13

The Replica Method*

In this chapter we will review another important tool to perform compact computations in random systems and in particular in random matrix theory, namely the "replica method". For example, one can use replicas to understand the R-transform addition rule when one adds two large, randomly rotated matrices.

Suppose that we want to compute the free energy of a large random system. The free energy is the logarithm of some partition function Z.[1] We expect that the free energy does not depend on the particular sample so we can average the free energy with respect to the randomness in the system to get the typical free energy of a given sample. Unfortunately, averaging the logarithm of a partition function is hard. What we can do more easily is to compute the partition function to some power n and later let $n \to 0$ using the "replica trick":

$$\log Z = \lim_{n \to 0} \frac{Z^n - 1}{n}. \tag{13.1}$$

The partition function Z^n is just the partition function of n non-interacting copies of the same system Z, these copies are called "replicas", hence the name of the technique. Averaging the logarithm is then equivalent to averaging Z^n and taking the limit $n \to 0$ as above. The averaging procedure will however couple the n copies and the resulting interacting system is in general hard to solve. In many interesting cases, the partition function can only be computed as the size of the system (say N) goes to infinity. Naturally one is tempted to interchange the limits ($n \to 0$ and $N \to \infty$) but there is no mathematical justification (yet) for doing so. Another problem is that we can hope to compute $\mathbb{E}[Z^n]$ for all integers n but is that really sufficient to do a proper $n \to 0$ limit?

For all these reasons, replica computations are not considered rigorous. Nevertheless, they are a precious source of intuition and they allow one to obtain results mathematicians would call conjectures, but that often turn out to be mathematically exact. Although a lot of progress has been made to understand why the replica trick works, there is still a halo of mystery and magic surrounding the method, and a nagging impression that an equivalent but more transparent formalism awaits revelation.

[1] See e.g. Section 13.4 for an explicit example.

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

In the present chapter, we will show how all the results obtained up to now can be rederived using replicas. We start by showing how the Stieltjes transform can be expressed using the replica method, and obtain once more the semi-circle law for Gaussian random matrices. We then discuss R-transforms and S-transforms in the language of replicas.

13.1 Stieltjes Transform

As we have shown in Chapter 2, the density of eigenvalues $\rho(\lambda)$ of a random matrix is encoded in the trace of the resolvent of that matrix, which defines the Stieltjes transform of $\rho(\lambda)$. Here we show how this quantity can be computed using the replica formalism.

13.1.1 General Set-Up

To use the replica trick in random matrix theory, we first need to express the Stieltjes transform as the average logarithm of a (possibly random) determinant. In the large N limit and for z sufficiently far from the real eigenvalues, the discrete Stieltjes transform $g_N(z)$ converges to a deterministic function $\mathfrak{g}(z)$. The replica trick will allow us to compute $\mathbb{E}[g_N(z)]$, which also converges to $\mathfrak{g}(z)$.

Using the definition Eq. (2.19) and dropping the N subscript, we have

$$\mathbb{E}[g_{\mathbf{A}}(z)] = \frac{1}{N}\mathbb{E}\left[\sum_{k=1}^{N}\frac{1}{z-\lambda_k}\right], \tag{13.2}$$

whereas the determinant of $z\mathbf{1}-\mathbf{A}$ is given by

$$\det(z\mathbf{1}-\mathbf{A}) = \prod_{k=1}^{N}(z-\lambda_k). \tag{13.3}$$

We can turn the product in the determinant into a sum by taking the logarithm and obtain $(z-\lambda_k)^{-1}$ from $\log(z-\lambda_k)$ by taking a derivative with respect to z. We then get

$$\mathbb{E}[g_{\mathbf{A}}(z)] = \frac{1}{N}\mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}z}\log\det(z\mathbf{1}-\mathbf{A})\right]. \tag{13.4}$$

To compute the determinant we may use the multivariate Gaussian identity

$$\int\frac{\mathrm{d}^N\boldsymbol{\psi}}{(2\pi)^{N/2}}\exp\left(-\frac{\boldsymbol{\psi}^T\mathbf{M}\boldsymbol{\psi}}{2}\right) = \frac{1}{\sqrt{\det\mathbf{M}}}, \tag{13.5}$$

which is exact for any N as long as the matrix $\mathbf{M}$ is positive definite. For z larger than the top eigenvalue of $\mathbf{A}$, $(z\mathbf{1}-\mathbf{A})$ will be positive definite. The Gaussian formula allows us to compute the inverse of the square-root of the determinant, but we can neutralize the power $-1/2$ by introducing an extra factor -2 in front of the logarithm. Applying the replica trick (13.1) we thus get

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

$$\mathbb{E}[g_{\mathbf{A}}(z)] = -2\mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}z}\lim_{n\to 0}\frac{Z^n - 1}{Nn}\right], \tag{13.6}$$

with

$$Z^n := \int \prod_{\alpha=1}^{n} \frac{\mathrm{d}^N \boldsymbol{\psi}_\alpha}{(2\pi)^{N/2}} \exp\left(-\sum_{\alpha=1}^{n} \frac{\boldsymbol{\psi}_\alpha^T (z\mathbf{1} - \mathbf{A})\boldsymbol{\psi}_\alpha}{2}\right), \tag{13.7}$$

where we have written Z^n as the product of n copies of the same Gaussian integral. This is all fine, except our Z^n is only defined for integer n and we need to take $n \to 0$. The limiting Stieltjes transform is defined as $\mathfrak{g}_{\mathbf{A}}(z) = \lim_{N\to\infty} \mathbb{E}[g_{\mathbf{A}}(z)]$, but in practice the replica trick will allow us to compute

$$\widehat{\mathfrak{g}}_{\mathbf{A}}(z) = -2\frac{\mathrm{d}}{\mathrm{d}z}\lim_{n\to 0}\lim_{N\to\infty}\frac{\mathbb{E}[Z^n] - 1}{Nn}, \tag{13.8}$$

and hope that the two limits $n \to 0$ and $N \to \infty$ commute, such that $\widehat{\mathfrak{g}}_{\mathbf{A}}(z) = \mathfrak{g}_{\mathbf{A}}(z)$. There is, however, no guarantee that these limits do commute.

13.1.2 The Wigner Case

As a detailed example of replica trick calculation, we now give all the steps necessary to compute the Stieltjes transform for the Wigner ensemble. We want to take the expectation value of Eq. (13.7) in the case where $\mathbf{A} = \mathbf{X}$, a symmetric Gaussian rotational invariant matrix:

$$\begin{aligned}
\mathbb{E}[Z^n] &= \int \prod_{\alpha=1}^{n} \frac{\mathrm{d}^N \boldsymbol{\psi}_\alpha}{(2\pi)^{N/2}} \mathbb{E}\left[\exp\left(-\frac{z}{2}\sum_{\alpha=1}^{n}\sum_{i=1}^{N}\psi_{\alpha i}^2 - \sum_{i<j}^{N}\mathbf{X}_{ij}\psi_{\alpha i}\psi_{\alpha j} - \frac{1}{2}\sum_{i}^{N}\mathbf{X}_{ii}\psi_{\alpha i}\psi_{\alpha i}\right)\right], \\
&= \int \prod_{\alpha=1}^{n} \frac{\mathrm{d}^N \boldsymbol{\psi}_\alpha}{(2\pi)^{N/2}} \exp\left(-\frac{z}{2}\sum_{\alpha=1}^{n}\sum_{i=1}^{N}\psi_{\alpha i}^2\right) \prod_{i<j}^{N} \mathbb{E}\left[\exp\left(-\mathbf{X}_{ij}\sum_{\alpha=1}^{n}\psi_{\alpha i}\psi_{\alpha j}\right)\right] \\
&\quad \times \prod_{i}^{N} \mathbb{E}\left[\exp\left(-\frac{1}{2}\mathbf{X}_{ii}\sum_{\alpha=1}^{n}\psi_{\alpha i}\psi_{\alpha i}\right)\right],
\end{aligned} \tag{13.9}$$

where we have isolated the products of expectation of independent terms and separated the diagonal and off-diagonal terms. We can evaluate the expectation values using the following identity: for a centered Gaussian variable x of variance σ^2, we have

$$\mathbb{E}[\mathrm{e}^{ax}] = \mathrm{e}^{\sigma^2 a^2/2}. \tag{13.10}$$

Using the fact that the diagonal and off-diagonal elements have a variance equal to, respectively, $2\sigma^2/N$ and σ^2/N (see Section 2.2.2), we get

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

$$\mathbb{E}[Z^n] = \int \prod_{\alpha=1}^{n} \frac{\mathrm{d}^N \boldsymbol{\psi}_\alpha}{(2\pi)^{N/2}} \exp\left(-\frac{z}{2}\sum_{\alpha=1}^{n}\sum_{i=1}^{N}\psi_{\alpha i}^2\right)\prod_{i<j}^{N}\exp\left(\frac{\sigma^2}{2N}\left(\sum_{\alpha=1}^{n}\psi_{\alpha i}\psi_{\alpha j}\right)^2\right)$$
$$\times \prod_{i}^{N}\exp\left(\frac{\sigma^2}{4N}\left(\sum_{\alpha=1}^{n}\psi_{\alpha i}\psi_{\alpha i}\right)^2\right). \tag{13.11}$$

