Comparing bounds for the Mean-field Gap for the Ising model

David Kewei Lin

A thesis presented for the degree of
Bachelor's of Science with Honors

Mathematics Department
Stanford University

Acknowledgements

I would like to thank Professor Andrea Montanari for taking me on as a student for the senior thesis. He has opened my eyes to the amazing world of statistical physics, and throughout the course of preparing this thesis I have learnt so much from him, from grand theories and principles to the how-tos of basic gritty computations, and so I offer him my heartfelt thanks.

A special shoutout goes to Jensen Wang, who taught me some basic statistical mechanics and discussed some of the probability questions that I had.

My remaining thanks go to my friends and family, who have been there to support me throughout this journey.

Contents

1 Introduction **3**

2 Notation **5**

3 Log-partition bounds from literature **6**

4 Overview of Results **10**
- 4.1 The Ising Model on the Hypercube . . . 10
- 4.2 Models . . . 11
- 4.3 Summary of Results . . . 14
- 4.4 Discussion . . . 15

5 Computation **16**
- 5.1 Ising Model Computations . . . 16
  - 5.1.1 Mean-width . . . 16
  - 5.1.2 Remaining results . . . 17
- 5.2 Specific Model Computations . . . 20
  - 5.2.1 ℓ_2-covering numbers . . . 20
  - 5.2.2 ℓ_1-covering numbers . . . 21
  - 5.2.3 Variational problem in Eldan's bound . . . 22
  - 5.2.4 Limiting spectral distributions . . . 23
  - 5.2.5 Limit version of Eldan's quantity . . . 24
  - 5.2.6 Limit version of covering number bounds . . . 24
  - 5.2.7 Computation for lattice model . . . 25

1 Introduction

Ising Model. The *Ising model* is a probability distribution over the hypercube $\{-1,+1\}^n$. We think of these as the spins of n particles. Furthermore, we allow pairs of particles can influence each other. This is captured by the (negative) Hamiltonian or energy $f : \{\pm 1\}^n \to \mathbb{R}$:

$$f(\sigma) = \sum_{i<j} J_{ij}\sigma_i\sigma_j = \frac{1}{2}\sigma^\top J\sigma$$

where $\sigma = (\sigma_1, \sigma_2, ..., \sigma_n) \in \{\pm 1\}^n$. The signs of the coefficients $\{J_{ij}\}$ represents the nature of the interaction between particles i and j: $J_{ij} > 0$ is a preference towards alignment and $J_{ij} < 0$ is a preference towards disalignment. This produces a probability measure μ (implicitly dependent on $\beta \in \mathbb{R}, J = \{J_{ij}\}_{i<j}$):

$$\mu(\sigma) \propto e^{\beta f(\sigma)}.$$

We would like to estimate the log-partition function

$$\log Z = \log \sum_{\sigma \in \{\pm 1\}^n} e^{\beta f(\sigma)}$$

and typically we would like this to be accurate up to $o(n)$.

Gibbs Variational Principle. By Jensen's inequality, we have the naive inequality

$$\begin{aligned}\log Z = \log \sum_{\sigma} e^{\beta f(\sigma)} &= n \log 2 + \log \mathop{\mathbb{E}}_{\sigma \sim \mathsf{Unif}\{\pm 1\}^n} e^{\beta f(\sigma)} \\ &\geq n \log 2 + \beta \cdot \mathop{\mathbb{E}}_{\sigma \sim \mathsf{Unif}\{\pm 1\}^n} f(\sigma) = n \log 2\end{aligned}$$

This is, unfortunately, potentially a very loose bound. In the limit $\beta \to \infty$, the sum in Z is dominated by $e^{\beta \max_\sigma f(\sigma)}$, so we expect $\log Z \approx \beta \max_\sigma f(\sigma)$, and this bound misses the dependence on β completely.

By taking more care with the choice of distribution, we can obtain a much better bound. Let μ be some distribution over $\{\pm 1\}^n$, then:

$$\begin{aligned}\log Z = \log \sum_{\sigma} e^{\beta f(\sigma)} &= \log \sum_{\sigma} e^{\beta f(\sigma) - \log \mu(\sigma)} \cdot \mu(\sigma) \\ &\geq \sum_{\sigma} (\beta f(\sigma) - \log \mu(\sigma)) \cdot \mu(\sigma) \\ &= \beta \cdot \mathop{\mathbb{E}}_{\sigma \sim \mu} f(\sigma) + H(\mu)\end{aligned}$$

where $H(\mu) = \mathbb{E}_{\sigma \sim \mu}[-\log \mu(\sigma)]$ is the entropy of μ. Equality is attained when $\mu(\sigma) \propto e^{\beta H(\sigma)}$, so this is often written as:

$$\log Z = \sup_{\mu} \left\{ \beta \cdot \mathop{\mathbb{E}}_{\sigma \sim \mu} f(\sigma) + H(\mu) \right\}$$

where we take supremum over all distributions over the hypercube $\{\pm 1\}^n$.

Mean-field hypothesis. The space of distributions over $\{\pm 1\}^n$ is extremely large. Is there a smaller distribution class that gives the correct answer?

The mean-field hypothesis is that under certain assumptions, the Gibbs inequality is approximately saturated by product measures, i.e.

$$\log Z \approx \sup_{\xi \text{ prod.}} \{ \beta \cdot \mathop{\mathbb{E}}_{\sigma \sim \xi} f(\sigma) + H(\xi) \}$$

There are many theorems in the literature that give an upper bound for the gap between $\log Z$ and the variational term on the RHS. For an accurate estimation of the free energy $\frac{1}{n} \log Z$, one would like this estimation to be $o(n)$.

Goal of the project. The goal of this project is to compare various upper bounds in the literature on the Ising model. One might hope for a single upper bound to be the tightest (up to order), but this is difficult to establish for quantities which are variational or covering-related.

For the sake of feasibility, we instead attempt to evaluate the upper bounds for specific J (which are either deterministic or random graph ensembles), and establish their behavior in terms of the number of particles n, the inverse temperature β and other parameters that J may depend on.

Applications. Estimation of the normalizing constant is a key step in the computations of probabilities in various random models (e.g. factor models). When the normalizing constant is computable, one can also compute, for instance, the marginal probabilities.

The first bound of this type appeared in [CD16], which demonstrated a large deviation principle for nonlinear functions. Suppose one had a nonlinear function $g : \{\pm 1\}^n \to \mathbb{R}$, and we would like to estimate the upper tails of $g(X)$ for $X \sim \mathsf{Unif}\{\pm 1\}^n$. By picking f to be a smooth "cut-off" function of the form

$$f(\sigma) = \begin{cases} 0 & \text{if } g(\sigma) \geq tn \\ \text{large negative number} & \text{if } g(\sigma) < tn \end{cases} \tag{1}$$

then $e^{f(\sigma)}$ is approximately the indicator $\mathbb{I}_{g(\sigma) \geq tn}$, so $\log Z \approx \log \mathbb{P}(g(X) \geq tn)$.

Using suitable g, one can encode bounds for various subgraph counts. The authors were able to obtain new results about the upper tails of triangle counts in the Erdős-Renyi graph, as well as the upper tail of the number of 3-term arithmetic progression in a random subset of integers.

Such bounds can also give better upper tails for random matrices. In [Aug20], a large deviation principle was established for the trace of Wigner matrices.

2 Notation

- For compactness, we will write $\{\pm 1\}$ instead of $\{-1, 1\}$ and $[\pm 1]$ instead of $[-1, 1]$.
- When we say that J has a limiting spectrum supported on $[-f(n), f(n)]$, we really mean that $\frac{1}{f(n)} J$ has a limiting spectrum supported on $[\pm 1]$.
- **Asymptotic notation.** All of the following are assumed to hold uniformly across other free parameters
  - $f(n) \asymp g(n)$ for $f(n) = \Theta(g(n))$
  - $f(n) \sim g(n)$ for $f(n) = (1 + o(1))g(n)$.
  - $f(n) \lesssim g(n)$ for $f(n) = O(g(n))$
  - $f(n) \ll g(n)$ for $f(n) = o(g(n))$
- **Matrix shorthands.**
  - J° denotes the matrix formed by the off-diagonal elements of J, so

$$J^\circ_{ij} = \begin{cases} J_{ij} & i \neq j \\ 0 & i = j \end{cases}$$

  - J_i denotes the i-th row of matrix J as a vector.
  - $\|J\|_F$ is the Frobenius norm of J.
  - $\lambda_i(J)$ is the i-th largest eigenvalue of J by absolute value. (For well-definedness, we always write $|\lambda_i(J)|$.)
- **Covering numbers.** $N(S, r, \ell)$ denotes the minimum number of balls of radius r (under the metric ℓ) required to cover the set S.
- **Standard functions.**
  - $(x)_+ = \max\{x, 0\}$
- **Convergence of random variables.**
  - $X_n \overset{d}{\Rightarrow} X$: $\{X_n\}$ converges to X in distribution.

3 Log-partition bounds from literature

Relative version of Gibbs. Instead of estimating $\log Z$, we can opt to estimate $\log \frac{Z}{2^n} = \log Z - n \log 2$. This can be interpreted as the log expected value of $e^{f(\sigma)}$ for σ drawn from the uniform distribution on $\{\pm 1\}^n$.

In general, for a product space $K = K_1 \times ... \times K_n$ with an ambient distribution ν and a test distribution $\mu \gg \nu$, we have the analogous bound

$$\log \int e^{f(\sigma)} d\nu(\sigma) = \log \int e^{\beta \cdot f(\sigma)} d\nu(\sigma) \tag{2}$$

$$= \log \int e^{f(\sigma) - \log \frac{d\nu}{d\mu}(\sigma)} d\mu(\sigma) \tag{3}$$

$$\geq \mathop{\mathbb{E}}_{\sigma \sim \mu} \left[f(\sigma) - \log \frac{d\mu}{d\nu}(\sigma) \right] \tag{4}$$

$$= \mathop{\mathbb{E}}_{\sigma \sim \mu} f + D_{\mathrm{KL}}(\mu || \nu) \tag{5}$$

This implies that the mean-field gap

$$\mathsf{GAP} := \log \int e^{f(\sigma)} d\nu(\sigma) - \sup_{\xi \text{ prod.}} \left\{ \mathop{\mathbb{E}}_{\xi} f + D_{\mathrm{KL}}(\xi || \nu) \right\} \tag{6}$$

is always non-negative. Inequalities presented below from the literature are upper bounds of GAP in this context under various assumptions.

We will use the following shorthand for these two recurring cases:

- $K = [-1, 1]^n, \nu = \mathsf{Unif}(\{\pm 1\}^n)$, which we call the hypercube.
- $K = \{-1, 1\}^n, \nu = \mathsf{Unif}(\{\pm 1\}^n)$, which we call the discrete hypercube.

