A FIRST LOOK AT THE CRITICAL TEMPERATURE OF A SPIN GLASS WITH RANDOM MATRIX THEORY

SIMONE MARIA GIANCOLA[*] [†]

Last Modified on March 26, 2025

Abstract

The Sherrington-Kirkpatrick model is the simplest example of a spin glass (Sherrington and Kirkpatrick 1975), where we just let the couplings of the Curie-Weiss model become Gaussian. The random interactions form a non-trivial behavior which is fertile ground for random matrix theory techniques. In this document we propose one classical computation of the critical temperature by Potters and Bouchaud (2020), with more theoretical context and explicit computations. This is an excuse to present the Harish-Chandra-Itzkynson-Zuber integral (Harish-Chandra 1957; Itzykson and Zuber 1980) in two forms taken from Potters and Bouchaud (2020) and Tao (2013).

CONTENTS

1 Introduction 1
- 1.I The simplest spin glass in a few lines 2
- 1.II Preliminary computation 2

2 The low-rank Harish-Chandra-Itzykson-Zuber asymptotic integral formula 3
- 2.I A theorem for the finite size HCIZ integral formula 5
  - 2.I.1 Route one 7
  - 2.I.2 Route two 8

3 Partial phase diagram of the model 8

References 12

A Details on theory and computations 13
- A.I Useful lemmas 13
- A.II Saddle-point method for low-rank HCIZ integrals 14
- A.III Rotating the R-transform 15
- A.IV Clever integrations with delta functions to decouple replicas 15

1 INTRODUCTION

In this short document we summarize the first computation of the critical temperature for a model of statistical physics. In doing so, we will use some ideas from the lectures on random matrix theory to understand a trick from physics that is as effective as it is non-rigorous. For the fun, it will touch on topics of courses from this semester. It will also allow us to have a basic vocabulary of fields that were ignored instead, such as spin glasses.
Computations, when performed, are explicit and pedagogical. References are non-exhaustive. We reroute the reader to the cited works for more context.

NOTATION Most of the symbols are standard. The only difference we make is between what is random and what is *not*, what is scalar, what is vectorial and what is matricial. For example, $a, b, c, x, y, z, \alpha, \beta, \gamma$ is a variable, while $\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{x}, \mathrm{y}, \mathrm{z}, \upalpha, \upbeta, \upgamma$ is a random variable. Similarly, $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ is a vector; $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\upalpha}, \boldsymbol{\upbeta}, \boldsymbol{\upgamma}$

[*]While at Université Paris-Saclay, Laboratoire de Mathématiques d'Orsay
[†]simonegiancola09@gmail.com

is a random vector. Again, $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Theta}$ is a matrix; $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{\Lambda}, \mathbf{\Psi}, \mathbf{\Theta}$ is a random matrix. An expectation $\mathbb{E}_{\mathrm{x}}\left[\mathrm{x}y\mathrm{z}\right] = \int xy\mathrm{z}\,\mathrm{d}\mathbb{P}\left[x\right]$ is such that y is deterministic, and we integrate out against x which is deterministic once it is expressed inside an integral, keeping z random throughout.

1.I *The simplest spin glass in a few lines*

{sec:intro

Without much regard to the motivation, let us present the Sherrington-Kirkpatrick model as a generalization of the fully-connected Ising model, also named Curie-Weiss model. We have an energy function:

$$\mathscr{E}(\boldsymbol{s}, \boldsymbol{J}) := \frac{1}{2}\boldsymbol{s}^\top \boldsymbol{J}\boldsymbol{s}, \qquad \boldsymbol{s} \in \{\pm 1\}^N. \tag{1.1}$$ {{eqn:hami

If we wanted to study the Ising model, we would take the Boltzmann measure proportional to $\mathrm{e}^{\beta\mathscr{E}(\boldsymbol{s})}$ at given $\boldsymbol{J}$ with each entry having the same sign. By construction, the Ising/Curie-Weiss model puts more weight on configurations $\mathbf{s}$ that align. However, we might as well say that $\mathbf{J}$ is a random Gaussian matrix! For scaling reasons, we will consider the GOE only. This (modulo higher order considerations) is an **example of a spin glass**. In other words, a random quadratic function with random coefficients. What happens is that we have two sources of randomness, in the spins and in their interactions. From now onwards, we will refer to the randomness of $\mathbf{J}$ as *disorder* (randomness). Ultimately the objects we wish to know are the same, and the free energy, now depending on disorder, is our favorite. It turns out that empirically and by concentration arguments the free energy concentrates exponentially fast to its expectation (see the discussion in appendix A.I). Without even introducing the former, we define the latter. We work with the probability density:

$$p(\boldsymbol{s}; \mathbf{J}, N, \beta) = \frac{1}{\mathcal{Z}(\mathbf{J}, N; \beta)}\mathrm{e}^{\beta \boldsymbol{s}^\top \mathbf{J}\boldsymbol{s}}, \qquad \mathcal{Z}(\mathbf{J}, N; \beta) = \int_{\{\pm 1\}^N} \mathrm{e}^{\beta \boldsymbol{s}^\top \mathbf{J}\boldsymbol{s}}\mathrm{d}\boldsymbol{s}, \qquad \beta \in (0, \infty), \tag{1.2}$$

where β is just an inverse temperature parameter and $\mathbf{J}$ is a Wigner matrix with the correct scaling. Then, the *quenched* free energy is the expectation of the log partition function (i.e. the cumulant generating function):

$$f(\beta) := \lim_{N\to\infty} \frac{1}{N}\mathcal{F}(N, \beta) = \lim_{N\to\infty} \frac{1}{\beta N}\mathbb{E}_{\mathbf{J}}\left[\ln \mathcal{Z}(\mathbf{J}, N; \beta)\right] \tag{1.3}$$

We say this is quenched because we have another version termed "annealed" which amounts to switching the expectation and the logarithm. The names come from a physical interpretation. To see a phase transition, we take the limit as $N \to \infty$ and study the behavior as β varies. In terms of questions, we are basically asked to compute the limit of a Gaussian integral of a log of a sum of exponentials.

Example 1.4 (Ubiquity of spin glasses)**.** *In many machine learning problems we take a square loss. If we optimize over a parameter* $\boldsymbol{\beta}$ *and the matrix of observations is modelled randomly then the loss is a spin-glass energy function with magnetic field (the linear term). If we take any optimization problem and properly randomize it we will see some version of a spin glass. The idea is that by studying a random problem one extracts insights on the deterministic problem, in exchange for tractability. For many examples, see the literature mentioned in (*Zdeborová and Krzakala 2016*).* {exm:spin

Our objective is to compute the quenched free energy quickly using random matrix theory techniques, exactly because $\mathbf{J}$ is random!

1.II *Preliminary computation*

{sec:compu

For the sake of this document we compute an expression that is exact below a certain critical parameter β_c. The full description is more complicated and is actually one of the reasons of recent Abel and Nobel Prizes (Parisi 2023; Sourav Chatterjee 2024). We highlight non-rigorous computations with a red equality for the sake of clarity. As a matter of fact, we may say that what follows is the essence of the distance from physics and mathematics: we use Taylor expansions, exchange limits and presume everything is well-behaved.

Remark 1.5 (Important remark)**.** *One non-rigorous step we do not highlight is the use of delta functions without regard to details.*

The starting point is a seemingly innocent relation:

$$\log a = \lim_{n\to 0} \frac{a^n - 1}{n} = \partial_n a^n \Big|_{n=0}. \tag{1.6}$$

The second equality is trivial, the first equality is obtained by letting $a^n = e^{n \log a}$ and expanding with the Taylor expression of the exponential. The magic is that the expectation of a logarithm is annoying, while the expectation of a derivative is nice, in the sense that:

$$\mathbb{E}\left[\log \mathrm{a}\right] = \partial_n \mathbb{E}\left[\mathrm{a}^n\right]\Big|_{n=0}, \tag{1.7}$$

where we exchanged derivative and expectation. In particular, the right-hand side is rather expressed for $n \in \mathbb{N}$ as an integral over n independent copies (replicas) of a. Eventually, we take $n = 0$, but this is a non-rigorous step we keep for later. Let us focus on the expectation of the product. If we replace a by the partition function $\mathcal{Z}$, we have n copies from the *same* Boltzmann measure dependent on the $\mathbf{J}$ disorder:

$$\mathbb{E}_{\mathbf{J}}\left[\mathcal{Z}^n\right] = \mathbb{E}_{\mathbf{J}}\left[\sum_{\{s^{(\alpha)}\}_{\alpha\in[n]},\, s^{(\alpha)}\in\{\pm1\}^N} \prod_{\alpha=1}^{n} e^{\beta/2\left[s^{(\alpha)}\right]^\top \mathbf{J} s^{(\alpha)}}\right]. \tag{1.8}$$

Now we just use the additivity of exponentials, matrix index summation, trace properties, and the decomposition of $\mathbf{J} = \mathbf{O}\boldsymbol{\Lambda}\mathbf{O}^\top$ to obtain:

$$\mathbb{E}_{\mathbf{J}}\left[\mathcal{Z}^n\right] = \mathbb{E}_{(\mathbf{O},\boldsymbol{\Lambda})}\left[\mathbb{E}_{\mathbf{M}^{(n)}\sim\mathrm{UnifRep}}\left[\exp\left\{N\beta/2\,\mathrm{Tr}\left(\mathbf{O}\boldsymbol{\Lambda}\mathbf{O}^\top\mathbf{M}^{(n)}\right)\right\}\right]\right], \qquad \mathrm{M}_{ij}^{(n)} := \sum_{\alpha=1}^{n} \frac{\mathrm{s}_i^{(\alpha)}\mathrm{s}_j^{(\alpha)}}{N}, \tag{1.9}$$ {{eqn:fina

where by $\mathbb{E}_{\mathbf{M}^{(n)}\sim\mathrm{UnifRep}}[\cdot]$ we mean that we just sum with equal weights over all configurations $\mathbf{s}^{(\alpha)} \in \{\pm1\}^N$ for all replicas $\alpha \in [n]$. Alternatively, we have $\mathbf{M}^{(n)} = 1/N \sum_{\alpha=1}^{n} \mathbf{s}^{(\alpha)}\left[\mathbf{s}^{(\alpha)}\right]^\top$. From this last expression we get automatically that $\mathbf{M}^{(n)}$ is at most of rank n, and we want to compute the limit as $N \to \infty$ of this integral. Now we make some informal observations:

- as $N \to \infty$, the random matrix $\boldsymbol{\Lambda}$ converges to the semicircular law;
- as $N \to \infty, n \to 0$ the matrix $\mathbf{M}^{(n)}$ is effectively low-rank compared to the others.

For said reasons, we might be interested in computing an integral of the type above for low-rank matrices. It turns out that we have a general expression, termed HCIZ integral (for Harish-Chandra-Itzykson-Zuber), that covers our special case. We present an asymptotic derivation for our low-rank case and a rigorous one for the full-rank finite size case.

Remark 1.10 (Is this the unique instance for HCIZ-type integrals?). *Differently from the "ubiquity of spin glasses", we cannot say so much. However, the HCIZ integral does appear in literature. There are some examples in mathematics and theory of machine learning. In particular, we reroute the reader to the works on matrix denoising (Maillard et al. 2024; Pourkamali, Barbier, and Macris 2024), the comments in the blog of Tao (2013) which also contain a discussion on the non-circularity of the next proof and the review work (McSwiggen 2021). For these reasons, it makes sense to know about it.*

2 THE LOW-RANK HARISH-CHANDRA-ITZYKSON-ZUBER ASYMPTOTIC INTEGRAL FORMULA

{sec:HCIZ

In this section we will present two derivations of the HCIZ integral formula. In particular, we will use the former to continue our computation, since it is a plug-in result. It is the book of Potters and Bouchaud (2020, chapter 10), and it should be an adaptation of a physics computation (Marinari, Parisi, and Ritort 1994).

