COMPUTER SCIENCE

Optimal structure and parameter learning of Ising models

Andrey Y. Lokhov,[1,2]* Marc Vuffray,[2] Sidhant Misra,[3] Michael Chertkov[1,2,4]

Reconstruction of the structure and parameters of an Ising model from binary samples is a problem of practical importance in a variety of disciplines, ranging from statistical physics and computational biology to image processing and machine learning. The focus of the research community shifted toward developing universal reconstruction algorithms that are both computationally efficient and require the minimal amount of expensive data. We introduce a new method, interaction screening, which accurately estimates model parameters using local optimization problems. The algorithm provably achieves perfect graph structure recovery with an information-theoretically optimal number of samples, notably in the low-temperature regime, which is known to be the hardest for learning. The efficacy of interaction screening is assessed through extensive numerical tests on synthetic Ising models of various topologies with different types of interactions, as well as on real data produced by a D-Wave quantum computer. This study shows that the interaction screening method is an exact, tractable, and optimal technique that universally solves the inverse Ising problem.

INTRODUCTION

The Ising model is a renowned model in statistical physics that was originally introduced to study the phase transition phenomenon in ferromagnetic materials (1). In modern applications, the Ising model is regarded as the most general graphical model describing stationary statistics of binary variables, called spins, that admit a pairwise factorization. The spins are associated with the nodes of a graph, and the edges specify pairwise interactions. Given a graph $G = (V, E)$, where V is the set of N nodes and E is the set of edges, the probability measure of an Ising model reads

$$P_{J^*, H^*}(\boldsymbol{\sigma}) = \frac{1}{Z} \exp\left(\sum_{(i,j) \in E} J_{ij}^* \sigma_i \sigma_j + \sum_{i \in V} H_i^* \sigma_i \right) \quad (1)$$

where $\boldsymbol{\sigma} = \{\sigma_i\}_{i \in V}$ denotes the vector of spin variables $\sigma_i \in \{-1, +1\}$, $\boldsymbol{J}^* = \{J_{ij}^*\}_{(i,j) \in E}$ is the vector of pairwise interactions, $\boldsymbol{H}^* = \{H_i^*\}_{i \in V}$ is the vector of magnetic fields, and Z, called the partition function, is a normalization factor that ensures $\sum_{\boldsymbol{\sigma}} P_{J^*, H^*}(\boldsymbol{\sigma}) = 1$. In this representation, the temperature is absorbed in $\boldsymbol{J}^*$ and $\boldsymbol{H}^*$. Regimes corresponding to small and large interactions and magnetic field intensities are respectively known as high- and low-temperature phases. Models in which couplings or magnetic fields are positive, negative, or have mixed signs are traditionally referred to as ferromagnet, antiferromagnet, and spin glass, respectively. In numerous application fields, such as statistical physics (2, 3), neuroscience (4, 5), biopolymers (6), gene regulatory networks (7), quantum computing (8), image segmentation (9), deep learning (10), and sociology (11), the underlying interaction graph and the values of couplings are often unknown a priori and have to be reconstructed from the data, which takes the form of several observed spin configurations. The learning problem that we consider in this paper, called the inverse Ising problem, is stated as follows: given M statistically independent samples $\{\boldsymbol{\sigma}^{(m)}\}_{m=1,\dots,M}$ generated by an unknown prob-

ability measure $P_{J^*, H^*}(\boldsymbol{\sigma})$, reconstruct the interaction graph G and the parameters $\{\boldsymbol{J}^*, \boldsymbol{H}^*\}$.

Over the past several decades, a considerable number of techniques have been developed in statistical physics, machine learning, and computer science communities to carry out this reconstruction task (12–24). A direct maximization of the log-likelihood of the data is generally intractable because it requires a repeated evaluation of the partition function Z for different trial values of the parameters $\{\boldsymbol{J}, \boldsymbol{H}\}$. Computing Z is, in general, a task of exponential complexity in the number of spins (25), with the exception of some special cases such as tree-structured Ising models (26) and planar Ising models with zero magnetic fields (27). Despite this difficulty, one may still try to maximize the log-likelihood using, for instance, Monte Carlo simulations, as done in the study by Ackley et al. (12) via the so-called learning for Boltzmann machines. In this method, one estimates all the magnetizations and pairwise correlation functions from samples and then maximizes the log-likelihood using a gradient ascent procedure over all couplings and magnetic fields. The Monte Carlo nature of the method makes it exponentially expensive in the number of runs required to achieve a predefined accuracy. Note, however, that this method is asymptotically exact as the number of samples goes to infinity, thus illustrating that "sufficient statistics"–based approaches that use only estimates of first moments and pair correlations of spins can achieve exact reconstruction albeit through computations with exponential complexity (28).

Following the observation that the first and second moments are sufficient to reconstruct Ising models, a number of mean-field approximations have been suggested to circumvent the difficulty of an analytical evaluation of magnetizations and pair correlation functions [see the study by Roudi et al. (14) for a review]. The applicability of these methods is limited: They perform weakly on systems embedded in a low-dimensional space and in the spin glass regime, where fluctuations are important and cannot be neglected. Some of the limitations of these naïve mean-field methods (13) are addressed in more advanced mean-field methods: The small correlation expansion (15) considers corrections to the mean-field in the high-temperature regime; Nguyen and Berg (16) exploits clustering of samples in the configuration space according to their mutual overlaps; and the Bethe approximation (17) is based on the tree-like approximation of the interaction graph. Nevertheless, the applicability of these approximate

[1]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [2]Theoretical Division T-4, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [3]Theoretical Division T-5, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [4]Skolkovo Institute of Science and Technology, 143026 Moscow, Russia.
*Corresponding author. Email: lokhov@lanl.gov

techniques remains limited to Ising models pertaining to specific classes.

Although sufficient statistics consisting of the first and second moments of the data carry all the information needed for estimating the couplings, the computations required to extract this information are expensive and prohibitive for large systems (28). This leaves the use of higher-order moments of the spin statistics as the only way to improve computational complexity. Several heuristic algorithms that use higher-order moments have been proposed on the basis of statistical physics arguments. Among other approximate methods, let us mention the adaptive cluster expansion (18), which controls the accuracy of the approximation at a cost of a higher computational complexity involving computation of entropies of growing clusters, and the probabilistic flow method (19), which introduces a relaxation dynamics to a certain trial distribution. However, both schemes remain computationally expensive and thus not suitable for large systems and rely on fine tuning of auxiliary parameters. An alternative method, which uses the full information contained in the samples, has been suggested and rigorously analyzed in the study by Ravikumar et al. (20). Although it has been shown by Montanari and Pereira (29) that this estimator is unable to correctly reproduce the underlying graph of the original model at low temperatures, until lately with certain modifications it remained the state-of-the-art practical method (21, 22, 30). Partly anticipating on our results, we show later in this paper that this regularized pseudolikelihood estimator (RPLE) can be turned into an exact and universal method if completed with a rather natural, but key, ingredient: a post-inference thresholding of reconstructed couplings.

The problem of designing a universal learning algorithm with polynomial computational complexity (28) that achieves exact graph topology reconstruction for arbitrary Ising models in all regimes was resolved only recently in the studies by Bresler et al. (23, 24). The biggest challenges addressed were the low-temperature regime and long-range correlations, which are known to be particularly difficult for learning. Nonetheless, the computational cost of these algorithms is still high and scales as a polynomial of high degree in the number of nodes (23) or double exponential in the maximum node degree $d_{\max}$ and in the maximum interaction strength (24). Moreover, both algorithms require prior information on the bounds on the interaction strengths, that is, positive α and β such that $\alpha \le |J_{ij}| \le \beta$ for all $(i, j) \in E$, as well as the knowledge of $d_{\max}$.