In the former case, some results deal with a slightly different gap (where the mean of the distribution ξ is passed directly into f):

$$\mathsf{GAP}^* := \log \int e^{f(\sigma)} d\nu(\sigma) - \sup_{\xi \text{ prod.}} \{ f(\mathop{\mathbb{E}}_{\xi} \sigma) + D_{KL}(\xi \| \nu) \} \tag{7}$$

$$= \log \int e^{f(\sigma)} d\nu(\sigma) - \sup_{\xi \text{ prod.}} \{ \mathop{\mathbb{E}}_{\xi} f + \frac{1}{2} \sum_{i=1}^{n} J_{ii}(1 - m_i^2) + D_{KL}(\xi \| \nu) \} \tag{8}$$

where $m = \mathbb{E}_\xi \sigma$ is the average magnetization in ξ.

This difference will not matter too much when we apply the result in Subsection 4.1, since

$$|\mathsf{GAP} - \mathsf{GAP}^*| \leq \sum_{i=1}^{n} \frac{1}{2} |J_{ii}|$$

which is typically an order smaller than GAP.

Results. We list the various results in literature below:

Theorem 1 ([CD16], Theorem 1.6)**.** *For the hypercube, if f twice continuously differentiable, we have*

$$\mathsf{GAP}^* \leq (\textit{complexity term}) + (\textit{smoothness term}) \tag{9}$$

where

$$(\textit{complexity term}) := \left(n\sum_{i=1}^n b_i^2\right)^{1/2}\varepsilon + \frac{3}{2}n\varepsilon + \log|\mathcal{D}(\varepsilon)| \tag{10}$$

$$(\textit{smoothness term}) := 8\left(\sum_{i=1}^n \left(ac_{ii}+b_i^2\right) + \sum_{i,j=1}^n \left(ac_{ij}^2 + b_ib_jc_{ij} + 2b_ic_{ij}\right)\right)^{1/2} + 2\left(\sum_{i=1}^n b_i^2\right)^{1/2}\left(\sum_{i=1}^n c_{ii}^2\right)^{1/2} + 12\sum_{i=1}^n c_{ii} + \log 2 \tag{11}$$

and

$$a = \|f\|, b_i = \left\|\frac{\partial f}{\partial\sigma_i}\right\|, c_{ij} = \left\|\frac{\partial^2 f}{\partial\sigma_i\partial\sigma_j}\right\|$$

where $\|\cdot\|$ *is the sup-norm for a function and* $\mathcal{D}(\varepsilon)$ *is a* $\sqrt{n}\varepsilon$*-covering of* $\nabla f(\{\pm1\}^n)$ *under the* ℓ_2*-distance.*

Theorem 2 ([Yan20])**.** *Suppose each* V_i *is a Banach space with a norm* $\|\cdot\|_{V_i}$ *and a probability measure* ν_i *supported on a convex compact set* $K_i \subset V_i$ *with diameter at most* M*. Let*

$$\|x\|_V := \max_{i\in[n]}\|x_i\|_{V_i}$$

be the induced norm on V*, and suppose that* $f : K \to \mathbb{R}$ *is twice Fréchet differentiable. Furthermore, let* $m(\xi_i) \in V_i$ *denote the mean of* ξ_i *for any probability distribution* $\xi_i \ll \mu_i$*, and for a product distribution on* V *define*

$$m(\xi) := (m(\xi_1), m(\xi_2), ..., m(\xi_n)).$$

Then, for any $\varepsilon > 0$*,*

$$\mathsf{GAP}^* := \log\int_K e^{f(x)}d\nu(x) - \max_{\xi\ll\nu,\xi\ \textit{prod.}}\left\{f(m(\xi)) - \sum_{i=1}^n D(\xi_i\|\nu_i)\right\} \tag{12}$$

$$\leq B_1 + B_2 + \log 2 + \log|\mathcal{D}(\varepsilon)| \tag{13}$$

where

$$B_1 := 4\left(M^2\left(a\sum_{i=1}^n c_{ii} + \sum_{i=1}^n b_i^2\right) + M^3\sum_{i,j=1}^n b_ic_{ij} + M^4\left(a\sum_{i,j=1}^n c_{ij}^2 + \sum_{i,j=1}^n b_ib_jc_{ij}\right)\right)^{\frac{1}{2}}, \tag{14}$$

$$B_2 := 4\left(\sum_{i=1}^n b_i^2 + \varepsilon^2 n\right)^{\frac{1}{2}}\left(M^3\left(\sum_{i=1}^n c_{ii}^2\right)^{\frac{1}{2}} + M^2n^{\frac{1}{2}}\varepsilon\right) + \sum_{i=1}^n M^2c_{ii} + Mn\varepsilon, \tag{15}$$

and the other parameters satisfy

$$a := \sup_{x\in K} f(x) \tag{16}$$

$$b_i := \sup_{x\in K}\sup_{\|r_i\|_{V_i}\leq1}|f_i(x)(r_i)| \tag{17}$$

$$c_{ij} = \sup_{x\in K}\sup_{\|z_i\|_{V_i},\|r_j\|_{V_i}\leq1}|f_{ij}(x)(z_i, r_j)| \tag{18}$$

where f_i*,* f_{ij} *are the first and second partial derivatives of* f *respectively, and interpreting* $\nabla f = (f_i(\cdot))_i$ *as a mapping* $\nabla f : K \to \prod_i L(V_i, \mathbb{R})$*,* $\mathcal{D}(\varepsilon)$ *is a* $\sqrt{n}\varepsilon$*-covering of* $\nabla f(K)$ *under the norm*

$$\|g\|_{\prod_i L(V_i,\mathbb{R})} = \left(\sum_{i=1}^n \sup_{\|r_i\|_{V_i}\leq1}|g_i(r_i)|^2\right)^{1/2}. \tag{19}$$

Theorem 3 ([Aug20], Theorem 1.1). *For f continuous differentiable on the hypercube,*

$$\mathsf{GAP}^* \leq \log|\mathcal{D}_\delta| + \delta \tag{20}$$

where $\mathcal{D}_\delta$ is a δ/D-net of the convex hull of $\nabla f(\operatorname{supp}\mu)$ and D is the diameter of $\operatorname{supp}\mu$.

Corollary 4. *There exists an absolute constant $\kappa > 0$ for which*

$$\mathsf{GAP}^* \leq \kappa n^{1/3}\, \mathsf{GW}(\nabla f(\operatorname{supp}\mu))^{2/3} \tag{21}$$

where $\mathsf{GW}(V)$ refers to the **Gaussian width** *of a set V, defined by*

$$\mathsf{GW}(V) := \mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,I)} \sup_{v\in V} \langle x, v\rangle \tag{22}$$

Theorem 5 ([Aus19], Prop 5.1). *If each K_i has diameter ≤ 1,*

$$\mathsf{GAP} \leq (\epsilon+\delta)n + \sqrt{\frac{\epsilon+\delta}{2}} \cdot L$$

where f is L-Lipschitz for the Hamming metric $d_n = \frac{1}{n}\sum_{i=1}^n d_{K_i}$, and the set $\{\nabla f(x,\cdot) : x \in K\}$ can be covered by $e^{\epsilon n}$ sets of $\|\cdot\|$-diameter less than δn.

Here, $\nabla f(x,\cdot)$ is the discrete derivative with respect to some fixed reference element $* \in K$:

$$\nabla f(x,y) := \sum_{i=1}^n \partial_i f(x, y_i) \tag{23}$$

$$\partial_i f(x,y_i) := (f(x_1,\dots,x_{i-1},y_i,x_{i+1},\dots,x_n) - f(x_1,\dots,x_{i-1},*_i,x_{i+1},\dots,x_n)) \tag{24}$$

and the norm $\|\cdot\|$ induced on $\{\nabla f(x,\cdot\}$ is the sup norm of functions on K, that is,

$$\|\nabla f(x,\cdot) - \nabla f(y,\cdot)\| = \sup_z |\nabla f(x,z) - \nabla f(y,z)| \tag{25}$$

Theorem 6 ([Eld18], Corollary 2). *For the discrete hypercube, we have*

$$\mathsf{GAP} \leq 64\, \mathsf{Lip}(f)^{2/3}\, \mathsf{GW}_+(\nabla f\{\pm 1\}^n) \tag{26}$$

where:

$$\begin{aligned}
\partial_i f(y) &:= \frac{1}{2}(f(y_1,\dots,y_{i-1},1,y_{i+1},\dots,y_n) - f(y_1,\dots,y_{i-1},-1,y_{i+1},\dots,y_n)) \\
\mathsf{Lip}(f) &:= \max_{i,y} |\partial_i f(y)| \\
\nabla f(y) &:= (\partial_i f(y))_i \\
\mathsf{GW}_+(V) &:= \mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,I)} \sup_{v\in V} (\langle x, v\rangle)_+
\end{aligned}$$

Theorem 7 ([Aug21], Theorem 1). *For the hypercube (and f differentiable),*

$$\mathsf{GAP} \leq 4 \cdot \mathsf{RW}(\nabla f(\{\pm 1\}^n))$$

where $\mathsf{RW}(V)$ denotes the **Rademacher mean-width** *of V:*

$$\mathsf{RW}(V) := \mathop{\mathbb{E}}_{x\sim\mathsf{Unif}\{\pm 1\}^n} \sup_{x,v} \langle x, v\rangle \tag{27}$$

Theorem 8 ([JKR19], Theorem 1.1). *For the Ising model ($f(\sigma) = \frac{1}{2}\sigma^\top J\sigma$) on the hypercube ($K = \{\pm 1\}^n$):*

$$\mathsf{GAP} \leq 3n^{2/3}\|J\|_F^{2/3}$$

Theorem 9 ([Eld20], Theorem 5 and Corollary 6)**.** *For $K = (\mathbb{R}^k)^n$ with underlying measure $\nu^{\otimes n}$ (where ν is a measure on $\mathbb{R}^k$ with compact support), and the Potts model*

$$f(\sigma) = \sum_{i,j=1}^{n} J_{ij} \langle \sigma_i, \sigma_j \rangle + \sum_{i=1}^{n} \langle h_i, \sigma_i \rangle \tag{28}$$

on (and $\mu \propto e^f$ the Gibbs distribution),

$$\mathsf{GAP} \leq 3 \log \det\Big(\mathrm{Cov}(\mu)\tilde{J} + I\Big)$$

where $\tilde{J} = (J^2)^{1/2}$ is the matrix absolute-value obtained by switching the eigenvalues for their absolute values.

