These notes are based on [4], which provides a new algorithm using Jensen's formula and a reweighted second moment method to give a new algorithm for computing the partition function of mean-field spin glasses in a regime of parameters beyond what was previously known.

1 Introduction

We'll be working with *mean-field spin glasses*, which are measures on $\mathcal{C}_n = \{-1, +1\}^n$ or $\mathcal{S}_n = \{\sigma \in \mathbb{R}^n : \|\sigma\|_2^2 = n\}$ with *disordered interactions*. Specifically, we will work with the following random hamiltonian:

$$\mathcal{H}_G = \sum_{p=2}^{p_{\max}} \frac{\gamma_p}{n^{\frac{p-1}{2}}} \sum_{i_1,\dots,i_p=1}^{n} G_{i_1,\dots,i_p} \prod_{j=1}^{p} \sigma_{i_j},$$

where $\{\gamma_p\}_{p=2}^{p_{\max}}$ are deterministic nonnegative real coefficients and $G_{i_1,\dots,i_p}$ are independent standard normal random variables. Then we consider the measure

$$\mu_{G,\beta}(d\sigma) \propto \exp(\beta \mathcal{H}_G(\sigma))\rho(d\sigma),$$

where ρ is the base measure, either uniform on $\mathcal{C}_n$ (the Ising case) or on $\mathcal{S}_n$ (the spherical case). This is called the *mixed p-spin model*, and an important special case is the *Sherrington–Kirkpatrick model* where $p_{\max} = 2$ and $\gamma_2 = 1/\sqrt{2}$, in the Ising case.

Two important questions arise here: sampling from $\mu_{G,\beta}$ and "counting" in the sense of calculating the partition function of $\mu_{G,\beta}$, which is defined as

$$Z_G(\beta) = \mathbb{E}_{\sigma\sim\rho}[\exp(\beta \mathcal{H}_G(\sigma))].$$

There is an interesting relationship between these two tasks, and they are equivalent in many settings. First of all, if one can sample from $\mu_{G,\beta}$ then one can approximate the partition function using formulas such as

$$(\log Z_G)'(\beta) = \mathbb{E}_{\sigma\sim\mu_{G,\beta}}[\mathcal{H}_G(\sigma)]. \tag{1}$$

On the other hand, if one can approximate the partition functions of an extended family of models including *conditioned* models, then one can sample (in the Ising case) bit-by-bit using formulas such as

$$\mu(\sigma_i = 1|\sigma_j = 1) = \frac{\sum_{\sigma:\sigma_i,\sigma_j=1} e^{\beta\mathcal{H}(\sigma)}}{\sum_{\sigma:\sigma_j=1} e^{\beta\mathcal{H}(\sigma)}} = \frac{Z(\beta;\sigma_i,\sigma_j=1)}{Z(\beta;\sigma_j=1)}.$$

However, note that in the present setting we will *not* be able to do this as the family of models we consider are not self-reducible (the partition functions on the right above are not of the same form). The main result of [4] is an algorithm for approximating $Z_G(\beta)$, the weaker of the above two tasks.

Theorem 1. *For any $\varepsilon, \eta \in (0,1)$, there is $C = \Theta(\varepsilon^{-1})$ and $\delta = \delta(\varepsilon, n) > 0$ with $\delta \to 0$ as $n \to \infty$ such that*

$$\mathbb{P}_G[|\log Z_G(\beta) - T(\beta)| \le \eta] \ge 1 - \delta, \qquad \forall \beta \in \mathbb{D}(0, (1-\varepsilon)\beta_{2nd}),$$

where $T(\beta)$ is the degree-$C\log(n/\eta)$ Taylor polynomial of $\log Z_G(\beta)$, which may be computed (deterministically) in $n^{C\log(n/\eta)}$ time.

The second-moment threshold β_{2nd} (defined below) is higher than previously known methods for approximating the partition function, and in the case of the Sherrington–Kirkpatrick model is actually expected to be optimal. The main step in the proof of Theorem 1 is the following zerofreeness result:

Theorem 2. *For any $\varepsilon \in (0,1)$, there exists $\delta = \delta(\varepsilon, n) > 0$ with $\delta \to 0$ as $n \to \infty$ such that with probability at least $1-\delta$, the partition function $Z_G(\beta)$ is nonzero for $\beta \in \mathbb{D}(0, (1-\varepsilon)\cdot\beta_{2nd})$.*

2 Previous work and thresholds

There are a variety of thresholds of β that are relevant for discussing prior work. Let us restrict to the spherical setting for ease of exposition. The first three thresholds represent natural points of transition intrinsic to the model:

