On the Difficulty of Selecting Ising Models with Approximate Recovery

Jonathan Scarlett and Volkan Cevher
Laboratory for Information and Inference Systems (LIONS)
École Polytechnique Fédérale de Lausanne (EPFL)
Email: {jonathan.scarlett,volkan.cevher}@epfl.ch

***Abstract*—We consider the problem of estimating the underlying graph associated with an Ising model given a number of independent and identically distributed samples. We adopt an *approximate recovery* criterion that allows for a number of missed edges or incorrectly-included edges, in contrast with the widely-studied exact recovery problem. Our main results provide information-theoretic lower bounds on the sample complexity for graphs having constraints on the number of edges and maximal degree. We identify a range of scenarios where, either up to constant factors or logarithmic factors, our lower bounds match the best known lower bounds for the exact recovery criterion, several of which are known to be tight or near-tight. Hence, in these cases, approximate recovery has a similar difficulty to exact recovery in the minimax sense.**

I. Problem Statement

The problem of *graphical model selection* consists of recovering the graph structure associated with a Markov random field given a number of independent samples from the underlying distribution.

Ising model: The ferromagnetic Ising model [2] is specified by a graph $G = (V, E)$ with vertex set $V = \{1, \dots, p\}$ and edge set E. Each vertex is associated with a binary random variable $X_i \in \{-1, 1\}$, and the corresponding joint distribution is

$$P_G(x) = \frac{1}{Z} \exp\left(\sum_{i,j} \lambda x_i x_j \right), \tag{1}$$

where Z is a normalizing constant called the partition function. Here $\lambda > 0$ is a parameter to the distribution, sometimes referred to the inverse temperature.

Let $\mathbf{X} \in \{0,1\}^{n \times p}$ be a matrix of n independent samples from this distribution, each row corresponding to one such sample of the p variables. Given $\mathbf{X}$, an *estimator* or *decoder* constructs an estimate $\hat{G}$ of the graph G, or equivalently, an estimate $\hat{E}$ of the edge set E.

Approximate recovery: The exact recovery problem has been considered in a variety of previous works such as [3]–[8]. In contrast, we consider the following approximate recovery criterion, for some maximum number of errors $q_{\max} \ge 0$:

$$P_{\rm e}(q_{\max}) := \max_{G \in \mathcal{G}} \mathbb{P}\big[|E \Delta \hat{E}| > q_{\max} \big], \tag{2}$$

where $E \Delta \hat{E} = (E \backslash \hat{E}) \cup (\hat{E} \backslash E)$, so that $|E \Delta \hat{E}|$ denotes the *edit distance*, i.e., the number of edge insertions and deletions required to transform one graph to another.

Graph class: We consider the class of graphs $\mathcal{G}_{k,d}$, consisting of graphs with at most k edges and maximal degree at most d.

II. Main Results

We provide algorithm-independent lower bounds on the number of samples required for the approximate recovery of graphs in $\mathcal{G}_{k,d}$, split into two theorems for convenience. Here and subsequently, H_2 denotes the binary entropy function in nats.

Theorem 1. (Class $\mathcal{G}_{k,d}$ with $k \le p/4$) *For any maximal degree $d > 2$ and number of edges k such that $k = \omega(d^2)$ and $k \le p/4$, and any distortion level $q_{\max} = \lfloor \theta k \rfloor$ for some $\theta \in \big(0, \frac{1}{4}\frac{d-2}{d}\big)$, it is necessary that*

$$n \ge \max\left\{ \frac{e^{\lambda(d-2)/4}\big(\log 2 - H_2\big(\frac{d}{d-2} \cdot 2\theta \big) \big)}{3\lambda d^2}, \frac{2(1-\theta)\log p}{\lambda \tanh \lambda} \right\} \Big(1 - \delta - o(1) \Big) \tag{3}$$

in order to have $P_{\rm e}(q_{\max}) \le \delta$ for all $G \in \mathcal{G}_{k,d}$.

The first term in (3) reveals that the sample complexity is exponential in λd. On the other hand, if $\lambda = O\big(\frac{1}{d}\big)$ then the second term gives a sample complexity of $\Omega(d^2 \log p)$.