We can now combine the last two sums in the exponential into a single sum over $\{ij\}$, which we can further transform into

$$\frac{\sigma^2}{4N}\sum_{i,j=1}^{N}\left(\sum_{\alpha=1}^{n}\psi_{\alpha i}\psi_{\alpha j}\right)^2 = \frac{\sigma^2}{4N}\sum_{\alpha,\beta=1}^{n}\left(\sum_{i=1}^{N}\psi_{\alpha i}\psi_{\beta i}\right)^2. \tag{13.12}$$

We would like to integrate over the variables $\psi_{\alpha i}$ but the argument of the exponential contains fourth order terms in the ψ's. To tame this term, one uses the Hubbard–Stratonovich identity:

$$\exp\left(\frac{ax^2}{2}\right) = \int \frac{\mathrm{d}q}{\sqrt{2\pi a}}\exp\left(-\frac{q^2}{2a}+xq\right). \tag{13.13}$$

Before we use Hubbard–Stratonovich, we need to regroup diagonal and off-diagonal terms in $\alpha\beta$:

$$\frac{\sigma^2}{4N}\sum_{\alpha,\beta=1}^{n}\left(\sum_{i=1}^{N}\psi_{\alpha i}\psi_{\beta i}\right)^2 = \frac{\sigma^2}{2N}\sum_{\alpha<\beta}^{n}\left(\sum_{i=1}^{N}\psi_{\alpha i}\psi_{\beta i}\right)^2 + \frac{\sigma^2}{N}\sum_{\alpha=1}^{n}\left(\sum_{i=1}^{N}\frac{\psi_{\alpha i}\psi_{\alpha i}}{2}\right)^2, \tag{13.14}$$

where in the diagonal terms we have pushed the factor $1/4$ in the squared quantity for later convenience. We can now use Eq. (13.13), introducing diagonal $q_{\alpha\alpha}$ and upper triangular $q_{\alpha\beta}$ to linearize the squared quantities. Writing the q's as a symmetric matrix $\mathbf{q}$ we have

$$\mathbb{E}[Z^n] \propto \int \mathrm{d}\mathbf{q}\int\prod_{\alpha=1}^{n}\frac{\mathrm{d}^N\boldsymbol{\psi}_\alpha}{(2\pi)^{N/2}}\exp-\left(\frac{N\,\mathrm{Tr}\,\mathbf{q}^2}{4\sigma^2}+\sum_{i=1}^{N}\sum_{\alpha,\beta=1}^{n}\frac{(z\delta_{\alpha\beta}-q_{\alpha\beta})\psi_{\alpha i}\psi_{\beta i}}{2}\right), \tag{13.15}$$

where $\mathrm{d}\mathbf{q}$ is the integration over the independent component of the $n\times n$ symmetric matrix $\mathbf{q}$; note that we have dropped z-independent constant factors. The integral of $\psi_{\alpha i}$ is now a multivariate Gaussian integral, actually N copies of the very same n-dimensional Gaussian integral:

$$\int\prod_{\alpha=1}^{n}\frac{\mathrm{d}\psi_\alpha}{\sqrt{2\pi}}\exp-\left(\sum_{\alpha,\beta=1}^{n}\frac{(z\delta_{\alpha\beta}-q_{\alpha\beta})\psi_\alpha\psi_\beta}{2}\right) = (\det(z\mathbf{1}-\mathbf{q}))^{-1/2}. \tag{13.16}$$

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

Raising this integral to the Nth power and using $\det \mathbf{M} = \exp \operatorname{Tr} \log \mathbf{M}$ we find

$$\mathbb{E}[Z^n] \propto \int \mathrm{d}\mathbf{q} \exp - \left[N \operatorname{Tr} \left(\frac{\mathbf{q}^2}{4\sigma^2} + \frac{1}{2} \log(z\mathbf{1} - \mathbf{q}) \right) \right] := \int \mathrm{d}\mathbf{q} \exp \left(-\frac{N}{2} F(\mathbf{q}) \right). \tag{13.17}$$

We now fix n and evaluate $\mathbb{E}[Z^n]$ for very large N, leaving the limit $n \to 0$ for later. In the large N limit, the integral over the matrix $\mathbf{q}$ can be done by the saddle point method. More precisely, we should find an extremum of $F(\mathbf{q})$ in the $n(n+1)/2$ elements of $\mathbf{q}$. Alternatively we can diagonalize $\mathbf{q}$, introducing the log of a Vandermonde determinant in the exponential (see Section 5.1.4).[2] In terms of the eigenvalues q_α of $\mathbf{q}$, one has

$$F(\{q_\alpha\}) = \sum_{\alpha=1}^{n} \frac{q_\alpha^2}{2\sigma^2} + \log(z - q_\alpha) - \frac{1}{N} \sum_{\alpha \neq \beta} \log |q_\alpha - q_\beta|. \tag{13.18}$$

To find the saddle point, we take the partial derivatives of $F\{q_\alpha\}$ with respect to the $\{q_\alpha\}$ and equate them to zero:

$$\frac{q_\alpha}{\sigma^2} - \frac{1}{z - q_\alpha} - \frac{1}{N} \sum_{\alpha \neq \beta} \frac{2}{q_\alpha - q_\beta} = 0. \tag{13.19}$$

The effect of the last term is to push the eigenvalues q_α away from one another, such that in equilibrium their relative distance is of order $1/N$ and the last term is of the same order as the first two. Since there are only n such eigenvalues, the total spread (from the largest to smallest) is of order n/N, which we will neglect when $N \to \infty$. Hence we can assume that all eigenvalues are identical and equal to $q^*(z)$, where $q^*(z)$ satisfies

$$z - q^* = \frac{\sigma^2}{q^*}. \tag{13.20}$$

We recognize the self-consistent equation for the Stieltjes transform of the Wigner (Eq. (2.35)) if we make the identification $q^*(z) = \sigma^2 \mathfrak{g}_{\mathbf{X}}(z)$. For N large and n small we indeed have

$$\mathbb{E}[Z^n] = \exp \left(-\frac{Nn}{2} F_1(z, q^*(z)) \right) \text{ with } F_1(z,q) = \frac{q^2}{2\sigma^2} + \log(z - q), \tag{13.21}$$

so

$$\lim_{n \to 0} \lim_{N \to \infty} \frac{\mathbb{E}[Z^n] - 1}{Nn} = -\frac{F_1(z, q^*(z))}{2}. \tag{13.22}$$

Finally, from Eq. (13.8), we should have

$$\mathfrak{g}_{\mathbf{X}}(z) = \frac{\mathrm{d}}{\mathrm{d}z} F_1(z, q^*(z)). \tag{13.23}$$

[2] A third method is used in spin-glass problems where the integrand has permutation symmetry but not necessarily rotational symmetry, see Section 13.4.

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

To finish the computation we need to take the derivative of $F_1(z, q^*(z))$ with respect to z, but since $q^*(z)$ is an extremum of F_1, the partial derivative of $F_1(z, q)$ with respect to q is zero at $q = q^*(z)$. Hence

$$\mathfrak{g}_{\mathbf{X}}(z) = \frac{\partial}{\partial z} F_1(z, q)\Big|_{q=q^*(z)} = \frac{1}{(z - q^*)} = \frac{q^*(z)}{\sigma^2}. \tag{13.24}$$

We thus recover the usual solution of the self-consistent Wigner equation.

13.2 Resolvent Matrix

13.2.1 General Case

We saw that the replica trick can be used to compute the average Stieltjes transform of a random matrix. The Stieltjes transform is the normalized trace of the resolvent matrix $\mathbf{G}_{\mathbf{A}}(z) = (z\mathbf{1} - \mathbf{A})^{-1}$. In Chapter 19 we will need to know the average of the elements of the resolvent matrix for free addition and multiplication. These averages can also be computed using the replica trick. An element of an inverse matrix can indeed be written as a multivariate Gaussian integral:

$$\left[\mathbf{M}^{-1}\right]_{ij} \equiv \frac{1}{Z} \int \frac{\mathrm{d}^N \boldsymbol{\psi}}{(2\pi)^{N/2}} \psi_i \psi_j \exp\left(-\frac{\boldsymbol{\psi}^T \mathbf{M} \boldsymbol{\psi}}{2}\right), \quad Z := \int \frac{\mathrm{d}^N \boldsymbol{\psi}}{(2\pi)^{N/2}} \exp\left(-\frac{\boldsymbol{\psi}^T \mathbf{M} \boldsymbol{\psi}}{2}\right), \tag{13.25}$$

which we can rewrite as

$$\left[\mathbf{M}^{-1}\right]_{ij} = \lim_{m \to -1} Z^m \int \mathrm{d}^N \boldsymbol{\psi}\, \psi_i \psi_j \exp\left(-\frac{\boldsymbol{\psi}^T \mathbf{M} \boldsymbol{\psi}}{2}\right). \tag{13.26}$$

If we express Z^m for $m \in \mathbb{N}^+$ as m Gaussian integrals and combine them with the integral with the $\psi_i \psi_j$ term (which we label number 1) we get, with $n = m + 1$:

$$\left[\mathbf{M}^{-1}\right]_{ij} = \lim_{n \to 0} \int \prod_{\alpha=1}^{n} \frac{\mathrm{d}^N \boldsymbol{\psi}_\alpha}{(2\pi)^{nN/2}} \psi_{1i} \psi_{1j} \exp\left(-\sum_{\alpha=1}^{n} \frac{\boldsymbol{\psi}_\alpha^T \mathbf{M} \boldsymbol{\psi}_\alpha}{2}\right) := \left\langle \psi_{1i} \psi_{1j} \right\rangle_{n=0}. \tag{13.27}$$

This equation can then be used to compute averages of elements of the resolvent matrix by using $\mathbf{M} = z\mathbf{1} - \mathbf{A}$ for the relevant random matrix $\mathbf{A}$. For example, in the case of Gaussian Wigner matrices, the correlation $\left\langle \psi_{1i} \psi_{1j} \right\rangle_{n=0}$ can be computed using the saddle point configuration of the ψ's. Since different i all decouple and play the same role, it is clear that

$$\mathbb{E}[\mathbf{G}_{\mathbf{X}}(z)]_{ij} = \left\langle \psi_{1i} \psi_{1j} \right\rangle_{n=0} = \delta_{ij} \mathfrak{g}_{\mathbf{X}}(z), \tag{13.28}$$

i.e. the average resolvent of a Gaussian matrix is simply the average Stieltjes transform times the identity matrix.

