Supplementary Materials for Interaction Screening: Efficient and Sample-Optimal Learning of Ising Models

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

[1]Theoretical Division T-4, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[2]Theoretical Division T-5, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[3]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[4]Skolkovo Institute of Science and Technology, 143026 Moscow, Russia

`{vuffray, sidhant, lokhov, chertkov}@lanl.gov`

1 Gradient Concentration

Lemma 1. *For any Ising model with p spins and for all $l \neq u \in V$*

$$\mathbb{E}\left[X_{ul}\left(\underline{\theta}^*_u\right)\right] = 0. \tag{1}$$

Proof. By direct computation, we find that

$$\begin{aligned}
\mathbb{E}\left[X_{ul}\left(\underline{\theta}^*_u\right)\right] &= \mathbb{E}\left[-\sigma_u\sigma_l \exp\left(-\sum_{i\in\partial u}\theta^*_{ui}\sigma_u\sigma_i\right)\right] \\
&= \frac{-1}{Z}\sum_{\underline{\sigma}}\sigma_u\sigma_l\exp\left(\sum_{(i,j)\in E}\theta^*_{ij}\sigma_i\sigma_j - \sum_{i\in\partial u}\theta^*_{ui}\sigma_u\sigma_i\right) = 0,
\end{aligned} \tag{2}$$

where in the last line we use the fact that the exponential terms involving σ_u cancel, implying that the sum over $\sigma_u \in \{-1,+1\}$ is zero. □

Lemma 2. *For any Ising model with p spins and for all $l \neq u \in V$*

$$\mathbb{E}\left[X_{ul}\left(\underline{\theta}^*_u\right)^2\right] = 1. \tag{3}$$

Proof. As a result of direct evaluation one derives

$$\begin{aligned}
\mathbb{E}\left[X_{ul}\left(\underline{\theta}^*_u\right)^2\right] &= \mathbb{E}\left[\exp\left(-2\sum_{i\in\partial u}\theta^*_{ui}\sigma_u\sigma_i\right)\right] \\
&= \frac{1}{Z}\sum_{\underline{\sigma}}\exp\left(\sum_{(i,j)\in E, i,j\neq u}\theta^*_{ij}\sigma_i\sigma_j - \sum_{i\in\partial u}\theta^*_{ui}\sigma_u\sigma_i\right) \\
&= \frac{1}{Z}\sum_{\underline{\sigma}}\exp\left(\sum_{(i,j)\in E, i,j\neq u}\theta^*_{ij}\sigma_i\sigma_j + \sum_{i\in\partial u}\theta^*_{ui}\sigma_u\sigma_i\right) \\
&= 1.
\end{aligned} \tag{4}$$

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

Notice that in the second line the first sum over edges (under the exponential) does not depend on σ_u. Furthermore, the first sum is invariant under the change of variables, $\sigma_u \to -\sigma_u$, while the second sum changes sign. This transformation results in appearance of the partition function in the numerator. $\square$

Lemma 3. *For any Ising model with p spins, with maximum degree d and maximum coupling intensity β, we guarantee that for all $l \neq u \in V$*

$$|X_{ul}(\underline{\theta}_u^*)| \leq \exp(\beta d). \tag{5}$$

Proof. Observe that components of $\underline{\theta}_u^*$ are smaller than β and at most d of them are non-zero. Recall that spins are binary, $\{-1, +1\}$, which results in the following estimate

$$\begin{aligned} |X_{ul}(\underline{\theta}_u^*)| &= \left| -\sigma_u \sigma_i \exp\left(-\sum_{i \in \partial u} \theta_{ui}^* \sigma_u \sigma_i \right) \right| \\ &\leq \exp\left(-\sum_{i \in \partial u} \theta_{ui}^* \sigma_u \sigma_i \right) \\ &\leq \exp(\beta d). \end{aligned} \tag{6}$$

$\square$

Lemma 4. *For any Ising model with p spins, with maximum degree d and maximum coupling intensity β. For any $\epsilon_3 > 0$, if the number of observation satisfies $n \geq \exp(2\beta d) \ln \frac{2p}{\epsilon_3}$, then the following bound holds with probability at least $1 - \epsilon_3$:*

$$\|\nabla \mathcal{S}_n(\underline{\theta}_u^*)\|_\infty \leq 2\sqrt{\frac{\ln \frac{2p}{\epsilon_3}}{n}}. \tag{7}$$

Proof. Let us first show that every term is individually bounded by the RHS of (7) with high-probability. We further use the union bound to prove that all components are uniformly bounded with high-probability. Utilizing Lemma 1, Lemma 2 and Lemma 3 we apply the Bernstein's Inequality

$$\mathbb{P}\left[\left| \frac{\partial}{\partial \theta_{ul}} \mathcal{S}_n(\underline{\theta}_u^*) \right| > t \right] \leq 2 \exp\left(-\frac{\frac{1}{2} t^2 n}{1 + \frac{1}{3} \exp(\beta d) t} \right). \tag{8}$$

Inverting the following relation

$$s = \frac{\frac{1}{2} t^2 n}{1 + \frac{1}{3} \exp(\beta d) t}, \tag{9}$$

and substituting the result in the Eq. (8) one derives

$$\mathbb{P}\left[\left| \frac{\partial}{\partial \theta_{ul}} \mathcal{S}_n(\underline{\theta}_u^*) \right| > \frac{1}{3} \left(u + \sqrt{\frac{18}{\exp(\beta d)} u + u^2} \right) \right] \leq 2 \exp(-s), \tag{10}$$

where $u = \frac{s}{n} \exp(\beta d)$.