In an attempt to determine the optimal number of samples needed for reconstructing the graph, information-theoretic bounds were derived by Santhanam and Wainwright (31). We emphasize three salient features of these bounds. First, the optimal number of samples M_{opt} for perfect graph recovery scales exponentially with the maximum interaction value and node degree, $M_{\mathrm{opt}} \propto e^{c\gamma}$, where $\gamma = \beta d_{\max} + h_{\max}$, and $h_{\max}$ denotes an upper bound on the absolute values of magnetic fields. Although it was shown that $c \in [1, 4]$, the precise value of c remains unknown; here, we refer to this range of c as to the optimal regime with respect to the dependence of the number of samples on γ. Intuitively, this exponential scaling requirement can be attributed to the typical waiting time for collecting a sufficient number of "nontrivial" samples, that is, those that are different from the ground state configurations. This waiting time is more pronounced in the low-temperature regime when γ is large. Second, for finite $d_{\max}$, the dependence on the number of variables N is very weak: $M_{\mathrm{opt}} \propto \ln N$. This logarithmic dependence represents the amount of information needed for hypothesis testing over the set of $C_N^{d_{\max}}$ candidate neighborhoods of a given vertex (32). Third, the number of required configurations grows as α decreases, because it is difficult to distinguish between the presence of a very

weak coupling and its absence. In particular, in the limit of small α, $M_{\mathrm{opt}} \propto 1/\alpha^2$.

In what follows, we discuss two exact methods for solving the inverse Ising problem. The first method is based on the RPLE of Ravikumar et al. (20) supplemented with a post-optimization parameter thresholding procedure. We prove that this ingredient makes the estimator exact, meaning that the algorithm can reconstruct an arbitrary Ising model with an appropriate number of samples. The second algorithm that we introduce is an exact estimator based on the interaction screening method. By setting up a framework for an empirical assessment of the performance of the algorithms guided by the information-theoretic arguments presented above, we show that our new estimator outperforms the pseudolikelihood-based algorithm and requires in all test cases a number of samples lying within the information-theoretically optimal regime.

RESULTS

Regularized pseudolikelihood estimator

A widely used approach aiming at achieving the optimal scalings was suggested in the study by Ravikumar et al. (20), where the estimation of model parameters is performed on the basis of the so-called pseudo-likelihood acting as a surrogate for the intractable log-likelihood function. The method is based on maximizing a set of local RPLEs. Each of them can be interpreted as a regularized probability of a single spin i conditioned on the remaining $N - 1$ spins in the system given by

$$\mathcal{L}_i(J_i, H_i) = \left\langle \ln \frac{1}{1 + e^{-2\sigma_i(H_i + \sum_{j \ne i} J_{ij}\sigma_j)}} \right\rangle_M - \lambda \|J_i\|_1 \qquad (2)$$

where $\langle f(\sigma) \rangle_M = M^{-1} \sum_{m=1}^{M} f(\sigma^{(m)})$ is the notation for the empirical average; J_i and H_i are the optimization parameters; and J_i is the shorthand notation for $\{J_{ij}\}_{j \ne i}$. The sparsity promoting ℓ_1 regularization term $\| J_i \|_1 = \sum_{j \ne i} |J_{ij}|$ is important because it discourages the minimizer $\hat{J}_i$ from being dense by effectively pushing the interaction values toward zero whenever an edge is absent. In the original version of the algorithm, the graph structure is identified as a set of edges carrying couplings that were not set to zero by the RPLE. Guarantees for perfect graph reconstruction with this procedure rely on a rather restrictive set of conditions that are not always satisfied and are hard to verify in practice (20). Models known to satisfy these conditions are particular ferromagnetic models at high temperature, but this procedure provably fails in other regimes, most noticeably at low temperatures (29). A natural extension of this algorithm that uses a postestimation thresholding of a part of nonzero couplings was introduced in the study by Aurell and Ekeberg (21). In this scheme, all recovered J_{ij} satisfying $|J_{ij}| < \delta$, where δ is a chosen threshold, are declared to be zero. However, the performance of the RPLE-based algorithm with thresholding has never been rigorously analyzed, and until now, it was believed that any RPLE scheme fails in the low-temperature regime, following theoretical indications (29) and experimental studies conducted in a framework that does not fully account for the sample complexity structure of the inverse Ising problem (21). The reason that previous numerical studies show failure of the RPLE with thresholding at low temperatures is most likely due to the hidden dependence of the required number of samples M^* on the strength of the couplings (inverse temperature β) in the original analysis (20), which resulted in

tests of the reconstruction quality as a function of inverse temperature assuming a constant number of samples only. At the same time, it is clear that in the low-temperature regime, the Boltzmann probability measure concentrates on the ground state samples, that is, most of the samples in a typical batch would correspond to less-informative ground state configurations. Hence, an assessment of empirical performance should be based on a setting where the number of provided samples is exponentially increasing, in agreement with information-theoretic dependencies (31). We take this fact into account in the numerical experiments presented below.

In the Supplementary Text, we prove that there exists a minimum number of samples M^* for which the error on the estimated couplings is bounded by $\alpha/2$, so that choosing $\delta = \alpha/2$ leads to a perfect reconstruction of the graph topology. Hence, our first result states that the RPLE with a post-evaluation thresholding is exact: In the worst case, the required number of samples scales at most as $M^* \propto \exp(8\gamma)\ln N/\alpha^2$, (see Supplementary Text for details). Note that the parameter estimation problem for each vertex is independent, and the optimization can be carried out separately for each spin. As we explain below, the symmetrized estimate of coupling associated with the edge (i, j) is obtained as an average of local estimates $(\hat{J}_{ij} + \hat{J}_{ji})/2$. This parallelization of local reconstructions is lost when the optimization is performed globally over the entire graph (22).

Interaction screening method

Recently, we introduced the first exact reconstruction algorithm having the same parametric dependence as the information-theoretic bound and termed the regularized interaction screening estimator (RISE) (33). Our theoretical analysis showed that the RISE has a lower theoretical sample complexity for perfect graph recovery compared to the one derived here for the RPLE with thresholding, guaranteeing that a number of samples $M^* \propto \exp(6\gamma)\ln N/\alpha^2$ is sufficient for reconstruction of the graph structure [see the Supplementary Text and (33) for details]. However, factors 6 and 8 in the exponents of the RISE and the RPLE, respectively, are likely to be an artifact of the used proof techniques and are not tight as indicated by the computational experiments in this paper.

The RISE is based on the minimization of the interaction screening objective (ISO)

$$S_i(\boldsymbol{J}_i, H_i) = \left\langle \exp\left(-\sum_{j \neq i} J_{ij}\sigma_i\sigma_j - H_i\sigma_i\right)\right\rangle_M \tag{3}$$

over the probe vector of couplings J_i and the probe magnetic field H_i for a given spin i. The ISO, as its name suggests, is constructed on the basis of the property of "interaction screening," which is illustrated in Fig. 1. As a consequence of this property, in the limit of a large number of samples, the unique minimizer of the convex ISO objective is achieved at $(\boldsymbol{J}_i, H_i) = (\boldsymbol{J}_i^*, H_i^*)$. A simple derivation of this fact is presented in the Materials and Methods. In the RISE construction, the ISO is appended with the ℓ_1 regularizer to promote sparsity (33). Here, we introduce a modification to the RISE that leads to a new exact learning method for the inverse Ising problem, which we call the logRISE and which takes the following form

$$(\hat{\boldsymbol{J}}_i, \hat{H}_i) = \underset{(\boldsymbol{J}_i, H_i)}{\mathrm{argmin}} \left[\ln S_i(\boldsymbol{J}_i, H_i) + \lambda\|\boldsymbol{J}_i\|_1\right] \tag{4}$$

The name logRISE comes from the fact that instead of the ISO itself, we use its logarithm to form the logRISE objective (4). Obviously, in the
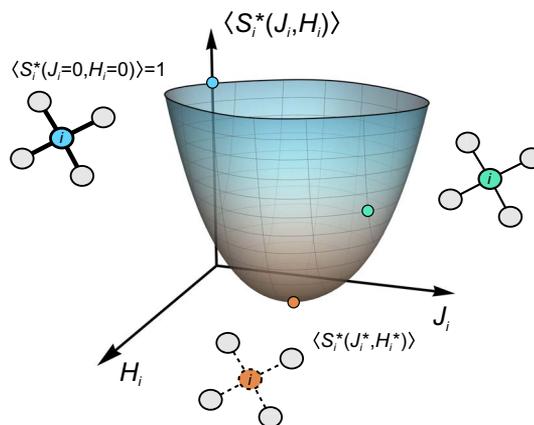

Fig. 1. ISO for different probe values of model parameters in the large M limit. The ISO is an empirical average of the inverse of the factors in the Gibbs measure, and its screening property becomes apparent in the limit of large number of samples. Changing the value of the probing parameters (J_i, H_i) in the ISO alters the effective interaction strength of σ_i with its neighbors. This mechanism is schematically represented in the figure, where the value of ISO for different values of probing parameters is depicted. When the probing parameters are equal to the true ones $(\boldsymbol{J}_i^*, H_i^*)$, the ISO completely screens this interaction making σ_i effectively independent of its neighbors. With some analysis, this can be shown to be equivalent to the ISO attaining its minimum at the true parameters of the model.

absence of the regularizer (for $\lambda = 0$), taking the logarithm of the ISO does not change its minimizer. However, this difference is crucial for nonzero values of the regularization term, which suggests that logRISE might have good properties for the reconstruction problem due to a particular form of its first and second derivatives (see the Supplementary Text for additional explanations and details).