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

13.2.2 Free Addition

In this section we will show how to use Eq. (13.27) to compute the average of the full resolvent for the sum of two randomly rotated matrices. Since we know these matrices are free in the large N limit, we expect to recover the additivity of R-transforms, but we will in fact obtain a slightly richer result. Also, the replica method is very convenient to manipulate and resum the perturbation theory alluded to in Section 12.2.

Consider two symmetric matrices $\mathbf{A}$ and $\mathbf{B}$ and the new matrix $\mathbf{C} = \mathbf{A} + \mathbf{OBO}^T$ where $\mathbf{O}$ is a random orthogonal matrix. We want to compute

$$\mathbb{E}[\mathbf{G}_\mathbf{C}(z)] = \mathbb{E}\left[(z\mathbf{1} - \mathbf{A} - \mathbf{OBO}^T)^{-1}\right], \tag{13.29}$$

where the expectation value is over the orthogonal matrix $\mathbf{O}$. We can always choose $\mathbf{B}$ to be diagonal. If $\mathbf{B}$ is not diagonal to start with, we just absorb the orthogonal matrix that diagonalizes $\mathbf{B}$ in the matrix $\mathbf{O}$. Expressing $\mathbf{G}_\mathbf{C}(z)$ in the eigenbasis of $\mathbf{A}$ is equivalent to choosing $\mathbf{A}$ to be diagonal. In that basis, the off-diagonal elements of $\mathbb{E}[\mathbf{G}_\mathbf{C}(z)]$ must be zero by the following argument: since both $\mathbf{A}$ and $\mathbf{B}$ are diagonal, for every matrix $\mathbf{O}$ that contributes to an off-diagonal element of $\mathbb{E}[\mathbf{G}_\mathbf{C}(z)]$ there exists an equally probable matrix $\mathbf{O}'$ with the same contribution but opposite sign, hence the average must be zero. Note that while the average matrix $\mathbb{E}[\mathbf{G}_\mathbf{C}(z)]$ commutes with $\mathbf{A}$, a particular realization of the random matrix $\mathbf{G}_\mathbf{C}(z)$ (corresponding to a specific choice for $\mathbf{O}$) will not in general commute with $\mathbf{A}$.

Now, let us use the replica formalism to compute $\mathbb{E}[\mathbf{G}_\mathbf{C}(z)]$, i.e. start with

$$\begin{aligned}\mathbf{G}_\mathbf{C}(z)_{ij} = \lim_{n\to 0} \int \prod_{\alpha=1}^{n} \frac{\mathrm{d}^N \boldsymbol{\psi}_\alpha}{(2\pi)^{N/2}} \psi_{1i}\psi_{1j} \exp\left(-\sum_{\alpha=1}^{n} \frac{\boldsymbol{\psi}_\alpha^T (z\mathbf{1} - \mathbf{A})\boldsymbol{\psi}_\alpha}{2}\right) \\ \times \mathbb{E}\left[\exp\left(\sum_{\alpha=1}^{n} \frac{\boldsymbol{\psi}_\alpha^T \mathbf{OBO}^T \boldsymbol{\psi}_\alpha}{2}\right)\right],\end{aligned} \tag{13.30}$$

where now we skip the i, j indices on $\boldsymbol{\psi}_\alpha$, treated as vectors.

The last term with the expectation value can be rewritten as

$$I = \mathbb{E}\left[\exp\left(\frac{N}{2}\sum_{\alpha=1}^{n} \mathrm{Tr}\, \mathbf{Q}_{\alpha,\alpha}\mathbf{OBO}^T\right)\right], \tag{13.31}$$

where $\mathbf{Q}_{\alpha,\beta} = \boldsymbol{\psi}_\alpha \boldsymbol{\psi}_\beta^T / N$ is an $n \times n$ symmetric matrix. We recognize the Harish-Chandra–Itzykson–Zuber integral discussed in Chapter 10, with one matrix ($\sum_{\alpha=1}^{n} \mathbf{Q}_{\alpha,\alpha}$) being at most of rank $n \ll N$, so we can use the low-rank formula Eq. (10.45). Our expectation value thus becomes

$$I = \exp\left(\frac{N}{2}\,\mathrm{Tr}_n\, H_\mathbf{B}(\mathbf{Q})\right), \tag{13.32}$$

where Tr_n denotes the trace of an $n \times n$ matrix and $H_\mathbf{B}$ is the anti-derivative of the R-transform of $\mathbf{B}$.

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

We now need to perform the integral of $\boldsymbol{\psi}_\alpha$ in Eq. (13.30). But in order to do so we must deal with $\mathrm{Tr}\, H_{\mathbf{B}}(\mathbf{Q})$, which is a non-linear function of the $\boldsymbol{\psi}_\alpha$. The trick is to make the matrix $\mathbf{Q}$ an integration variable that we fix to its definition using a delta function. The delta function is itself represented as an integral over another $(n \times n)$ symmetric matrix $\mathbf{Y}$ along the imaginary axis. In other words, we introduce the following representation of the delta function in Eq. (13.30):

$$\int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{N^{n(n+1)/2} \mathrm{d}\mathbf{Y}}{2^{3n/2}\pi^{n/2}} \exp\left(-\frac{N}{2}\,\mathrm{Tr}_n\, \mathbf{QY} + \frac{1}{2}\sum_{\alpha,\beta=1}^{n} \mathbf{Y}_{\alpha,\beta}\boldsymbol{\psi}_\alpha \boldsymbol{\psi}_\beta^T\right), \tag{13.33}$$

where the integrals over $\mathrm{d}\mathbf{Q}$ and $\mathrm{d}\mathbf{Y}$ are over symmetric matrices. We have absorbed a factor of N in $\mathbf{Y}$ and a factor of 2 on its diagonal, hence the extra factors of 2 and N in front of $\mathrm{d}\mathbf{Y}$. We can now perform the following Gaussian integral over $\boldsymbol{\psi}_\alpha$:

$$J_{ij} = \int \prod_{k=1}^{N}\prod_{\alpha=1}^{n} \frac{\mathrm{d}\psi_{\alpha k}}{\sqrt{2\pi}} \psi_{1i}\psi_{1j} \exp\left(-\frac{1}{2}\sum_{k=1}^{N}\sum_{\alpha,\beta=1}^{n} \psi_{\alpha k}(z\delta_{\alpha,\beta} - a_k\delta_{\alpha,\beta} - \mathbf{Y}_{\alpha\beta})\psi_{\beta k}\right), \tag{13.34}$$

where we have written the vectors $\boldsymbol{\psi}_\alpha$ in terms of their components $\psi_{\alpha k}$, and where a_k are the eigenvalues of $\mathbf{A}$. We notice that the Gaussian integral is diagonal in the index k, so we have $N-1$ n-dimensional Gaussian integrands differing only by their value of a_k, and a last integral for $k = i = j$, where the term ψ_{1i}^2 is in front of the Gaussian integrand (the integral is zero if $i \neq j$, meaning that $\mathbf{G}_{\mathbf{C}}(z)$ is diagonal, as expected). The result is then

$$J_{ij} = \delta_{ij}\left[((z-a_i)\mathbf{1}_n - \mathbf{Y})^{-1}\right]_{11} \prod_{k=1}^{N} (\det((z-a_k)\mathbf{1}_n - \mathbf{Y}))^{-1/2}, \tag{13.35}$$

where the first term is the 11 element of an $n \times n$ matrix, coming from the term ψ_{1i}^2.