Since μ is difficult to compute directly, the following three corollaries are also true:

$$\mathsf{GAP} \leq 3 \max\left\{\sum_{i\in[n]} \log(\beta_i|\lambda_i(J)| + 1); \beta_i \geq 0, \sum \beta_i \leq S\right\} \tag{29}$$

$$\mathsf{GAP} \leq 10 \cdot \frac{p+1}{p}(D^2 n\|J\|_{S_p}^{p/(p+1)}) \tag{30}$$

$$\mathsf{GAP} \leq 3\,\mathrm{rank}(J) \log\big(D^2 n\|J\|_{S_\infty} + 1\big) \tag{31}$$

where D refers to the diameter of the support of ν, and $\|J\|_{S_p}$ refers to the p-th **Schatten norm**

$$\|J\|_{S_p} := \left(\sum_{i\in[n]} |\lambda_i(J)|^p\right)^{1/p} \tag{32}$$

and $\{\lambda_i(J)\}$ are the eigenvalues of J.

4 Overview of Results

The goal here is to determine is to attempt to compare the different bounds (at least under various specific assumptions) up to order. A summary of the methodology is as follows:

4.1 The Ising Model on the Hypercube

The common denominator of all the theorems gathered in Section 3 is that they all specialize to statements about the Ising model on the hypercube, so we will do so for the sake of comparison. For each result, we would like to establish the upper bound for GAP (or GAP^*) up to order given the (symmetric) interaction matrix J.

We summarize the results in the following table, and defer the computations to Subsection 5.1.

Result	Upper bound for GAP (up to order)
[CD16]	$\inf_{\varepsilon>0}\left\{\left(\frac{1}{n}\sum_{i=1}^n \\|J_i\\|_1^2\right)^{1/2}\varepsilon+\varepsilon+\log N(J\{\pm1\},\varepsilon/\sqrt{n},\ell_2)\right\}+$ $\Big(\sum_{i=1}^n \\|J_i\\|_1^2+\sqrt{n}\\|J\\|_{\mathrm{op}}(\sum_{i=1}^n \lvert J_{ii}\rvert+\\|J\\|_F)+\sum_{i,j=1}^n \\|J_i\\|_1\\|J_j\\|_1\lvert J_{ij}\rvert$ $+\left(\sum_{i=1}^n \\|J_i\\|_1^2\right)\left(\sum_{i=1}^n J_{ii}^2\right)+\left(\sum_{i=1}^n J_{ii}^2\right)^2+1\Big)^{1/2}$
[Yan20]	$\inf_{\varepsilon>0}\left\{\frac{\varepsilon^2}{n}+\varepsilon\left(1+\frac{1}{n}\sum_i\left(J_{ii}^2+\\|J_i\\|_1^2\right)\right)^{1/2}+\log N(J[\pm1]^n,\varepsilon/\sqrt{n},\ell_2)\right\}$ $+\Big(\sqrt{n}\\|J\\|_{\mathrm{op}}\left(\sum_i \lvert J_{ii}\rvert+\\|J\\|_F^2\right)+\sum_i \\|J_i\\|_1^2$ $+\sum_{i,j}\\|J_i\\|_1\\|J_j\\|_1\lvert J_{ij}\rvert+\left(\sum_{i=1}^n \\|J_i\\|_1^2\right)\left(\sum_{i=1}^n J_{ii}^2\right)+1\Big)^{1/2}$
[BM17]	$o(n)$ if $\\|J\\|_F^2=o(n)$
[JKR19]	$n^{2/3}\\|J\\|_F^{2/3}$
[Aug20]	$\inf_{\delta}\{\delta+\log N(J[-1,1]^n,\frac{\delta}{2\sqrt{n}},\ell_2)\}$
[Aus19]	$C+\sqrt{C/n}\max_i \\|J_i^\circ\\|_1$ for $C=\inf_\delta\{\delta+\log N(J\{\pm1\}^n,\delta/2,\ell_1)\}$
[Aug20] Corr.	$n^{1/3}(\sum_i \\|J_i\\|_2)^{2/3}$
[Eld18]	$n^{2/3}(\max_i \\|J_i^\circ\\|_1)^{2/3}(\sum_i \\|J_i\\|_2)^{1/3}$
[Aug21]	$\sum_i \\|J_i\\|_2$
[Eld20] C1	$\max\{\sum_i \log(c_i\lvert\lambda_i(J)\rvert+1); c_i\ge 0,\sum c_i\le 4n\}$
[Eld20] C2	$\inf_{p>0}\frac{p+1}{p}n\\|J\\|_{S_p}^{p/(p+1)}$
[Eld20] C3	$\mathrm{rank}(J)\log(1+n\lvert\lambda_1(J)\rvert)$

To recap the relevant notation:

- J° denotes the off-diagonal matrix of J.
- J_i denotes the i-th row of matrix J as a vector.
- $N(S,r,\ell)$ denotes the minimum number of balls of radius r (under the metric ℓ) required to cover the set S.

Some bounds require us to compute the log-covering number for $J\{\pm1\}^n$ or $J[\pm1]^n$. This is usually difficult to do exactly (even up to exact order), so for the sake of tractability we will relax the set into an ellipsoid E defined by

$$E := J(\sqrt{n}B_n) \supset J[-1,1]^n. \tag{33}$$

From the perspective of finding the "best" bound, we can already make some statements about relative effectiveness:

- [Eld20] C2 and C3 follow from C1 (since they are both corollaries in the original paper).
- [BM17] follows from [JKR19], which in turn is a special case of [Eld20] Corr. 2 when $p = 2$.
- [CD16] contains [Aug20] as a term, assuming we use the same ellipsoidal relaxation for both $J\{\pm 1\}^n$ and $J[\pm 1]^n$.
- By Cauchy-Schwarz, [Aug20] Corr. is a better bound than [JKR19]. Furthermore, on every model considered, each $\|J_i\|_2$ will be approximately the same value, so equality holds up to order.
- In Subsubsection 5.2.2, we argue that

  $$\inf_\delta\{\delta + \log N(E, \delta, \ell_1)\} \asymp \inf_\delta\{\delta + \log N(E, \delta/\sqrt{n}, \ell_2)\}$$

  where E is the ellipsoidal relaxation of $J\{\pm 1\}$.

  For the special case where the spectra is composed on one large eigenvalue and eigenvalues of the same order, this value also matches the quantity in [Eld20] Corr. 1 (see Subsubsection 5.2.3).

4.2 Models

Since these bounds are difficult to compare directly, we would like to apply them on specific Ising models (by selecting the matrix J).

These models will be implicitly indexed by n, and we will be interested solely in the $n \to \infty$ limit. We will also impose scaling so that each particle experiences $\Theta(1)$ influence from other particles in total (i.e. $1/\deg$ for ferromagnetic models and $1/\sqrt{\deg}$ for spin glasses, where deg is degree).

For each of these models, we describe the matrix and its spectral properties. For our purposes, it will be important to know the order of the top eigenvalue and the majority (e.g. $\Theta(1)$ fraction) of eigenvalues.

Curie-Weiss Model. This corresponds to selecting

$$J = \frac{\beta}{n}\mathbf{1}\mathbf{1}^\top \tag{34}$$

This is a rank 1 matrix with eigenvalue β. This also corresponds to the adjacency matrix of a complete graph.

Sherrington-Kirkpatrick Model. Let GOE(n) describe the random symmetric matrix G with

$$G_{ij} \begin{cases} \sim \mathcal{N}(0,1) & i < j \\ \sim \mathcal{N}(0,2) & i = j \\ = G_{ji} & i > j \end{cases} \tag{35}$$

where $\{G_{ij}\}_{i \le j}$ are drawn independently.

Set

$$J \sim \frac{\beta}{\sqrt{n}} \cdot \mathrm{GOE}(n) \tag{36}$$

It is known that the top eigenvalue of J is $(2 + o(1))\beta$ with high probability, and the limit spectral density follows the semicircle law supported on $[-2\beta, 2\beta]$. This tells us that with high probability, we expect $\Omega(n)$ eigenvalues of order β.

Ferromagnetic graph models. Let G be a graph. We can define the adjacency matrix $\mathsf{Adj}(G)$ by

$$(\mathsf{Adj}(G))_{ij} = \begin{cases} 1 & \text{if } (ij) \text{ is an edge} \\ 0 & \text{otherwise} \end{cases} \tag{37}$$

With this assumption, we can set

$$J = \frac{\beta}{(\text{average degree})} \mathsf{Adj}(G)$$

for a variety of different graphs G:

- Complete graph; coincides with the Curie-Weiss model.
- The Erdős-Rényi graph $G(n, p)$, where each edge is present independently with probability p.
- Random regular graph of degree d
- For $n = m^d$, the k-dimensional lattice with periodic boundary conditions.

Erdős-Rényi graph. An Erdős-Rényi graph $G(n, p)$ is formed by manifesting each of the $\binom{n}{2}$ edges independently with probability p. We will also let $d \sim np$ denote the average degree.

The typical qualitative structure of $G(n, p)$ was first described in [ER+60]. As the average degree np increases, three "regimes" of an Erdős-Rényi graph:

- For $d \ll 1$, the graph is formed by trees of size $O(\log n)$.
- For $d = 1$, the graph has one large connected component of size $\Theta(n^{2/3})$.
- For constant $d > 1$, there is a unique component of size linear in n, and all other vertices are in a tree of size $O(\log n)$.
- For $d > \log n$, the graph is connected with high probability.

This has the following implications for the spectra [Zha12]:

- In all cases, there is a limiting spectra of order β (assuming $p \leq 1/2$).
- For $np = \alpha$ constant:
  - For $\alpha < 1$, the spectra is discrete and explained by the small trees.
  - For $\alpha > 1$, then the spectra is a mix of a discrete spectrum and a continuous spectrum. In general, the continuous spectrum comes from the giant connected component, and the discrete spectrum comes from trees / stubs of trees connected to the giant component.
- For $\alpha \to \infty$, the limiting spectra tends to the semicircular law.
- For $\alpha \gg \log n$, [BBK20] gives that the second largest eigenvalue and the smallest eigenvalues to the edges of the support of the asymptotic eigenvalue distribution. This does not hold when $np \ll \log n$ or $np \asymp \log n$ but [BBK19; ADK21] provides some qualitative understanding of the tail in excess of the asymptotic distribution.

Random regular graph of degree d. Clearly, for a regular graph G of degree d, $\mathsf{Adj}(G)$ has an eigenvalue of size d. If G was a random regular graph for fixed $d \geq 2$, as $n \to \infty$ $\mathsf{Adj}(G)$ has a continuous spectrum following the Kesten-McKay law supported on $[\pm 2\sqrt{d-1}]$ [McK81]. Furthermore, the second eigenvalue is $2\sqrt{d-1} + o(1)$ with high probability [Bor15], so the spectra captures all the other eigenvalues.

If we allow $d \to \infty$, the Kesten-McKay law tends to Wigner's semicircle law, and a result by [TVW13] confirms that it is indeed the limiting distribution for $\mathsf{Adj}(G)$ for a random regular graph with degree $d = d(n)$ where $d(n) \to \infty$ as $n \to \infty$.