The HCIZ formula has many complicated interpretations (see e.g. (McSwiggen 2021)), let us use a simpler one from (Potters and Bouchaud 2020). A classical property of random variables is that the distribution of the sum is easy to compute when Fourier transforms are available. In particular, the Fourier transform of a sum of independent variables is the product of Fourier transforms. We seek an analogue for free random variables taking values in a matrix space. A good candidate for the Fourier transform is a function that:

(D1) is the exponential of a scalar term;

(D2) depends only on the spectrum of $\mathbf{A}$;

(D3) its logarithm is additive for free random variables.

Let us propose the following formula:

$$\phi(\boldsymbol{T};\mathbf{A}) := \mathbb{E}_{\mathbf{O}}\left[\exp\left\{{}^{N}\!/\!_{2}\mathrm{Tr}\left(\boldsymbol{T}\mathbf{O}\mathbf{A}\mathbf{O}^\top\right)\right\}\right]. \tag{2.1}$$

Item (D1) is automatic. Item (D2) follows by applying a rotation to $\boldsymbol{T}$ or $\mathbf{A}$ inside the expectations, since the trace is invariant to rotations, products of rotations remain rotations, and the Haar distribution is the unique left-right invariant distribution. The third term is more subtle. If we take a further random rotation, we know that the spectrum of $\mathbf{C} = \mathbf{A} + \mathbf{O}_1\mathbf{B}\mathbf{O}_1^\top$ will not depend on $\mathbf{O}_1$. In other words, taking an average of the limiting object with respect to $\mathbf{O}_1$ we will have any realization. Checking the expression for the law of $\mathbf{A} + \mathbf{O}_1\mathbf{B}\mathbf{O}_1^\top$ integrated over $\mathbf{O}_1$ we find additivity of the logarithm since $\mathbf{O}\mathbf{O}_1 \sim \mathrm{Haar}(\mathbb{O}_N)$, and it is independent of $\mathbf{O}$ (see lemma A.4). Basically, the integrals decouple.

Remark 2.2. *Notice that unlike the scalar Fourier transform this expression is random! We will see later that it concentrates, but for now it is a random variable.*

The expression for general $\boldsymbol{T}$ is more advanced. Let us start from a rank-1 appetizer. We take $\boldsymbol{T} = t\boldsymbol{v}\boldsymbol{v}^\top$. Without loss of generality, $\mathbf{A}$ is diagonal. Like in large deviations, a statement of the form:

$$\phi(t\boldsymbol{v}\boldsymbol{v}^\top;\mathbf{A}) = \mathrm{e}^{N/2\,\frac{2}{N}\log\phi(t\boldsymbol{v}\boldsymbol{v}^\top;\mathbf{A})} \overset{N\to\infty}{\sim} \mathrm{e}^{N/2\,h_{\mathbf{A}}(t)}, \tag{2.3}$$

where

$$h_{\mathbf{A}}(t) := \lim_{N\to\infty}\frac{2}{N}\log\mathbb{E}_{\mathbf{O}}\exp\left\{\frac{tN}{2}\mathrm{Tr}\left(\boldsymbol{v}\boldsymbol{v}^\top\mathbf{O}\mathbf{A}\mathbf{O}^\top\right)\right\}, \tag{2.4}$$ {{eqn:ha t

makes sense if $h_{\mathbf{A}}(t)$ has a well-defined limit. Then, we will look at the normalized log-exp in the argument of the exponential, which we saw is additive. As a first observation, we use again cyclicity of the trace to interpret $\mathbf{O}^\top t\boldsymbol{v}\boldsymbol{v}^\top\mathbf{O} = \mathbf{p}\mathbf{p}^\top$ as a random projection with $\|\mathbf{p}\|^2 = t$ and $\mathbf{p}/\|\mathbf{p}\| \sim \mathrm{Unif}(\mathbb{S}^{N-1})$, which justifies ignoring the dependence on $\boldsymbol{v}$ in the definition. To compute the inner expectation with respect to $\mathbf{O}$, we may then integrate against $\mathbf{p}$. The expression is a Gaussian integral, since the trace is a quadratic form $\mathbf{p}^\top\mathbf{A}\mathbf{p}$. We add convenient factors and perform a change of variables from $\mathbf{z} = \sqrt{N}\mathbf{p}$ to find:

$$\mathbb{E}_{\mathbf{O}}\left[\exp\left\{\frac{tN}{2}\mathrm{Tr}\left(\boldsymbol{v}\boldsymbol{v}^\top\mathbf{O}\mathbf{A}\mathbf{O}\right)\right\}\right] = \frac{I_t(\mathbf{A})}{I_t(\mathbf{0}_{N\times N})}, \qquad I_t(\mathbf{A}) := \int\frac{1}{(2\pi)^{N/2}}\delta(\|\boldsymbol{z}\|^2 - Nt)\mathrm{e}^{1/2\mathbf{z}^\top\mathbf{A}\mathbf{z}}\mathrm{d}\boldsymbol{z}. \tag{2.5}$$

This passage is quite strange but just requires some calculus: the Dirac delta enforces the norm to be consistent with the random projection, and the $I_t(\mathbf{0}_{N\times N})$ term normalizes the expression. The convenient thing is that now we can write down the Dirac delta via its integral expression and apply the saddle-point method. These will return the asymptotic behavior of $h_{\mathbf{A}}(t)$. Let us provide some details.

Recall that the Dirac delta admits the representation:

$$\delta(x) = \int_{-\infty}^{\infty}\frac{\mathrm{e}^{\mathrm{i}zx}}{2\pi}\,\mathrm{d}z = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty}\frac{\mathrm{e}^{-zx/2}}{4\mathrm{i}\pi}\,\mathrm{d}z, \tag{2.6}$$

where the second equality is just a change of variables, and $i\infty$ means that we take complex numbers with large imaginary part. We combine this with the fact that since we force $\|\mathbf{z}\|^2 = Nt$ we may represent unity as:

$$1 = \mathrm{e}^{-\lambda(\|\mathbf{z}\|^2 - Nt)/2}, \qquad \lambda > \lambda_{\max}^{\mathbf{A}}. \tag{2.7}$$

Plugging these two clever equations into the integral we find:

$$\begin{aligned} I_t(\mathbf{A}) &= \int_{\lambda-\mathrm{i}\infty}^{\lambda+\mathrm{i}\infty}\frac{1}{4\pi}\int\frac{1}{(2\pi)^{N/2}}\exp\left\{-\frac{1}{2}\boldsymbol{z}^\top(z\boldsymbol{I}_N - \mathbf{A})\boldsymbol{z} + \frac{Nzt}{2}\right\}\mathrm{d}\boldsymbol{z}\,\mathrm{d}z & (2.8)\\ &= \int_{\lambda-\mathrm{i}\infty}^{\lambda+\mathrm{i}\infty}\frac{1}{4\pi}\det\left(z\boldsymbol{I}_N - \mathbf{A}\right)^{-1/2}\exp\frac{Nzt}{2}\,\mathrm{d}z \qquad \text{Gaussian integration} & (2.9)\\ &= \int_{\lambda-\mathrm{i}\infty}^{\lambda+\mathrm{i}\infty}\frac{1}{4\pi}\exp\left\{\frac{N}{2}f_t(z,\mathbf{A})\right\}\mathrm{d}z, & (2.10)\end{aligned}$$

where we defined the function by putting everything in the exponential. Explicitly:

$$f_t(z, \mathbf{A}) := zt - \frac{1}{N} \sum_k \log(z - \lambda_k^{\mathbf{A}}). \tag{2.11}$$

{{eqn:defi

Now that we have a classical equation in the saddle-point sense, we perform the derivatives. By the stationary phase approximation we know that dominating points are those that are constant in the imaginary variable z at the saddle. Simple algebra (differentiating $f_t(z, \mathbf{A})$) gives us:

$$t - \frac{1}{N} \sum_k \frac{1}{z - \lambda_k^{\mathbf{A}}} = t - s_N^{\mathbf{A}}(z) = 0. \tag{2.12}$$

{{eqn:opti

In particular, the optimal point in z depends on t implicitly. Assuming the Stieltjes-Cauchy transform $s_N^{\mathbf{A}}(z)$ is invertible, which is true since we took $\lambda > \lambda_{\max}^{\mathbf{A}}$, we can take some time to deform the contour to pass through $\zeta(t)$ the inverse and apply a nice saddle-point approximation.[1] We find an initial implicit expression:

$$h_{\mathbf{A}}(t) = \zeta(t)t - 1 - \log t - \frac{1}{N} \sum_k \log(\zeta(t) - \lambda_k^{\mathbf{A}}) \tag{2.13}$$

{{eqn:impl

$$= \hbar(\zeta(t), t), \tag{2.14}$$

where we are quite sloppy in keeping an N factor after taking the limit, but recognize that the sum is order N so it should be fine. Alternatively, we see it as its limiting object, i.e. an integral over the limiting spectral measure of $\mathbf{A}$. In this sense, the following discussion will hold only in the limit $N \to \infty$.
While seemingly complicated, we can still extract information from the approximation we made. We just said that the result is stationary in z. This allows us to find a good expression for the derivative of $h_{\mathbf{A}}(t)$. Indeed:

$$\frac{\mathrm{d}h_{\mathbf{A}}(t)}{\mathrm{d}t} = \underbrace{\partial_z \hbar(\zeta(t), t)}_{=0} \frac{\mathrm{d}\zeta(t)}{\mathrm{d}t} + \partial_t \hbar(\zeta(t), t) = \zeta(t) - \frac{1}{t} = R_{\mathbf{A}}(t), \tag{2.15}$$

where in the last equality we recognized one of the definitions of the R-transform of a matrix (Potters and Bouchaud 2020, eqn. 10.10). From the identity $h_{\mathbf{A}}(0) = 0$ we may apply the fundamental theorem of calculus to write:

$$h_{\mathbf{A}}(t) = \int_0^t R_{\mathbf{A}}(x)\, \mathrm{d}x, \tag{2.16}$$

and by additivity of $h_{\mathbf{A}}(t)$ for free matrices (by construction), we obtain additivity of its derivative![2] Having approximated the normalized log exp, we may say that:

$$\phi(t\boldsymbol{v}\boldsymbol{v}^\top; \mathbf{A}) \overset{N \to \infty}{\simeq} \exp\left\{ \frac{N}{2} h_{\mathbf{A}}(t) \right\}. \tag{2.17}$$

More importantly, for any low-rank matrix $\boldsymbol{T}$, the result is the same, with just more calculus, i.e. we find:

$$\phi(\boldsymbol{T}; \mathbf{A}) \overset{N \to \infty}{\approx} \exp\left\{ \frac{N}{2} \mathrm{Tr}\left(h_{\mathbf{A}}(\boldsymbol{T}) \right) \right\}. \tag{2.18}$$

Having made some non-rigorous reasonings, let us propose a theorem from the blog of Tao (2013).