Returning to our main expression Eq. (13.30) and dropping constants that are 1 as $n \to 0$,

$$\begin{aligned}\mathbb{E}[\mathbf{G}_{\mathbf{C}}(z)]_{ij} = \lim_{n\to 0} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \mathrm{d}\mathbf{Q} \int \mathrm{d}\mathbf{Y}\delta_{ij} \left[((z-a_i)\mathbf{1}_n - \mathbf{Y})^{-1}\right]_{11} \\ \times \exp\left[\frac{N}{2}\left(-\mathrm{Tr}_n\, \mathbf{QY} + \mathrm{Tr}_n\, H_{\mathbf{B}}(\mathbf{Q}) - \frac{1}{N}\sum_{k=1}^{N} \mathrm{Tr}_n \log((z-a_k)\mathbf{1}_n - \mathbf{Y})\right)\right].\end{aligned} \tag{13.36}$$

For large N the integral over $\mathbf{Y}$ and $\mathbf{Q}$ is dominated by the saddle point, i.e. the extremum of the argument of the exponential. The inverse-matrix term in front of the exponential does not contain a power of N so it does not contribute to the determination of the saddle point. The extremum is over a function of two $n \times n$ symmetric matrices. Taking derivatives with respect to $\mathbf{Q}_{\alpha\beta}$ and equating it to zero gives

$$\mathbf{Y}_{\alpha\beta} = [R_{\mathbf{B}}(\mathbf{Q})]_{\alpha\beta}, \tag{13.37}$$

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

and similarly when taking derivatives with respect to $\mathbf{Y}_{\alpha\beta}$:

$$\mathbf{Q}_{\alpha\beta} = \frac{1}{N}\sum_{k=1}^{N}\left[((z-a_k)\mathbf{1}_n - \mathbf{Y})^{-1}\right]_{\alpha\beta}. \tag{13.38}$$

Let us look at these equations in a basis where $\mathbf{Q}$ is diagonal. The first equation shows that $\mathbf{Y}$ is also diagonal, so that the second equation reads

$$\mathbf{Q}_{\alpha\alpha} = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{z-a_k-\mathbf{Y}_{\alpha\alpha}} \equiv g_{\mathbf{A}}(z-\mathbf{Y}_{\alpha\alpha}). \tag{13.39}$$

Hence all n diagonal elements $\mathbf{Q}_{\alpha\alpha}$ and $\mathbf{Y}_{\alpha\alpha}$, $\alpha = 1,\ldots,n$ satisfy the same pair of equations. For large z, there is a unique solution to these equations, hence $\mathbf{Q}$ and $\mathbf{Y}$ must be multiples of the identity $\mathbf{Q} = q^*\mathbf{1}_n$ and $\mathbf{Y} = y^*\mathbf{1}_n$, as expected from the rotational symmetry of the argument of the exponential that we are maximizing. The quantities q^* and y^* are the unique solutions of

$$y = R_{\mathbf{B}}(q) \qquad \text{and} \qquad q = \mathfrak{g}_{\mathbf{A}}(z-y). \tag{13.40}$$

The saddle point for $\mathbf{Y}$ is real while our integral representation Eq. (13.33) was over purely imaginary matrices; but for large values of z, the solutions of Eqs. (13.40) give small values for q^* and y^*, and for such small values the integral contour can be deformed without encountering any singularities. We also justify the use of $R_{\mathbf{B}}(q^*) = H'_{\mathbf{B}}(q^*)$ as q^* can be made arbitrarily small by choosing a large enough z.

The expectation of the resolvent is thus given, for large enough z and N, by

$$\mathbb{E}[\mathbf{G}_{\mathbf{C}}(z)_{ij}] \approx \lim_{n\to 0}\frac{\delta_{ij}}{(z-a_i-y^*)}\exp\left[\frac{nN}{2}\left(-q^*y^* + H_{\mathbf{B}}(q^*) - \frac{1}{N}\sum_{k=1}^{N}\log(z-a_k-y^*)\right)\right]. \tag{13.41}$$

As $n \to 0$ the exponential drops out and we obtain, in matrix form,

$$\mathbb{E}[\mathbf{G}_{\mathbf{C}}(z)] = \mathbf{G}_{\mathbf{A}}(z - R_{\mathbf{B}}(q^*)) \quad \text{with} \quad q^* = \mathfrak{g}_{\mathbf{A}}(z - R_{\mathbf{B}}(q^*)). \tag{13.42}$$

13.2.3 Resolvent Subordination for Addition and Multiplication

Equation (13.42) relates the average resolvent of $\mathbf{C}$ to that of $\mathbf{A}$. By taking the normalized trace on both sides we find

$$q^* = \mathfrak{g}_{\mathbf{C}}(z) = \mathfrak{g}_{\mathbf{A}}(z - R_{\mathbf{B}}(q^*)), \tag{13.43}$$

which is precisely the subordination relation for the Stieltjes transform of a free sum that we found in Section 11.3.8. We just have rederived this result once again with replicas.

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

But what is more interesting is that we have found a relationship for the average of a matrix element of the full resolvent matrix, namely

$$\mathbb{E}[\mathbf{G}_{\mathbf{C}}(z)] = \mathbf{G}_{\mathbf{A}}\left(z - R_{\mathbf{B}}(\mathfrak{g}_{\mathbf{C}}(z))\right). \tag{13.44}$$

This relation will give precious information on the overlap between the eigenvalues of $\mathbf{A}$ and $\mathbf{B}$ and those of $\mathbf{C}$. Note that, by symmetry, one also has

$$\mathbb{E}[\mathbf{G}_{\mathbf{C}}(z)] = \mathbf{G}_{\mathbf{B}}\left(z - R_{\mathbf{A}}(\mathfrak{g}_{\mathbf{C}}(z))\right). \tag{13.45}$$

In the free product case, namely $\mathbf{C} = \mathbf{A}^{\frac{1}{2}}\mathbf{B}\mathbf{A}^{\frac{1}{2}}$ where $\mathbf{A}$ and $\mathbf{B}$ are large positive definite matrices whose eigenvectors are mutually random, a very similar replica computation gives a subordination relation for the average T matrix:

$$\mathbb{E}[\mathbf{T}_{\mathbf{C}}(\zeta)] = \mathbf{T}_{\mathbf{A}}[S_{\mathbf{B}}(\mathfrak{t}_{\mathbf{C}}(\zeta))\zeta], \tag{13.46}$$

with $S_{\mathbf{B}}(t)$ the S-transform of the matrix $\mathbf{B}$. If we take the normalized trace on both sides, we recover the subordination relation Eq. (11.109). Equation (13.46) can then be turned into a subordination relation for the full resolvent:

$$\mathbb{E}[\mathbf{G}_{\mathbf{C}}(z)] = S^*\mathbf{G}_{\mathbf{A}}(zS^*) \quad \text{with} \quad S^* := S_{\mathbf{B}}(z\mathfrak{g}_{\mathbf{C}}(z) - 1). \tag{13.47}$$

13.2.4 "Quenched" vs "Annealed"

The replica trick is quite burdensome as one has to keep track of n copies of an integration vector $\boldsymbol{\psi}_n$ and these vectors interact through the averaging process. At large N one typically has to do a saddle point over one or several $n \times n$ matrices (e.g. $\mathbf{Q}$ and $\mathbf{Y}$ in the free addition computation of the previous section), and at the end take the $n \to 0$ limit. But in all computations of Stieltjes transforms so far, taking $n = 1$ instead of $n \to 0$ gives the correct saddle point and the correct final result. In other words, assuming that

$$\mathbb{E}[\log Z] \approx \log \mathbb{E}[Z] \tag{13.48}$$

leads to the correct result. For historical reasons coming from physics, $\mathbb{E}[\log Z]$ is called a *quenched average* whereas $\log \mathbb{E}[Z]$ is called an *annealed average*.

For example, if we go back to Eq. (13.21) we see that taking the logarithm of the $n = 1$ result gives the same result as the correct $n \to 0$ limit. The same is true for the Wishart case. For the free addition and multiplication one can also compute the Stieltjes transform using $n = 1$. This is a general result for bulk properties of random matrices. Most natural ensembles of random symmetric matrices (such as those from Chapter 5 and those arising from free addition and multiplication) feature a strong repulsion of eigenvalues. Because of this repulsion, eigenvalues do not fluctuate much around their *classical positions* – see the detailed discussion in Section 5.4.1. It is the absence of eigenvalue fluctuations on the global scale that makes the $n = 1$ and $n \to 0$ saddle points equivalent.

For the rank-1 HCIZ integral, on the other hand, things are more subtle. As we show in the next section, the annealed average $n = 1$ gives the right answer in some interval

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

of parameters, when the integral is dominated by the bulk properties of eigenvalues. Outside this regime, fluctuations of the largest eigenvalue matter and the $n = 1$ result is no longer correct.

13.3 Rank-1 HCIZ and Replicas

In Chapter 10 we studied the rank-1 HCIZ integral and defined the function $H_{\mathbf{B}}(t)$ as

$$H_{\mathbf{B}}(t) := \lim_{N\to\infty} \frac{2}{N} \log \left\langle \exp\left(\frac{N}{2} \operatorname{Tr} \mathbf{TOBO}^T\right)\right\rangle_{\mathbf{O}}, \tag{13.49}$$

where the averaging is done over the orthogonal group for $\mathbf{O}$, $\mathbf{T}$ is a rank-1 matrix with eigenvalue t and $\mathbf{B}$ a fixed matrix. If $\mathbf{B}$ is a member of a random ensemble, such as the Wigner ensemble, the averaging over $\mathbf{O}$ should be done for a fixed $\mathbf{B}$ and only later the function $H_{\mathbf{B}}(t)$ can be averaged, if needed, over the randomness of $\mathbf{B}$ (quenched average). One could also do an annealed average over $\mathbf{B}$, defining another function $\hat{H}(t)$ as

$$\hat{H}(t) := \lim_{N\to\infty} \frac{2}{N} \log \left\langle \exp\left(\frac{N}{2} \operatorname{Tr} \mathbf{TOBO}^T\right)\right\rangle_{\mathbf{O},\mathbf{B}}. \tag{13.50}$$

It turns out that for small enough values of t, the two quantities are equal, i.e. $\hat{H}(t) = H_{\mathbf{B}}(t)$. For larger values of t, however, these two quantities differ. The aim of this section is to compute explicitly these quantities in the Wigner case using the replica trick, and show that there is a phase transition for a well-defined value $t = t_c$ beyond which quenched and annealed averages do not coincide.