When $n \geq s \exp(2\beta d)$ Eq. (10) can be simplified to become independent of β and d

$$\mathbb{P}\left[\left| \frac{\partial}{\partial \theta_{ul}} \mathcal{S}_n(\underline{\theta}_u^*) \right| > 2\sqrt{\frac{s}{n}} \right] \leq 2 \exp(-s). \tag{11}$$

Using $s = \ln \frac{2p}{\epsilon_3}$ and the union bound on every component of the gradient leads to the desired result. $\square$

1.1 Restricted Strong-Convexity

We recall that the remainder of the first-order Taylor-expansion of the ISO reads

$$\delta\mathcal{S}_n\left(\Delta_u,\theta^*\right)=\frac{1}{n}\sum_{k=1}^{n}\exp\left(-\sum_{i\in\partial u}\theta^*_{ui}\sigma_u^{(k)}\sigma_i^{(k)}\right)\left(\exp\left(-Y_u^{(k)}\left(\Delta_u\right)\right)-1+Y_u^{(k)}\left(\Delta_u\right)\right), \quad (12)$$

where the random variables $Y_u^{(k)}\left(\Delta_u\right)$ are i.i.d and are related to the spin configurations according to

$$Y_u\left(\Delta_u\right)=\sum_{i\in V\setminus u}\Delta_{ui}\sigma_u\sigma_i. \quad (13)$$

Lemma 5. *Consider an Ising model with p spins, with maximum degree d and maximum coupling intensity β. For all $\Delta_u\in\mathbb{R}^{p-1}$ the following bound holds*

$$\mathbb{E}\left[Y_u\left(\Delta_u\right)^2\right]\geq\frac{e^{-2\beta d}}{d+1}\left\|\Delta_u\right\|_2^2. \quad (14)$$

Proof. Our proof strategy here follows [1, Cor. 3.1]. Notice that the probability measure of the Ising model is symmetric with respect to the sign flip, i.e. $\mu\left(\sigma_1,\ldots,\sigma_p\right)=\mu\left(-\sigma_1,\ldots,-\sigma_p\right)$. Thus any spin has zero mean, which implies that for every $\Delta_u\in\mathbb{R}^{p-1}$

$$\mathbb{E}\left[\left(\sum_{i\in V\setminus u}\Delta_{ui}\sigma_i\right)\right] = 0. \quad (15)$$

This allows to reinterpret (14) as a variance, using that $\sigma_u^2=1$,

$$\begin{aligned}\mathbb{E}\left[Y_u\left(\Delta_u\right)^2\right] &= \mathbb{E}\left[\left(\sum_{i\in V\setminus u}\Delta_{ui}\sigma_i\right)^2\right]\\ &= \mathrm{Var}\left[\sum_{i\in V\setminus u}\Delta_{ui}\sigma_i\right]. \quad (16)\end{aligned}$$

Construct a subset $A\subset V$ recursively as follows: (i) let $i_0=\mathrm{argmax}_{j\in V\setminus u}\Delta_{uj}^2$ and define $A_0=\{i_0\}$, (ii) given $A_t=\{i_0,\ldots,i_t\}$, let $B_t=\{j\in V\setminus A_t\mid\partial j\cap A_t=\emptyset\}$ and $i_{t+1}=\mathrm{argmax}_{j\in B_t\setminus u}\Delta_{uj}^2$ and set $A_{t+1}=A_t\cup\{i_{t+1}\}$, (iii) terminate when $B_t\setminus u=\emptyset$ and declare $A=A_t$.

The set A possesses the following two main properties. First, every node $i\in A$ does not have any neighbors in A and, second,

$$(d+1)\sum_{i\in A}\Delta_{ui}^2\geq\sum_{i\in V\setminus u}\Delta_{ui}^2. \quad (17)$$

We apply the law of total variance to (16) by conditioning on the set of spins $\underline{\sigma}_{A^c}$ whose indexes are from the complementary set A^c.

$$\begin{aligned}\mathrm{Var}\left[\sum_{i\in V\setminus u}\Delta_{ui}\sigma_i\right] &\geq \mathbb{E}\left[\mathrm{Var}\left[\sum_{i\in V\setminus u}\Delta_{ui}\sigma_i\mid\underline{\sigma}_{A^c}\right]\right]\\ &= \sum_{i\in A}\Delta_{ui}^2\mathbb{E}\left[\mathrm{Var}\left[\sigma_i\mid\underline{\sigma}_{A^c}\right]\right], \quad (18)\end{aligned}$$

where in the last line one uses that the spins in A are conditionally independent given their neighbors $\underline{\sigma}_{A^c}$. One concludes the proof by using relation (17) and the fact that the conditional variance of a spin given its neighbors are bounded from below:

$$\begin{aligned}\mathrm{Var}\left[\sigma_i\mid\underline{\sigma}_{A^c}\right] &= 1-\tanh^2\left(\sum_{j\in\partial i}\theta^*_{ij}\sigma_j\right)\\ &\geq \exp\left(-2\beta d\right). \quad (19)\end{aligned}$$

