to carry out the mapping $\boldsymbol{\phi}$ explicitly, rather than

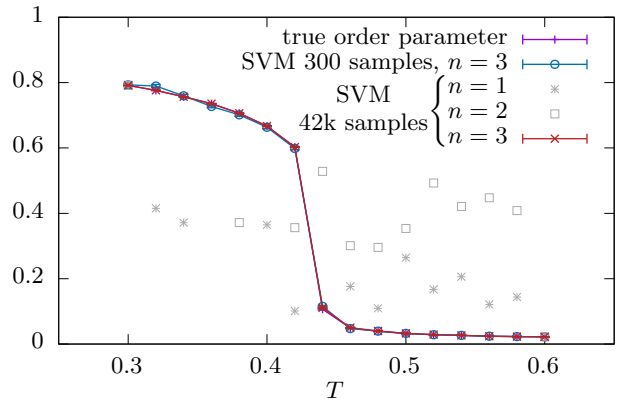



FIG. 1. The square root of the decision function, $\sqrt{d(\mathbf{x})}$, trained at different ranks for the tetrahedral order. The true order parameter curve is shown for comparison, and $d(\mathbf{x})$ has been rescaled linearly, such that their endpoints match up. Insufficient tensor ranks do not result in a meaningful order parameter.

extending the kernel trick by composing the mapping $\boldsymbol{\phi}$ with the quadratic kernel $K$. (Refer to the SM for details.)

We can then apply the aforementioned quadratic kernel to $\boldsymbol{\phi}$,

$$d(\mathbf{x}) = \sum_k \lambda_k y_k \left[\boldsymbol{\phi}(\mathbf{x}^{(k)}) \cdot \boldsymbol{\phi}(\mathbf{x})\right]^2 = \sum_{\mu\nu} C_{\mu\nu}\phi_\mu\phi_\nu, \quad (5)$$

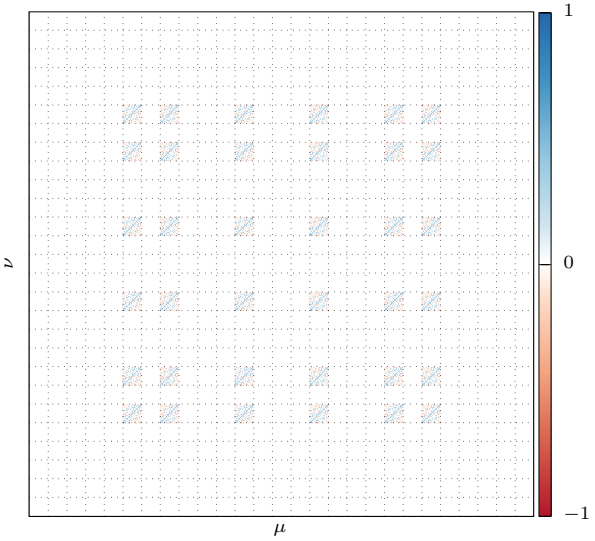
$$C_{\mu\nu} = \sum_k \lambda_k y_k \langle S^{\alpha_1}_{a_1} \dots S^{\alpha_n}_{a_n}\rangle^{(k)}_{cl} \langle S^{\alpha'_1}_{a'_1} \dots S^{\alpha'_n}_{a'_n}\rangle^{(k)}_{cl}, \quad (6)$$

where $\mu, \nu$ correspond to collective indices, $\mu = (\alpha_1, \dots, \alpha_n; a_1, \dots, a_n)$, $\nu = (\alpha'_1, \dots, \alpha'_n; a'_1, \dots, a'_n)$. Indeed, the resulting decision function $d(\mathbf{x})$ formally appears identical to the ordinary quadratic kernel, but is now capable of capturing general orientational spin orders with the same raw input $\mathbf{x} = \{S^\alpha_{i,a}\}$. Moreover, the analytical expression of the underlying order parameter is again encoded in the coefficient matrix $C_{\mu\nu}$ and can be systematically inferred.