2.1 *A theorem for the finite size HCIZ integral formula*

This derivation is quite shallow in depth of meaning. It is not due to the original author, but rather that the connections are very subtle. We will write the bilinear form of the heat flow on the space of Hermitian matrices (i.e. imagine heat flow on the reals and you get Brownian motion), which is strongly connected with the matrix-analogue of Brownian motion: Dyson Brownian motion. This short note is not suited to give a complete account of what is going on, but the blog-posts (Tao 2010, 2013), the book chapters (Potters and Bouchaud 2020, chap. 11-13) and surprisingly the comments by Zuber himself in the blog should be a good starting point. Let us now move to the actual computation. We want to prove the following theorem.

[1] See appendix A.II for details.
[2] Notice how we made a quite informal step when taking the limit, but the fact that additivity of the R-transform holds only in the limit is non-contradicting, so we are not taking nonsensical conclusions.

Theorem 2.19. *Let A, B be Hermitian matrices with simple eigenvalues ordered increasingly.*[3] *For $\mathbf{U} \sim \mathrm{Haar}(\mathbb{U}_N)$ a unitary matrix uniformly distributed, and $t \in \mathbb{C}\backslash\{0\}$ we have:*

$$\mathbb{E}_{\mathbf{U}}\left[\exp\left\{t\mathrm{Tr}\left(\boldsymbol{A}\mathbf{U}\boldsymbol{B}\mathbf{U}^\dagger\right)\right\}\right] = c_N \frac{\det\left(\exp\left\{t\lambda_i^{\boldsymbol{A}}\lambda_j^{\boldsymbol{B}}\right\}\right)_{i,j\in[N]}}{t^{N^2-N/2}\Delta(\lambda^{\boldsymbol{A}})\Delta(\lambda^{\boldsymbol{B}})}, \qquad \begin{cases} c_N := \prod_{i=1}^{N-1} i!, \\ \Delta(\lambda^{\boldsymbol{A}}) & \textit{Vandermonde determinant.}\end{cases} \tag{2.20}$$

Remark 2.21. *From the assumption of a simple spectrum, we cannot conclude directly that the previous formula generalizes to this one. Moreover, we take an asymptotic $N \to \infty$ that is not evident from this formula. For these reasons, this section is a mathematical vindication of the arguments we will use, rather than a justification.*

We first set some notation and reminders from class. The space of $N \times N$ Hermitian matrices is $\mathscr{H}_{\mathbb{C}}(N)$, where we avoid mentioning the complex plane field as it is superfluous. It is equipped with the Haar uniform measure $\mathrm{d}\mu(\boldsymbol{M}_N)$. Throughout, normalizations do not matter as they cancel out.[4] We endow the measure space with a functional calculus in the following sense. Functions act on matrices invariantly by rotations, so that they only depend on their spectrum. Mathematically, we say:

$$f : \mathscr{H}_{\mathbb{C}}(N) \to \mathbb{C}, \qquad f(\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^\dagger) = f(\boldsymbol{A}), \qquad \text{for all } \boldsymbol{A} \in \mathscr{H}_{\mathbb{C}}(N),\ \boldsymbol{U} \in \mathbb{U}_N. \tag{2.22}$$

In this setting, we may identify $f(\boldsymbol{A}) = f(\boldsymbol{\lambda}^{\boldsymbol{A}})$ without loss of generality. By convention, we order eigenvalues increasingly as $\lambda_1^{\boldsymbol{A}} \leqslant \cdots \leqslant \lambda_N^{\boldsymbol{A}}$ and will therefore see the function with a domain on the *Weyl chamber*:[5]

$$f : \mathbb{R}^N_{\leqslant} \to \mathbb{C}, \qquad \mathbb{R}^N_{\leqslant} := \left\{\boldsymbol{\lambda} \in \mathbb{R}^N \mid \lambda_1 \leqslant \cdots \leqslant \lambda_N\right\}. \tag{2.23}$$

We will also need two formulas we saw in class for the GUE ensemble. Namely, its density is:

$$p_{\mathrm{GUE}}(\boldsymbol{M}_N) = C_N \exp\left\{-\frac{\mathrm{Tr}\left(\boldsymbol{M}_N^2\right)}{2}\right\} \mathrm{d}\mu(\boldsymbol{M}_N), \tag{2.24}$$

where C_N is a normalizing constant. Alternatively, we may express it via the density of the entries, which are independent, obtaining a distribution absolutely continuous with respect to Lebesgue. From this expression, we derived the density of the eigenvalues, which presents the classical repulsion phenomenon:

$$p_{\mathrm{GUE;eigen}}(\boldsymbol{\lambda}) = p_{\mathrm{GUE;eigen}}(\lambda_1, \ldots, \lambda_N) = \frac{1}{(2\pi)^{N/2} c_N}\Delta(\boldsymbol{\lambda})^2 \exp\left\{-\frac{\|\boldsymbol{\lambda}\|_2^2}{2}\right\} \mathrm{d}\boldsymbol{\lambda}. \tag{2.25}$$

Having these in hand, we can start the argument. All the derivations below hold for a "sufficiently nice" function, which we can take to be smooth and exponentially decaying when the arguments are large. First, by an application of Riesz' representation theorem,[6] there is a density function $w : \mathbb{R}^n_{\leqslant} \to \mathbb{R}_+$ such that we can move from integrating over matrices to integrating over eigenvalues, i.e. :

$$\int_{\mathscr{H}_{\mathbb{C}}(N)} f(\boldsymbol{M}_N)\,\mathrm{d}\mu(\boldsymbol{M}_N) = \int_{\mathbb{R}^N_{\leqslant}} f(\boldsymbol{\lambda})w(\boldsymbol{\lambda})\,\mathrm{d}\boldsymbol{\lambda}. \tag{2.26}$$

If instead of a uniform measure we placed the GOE density, we would intuitively obtain the density of eigenvalues of the GOE on the RHS. Thanks to this, we have the following expression:

$$\int_{\mathscr{H}_{\mathbb{C}}(N)} f(\boldsymbol{M}_N) C_N \exp\left\{-\frac{\mathrm{Tr}\left(\boldsymbol{M}_N\right)^2}{2}\right\} \mathrm{d}\mu(\boldsymbol{M}_N) = \int_{\mathbb{R}^N_{\leqslant}} f(\boldsymbol{\lambda})\frac{1}{(2\pi)^{N/2}c_N}\Delta(\boldsymbol{\lambda})^2 \exp\left\{-\frac{\|\boldsymbol{\lambda}\|_2^2}{2}\right\} \mathrm{d}\boldsymbol{\lambda}. \tag{2.27}$$

The exponential terms are the same, and we identify the weight function by regrouping:

$$w(\boldsymbol{\lambda}) = \frac{1}{C_N c_N (2\pi)^{N/2}}\Delta(\boldsymbol{\lambda})^2. \tag{2.28}$$

[3] A simple eigenvalue in this document is distinct from all the others. A simple spectrum then has all distinct eigenvalues.
[4] This is evident at the end of the proof.
[5] By abuse of notation.
[6] See appendix A.I for hints.

In words, up to constants depending on N, the integral of nice functions over the GUE ensemble has expression:

$$\mathbb{E}_{\mathbf{M}_N\sim\mathrm{Haar}(N)}\left[f(\mathbf{M}_N)\right] \propto \int_{\mathbb{R}^N_{\leqslant}} f(\boldsymbol{\lambda})\Delta(\boldsymbol{\lambda})^2\,\mathrm{d}\boldsymbol{\lambda}. \tag{2.29}$$

{{eqn:GUE

What we will do next is use this result to do a double integral. For this purpose, let f, g be "sufficiently nice" and $t > 0$ be a positive number. We wish to compute:[7]

$$\mathcal{I} := \int_{\mathscr{H}_{\mathbb{C}}(N)}\int_{\mathscr{H}_{\mathbb{C}}(N)} f(\boldsymbol{A}_N)g(\boldsymbol{B}_N)\frac{C_N}{t^{N^2/2}}\exp\left\{-\frac{\mathrm{Tr}\left((\boldsymbol{A}_N-\boldsymbol{B}_N)^2\right)}{2t}\right\}\mathrm{d}\mu(\boldsymbol{A}_N)\,\mathrm{d}\mu(\boldsymbol{B}_N). \tag{2.30}$$

{{eqn:inte

We do so in two ways, and by equating them we will get our result.

2.1.1 *Route one*

We see from the trace that we have an interaction term when $\boldsymbol{A}, \boldsymbol{B}$ multiply, so we may redefine the f, g functions as:

$$\tilde{f}(\boldsymbol{A}_N) := f(\boldsymbol{A}_N)\exp\left\{-\frac{\mathrm{Tr}\left(\boldsymbol{A}_N^2\right)}{2t}\right\}; \tag{2.31}$$

$$\tilde{f}(\boldsymbol{B}_N) := f(\boldsymbol{B}_N)\exp\left\{-\frac{\mathrm{Tr}\left(\boldsymbol{B}_N^2\right)}{2t}\right\}. \tag{2.32}$$

Then, we make a decoupling trick: since the GUE is invariant to rotations the interaction term is equal to the interaction term integrated over the Haar measure on the unitary group $\mathbb{U}_N$. In equations, we have that:

$$\begin{aligned}
\exp\left\{\frac{\mathrm{Tr}\left(\boldsymbol{A}_N\boldsymbol{B}_N\right)}{t}\right\} &= \mathbb{E}_{\mathbf{U}_N\sim\mathrm{Haar}(\mathbb{U}_N)}\left[\exp\left\{\frac{\mathrm{Tr}\left(\boldsymbol{A}_N\mathbf{U}_N\boldsymbol{B}_N\mathbf{U}_N^{\dagger}\right)}{t}\right\}\right] && (2.33)\\
&= \int_{\mathbb{U}_N}\exp\left\{\frac{\mathrm{Tr}\left(\boldsymbol{A}_N\boldsymbol{U}_N\boldsymbol{B}_N\boldsymbol{U}_N^{\dagger}\right)}{t}\right\}\mathrm{d}\mu(\boldsymbol{U}_N) && (2.34)\\
&=: \mathcal{K}_t(\boldsymbol{A}_N, \boldsymbol{B}_N). && (2.35)
\end{aligned}$$