Unfortunately, the proof techniques used for deriving bounds on scaling for the RPLE and the RISE provide less tight expressions when applied to the estimator logRISE, because it no longer can be represented in a form of finite functional sum over individual samples. Our analysis states that the number of required samples for logRISE in the worst case scales as $M^* \propto \exp(10\gamma)\ln N/\alpha^2$ for guaranteeing the reconstruction of the structure of the underlying Ising model with a high probability. Given the looseness of the theoretical analysis in this case, the empirical assessment of the performance of the logRISE and its comparison with the RPLE are required. We provide a detailed numerical study of the quality of different estimators below.

As we show through a rigorous analysis in the Supplementary Text, the ℓ_1 regularizer plays an important role for all of the estimators because it reduces the required sample complexity for perfect topology reconstruction from quasi-linear to logarithmic in the number of spins N. However, the performance of the RPLE, the RISE, and the logRISE and hence the number of required samples M^* are dependent on the regularization coefficient λ. The choice of λ needs to account for the following tradeoff: If λ is too small, then the estimation is prone to noise, and if λ is too large, then it introduces a bias in the estimated couplings toward zero. The optimal value of λ is unknown a priori. In the Supplementary Text, we present detailed simulations for different topologies, which show that for achieving correct graph reconstruction with probability $1 - \varepsilon$, the choice $\lambda = c_\lambda\sqrt{\ln(N^2/\varepsilon)/M}$ is appropriate when no additional information about the model is available, with $c_\lambda \simeq 0.2$ for the RPLE, $c_\lambda \simeq 0.4$ for the RISE, and $c_\lambda \simeq 0.8$ for the logRISE. We use these

values for λ in all numerical experiments reported below. Given a sufficient number of samples, other techniques such as consistency cross-validation can be used for selecting the optimal value of the regularization coefficient on a case-by-case basis. An illustration of this approach alongside some practical remarks is provided in the Supplementary Text.

Learning structure and parameters of the model

We state our three-step algorithm for learning the underlying graph and the parameter values of the Ising model using the RPLE or the logRISE (the same algorithm applies to the RISE). First, given M samples, we find the optimizer of the objective (2) or (4), respectively, at each node $i \in V$ and obtain a collection of estimated parameters $(\hat{J}_i, \hat{H}_i)$. Given that both estimators are convex, any appropriate convex optimization method can be used to find the minimizer of the objective function, the simplest one being a plain gradient descent supplemented with an additional projection step due to nondifferentiability of the ℓ_1 regularization term. For our numerical experiments, we used the Ipopt optimization software (34); however, as we comment in the Supplementary Text, better choices such as composite-type gradient descent methods exist for experiments with very large networks (35, 36).

Given a sufficient number of samples M, a typical histogram of couplings estimated by the RPLE, the RISE, or the logRISE takes the form shown in Fig. 2A. Notice the emergence of gaps separating a group of inferred couplings that are close to zero from those with significantly bigger intensities in absolute value. In the second step, we threshold the inferred couplings below the observed gaps to zero. The edges associated with the remaining nonzero couplings form the reconstructed graph $\hat{G}$. Finally, we optimize the unregularized objective for each of the three estimators by setting $\lambda = 0$, but only over the couplings corresponding to the edges in $\hat{G}$, and obtain our final estimates $(\hat{J}_i, \hat{H}_i)$. This procedure is illustrated in Fig. 2B for the logRISE on an Erdös-Rényi graph with $N = 25$ nodes and spin glass couplings, where the scatter plot of predicted versus true values of the model parameters is presented, and only parameters over the already reconstructed graph from $M = M^* = 5000$ samples have been accounted for. We see that even using a small number of samples (in this example, the minimal amount required for a correct structure recovery), the numerical values of the parameters are also reconstructed with a very good accuracy that increases when more samples are provided.

To have statistical confidence in our results, we determine M^* as follows. Progressively increasing values of M, the reconstruction experiment runs L times, using L independent sets of M samples. On the basis of the number of successful topology reconstructions L_{succ}, one can define the empirical probability of reconstruction $P_{emp} = L_{succ}/L$. We define M^* as the minimum M for which $P_{emp} = 1$ (see Fig. 2C for a typical example). The value L that we use in our computations, $L = 45$, comes from the requirement of a perfect topology reconstruction with probability greater than $1 - \epsilon$, where we fix $\epsilon = 0.05$. That is, it is essential to get $L = 45$ successful reconstructions in a row to make sure that the probability of correct topology recovery is above 0.95 with confidence of at least 90%, as we explain in the Supplementary Text. We use this value of L in the computations throughout the text.

We performed extensive numerical experiments to empirically obtain the minimal number of samples M^* required for perfect graph reconstruction for different topologies and types of interactions. We carried out numerical experiments for all of the three estimators considered in this paper. However, for the sake of simplicity and for the clarity of presentation, in what follows in the main text, we present
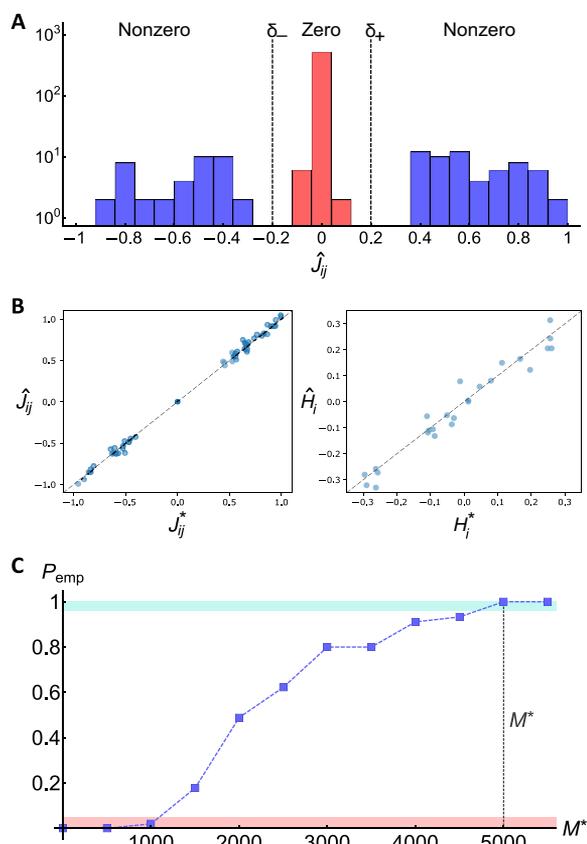

Fig. 2. Reconstruction of the graph topology and the values of parameters with the logRISE. Reconstruction procedure for an Erdös-Rényi graph with $N = 25$ and average degree $\langle d \rangle = 4$ given $M = 5000$ configurations. The couplings and magnetic fields are generated uniformly at random in the range $[-1.0, -0.4] \cup [0.4, 0.1]$ and $[-0.3, 0.3]$, respectively. (**A**) The symmetrized estimate of coupling $\hat{J}_{ij}$ associated with the edge (i, j) is obtained as an average of local estimates $(\hat{J}_{ij} + \hat{J}_{ji})/2$. When the regularizing parameter λ is appropriately chosen and the number of samples is sufficient, gaps emerge in the estimated couplings $\hat{J}_{ij}$ around $\delta_+ > 0$ and $\delta_- < 0$, separating the estimated couplings that are close to zero and those with higher intensities in absolute value. The values below the threshold are then set to zero to obtain an estimate of the graph structure. (**B**) Once the graph structure is learned, the parameters are reestimated by optimizing the unregularized objective only over the edges in the reconstructed graph. The reduction in the number of free optimization variables from N to $d_{max} + 1$ greatly improves the estimates. The resulting values are shown in the scatter plot. (**C**) Empirical probability of successful structure recovery P_{emp} over $L = 45$ independent runs as a function of the number of samples M. For the logRISE, the smallest number of samples for which $P_{emp} = 1$ is given by $M^* = 5000$.

numerical results only for the logRISE, which is the central object of the present study, and for the RPLE, which is the state-of-the-art method for the inverse Ising problem. Note that throughout the manuscript, we present comparisons of the logRISE with the exact and universal version of the RPLE, that is, corrected through our thresholding procedure. The corresponding scalings for the RISE are available in the Supplementary Text.