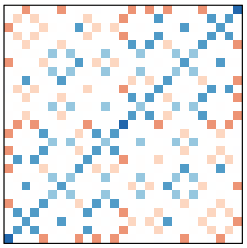
*High-rank orders.*—Here we present the results for the most intricate orders.

We start with the tetrahedral ($T_d$) order. The sample data are generated by taking $G = T_d$ in the gauge model Eq. (1), and the decision function Eq. (5) is trained with different values for the rank $n$ of $\boldsymbol{\phi}$ and with a given $T_{\text{disc}}$. We take the ideal $T_{\text{disc}} = T_c$ for now (learnt from Ref. [49]), and save the discussion of $T_{\text{disc}} \neq T_c$ for later.

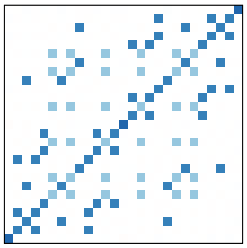
After training, the decision function obtained for different ranks $n$ is measured for new testing samples. The results are shown in Fig. 1 by plotting $\sqrt{d(\mathbf{x})}$ against the true order parameter curve where $d(\mathbf{x})$ has been linearly rescaled for comparison. Clearly, the decision function exhibits only noise for lower ranks, but reproduces the order parameter curve for $n = 3$, indicating that the order has been captured at this rank.




(a) Full $C_{\mu\nu}$ matrix



(b) Full [lmn; lmn] block

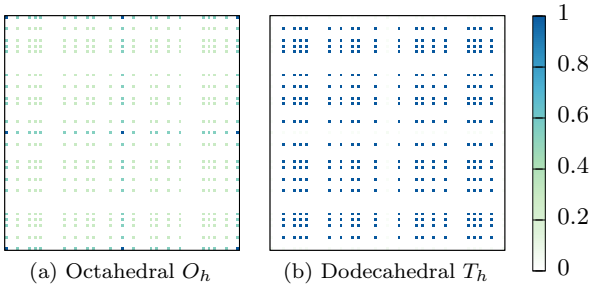


(c) Non-trivial contractions

FIG. 2. The coefficient matrix $C_{\mu\nu}$ for the tetrahedral order learnt using the rank-3 kernel. (a) Full $C_{\mu\nu}$ matrix, where the multi-indices $\mu, \nu = (\alpha_1, \alpha_2, \alpha_3; a_1, a_2, a_3)$ are lexicographically ordered. Each block is assigned coordinates $[\alpha_1\alpha_2\alpha_3; \alpha'_1\alpha'_2\alpha'_3]$. Non-trivial blocks have mutually exclusive color indices. (b) Details of the [lmn; lmn] block, in comparison with (c) where trivial contractions have been removed.

We then extract the order parameter from the corresponding $C_{\mu\nu}$. At rank 3, $\boldsymbol{\phi}$ encodes 27 tensors of the form $\mathbf{T}^{\alpha_1\alpha_2\alpha_3} = \mathbf{S}^{\alpha_1} \otimes \mathbf{S}^{\alpha_2} \otimes \mathbf{S}^{\alpha_3}$, dividing $C_{\mu\nu}$ into 27-by-27 blocks, as shown in Fig. 2(a). Each block relates to the tensor-analogue of the Frobenius inner product $\text{Tr}\left(\mathbf{T}^{\alpha_1\alpha_2\alpha_3} \cdot \mathbf{T}^{\alpha'_1\alpha'_2\alpha'_3}\right)$, and can be assigned coordinates $[\alpha_1\alpha_2\alpha_3; \alpha'_1\alpha'_2\alpha'_3]$. The underlying order parameter is consequently described by blocks with a non-trivial weight. Indeed, as can be inferred from Fig. 2(a), only blocks satisfying $\alpha_1 \neq \alpha_2 \neq \alpha_3$ and $\alpha'_1 \neq \alpha'_2 \neq \alpha'_3$ have non-vanishing entries, and the entire set of these blocks can be attributed to the Frobenius norm $\|\mathbb{O}^{(T_d)}\|^2_F := \text{Tr}\left(\mathbb{O}^{(T_d)} \cdot \mathbb{O}^{(T_d)}\right)$, with $\mathbb{O}^{(T_d)} = \sum_{\alpha_1\neq\alpha_2\neq\alpha_3} \mathbf{T}^{\alpha_1\alpha_2\alpha_3}$. $\mathbb{O}^{(T_d)}$ is exactly the tetrahedral order parameter [50], and consistently, the decision function is again related to its norm squared, $d(\mathbf{x}) \sim \|\mathbb{O}^{(T_d)}\|^2_F$, up to linear rescaling.

