Which graphical models are difficult to learn? Supplementary material

José Bento
Department of Electrical Engineering
Stanford University
Stanford, CA 94305
jbento@stanford.edu

Andrea Montanari
Department of Electrical Engineering and
Department of Statistics
Stanford University
Stanford, CA 94305
montanari@stanford.edu

In this document we report some of the calculations and proofs omitted from the main paper. References to equations and theorems in the main paper are in boldface.

1 Thresholding algorithm

In the following we let $C_{ij} \equiv \mathbb{E}_{G,\theta}\{x_i x_j\}$ where expectation is taken with respect to the Ising model **(1)**.

Before proving Theorem **1.1** we start with an easy related lemma.

Lemma 1.0. If G is a tree, and $\tau(\theta) = (\tanh\theta + \tanh^2\theta)/2$, then

$$n_{\mathsf{Thr}(\tau)}(G,\theta) \leq \frac{8}{(\tanh\theta - \tanh^2\theta)^2}\ \log\frac{2p}{\delta}\,. \tag{1}$$

Proof. (Lemma 1.0) If G is a tree then $C_{ij} = \tanh\theta$ for all $(ij) \in E$ and $C_{ij} \leq \tanh^2\theta$ for all $(ij) \notin E$. The probability that $\mathsf{Thr}(\tau)$ fails is

$$1 - \mathrm{P}_{\rm succ} = \mathbb{P}_{n,G,\theta}\{\widehat{C}_{ij} < \tau \text{ for some } (i,j) \in E \textit{ or } \widehat{C}_{ij} \geq \tau \text{ for some } (i,j) \notin E\}\,. \tag{2}$$

Let $\tau = (\tanh\theta + \tanh^2\theta)/2$. Applying Azuma-Hoeffding inequality to $\widehat{C}_{ij}$ followed by union bound over the edges, we bound this probability by

$$\mathrm{P}_{\rm succ} \geq 1 - p^2\, e^{-\frac{1}{8}n(\tanh\theta - \tanh^2\theta)^2}\,. \tag{3}$$

Imposing the right hand side to be larger than δ proves our result. □

Proof. (Theorem **1.1**) We will prove that, for $\theta < \operatorname{arctanh}(1/(2\Delta))$, $C_{ij} \geq \tanh\theta$ for all $(i,j) \in E$ and $C_{ij} \leq 1/(2\Delta)$ for all $(ij) \notin E$. In particular $C_{ij} < C_{kl}$ for all $(i,j) \notin E$ and all $(k,l) \in E$. The theorem follows from this fact via union bound and Azuma-Hoeffding inequality as in the proof of Theorem 1.0.

The bound $C_{ij} \geq \tanh\theta$ for $(ij) \in E$ is a direct consequence of Griffiths inequality: compare the expectation of $x_i x_j$ in G with the same expectation in the graph that only includes edge (i,j).

The second bound is derived using the technique of **[16]**, i.e., bound C_{ij} by the generating function for self-avoiding walks on the graphs from i to j. More precisely, assume $l = \operatorname{dist}(i,j)$ and denote by $N_{ij}(k)$ the number of self avoiding walks of length k between i and j on G. Then **[16]** proves that

$$C_{ij} \leq \sum_{k=l}^{\infty}(\tanh\theta)^k N_{ij}(k) \leq \sum_{n=l}^{\infty}\Delta^{k-1}(\tanh\theta)^k \leq \frac{\Delta^{l-1}(\tanh\theta)^l}{1-\Delta\tanh\theta} \leq \frac{\Delta(\tanh\theta)^2}{1-\Delta\tanh\theta} \tag{4}$$

If $\theta < \operatorname{arctanh}(1/(2\Delta))$ the above implies $C_{ij} \leq 1/(2\Delta)$ which is our claim. □

Proof. (Theorem **1.2**) The theorem is proved by constructing G as follows: sample a uniformly random regular graph of degree Δ over the $p-2$ vertices $\{1,2,\dots,p-2\} \equiv [p-2]$. Add an extra edge between nodes $p-1$ and p. The resulting graph is not connected. We claim that for $\theta > K/\Delta$ and with probability converging to 1 as $p \to \infty$, there exist $i,j \in [p-2]$ such that $(i,j) \notin E$ and $C_{ij} > C_{p-1,p}$. As a consequence, thresholding fails.