- Below β_{unique}, the Langevin dynamics is believed to have a spectral gap, and there should be no metastable states. Above this threshold, it is known that worst-case mixing is exponentially slow.
- Below β_{shatter}, most of the mass is still in the main metastable well, and Langevin dynamics mixes rapidly from a warm start. Above this threshold, there will be exponentially many metastable wells with exponentially small mass.
- Below β_{RS}, the model exhibits *replica symmetry*, meaning that the overlap $\frac{\langle\sigma,\tau\rangle}{n}$ between independent samples σ and τ will be very close to zero. Above this threshold, the measure concentrates on bands near the two polar opposite global minimizers of the Hamiltonian, and the spherical spin glass exhibits *one-step replica symmetry breaking*. This is the low temperature phase.

There are also two other thresholds which seem to be more related to specific proof or sampling techniques:

- Below β_{SL}, algorithmic stochastic localization successfully samples from $\mu_{G,\beta}$ and above which algorithmic stochastic localization fails. Simulated annealing is also proven to succeed below this threshold, but it may succeed beyond it as well.
- The threshold of this article, β_{2nd}, which delineates the so-called "second-moment regime" where Theorem 1 applies.

It is known that $0 \le \beta_{\text{unique}} \le \beta_{\text{SL}} \le \beta_{\text{shatter}} \le \beta_{\text{RS}}$, and I believe that β_{2nd} goes between β_{shatter} and β_{RS}, but I'm not sure about this. In any case, for pure p-spin models (where only one value of p appears in the Hamiltonian), one has the following asymptotic expressions:

$$\beta_{\text{unique}} = \Theta\left(\frac{1}{\sqrt{\log p}}\right), \qquad \beta_{\text{SL}}, \beta_{\text{shatter}} = \Theta(1) \text{ with } \frac{\beta_{\text{SL}}}{\beta_{\text{shatter}}} \approx \frac{\sqrt{e}}{2}, \qquad \beta_{\text{2nd}}, \beta_{\text{RS}} = \Theta\left(\sqrt{\log p}\right).$$

Additionally, note that in the Sherrington–Kirkpatrick model, we have $\beta_{\text{2nd}} = \beta_{\text{RS}} = 1$, meaning that Theorem 1 is expected to be optimal in this case.

To compare with previous work, let us just mention a few things that have been proven:

- For the SK model, a cluster expansion approximation (not completely unlike part of the proof of Theorem 1) can be used to get an $o_n(1)$-additive approximation of $\log Z_G(\beta)$ for all $\beta < \beta_{\text{2nd}}$ [1]. Note that Theorem 1 improves upon this by applying to more general settings as well as letting the error be arbitrarily small and not a specific function of n, at the cost of longer runtime.
- For the SK model, Langevin dynamics mixes rapidly (in time $O(n \log n)$) for $\beta < 0.295$ [2, 3].
- In the general mixed p-spin spherical setting, algorithmic stochastic localization can sample with $o_n(1)$ total variation error for $\beta < \beta_{\text{SL}}$ [5].

Both of the above sampling results can be used to derive some version of Theorem 1 as mentioned above. There are also other sampling results that work up to β_{RS} for the SK model but give worse bounds that cannot be used to replicate Theorem 1.

2.1 Definition of second-moment threshold

To define β_{2nd}, let us introduce the mixture function

$$\xi(s) = \sum_{p=2}^{p_{\max}} \gamma_p^2 s^p,$$

which satisfies

$$\mathbb{E}_G[\mathcal{H}_G(\tau) \cdot \mathcal{H}_G(\sigma)] = n \cdot \xi\left(\frac{\langle \tau, \sigma \rangle}{n}\right), \tag{2}$$

which expresses things as a function of the *overlap* $\frac{\langle \tau,\sigma \rangle}{n}$. Let us also denote by $h(m)$ the following entropy function on $[-1, 1]$:

$$h(m) = \begin{cases} -\frac{1}{2}((1-m)\log(1-m) + (1+m)\log(1+m)) & \text{for } \mathcal{C}_n, \\ \frac{1}{2}\log(1-m^2) & \text{for } \mathcal{S}_n. \end{cases}$$

We say we are in the *second moment regime* $\beta < \beta_{\text{2nd}}$ if β satisfies that

$$\varphi(m) \coloneqq \beta^2 \cdot \xi(m) + h(m)$$

has 0 as its unique global maximizer, and $\varphi''(0) < 0$.