Elements inside a block correspond to the spin component indices. In the $T_d$ case specifically, non-vanishing blocks exhibit an identical pattern, since each $\mathbf{T}^{\alpha_1\alpha_2\alpha_3}$




(a) Octahedral $O_h$ (b) Dodecahedral $T_h$

FIG. 3. Block structure of $C_{\mu\nu}$ for (a) the octahedral and (b) the dodecahedral order, learnt using the rank-4 kernel. Each pixel corresponds to a block of $C_{\mu\nu}$, identified by coordinates $[\alpha_1\alpha_2\alpha_3\alpha_4;\alpha'_1\alpha'_2\alpha'_3\alpha'_4]$. The value of each pixel is given by the squared Frobenius norm of the corresponding block.

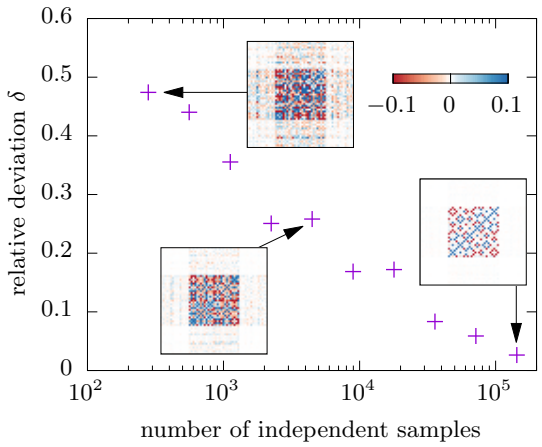



FIG. 4. Deviation $\delta$ for the tetrahedral order as a function of the number of training samples. The insets show excerpts of the coefficient tensor for selected points.

for $\alpha_1 \neq \alpha_2 \neq \alpha_3$ contributes equally in defining $\mathbb{O}^{(T_d)}$. In Fig. 2(b), we show the pattern of the [lmn; lmn] block for instance. The locations of non-vanishing elements are identified with all possible manners of contracting component indices of $T^{lmn}_{a_1a_2a_3}T^{lmn}_{a'_1a'_2a'_3}$. These include self-contractions such as $(\overbrace{a_1a_1}a_3)(\overbrace{a_3a'_2}a'_2)$ which are incompatible with the Frobenius norm, but contribute only a trivial constant to the decision function, thus not interfering with the relation $d(\mathbf{x}) \sim \|\mathbb{O}^{(T_d)}\|^2_F$. In particular, this constant is zero for the tetrahedral order due to $\mathbb{O}^{(T_d)}$ being traceless. Contractions having non-trivial contributions to $d(\mathbf{x})$ can be readily distinguished from the trivial ones, as shown in Fig. 2(c).