Using cyclic invariance of the trace and lemma A.4, the integral $\mathcal{K}_t(\boldsymbol{A}_N, \boldsymbol{B}_N)$ is invariant with respect to rotations in both arguments. Therefore, we have a GUE integration of an invariant function in both variables. Basically, we can use equation 2.29 twice: first in $\mathrm{d}\mu(\boldsymbol{A}_N)$, then in $\mathrm{d}\mu(\boldsymbol{B}_N)$, keeping the constants in front. Unrolling the steps:

$$\begin{aligned}
\mathcal{I} &= \frac{C_N}{t^{N^2/2}}\int_{\mathscr{H}_{\mathbb{C}}(N)}\tilde{g}(\boldsymbol{B}_N)\left[\int_{\mathscr{H}_{\mathbb{C}}(N)}\tilde{f}(\boldsymbol{A}_N)\mathcal{K}_t(\boldsymbol{A}_N,\boldsymbol{B}_N)\,\mathrm{d}\mu(\boldsymbol{A}_N)\right]\mathrm{d}\mu(\boldsymbol{B}_N) && (2.36)\\
&= \frac{C_N}{t^{N^2/2}}\frac{1}{(2\pi)^{N/2}C_Nc_N}\int_{\mathscr{H}_{\mathbb{C}}(N)}\tilde{g}(\boldsymbol{B}_N)\left[\int_{\mathbb{R}^N_{\leqslant}}\tilde{f}(\boldsymbol{\lambda}^{\boldsymbol{A}_N})\mathcal{K}_t(\boldsymbol{\lambda}^{\boldsymbol{A}_N},\boldsymbol{B}_N)\Delta(\boldsymbol{\lambda}^{\boldsymbol{A}_N})^2\,\mathrm{d}\boldsymbol{\lambda}^{\boldsymbol{A}_N}\right]\mathrm{d}\mu(\boldsymbol{B}_N) && (2.37)\\
&= \frac{C_N}{t^{N^2/2}}\frac{1}{(2\pi)^{N/2}C_Nc_N}\frac{1}{(2\pi)^{N/2}C_Nc_N}\int_{\mathbb{R}^N_{\leqslant}}\int_{\mathbb{R}^N_{\leqslant}}\tilde{g}(\boldsymbol{\lambda}^{\boldsymbol{B}_N})\tilde{f}(\boldsymbol{\lambda}^{\boldsymbol{A}_N})\mathcal{K}_t(\boldsymbol{\lambda}^{\boldsymbol{A}_N},\boldsymbol{\lambda}^{\boldsymbol{B}_N})\Delta(\boldsymbol{\lambda}^{\boldsymbol{A}_N})^2\Delta(\boldsymbol{\lambda}^{\boldsymbol{B}_N})^2\,\mathrm{d}\boldsymbol{\lambda}^{\boldsymbol{A}_N}\,\mathrm{d}\boldsymbol{\lambda}^{\boldsymbol{B}_N}. && (2.38)
\end{aligned}$$

Noticing that the eigenvalues are integrated out, we could simplify the notation by writing $\boldsymbol{\lambda}, \boldsymbol{\nu}$. We write down the final result to use it later:

$$\mathcal{I} = \frac{1}{(2\pi)^N C_N c_N^2 t^{N^2/2}}\int_{\mathbb{R}^N_{\leqslant}}\int_{\mathbb{R}^N_{\leqslant}}\tilde{f}(\boldsymbol{\lambda})\tilde{g}(\boldsymbol{\nu})\Delta(\boldsymbol{\lambda})^2\Delta(\boldsymbol{\nu})^2\mathcal{K}_t(\boldsymbol{\lambda},\boldsymbol{\nu})\,\mathrm{d}\boldsymbol{\lambda}\,\mathrm{d}\boldsymbol{\nu}. \tag{2.39}$$

{{eqn:inte

[7] In particular, the equation is the bilinear form of the heat flow in $\mathscr{H}_{\mathbb{C}}(N)$ but we said we will gloss over this aspect.

2.1.2 *Route two*

Let us come back to equation 2.30. We first integrate out in $\mathrm{d}\mu(\boldsymbol{A}_N)$ a seemingly complicated expression:

$$\mathcal{E}_{\boldsymbol{B}_N} := \frac{C_N}{t^{N^2/2}} \int_{\mathscr{H}_{\mathbb{C}}(N)} f(\boldsymbol{A}_N) \exp\left\{ -\frac{\operatorname{Tr}\left((\boldsymbol{A}_N - \boldsymbol{B}_N)^2\right)}{2t} \right\} \mathrm{d}\mu(\boldsymbol{A}_N). \tag{2.40}$$

Notice how $\boldsymbol{B}_N$ is fixed inside. We aim to express this as an expectation over an evaluation of f a function of $\boldsymbol{B}_N$. It suffices to notice that setting $\mathbf{G}_N = 1/\sqrt{t}(\mathbf{A}_N - \mathbf{B}_N)$ for which $\mathbf{A}_N = \mathbf{B}_N + \sqrt{t}\mathbf{G}_N$ we have that:

- the change of variables is a dilation cancelling the $t^{N^2/2}$ term;
- the exponential becomes $\exp\left\{-\operatorname{Tr}(G_N^2)/2\right\}$ since the t cancels out.

Thus, we have a representation of the integral as an expectation over the GUE ensemble for $\mathbf{G}_N$ independent of $\mathbf{B}_N$. Mathematically:

$$\mathcal{E}_{\boldsymbol{B}_N} = \mathbb{E}_{\mathbf{G}_N \sim \mathrm{GUE}(N)}\left[f(\mathbf{B}_N + \sqrt{t}\mathbf{G}_N) \right], \qquad \text{where } \mathbf{G}_N \perp\!\!\!\perp \mathbf{B}_N. \tag{2.41}$$

What comes to the rescue is the Brezin-Hikami-Johansson formula (Brézin and Hikami 1996; Johansson 2001; Tao 2010, 2013), which evaluates the density of the eigenvalues exactly. For our case, we will have that:

$$\mathcal{E}_{\boldsymbol{B}_N} = \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N_{\leqslant}} f(\boldsymbol{\lambda}) \frac{\Delta(\boldsymbol{\lambda})}{\Delta(\boldsymbol{\lambda}^{\boldsymbol{B}_N})} \det\left(\left\{ \exp\left\{ -\frac{(\lambda_i - \lambda_j^{\boldsymbol{B}_N})^2}{2t} \right\} \right\}_{i,j\in[n]} \right) \mathrm{d}\boldsymbol{\lambda}. \tag{2.42}$$

Now that we flattened the integration in $\mathrm{d}\mu(\boldsymbol{A}_N)$, we want to flatten the integration in the other matrix. We have:

$$\mathcal{I} = \int_{\mathscr{H}_{\mathbb{C}}(N)} g(\boldsymbol{B}_N)\mathcal{E}_{\boldsymbol{B}_N}\, \mathrm{d}\mu(\boldsymbol{B}_N). \tag{2.43}$$

Again, this function is invariant to rotations, and we can use equation 2.29 to express it as an integral over the eigenvalues. We have the weight function coming in and g will take an element of $\mathbb{R}^N_{\leqslant}$ as input. After some rearranging, we find:

$$\mathcal{I} = \frac{1}{C_N c_N (2\pi)^N t^{N/2}} \int_{\mathbb{R}^N_{\leqslant}} \int_{\mathbb{R}^N_{\leqslant}} f(\boldsymbol{\lambda}) g(\boldsymbol{\nu}) \Delta(\boldsymbol{\lambda})\Delta(\boldsymbol{\nu}) \det\left(\left\{ \exp\left\{ -\frac{(\lambda_i - \lambda_j^{\boldsymbol{B}_N})^2}{2t} \right\} \right\}_{i,j\in[n]} \right) \mathrm{d}\boldsymbol{\lambda}\, \mathrm{d}\boldsymbol{\nu}. \tag{2.44}$$ {{eqn:inte

Notice that $1/C_N c_N$ appears when we apply equation 2.29 and indeed here we did so only once, but the C_N term at the numerator was included in the expectation to apply the Brezin-Hikami-Johansson formula. What do we get from this? The two representations for the integral return a point-wise expression of the $\mathcal{K}_t(\boldsymbol{A}_N, \boldsymbol{B}_N)$ integral, since the rest is explicit! Reordering the terms in equations 2.39-2.44:

$$\mathcal{K}_t(\boldsymbol{A}_N, \boldsymbol{B}_N) = \int_{\mathbb{U}_N} \exp\left\{ \frac{\operatorname{Tr}\left(\boldsymbol{A}_N \boldsymbol{U}_N \boldsymbol{B}_N \boldsymbol{U}_N^\dagger\right)}{t} \right\} \mathrm{d}\mu(\boldsymbol{U}_N) = c_N t^{(n^2-n)/2} \frac{\det\left(\{\exp\{\lambda_j \nu_i / t\}\}_{i,j\in[n]}\right)}{\Delta(\boldsymbol{\lambda})\Delta(\boldsymbol{\nu})}. \tag{2.45}$$

Notice how the formula differs from the theorem in the dependence with respect to t. This is by purpose. Since t was positive and real, we just need to do analytic continuation to conclude that it holds for all $t \in \mathbb{C}\backslash\{0\}$ as claimed in the theorem, using the fact that $\frac{1}{t}$ is holomorphic in the punctured disk, and it is composed with entire functions.

3 PARTIAL PHASE DIAGRAM OF THE MODEL

{sec:parti

Since we know the integral has a nice form, we return to our objective. Let us report equation 1.9, which is the last one in section 1.II:

$$\mathbb{E}_{\mathbf{J}}\left[\mathcal{Z}^n\right] = \mathbb{E}_{(\mathbf{O},\mathbf{\Lambda})}\left[\mathbb{E}_{\mathbf{M}^{(n)} \sim \mathrm{UnifRep}}\left[\exp\left\{ N\beta/2 \operatorname{Tr}\left(\mathbf{O}\mathbf{\Lambda}\mathbf{O}^\top \mathbf{M}^{(n)} \right) \right\} \right] \right], \qquad \mathrm{M}_{ij}^{(n)} := \sum_{\alpha=1}^{n} \frac{\mathrm{s}_i^{(\alpha)} \mathrm{s}_j^{(\alpha)}}{N}. \tag{3.1}$$

Switching the two integrals, we can apply the low-rank HCIZ formula to the inner expectation, finding:

$$\mathbb{E}_{(\mathbf{O},\mathbf{\Lambda})}\left[\exp\{N\beta/2\}\operatorname{Tr}\left(\mathbf{O}\mathbf{\Lambda}\mathbf{O}^\top\mathbf{M}^{(n)}\right)\right] \approx \exp\left\{N/2\operatorname{Tr}\left(h_{\mathbf{\Lambda}}(\beta\mathbf{M}^{(n)})\right)\right\}. \tag{3.2}$$

A word of caution is needed here, the $h_{\mathbf{\Lambda}}$ function we write should not depend on $\mathbf{\Lambda}$, and indeed it does not: we first integrate with respect to $\mathbf{O}$ using the HCIZ formula at fixed[8] $\mathbf{\Lambda}$, and by the notation we mean that the $h_{\mathbf{\Lambda}}$ function is the antiderivative of the R-transform of $\mathbf{\Lambda}$, which at large size is just the deterministic semi-circular law. Moreover, by the rotational invariance of R-transforms in the limit, this is also the R-transform of the original $\mathbf{J}$ when $N \to \infty$.