For a bit of motivation, note that we will be attempting to apply the second-moment method to $Z_G(\beta)$ to show that it's nonzero, which means we'd like to have $\mathbb{E}[Z_G(\beta)^2] = (1 + o(1))\mathbb{E}[Z_G(\beta)]^2$. From (2) and the definition of $Z_G(\beta)$, we can see that

$$\begin{aligned} \mathbb{E}[Z_G(\beta)^2] &= \mathbb{E}_{\sigma,\tau\sim\rho}\left[\mathbb{E}_G\left[e^{\beta\mathcal{H}_G(\sigma)+\beta\mathcal{H}_G(\tau)}\right]\right] \\ &= \mathbb{E}_{\sigma,\tau\sim\rho}\left[e^{\frac{1}{2}\left(2\beta^2\mathbb{E}[\mathcal{H}_G(\mathbf{1})^2]+2\beta^2\mathbb{E}[\mathcal{H}_G(\sigma)\mathcal{H}_G(\tau)]\right)}\right] \\ &= \mathbb{E}[Z_G(\beta)]^2 \cdot \mathbb{E}_{\tau\sim\rho}\left[e^{n\beta^2\xi\left(\frac{\langle \mathbf{1},\tau\rangle}{n}\right)}\right]. \end{aligned}$$

Now, the number of τ having overlap m with $\mathbf{1}$ is approximately $e^{nh(m)}$, so we can see that the expectation on the right-hand side above is approximately

$$\int_{-1}^{1} e^{n(\beta^2\xi(m)+h(m))}\, dm.$$

Thus $\beta < \beta_{\text{2nd}}$ means that the above integral is *not* exponentially large in m, since the maximizer of the $n\varphi(m)$ in the exponent is $m = 0$, and the maximum there is $\varphi(0) = 0$. So we at least have a chance for the second moment method to work, but it will need some modification.

3 Taylor approximation via zerofreeness

Let us assume Theorem 2 for now, and see how that can be used to prove Theorem 1. The main tool is the following result, which says that if f is zerofree then we can approximate its logarithm using a Taylor expansion. Interestingly, one way to prove this is to use Jensen's formula, which is also used at another point in the proof for different reasons.

Proposition 3. *If f is zerofree in $\mathbb{D}(0, R)$ and analytic in a region that contains the closure of the disc, and there is some $L > 1$ such that $\left|\frac{f(w)}{f(0)}\right| \le L$ uniformly for $w \in \overline{\mathbb{D}(0, R)}$, then for any $r \in (0, R)$ and $\eta > 0$ there is some $C = C(r/R)$ such that for all $w \in \mathbb{D}(0, r)$,*

$$\left|\log f(w) - \sum_{k=0}^{C\log\left(\frac{\pi+\log L}{\eta}\right)} \frac{(\log f)^{(k)}(0)}{k!} w^k\right| \le \eta.$$

In our setting, we may take $L = e^{O(n)}$, which means we only need to consider $m = O(\log(n/\eta))$ terms of the Taylor series for $\log Z_G(\beta)$. It turns out we can actually calculate the coefficients of this series deterministically, once we have sampled the Gaussians $G_{i_1,\ldots,i_p}$. To do this, note that the derivatives of $\log Z_G(\beta)$ at $\beta = 0$ can be calculated in terms of the moments of $\mathcal{H}_G(\sigma)$ for $\sigma \sim \rho$, as in (1), since with $\beta = 0$ the measure $\mu_{G,\beta} = \rho$. The algorithm to get m Taylor coefficients of $\log Z_G(z)$ from the moments runs in $O(m^2)$ time, so this portion is easy relative to the next part, calculating the moments.

Lemma 4. *The first m moments of $\mathcal{H}_G(\sigma)$ for $\sigma \sim \rho$ can be computed deterministically in $n^{O(m)}$ time.*

To prove this, note that

$$\mathbb{E}_{\sigma\sim\rho}\left[\mathcal{H}_G(\sigma)^k\right] = \sum_{p_1,\ldots,p_k=2}^{p_{\max}} \sum_{\substack{i_l^j=1 \\ \text{for } 1\le l\le p_j \\ \text{for } 1\le j\le k}}^{n} \left(\prod_{j=1}^{k} \frac{\gamma_{p_j}}{n^{\frac{p_j-1}{2}}} G_{i_1^j,\ldots,i_{p_j}^j}\right) \cdot \mathbb{E}_{\sigma\sim\rho}\left[\prod_{j=1}^{k}\prod_{l=1}^{p_j} \sigma_{i_l^j}\right].$$

Now in the Ising case (base measure being the hypercube $\mathcal{C}_n$), all of the expectations on the right may be calculated easily: they are 1 if each σ_i appears an even number of times and 0 otherwise. There is also an explicit formula for the expectation in terms of the indices $\{i_l^j\}$ in the spherical case. The algorithm is to simply enumerate the sum and compute by brute force; there are $n^{O(k)}$ terms in the inner sum since $p_{\max}$ is a constant, and in the outer sum there are $e^{O(k)}$ terms, leading to the desired runtime.