13.3.1 Annealed Average

Let us compute directly the "annealed" average when $\mathbf{B} = \mathbf{X}$ is a Wigner matrix and $\mathbf{T} = t\, \mathrm{e}_1 \mathrm{e}_1^T$, where t is the only non-zero eigenvalue of $\mathbf{T}$ and e_1 is the unit vector $(1, 0, \ldots, 0)^T$. Then

$$\begin{aligned}
\left\langle \exp\left(\frac{N}{2} \operatorname{Tr} \mathbf{TX}\right)\right\rangle_{\mathbf{X}} &= \left\langle \exp\left(\frac{Nt}{2} \mathrm{e}_1^T \mathbf{X} \mathrm{e}_1\right)\right\rangle_{\mathbf{X}} \\
&= \int \frac{\mathrm{d}X_{11}}{\sqrt{4\pi\sigma^2/N}} \exp\left(\frac{NtX_{11}}{2} - \frac{N}{4\sigma^2} X_{11}^2\right) \\
&= \exp\left(\frac{N}{2} \frac{t^2\sigma^2}{2}\right),
\end{aligned} \tag{13.51}$$

so the annealed $\hat{H}_{\mathrm{wig}}(t)$ is given by

$$\hat{H}_{\mathrm{wig}}(t) = \frac{\sigma^2}{2} t^2, \tag{13.52}$$

which, at least superficially, coincides with the integral of the R-transform of a Wigner matrix, Eq. (10.61).

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

13.3.2 Quenched Average

The annealed average corresponds, in the replica language, to $n = 1$. Let us now turn to arbitrary (integer) n. To keep notation light we will set $\sigma^2 = 1$. As in Eq. (10.31) we define the partition function

$$Z_t(\mathbf{X}) = \int \frac{\mathrm{d}^N \boldsymbol{\psi}}{(2\pi)^{N/2}} \delta\left(\|\boldsymbol{\psi}\|^2 - Nt\right) \exp\left(\frac{1}{2}\boldsymbol{\psi}^T \mathbf{X} \boldsymbol{\psi}\right), \tag{13.53}$$

seeking to compute, at the end of the calculation,

$$\mathbb{E}[H_{\mathbf{X}}(t)] = \lim_{N\to\infty} \frac{2}{N} \lim_{n\to 0} \left(\frac{Z_t^n(\mathbf{X}) - 1}{n}\right) - 1 - \log t, \tag{13.54}$$

where $1 + \log t$ is the large N limit of $2/N \log Z_t(\mathbf{X} = 0)$, with $Z_t(0)$ given by Eq. (10.38). If we write $Z_t^n(\mathbf{X})$ as multiple copies of the same integral and express the Dirac deltas as Fourier integrals over z_α, we get

$$Z_t^n(\mathbf{X}) = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \prod_{\alpha=1}^n \mathrm{d}z_\alpha \int \prod_{\alpha=1}^n \frac{\mathrm{d}^N \boldsymbol{\psi}_\alpha}{(2\pi)^{N/2}} \exp\left(\frac{1}{2}\sum_{\alpha=1}^n \left(N z_\alpha t - z_\alpha \boldsymbol{\psi}_\alpha^T \boldsymbol{\psi}_\alpha\right) + \frac{1}{2}\sum_{\alpha=1}^n \boldsymbol{\psi}_\alpha^T \mathbf{X} \boldsymbol{\psi}_\alpha\right). \tag{13.55}$$

In order to take the expectation value over the Gaussian random matrix $\mathbf{X}$, we need as always to separate the diagonal and off-diagonal elements of $\mathbf{X}$. The steps are the same as those we took in Section 13.1.2:

$$\begin{aligned}\mathbb{E}\left[\exp\left(\frac{1}{2}\sum_{\alpha=1}^n \boldsymbol{\psi}_\alpha^T \mathbf{X} \boldsymbol{\psi}_\alpha\right)\right] &= \exp\left(\sum_{i,j=1}^N \frac{1}{4N}\left(\sum_{\alpha=1}^n \psi_{\alpha i}\psi_{\alpha j}\right)^2\right) \\ &= \exp\left(\frac{1}{4N}\sum_{\alpha,\beta=1}^n \left(\sum_{i=1}^N \psi_{\alpha i}\psi_{\beta i}\right)^2\right),\end{aligned} \tag{13.56}$$

which can be rewritten as a Hubbard–Stratonovich integral over an $n \times n$ matrix $\mathbf{q}$:

$$\mathbb{E}[\ldots] = C(n) \int \mathrm{d}\mathbf{q} \exp\left(-N\frac{\mathrm{Tr}\,\mathbf{q}^2}{4} + \sum_{i=1}^N \sum_{\alpha,\beta=1}^n \frac{q_{\alpha\beta}\psi_{\alpha i}\psi_{\beta i}}{2}\right), \tag{13.57}$$

where $C(n)$ is a numerical coefficient. After Gaussian integration, one thus finds

$$\mathbb{E}[Z_t^n] = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \prod_{\alpha=1}^n \mathrm{d}z_\alpha \int \mathrm{d}\mathbf{q} C(n) \exp\left[\frac{N}{2}\left(t\,\mathrm{Tr}\,\mathbf{z} - \frac{\mathrm{Tr}\,\mathbf{q}^2}{2} - \mathrm{Tr}\log(\mathbf{z} - \mathbf{q})\right)\right], \tag{13.58}$$

which makes sense provided that the real part of z is larger than all the eigenvalues of $\mathbf{q}$. We now define

$$F_n(\mathbf{q}, \mathbf{z}; t) = t\,\mathrm{Tr}\,\mathbf{z} - \frac{\mathrm{Tr}\,\mathbf{q}^2}{2} - \mathrm{Tr}\log(\mathbf{z} - \mathbf{q}), \tag{13.59}$$

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

where $\mathbf{z}$ is the vector of z_α treated as a diagonal matrix. For $n \geq 1$ we need to find a saddle point in the space of $n \times n$ matrices $\mathbf{z}$ and $\mathbf{q}$, i.e. a point in that space where the first derivatives of $F_t(\mathbf{q}, \mathbf{z})$ are zero with a negative Hessian.

As a check, for $n = 1$ we have at the saddle point

$$q^* = \frac{1}{z^* - q^*} \quad \text{and} \quad t = \frac{1}{z^* - q^*} \quad \Rightarrow \quad z^* = t + \frac{1}{t} \quad \text{and} \quad q^* = t. \tag{13.60}$$

Hence

$$F_1(q^*, z^*; t) = t^2 + 1 - \frac{t^2}{2} + \log t, \tag{13.61}$$

or

$$\hat{H}_{\text{wig}}(t) = \frac{t^2}{2}, \tag{13.62}$$

as it should be.

We now go back to the general n case. Using Eq. (1.37) we can take a matrix derivative of Eq. (13.59) with respect to $\mathbf{q}$ and $\mathbf{z}$:

$$\mathbf{q} = (\mathbf{z} - \mathbf{q})^{-1} \qquad \text{and} \qquad \left[(\mathbf{z} - \mathbf{q})^{-1}\right]_{\alpha\alpha} = t. \tag{13.63}$$

In the following technical part, we solve these equations for integer $n \geq 1$. We discuss the final result at the end of this subsection.

The second equation in (13.63) comes from the derivative with respect to z; remember $\mathbf{z}$ is only a diagonal matrix, the derivative with respect to $\mathbf{z}$ tells us only about the diagonal elements.

From this we can argue that $\mathbf{z}$ must be a multiple of the identity and $\mathbf{q}$ of the form[3]

$$\mathbf{q} = \begin{pmatrix} t & b & \dots & b \\ b & t & \dots & b \\ \vdots & \vdots & \ddots & b \\ b & b & \dots & t \end{pmatrix}, \tag{13.64}$$

for some b to be determined. To find an equation for b and z we need to express Eqs. (13.63) in terms of those quantities. To do so we first write the matrix $\mathbf{q}$ as a rank-1 perturbation of a multiple of the identity matrix:

$$\mathbf{q} = (t - b)\mathbf{1} + nb\mathbf{P}_1, \tag{13.65}$$

where $\mathbf{P}_1 = \mathbf{e}\mathbf{e}^T$ is the projector onto the normalized vector of all 1:

$$\mathbf{e} = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}. \tag{13.66}$$

[3] More complicated, block diagonal structures for $\mathbf{q}$ sometimes need to be considered in the limit $n \to 0$. This is called "replica symmetry breaking", a phenomenon that occurs in many "complex" optimization problems, such as spin-glasses – see Section 13.4. Fortunately, in the present case, these complications are not present.