We first verify the logarithmic scaling of M^*, claimed in our theoretical analysis for RPLE and logRISE, with respect to the number of spins N in ferromagnetic Ising models without magnetic fields ($J_{ij}^* > 0$, $H_i^* = 0$), defined on two topologies: square lattice with periodic boundary conditions and random regular (RR) graphs with degree $d = 3$. The choice of ferromagnetic models has been dictated by the need to generate

independent samples for large values of N, and given that for spin glass models, this is a nontrivial task (37). For the two aforementioned topologies, we generate independent samples using Glauber dynamics for different values of N in the low-temperature regime, where the correlations are long-range: we have used $J_{ij}^* = 0.7$ for the lattice ensemble and $J_{ij}^* = 1.0$ for the RR graph ensemble. The minimal required sample size M^* on both topologies is presented in Fig. 3. We see that M^* exhibits a logarithmic dependence on N for both estimators: the logRISE and the RPLE.

The major difference in performance between the estimators is observed in the scaling with respect to $\gamma = \beta d_{\max} + h_{\max}$. This is critical because a favorable exponent allows the algorithm to have a lower sample complexity in the low-temperature regime, where known algorithms either do not work or exhibit poor scaling. An extensive numerical study is presented in Fig. 4, where we study quasi-homogeneous systems with ferromagnetic-type couplings (Fig. 4A, B, and E) and spin glass–type couplings (Fig. 4C and D) on two topologies: square lattice with double-periodic boundary conditions (Fig. 4A, C, and E) and random regular graphs (Fig. 4B and D). This choice of topologies eliminates fluctuations with respect to the heterogeneity of node degrees, so that it becomes easier to extract the right scaling with respect to β and d. To disentangle the effects of α and β, we always fix one (for ferromagnets) or two (for spin glass systems) couplings to α and $-\alpha$, which is different from the interaction values $\pm\beta$ carried by the rest of the edges. Therefore, β can be conveniently thought of as the inverse temperature of the model. To investigate the effect of temperature on the scalings, we deliberately set magnetic fields to zero and fix the thresholding parameter to $\delta = \alpha/2$. The test cases represented in Fig. 4 (A to D) show that, overall, the RPLE and the logRISE demonstrate similar scaling properties. Notice that there exists a qualitative difference in the scaling behavior between the low- and high-temperature regimes,

with an exponential scaling for both estimators observed for large β. Our numerical study shows that from the learning perspective, the ferromagnetic model on the two-dimensional lattice appears to be the most challenging class of Ising models for both the logRISE and the RPLE. It has the highest scaling exponent with respect to γ and, hence, the largest sample complexity for the inverse Ising problem. This observation supports theoretical evidence that this case belongs to the hardest class of models with respect to learning (38). In particular, this finding shows that, paradoxically, the inverse Ising problem is significantly harder for planar ferromagnetic models compared to spin glass models on random graphs, whereas the direct problem of drawing independent samples from the former can be incomparably easier than from the latter (37).

The ultimately hardest case for the reconstruction problem is unknown. However, we were able to construct a slight variant of the ferromagnetic model on a lattice that appears to be even harder for all algorithms considered: a ferromagnetic model with a weak antiferromagnetic interaction, that is, an edge carrying a negative coupling $-\alpha$. In Discussion, we present intuitive arguments why this case should be fundamentally hard. The results for the extraction of the M^* in this model instance are presented in Fig. 4E. We see that the logRISE has a markedly better scaling exponent compared to the RPLE. In this test case, the scaling exponent of the RPLE is significantly larger than the information-theoretic upper bound, whereas the corresponding value for the logRISE lies within the optimal regime in terms of the information-theoretic predictions. We summarize the scaling behavior of the estimators in Discussion.

Application to a real system: D-Wave quantum computer

To evaluate the performance and robustness of the estimators in a nonsynthetic case, we apply the logRISE and the RPLE to real data produced by the D-Wave 2X quantum annealer "Ising" at Los Alamos National Laboratory. The D-Wave computer (39) has been designed for solving binary quadratic optimization problems in the form of Ising models through quantum annealing, that is, slowly transforming an initially prepared state of the system to the ground state of the desired input Ising Hamiltonian encoded on its chip. Because of the thermal noise in the system, a single annealing run may end in one of the excited states instead of the desired ground state. In practice, the device attempts to find the target ground state by rerunning the annealing multiple times and producing as output the best solution found. Previous experiments with D-Wave report that the produced samples are distributed according to the Boltzmann distribution at some effective temperature (40) related but not equal to the native temperature at which D-Wave operates. This effective temperature is naturally low because D-Wave contains superconducting elements as a part of its architecture. Because of the temperature rescaling effect, as well as inevitable biases present in this analog device, the effective Ising model from which the samples are produced does not exactly correspond to the input Ising model. It then becomes interesting to see how the structure of the distorted effective Ising model is related to the one encoded in the chip. This task is exactly what the methods presented in our paper are designed to solve, making it a good real-world application for testing their performance.

Let us describe the procedure that we followed for generating the data. Our goal was to check the performance of the algorithms on a noisy heterogeneous instance, both in node degrees and couplings as well as in magnetic fields. Hence, we encoded an Ising model with random couplings and magnetic fields, distributed uniformly in the range
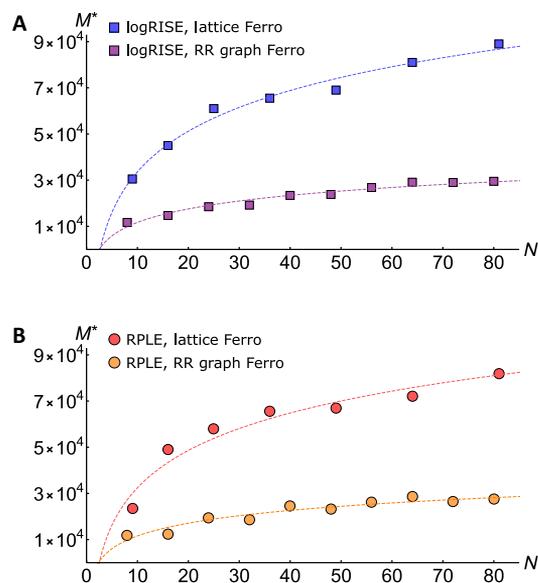