Regardless, the key insight here is that, although a closer look at $C_{\mu\nu}$ will serve as a consistency check, the main information of the target order parameter is already revealed by the global block structure of $C_{\mu\nu}$, hence its interpretation is rather straightforward. This also holds true for the more complicated orders. In Fig. 3, we show the results of the octahedral ($O_h$) and dodecahedral ($T_h$) order by highlighting the block structure of $C_{\mu\nu}$. For both cases, the order is learnt at rank 4, and each block is again identified by the spin color indices of the inner product $\mathrm{Tr}\big(\mathbf{T}^{\alpha_1\alpha_2\alpha_3\alpha_4}\cdot\mathbf{T}^{\alpha'_1\alpha'_2\alpha'_3\alpha'_4}\big)$. The dominant contributions stem from the blocks featuring four identical color indices ($O_h$), and two mutually exclusive pairs of identical color indices ($T_h$), respectively. Correspondingly, these give rise to the ordering tensors, $\mathbb{O}^{(O_h)} = \mathbf{T}^{\mathrm{llll}} + \mathbf{T}^{\mathrm{mmmm}} + \mathbf{T}^{\mathrm{nnnn}}$, and $\mathbb{O}^{(T_h)} = \mathbf{T}^{\mathrm{llmm}} + \mathbf{T}^{\mathrm{mmnn}} + \mathbf{T}^{\mathrm{nnll}}$ [47]. In particular, $\mathbb{O}^{(T_h)}$ is a partially symmetric tensor, and has six equivalent definitions generated by permuting its color indices. Interestingly, SVM captures all these variants exhaustively. Beyond that, the subdominant weights in Fig. 3 arise from self-contractions and a non-vanishing trace of $\mathbb{O}^G$. As in the $T_d$ case, they again only result in a trivial constant, and are not relevant for the order parameter.

We also examined the extreme case of the icosahedral ($I_h$) order. This corresponds to the largest polyhedral subgroup of $O(3)$ and requires a rank-6 tensor. Owing to the high rank, we observe a proliferation of self-contractions, which complicates the extraction of the analytical order parameter which we will leave for a future work. However, the learnt $C_{\mu\nu}$ still serves as a numerical representation of the order parameter, and its block structure reveals the $I_h$ symmetry.

*Performance.*—To quantify the performance of SVM, we introduce a deviation metric, $\delta$, defined by the element-wise discrepancy between the learnt $C_{\mu\nu}$ and the theoretical one, $\tilde{C}_{\mu\nu}$, $\delta \coloneqq \frac{\|\mathbf{C}-\tilde{\mathbf{C}}\|_F}{\|\tilde{\mathbf{C}}\|_F} \geq 0$. The tetrahedral order is taken as an example, but the general features are also valid for the other aforementioned orders.

In Fig. 4 we demonstrate the dependence of $\delta$ on the number of training samples. Aside from the convergence of $\delta$ as more training samples are accumulated, interestingly, the expected block structure of $C_{\mu\nu}$ has already emerged at as little as 300 samples, which is sufficient to infer the underlying order parameter. We emphasize that $\delta$ is a rather sensitive deviation metric. Empirically, with a deviation $\delta \approx 0.5$, the measured decision function $d(\mathbf{x})$ remains in decent agreement with the true order parameter curve (Fig. 1).

Fig. 5 shows $\delta$ against the discrepancy of the assumed $T_{\mathrm{disc}}$ from the real $T_c$. We remark the low level of error even for an estimate that is off by as much as $|\tau| \sim 40\%$, where $\tau = \frac{T_{\mathrm{disc}}-T_c}{T_c}$. This robustness of the SVM facilitates applications where the locus of the phase transition is not known *a priori*. Moreover, a crude $C_{\mu\nu}$ learnt with large $|\tau|$ can in turn guide a better estimate of $T_c$, as a well behaved $d(\mathbf{x})$ is still obtained, reminiscent of the learning-by-confusion scheme [5]. Additionally, as seen from Fig. 4, a crude $C_{\mu\nu}$ may already suffice for an appropriate inference of the potential order parameter by which one could further derive more sensitive measure-