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

Note that the eigenvalues of the matrix $\mathbf{q}$ are $(t-b)+nb$ (with multiplicity 1) and $(t-b)$ (with multiplicity $(n-1)$). Since $\mathbf{z}$ is a multiple of the identity, the matrix $\mathbf{z}-\mathbf{q}$ is a rank-1 perturbation of a multiple of the identity and it can be inverted using the Sherman–Morrison formula, Eq. (1.28). The first of Eqs. (13.63) becomes

$$(t-b)\mathbf{1}+nb\mathbf{P}_1=\frac{\mathbf{1}}{z-t+b}+\frac{nb\mathbf{P}_1}{(z-t+b)^2(1-nb(z-t+b)^{-1})}. \tag{13.67}$$

We can now equate the prefactors in front of the identity matrix $\mathbf{1}$ and of the projector $\mathbf{P}_1$ separately, to get two equations for our two unknowns (z and b). For the identity matrix we get

$$(t-b)=\frac{1}{z-t+b} \quad \Rightarrow \quad z=(t-b)+\frac{1}{t-b}. \tag{13.68}$$

For the second equation, we first replace $(z-t+b)^{-1}$ by $t-b$ and get

$$nb=\frac{(t-b)^2nb}{1-nb(t-b)}. \tag{13.69}$$

We immediately find one solution: $b=0$. For this solution both $\mathbf{q}=q_0\mathbf{1}$ and $\mathbf{z}=z_0\mathbf{1}$ are multiples of the identity and we have $q_0=t$ and $z_0=t+t^{-1}$. This coincides with the (unique) solution we found in the annealed ($n=1$) case. For general n, there are potentially other solutions. Simplifying off nb, we find a quadratic equation for b:

$$1-nb(t-b)=(t-b)^2, \tag{13.70}$$

whose solutions we write as

$$b_\pm=\frac{(n-2)t\pm\sqrt{(n^2t^2-4(n-1))}}{2(n-1)}. \tag{13.71}$$

From the two solutions for b we can compute the corresponding values of z using Eq. (13.68). We get a term with a square-root on the denominator that we simplify using $(c\pm\sqrt{d})^{-1}\equiv(c\mp\sqrt{d})/(c^2-d)$. After further simplification we find

$$z_\pm=\frac{n^2t\pm(n-2)\sqrt{(n^2t^2-4(n-1))}}{2(n-1)}. \tag{13.72}$$

We now need to choose one of the three solutions z_0, z_+ and z_-. First, we consider integer $n\geq 1$ where the replica method is perfectly legitimate. We will later deal with the $n\to 0$ limit.

For $n=1$, z_- is ill defined while z_+ becomes identical to z_0. The only solution for all t is therefore z_0 and we recover the annealed result discussed in the previous subsection.

For $n\geq 2$, we first notice that the solutions $z_\pm$ do not exist for $t<t_{\rm s}:=2\sqrt{n-1}/n$; they yield a complex result when the result must be real. So for $t<t_{\rm s}$, z_0 is the solution. For larger values of t we should compare the values of $F_n(\mathbf{q}^*,\mathbf{z}^*;t)$ and choose the maximum one. For $t>t_{\rm s}$, the z_+ solution always dominates z_-, so we only consider z_+ and z_0 henceforth.

For $n=2$, the analysis is easy, $t_{\rm s}=1$ and $t_{\rm c}=1$: at $t=1$, $z_+=z_0$ and for $t>1$ the z_+ solution dominates. For $n>2$, the situation is a bit more subtle. The z_+ solution appears at $t_{\rm s}<1$ but at that point z_0 still dominates. At $t=1$, z_+ dominates. At some $t=t_{\rm c}$, with $t_{\rm s}<t_{\rm c}<1$, we must have $F_n(\mathbf{q}_0,\mathbf{z}_0;t_{\rm c})=F_n(\mathbf{q}_+,\mathbf{z}_+;t_{\rm c})$. This point could in principle be shown analytically but it is easier numerically. In particular we do not have an analytical expression for $t_{\rm c}(n)$ except the above bound $t_{\rm c}(n>2)<1$ and the value $t_{\rm c}(2)=1$ (see Fig. 13.1).

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

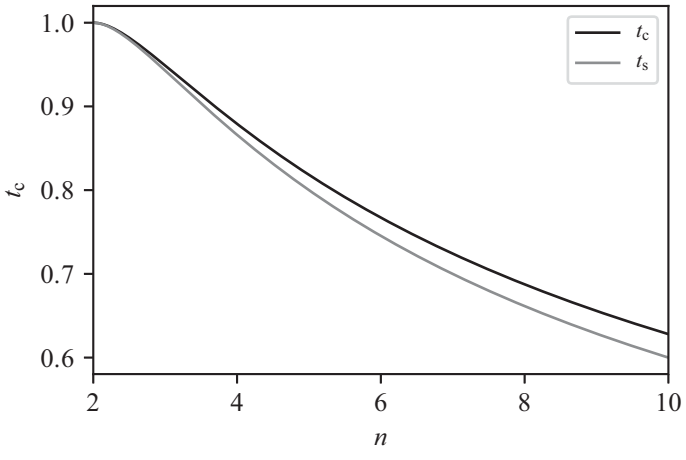

Figure 13.1 The point $t_c(n)$ where the $F_n(\mathbf{q}_0, \mathbf{z}_0; t_c) = F_n(\mathbf{q}_+, \mathbf{z}_+; t_c)$ and beyond which the z_+ solution dominates. Also shown is the point $t_s(n) = 2\sqrt{n-1}/n$ where the z_+ solution starts to exist. Note that $t_s \leq t_c \leq 1$ for all $n \geq 2$. Hence the transition appears below $t = 1$. Note also that $t_c \to 0$ as $n \to \infty$.

We can now put everything together but there is a trick to save computation effort. Given that our solutions cancel the partial derivatives of $F_t(\mathbf{q}, \mathbf{z}; t)$ with respect to $\mathbf{q}$ and $\mathbf{z}$, we can easily compute its derivative with respect to t:

$$\frac{\mathrm{d}}{\mathrm{d}t} F_n(\mathbf{q}^*, \mathbf{z}^*; t) = \frac{\partial}{\partial t} F_n(\mathbf{q}^*, \mathbf{z}^*; t) = \operatorname{Tr} \mathbf{z}^*(t) = n z^*(t). \tag{13.73}$$

Note that we can follow the value of $F_n(\mathbf{q}^*, \mathbf{z}^*; t)$ through the critical point t_c because $F_n(\mathbf{q}^*, \mathbf{z}^*; t)$ is continuous at that point (even if its derivative is not).

The above analysis therefore allows us to find, for $n \geq 1$,

$$\log \mathbb{E}[Z_t^n] \sim \frac{Nn}{2} \mathcal{F}_n(t), \tag{13.74}$$

where

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{F}_n(t) = \begin{cases} t + \frac{1}{t} & \text{for } t \leq t_c(n), \\ \frac{n^2 t + (n-2)\sqrt{(n^2 t^2 - 4(n-1))}}{2(n-1)} & \text{for } t > t_c(n), \end{cases} \tag{13.75}$$

with the boundary condition $\mathcal{F}_n(0) = 0$.

We can now analytically continue this solution down to $n \to 0$. The first regime, for small t, is easy as it does not depend on n. In the large t regime, the extrapolation of $z_+(t)$ to $n \to 0$ gives the very simple result (see Eq. (13.72)): $z_+ = 2$ for all t. The most tricky part is to find the critical point where one goes from the $z_0 = t + 1/t$ solution to the $z_+ = 2$ solution. We cannot analytically continue $t_c(n)$ to $n \to 0$, as we have no explicit formula for it. On the other hand, we can directly find the point $t_c(n = 0)$ at which the two solutions

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

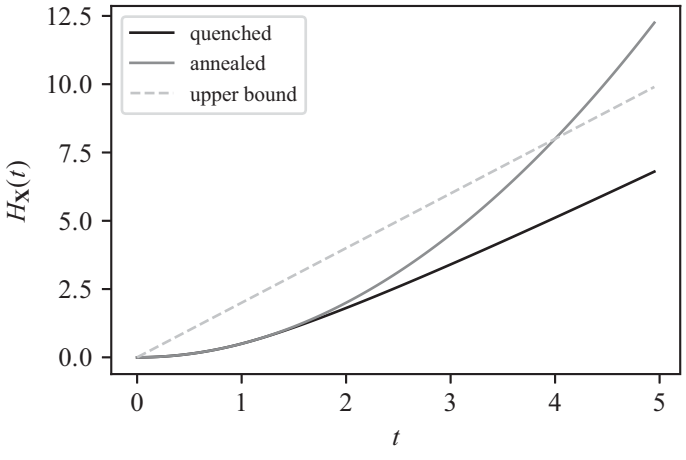

Figure 13.2 The function $H_{\mathbf{X}}(t)$ for a unit Wigner computed with a "quenched" average (HCIZ integral) and an "annealed" average. We also show the upper bound given by Eq. (10.59). The annealed and quenched averages are identical up to $t = t_c = 1$ and differ for larger t. The annealed average violates the bound, which is expected as in this case $\lambda_{\max}$ fluctuates and exceptionally large values of $\lambda_{\max}$ dominate the average exponential HCIZ integral.

lead to the same $\mathcal{F}_0$. It is relatively straightforward to show that this point is $t_c = 1$.[4] Correspondingly,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}_0(t) = \begin{cases} t + \frac{1}{t} & \text{for } t \leq t_c(0) = 1, \\ 2 & \text{for } t > t_c(0) = 1. \end{cases} \tag{13.76}$$

We can now go back to the definition of the function $\mathbb{E}[H_{\mathbf{X}}(t)]$ (Eq. (13.54)). After taking the $n \to 0$ and $N \to \infty$ limits we find (see Fig. 13.2)

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[H_{\mathbf{X}}(t)] = \begin{cases} t & \text{for } t \leq 1, \\ 2 - \frac{1}{t} & \text{for } t > 1, \end{cases} \tag{13.77}$$

with the condition $\mathbb{E}[H_{\mathbf{X}}(t = 0)] = 0$.