Both cases translate to J having one large eigenvalue at β and a spectra of order $\asymp \beta/\sqrt{d}$.

d-dimensional lattice with periodic boundary conditions. The **Cartesian product** of graphs G and H, denoted $G\square H$, is the graph on $V(G)\times V(H)$ formed by connecting $(u_1,v)\sim(u_2,v)$ iff $(u_1u_2)\in E(G)$ and $(u,v_1)\sim(u,v_2)$ iff $(v_1v_2)\in E(H)$. In particular, the (periodic) d-dimensional lattice with m^d vertices is the d-fold Cartesian product of the cycle graph with m vertices C_m.

The spectral properties follows from the fact that the eigenvalues of $G\square H$ are precisely the pairwise sums of an eigenvalue of G with an eigenvalue of H [BH11]. Furthermore, the cycle graph C_m has eigenvalues exactly $2\cos\frac{2\pi k}{m}, k=0,1,...,m-1$, so the set of eigenvalues of J are precisely

$$\left\{\frac{\beta}{d}\sum_{i=1}^{d}\cos\frac{2\pi k_i}{m} : (k_i)_i\in[m-1]^d\right\}$$

For $d=1$, a simple computation shows that the limiting spectral distribution matches the Kesten-McKay distribution with degree 2 (i.e. $\rho(x)\propto\frac{1}{\sqrt{4-x^2}}$), so the bulk of eigenvalues of J have order β. It thus follows that for fixed d, we expect the spectral distribution to be a d-fold convolution of the $d=1$ distribution, so ignoring the dependence on d we again have that the bulk of eigenvalues of J has order β.

If $d\to\infty$, this set approximates the distribution of $\frac{\beta}{\sqrt{d}}\mathcal{N}(0,1/2)$ for some constant c depending only on m, so a constant fraction of the eigenvalues have order at most $\frac{\beta}{\sqrt{d}}$ and decaying double-exponentially thereafter.

4.3 Summary of Results

For the following results, we assume that we only keep the leading order term as $n \to \infty$ and keep the dependence on all other variables except β. We justify these results in subsec

Result	CW	SK	RR	ER (d fixed)	ER ($d \to \infty$)	d-dimensional grid (fixed d)	d-dimensional grid ($d \to \infty$)
[CD16]	$(\beta+\beta^{3/2})n^{1/2}$	$n^{5/4}\beta^3$	$n\log\left(1+\frac{\beta}{\sqrt{d}}\right)+\log(1+n\beta)$	$n\log(1+\beta)$	$n\log\left(1+\frac{\beta}{\sqrt{d}}\right)+\log(1+n\beta)$	$n\log(1+\beta)$	$n\beta\sqrt{\frac{\log d}{d}}$
[Yan20]	$(\beta+\beta^{3/2})n^{1/2}$	$n^{5/4}\beta^3$	$n\log\left(1+\frac{\beta}{\sqrt{d}}\right)+\log(1+n\beta)$	$n\log(1+\beta)$	$n\log\left(1+\frac{\beta}{\sqrt{d}}\right)+\log(1+n\beta)$	$n\log(1+\beta)$	$n\beta\sqrt{\frac{\log d}{d}}$
[BM17]	$o(n)$	-	$o(n)$ if $d \to \infty$	-	$o(n)$	-	$o(n)$
[Aug20]	$\log n$	$n\log(1+\beta)$	$n\log\left(1+\frac{\beta}{\sqrt{d}}\right)+\log(1+n\beta)$	$n\log(1+\beta)$	$n\log\left(1+\frac{\beta}{\sqrt{d}}\right)+\log(1+n\beta)$	$n\log(1+\beta)$	$n\beta\sqrt{\frac{\log d}{d}}$
[Eld20] C1	$\log n$	$n\log(1+\beta)$	$n\log\left(1+\frac{\beta}{\sqrt{d}}\right)+\log(1+n\beta)$	$n\log(1+\beta)$	$n\log\left(1+\frac{\beta}{\sqrt{d}}\right)+\log(1+n\beta)$	$n\log(1+\beta)$	$n\beta\sqrt{\frac{\log d}{d}}$
[Aus19]	$\log n$	$n\log(1+\beta)$	$n\log\left(1+\frac{\beta}{\sqrt{d}}\right)+\log(1+n\beta)$	$n\log(1+\beta)$	$n\log\left(1+\frac{\beta}{\sqrt{d}}\right)+\log(1+n\beta)$	$n\log(1+\beta)$	$n\beta\sqrt{\frac{\log d}{d}}$
[JKR19] / [Aug20] Corr.	$n^{2/3}\beta^{2/3}$	$n\beta^{2/3}$	$\frac{n\beta 2/3}{d^{1/3}}$	$n\beta 2/3$	$\frac{n\beta^{2/3}}{d^{1/3}}$	$n\beta 2/3$	$\frac{n\beta^{2/3}}{d^{1/3}}$
[Eld18]	$n^{5/6}\beta$	$n^{4/3}\beta$	$\frac{n\beta}{d^{1/6}}$	$n\beta$	$\frac{n\beta}{d^{1/6}}$	$n\beta$	$\frac{n\beta}{d^{1/6}}$
[Aug21]	$n^{1/2}\beta$	$n\beta$	$\frac{n\beta}{\sqrt{d}}$	$n\beta$	$\frac{n\beta}{\sqrt{d}}$	$n\beta$	$\frac{n\beta}{d^{1/2}}$

4.4 Discussion

Curie-Weiss model. All the bounds are sublinear in n, but a direct computation following [MM09] yields that the optimal gap is order $\log n$. Possibly, the low rank property is not encoded by the expressions which only depend on the size of the entries in J.

In the bounds of [CD16], the complexity term is $O(\log n)$, so we see in this case the inefficiency comes from the smoothness term.

Sherrington-Kirkpatrick Model. A priori, we expect that the bounds which only depend on $\{|J_{ij}|\}$ (e.g. only on $\|J_i\|_\bullet$ or $\|J\|_F$) will be insensitive to cancellation between spins and thus give a much worse result than spectrum-based bounds.

Surprisingly, this is not true, and the best bound comes from [Aug21] (alongside [Eld20], [Aus19], [Aug20], which are spectrum-based), which is $O(\beta n)$. The optimal gap below the critical temperature is $O(\beta^2 n)$, so while these bounds give the correct order in n, it does not give the correct order in β for the high-temperature limit $\beta \to 0$.

For [CD16], the main contributor is still the smoothness term.

Random regular graphs. This case is particularly easy to compute because all eigenvalues except for the top lie in/near the support of the continuous limit spectrum.

For fixed $\sqrt{d}$, all bounds give the answer $\Theta(n\beta)$. For $d \to \infty$, we get an extra $\frac{1}{\sqrt{d}}$ factor for the best possible dependence on d.

For [CD16], the main contributor is now the complexity term.

Erdős-Renyi graphs. The bounds match those for the random regular graphs with corresponding d.

d-dimensional grid. For fixed d, the bounds match those of the random regular graphs (which isn't surprising since this is a $2d$-regular graph). However, for $d \to \infty$, we get an extra $\sqrt{\log d}$ factor on the spectra because the limiting spectrum is a normal distribution (instead of having bounded support, like the semicircular law).

In the last three cases, we get that in the limit $d \to \infty$, the gap is bounded above by $o(n)$ (i.e. the mean-field approximation holds), which is agreement with physical predictions.

5 Computation

5.1 Ising Model Computations

Here we justify the computations in Subsection 4.1. We note that the results from [BM17; JKR19; Eld20] are already in the desired form and do not need further simplification. The corollary of [Aug20] and the result in [Aug21] follow immediately by applying Subsubsection 5.1.1.

5.1.1 Mean-width

We would like to simplify the Gaussian mean-width and Rademacher mean-width. We start by noticing that

$$\nabla f(\sigma) = J\sigma \tag{38}$$

so the Gaussian mean-width is simply

$$\begin{aligned}
\mathsf{GW}(\nabla f\{\pm 1\}^n) &= \mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,1)} \sup_{\sigma} \langle x, J\sigma\rangle && (39)\\
&= \mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,1)} \sup_{\sigma} \langle Jx, \sigma\rangle && (40)\\
&= \mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,1)} \|Jx\|_1 && (41)\\
&= \sum_{i=1}^{n} \mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,1)} |\langle J_i, x\rangle| && (42)\\
&= \sqrt{\frac{2}{\pi}} \cdot \sum_{i=1}^{n} \|J_i\|_2 && (43)
\end{aligned}$$

where J_i is the i-th row of J. The last line is true because $\langle J_i, x\rangle$ has distribution $N(0, \|J_i\|)_2^2)$, and we use the fact that $\mathbb{E}_{x\sim\mathcal{N}(0,1)} |x| = \sqrt{2/\pi}$.

We expect the same to hold for Rademacher width:

Lemma 10. *For any $v \in \mathbb{R}^n$,*

$$\mathop{\mathbb{E}}_{x\in\mathsf{Unif}\{\pm 1\}^n} |\langle x, v\rangle| \asymp \|v\|_2$$

Proof. The upper bound is a straightforward application of Cauchy-Schwarz:

$$\begin{aligned}
\mathop{\mathbb{E}}_{x\sim\mathsf{Unif}\{\pm 1\}^n} |\langle x, v\rangle| &\le \left(\mathop{\mathbb{E}}_{x\sim\mathsf{Unif}\{\pm 1\}^n} \langle x, v\rangle^2\right)^{1/2}\\
&= \left(\sum_{i,j=1}^{n} v_i v_j \mathop{\mathbb{E}}_{x\sim\mathsf{Unif}\{\pm 1\}^n} x_i x_j\right)^{1/2}\\
&= \left(\sum_{i=1}^{n} v_i^2\right)^{1/2} = \|v\|_2^2
\end{aligned}$$

where we used the fact that

$$\mathop{\mathbb{E}}_{x\sim\mathsf{Unif}\{\pm 1\}^n} x_i x_j = \begin{cases} 1 & \text{if } i = j\\ 0 & \text{otherwise}\end{cases}$$