To compute the trace inside this expectation, the main idea is that we only care about the *eigenvalues* of $\mathbf{M}^{(n)} \in \mathbb{R}^{N\times N}$. Fortunately, these coincide with the eigenvalues of the *overlap matrix*, defined as:[9]

$$\mathbf{Q}^{(n)} := (\mathrm{Q}^{(n)}_{\alpha,\beta})_{\alpha,\beta\in[n]} = \frac{1}{N}\sum_{i=1}^{N}\left[\mathrm{s}_i^{(\alpha)}\right]_{\alpha\in[n]}\left(\left[\mathrm{s}_i^{(\alpha)}\right]_{\alpha\in[n]}\right)^\top, \qquad \mathrm{Q}^{(n)}_{\alpha,\beta} := \frac{1}{N}\sum_{i=1}^{N}\mathrm{s}_i^{(\alpha)}\mathrm{s}_i^{(\beta)}. \tag{3.3}$$

To decouple the integral, we need to create a hierarchy of clever integrations (see appendix A.IV). The key step is introducing a representation of unity that decouples sums over $i \in [n]$; essentially, it is a rewriting of Dirac's delta integrating over the space of all $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ overlap matrices and enforcing the observed value $\mathbf{Q}^{(n)}$, which we expect to enjoy a law of large numbers. Mathematically:

$$1 = \int_{\mathscr{H}_\mathbb{C}(n)} \frac{N^{\binom{n}{2}}}{2^{3n/2}\pi^{n/2}}\left[\int_{\mathscr{H}_\mathbb{R}(n)} \exp\left\{-N\operatorname{Tr}(\boldsymbol{Q}\boldsymbol{Y}) + \sum_{\alpha,\beta=1}^{n} Y_{\alpha,\beta}\sum_{i=1}^{N}\mathrm{s}_i^{(\alpha)}\mathrm{s}_i^{(\beta)}\right\}\mathrm{d}\boldsymbol{Q}\right]\mathrm{d}\boldsymbol{Y}. \tag{3.4}$$ {{eqn:repr

It requires some time to digest: the immediate observation is that the exponential is one if and only if $\mathbf{Q}^{(n)} = \boldsymbol{Q}$, and for all the other cases, it will be zero. Plugging this inside our integrals, after some computations deferred to appendix A.IV, we recover a nicer form:

$$\mathbb{E}_\mathbf{J}[\mathcal{Z}^n] = C_{N,n}\int_{\mathscr{H}_\mathbb{C}(n)}\left[\int_{\mathscr{H}_\mathbb{R}(n)}\exp\left\{\frac{N}{2}\operatorname{Tr}(h_{\mathbf{\Lambda}}(\beta\boldsymbol{Q})) - N\operatorname{Tr}(\boldsymbol{Q}\boldsymbol{Y}) + N\mathcal{H}(\boldsymbol{Y})\right\}\mathrm{d}\boldsymbol{Q}\right]\mathrm{d}\boldsymbol{Y}, \tag{3.5}$$ {{eqn:inte

where $C_{N,n}$ depends only on (N,n) and:

$$\mathcal{H}(\boldsymbol{Y}) := \log\mathscr{Z}_{\boldsymbol{Y}}, \qquad \mathscr{Z}_{\boldsymbol{Y}} := \sum_{\boldsymbol{s}\in\{\pm1\}^n}\exp\left\{\boldsymbol{s}^\top\boldsymbol{Y}\boldsymbol{s}\right\}. \tag{3.6}$$

As $N \gg 1$ diverges, we can again apply the saddle-point method, twice. Notice that this requires taking $N \to \infty$ before $n \to 0$ which means switching the integrals. The saddle-point in $\mathrm{d}\boldsymbol{Q}$ gives a set of derivatives in $(\alpha,\beta) \in [n]\times[n]$ indices of $\boldsymbol{Q}$ of the form:[10]

$$\boldsymbol{Y}_{\alpha,\beta} = \frac{\beta}{2}\left[R_{\mathbf{\Lambda}}(\beta\boldsymbol{Q})\right]_{\alpha,\beta} = \frac{\beta}{2}R_{\mathbf{\Lambda}}(\beta\mathrm{Q}_{\alpha,\beta}). \tag{3.7}$$ {{eqn:sadd

Similarly, for $\partial Y_{\alpha,\beta}$ we derive the second and third terms in the exponential:

$$\mathrm{Q}_{\alpha,\beta} = \frac{1}{\mathscr{Z}_{\boldsymbol{Y}}}\sum_{\boldsymbol{s}\in\{\pm1\}^n} s^{(\alpha)}s^{(\beta)}e^{\boldsymbol{s}^\top\boldsymbol{Y}\boldsymbol{s}}. \tag{3.8}$$ {{eqn:sadd

The upshot is that the integral is dominated by configurations satisfying equations 3.7-3.8 jointly. We use this to derive a self-consistent equation for $\boldsymbol{Q}$, since we do not care about $\boldsymbol{Y}$ (it is an auxiliary matrix). For given (α,β), we have that:

$$\mathrm{Q}_{\alpha,\beta} = \mathbb{E}_{\mathrm{Boltz}(R_{\mathbf{\Lambda}}(\beta\boldsymbol{Q}))}\left[\mathrm{s}^{(\alpha)}\mathrm{s}^{(\beta)}\right], \tag{3.9}$$ {{eqn:self

where by $\mathrm{Boltz}(\cdot)$ we mean the Boltzmann measure associated to the coupling given by the argument, which in this case is the R-transform of a function of $\boldsymbol{Q}$ itself of the semi-circular law of $\mathbf{\Lambda}$. This is so far still a system in n replicas, i.e. the dimension of $\boldsymbol{Q}$. Solving this self-consistent equation is not easy. Numerical experiments suggest us the following:

> "At β low enough, say below a critical β_c the solutions of the self-consistent equation **seem to have** a simple structure, with a common diagonal."

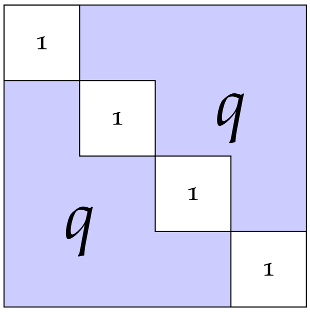

Figure 1: Replica-symmetric ansatz
Here 1 is only in the diagonal: the block is large for aesthetic reasons. {fig: RS

If we trust this observation, we might as well try to plug in this "replica symmetric" solution inside equation 3.9:[11]

$$Q^{(RS)}_{\alpha,\beta} = \mathbb{1}_{\alpha=\beta}(1-q) + q. \tag{3.10}$$ {{eqn:RS a

Such matrix has two eigenvalues:

- one has multiplicity one and is equal to $1 + (n-1)q$;
- another has multiplicity $n-1$ and is $1-q$.

This translates directly into eigenvalues of $R_{\mathbf{\Lambda}}$ (see appendix A.III). After some calculus on $\mathscr{Z}_Y$ (see (Potters and Bouchaud 2020, eqns. 13.92-13.95)) the result is a new equation that relates only q and the off-diagonal entries of the R-transform:

$$r := \left[R_{\mathbf{\Lambda}}(\mathbf{Q}^{(RS)})\right]_{\alpha,\beta} = \frac{1}{n}\left[R_{\mathbf{\Lambda}}(\beta(1+(n-1)q)) - R_{\mathbf{\Lambda}}(\beta(1-q))\right] \qquad \text{where } \alpha \neq \beta; \tag{3.11}$$ {{eqn:r de

$$q = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tanh^2\left\{x\sqrt{\beta r}\right\} dx. \tag{3.12}$$

Now some informal steps need specification:

- to make the R-transform deterministic, we said it needs to be at $N = \infty$, so we have to exchange the n, N order of the limits, and take first $N \to \infty$;
- we believe that the R-transform in the definition of r has nice properties, so that as $n \to 0$ equation 3.11 is a derivative of the R-transform;
- we have to believe in the fact that replicas (i.i.d. copies of random variables), can be a non-integer number, and can even tend to zero.

Let us just make a leap of faith. At $N \to \infty$ and then $n \to 0$, we find a self-consistent equation in q:

$$q = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tanh^2\left\{x\sqrt{\beta^2 q R'_{sc}(\beta(1-q))}\right\} dx, \tag{3.13}$$ {{eqn:self

where we stress again that R_{sc} is not random and is the limit of the R-transform of $\mathbf{\Lambda}$.

While difficult at first sight, this equation is at least easy to simulate or analyze. We have the following basic observation: $q = 0$ is a solution, which should correspond to a high-temperature point in the phase diagram. Taylor expanding around $q = 0$, we can see that at least for the GOE case (and by simulation for others), we find a parabola equation in q where the coefficients change curvature when:

$$\frac{1}{\beta_c} = R'_{sc}(\beta_c). \tag{3.14}$$

[8] Recall that in the GOE these are independent.

[9] A moment of thought shows that we are just transposing the $\mathbf{M}^{(n)}$ matrix seeing the n replicas as elements of $\mathbb{R}^N$ or the N spins of the replicas as elements of $\mathbb{R}^n$. In particular, now it is a sample mean of interactions.

[10] In words: derive the first two terms in the exponential.

[11] Notice how it is also the easiest possible non-trivial matrix!

The interpretation is that $q \neq 0$ when we are above such β_c, and in this case, for two randomly sampled configurations $\mathbf{s}^{(\alpha)}, \mathbf{s}^{(\beta)}$ from the Boltzmann measure of the spin-glass there is a regime in which their overlap is $1/N \left\langle \mathbf{s}^{(\alpha)}, \mathbf{s}^{(\beta)} \right\rangle = q$.

Unfortunately, further analysis tells us that this solution is *unstable* and that symmetry is *broken* (Sherrington and Kirkpatrick 1975). The true solution is an advanced topic that we do not cover here, but is the natural continuation of this line of reasoning. At least for the case in which $\mathbf{J}$ is a GOE matrix we have that $\beta_c = 1$ is the correct value of the phase transition. In retrospect, we guessed it from first principles using random matrix theory.

Another critique is that we did not really compute the free energy: it is true. As a matter of fact, before finding the expression for the free energy, we identified a self-consistent equation in the order parameter q, namely equation 3.13, that we claim is a full descriptor of the statistics of the model, i.e. in physics jargon an *order parameter*. The main reason is that upon knowing the statistics of q, which depend on β, one can fully recover the free energy as a function of β. We do not do it here for the sake of space, but it is just a matter of crunching together the last equations. The interested reader can find the result in the work of Sherrington and Kirkpatrick (1975), with the caveat that the free energy is wrong, as we just said. The true expression of the free energy is derived and explained in a series of letters (Parisi 1979a,b, 1980a,b, 1983) and the works that followed.

Anderson, Greg W., Alice Guionnet, and Ofer Zeitouni (2009). *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press. ISBN: 978-0-521-19452-5. DOI: 10.1017/CBO9780511801334 (cit. on p. 13).

Brézin, E. and S. Hikami (Nov. 1996). "Correlations of Nearby Levels Induced by a Random Potential". In: *Nuclear Physics B* 479.3, pp. 697–706. ISSN: 0550-3213. DOI: 10.1016/0550-3213(96)00394-X (cit. on p. 8).

Harish-Chandra (1957). "Differential Operators on a Semisimple Lie Algebra". In: *American Journal of Mathematics* 79.1, pp. 87–120. ISSN: 0002-9327. DOI: 10.2307/2372387. JSTOR: 2372387 (cit. on p. 1).

Itzykson, C. and J.-B. Zuber (Mar. 1980). "The Planar Approximation. II". In: *Journal of Mathematical Physics* 21.3, pp. 411–421. ISSN: 0022-2488. DOI: 10.1063/1.524438 (cit. on p. 1).

Johansson, Kurt Johansson (Jan. 2001). "Universality of the Local Spacing Distribution¶in Certain Ensembles of Hermitian Wigner Matrices". In: *Communications in Mathematical Physics* 215.3, pp. 683–705. ISSN: 1432-0916. DOI: 10.1007/s002200000328 (cit. on p. 8).

Maillard, Antoine et al. (Aug. 2024). *Bayes-Optimal Learning of an Extensive-Width Neural Network from Quadratically Many Samples*. DOI: 10.48550/arXiv.2408.03733. arXiv: 2408.03733 [stat] (cit. on p. 3).

Marinari, Enzo, Giorgio Parisi, and Felix Ritort (Dec. 1994). "Replica Field Theory for Deterministic Models (II): A Non-Random Spin Glass with Glassy Behavior". In: *Journal of Physics A: Mathematical and General* 27.23, pp. 7647–7668. ISSN: 0305-4470, 1361-6447. DOI: 10.1088/0305-4470/27/23/011. arXiv: cond-mat/9406074 (cit. on p. 3).

McSwiggen, Colin (May 2021). *The Harish-Chandra Integral: An Introduction with Examples*. DOI: 10.48550/arXiv.1806.11155. arXiv: 1806.11155 [math-ph] (cit. on p. 3).