The upshot of this (rather complex) calculation is that, as announced in the introduction, for $t \leq t_c = 1$ the quenched and annealed results coincide, i.e. $\hat{H}_{\mathrm{wig}}(t) = \mathbb{E}[H_{\mathbf{X}}(t)]$. For $t > t_c$, on the other hand, the two quantities are different. The reason is that for sufficiently large values of t, the average HCIZ integral becomes dominated by very rare Wigner matrices that happen to have a largest eigenvalue significantly larger than the Wigner semi-circle edge $\lambda = 2$. This allows $\hat{H}_{\mathrm{wig}}(t)$ to continue its quadratic progression, while $\mathbb{E}[H_{\mathbf{X}}(t)]$ is dominated by the edge of the average spectrum and its growth with t is contained (see Fig. 13.2). When one computes higher and higher moments of the HCIZ

[4] Something peculiar happens as $n \to 0$, namely the minimum solution becomes the maximum one and vice versa. In other words for small t, where we know that z_0 is the right solution, we have $\mathcal{F}_0(z_0; t) < \mathcal{F}_0(z_+; t)$ and the opposite at large t. This paradox is always present within the replica method when $n \to 0$.

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

integral (i.e. as the number of replicas n increases), the dominance of extreme eigenvalues becomes more and more acute, leading to a smaller and smaller transition point $t_c(n)$.[5]

13.4 Spin-Glasses, Replicas and Low-Rank HCIZ

"Spin-glasses" are disordered magnetic materials exhibiting a freezing transition as the temperature is reduced. Typical examples are silver–manganese (or copper–manganese) alloys, where the manganese atoms carry a magnetic spin and are randomly dispersed in a non-magnetic matrix. Contrary to a ferromagnet (i.e. usual magnets like permalloy), where all microscopic spins agree to point more or less in the same direction when the temperature is below a certain transition temperature (872 K for permalloy), spins in spin-glasses freeze, but the configuration they adopt is disordered, "amorphous", with zero net magnetization.

A simple model to explain the phenomenon is the following. The energy of N spins $S_i = \pm 1$ is given by

$$\mathcal{H}(\{S\}) = \frac{1}{2} \sum_{i,j=1}^{N} \mathbf{J}_{ij} S_i S_j, \tag{13.78}$$

where $\mathbf{J}$ is a random matrix, which we take to be drawn from a rotational invariant ensemble, i.e. $\mathbf{J} = \mathbf{O}\Lambda\mathbf{O}^T$ where $\mathbf{O}$ is chosen according to the (flat) Haar measure over $O(N)$ and Λ is a certain fixed diagonal matrix with $\tau(\Lambda) = 0$, such that any pair of spins is as likely to want to point in the same direction or in opposite directions. The simplest case corresponds to $\mathbf{J} = \mathbf{X}$, a Wigner matrix, in which case the spectrum of Λ is the Wigner semi-circle. This case corresponds to the celebrated Sherrington–Kirkpatrick (SK) model, but other cases have been considered in the literature as well.

The physical properties of the system are encoded in the average free energy F, defined as

$$F := -T\mathbb{E}_{\mathbf{J}}\left[\log Z\right]; \qquad Z := \sum_{\{S\}} \exp\left(\frac{\mathcal{H}(\{S\})}{T}\right), \tag{13.79}$$

where the partition function Z is obtained as the sum over all 2^N configurations of the N spins, and T is the temperature. One of the difficulties of the theory of spin-glasses is to perform the average over the interaction matrix $\mathbf{J}$ of the *logarithm* of Z. Once again, one can try to use the replica trick to perform this average, to wit,

$$\mathbb{E}_{\mathbf{J}}\left[\log Z\right] = \frac{\partial}{\partial n}\mathbb{E}_{\mathbf{J}}\left[Z^n\right]\Big|_{n=0}. \tag{13.80}$$

One then computes the right hand side for integer n and hopes that the analytic continuation to $n \to 0$ makes sense. Introducing n replicas of the system, one has

$$\mathbb{E}_{\mathbf{J}}\left[Z^n\right] = \mathbb{E}_{\mathbf{J}}\left[\sum_{\{S^1, S^2, \ldots, S^n\}} \exp\left(\frac{\sum_{\alpha=1}^{n}\sum_{i,j=1}^{N} \mathbf{J}_{ij} S_i^\alpha S_j^\alpha}{2T}\right)\right]. \tag{13.81}$$

Now, for a fixed configuration of all nN spins $\{S^1, S^2, \ldots, S^n\}$, the $N \times N$ matrix $\mathbf{K}_{ij}^{(n)} := \sum_{\alpha=1}^{n} S_i^\alpha S_j^\alpha / N$ is at most of rank n, which is small compared to N which we will

[5] A very similar mechanism is at play in Derrida's random energy model, see Derrida [1981].

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

take to infinity. So we have to compute a low-rank HCIZ integral, which is given by Eq. (10.45):

$$\mathbb{E}_{\mathbf{J}}\left[\exp\left(\frac{N}{2T}\,\mathrm{Tr}\,\mathbf{O}\Lambda\mathbf{O}^T\mathbf{K}^{(n)}\right)\right] \approx \exp\left(\frac{N}{2}\,\mathrm{Tr}\,H_{\mathbf{J}}(\mathbf{K}^{(n)}/T)\right), \tag{13.82}$$

where $H_{\mathbf{J}}$ is the anti-derivative of the R-transform of $\mathbf{J}$ (or Λ). Now the non-zero eigenvalues of the $N \times N$ matrix $\mathbf{K}^{(n)}$ are the same as those of the $n \times n$ matrix $\mathbf{Q}_{\alpha\beta} = \sum_{i=1}^{N} S_i^\alpha S_i^\beta/N$, called the *overlap* matrix, because its elements measure the similarity of the configurations $\{S^a\}$ and $\{S^b\}$. Hence, we have to compute

$$\mathbb{E}_{\mathbf{J}}\left[Z^n\right] = \sum_{\{S^1,S^2,\ldots,S^n\}} \exp\left(\frac{N}{2}\,\mathrm{Tr}_n\, H_{\mathbf{J}}(\mathbf{Q}/T)\right). \tag{13.83}$$

It should be clear to the reader that all the steps above are very close to the ones followed in Section 13.2.2. We continue in the same vein, introducing a new $n \times n$ matrix for imposing the constraint $\mathbf{Q}_{\alpha\beta} = \sum_{i=1}^{N} S_i^\alpha S_i^\beta/N$:

$$1 = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{N^{n(n+1)/2}\mathrm{d}\mathbf{Y}}{2^{3n/2}\pi^{n/2}} \int \mathrm{d}\mathbf{Q} \exp\left(-N\,\mathrm{Tr}_n\,\mathbf{QY} + \sum_{\alpha,\beta=1}^{n} \mathbf{Y}_{\alpha,\beta} \sum_{i=1}^{N} S_i^\alpha S_i^\beta\right). \tag{13.84}$$

The nice thing about this representation is that sums over i become totally decoupled. So we get

$$\mathbb{E}_{\mathbf{J}}\left[Z^n\right] = C \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \mathrm{d}\mathbf{Y} \int \mathrm{d}\mathbf{Q} \exp\left(\frac{N}{2}\,\mathrm{Tr}_n\, H_{\mathbf{J}}(\mathbf{Q}/T) - N\,\mathrm{Tr}_n\,\mathbf{QY} + N\Xi(\mathbf{Y})\right), \tag{13.85}$$

where C is an irrelevant constant and

$$\Xi(\mathbf{Y}) := \log \mathcal{Z} \quad \text{with} \quad \mathcal{Z} := \left[\sum_S \mathrm{e}^{\sum_{\alpha,\beta=1}^{n} \mathbf{Y}_{\alpha,\beta} S^\alpha S^\beta}\right]. \tag{13.86}$$