4 Starting on the proof of zerofreeness

To prove Theorem 2, we will show that

$$\mathbb{E}_G[\text{number of zeroes of } Z_G \text{ in } \mathbb{D}(0,(1-\varepsilon)\beta_{2\text{nd}})] = o_n(1),$$

and then use Markov's inequality. To get a handle on this expectation, we will use Jensen's formula. For a bit of intuition here, normally we can use the second moment method to show that a single random variable is nonzero. But Jensen's formula will allow us to encode all the information about zeroes of a *function* inside of a disk in a single random variable, which allows us to use the second moment method to show that the function is nonzero in the entire disk.

Theorem 5 (Jensen's formula)**.** *Let f be a complex function which is analytic on an open set containing $\overline{\mathbb{D}(0,R)}$, with $f(0)\neq 0$. Then we have the following equality:*

$$\sum_{w\in\mathbb{D}(0,R):f(w)=0} \log\left(\frac{R}{|w|}\right) = \frac{1}{2\pi}\int_0^{2\pi} \log\left|\frac{f(Re^{i\theta})}{f(0)}\right| d\theta.$$

I previously thought that Jensen's formula was a consequence of Cauchy's integral formula, but it actually seems to be a more direct consequence of the fact that if f is a zerofree analytic function then $\log|f(z)|$ is harmonic, as it is the real part of $\log f(z)$. In this case, we find that

$$\log|f(0)| = \frac{1}{2\pi}\int_0^{2\pi} \log|f(Re^{i\theta})|\, d\theta.$$

And if f has some zeroes $w_1,\ldots,w_k\in\mathbb{D}(0,R)$, then apply the above to the function

$$g(z) = \frac{f(z)}{(z-w_1)\cdots(z-w_k)},$$

which has no zeroes. We obtain

$$\log|f(0)| - \sum_{w\in\mathbb{D}(0,R):f(w)=0} \log|w| = \frac{1}{2\pi}\int_0^{2\pi}\log|f(Re^{i\theta})|\,d\theta - \sum_{w\in\mathbb{D}(0,R):f(w)=0} \frac{1}{2\pi}\int_0^{2\pi}\log|Re^{i\theta}-w|\,d\theta,$$

and each integral in the sum on the right-hand side is equal to

$$\log R + \frac{1}{2\pi}\int_0^{2\pi}\log|e^{i\theta}-w/R|\,d\theta.$$

Finally, the integral above is zero regardless of the value of w as it is the real part of the integral

$$\frac{1}{2\pi}\int_0^{2\pi}\log(e^{i\theta}-|w/R|)\,d\theta,$$

which is purely imaginary, as can be seen using integration by parts and the residue theorem.

Now let us define $r = (1-\varepsilon)\beta_{2\mathrm{nd}}$ and $R = (1-\frac{\varepsilon}{2})\beta_{2\mathrm{nd}}$. With Jensen's formula in hand, we see that

$$\begin{aligned}\text{number of zeroes of } f \text{ in } \mathbb{D}(0,r) &\le \left(\log\frac{R}{r}\right)^{-1} \sum_{w\in\mathbb{D}(0,r): f(w)=0} \log\left(\frac{R}{|w|}\right) \\ &\le \frac{1}{\varepsilon}\frac{1}{2\pi}\int_0^{2\pi} \log\left|\frac{f(Re^{i\theta})}{f(0)}\right| d\theta.\end{aligned}$$

Now applying the above to a random function $f = f_G$, taking expectations and using Jensen's *inequality* with the concave function $\log\sqrt{x}$, we find

$$\mathbb{E}[\text{number of zeroes of } f_G \text{ in } \mathbb{D}(0,r)] \le \frac{1}{4\pi\varepsilon}\int_0^{2\pi} \log\mathbb{E}_G\left[\left|\frac{f_G(Re^{i\theta})}{f_G(0)}\right|^2\right].$$