Parisi, Giorgio (Dec. 1979a). "Infinite Number of Order Parameters for Spin-Glasses". In: *Physical Review Letters* 43.23, pp. 1754–1756. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.43.1754 (cit. on p. 11).

— (Sept. 1979b). "Toward a Mean Field Theory for Spin Glasses". In: *Physics Letters A* 73.3, pp. 203–205. ISSN: 03759601. DOI: 10.1016/0375-9601(79)90708-4 (cit. on p. 11).

— (Apr. 1980a). "A Sequence of Approximated Solutions to the S-K Model for Spin Glasses". In: *Journal of Physics A: Mathematical and General* 13.4, pp. L115–L121. ISSN: 0305-4470, 1361-6447. DOI: 10.1088/0305-4470/13/4/009 (cit. on p. 11).

— (May 1980b). "Magnetic Properties of Spin Glasses in a New Mean Field Theory". In: *Journal of Physics A: Mathematical and General* 13.5, pp. 1887–1895. ISSN: 0305-4470, 1361-6447. DOI: 10.1088/0305-4470/13/5/047 (cit. on p. 11).

— (June 1983). "Order Parameter for Spin-Glasses". In: *Physical Review Letters* 50.24, pp. 1946–1948. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.50.1946 (cit. on p. 11).

— (Aug. 2023). *Nobel Lecture: Multiple Equilibria*. DOI: 10.48550/arXiv.2304.00580. arXiv: 2304.00580 [cond-mat] (cit. on p. 2).

Potters, Marc and Jean-Philippe Bouchaud (2020). *A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists*. Cambridge: Cambridge University Press. ISBN: 978-1-108-48808-2. DOI: 10.1017/9781108768900 (cit. on pp. 1, 3, 5, 10, 14, 16).

Pourkamali, Farzad, Jean Barbier, and Nicolas Macris (Nov. 2024). "Matrix Inference in Growing Rank Regimes". In: *IEEE Transactions on Information Theory* 70.11, pp. 8133–8163. ISSN: 1557-9654. DOI: 10.1109/TIT.2024.3422263 (cit. on p. 3).

Sherrington, David and Scott Kirkpatrick (Dec. 1975). "Solvable Model of a Spin-Glass". In: *Physical Review Letters* 35.26, pp. 1792–1796. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.35.1792 (cit. on pp. 1, 11).

Sourav Chatterjee (June 2024). *The Abel Lectures 2024: Michel Talagrand's Works on Spin Glasses* (cit. on p. 2).

Talagrand, Michel (2006). "The Parisi Formula". In: *Annals of Mathematics* 163.1, pp. 221–263. ISSN: 0003-486X. JSTOR: 20159953.

— (2011). *Mean Field Models for Spin Glasses: Volume I: Basic Examples*. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 978-3-642-15201-6 978-3-642-15202-3. DOI: 10.1007/978-3-642-15202-3 (cit. on p. 13).

Tao, Terence (Jan. 2010). *254A, Notes 3b: Brownian Motion and Dyson Brownian Motion* (cit. on pp. 5, 8).

— (Feb. 2013). *The Harish-Chandra-Itzykson-Zuber Integral Formula* (cit. on pp. 1, 3, 5, 8).

Zdeborová, Lenka and Florent Krzakala (Sept. 2016). "Statistical Physics of Inference: Thresholds and Algorithms". In: *Advances in Physics* 65.5, pp. 453–552. ISSN: 0001-8732, 1460-6976. DOI: 10.1080/00018732.2016.1211393. arXiv: 1511.02476 [cond-mat, stat] (cit. on p. 2).

{app:detai

We provide context on theorems and secondary computations divided by topic.

A.1 *Useful lemmas*

{subsec:us

In the main document, we avoided stating some results which are standard in the theory.

QUENCHED FREE ENERGY CONCENTRATION To show concentration of the quenched free energy we just need to show that the log-sum-exp function is Lipschitz, and apply concentration of Lipschitz functions of Gaussian matrices as in (Anderson, Guionnet, and Zeitouni 2009, chap. 2, especially sec. 3). In particular, the fully formal statement we sketch is in (Talagrand 2011, thm. 1.3.4, prop. 1.3.5) and the discussion just after. Let us rewrite the partition function:

$$\mathcal{Z}(\mathbf{J}, N; \beta) = \int_{\{\pm 1\}^N} \mathrm{e}^{\beta \boldsymbol{s}^\top \mathbf{J} \boldsymbol{s}} \,\mathrm{d}\boldsymbol{s} = \sum_{\boldsymbol{s} \in \{\pm 1\}^N} \mathrm{e}^{\beta \boldsymbol{s}^\top \mathbf{J} \boldsymbol{s}}. \tag{A.1}$$

Seen as a random variable in $\mathbf{J} \in \mathbb{R}^{N \times N}$ it is such that for any matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ the derivative with respect to one entry is:

$$\partial_{ij} \frac{1}{\beta} \log \mathcal{Z}(\boldsymbol{A}, N; \beta) = \frac{1}{\beta} \frac{\partial_{ij} \mathcal{Z}(\boldsymbol{A}, N; \beta)}{\mathcal{Z}(\boldsymbol{A}, N; \beta)} = \frac{1}{\beta} \frac{1}{\mathcal{Z}(\boldsymbol{A}, N; \beta)} \sum_{\boldsymbol{s} \in \{\pm 1\}^N \beta} s_i s_j \mathrm{e}^{\beta \boldsymbol{s}^\top \boldsymbol{A} \boldsymbol{s}} = \mathbb{E}_{\boldsymbol{s} \sim \mathrm{Boltz}(\boldsymbol{A})} \left[\boldsymbol{s}_i \boldsymbol{s}_j \right] \leqslant 1, \tag{A.2}$$

where by $\mathrm{Boltz}(\boldsymbol{A})$ we mean the Boltzmann measure with couplings $\boldsymbol{A}$ and we simply bounded the product of two spins by one, since they are in the hypercube. By the fact that the derivative is bounded, the function is 1-Lipschitz in all entries of $\boldsymbol{A}$, and we may apply the Gaussian concentration inequalities for Lipschitz functions. The correction is $1/\sqrt{N}$ by the fact that in our scaling the GOE has $1/\sqrt{N}$ standard deviation at each entry. Therefore, rearranging terms, we have $\sim N^2$ random variables but the Lipschitz order gives us $1/\sqrt{N}$ exponential concentration.

RIESZ' REPRESENTATION THEOREM To apply Riesz' representation theorem on unitarily invariant functions we use the fact that the functions we consider are also assumed to be fast decaying. The simple case would be to add that they map to the positive real line, and by an approximation argument (e.g. Weierstrass), restrict to polynomials over a compact interval. We sketch the fully general statement without these restrictions, leaving some details as exercises. The space $\mathbb{R}^N_\leqslant$ is a locally compact Hausdorff space (exercise). Then, it suffices to show that $\psi(f) := \int_{\mathscr{H}_\mathbb{C}(N)} f(\boldsymbol{M}_N) \,\mathrm{d}\boldsymbol{M}_N$ is a linear (exercise) functional for $f : \mathbb{R}^N_\leqslant \to \mathbb{C}$ by the invariance with respect to unitary rotations. Moreover, such $\psi(\cdot)$ are continuous in the space $C_0(\mathbb{R}^N_\leqslant)$, the space of continuous linear functionals on $\mathbb{R}^N_\leqslant$ that vanish at infinity (exercise). Then there is a unique real valued regular Borel measure μ on $\mathbb{R}^N_\leqslant$ such that:

$$\psi(f) = \int_{\mathbb{R}^N_\leqslant} f(\boldsymbol{\lambda}) \,\mathrm{d}\mu(\boldsymbol{\lambda}), \qquad \forall f \in C_0(\mathbb{R}^N_\leqslant), \tag{A.3}$$

as a consequence of Riesz' representation theorem. All the functions we are interested in fall under $C_0(\mathbb{R}^N_\leqslant)$, as we just take them to be sufficiently nice. We leave to the reader the exercise to prove that the measure μ is σ-finite, so that by the Radon-Nykodim theorem we may write it as $w(\boldsymbol{\lambda}) \,\mathrm{d}\boldsymbol{\lambda} = \mathrm{d}\mu(\boldsymbol{\lambda})$ for $\mathrm{d}\boldsymbol{\lambda}$ the Lebesgue measure on $\mathbb{R}^N_\leqslant$, which exists again by an exercise.

Lemma A.4. *Let* $\mathbf{O}, \mathbf{R} \sim \mathrm{Haar}(\mathbb{O}_N)$ *be independent. Then,* $\mathbf{P} = \mathbf{OR} \sim \mathrm{Haar}(\mathbb{O}_N)$ *and* $\mathbf{O}, \mathbf{R}, \mathbf{P}$ *are independent. The same holds for unitary matrices.* {lem:rotat

Proof. Suppose $\mathbf{O}, \mathbf{R} \sim \mathrm{Haar}(\mathbb{O}_N)$. Their product $\mathbf{P} = \mathbf{OR}$ is orthogonal. Moreover, the Haar (probability) measure is the unique left-right invariant measure over $\mathbb{O}_N$. It follows that $\mathbf{OR} \sim \mathrm{Haar}(\mathbb{O}_N)$. For independence, consider two continuous bounded function $f, g : \mathbb{O}_N \to \mathbb{R}$ we inspect the integral:

$$\mathbb{E}_{(\mathbf{P},\mathbf{R})} \left[f(\mathbf{P}) g(\mathbf{R}) \right] = \int f(\mathbf{P}) g(\mathbf{R}) \mathrm{d}\mu_{(\mathbf{P},\mathbf{R})}. \tag{A.5}$$

Let us disintegrate μ into its conditional:

$$\mathbb{E}_{(\mathbf{P},\mathbf{R})}\left[f(\mathbf{P})g(\mathbf{R})\right] = \int\int f(\mathbf{P})\mathrm{d}\mu_{\mathbf{P}|\mathbf{R}}g(\mathbf{R})\mathrm{d}\mu_{\mathbf{R}}. \tag{A.6}$$

The conditional $\mathbf{P} \mid \mathbf{R}$ is Haar distributed since $\mathbf{P}$ is Haar and for fixed $\mathbf{R}$ it is again Haar as $\mathbf{R}$ is itself a rotation. This means that independently of $\mathbf{R}$ the inner integral is a Haar Integral over the unique (probability) measure. Having this, we may compute the inner expectation, take it out, and compute the outer expectation left. The product decouples, making the random variables independent. □

A.II *Saddle-point method for low-rank HCIZ integrals*

{subsec:sa

For the sake of completeness, we just report the reasoning (see (Potters and Bouchaud 2020, eqns. 10.36-10.39)). Recall the definition of $f_t(z, \boldsymbol{A})$ in equation 2.11 and the saddle-point condition in equation 2.12. We compute $I_t(\boldsymbol{A})$ and $I_t(\mathbf{0}_{N\times N})$ by deforming the contour. To begin, we notice that since $\partial_z f_t(z, \boldsymbol{A}) = t - s_N^{\boldsymbol{A}}(z)$, then:

$$\partial_z^2 f_t(z, \boldsymbol{A}) = \partial_z s_N^{\boldsymbol{A}}(z). \tag{A.7}$$