Fig. 3. Verification of the logarithmic scaling with the size of the system. Scaling of M^* with the number of spins N for (**A**) the logRISE and (**B**) the RPLE obtained using samples produced in the cases of the ferromagnetic Ising model over a double periodic two-dimensional lattice with $\beta = 0.7$ and ferromagnetic random regular graphs with degree $d = 3$ for $\beta = 1.0$. In all cases, we observe a logarithmic growth of M^* with respect to N, which is in agreement with the information-theoretic bounds and our theoretical analysis for the estimators.
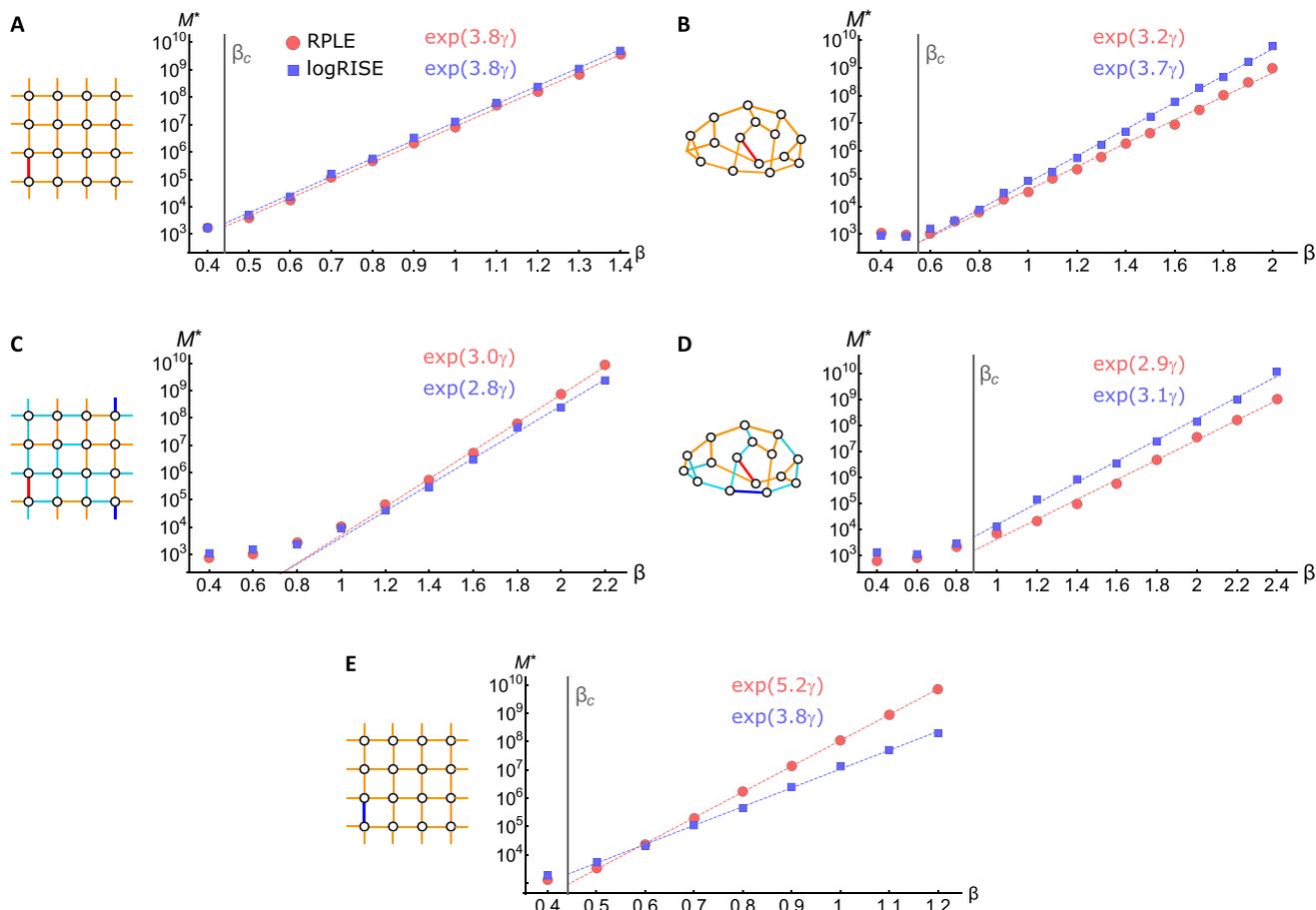
Fig. 4. Scaling of M^* with the couplings strength. Comparison of the performance of the logRISE and the RPLE is presented for five different ensembles of Ising models. Because of a weak dependence $M^* \propto \ln N$, we consider graphs of size $N = 16$, which allowed us to produce independent samples through an exhaustive enumeration of spin configurations. The first four cases correspond to (**A**) a ferromagnet on a square lattice with double-periodic boundary conditions, (**B**) a ferromagnet on a random 3-regular graph, (**C**) a spin glass on a periodic lattice, and (**D**) a spin glass on a random 3-regular graph. The most difficult reconstruction test case for both algorithms, a ferromagnetic lattice with a weak antiferromagnetic impurity, is presented in (**E**). The phase transition points in the corresponding infinite systems are indicated as β_c. An exact pictorial representation of the corresponding Ising model is portrayed on the left-hand side of each plot. Ferromagnetic couplings equal to β and $\alpha = 0.4$ are colored in orange and red, respectively. Antiferromagnetic couplings equal to $-\beta$ and $-\alpha$, respectively, are colored in turquoise and blue.

$[-0.16, -0.02] \cup [0.02, 0.16]$. We also chose to encode these couplings in a region of the chip with the highest concentration of broken qubits that are inevitably present and can potentially create additional noise. The topology of this portion of the chip containing $N = 62$ qubits is illustrated in gray in Fig. 5A. We observed that the initial Ising model got distorted while being implemented on the chip. From several trial tests, we inferred that the effective rescaling factor in this regime roughly fluctuates around $\beta_{eff} \approx 12$, although this factor is different for individual model parameters. Because the precise values of the couplings and magnetic fields actually implemented on the chip are unknown, the only "ground truth" available to us in this experiment is the topology of the portion of the chip that we encoded our model on. However, let us point out that because of the complexity of the D-Wave architecture and a possible conflict of superconducting loops representing couplers between qubits, it is a priori unclear whether the resulting topology of the effective Ising model will necessarily remain unchanged.

The maximum number of annealing runs for a given Ising model implementation is limited to 10^4 by standard system settings on the D-Wave. We collected 5×10^5 samples corresponding to the same input

model specified above by obtaining 50 batches of 10^4 samples each and provided them as an input to the logRISE and the RPLE. Notice that each additional implementation of the same chosen Ising model for each batch in principle corresponds to a different actual Ising Hamiltonian owing to a different concrete realization of random biases; this creates an additional source of noise in our data. The reconstructed model parameters are presented in Fig. 5 (A and B). We emphasize that it is difficult to disentangle the effects of statistical errors due to the finiteness of the number of samples and the errors due to noise.