□

Lemma 6. *Consider an Ising model with p spins, with maximum degree d and maximum coupling intensity β. For all $\Delta_u \in \mathbb{R}^{p-1}$ and for all $\epsilon_4 > 0$, the remainder of the Taylor expansion* (12) *satisfies with probability at least $1 - \epsilon_4$, the following inequality*

$$\delta \mathcal{S}_n \left(\Delta_u, \theta_u^*\right) \geq \left(2 + \|\Delta_u\|_1\right)^{-1} \left(\frac{e^{-3\beta d}}{d+1} \|\Delta_u\|_2^2 - 2\sqrt{\frac{\ln \frac{1}{\epsilon_4}}{n}} \|\Delta_u\|_1^2 \right), \tag{20}$$

whenever $n \geq 4 \exp\left(2\beta d\right) \ln \frac{1}{\epsilon_4}$.

Proof. This concentration property is based on the Bernstein's inequality. First of all, observe that for all $z \in \mathbb{R}$, the following bound holds

$$\left(2 + |z|\right) \left(e^{-z} - 1 + z\right) \geq z^2. \tag{21}$$

This implies that the remainder of the Taylor expansion (12) is lower-bounded

$$\delta \mathcal{S}_n \left(\Delta_u, \theta_u^*\right) \geq \frac{1}{n} \sum_{k=1}^{n} \exp\left(- \sum_{i \in \partial u} \theta_{ui}^* \sigma_u \sigma_i \right) \frac{Y_u^{(k)} \left(\Delta_u\right)^2}{2 + \left| Y_u^{(k)} \left(\Delta_u\right) \right|}. \tag{22}$$

Notice that the support of $Y_u \left(\Delta_u\right)$ is trivially upper bounded for all $\Delta_u \in \mathbb{R}^{p-1}$

$$\left| Y_u \left(\Delta_u\right) \right| \leq \|\Delta_u\|_1 . \tag{23}$$

It implies that the expression (22) can be lower-bounded by a quadratic expression in $Y_u^{(k)}$

$$\delta \mathcal{S}_n \left(\Delta_u, \theta_u^*\right) \geq \left(2 + \|\Delta_u\|_1\right)^{-1} \frac{1}{n} \sum_{k=1}^{n} \exp\left(- \sum_{i \in \partial u} \theta_{ui}^* \sigma_u \sigma_i \right) Y_u^{(k)} \left(\Delta_u\right)^2 . \tag{24}$$

Then extending the technique of the type used in Lemma 2, one shows that

$$\begin{aligned} \mathbb{E} \left[\exp\left(-2 \sum_{i \in \partial u} \theta_{ui}^* \sigma_u \sigma_i \right) Y_u \left(\Delta_u\right)^4 \right] =& \mathbb{E} \left[Y_u \left(\Delta_u\right)^4 \right] \\ \leq& \|\Delta_u\|_1^4 . \end{aligned} \tag{25}$$

To finish the proof we apply the Bernstein's inequality to the right-hand side of (24), thus combining relations (25), (23) and Lemma 5. Moreover when $n \geq 4e^{2\beta d} \ln \frac{1}{\epsilon_4}$ further simplifications can be made in the way similar to the one used in Lemma 4. $\square$

Lemma 7. *Consider an Ising model with p spins, with maximum degree d and maximum coupling intensity β. For all $\epsilon_4 > 0$, when $n \geq 2^{12} d^2 \left(1 + d\right)^2 e^{6\beta d} \ln \frac{1}{\epsilon_4}$ the ISO satisfies, with probability at least $1 - \epsilon_4$, the restricted strong convexity condition*

$$\delta \mathcal{S}_n \left(\Delta_u, \theta_u^*\right) \geq \frac{e^{-3\beta d}}{4\left(d+1\right)\left(1 + 2\sqrt{d} R\right)} \|\Delta_u\|_2^2 , \tag{26}$$

for all $\Delta_u \in \mathbb{R}^{p-1}$ such that $\|\Delta_u\|_1 \leq 4\sqrt{d} \|\Delta_u\|_2$ and $\|\Delta_u\|_2 \leq R$ with $R > 0$.

Proof. We prove it applying Lemma 6 directly to $\|\Delta_u\|_1 \leq 4\sqrt{d} \|\Delta_u\|_2$ and $\|\Delta_u\|_2 \leq R$ when $n \geq 2^{12} d^2 \left(1 + d\right)^2 e^{6\beta d} \ln \frac{1}{\epsilon_4}$. $\square$

References

[1] A. Montanari, "Computational implications of reducing data to sufficient statistics," *Electron. J. Statist.*, vol. 9, no. 2, pp. 2370–2390, 2015.