First we compute the easy one. By definition, the Stieltjes transform of the null matrix is $g_{\mathbf{0}_{N\times N}}(z) = 1/z$, so its inverse is $\zeta(t) = 1/t$. The optimal point is explicit in this case, and satisfies:

$$\begin{aligned}
f_t(\zeta(t), \mathbf{0}_{N\times N}) &= \zeta(t)t - \frac{1}{N}\sum_k \log(z - \lambda_k^{\mathbf{A}}) && \text{(A.8)}\\
&= 1 - \log(1/t) && \text{(A.9)}\\
&= 1 + \log t && \text{(A.10)}\\
\partial_z^2 f_t(\zeta(t), \mathbf{0}_{N\times N}) &= \partial_z s_N^{\mathbf{A}}(z) && \text{(A.11)}\\
&= \partial_z \frac{1}{\zeta(t)} && \text{(A.12)}\\
&= t^2. && \text{(A.13)}
\end{aligned}$$

The saddle-point approximation tells us that for a proper contour integral as in this case:

$$\begin{aligned}
I_t(\mathbf{0}_{N\times N}) &= \frac{1}{4\pi}\int_{\lambda-\mathrm{i}\infty}^{\lambda+\mathrm{i}\infty} \exp\left\{\frac{N}{2}f_t(z, \mathbf{0}_{N\times N})\right\} && \text{(A.14)}\\
&\overset{N\to\infty}{\sim} \frac{1}{4\pi}\frac{\sqrt{4\pi}}{\sqrt{|N\,\partial_z^2\left[f(\zeta(t), \mathbf{0}_{N\times N})\right]|}} \exp\left\{\frac{N}{2}f_t(\zeta(t), \mathbf{0}_{N\times N})\right\} && \text{(A.15)}\\
&= \frac{1}{2t\sqrt{N\pi}} \exp\left\{\frac{N}{2}(1 + \log t)\right\}. && \text{(A.16)}
\end{aligned}$$

For the numerator term $I_t(\mathbf{A})$, we need more work. The Stieltjes transform is invertible since above $\lambda_{\max}^{\boldsymbol{A}}$ it is monotonic, so the existance of $\zeta(t)$ the inverse of $s_n^{\boldsymbol{A}}(z)$ is immediate from some analysis (Potters and Bouchaud 2020, sec. 10.4). Moreover, we find in this regime of λ that for $t < s_N^{\boldsymbol{A}}(\lambda_{\max}^{\boldsymbol{A}})$ the inverse satisfies $\zeta(t) > \lambda_{\max}$. Therefore, the integration in $(\lambda - \mathrm{i}\infty, \lambda + \mathrm{i}\infty)$ can deform to touch $\zeta(t)$, since $f_t(z, \boldsymbol{A})$ is analytic above $\lambda_{\max}^{\boldsymbol{A}}$. The $\zeta(t)$ point will dominate the integral. The auxiliary terms $f_t(\zeta(t), \mathbf{A})$ and $\partial_z^2 f_t(\zeta(t), \mathbf{A})$ are now implicit; we just use the fact that the second partial derivative gives the Stieltjes transform. The saddle-point approximation tells us that the integral is dominated by the exponential evaluated at its maximum rescaled by fluctuations. Reordering terms:

$$I_t(\mathbf{A}) \overset{N\to\infty}{\sim} \frac{1}{2\sqrt{N\pi|\,\partial_z s_N^{\mathbf{A}}(\zeta(t))|}} \exp\left\{\frac{N}{2}f_t(\zeta(t), \mathbf{A})\right\}. \tag{A.17}$$

Originally, we wanted to give an expression for equation 2.4, which is:

$$\begin{aligned}
h_{\mathbf{A}}(t) &= \lim_{N\to\infty}\frac{2}{N}\log\left(\frac{I_t(A)}{I_t(\mathbf{0}_{N\times N})}\right) && \text{(A.18)}\\
&= \lim_{N\to\infty}\frac{2}{N}\left\{\frac{N}{2}\left[f_t(\zeta(t), \mathbf{A}) - f_t(\zeta(t), \mathbf{0}_{N\times N})\right] + o\,(N)\right\} && \text{(A.19)}\\
&= f_t(\zeta(t), \mathbf{A}) - 1 - \log t., && \text{(A.20)}
\end{aligned}$$

where in the $o(N)$ terms we put the denominators in front of the exponentials. Using equation 2.11 for $f_t(\zeta(t), \mathbf{A})$ gives the claimed equation 2.13.

A.III *Rotating the R-transform*

The R-transform of the matrix $\boldsymbol{Q}^{(\mathrm{RS})}$ is simple. Let us note that we used the notation $[R_{\boldsymbol{\Lambda}}(\boldsymbol{Q})]_{\alpha,\beta}$ to mean the function $R_{\boldsymbol{\Lambda}}(Q_{\alpha,\beta})$. Then, we have to compute only two values, namely $R_{\boldsymbol{\Lambda}}(1)$ and $R_{\boldsymbol{\Lambda}}(q) \not\equiv r$ in the notation we used in equation 3.11 for reasons that will be clear at the end of the paragraph. From the structure imposed in equation 3.10 we said that $\boldsymbol{Q}^{(\mathrm{RS})}$ has two eigenvalues: this is immediate and is generalized in the following discussion. {subsec:ro

The authors claim that the matrix $R_{\boldsymbol{\Lambda}}(\boldsymbol{Q})$ has two eigenvalues, with a specific expression. Indeed, if we shorthand R for the R-transform, we have:

$$R_{\boldsymbol{\Lambda}}(\boldsymbol{Q}) = \begin{bmatrix} R(1) & R(q) & R(q) & \cdots & R(q) \\ R(q) & R(1) & R(q) & \cdots & R(q) \\ R(q) & R(q) & R(1) & \cdots & R(q) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(q) & R(q) & R(q) & \cdots & R(1) \end{bmatrix}, \tag{A.21}$$ {{eqn:prop

and imposing the characteristic equation $R_{\boldsymbol{\Lambda}}\boldsymbol{v} = \varrho \boldsymbol{v}$ we find the condition:

$$R(1)v_i + R(q)\sum_{j\neq i} v_j = \varrho v_i, \qquad \forall i \in [n]. \tag{A.22}$$

Some attempts give us the two eigenvalues:

- there is a unique solution $v_i \equiv 1$ associated to $\varrho = R(1) + (n-1)R(q)$;
- there is an $n-1$ dimensional space spanned by $\sum_{i=1}^{n} v_i = 0$ associated to $\varrho = R(1) - R(q)$.

Now the main observation is that the R-transform is invariant to rotations (assuming we took $N \to \infty$ before $n \to 0$). Therefore, we might as well orient in the diagonalization of $\boldsymbol{Q}$ and obtain that the R-transform has two eigenvalues:

- $R_{\boldsymbol{\Lambda}}(1 + (n-1)q)$ for the simple one;
- $R_{\boldsymbol{\Lambda}}(1 - q)$ for the degenerate one.

It is then a matter of taste (and cleverness), to choose a matrix of the form:

$$R_{\boldsymbol{\Lambda}}(\boldsymbol{Q}) \overset{\text{rotation}}{\rightsquigarrow} = \begin{cases} \frac{1}{n}\left(R_{\boldsymbol{\Lambda}}(\beta(1+(n-1)q)) - R_{\boldsymbol{\Lambda}}(\beta(1-q))\right) & \text{if } \alpha \neq \beta \\ R_{\boldsymbol{\Lambda}}(\beta(1-q)) + \frac{1}{n}\left(R_{\boldsymbol{\Lambda}}(\beta(1+(n-1)q)) - R_{\boldsymbol{\Lambda}}(\beta(1-q))\right) & \text{if } \alpha = \beta \end{cases}; \tag{A.23}$$

where now we stress that we have recovered in the first line equation 3.11. A little calculus gives that this candidate has the right eigenvalues, and is a rotation of the first one proposed in equation A.21. The importance of this expression is that the r term is a finite difference equation that as $n \to 0$ becomes a derivative! Magically everything aligns. Notice also that we will use this construction as a plug in for $\mathbf{Y}$ in our expression of $\mathscr{Z}_{\mathbf{Y}}$, which is rotationally invariant, and for the computation of q which is again an overlap invariant to rotations. Basically we are at a stage of the problem where we may take the best representation.

A.IV *Clever integrations with delta functions to decouple replicas*

{subsec:de

We want to compute:

$$\mathcal{I} := \mathbb{E}_{\mathbf{M}^{(n)}\sim\text{UnifRep}}\left[\exp\left\{{}^{N}\!/\!{}_{2}\mathrm{Tr}\left(h_{\boldsymbol{\Lambda}}(\beta\mathbf{M}^{(n)})\right)\right\}\right] = \sum_{\boldsymbol{s}^{(\alpha)}\in\{\pm1\}^N,\alpha\in[n]} \exp\left\{{}^{N}\!/\!{}_{2}\mathrm{Tr}\left(h_{\boldsymbol{\Lambda}}(\beta\boldsymbol{M}^{(n)})\right)\right\}, \tag{A.24}$$

where we will use the "flipped" $\mathbf{M}^{(n)}$ into $\mathbf{Q}^{(n)}$. The main issue is that the $\mathbf{M}^{(n)}$ matrix is a collection of replicas $\{\mathbf{s}^{(\alpha)}\}_{\alpha\in[n]}$ where each $\mathbf{s}^{(\alpha)} \in \mathbb{R}^N$. This is clearly highly mixed with many interactions between the random variables. The key point is that summing over $\{s^{(\alpha)}\}_{\alpha\in[n]}$ and over $\{s_i\}_{i\in[N]}$ where $s^{(\alpha)} \in \{\pm 1\}^N, s_i \in \{\pm 1\}^n$ is the same: we just traverse spins or replicas. In the next equations, we will use them interchangeably. Following (Potters and Bouchaud 2020, sec. 13.2.2), which takes inspiration from the classical treatment of this problem, we let $\mathbf{Q}^{(n)}$ be a variable to integrate out with a Dirac delta representation that forces it to be $\mathbf{Q}^{(n)}$ for given spins. This is the idea of equation 3.4, which we report here for convenience:

$$1 = \int_{\mathscr{H}_n(\mathbb{C})} \frac{N^{\binom{n}{2}}}{2^{3n/2}\pi^{n/2}} \left[\int_{\mathscr{H}_{\mathbb{R}}(n)} \exp\left\{ -N\mathrm{Tr}\left(\boldsymbol{Q}\boldsymbol{Y}\right) + \sum_{i=1}^{N} \boldsymbol{s}_i^\top \boldsymbol{Y} \boldsymbol{s}_i \right\} \mathrm{d}\boldsymbol{Q} \right] \mathrm{d}\boldsymbol{Y}, \qquad \forall \{\boldsymbol{s}_i\}_{i=1}^N,\ \boldsymbol{s}_i \in \mathbb{R}^n. \tag{A.25}$$

In particular, differently from (Potters and Bouchaud 2020, eqn. 13.33) we lose the $1/2$ factor because the models are slightly different, and we see $\boldsymbol{s}_i = (\boldsymbol{s}_i^{(1)}, \dots, \boldsymbol{s}_i^{(n)}) \in \mathbb{R}^n$ as a collection of the i^{th} spin for each replica $\alpha \in [n]$, so logically $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{Y} \in \mathbb{C}^{n\times n}$ because it uses the integral representation of Dirac deltas. We take them to be symmetric, so they have $n(n+1)/2$ integrands hidden in the differentials of matrices. To give more details let us fix a given collection of spins $\{\boldsymbol{s}_i\}_{i\in[N]}$, so that the matrix $\boldsymbol{Q}^{(n)} \in \mathbb{R}^{n\times n}$ is fixed. We introduce for every $(\alpha, \beta) \in [n] \times [n]$ a constraint with the following flavour. If we had to constrain only one entry then we would:

- force the scalar with a Dirac delta measure

$$1 = \int_{\mathbb{R}} \delta(Q_{\alpha,\beta} - Q_{\alpha,\beta}^{(n)})\,\mathrm{d}Q; \tag{A.26}$$

- use the complex representation of Dirac delta functions, to find:

$$1 = C \int_{\mathbb{R}} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \exp\left\{ -\frac{1}{2} y (Q_{\alpha,\beta} - Q_{\alpha,\beta}^{(n)}) \right\} \mathrm{d}y\,\mathrm{d}Q_{\alpha,\beta}, \tag{A.27}$$

  where C is a constant.