For structure reconstruction, we chose to threshold the parameters J_{ij} in the tail of a set of couplings reconstructed in the vicinity of zero. Given this choice of threshold, we found that both algorithms are quite robust to noise and are able to accurately reconstruct the graph topology, making only a few false positives and false negatives. The reconstructed topologies are shown in the left of Fig. 5 (A and B). Notice that although the RPLE makes local errors, detecting one false-positive and one false-negative connections between neighboring spins, the logRISE misclassifies a nonexisting edge as existing in a clearly nonlocal fashion, meaning that the vertices it misclassifies as neighbors are
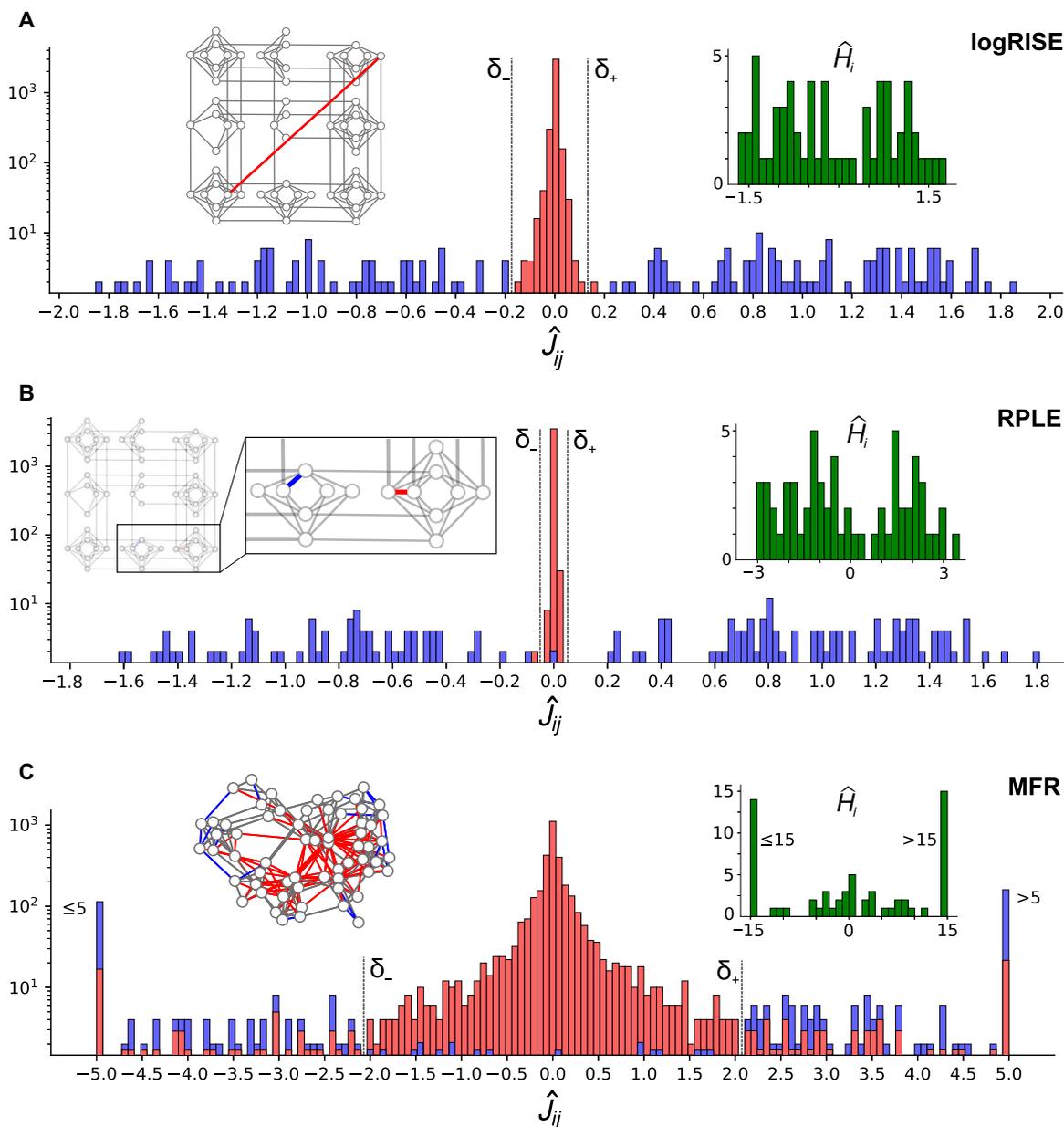
Fig. 5. Reconstruction of the structure of a portion of the D-Wave annealer chip using 5×10^5 samples. This part contains 62 qubits with heterogeneous connectivity, couplings, and magnetic fields. Reconstructed couplings are presented for (**A**) the logRISE, (**B**) the RPLE, and (**C**) the mean-field regime (MFR) of the RPLE and the logRISE. On each histogram in the main plots, bars corresponding to the edges actually present on the chip are colored in blue, whereas nonexistent connections are colored in red. The reconstructed magnetic fields are shown in green in a separate histogram on the right-hand side. A topology of the reconstructed structure is depicted on the left-hand side, with correctly reconstructed edges, missing edges, and incorrectly reconstructed edges colored in gray, blue, and red, respectively. Although the MFR exhibits a poor behavior, as expected at such low temperatures, both the logRISE and the RPLE are achieving similarly good performance. Notice that whereas there exists a thresholding procedure that produces a perfect network recovery with the logRISE, it is not the case for the RPLE as one existing coupling has been set to zero.

far away in graph theoretic distance on the D-Wave chip. Although in the case of the logRISE it is possible to choose an optimal threshold that allows one to completely separate zero couplings from nonzero ones and, thus, reconstruct the structure of the chip perfectly, no such thresholding is possible for the result produced by the RPLE, suggesting that the RPLE needs more samples before this separation becomes possible. Finally, notice that according to the histograms on reconstructed magnetic fields in the right insets of Fig. 5 (A and B), the RPLE seems to make larger errors in the reconstruction of magnetic fields

that should be of the same order as couplings according to our input Hamiltonian.

As we pointed out in the Introduction, a plethora of other methods have been proposed for the inverse Ising problem, but the majority of them either are too computationally expensive for practical applications or fail at low temperatures, sometimes even when an infinite number of samples are provided. To illustrate the value of exact algorithms, especially for problems at low temperatures (such as this application), we compare the results obtained from the logRISE and

the RPLE to those from mean-field–type methods (see Fig. 5C). The particular scheme that we used for comparison is obtained from a high-temperature expansion of our estimators and is closely related to the naïve mean-field method of statistical physics, which performs well at high temperatures. See Materials and Methods for a detailed description of the method and related discussions. As expected for such systems with strong and long-range correlations, this method using only information contained in magnetization and pairwise correlations behaves poorly, incurring a very large number of false positives and false negatives. This illustrates an importance of taking into account higher-order interaction in data samples for a reliable reconstruction in the low-temperature regime.

DISCUSSION

In Fig. 6, we comparatively present the algorithmic scalings that summarize the main theoretical and empirical results of our paper. All of the three considered estimators for the inverse Ising problem have a better worst-case empirical scaling compared to their theoretical estimates. The empirical sample complexity of the logRISE algorithm introduced in this paper lies in the optimal regime with respect to the information-theoretic predictions, outperforming all existing methods. The worst-case scalings are based on the hardest case for the learning problem that we were able to construct. To describe the logic behind this case, we first mention observations in existing literature and then provide intuitive arguments regarding the way the structure of the underlying graph and the nature of interactions affect the hardness of the reconstruction. There are strong theoretical indications that ferromagnetic-type spin systems are among the models requiring a maximum number of samples to be learned. Information-theoretic bounds suggest that these models are at least as hard to learn as any other model (38). Moreover, the presence of strong long-range correlations is known to be a challenging situation to deal with (24). This hardness of learning of ferromagnetic models is consistent with our numerical studies in which ferromagnetic random graphs and especially ferromagnetic lattices are the cases requiring the largest amount of samples. Intuitive explanations for this behavior are twofold. As mentioned earlier, ferromagnetic models are more prompt to develop strong long-range correlations at low temperatures, especially on lattices, and they tend to favor two configurations that are the ground states. Long-range correlations make it less likely to obtain nontrivial samples, that is, fluctuations around ground states that are crucial to obtain information about the detailed structure of the graph that is crucial for the reconstruction. This translates into a need for a larger number of samples, proportional to the likeliness of such fluctuations that is typically exponentially suppressed in γ. Moreover, when several similar models share identical ground states, it becomes very hard to make a distinction between them solely based on configurations close to their ground states. This mechanism can be illustrated very simply using an extreme example of three spins with homogeneous couplings forming a chain that is either open or closed, forming a triangle. Deciding which chain is formed is impossible for a ferromagnetic system when only the ground states ±(1, 1, 1) are observed. However, it is an easy task for an antiferromagnetic system because an open chain has two ground states ±(1, −1, 1), whereas the close chain has six ground states ±(1, 1, −1), ±(1, −1, 1), and ±(−1, 1, 1).

The hardest test case studied in our numerical experiments contains an extra ingredient that makes the inverse Ising problem even more challenging: an additional weak negative coupling or an "antiferromagnetic impurity" added on top of the ferromagnetic model on a lattice. This weak antiferromagnetic bond has the effect of weakening or cancelling the correlation between the two spins that it connects. Consequently, it becomes difficult to distinguish between the presence of this weak negative coupling from its absence. Although we do not claim with certainty that this model is the hardest to learn, we believe that any such difficult-to-learn model is likely to include the features outlined above.

We proved that the three techniques explored in this paper, the logRISE, the RISE, and the RPLE, are exact and universal methods to solve the inverse Ising problem. Exactness and universality in this context mean that these methods reconstruct couplings and magnetic fields up to any given accuracy with a sufficient but finite number of samples and for every Ising model regardless of its structure, density, temperature, or any other property that characterizes it. Although in the present article we focused on the quantification of the scaling of the number of required samples with structural properties and temperature of sparse systems, it remains an interesting question left for exploration in dense models, for instance, in the Curie-Weiss or the Sherrington-Kirkpatrick type (37). In these models, the exponential scaling with coupling intensities and degrees, denoted by γ for sparse models, will be more intricate. It seems reasonable to expect that the sample requirement scales exponentially with the typical "energy per spin." For instance, in the Curie-Weiss–type models with all
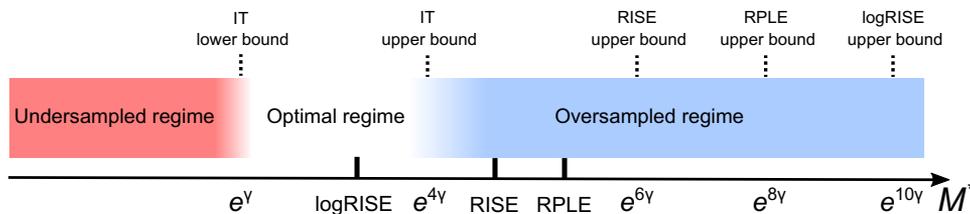