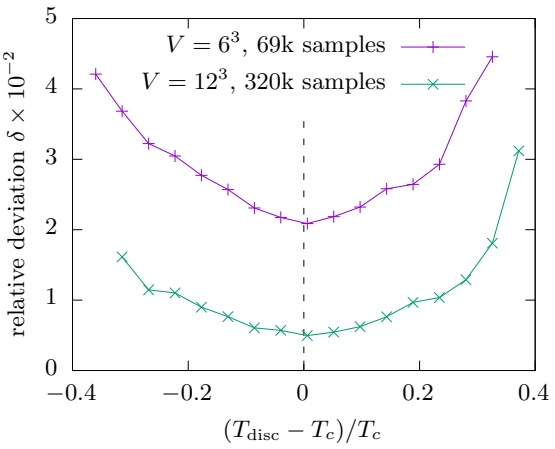



FIG. 5. Deviation $\delta$ for the tetrahedral order against different discriminatory temperatures $T_{\rm disc}$ used to classify training samples.

ments of a phase transition, such as the susceptibility and Binder cumulant.

*Concluding remarks.*—We have presented a kernel scheme for a SVM to probe and identify generic orientational order breaking the $O(3)$ symmetry. We demonstrated the power of this method by explicitly addressing the most intricate orders, and examined its robustness against uncertainty regarding the locus of the phase transition. These orders are representative of a myriad of exotic states which may potentially emerge in frustrated spin and orbital systems, in particular in the pyrochlore and three-dimensional Kitaev lattices which are often built on tetrahedral or octahedral motifs [31, 51]. However, owing to the highly nonlinear nature, those orders may not be captured by traditional techniques, and may hence be mistaken for spin liquids. We hope that our approach will prove useful in identifying novel magnetic orders and in excluding spurious spin liquids. Finally, although we exclusively work with classical Monte Carlo data, it would be interesting to apply the algorithm to path-integral representations of (frustrated) quantum magnets.

This work is supported by FP7/ERC Consolidator grant No. 771891 and the Nanosystems Initiative Munich. We would like to thank Lei Wang for enlightening discussions. Our simulations make use of the $\nu$-SVM formulation [52], the LIBSVM library [53, 54], and the ALPSCore library [55].