The lower bound proceeds using Hölder's inequality: writing $Z = \langle x, v\rangle$,

$$\mathbb{E}[|Z|]^{2/3}\,\mathbb{E}[Z^4]^{1/3} \ge \mathbb{E}[Z^2] \tag{44}$$

therefore $\mathbb{E}[|Z|] \geq \frac{\mathbb{E}[Z^2]^{3/2}}{\mathbb{E}[Z^4]^{1/2}}$, so it suffices to show that $\mathbb{E}[Z^4] \leq c\|v\|_2^4$ for some constant $c > 0$. Now we expand:

$$\begin{aligned}
\mathbb{E}[Z^4] &= \sum_{i,j,k,\ell=1}^{n} v_i v_j v_k v_\ell \mathop{\mathbb{E}}_{x\sim\mathsf{Unif}\{\pm 1\}^n} x_i x_j x_k x_\ell \\
&= \sum_{i=1}^{n} v_i^4 + 3\sum_{i\neq j} v_i^2 v_j^2 \\
&\leq 3\left(\sum_{i=1}^{n} v_i^2\right)^2 = 3\|v\|_2^4
\end{aligned}$$

and in particular, $\mathbb{E}[|Z|] \geq \frac{1}{\sqrt{3}}\|v\|_2$. □

The estimate of the Rademacher mean-width follows immediately from

$$\begin{aligned}
\mathsf{RW}(\nabla f\{\pm 1\}^n) &= \mathop{\mathbb{E}}_{x\sim\mathsf{Unif}\{\pm 1\}^n} \sup_{\sigma} \langle x, J\sigma\rangle && (45)\\
&= \mathop{\mathbb{E}}_{x\sim\mathsf{Unif}\{\pm 1\}^n} \sup_{\sigma} \langle Jx, \sigma\rangle && (46)\\
&= \mathop{\mathbb{E}}_{x\sim\mathsf{Unif}\{\pm 1\}^n} \sup_{\sigma} \|Jx\|_1 && (47)\\
&= \sum_{i=1}^{n} \mathop{\mathbb{E}}_{x\sim\mathsf{Unif}\{\pm 1\}^n} |\langle J_i, x\rangle| && (48)\\
&\asymp \sum_{i=1}^{n} \|J_i\|_2 && (49)
\end{aligned}$$

5.1.2 Remaining results

[CD16] Firstly, we have

$$b_i = \sup_{\sigma\in[-1,1]^n} \langle J_i, \sigma\rangle = \|J_i\|_1, c_{ij} = |J_{ij}|$$

and for the sake of tractability we will adopt the relaxation

$$a = \sup_{\sigma\in[-1,1]^n} \left|\frac{1}{2}\sigma^\top J\sigma\right| \leq \frac{1}{2}\sup_{\|\sigma\|_2\leq n} \left|\sigma^\top J\sigma\right| = \sqrt{n}\|J\|_{\text{op}}$$

Plugging these in, we get

$$\begin{aligned}
(\textit{complexity}) &= \inf_{\varepsilon>0}\left\{\left(n\sum_{i=1}^{n}\|J_i\|_1^2\right)^{1/2}\varepsilon + \frac{3}{2}n\varepsilon + \log N(J\{\pm 1\}, \sqrt{n}\varepsilon, \ell_2)\right\} && (50)\\
&\asymp \inf_{\varepsilon>0}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\|J_i\|_1^2\right)^{1/2}\varepsilon + \frac{3}{2}\varepsilon + \log N(J\{\pm 1\}, \varepsilon/\sqrt{n}, \ell_2)\right\} && (51)\\
(\textit{smoothness})^2 &\asymp \sum_{i=1}^{n}\|J_i\|_1^2 + \sqrt{n}\|J\|_{\text{op}}\left(\sum_{i=1}^{n}|J_{ii}| + \|J\|_F\right) + \sum_{i,j=1}^{n}\|J_i\|_1\|J_j\|_1|J_{ij}| && (52)\\
&\quad + \left(\sum_{i=1}^{n}\|J_i\|_1^2\right)\left(\sum_{i=1}^{n}J_{ii}^2\right) + \left(\sum_{i=1}^{n}J_{ii}^2\right)^2 + 1 && (53)
\end{aligned}$$

[Yan20] We pick $V_i = \mathbb{R}$ and $K_i = [-1, 1]$, so $M = 2$ and can be ignored for order computations. The Fréchet derivatives reduce to the usual partial derivatives, so the quantities a, b_i, c_{ij} match up with those in [CD16]. $\mathcal{D}(\varepsilon)$ becomes a $\sqrt{n}\varepsilon$-covering of $\nabla f(\{\pm 1\})$ under the ℓ_2-distance, so replacing $\varepsilon \to \varepsilon/n$,

$$B_1^2 \asymp \sqrt{n}\|J\|_{\text{op}}\left(\sum_i |J_{ii}| + \|J\|_F^2\right) + \sum_i \|J_i\|_1^2 + \sum_{i,j} \|J_i\|_1\|J_j\|_1|J_{ij}| \tag{54}$$

$$B_2 \asymp \frac{\varepsilon^2}{n} + \varepsilon\left(1 + \frac{1}{n}\sum_i \left(J_{ii}^2 + \|J_i\|_1^2\right)\right)^{1/2} + \left(\sum_{i=1}^n \|J_i\|_1^2\right)^{1/2}\left(\sum_{i=1}^n J_{ii}^2\right)^{1/2} \tag{55}$$

so the overall bound is

$$\mathsf{GAP}^* \leq \inf_{\varepsilon>0}\left\{\frac{\varepsilon^2}{n} + \varepsilon\left(1 + \frac{1}{n}\sum_i \left(J_{ii}^2 + \|J_i\|_1^2\right)\right)^{1/2} + \log N(J[\pm 1]^n, \varepsilon/\sqrt{n}, \ell_2)\right\} \tag{56}$$

$$+ \left(\sqrt{n}\|J\|_{\text{op}}\left(\sum_i |J_{ii}| + \|J\|_F^2\right) + \sum_i \|J_i\|_1^2\right. \tag{57}$$

$$\left. + \sum_{i,j} \|J_i\|_1\|J_j\|_1|J_{ij}| + \left(\sum_{i=1}^n \|J_i\|_1^2\right)\left(\sum_{i=1}^n J_{ii}^2\right) + 1\right)^{1/2} \tag{58}$$

[Aus19] First, we compute the Lipschitz constant L for the Hamming metric $d_n = \frac{1}{n}\sum_{i=1}^n d_{K_i}$, i.e.

$$L := \frac{1}{2}\max_{\sigma\neq\sigma'} \frac{|\sigma^\top J\sigma - (\sigma')^\top J\sigma'|}{d_n(\sigma,\sigma')} \tag{59}$$

$$= \frac{1}{2}\max_{\sigma,\sigma' \text{ adj.}} |\sigma^\top J\sigma - (\sigma')^\top J\sigma'| \tag{60}$$

$$= \max_{i,\sigma} |\sum_{j\neq i} J_{ij}\sigma_i| = \max_i \|J_i^\circ\|_1 \tag{61}$$

Next, we would like to interpret the $\|\cdot\|$ norm on the set $\{\nabla f(x,\cdot)\}$. Note that

$$\|\nabla f(x,\cdot) - \nabla f(y,\cdot)\| = \sup_z \left|\sum_i (\partial_i f(x,z) - \partial_i f(y,z)\right| \tag{62}$$

$$= \sup_z \left|\sum_i (z_i - *_i)\sum_{j\neq i} J_{ij}(x_j - y_j)\right| \tag{63}$$

$$\asymp \|J^\circ(x-y)\|_1 \tag{64}$$

For the last line, we would like to show that in general,

$$\|t\|_1 \leq \sup_z |\langle z - *, t\rangle| \leq 2\|t\|_1 \tag{65}$$

Let $z^{(1)}$ and $z^{(2)}$ maximize and minimize $\langle z - *, t\rangle$ respectively. Then, by checking cases we have

$$\left|\left\langle z^{(1)} - *, t\right\rangle\right| + \left|\left\langle z^{(2)} - *, t\right\rangle\right| = 2\|t\|_1$$

so Equation 65 holds since the supremum is attained on $z^{(1)}$ or $z^{(2)}$.

Recall that we would like to cover the set $\{J^\circ x : x \in \{\pm 1\}^n\}$ with $e^{\varepsilon n}$ sets with $\|\cdot\|$-diameter δn, and we note that it suffices to minimize the quantity $C = \varepsilon + \delta$.

Let $N^d(S, r, \ell)$ denote the minimum number of sets of diameter r needed to cover set S under metric ℓ. Then $C = \inf_{\delta>0}\{\delta + \log N^d(J\{\pm 1\}^n, \delta n, \|\cdot\|)\}$. We make two observations:

- **The $\|\cdot\|$-norm can be replaced by the ℓ_1-norm.**

  By the above argument, we have

$$N^d(J\{\pm 1\}^n, 2\delta n, \ell_1) \leq N^d(J\{\pm 1\}^n, \delta n, \|\cdot\|) \leq N^d(J\{\pm 1\}^n, \delta n, \ell_1)$$

  This means that if we only need to estimate C up to a constant factor, we can freely replace $\|\cdot\|$ with ℓ_1.

- **Diameter can be replaced with radius.**

  Again, note that

$$N(J\{\pm 1\}^n, \delta n, \ell_1) \leq N^d(J\{\pm 1\}^n, \delta n, \ell_1) \leq N(J\{\pm 1\}^n, \delta n/2, \ell_1)$$

  so similarly as above we can replace N^d with N.

We thus conclude that

$$C \asymp \inf_{\delta > 0} \{\delta n + \log N(J\{\pm 1\}^n, \delta n, \ell_1)\} \tag{66}$$

since a constant factor in the covering radius can be pulled out in the infimum term.

[Eld20] Recall that the result we require is that

$$\mathsf{GAP} \leq 64\, \mathsf{Lip}(f)^{2/3}\, \mathsf{GW}_+(\nabla f\{\pm 1\}^n) n^{2/3} \tag{67}$$

First we compute ∇f (which in this case refers to the discrete derivative of f):

$$\begin{aligned}
\partial_i f(y) &= \frac{1}{2}(f(y_1, ..., y_{i-1}, 1, y_{i+1}, ..., y_n) - f(y_1, ..., y_{i-1}, -1, y_{i+1}, ..., y_n)) && (68)\\
&= \sum_{j \neq i} J_{ij} y_j = \langle J_i^\circ, y\rangle && (69)
\end{aligned}$$

This implies that $\nabla f(y) = J^\circ y$, hence $\nabla f\{\pm 1\}^n$ is symmetric about the origin and so we can swap out GW_+ for GW.

Next, we compute $\mathsf{Lip}(f)$. Note that

$$\begin{aligned}
\mathsf{Lip}(f) &:= \max_{i,y} |\partial_i f(y)| && (70)\\
&= \max_i \|J_i^\circ\|_1 && (71)
\end{aligned}$$

Finally, from Subsubsection 5.1.1 we have

$$\mathsf{GW}(\nabla f\{\pm 1\}^n) \asymp \sum_{i=1}^n \|J_i\|_2$$

so putting it all together gives

$$\mathsf{GAP} \lesssim n^{2/3} (\max_i \|J_i^\circ\|_1)^{2/3} \left(\sum_{i=1}^n \|J_i\|_2\right)^{1/3} \tag{72}$$

5.2 Specific Model Computations

In this subsection, we discuss how to compute various tricky quantities: the bound of [Eld20] has a variational problem, while some other bounds depends on various covering numbers ([CD16], [Yan20], [Aug20], [Aus19]).