Fig. 6. Theoretical and empirical worst-case scaling of M^* with respect to γ. This figure summarizes the main theoretical and empirical results of this paper for the inverse Ising problem. The red region represents the undersampled regime where the number of samples is insufficient for perfect graph reconstruction from the information theory perspective. The existence of an exact algorithm, albeit with an exponential computational complexity, has been proven for $M \propto e^{4\gamma}$, and thus represents an upper bound on the optimal number of samples M_{opt} that must lie in the white region, named the optimal regime. The quantities $e^{6\gamma}$, $e^{8\gamma}$, and $e^{10\gamma}$ denote our theoretical upper bounds on the scaling for the RISE, the RPLE, and the logRISE, respectively. However, these bounds are not tight, and the worst-case empirical scalings observed in our numerical experiments were much lower; these values are indicated in the chart as "RISE," "RPLE," and "logRISE" and correspond to $e^{4.5\gamma}$, $e^{5.2\gamma}$, and $e^{3.8\gamma}$, respectively (see the Supplementary Text for additional details on the scaling of the RISE). The empirical scaling for the logRISE lies within the optimal regime.

$J_{ij} \geq 0$, this quantity is $\gamma_{CW} \approx h_i + \frac{1}{N}\Sigma_{j \neq i}J_{ij}$, whereas in the Sherrington-Kirkpatrick–type models, where J_{ij} are centered random variables, it reads $\gamma_{SK} \approx h_i + \frac{1}{\sqrt{N}}\Sigma_{j \neq i}\sqrt{\text{Var}(J_{ij})}$. We also note that, in these dense models, there is no longer any reason to expect that the sample complexity requirement scales logarithmically with the system size (31); instead, we expect it to exhibit a polynomial dependence. Note that, in this case, the inclusion of the ℓ_1 regularizer in the logRISE and the RPLE is no longer necessary because there is no sparsity pattern to promote.

In conclusion, in this paper, we showed both theoretically and experimentally that an arbitrary Ising model can be reconstructed exactly with an information-theoretically minimum number of samples using the introduced interaction screening method. In addition, no prior knowledge on the graph and associated parameters is required to implement the algorithm, making it a very practical choice for applications. The practical advantages of our methods have been illustrated on real data coming from a D-Wave quantum computer. We also provided a sample complexity analysis of the popular RPLE showing the logarithmic scaling in system size for arbitrary Ising models, albeit with a higher worse-case scaling with respect to the inverse temperature when compared to the logRISE. We demonstrated the paradoxical relation between sampling and learning, showing that the instances that are easier for one task are harder for the other. In Materials and Methods, we point out a curious connection to the mean-field approximation at high temperatures. The second-order high-temperature expansion of all exact estimators considered in this paper provides an identical reconstruction scheme, valid in the limit of weak couplings. This high-temperature regime is related to learning methods based on the well-known naïve mean-field approximation in statistical physics. Finally, although this paper is dedicated to the reconstruction of Ising models, the interaction screening method can be generalized to graphical models with higher-order interactions and nonbinary alphabets, including those described by Hamiltonians over continuous variables. Exploration of these research directions is underway.

MATERIALS AND METHODS
Interaction screening property
Here, we presented a simple argument that illustrated the fact that in the limit of a large number of samples, the unique minimizer of the convex ISO objective (Eq. 3) was achieved at $(J_i, H_i) = (J_i^*, H_i^*)$, meaning that the true interactions present in the model were fully "screened." The ISO was an empirical average of the inverse of the factors in the Gibbs measure; if $\mathcal{F}_i(J_i, H_i) = \exp(\Sigma_{j \neq i}J_{ij}\sigma_i\sigma_j + H_i\sigma_i)$, then $S(J_i, H_i) = \langle \mathcal{F}_i^{-1}(J_i, H_i)\rangle_M$. In the limit of a large number of samples $S(J_i, H_i) \rightarrow S^*(J_i, H_i) = \langle 1/\mathcal{F}_i(J_i, H_i)\rangle$. The derivative of the ISO corresponded to weighted pairwise correlations, $\partial S^*/\partial J_{ij} = \langle \sigma_i\sigma_j/\mathcal{F}_i(J_i, H_i)\rangle$, and this shed light on its key property. When $(J_i, H_i) = (J_i^*, H_i^*)$, $\partial S^*/\partial J_{ij}|_{J_i^*, H_i^*} = 0$, meaning that the minimum of ISO was achieved at $(J_i, H_i) = (J_i^*, H_i^*)$ as $M \rightarrow \infty$.

High-temperature expansion of exact estimators and connection to the mean-field method
Among all heuristics undertaken to solve the inverse Ising problem, a large fraction of methods is based on mean-field approximations using various level of sophistication [see the study by Roudi et al. (14) for a review]. In particular, the first such attempt to solve the inverse Ising problem was based on a naïve mean-field approach, where inferred couplings are related to the inverse reduced correlation matrix (13). Although these techniques provided satisfactory estimates in the

high-temperature regime, they were known to exhibit poor behaviors at low temperatures when the model developed long-range correlations, even for an infinite number of samples (29).

It is interesting to observe that there exists a connection between mean-field approaches and the high-temperature expansion of the exact estimators RISE and RPLE. A second-order Taylor expansion of the pseudolikelihood objective function (without regularizer) and the ISO around the high-temperature point $(J_i, H_i) = (0, 0)$ produced an explicitly solvable minimization problem (see Supplementary Text for an exact derivation). It is remarkable that, in this regime, both objective functions produced identical estimates for the model parameters. Couplings and magnetic fields reconstructed in this mean-field regime (MFR) were expressed as functions of the inverse connected correlation matrix and local magnetizations

$$\hat{J}_{ik}^{\text{MFR}} = -\frac{\left[\bar{C}^{-1}\right]_{ik}}{\left[\bar{C}^{-1}\right]_{ii}}, \quad \hat{H}_i^{\text{MFR}} = -m_i + \sum_{j \neq i}\frac{\left[\bar{C}^{-1}\right]_{jk}}{\left[\bar{C}^{-1}\right]_{ii}}m_j \quad (5)$$

where the matrix of empirical connected correlations and local magnetizations were directly computed from samples using the formulae $\bar{C}_{ij} = \langle \sigma_i\sigma_j\rangle_M - \langle \sigma_i\rangle_M\langle \sigma_j\rangle_M$ and $m_i = \langle \sigma_i\rangle_M$. Note that there was a subtle difference between the MFR estimates in Eq. (5) and the naïve mean-field estimates in (13). The values produced by the naïve mean-field method were directly equal to the inverse connected correlation matrix, whereas the MFR estimates were rescaled by the diagonal entries of this matrix. As a result, although both estimators provided similar answers, the choice of symmetrization and thresholding procedures for the reconstructed parameters can lead to significant discrepancies in the final estimates of the graph structure. It is worth noticing that the exact same expression producing the MFR estimates arose in the context of reconstructing multivariate Gaussian distributions (41). This parallel suggests that an optimal thresholding and symmetrization procedure for the MFR estimates was likely to be based on the geometrical mean rather than the arithmetic average.

SUPPLEMENTARY MATERIALS
Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/4/3/e1700791/DC1

Supplementary Text
section S1. Analysis of the estimators RPLE, RISE, and logRISE
section S2. On optimization techniques for minimizing the estimators
section S3. On the procedure for M^* selection
section S4. On the theoretical predictions for λ selection
section S5. Empirical selection of the regularization parameter λ
section S6. Hyperparameter λ selection through cross-validation
section S7. Scalings of the RISE with respect to γ
section S8. High-temperature expansion of the RISE and the RPLE
fig. S1. Schematic representation of the proof strategy for bounding the reconstruction error.
fig. S2. Dependence of the number of samples M^* required for a correct structure recovery on the regularization coefficient.
fig. S3. Selection of the hyperparameter λ through the K-fold cross-validation method on a 4×4 spin glass system on a square lattice.
fig. S4. Values of M^* and γ exponents for the RISE across different test cases.
References (42–45)

REFERENCES AND NOTES
1. G. Gallavotti, *Statistical Mechanics: A Short Treatise* (Springer Science & Business Media, 2013).
2. W. Kunkin, H. L. Frisch, Inverse problem in classical statistical mechanics. *Phys. Rev.* **177**, 282 (1969).