* ke.liu@lmu.de

[1] L.-F. m. c. Arsenault, A. Lopez-Bezanilla, O. A. von Lilienfeld, and A. J. Millis, Phys. Rev. B **90**, 155136 (2014).
[2] L. Wang, Phys. Rev. B **94**, 195105 (2016).
[3] T. Ohtsuki and T. Ohtsuki, Journal of the Physical Society of Japan **85**, 123706 (2016), https://doi.org/10.7566/JPSJ.85.123706.
[4] J. Carrasquilla and R. G. Melko, Nature Physics **13**, 431 EP (2017).
[5] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Nature Physics **13**, 435 EP (2017).
[6] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, Phys. Rev. X **7**, 031038 (2017).
[7] D.-L. Deng, X. Li, and S. Das Sarma, Phys. Rev. B **96**, 195145 (2017).
[8] Y. Zhang, R. G. Melko, and E.-A. Kim, Phys. Rev. B **96**, 245119 (2017).
[9] F. Schindler, N. Regnault, and T. Neupert, Phys. Rev. B **95**, 245134 (2017).
[10] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, Scientific Reports **7**, 8823 (2017).
[11] S. J. Wetzel and M. Scherzer, Phys. Rev. B **96**, 184410 (2017).
[12] W. Hu, R. R. P. Singh, and R. T. Scalettar, Phys. Rev. E **95**, 062122 (2017).
[13] P. Zhang, H. Shen, and H. Zhai, Phys. Rev. Lett. **120**, 066401 (2018).
[14] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, Phys. Rev. X **8**, 011006 (2018).
[15] Z. Cai and J. Liu, Phys. Rev. B **97**, 035116 (2018).
[16] C. Lacroix, P. Mendels, and F. Mila, *Introduction to Frustrated Magnetism: Materials, Experiments, Theory*, Springer Series in Solid-State Sciences (Springer Berlin Heidelberg, 2011).
[17] A. V. Chubukov, Journal of Physics: Condensed Matter **2**, 1593 (1990).
[18] J. T. Chalker, P. C. W. Holdsworth, and E. F. Shender, Phys. Rev. Lett. **68**, 855 (1992).
[19] R. Moessner and J. T. Chalker, Phys. Rev. B **58**, 12049 (1998).
[20] N. Shannon, T. Momoi, and P. Sindzingre, Phys. Rev. Lett. **96**, 027213 (2006).
[21] A. Läuchli, F. Mila, and K. Penc, Phys. Rev. Lett. **97**, 087205 (2006).
[22] C. Wu, Phys. Rev. Lett. **100**, 200406 (2008).
[23] J. Yamaura, K. Ohgushi, H. Ohsumi, T. Hasegawa, I. Yamauchi, K. Sugimoto, S. Takeshita, A. Tokuda, M. Takata, M. Udagawa, M. Takigawa, H. Harima, T. Arima, and Z. Hiroi, Phys. Rev. Lett. **108**, 247205 (2012).
[24] M. Mourigal, M. Enderle, B. Fåk, R. K. Kremer, J. M. Law, A. Schneidewind, A. Hiess, and A. Prokofiev, Phys. Rev. Lett. **109**, 027203 (2012).
[25] O. Janson, S. Furukawa, T. Momoi, P. Sindzingre, J. Richter, and K. Held, Phys. Rev. Lett. **117**, 037206 (2016).
[26] R. Wawrzyńczak, Y. Tanaka, M. Yoshida, Y. Okamoto, P. Manuel, N. Casati, Z. Hiroi, M. Takigawa, and G. J. Nilsen, Phys. Rev. Lett. **119**, 087201 (2017).
[27] A. Orlova, E. L. Green, J. M. Law, D. I. Gorbunov, G. Chanda, S. Krämer, M. Horvatić, R. K. Kremer, J. Wosnitza, and G. L. J. A. Rikken, Phys. Rev. Lett. **118**, 247201 (2017).
[28] H. Yan, O. Benton, L. Jaubert, and N. Shannon, Phys. Rev. B **95**, 094422 (2017).
[29] D. Bergman, J. Alicea, E. Gull, S. Trebst, and L. Balents, Nature Physics **3**, 487 EP (2007).
[30] L. Balents, Nature **464**, 199 EP (2010).

[31] J. S. Gardner, M. J. P. Gingras, and J. E. Greedan, Rev. Mod. Phys. **82**, 53 (2010).

[32] C. L. Henley, Annual Review of Condensed Matter Physics **1**, 179 (2010), https://doi.org/10.1146/annurev-conmatphys-070909-104138.

[33] C. Castelnovo, R. Moessner, and S. Sondhi, Annual Review of Condensed Matter Physics **3**, 35 (2012), https://doi.org/10.1146/annurev-conmatphys-020911-125058.

[34] Y. Zhou, K. Kanoda, and T.-K. Ng, Rev. Mod. Phys. **89**, 025003 (2017).

[35] M. E. Zhitomirsky, Phys. Rev. Lett. **88**, 057204 (2002).

[36] M. E. Zhitomirsky, Phys. Rev. B **78**, 094423 (2008).

[37] T. Momoi, P. Sindzingre, and N. Shannon, Phys. Rev. Lett. **97**, 257204 (2006).

[38] M. Taillefumier, O. Benton, H. Yan, L. D. C. Jaubert, and N. Shannon, Phys. Rev. X **7**, 041057 (2017).

[39] Y.-D. Li, X. Wang, and G. Chen, Phys. Rev. B **94**, 201114 (2016).

[40] Q. Luo, S. Hu, B. Xi, J. Zhao, and X. Wang, Phys. Rev. B **95**, 165110 (2017).

[41] V. Vapnik, *Statistical learning theory*, Adaptive and learning systems for signal processing, communications, and control (Wiley, 1998).

[42] P. Ponte and R. G. Melko, Phys. Rev. B **96**, 205146 (2017).