As a summary of the main results:

- The ℓ_1-covering number is essentially equivalent to the ℓ_2-covering number (for our purposes)
- When the spectra consists of one large eigenvalue and the rest forming a continuous spectrum in the limit, we can compute covering numbers and Eldan's quantity up to order
- More generally, when the limiting spectrum is continuous and bounded, one obtains that the size of these quantities are $n \cdot$ (order of bulk of spectrum).
- A separate ad-hoc computation is needed for the lattice case, since the limiting spectrum is normal and thus not bounded.

5.2.1 ℓ_2-covering numbers

For the sake of tractability, we will use

$$J[-1,1]^n \subset J(\sqrt{n}B_2)$$

We would like to cover $E = \{Jx : \|x\|_2 \le 1\}$ with balls of radius r. Note that E has semi-axes of length $\{|\lambda_i|\}$ respectively.

Standard volume bounds give that for $r < |\lambda_n|$ (so that $E \supset r \cdot B_n$),

$$\frac{\mathrm{Vol}(E)}{\mathrm{Vol}(rB_n)} \le N(E, r, \ell_2) \le \frac{\mathrm{Vol}(E + \frac{r}{2}B_n)}{\mathrm{Vol}(\frac{r}{2}B_n)} \underset{E \subset \frac{r}{2}B_n}{\le} \frac{\mathrm{Vol}(2E)}{\mathrm{Vol}(\frac{r}{2}B_n)} \tag{73}$$

which gives us an estimate of $\log N(E, r, \ell_2)$ up to an additive estimate of $n \log 4$. To deal with the condition on the RHS, we note that rB_n contains $\frac{r}{2}B_k \times \frac{r}{2}B_{n-k}$, so instead we "flatten" along all semi-axes shorter than length $\frac{r}{2}$ and cover the hyperellipsoid along the remaining axes. Denote E_r to be the ellipsoid with precisely only semi-axes of E which are at least r. Then we have the following bounds (denote $k(r) = \#\{i : |\lambda| > r\}$):

$$\begin{aligned}
\log N(E, r, \ell_2) &\ge \log N(E_r, r, \ell_2) && (74)\\
&\ge \log \frac{\mathrm{Vol}(E_r)}{\mathrm{Vol}(r \cdot B_{k(r)})} && (75)\\
&= \sum_{|\lambda| \ge r} \log \frac{|\lambda|}{r} = \sum_{\lambda} \left(\log \frac{|\lambda|}{r}\right)_+ && (76)\\
\log N(E, r, \ell_2) &\le \log N(E_{r/2}, \tfrac{r}{2}, \ell_2) && (77)\\
&\le \log \frac{\mathrm{Vol}(E_{r/2} + \frac{r}{4}B_{k(r/2)})}{\mathrm{Vol}(\frac{r}{4}B_{k(r/2)})} && (78)\\
&= \sum_{|\lambda| \ge r/2} \log\left(1 + \frac{4|\lambda|}{r}\right) && (79)
\end{aligned}$$

As a first step, we compute the key optimization up to order in the case of one large eigenvalue and the rest of the same order.

Lemma 11. *Suppose that J has some spectra around λ_2 (that is, there exists an absolute constants $\eta > 0$ where ηn eigenvalues are in the interval $(\lambda_2/2, \lambda_2]$. Then,*

$$\inf_{\delta > 0} (\delta n + \log_2 N(E, \delta, \ell_2)) \asymp n \log(1 + \lambda_2) + \log(1 + n\lambda_1) + 1$$

Proof. Write $S = \delta n + \log_2 N(E, \delta, \ell_2)$. We first establish the lower bound.

- If $\delta \geq \lambda_1$, $S \geq n\lambda_1$, and we have both $n\lambda_1 \geq n\log(1+\lambda_1) \geq n\log(1+\lambda_2)$ and $n\lambda_1 \geq \log(1+n\lambda_1)$, so
$$S \geq n\lambda_1 \gtrsim RHS$$
- If $\delta \in (\lambda_2, \lambda_1)$, firstly note that $\log N > 0$ so $\log N \geq \log 2$. Then,
$$S \geq n\delta + \log\left(\frac{\lambda_1}{\delta}\right)_+ \tag{80}$$
$$= n\delta + (\log(n\lambda_1) - \log(n\delta))_+ \tag{81}$$
$$\geq n\delta - \log(n\delta)_+ + \log(n\lambda_1)_+ \geq \log(n\lambda_1)_+ \tag{82}$$
For the other terms, we are required to analyze the function $f(x) = nx + \log\frac{1}{x}$, which is minimized at $x \asymp 1/n$:
  - If $\lambda_2 \gtrsim \frac{1}{n}$, then f is increasing on the interval $(c\lambda_2, \lambda_1)$ for some constant c, so
$$S \geq n\lambda_2 \geq n\log(1+\lambda_2)$$
  - If $\lambda_2 \lesssim 1/n$, then $n\log(1+\lambda_2) \lesssim 1$, so this case is already covered.

  In summary,
$$S \gtrsim 1 + \log(n\lambda_1)_+ + n\lambda_2 \gtrsim 1 + (\log n\lambda_1) + n\lambda_2 \gtrsim RHS.$$
- If $\delta < \lambda_2$, then $S \gtrsim n\delta + n\log(\lambda_2/\delta) = n(\log\lambda_2 + \delta + \log(1/\delta) \asymp n(\log\lambda_2 + 1)) \gtrsim n\log(1+\lambda_2)$. Similarly, $LHS \geq n\delta + \log(\lambda_1/\delta) \gtrsim \log(1+n\lambda_1)$.

For the upper bound, we consider two choices of δ:

- $\delta = 2\lambda_2 + \varepsilon$ as $\varepsilon \to 0$. This gives
$$S \leq 2\lambda_2 n + \log\left(1 + \frac{2\lambda_1}{\lambda_2}\right) \lesssim \lambda_2 n + \log\left(1 + \frac{\lambda_1}{\lambda_2}\right) + 1$$
- $\delta = 1$:
$$S \leq n + n\log(1+4\lambda_2) + \log(1+4\lambda_1) \lesssim n(1+\log(1+\lambda_2)) + \log(1+\lambda_1)$$

Roughly speaking, this suggests using $\delta = 1$ when $\lambda_2 \gtrsim 1$ and $\delta = 2\lambda_2$ when $\lambda_2 \lesssim 1$. This is sufficient provided that $\lambda_2 \gtrsim \frac{1}{n}$. However, if $\lambda_2 \ll \frac{1}{n}$, then instead we substitute $\delta = \frac{c}{n} > 2\lambda_2$ to get $S \lesssim 1 + \log(1+n\lambda_1)$. □

5.2.2 ℓ_1-covering numbers

In this section, we show that the ℓ_1 covering number is on the same order as the ℓ_2 covering numbers when the bound in Subsubsection 5.2.1 is effective up to order.

An easy start is that since $\|x\|_1 \leq \sqrt{n}\|x\|_2$, the ℓ_1-ball of radius r contains the ℓ_2-ball of radius $\frac{r}{\sqrt{n}}$, so
$$\log N(S, r, \ell_1) \leq \log N(S, r/\sqrt{n}, \ell_2). \tag{83}$$
On the other hand, we can use the same axis cutoff strategy to bound the other side. First, we

recall the formulas for ℓ_1 and ℓ_2 ball volumes and apply Stirling's approximation:

$$\begin{aligned}\mathrm{Vol}(B_2) &= \frac{\pi^{n/2}}{\Gamma(n/2+1)} && (84)\\ &\sim \frac{\pi^{n/2}}{\sqrt{\pi n}(\frac{n}{2e})^{n/2}} && (85)\\ &= \frac{1}{\sqrt{\pi n}}\left(\frac{2\pi e}{n}\right)^{n/2} && (86)\\ \mathrm{Vol}(\sqrt{\tfrac{2\pi n}{e}}\cdot B_1) &= \left(\frac{2\pi n}{e}\right)^{n/2}\cdot\frac{1}{n!} && (87)\\ &\sim \frac{(2\pi n/e)^{n/2}}{\sqrt{2\pi n}(\frac{n}{e})^n} && (88)\\ &= \frac{1}{\sqrt{2\pi n}}\left(\frac{2\pi e}{n}\right)^{n/2} && (89)\end{aligned}$$

so we can conclude that asymtotically,

$$\mathrm{Vol}(B_2) \sim \frac{1}{\sqrt{2}}\mathrm{Vol}(\sqrt{\tfrac{2\pi n}{e}}\cdot B_1).$$

Thus, if we apply the semiaxes-cutoff argument from before at level s, we get

$$\begin{aligned}\log N(E,r,\ell_1) &\geq \log\frac{\mathrm{Vol}\,E_s}{\mathrm{Vol}(r\cdot B_{\ell_1;k(s)})} && (90)\\ &= \log\frac{\mathrm{Vol}\,E_s}{\mathrm{Vol}(\sqrt{2\pi n/e}\cdot r\cdot B_{\ell_2;k(s)})} + O(1) && (91)\end{aligned}$$

Setting $s = \sqrt{2\pi n/e}\cdot r$, we realize that this is precisely the lower bound for $\log N(E,s,\ell_2)$. Thus, if we had the bounds

$$\log N^-(E,r,\ell_2) \leq \log N(E,r,\ell_2) \leq \log N_+(E,r,\ell_2)$$

we must also have

$$\log N^-(E,cr/\sqrt{n},\ell_2) \leq \log N(E,r,\ell_1) \leq \log N^-(E,r/\sqrt{n},\ell_2)$$

for $c = \sqrt{\frac{e}{2\pi}}$. This essentially suggests that any time the ℓ_2-covering number can be bounded up to order, the same goes for the ℓ_1-covering number.