However, we do not have a single entry $Q_{\alpha,\beta}^{(n)}$ to enforce, but rather the whole symmetric matrix $\boldsymbol{Q}^{(n)}$, which has $n(n+1)/2$ terms. Therefore, we will integrate over the space of real symmetric matrices enforcing a single point measure:

$$1 = \int_{\mathscr{H}_{\mathbb{R}}(n)} \delta(\boldsymbol{Q}^{(n)} - \boldsymbol{Q})\,\mathrm{d}\boldsymbol{Q}, \qquad \text{as an } \frac{n(n+1)}{2} \text{ fold integral,} \tag{A.28}$$

and then use the complex representation of the Dirac delta once again, to find that the $\mathrm{d}y$ auxiliary integration is over complex Hermitian matrices, and we can write:

$$1 = C_{N,n} \int_{\mathscr{H}_{\mathbb{C}}(n)} \int_{\mathscr{H}_{\mathbb{R}}(n)} \prod_{\alpha,\beta=1}^{n} \exp\left\{ Y_{\alpha,\beta}(Q_{\alpha,\beta} - Q_{\alpha,\beta}^{(n)}) \right\}. \tag{A.29}$$

To find back equation 3.4, it suffices to collect the product of exponentials, express $Q_{\alpha,\beta}^{(n)} = \frac{1}{N}\left\langle \boldsymbol{s}^{\alpha}, \boldsymbol{s}^{(\beta)} \right\rangle$ since we fixed the spins, and do some rescaling.

Using this trick, we carefully compute the integrals:

$$\begin{aligned}
\mathcal{I} &= \sum_{s^{(\alpha)}\in\{\pm1\}^N, \alpha\in[n]} \exp\left\{ N/2\,\mathrm{Tr}\left(h_{\boldsymbol{\Lambda}}(\beta \boldsymbol{M}^{(n)}) \right) \right\}, && \text{(A.30)}\\
&= \sum_{s^{(\alpha)}\in\{\pm1\}^N, \alpha\in[n]} \exp\left\{ N/2\,\mathrm{Tr}\left(h_{\boldsymbol{\Lambda}}(\beta \boldsymbol{Q}^{(n)}) \right) \right\}, && \text{(A.31)}\\
&= \sum_{s^{(\alpha)}\in\{\pm1\}^N, \alpha\in[n]} \exp\left\{ N/2\,\mathrm{Tr}\left(h_{\boldsymbol{\Lambda}}(\beta \boldsymbol{Q}^{(n)}) \right) \right\} \left[\int_{\mathscr{H}_{\mathbb{C}}(n)} \frac{N^{\binom{n}{2}}}{2^{3n/2}\pi^{n/2}} \left[\int_{\mathscr{H}_{\mathbb{R}}(n)} \exp\left\{ -N\mathrm{Tr}\left(\boldsymbol{Q}\boldsymbol{Y}\right) + \sum_{i=1}^{N} \boldsymbol{s}_i^\top \boldsymbol{Y} \boldsymbol{s}_i \right\} \mathrm{d}\boldsymbol{Q} \right] \mathrm{d}\boldsymbol{Y} \right]. && \text{(A.32)}
\end{aligned}$$

Now we take the scaling factors out into a constant $C_{N,n}$ and push the sum over all spins inside the integral. It remains to realize that in the integration over $\boldsymbol{Q}$ we will eventually hit $\boldsymbol{Q}^{(n)}$ only once at fixed $\{s^{(\alpha)}\}_{\alpha\in[n]}$, justifying the idea that we may integrate over all of the $\boldsymbol{Q}$ also the exponential of $h_{\boldsymbol{\Lambda}}$, provided that we sum the $\{\boldsymbol{s}_i\}_{i\in[n]}$ dependent term over all possible configurations. Let us do it for one single set of configurations to be clearer. We fix $\{\boldsymbol{s}_i\}_{i\in[N]}$ where $\boldsymbol{s}_i \in \{\pm 1\}^n$ and find that (the integral, the sum and constant factors are at the left of this expression):

$$\exp\left\{{}^{N}\!/\!{}_{2}\mathrm{Tr}\left(h_{\boldsymbol{\Lambda}}(\beta \boldsymbol{Q}^{(n)})\right)\right\}\left[\int \exp\left\{-N\mathrm{Tr}\left(\boldsymbol{Q}\boldsymbol{Y}\right)+\sum_{i=1}^{N}\boldsymbol{s}_i^\top \boldsymbol{Y}\boldsymbol{s}_i\right\}\,\mathrm{d}\boldsymbol{Q}\right] \tag{A.33}$$

$$=\left[\int \exp\left\{{}^{N}\!/\!{}_{2}\mathrm{Tr}\left(h_{\boldsymbol{\Lambda}}(\beta \boldsymbol{Q}^{(n)})\right)\right\}\exp\left\{-N\mathrm{Tr}\left(\boldsymbol{Q}\boldsymbol{Y}\right)\right\}\,\mathrm{d}\boldsymbol{Q}\right]\exp\left\{\sum_{i=1}^{N}\boldsymbol{s}_i^\top \boldsymbol{Y}\boldsymbol{s}_i\right\} \tag{A.34}$$

$$=\left[\int \exp\left\{{}^{N}\!/\!{}_{2}\mathrm{Tr}\left(h_{\boldsymbol{\Lambda}}(\beta \boldsymbol{Q})\right)\right\}\exp\left\{-N\mathrm{Tr}\left(\boldsymbol{Q}\boldsymbol{Y}\right)\right\}\,\mathrm{d}\boldsymbol{Q}\right]\exp\left\{\sum_{i=1}^{N}\boldsymbol{s}_i^\top \boldsymbol{Y}\boldsymbol{s}_i\right\} \tag{A.35}$$

$$=:\mathcal{I}_{\boldsymbol{Y},\{\boldsymbol{s}_i\}_{i\in[N]}}, \tag{A.36}$$

where in the last step we used the fact that for given $\{\boldsymbol{s}_i\}_{i\in[N]}$ the representation of one we introduced will realize uniquely at $\boldsymbol{Q}^{(n)}$. In simple words, we integrate over all of $\mathrm{d}\boldsymbol{Q}$ an indicator of $\boldsymbol{Q}^{(n)}$. We only need to integrate over $\mathrm{d}\boldsymbol{Y}$ and the n hypercubes of spin replicas. We can exchange the integrals because the terms are positive for each summand and integrand, so:

$$\mathcal{I}=C_{N,n}\sum_{\boldsymbol{s}^{(\alpha)}\in\{\pm1\}^N,\alpha\in[n]}\left[\int_{\mathscr{H}_{\mathbb{C}}(n)}\mathcal{I}_{\boldsymbol{Y},\{\boldsymbol{s}_i\}_{i\in[N]}}\,\mathrm{d}\boldsymbol{Y}\right]=C_{N,n}\int_{\mathscr{H}_{\mathbb{R}}(n)}\left[\sum_{\{\boldsymbol{s}^{(\alpha)}\}_{\alpha\in[n]}}\mathcal{I}_{\boldsymbol{Y},\{\boldsymbol{s}_i\}_{i\in[N]}}\right]\mathrm{d}\boldsymbol{Y}. \tag{A.37}$$

A closer look tells us that we are summing *the same quantity*, which is the integral in $\boldsymbol{Q}$ over given exponential weights $\exp\left\{\boldsymbol{s}_i^\top \boldsymbol{Y}\boldsymbol{s}\right\}$, where $\boldsymbol{Y}$ is fixed. The magic is that the sum is now decoupled, in the sense that the $\{\boldsymbol{s}_i\}_{i\in[N]} \equiv \{\boldsymbol{s}^{(\alpha)}\}_{\alpha\in[n]}$, which were once highly interacting in the $h_{\boldsymbol{\Lambda}}$ function, are now independent (or decoupled in physics jargon). Indeed, regarding the integral term as a constant for given $\boldsymbol{Y}$ inside the $\mathrm{d}\boldsymbol{Y}$ integral, we perform a three line manipulation, which we explicit to six for clarity:

$$\sum_{\{\boldsymbol{s}^{(\alpha)}\}_{\alpha\in[n]}}\exp\left\{\sum_{i=1}^{N}\boldsymbol{s}_i^\top \boldsymbol{Y}\boldsymbol{s}_i\right\}=\sum_{\{\boldsymbol{s}_i\}_{i\in[N]}}\exp\left\{\sum_{i=1}^{N}\boldsymbol{s}_i^\top \boldsymbol{Y}\boldsymbol{s}_i\right\} \tag{A.38}$$

$$=\sum_{\{\boldsymbol{s}_i\}_{i\in[N]}}\prod_{i=1}^{N}\exp\left\{\boldsymbol{s}_i^\top \boldsymbol{Y}\boldsymbol{s}_i\right\} \tag{A.39}$$

$$=\prod_{i=1}^{N}\sum_{\boldsymbol{s}\in\{\pm1\}^n}\exp\left\{\boldsymbol{s}^\top \boldsymbol{Y}\boldsymbol{s}\right\} \tag{A.40}$$

$$=\left(\sum_{\boldsymbol{s}\in\{\pm1\}^n}\exp\left\{\boldsymbol{s}^\top \boldsymbol{Y}\boldsymbol{s}\right\}\right) \tag{A.41}$$

$$=\exp\left\{N\log\sum_{\boldsymbol{s}\in\{\pm1\}^n}\exp\left\{\boldsymbol{s}^\top \boldsymbol{Y}\boldsymbol{s}\right\}\right\} \tag{A.42}$$

$$=\exp\left\{N\log\mathscr{Z}_{\boldsymbol{Y}}\right\} \tag{A.43}$$

where we just used the fact that the exponential of a sum is a product of exponentials, and that these sums depend on different terms. In the meantime, we changed $\boldsymbol{s}_i$ to $\boldsymbol{s}$ is just because the index $i \in [N]$ is a dummy index, and we put an N-fold product. The rest is an exp-log rearrangement. Putting it all together, we found that:

$$\mathcal{I}=C_{N,n}\int_{\mathscr{H}_{\mathbb{C}}(n)}\left[\int_{\mathscr{H}_{\mathbb{R}}(n)}\exp\left\{{}^{N}\!/\!{}_{2}\mathrm{Tr}\left(h_{\boldsymbol{\Lambda}}(\beta\boldsymbol{Q})\right)-N\mathrm{Tr}\left(\boldsymbol{Q}\boldsymbol{Y}\right)+N\log(\mathscr{Z}_{\boldsymbol{Y}})\right\}\,\mathrm{d}\boldsymbol{Q}\right]\mathrm{d}\boldsymbol{Y}, \tag{A.44}$$

which coincides with equation 3.5, upon computing the right constant.