[43] K. Liu, J. Nissinen, R.-J. Slager, K. Wu, and J. Zaanen, Phys. Rev. X **6**, 041025 (2016).

[44] In general form, this Hamiltonian could also include a plaquette term describing the interaction between gauge fields. However, the plaquette term is not required to mimic nematic orders, and thereby not considered here for simplicity; for more details see Ref. [43].

[45] S. Elitzur, Phys. Rev. D **12**, 3978 (1975).

[46] J. C. Platt, in *Advances in Kernel Methods – Support Vector Learning*, edited by B. Schölkopf, C. J. C. Burges, and A. J. Smola (MIT Press, 1998).

[47] J. Nissinen, K. Liu, R.-J. Slager, K. Wu, and J. Zaanen, Phys. Rev. E **94**, 022701 (2016).

[48] L. Michel and B. Zhilinskiı, Physics Reports **341**, 11 (2001).

[49] K. Liu, J. Greitemann, and L. Pollet, Phys. Rev. E **97**, 012706 (2018).

[50] L. G. Fel, Phys. Rev. E **52**, 702 (1995).

[51] S. Trebst, arXiv preprint arXiv:1701.07056 (2017).

[52] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, Neural computation **12**, 1207 (2000).

[53] C.-C. Chang and C.-J. Lin, Neural computation **13**, 2119 (2001).

[54] C.-C. Chang and C.-J. Lin, ACM Trans. Intell. Syst. Technol. **2**, 27:1 (2011).

[55] A. Gaenko, A. Antipov, G. Carcassi, T. Chen, X. Chen, Q. Dong, L. Gamper, J. Gukelberger, R. Igarashi, S. Iskakov, M. Knz, J. LeBlanc, R. Levy, P. Ma, J. Paki, H. Shinaoka, S. Todo, M. Troyer, and E. Gull, Computer Physics Communications **213**, 235 (2017).

# — Supplemental Materials —

# Probing Hidden Spin Order with Interpretable Machine Learning

Jonas Greitemann, Ke Liu, and Lode Pollet

## REDUNDANCY IN THE MONOMIAL MAPPING

The explicit mapping of the configuration vector $\mathbf{x}$ to monomials $\boldsymbol{\phi}(\mathbf{x})$ introduced in Eq. (3) of the main text, comes with the downside that $\dim \boldsymbol{\phi}(\mathbf{x})$ depends on the tensor rank $n$ which needs to be fixed when sampling configurations from the Monte Carlo simulation. This is to be contrasted with the alternative strategy of sampling the raw $\mathbf{x}$ and applying a composite kernel $\tilde{K}(\mathbf{x}', \mathbf{x}) = (\boldsymbol{\phi}(\mathbf{x}') \cdot \boldsymbol{\phi}(\mathbf{x}))^2$ at training time.

Table S1 explicitly demonstrates the growth of the configuration vector with rank. In particular, by eliminating redundant elements, i.e. including the monomial

$$\langle S^{\alpha_1}_{a_1} S^{\alpha_2}_{a_2} \dots S^{\alpha_n}_{a_n} \rangle_{cl} \tag{S1}$$

in $\boldsymbol{\phi}(\mathbf{x})$ if and only if $(\alpha_1, a_1) \leq (\alpha_2, a_2) \leq \cdots \leq (\alpha_n, a_n)$ with some arbitrary ordering imposed on color-component tuples, the dimension of $\boldsymbol{\phi}(\mathbf{x})$ is given by the multiset coefficient

$$\dim \boldsymbol{\phi}(\mathbf{x}) = \left(\!\!\binom{r}{n}\!\!\right) = \binom{r+n-1}{n} = \frac{(r+n-1)!}{n!(r-1)!}. \tag{S2}$$

By contrast, $\dim \mathbf{x} = rV$ is extensive and will outgrow $\dim \boldsymbol{\phi}(\mathbf{x})$ for lattices as small as $L \geq 7$ even when the rank-6 mapping is used. Thus, the mapping $\boldsymbol{\phi}$ actually reduces the dimension and it is therefore prudent to rely on the kernel trick only for the implicit mapping to a higher-dimensional space due to the quadratic kernel.