In the large N limit, Eq. (13.85) can be estimated using a saddle point method over $\mathbf{Y}$ and $\mathbf{Q}$. As in Section 13.2.2 the first equation reads

$$\mathbf{Y}_{\alpha\beta} = \frac{1}{2T}[R_{\mathbf{J}}(\mathbf{Q}/T)]_{\alpha\beta}, \tag{13.87}$$

and taking derivatives with respect to $\mathbf{Y}_{\alpha\beta}$, we find

$$\mathbf{Q}_{\alpha\beta} = \frac{1}{\mathcal{Z}} \sum_S S^\alpha S^\beta \mathrm{e}^{\sum_{\alpha',\beta'=1}^{n} \mathbf{Y}_{\alpha',\beta'} S^{\alpha'} S^{\beta'}}, \tag{13.88}$$

which leads to the following self-consistent equation for $\mathbf{Q}$:

$$\mathbf{Q}_{\alpha\beta} = \frac{\sum_S S^\alpha S^\beta \mathrm{e}^{\frac{1}{2T}\sum_{\alpha',\beta'=1}^{n} [R_{\mathbf{J}}(\mathbf{Q}/T)]_{\alpha'\beta'} S^{\alpha'} S^{\beta'}}}{\sum_S \mathrm{e}^{\frac{1}{2T}\sum_{\alpha',\beta'=1}^{n} [R_{\mathbf{J}}(\mathbf{Q}/T)]_{\alpha'\beta'} S^{\alpha'} S^{\beta'}}} := \langle S^\alpha S^\beta \rangle_T. \tag{13.89}$$

At sufficiently high temperatures, one can expect the solution of these equations to be "replica symmetric", i.e.

$$\mathbf{Q}_{\alpha\beta} = \delta_{\alpha\beta}(1-q) + q. \tag{13.90}$$

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

This matrix has two eigenvalues, one non-degenerate equal to $1 + (n-1)q$ and another $(n-1)$-fold degenerate equal to $1-q$. Correspondingly, $R_{\mathbf{J}}(\mathbf{Q})$ has eigenvalues $R_{\mathbf{J}}((1+(n-1)q)/T)$ and $R_{\mathbf{J}}((1-q)/T)$, from which we reconstruct the diagonal and off-diagonal elements of $R_{\mathbf{J}}(\mathbf{Q})$:

$$\begin{aligned} r &:= [R_{\mathbf{J}}(\mathbf{Q})]_{\alpha\beta} = \frac{1}{n}\left[R_{\mathbf{J}}((1+(n-1)q)/T) - R_{\mathbf{J}}((1-q)/T)\right], \\ r_d &:= [R_{\mathbf{J}}(\mathbf{Q})]_{\alpha\alpha} = R_{\mathbf{J}}((1-q)/T) + r. \end{aligned} \tag{13.91}$$

Injecting in the definition of $\mathcal{Z}$, we find

$$\mathcal{Z} = \sum_S \exp\left(\frac{1}{2T}\left[nr_d + r\sum_{\alpha\neq\beta=1}^{n} S^\alpha S^\beta\right]\right), \tag{13.92}$$

where we have used $S_\alpha^2 \equiv 1$. Writing $\sum_{\alpha\neq\beta=1}^n S^\alpha S^\beta = \left(\sum_\alpha^n S^\alpha\right)^2 - n$ and using a Hubbard–Stratonovich transformation, one gets

$$\mathcal{Z} = \sum_S \mathrm{e}^{n\frac{r_d-r}{2T}} \int_{-\infty}^{+\infty} \mathrm{d}x \exp\left(-\frac{x^2}{2} + x\sqrt{\frac{r}{T}}\sum_\alpha^n S^\alpha\right). \tag{13.93}$$

The sums of different S^α now again decouple, leading to

$$\mathcal{Z} = \mathrm{e}^{n\frac{r_d-r}{2T} + n\log 2} \int_{-\infty}^{+\infty} \mathrm{d}x \frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \cosh^n\left[x\sqrt{\frac{r}{T}}\right]. \tag{13.94}$$

Now, one can notice that

$$\sum_{\alpha\neq\beta=1}^{n} \langle S^\alpha S^\beta\rangle_T = n(n-1)q = 2T\frac{\partial \log \mathcal{Z}}{\partial r}, \tag{13.95}$$

to get, in the limit $n \to 0$ and a few manipulations (including an integration by parts), an equation involving only q:

$$q = \int_{-\infty}^{+\infty} \frac{\mathrm{d}x}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \tanh^2\left[x\sqrt{\frac{r}{T}}\right], \tag{13.96}$$

where, in the limit $n \to 0$,

$$r = \frac{q}{T} R_{\mathbf{J}}'\left(\frac{1-q}{T}\right). \tag{13.97}$$

Clearly, $q = 0$ is always a solution of this equation. The physical interpretation of q is the following: choose randomly two microscopic configurations of spins $\{S_i^\alpha\}$ and $\{S_i^\beta\}$, each with a weight given by $\exp\left(\frac{\mathcal{H}(\{S\})}{T}\right)/Z$. Then, the average overlap between these configurations, $\sum_{i=1}^N S_i^\alpha S_i^\beta/N$, is equal to q. When $q = 0$, these two configurations are thus uncorrelated. One expects this to be the case at high enough temperature, where the system explores randomly all configurations.

When the spins start freezing, on the other hand, one expects the system to strongly favor some (amorphous) configurations over others. Hence, one expects that $q > 0$

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

in the spin-glass phase. Expanding the right hand side of Eq. (13.96) for small q gives

$$\int_{-\infty}^{+\infty} \frac{\mathrm{d}x}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \tanh^2\left[x\sqrt{\frac{r}{T}}\right] = q\frac{R'_{\mathbf{J}}(1/T)}{T^2} - q^2\left(\frac{R''_{\mathbf{J}}(1/T)}{T^3} + 2\left(\frac{R'_{\mathbf{J}}(1/T)}{T^2}\right)^2\right) + O(q^3). \tag{13.98}$$

Assuming that the coefficient in front of the q^2 term is negative, we see that a non-zero q solution appears continuously below a critical temperature T_c given by

$$T_c^2 = R'_{\mathbf{J}}\left(\frac{1}{T_c}\right). \tag{13.99}$$

When $\mathbf{J}$ is a Wigner matrix, the spin-glass model is the one studied originally by Sherrington and Kirkpatrick in 1975. In this case, $R_{\mathbf{J}}(x) = x$ and therefore $T_c = 1$. There are cases, however, where the transition is discontinuous, i.e. where the overlap q jumps from zero for $T > T_c$ to a non-zero value at T_c. In these cases, the small q expansion is unwarranted and another method must be used to find the critical temperature. One example is the "random orthogonal model", where the coupling matrix $\mathbf{J}$ has $N/2$ eigenvalues equal to $+1$ and $N/2$ eigenvalues equal to -1.

The spin-glass phase $T < T_c$ is much more complicated to analyze, because the replica symmetric ansatz $\mathbf{Q}_{\alpha\beta} = \delta_{\alpha\beta}(1-q) + q$ is no longer valid. One speaks about "replica symmetry breaking", which encodes an exquisitely subtle, hierarchical organization of the phase space of these models. This is the physical content of the celebrated Parisi solution of the SK model, but is much beyond the scope of the present book, and we encourage the curious reader to learn more from the references given below.

Bibliographical Notes

- The replica trick was introduced by
  - R. Brout. Statistical mechanical theory of a random ferromagnetic system. *Physical Review*, 115:824–835, 1959,

  and popularized in
  - S. F. Edwards and P. W. Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965, 1975,

  for an authoritative book on the subject, see
  - M. Mézard, G. Parisi, and M. A. Virasoro. *Spin Glass Theory and Beyond*. World Scientific, Singapore, 1987,

  see also
  - M. Mézard and G. Parisi. Replica field theory for random manifolds. *Journal de Physique I, France*, 1(6):809–836, 1991.
- The replica method was first applied to random matrices by
  - S. F. Edwards and R. C. Jones. The eigenvalue spectrum of a large symmetric random matrix. *Journal of Physics A: Mathematical and General*, 9(10):1595, 1976,

  it is rarely discussed in random matrix theory books with the exception of

https://doi.org/10.1017/9781108768900.016 Published online by Cambridge University Press

  - G. Livan, M. Novaes, and P. Vivo. *Introduction to Random Matrices: Theory and Practice*. Springer, New York, 2018.
- For a replica computation for the average resolvent for free addition and free product, see
  - J. Bun, R. Allez, J.-P. Bouchaud, and M. Potters. Rotational invariant estimator for general noisy matrices. *IEEE Transactions on Information Theory*, 62:7475–7490, 2016,
  - J. Bun, J.-P. Bouchaud, and M. Potters. Cleaning large correlation matrices: Tools from random matrix theory. *Physics Reports*, 666:1–109, 2017.
- For Derrida's random energy model where different moments of Z have different transition temperatures, see
  - B. Derrida. Random-energy model: An exactly solvable model of disordered systems. *Physical Review B*, 24:2613–2626, 1981.
- For an introduction to experimental spin-glasses, see
  - K. H. Fischer and J. A. Hertz. *Spin Glasses*. Cambridge University Press, Cambridge, 1991,
  - A. P. Young. *Spin Glasses and Random Fields*. World Scientific, Singapore, 1997,

  see also M. Mézard, G. Parisi, M. Virasoro, *op. cit.*
- For general orthogonal invariant spin-glasses, see
  - R. Cherrier, D. S. Dean, and A. Lefèvre. Role of the interaction matrix in mean-field spin glass models. *Physical Review E*, 67:046112, 2003.