Theorem 2. (Class $\mathcal{G}_{k,d}$ with $k = \Omega(p)$) *For any maximal degree $d > 2$ and number of edges k such that $k = \omega(d^2)$ and $k \le \frac{1}{2}p(d'-1)$ for some $d' \le d$, and any distortion level $q_{\max} = \lfloor \theta k \rfloor$ for some $\theta \in \big(0, \frac{1}{4}\frac{d-2}{d}\big)$, it is necessary that*

$$n \ge \max\left\{ \frac{e^{\lambda(d-2)/4}\big(\log 2 - H_2\big(\frac{d}{d-2} \cdot 2\theta \big) \big)}{3\lambda d^2}, \frac{\log 2 - H_2(\theta)}{\lambda \frac{e^{2\lambda}\cosh(2\lambda d') - 1}{e^{2\lambda}\cosh(2\lambda d') + 1}} \right\} \Big(1 - \delta - o(1) \Big) \tag{4}$$

in order to have $P_{\rm e}(q_{\max}) \le \delta$ for all $G \in \mathcal{G}_{k,d}$.

By the first term, the sample complexity remains exponential in λd, and standard asymptotic expansions [1] reveal that the second term yields a sample complexity of $n = \Omega\big(\min\big\{ d^2, \frac{d^3 p^2}{k^2} \big\} \big)$.

III. Comparisons to Exact Recovery

We make the following comparisons to the information-theoretic lower bounds on exact recovery from [3], [4]:

- In all of the known cases where exact recovery is known to be difficult in the sense of being exponential in $\lambda d \to \infty$, the same difficulty is observed for approximate recovery
- In many of the cases where the necessary conditions for exact recovery lack exponential terms, the corresponding necessary conditions for approximate recovery are the same up to either constant or logarithmic factors; in particular, this is true under the scaling of Theorem 1, as well as Theorem 2 with $k \ll p\sqrt{d}$.
- In contrast, more significant gaps remain in other cases, including Theorem 2 with $p\sqrt{d} \ll k \ll pd$.

Further details on these comparisons can be found in [1], along with analyses of other graph ensembles.

This work was supported in part by the European Commission under Grant ERC Future Proof, SNF 200021-146750 and SNF CRSII2-147633. The full version can be found in [1].

REFERENCES

[1] J. Scarlett and V. Cevher, "On the difficulty of selecting Ising models with approximate recovery," 2016, accepted to *IEEE Trans. Sig. Inf. Proc. over Networks*.

[2] E. Ising, "Beitrag zur theorie des ferromagnetismus," *Zeitschrift für Physik A Hadrons and Nuclei*, vol. 31, no. 1, pp. 253–258, 1925.

[3] N. Santhanam and M. Wainwright, "Information-theoretic limits of selecting binary graphical models in high dimensions," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4117–4134, July 2012.

[4] K. Shanmugam, R. Tandon, A. Dimakis, and P. Ravikumar, "On the information theoretic limits of learning Ising models," in *Adv. Neur. Inf. Proc. Sys. (NIPS)*, 2014.

[5] A. Anandkumar, V. Y. F. Tan, F. Huang, and A. S. Willsky, "High-dimensional structure estimation in Ising models: Local separation criterion," *Ann. Stats.*, vol. 40, no. 3, pp. 1346–1375, 2012.

[6] P. Ravikumar, M. J. Wainwright, J. D. Lafferty, and B. Yu, "High-dimensional Ising model selection using ℓ_1-regularized logistic regression," *Ann. Stats.*, vol. 38, no. 3, pp. 1287–1319, 2010.

[7] G. Bresler, E. Mossel, and A. Sly, "Reconstruction of Markov random fields from samples: Some observations and algorithms," in *Appr., Rand. and Comb. Opt. Algorithms and Techniques*. Springer Berlin Heidelberg, 2008, pp. 343–356.

[8] G. Bresler, "Efficiently learning Ising models on arbitrary graphs," in *ACM Symp. Theory Comp. (STOC)*, 2015.