Obviously $C_{p-1,p} = \tanh\theta$. Choose $i \in [p-2]$ uniformly at random, and j a node at a fixed distance t from i. We can compute C_{ij} as $p \to \infty$ using the same local weak convergence result as in the proof of Lemma **3.3**. Namely, C_{ij} converges to the correlation between the root and a leaf node in the tree Ising model **(16)**. In particular one can show that

$$\lim_{p\to\infty} C_{ij} \geq m(\theta)^2 \,, \tag{5}$$

where $m(\theta) = \tanh(\Delta h^*/(\Delta-1))$ and h^* is the unique positive solution of $h = (\Delta - 1)\,\mathrm{atanh}\,\{\tanh\theta\,\tanh h\}$.

The proof is completed by showing that $\tanh\theta < m(\theta)^2$ for all $\theta > K/\Delta$. $\square$

2 Regularized logistic regression

Proof. (Lemma **3.1**) We outline here the upper bound on the term R^n.

Since $\hat{\underline{\theta}}_{S^C} = 0$ an application of the mean value theorem yields $|[R^n]_j| \leq 2\Delta||\hat{\underline{\theta}}_S - \underline{\theta}^*_S||_2^2$. Now $\sigma_{\min}(Q^*) \leq 1$ so the event $\mathcal{E}$ guarantees that $\sigma_{\min}(Q^{n*}) \leq 2$. Using Lemma 3 from **[7]** we can write

$$||\hat{\underline{\theta}}_S - \underline{\theta}^*_S||_2 \leq \frac{1}{\Delta^{3/2}}\left(1 - \sqrt{1 - \lambda\frac{8\Delta^2}{C_{\min}}(1 + ||\frac{W^n_S}{\lambda}||_\infty)}\right) \,. \tag{6}$$

If $\mathcal{E}$ holds we can assume without loss of generality $||\frac{W^n_S}{\lambda}||_\infty < 1$ and since $1 - \sqrt{1-x} \leq x, x \in [0,1]$ the theorem's assumption on λ makes both $\frac{\Delta}{C_{\min}}||\frac{R^n_S}{\lambda}||_\infty$ and $|\frac{R^n_v}{\lambda}|$ smaller than $\epsilon/8$. $\square$

Proof. (Lemma **3.3**) We outline here some of the calculations with respect to the tree model **(16)**. An important property that follows from the fixed point equation $h = (\Delta-1)\,\mathrm{atanh}\,\{\tanh\theta\,\tanh h\}$ is that, if $g(\underline{x}_{\mathsf{T}(t)})$ is a function of the variables in $\mathsf{T}(t)$ then

$$\mathbb{E}_{\mathsf{T}(t),\theta,+}\{g(\underline{X}_{\mathsf{T}(t)})\} = \mathbb{E}_{\mathsf{T}(t+1),\theta,+}\{g(\underline{X}_{\mathsf{T}(t)})\}\;,, \tag{7}$$

with the obvious identification of $\mathsf{T}(t)$ as a subtree of $\mathsf{T}(t+1)$.

Let r be a uniformly random vertex in G and $i \neq j$ two neighbors of r. Using the local weak convergence property **(17)** with $t=1$ we get

$$\lim_{p\to\infty}(Q^*_{SS})_{ii} \quad\equiv\quad a = \mathbb{E}_{\mathsf{T}(1),\theta,+}\Big(\frac{1}{\cosh^2\theta M}\Big)\,, \tag{8}$$

$$\lim_{p\to\infty}(Q^*_{SS})_{ij} \quad\equiv\quad b = \mathbb{E}_{\mathsf{T}(1),\theta,+}\Big(\frac{X_iX_j}{\cosh^2\theta M}\Big)\,, \tag{9}$$

where $M \equiv \sum_{i\in\partial\mathsf{T}(1)} X_i$ is the sum of the variables on the leaves of a depth 1 tree, and $i,j \in \partial\mathsf{T}(1)$. Let c and d be defined by

$$\lim_{p\to\infty}({Q^*_{SS}}^{-1})_{ii} \quad\equiv\quad c\,, \tag{10}$$

$$\lim_{p\to\infty}({Q^*_{SS}}^{-1})_{ij} \quad\equiv\quad d\,. \tag{11}$$

Finally, for r' at distance t from r, consider the Δ-dimensional vector in

$$\lim_{p\to\infty}(Q^*_{S^cS})_{r'} = F_S(t)\,. \tag{12}$$

It can be shown that

$$F_S(t) = \mathbb{E}_{\mathsf{T}(1),\theta,+}(X_{r'})\mathbb{E}_{\mathsf{T}(1),\theta,+}\Big(\frac{X_i}{\cosh^2\theta M}\Big) + o_t(1) = f + o_t(1)\,, \tag{13}$$

where r' is the root of a first tree, and $i \in \partial\mathsf{T}(1)$ is a leaf of the second tree. In particular $F_S(t)$ has, for large t, asymptotically equal entries.

The final result is obtained by computing the quantities a, b, c, d, f. $\square$