5.2.3 Variational problem in Eldan's bound

In the case where J has one large eigenvalue λ_1 and some spectra of order λ_2 (the second largest eigenvalue), we can compute the variational quantity in Eldan's upper bound up to order:

Lemma 12. *If there are universal constants $\delta,\eta > 0$ for which there exists at least δn eigenvalues of size $\eta\lambda$, then*

$$\max\left\{\sum_i \log(c_i|\lambda_i(J)|+1); c_i \geq 0, \sum c_i \leq 4n\right\} \asymp n\log(1+|\lambda_2|) + \log(1+n|\lambda_1|). \quad (92)$$

Proof. For convenience, we suppress the J and just write λ_i for the i-th eigenvalue by absolute value. The valid allocation $c_1 = n$, $c_2 = ... = c_n = 1$ gives the lower bound. As for the upper bound, we note that

$$\begin{aligned}\sum_i \log(c_i|\lambda_i|+1) &\leq \log(c_1|\lambda_1|+1) + \sum_{i=1}^n \log(c_i|\lambda_2|+1) && (93)\\ &\leq \log(n|\lambda_1|+1) + n\log(4|\lambda_2|+1) && (94)\end{aligned}$$

where we used Jensen's inequality in the last line. □

For more general cases, we can establish the following lemma:

Lemma 13. *The variational quantity in [Eld20] is equivalent to*

$$\sup_{x>0}\left\{\sum_{|\lambda|\geq x}\log|\lambda_i|+\ell\log m;\quad m:=\frac{1}{\ell}\left(\sum_{|\lambda|\geq x}\frac{1}{|\lambda_i|}+4n\right)\geq\frac{1}{x}\right\} \tag{95}$$

Proof. Note that so long as

$$\sum_i\log(c_i|\lambda_i|+1)=\sum_i\log|\lambda_i|+\sum_i\log\left(c_i+\frac{1}{|\lambda_i|}\right) \tag{96}$$

where the sum is over all i where $|\lambda_i|\neq 0$. Notice that the maximum must be attained at $\sum_i c_i=4n$. Since the sum is fixed (and log is concave), at maximum every two terms are equal or blocked by a constraint. That is, for distinct i,j, if $c_i+\frac{1}{|\lambda_i|}>c_j+\frac{1}{|\lambda_j|}$ then $c_i=0$.

Hence, we can find some threshold ℓ such that

$$c_i+\frac{1}{|\lambda_i|}=\begin{cases}m:=\frac{1}{\ell}\left(\sum_{i\leq\ell}\frac{1}{|\lambda_i|}+4n\right) & i\leq\ell\\ \frac{1}{|\lambda_i|} & i>\ell\end{cases} \tag{97}$$

However, for this to be valid, we require $m\geq\frac{1}{|\lambda_\ell|}$. Finally, we swap out the index threshold ℓ for a value threshold, which only makes the condition tighter but keeps the equality cases at $x=|\lambda_\ell|$. This completes the proof. □

This specific characterization will be useful later in Subsubsection 5.2.5.

5.2.4 Limiting spectral distributions

In the cases where a continuous limiting spectral distribution is known, the quantities we are concerned about often reduce to computation about the limiting spectrum. Here, we establish some lemmas to make this process easier.

We first define some notation. Let Λ_n denote the empirical spectral distribution of J (suppressing the indexing by n), and suppose $\Lambda_n\overset{d}{\Rightarrow}\Lambda$ almost surely.

Lemma 14. *If* $\sup_n\mathbb{E}|\Lambda_n|^2<\infty$ *almost surely, then for any function* f *with* $f(x)=o(x^2)$ *as* $x\to\infty$*,*

$$\mathbb{E}[f(|\Lambda_n|)\mathbb{I}_{|\Lambda_n|\geq x})]\to\mathbb{E}[f(|\Lambda|)\mathbb{I}_{|\Lambda_n|\geq x}] \tag{98}$$

as $n\to\infty$ *uniformly over* x*.*

Proof. It suffices to check that the sequence of random variables $X_n=f(|\Lambda_n|)\mathbb{I}_{|\Lambda_n|\geq x}$ is uniformly integrable, i.e. $\lim_{x\to\infty}\sup_n\mathbb{E}[|X_n|\mathbb{I}_{|X_n|\geq x}]=0$.

$$\sup_n\mathbb{E}[f(|\Lambda_n|)\mathbb{I}_{|\Lambda_n|\geq x}]\leq\sup_n\mathbb{E}\left[\frac{f(|\Lambda_n|)}{|\Lambda_n|^2}\cdot|\Lambda_n|^2\mathbb{I}_{|\Lambda_n|\geq x}\right] \tag{99}$$

$$\leq\sup_{y\geq x}\frac{f(y)}{y^2}\cdot\sup_n\mathbb{E}|\Lambda_n|^2\to 0\qquad\text{as }x\to\infty \tag{100}$$

□

5.2.5 Limit version of Eldan's quantity

Denote

$$\ell_n(x) := \#\{\lambda : |\lambda| \geq x\} \tag{101}$$

$$\mathcal{G}_n(x) := \frac{1}{\ell_n(x)} \left(\sum_{|\lambda| \geq x} \frac{1}{|\lambda_i|} + 4n \right), \tag{102}$$

$$\mathcal{F}_n(x) := \sum_{|\lambda| \geq x} \log |\lambda_i| + \ell_n(x) \log \mathcal{G}_n(x). \tag{103}$$

These quantities can be rewritten using Λ_n:

$$\frac{1}{n}\ell_n(x) = \mathbb{P}(|\Lambda_n| \geq x), \tag{104}$$

$$\mathcal{G}_n(x) = \frac{\mathbb{E}\,\frac{1}{|\Lambda_n|} + 4}{\mathbb{P}(|\Lambda_n| \geq x)}, \tag{105}$$

$$\frac{1}{n}\mathcal{F}_n(x) = \mathbb{E} \log |\Lambda_n| \mathbb{I}_{|\Lambda_n| \geq x} + \mathbb{P}(|\Lambda_n| \geq x) \log \mathcal{G}_n(x). \tag{106}$$

In the variational problem, we have the condition that $\mathcal{G}_n(x) \geq \frac{1}{x}$. With a little manipulation, this can be more succinctly rewritten as

$$\mathcal{H}_n(x) := \mathbb{E}\left[\left(\frac{1}{x} - \frac{1}{|\Lambda_n|} \right)_+ \right] \leq 4. \tag{107}$$

Note that $\mathcal{H}_n$ is decreasing in x, and furthermore $H_n(x) \to 0$ as $x \to \infty$ and $H_n(x) \to \infty$ as $x \to 0$. Thus, we expect that (1) $H_n(x) \leq 4$ for all $x > x_*$, where $x_* = \inf\{x : H_n(x) \leq 4\}$, and $\mathcal{H}_n$ is continuous at x if and only if $\mathbb{P}(|\Lambda_n| = x) = 0$.

We now introduce the limit version of the above quantities:

$$\mathcal{F}_\infty(x) := \mathbb{E} \log |\Lambda| \mathbb{I}_{|\Lambda| \geq x} + \mathbb{P}(|\Lambda| \geq x) \tag{108}$$

$$\mathcal{H}_\infty(x) := \mathbb{E}\left[\left(\frac{1}{x} - \frac{1}{|\Lambda|} \right)_+ \right] \tag{109}$$

We hope that

$$\lim_{n\to\infty} \frac{1}{n} \sup_x \{\mathcal{F}_n(x); \mathcal{H}_n(x) \geq 4\} \stackrel{?}{=} \lim_{n\to\infty} \sup_x \{F_\infty(x); H_\infty(x) \leq 4\}. \tag{110}$$

If this is true, and the value on the RHS is nonzero, then we've just successfully computed Eldan's quantity.

5.2.6 Limit version of covering number bounds

Using this new notation, we have

$$\mathbb{E}\left[\left(\log \frac{|\Lambda_n|}{\delta}) \right)_+ \right] \leq \frac{1}{n} \log N(E, \delta, \ell_2) \leq \mathbb{E}\left[\log\left(1 + \frac{4|\Lambda_n|}{\delta} \right) \mathbb{I}_{|\Lambda_n| \geq \delta/2} \right]. \tag{111}$$

Roughly, when Λ_n has a limiting distribution (with some suitable normalization), we expect these values to be constants when δ is of the same order as $|\Lambda_n|$, which means that $\inf_{\delta>0}\{\delta n + \log N(E, \delta, \ell_2)\} \asymp n \cdot (\text{order of } \Lambda_n)$.

5.2.7 Computation for lattice model

Write $\lambda_i = \lambda_i(J)$ for convenience. We will compute the free energy bound quantity up to order for the lattice model.

Recall that the set of eigenvalues of J are

$$\left\{\frac{\beta}{d}\sum_{i=1}^{d}\cos\frac{2\pi k_i}{m} : (k_i)_i \in [m-1]^d\right\}$$

We recall the limiting distribution of $(\beta/\sqrt{d})^{-1}J$: for $d = 1$, the limiting distribution has density $\frac{1}{\pi(1-x^2)}$. Let Λ_1 be a random variable drawn from this distribution. For general d, the limiting distribution is the law of the sum of d i.i.d. copies of Λ_1 normalized by $\frac{1}{\sqrt{d}}$ (which we call Λ_d). Furthermore,

$$\Lambda_d \overset{d}{\Rightarrow} \Lambda \sim \mathcal{N}(0, 1/2) \qquad \text{as } d \to \infty.$$

Warning. This is different from the previous definition of Λ_n in Subsubsection 5.2.4 in the sense that we add additional normalization for the convergence to hold.

We refer back to lemma 13 for the exact free energy and the limiting analogues:

$$\mathcal{F}_{n,d}(x) := \sum_{|\lambda|\geq x}\log|\lambda_i| + \ell\log\left(\frac{1}{\ell}\left(\sum_{|\lambda|\geq x}\frac{1}{|\lambda_i|} + 4n\right)\right) \tag{112}$$

$$\mathcal{F}_d(x) := \lim_{n\to\infty}\frac{\mathcal{F}_{n,d}((\beta/\sqrt{d})x)}{n} \tag{113}$$

$$= \mathbb{E}[\log|\Lambda_d| \cdot \mathbb{I}_{|\Lambda_d|\geq x}] + \mathbb{P}(|\Lambda_d| \geq x)\log\left(\frac{\mathbb{E}\,\frac{1}{|\Lambda_d|}\mathbb{I}_{|\Lambda_d|\geq x} + \frac{4\beta}{\sqrt{d}}}{\mathbb{P}(|\Lambda_d| \geq x)}\right) \tag{114}$$

$$\mathcal{F}_\infty(x) = \mathbb{E}[\log|\Lambda| \cdot \mathbb{I}_{|\Lambda|\geq x}] + \mathbb{P}(|\Lambda| \geq x)\log\left(\frac{\mathbb{E}\,\frac{1}{|\Lambda|}\mathbb{I}_{|\Lambda|\geq x}}{\mathbb{P}(|\Lambda| \geq x)}\right) \tag{115}$$

We also have the following restriction in the limit (in the variational quantity):

$$\frac{1}{\ell}\left(\sum_{|\ell|\geq x}\frac{1}{|\lambda_i|} + 4n\right) \geq \frac{1}{x} \tag{116}$$

$$\Leftrightarrow 4 \geq \frac{1}{n}\left(\sum_{|\ell|\geq x}\left(\frac{1}{x} - \frac{1}{|\lambda_i|}\right)\right) \tag{117}$$

$$\frac{4\beta}{\sqrt{d}} \geq \mathbb{E}\left[\left(\frac{1}{x} - \frac{1}{|\Lambda_d|}\right)_+\right] \tag{118}$$

It is easy to check that RHS is decreasing in x, $RHS \to \infty$ as $x \to 0$ and $RHS \to 0$ as $x \to \infty$. This implies the existence of a unique x_* such that equality holds, and the inequality is true iff $x \geq x_*$.