3. J. T. Chayes, L. Chayes, E. H. Lieb, The inverse problem in classical statistical mechanics. *Commun. Math. Phys.* **93**, 57–121 (1984).

4. E. Schneidman, M. J. Berry II, R. Segev, W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440**, 1007–1012 (2006).

5. S. Cocco, S. Leibler, R. Monasson, Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14058–14062 (2009).

6. F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, M. Weigt, Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1293–E1301 (2011).

7. D. Marbach, J. C. Costello, R. Küffner, N. M. Vega, R. J. Prill, D. M. Camacho, K. R. Allison; DREAM5 Consortium, M. Kellis, J. J. Collins, G. Stolovitzky, Wisdom of crowds for robust gene network inference. *Nat. Methods* **9**, 796–804 (2012).

8. T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, M. Troyer, Defining and detecting quantum speedup. *Science* **345**, 420–424 (2014).

9. D. K. Panjwani, G. Healey, Markov random field models for unsupervised segmentation of textured color images. *IEEE Trans. Pattern Anal. Mach. Intell.* **17**, 939–954 (1995).

10. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436–444 (2015).

11. N. Eagle, A. S. Pentland, D. Lazer, Inferring friendship network structure by using mobile phone data. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15274–15278 (2009).

12. D. H. Ackley, G. E. Hinton, T. J. Sejnowski, A learning algorithm for boltzmann machines. *Cogn. Sci.* **9**, 147–169 (1985).

13. H. J. Kappen, F. B. Rodríguez, Boltzmann machine learning using mean field theory and linear response correction, in *Advances in Neural Information Processing Systems*, M. I. Jordan, M. J. Kearns, S. A. Solla, Eds. (MIT Press, 1998), pp. 280–286.

14. Y. Roudi, E. Aurell, J. A. Hertz, Statistical physics of pairwise probability models. *Front. Comput. Neurosci.* **3**, 22 (2009).

15. V. Sessak, R. Monasson, Small-correlation expansions for the inverse Ising problem. *J. Phys. A* **42**, 055001 (2009).

16. H. C. Nguyen, J. Berg, Mean-field theory for the inverse Ising problem at low temperatures. *Phys. Rev. Lett.* **109**, 050602 (2012).

17. F. Ricci-Tersenghi, The Bethe approximation for solving the inverse Ising problem: A comparison with other inference methods. *J. Stat. Mech.* **2012**, P08015 (2012).

18. S. Cocco, R. Monasson, Adaptive cluster expansion for inferring Boltzmann machines with noisy data. *Phys. Rev. Lett.* **106**, 090601 (2011).

19. J. Sohl-Dickstein, P. B. Battaglino, M. R. DeWeese, New method for parameter estimation in probabilistic models: Minimum probability flow. *Phys. Rev. Lett.* **107**, 220601 (2011).

20. P. Ravikumar, M. J. Wainwright, J. D. Lafferty, High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Stat.* **38**, 1287–1319 (2010).

21. E. Aurell, M. Ekeberg, Inverse Ising inference using all the data. *Phys. Rev. Lett.* **108**, 090201 (2012).

22. A. Decelle, F. Ricci-Tersenghi, Pseudolikelihood decimation algorithm improving the inference of the interaction network in a general class of Ising models. *Phys. Rev. Lett.* **112**, 070603 (2014).

23. G. Bresler, E. Mossel, A. Sly, Reconstruction of Markov random fields from samples: Some observations and algorithms, in *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, A. Goel, K. Jansen, J. D. P. Rolim, R. Rubinfeld, Eds. (Springer, 2008), pp. 343–356.

24. G. Bresler, Efficiently learning Ising models on arbitrary graphs, in *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, Portland, OR, 14 to 17 June 2015.

25. G. F. Cooper, The computational complexity of probabilistic inference using bayesian belief networks. *Artif. Intell.* **42**, 393–405 (1990).

26. C. Chow, C. Liu, Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inf. Theory* **14**, 462–467 (1968).

27. J. K. Johnson, D. Oyen, M. Chertkov, P. Netrapalli, Learning planar Ising models. *J. Mach. Learn. Res.* **17**, 1–26 (2016).

28. A. Montanari, Computational implications of reducing data to sufficient statistics. *Electron. J. Stat.* **9**, 2370–2390 (2015).

29. A. Montanari, J. A. Pereira, Which graphical models are difficult to learn? in *Advances in Neural Information Processing Systems*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, A. Culotta, Eds. (Curran Associates Inc., 2009), pp. 1303–1311.

30. M. Ekeberg, C. Lövkvist, Y. Lan, M. Weigt, E. Aurell, Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. *Phys. Rev. E* **87**, 012707 (2013).

31. N. P. Santhanam, M. J. Wainwright, Information-theoretic limits of selecting binary graphical models in high dimensions. *IEEE Trans. Inf. Theory* **58**, 4117–4134 (2012).

32. B. Yu, Assouad, Fano, and Le Cam, in *Festschrift for Lucien Le Cam*, D. Pollard, E. Torgersen, G. L. Yang, Eds. (Springer, 1997), pp. 423–435.

33. M. Vuffray, S. Misra, A. Lokhov, M. Chertkov, Interaction screening: Efficient and sample-optimal learning of Ising models, in *Advances in Neural Information Processing Systems*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett, Eds. (Curran Associates Inc., 2016), pp. 2595–2603.

34. L. T. Biegler, V. M. Zavala, Large-scale nonlinear programming using IPOPT: An integrating framework for enterprise-wide dynamic optimization. *Comput. Chem. Eng.* **33**, 575–582 (2009).

35. Y. Nesterov, Gradient methods for minimizing composite objective function, Technical Report, UCL (2007).

36. A. Agarwal, S. Negahban, M. J. Wainwright, Fast global convergence rates of gradient methods for high-dimensional statistical recovery, in *Advances in Neural Information Processing Systems*, J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, A. Culotta, Eds. (Curran Associates Inc., 2010), pp. 37–45.

37. M. Mezard, A. Montanari, *Information, Physics, and Computation* (Oxford Univ. Press, 2009).

38. R. Tandon, K. Shanmugam, P. K. Ravikumar, A. G. Dimakis, On the information theoretic limits of learning Ising models, in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger, Eds. (Curran Associates Inc., 2014), pp. 2303–2311.

39. P. I. Bunyk, E. M. Hoskinson, M. W. Johnson, E. Tolkacheva, F. Altomare, A. J. Berkley, R. Harris, J. P. Hilton, T. Lanting, A. J. Przybysz, J. Whittaker, Architectural considerations in the design of a superconducting quantum annealing processor. *IEEE Trans. Appl. Supercond.* **24**, 1700110 (2014).

40. M. Benedetti, J. Realpe-Gómez, R. Biswas, A. Perdomo-Ortiz, Estimation of effective temperatures in quantum annealers for sampling applications: A case study with possible applications in deep learning. *Phys. Rev. A* **94**, 022308 (2016).

41. N. Meinshausen, P. Bühlmann, High-dimensional graphs and variable selection with the lasso. *Ann. Stat.* **34**, 1436–1462 (2006).

42. S. Negahban, B. Yu, M. J. Wainwright, P. K. Ravikumar, A unified framework for high-dimensional analysis of *M*-estimators with decomposable regularizers, in *Advances in Neural Information Processing Systems*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, A. Culotta, Eds. (Curran Associates Inc., 2009), pp. 1348–1356.

43. W. Hoeffding, Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58**, 13–30 (1963).

44. A. M. Polyakov, *Gauge Fields and Strings* (Harwood Academic Publishers, 1987), vol. 140.

45. A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**, 183–202 (2009).

Science Advances

Optimal structure and parameter learning of Ising models

Andrey Y. Lokhov, Marc Vuffray, Sidhant Misra and Michael Chertkov

ARTICLE TOOLS	http://advances.sciencemag.org/content/4/3/e1700791
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2018/03/12/4.3.e1700791.DC1
REFERENCES	This article cites 32 articles, 4 of which you can access for free http://advances.sciencemag.org/content/4/3/e1700791#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service