For the interpretation of the coefficient tensor, we find it more beneficial to include redundant elements to avoid obfuscating the block structure discussed in the main text. The multiplicity, i.e. the number of equivalent permutations, is given by the multinomial coefficients

$$m_{(\alpha_1,a_1)\dots(\alpha_n,a_n)} = \binom{n}{k_1, k_2, \dots} = \frac{n!}{k_1! k_2! \dots}, \tag{S3}$$

where $k_1 + k_2 \cdots + k_r = n$ count the occurrences of each of the $r$ possible index values. We include the square roots of the multiplicities in the configuration, e.g. at rank 2,

$$\boldsymbol{\phi}(\mathbf{x}) = \{\sqrt{m_{\dots}} \langle S^{\alpha_1}_{a_1} S^{\alpha_2}_{a_2} \rangle_{cl} \mid (\alpha_1, a_1) \leq (\alpha_2, a_2)\}. \tag{S4}$$

| rank $n$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $r^n$ | 9 | 81 | 729 | 6561 | 59049 | 531441 |
| $\left(\!\!\binom{r}{n}\!\!\right)$ | 9 | 45 | 165 | 495 | 1287 | 3003 |

TABLE S1. Dimensions of the configuration vector $\boldsymbol{\phi}(\mathbf{x})$ before and after eliminating redundant monomials. $r = 9$ is the range of the spin indices $(\alpha, a)$ (3 colors, 3 components).

That way, when using the above configuration in conjunction with the quadratic kernel, we learn the same decision function that one would have gotten if all $r^n$ monomials had been considered regardless of their redundancy.

## REGULARIZATION PARAMETER

SVMs involve a regularization parameter $C$ and its choice, as it applies to phase classification, has been discussed previously [42]. In principle, one has to validate the learnt model with respect to independent test data for different values of $C$ which can span many orders of magnitude.

There exists however an alternative reparametrization of the SVM optimization problem in terms of a regularization parameter $\nu \in [0, 1)$ which has been shown to impose a lower bound on the fraction of training samples that serve as support vectors [52]. $\nu$-SVM thus admits a more universal interpretation and we found it to simplify the selection of an appropriate regularization.

For the present work, we found a stronger regularization in terms of $\nu$ to improve the quality of the learnt order parameter as demonstrated in Fig. S1 for the tetrahedral order. This is consistent with the fact the ensembles of micro-states in either phase near the transition temperature have a significant overlap. Thus, we picked a rather large value of $\nu = 0.6$ for the data presented in Figs. 1-4, and $\nu = 0.4$ in Fig. 5 to allow for more imbalanced training data [54].

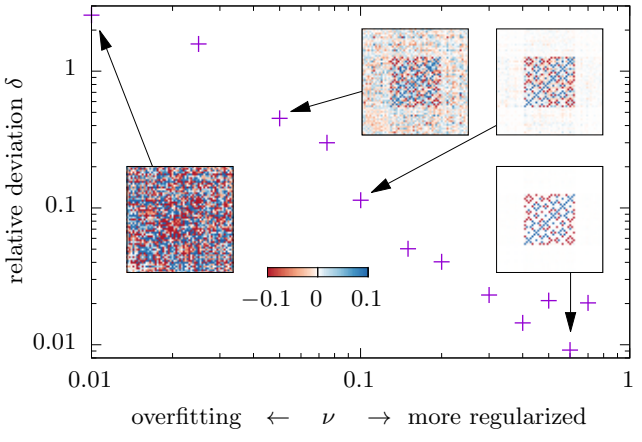



FIG. S1. Deviation $\delta$ for the tetrahedral order, computed with different levels of regularization $\nu$. The insets show excerpts of the coefficient tensor for selected points.