It would be tempting at this point to plug in $f_G = Z_G$ and continue. For this, we would like the integral on the right-hand side to be close to zero, meaning that we'd need

$$\mathbb{E}_G\left[|Z_G(\beta)|^2\right] \approx Z_G(0)^2 = 1 \tag{3}$$

for most values of β on the circle of radius R. However, recall that

$$Z_G(\beta) = \mathbb{E}_{\sigma\sim\rho}[\exp(\beta\mathcal{H}_G(\sigma))],$$

and $\mathcal{H}_G(\sigma)$ has Gaussian-type fluctuations at scale 1 (in terms of G) around a mean of order n. Thus we should expect that $Z_G(\beta) \approx \exp(cn + \Phi_G(\beta))$, where $\Phi_G(\beta)$ also has Gaussian-type fluctuations at scale 1. Note that for $Z \sim \mathcal{N}(0,1)$, we have

$$\mathbb{E}\left[e^{aZ}\right] = e^{\frac{a^2}{2}},$$

so we should in fact not have (3), even if we divide out by the constant e^{cn} first.

To fix this, we will need to study $Z_G(\beta)$ more closely. In fact, we will have to extract the part of $\Phi_G(\beta)$ which fluctuates at scale 1, and write

$$Z_G(\beta) = \exp\left(\Phi_G^1(\beta) + \Phi_G^2(\beta)\right),$$

where $\Phi_G^1(\beta)$ fluctuates at scale 1 and $\Phi_G^2(\beta)$ has fluctuations tending to zero with n. Now defining the reweighting factor $A_G(\beta) = e^{-\Phi_G^1(\beta)}$, we will thus have

$$\mathbb{E}_G\left[\left|A_G(\beta)Z_G(\beta)\right|^2\right] = 1 + o_n(1),$$

allowing us to show that $f = A_G Z_G$ has no zeroes in $\mathbb{D}(0,r)$ with high probability. Thus, the remaining work and the main contribution of [4] is to determine $A_G(\beta)$ and prove that it is analytic in $\overline{\mathbb{D}(0,R)}$.

4.1 The reweighting factor

Here we write down the formula for $A_G(\beta)$, leaving the derivation for next time. First, define the matrix $M = \nabla^2\mathcal{H}_G(0)$, which is just the same thing as $\mathcal{H}_G$ in the SK case. We will think of the index set of M as edges $\{i,j\}$ of the complete graph K_n. Define

$$\mathrm{UC}(n) = \{\Gamma \subseteq K_n : \Gamma \text{ is a disjoint union of cycles and } |\Gamma| \le \log\log n\},$$

and for $\Gamma \in \mathrm{UC}(n)$, let $c(\Gamma)$ denote the number of connected components of Γ. Finally, define $\zeta = \beta\sqrt{\xi''(0)}$. Then the formula for the reweighting factor $A_G(\beta)$ is as follows:

$$\begin{aligned}A_G(\beta) = {}& \frac{1}{\sqrt{1-\zeta^2}}\exp\left(-\frac{n\beta^2}{2}\xi(1) + \frac{n\zeta^2}{4} - \frac{\zeta^2}{2} - \frac{\zeta^4}{8}\right) \\ &\times \exp\left(-\frac{\beta}{2}\mathrm{Tr}(M) - \frac{\beta^2}{2}\|M\|_2^2\right)\sum_{\Gamma\in\mathrm{UC}(n)}(-1)^{c(\Gamma)}\prod_{e\in\Gamma}\beta M_e.\end{aligned}$$

Note that the definition of $\beta_{2\mathrm{nd}}$ implies that $|\zeta| \le 1 - \frac{\varepsilon}{2}$ for $\beta \in \overline{\mathbb{D}(0,R)}$, so that this formula for $A_G(\beta)$ is analytic in the required domain (the factor in front is the only problematic one).

References

[1] Michael Aizenman, Joel L Lebowitz, and David Ruelle. Some rigorous results on the sherrington-kirkpatrick spin glass model. *Communications in mathematical physics*, 112(1):3–20, 1987.

[2] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence: optimal mixing of down-up random walks. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1418–1430, 2022.

[3] Nima Anari, Frederic Koehler, and Thuy-Duong Vuong. Trickle-down in localization schemes and applications. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1094–1105, 2024.

[4] Ferenc Bencs, Brice Huang, Daniel Z Lee, Kuikui Liu, and Guus Regts. On zeros and algorithms for disordered systems: mean-field spin glasses. *arXiv preprint arXiv:2507.15616*, 2025.

[5] Brice Huang, Andrea Montanari, and Huy Tuan Pham. Sampling from spherical spin glasses in total variation via algorithmic stochastic localization. *arXiv preprint arXiv:2404.15651*, 2024.