If we now track the dependence of x_* on d, since $LHS \to 0$ as $d \to \infty$, we get that $x_* \to \infty$ as $d \to \infty$.

Now we try to establish x_* up to order. We need the following lemma:

Lemma 15. *For fixed integer $n \geq 0$, as $x \to \infty$ we have*

$$\mathbb{E}[(|\Lambda| - x)^n \mathbb{I}_{|\Lambda|\geq x}] \asymp x^{-(n+1)}e^{-x^2/4} \tag{119}$$

where the constants may depend on n.

Proof. Lower bound:

$$\mathbb{E}[(|\Lambda| - x)^n \mathbb{I}_{|\Lambda|\geq x}] = \frac{1}{\sqrt{\pi}} \int_x^\infty (t-x)^n e^{-t^2/4}\,dt \tag{120}$$

$$\geq \frac{1}{\sqrt{\pi}} \int_{x+1/x}^{x+2/x} (t-x)^n e^{-t^2/4}\,dt \tag{121}$$

$$\gtrsim \frac{1}{x^{n+1}} e^{-(x+2/x)^2/4} \gtrsim \frac{1}{x^{n+1}} e^{-x^2/4} \tag{122}$$

Upper bound:

$$\mathbb{E}[(|\Lambda| - x)^n \mathbb{I}_{|\Lambda|\geq x}] = \frac{1}{\sqrt{\pi}} \int_x^\infty (t-x)^n e^{-t^2/4}\,dt \tag{123}$$

$$\leq e^{-x^2/4} \frac{1}{\sqrt{\pi}} \int_x^\infty (t-x)^n e^{-x(t-x)/2}\,dt \tag{124}$$

$$= e^{-x^2/4} \frac{1}{\sqrt{\pi}} \frac{\Gamma(n+1)}{(x/2)^{n+1}} \tag{125}$$

□

Lemma 16. *For f twice differentiable with $|f''|$ decreasing eventually,*

$$\mathbb{E}[f(|\Lambda|)\mathbb{I}_{|\Lambda|\geq x}] = f(x)\mathbb{P}(|\lambda| \geq x) + f'(x)\,\mathbb{E}[(|\Lambda| - x)_+] + O(|f''(x)|x^{-3}e^{-x^2/4}) \tag{126}$$

Proof. This is immediate by using the Taylor expansion:

$$|f(y) - f(x) - (y-x)f'(x)| \leq \frac{(y-x)^2}{2} \cdot \sup_{z\in[x,y]} |f''(x)| \tag{127}$$

Taking expectation of $y = \Lambda$ over the support of $(|\Lambda| \geq x)$ we get the result. □

Lemma 17. *For f monotone and $x \leq \sqrt{d}$,*

$$\left|\mathbb{E}[f(|\Lambda_d|)\mathbb{I}_{|\Lambda_d|\geq x}] - \mathbb{E}[f(|\Lambda|)\mathbb{I}_{|\Lambda|\geq x}]\right| \lesssim \frac{1}{\sqrt{d}} \max\{f(x), f(\sqrt{d})\} + e^{-d/4} \sup_{t\geq\sqrt{d}} f(t)e^{-\sqrt{d}(t-\sqrt{d})/2} \tag{128}$$

Proof. Berry-Esseen gives $\sup_x |(\mathbb{P}(|\Lambda| \geq x) - \mathbb{P}(|\Lambda_d| \leq x))| \lesssim 1/\sqrt{d}$, so

$$LHS = \left|\int_x^\infty f'(t)(\mathbb{P}(|\Lambda| \geq x) - \mathbb{P}(|\Lambda_d| \leq x))\,dt\right| \tag{129}$$

$$\lesssim \frac{1}{\sqrt{d}} \int_x^{\sqrt{d}} |f'(t)| + \mathbb{E}[f(|\Lambda|)\mathbb{I}_{|\Lambda|\geq\sqrt{d}}] \tag{130}$$

$$= \frac{1}{\sqrt{d}} \int_x^{\sqrt{d}} |f'(t)| + \int_{\sqrt{d}}^\infty f(t)e^{-t^2/4}\,dt \tag{131}$$

$$\lesssim \frac{1}{\sqrt{d}} \int_x^{\sqrt{d}} |f'(t)| + e^{-x^2/4} \int_{\sqrt{d}}^\infty f(t)e^{-2x(t-x)}e^{-(t-x)^2/4}\,dt \tag{132}$$

$$\lesssim \frac{1}{\sqrt{d}} \int_x^{\sqrt{d}} |f'(t)| + e^{-d/4} \sup_{t\geq x} f(t)e^{-2x(t-x)} \tag{133}$$

□

Now we are ready to compute the order of x_*. Note that

$$\mathbb{E}\left[\left(\frac{1}{x_*} - \frac{1}{|\Lambda_d|}\right)_+\right] = \mathbb{E}\left[\left(\frac{1}{x_*} - \frac{1}{|\Lambda|}\right)_+\right] + O\left(\frac{1}{x\sqrt{d}} + \frac{e^{-d/4}}{\sqrt{d}}\right) \tag{134}$$

$$= \mathbb{E}\left[\left(\frac{1}{x_*} - \frac{1}{|\Lambda|}\right)_+\right] + o(1/\sqrt{d}) = \frac{4\beta}{\sqrt{d}} \tag{135}$$

However,

$$\mathbb{E}\left[\left(\frac{1}{x_*} - \frac{1}{|\Lambda|}\right)_+\right] \asymp \frac{1}{x_*^4} e^{-x_*^2/4} \tag{136}$$

so $x_* \asymp \sqrt{\log d}$, and

$$\mathbb{E}[(|\Lambda| - x_*)^n \mathbb{I}_{|\Lambda| \geq x_*}] \asymp \frac{1}{x_*^{n-3}} \cdot \frac{\beta}{\sqrt{d}} \tag{137}$$

Now we compute:

$$|\mathbb{E}[(|\Lambda_d| - x)_+] - \mathbb{E}[(|\Lambda| - x)_+]| = \left|\int_x^\infty (\mathbb{P}(|\Lambda_d| \geq t) - \mathbb{P}(|\Lambda| \geq t))\, dt\right| \tag{138}$$

$$\leq \frac{a}{\sqrt{d}} + \int_a^\infty |\mathbb{P}(|\Lambda_d| \geq t) - \mathbb{P}(|\Lambda| \geq t)|\, dt \tag{139}$$

$$\lesssim \frac{a}{\sqrt{d}} + \int_a^\infty e^{-t^2/2}\, dt \tag{140}$$

$$\lesssim \frac{a}{\sqrt{d}} + \frac{e^{-a^2/2}}{a} \tag{141}$$

Taking $a = \sqrt{\log d}$, the above is $O(\sqrt{\log d/d})$, while $\mathbb{E}(|\Lambda| - x_*)_+ \asymp \beta x_*^2/\sqrt{d}$, so

$$\mathbb{E}(|\Lambda_d| - x_*)_+ \sim \mathbb{E}(|\Lambda| - x_*)_+ \asymp \beta x_*^2/\sqrt{d}.$$

Similarly,

$$\mathbb{P}(|\Lambda_d| \geq x_*) = \mathbb{P}(|\Lambda| \geq x_*) + O(1/\sqrt{d}) \tag{142}$$

$$\asymp \frac{\beta x_*^3}{\sqrt{d}} + O(1/\sqrt{d}) \asymp \frac{\beta x_*^3}{\sqrt{d}} \tag{143}$$

Hence it follows that

$$\frac{\mathbb{E}(|\Lambda_d| - x_*)_+}{x_* \mathbb{P}(|\Lambda_d| \geq x_*)} \asymp \frac{1}{x_*^2} \tag{144}$$

This is relevant since

$$\frac{\mathcal{F}_d(x_*)}{\mathbb{P}(|\Lambda_d| \geq x_*)} = \frac{\mathbb{E}(|\Lambda_d| - x_*)_+}{x_* \mathbb{P}(|\Lambda_d| \geq x_*)} - \log x_* + \log\left(\frac{\mathbb{E}[\frac{1}{|\Lambda_d|}\mathbb{I}_{|\Lambda_d| \geq x_*}] + 4\beta/\sqrt{d}}{\mathbb{P}(|\Lambda_d| \geq x_*)}\right) \tag{145}$$

$$= \frac{\mathbb{E}(|\Lambda_d| - x_*)_+}{x_* \mathbb{P}(|\Lambda_d| \geq x_*)} - \log x_* + \log\left(\frac{1}{x_*} - \frac{1}{x_*^2} \cdot \frac{\mathbb{E}(|\Lambda_d| - x_*)_+}{\mathbb{P}(|\Lambda_d| \geq x_*)} + \Theta(x_*^{-3})\right) \tag{146}$$

$$= \frac{\mathbb{E}(|\Lambda_d| - x_*)_+}{x_* \mathbb{P}(|\Lambda_d| \geq x_*)} - \log\left(1 - \frac{\mathbb{E}(|\Lambda_d| - x_*)_+}{x_* \mathbb{P}(|\Lambda_d| \geq x_*)} + \Theta(x_*^{-2})\right) \tag{147}$$

$$= \Theta(x_*^{-2}) \tag{148}$$

so we conclude that $\mathcal{F}_d(x_*) \asymp x_*^{-2}\mathbb{P}(|\Lambda_d| \geq x_*) \asymp \frac{\beta x_*}{\sqrt{d}} \asymp \frac{\beta\sqrt{\log d}}{\sqrt{d}}$.

We briefly sketch the corresponding bound for the ℓ_2 covering number. We have the following bounds:

$$\mathbb{E}\log\left(1 + \frac{4|\Lambda_d|}{\delta}\mathbb{I}_{|\Lambda_n| \geq \delta/2}\right) \geq \lim_{n\to\infty} \frac{1}{n} \log N(E, \delta \cdot \tfrac{\beta}{\sqrt{d}}, \ell_2) \geq \mathbb{E}\left(\log \frac{|\Lambda_d|}{\delta}\right)_+ \tag{149}$$

By lemmas developed earlier, it is easy to check that as $\delta \to \infty$, the upper bound is of order $\delta^{-1}e^{-\delta^2/4}$ and the lower bound is of order $\delta^{-3}e^{-\delta^2/4}$. However, for any fixed $c > 0$,

$$\inf_{\delta > 0} \left\{\frac{\beta}{\sqrt{d}} \cdot \delta + \frac{1}{\delta^c} e^{-\delta^2/4}\right\} \asymp \frac{\beta}{\sqrt{d}} \cdot \sqrt{\log d} \tag{150}$$

so

$$\inf_{\delta>0}\{\delta n+\log N(E,\delta,\ell_2\} \asymp \frac{\beta}{\sqrt{d}}\cdot\sqrt{\log d} \tag{151}$$

which also matches the other bound.