Lecture notes on the replica method for Wishart matrix eigenvalues

Jacob A. Zavatone-Veth*

September 26, 2022

Abstract

In these lecture notes, we introduce the *replica method* from the statistical mechanics of disordered systems as a powerful heuristic tool for studying models that incorporate random data or parameters. We take as an illustrative example the computation of the minimum and maximum eigenvalues of an infinitely large Wishart random matrix, with the goal of providing step-by-step introduction to the basics of the replica method in cases where no replica symmetry breaking occurs.

Contents

*Department of Physics, Harvard University, Cambridge MA 02138, jzavatoneveth@g.harvard.edu

1 Background

In statistical physics and machine learning, one often encounters probability distributions or optimization problems that depend on external, fixed data. For example, consider least-squares linear regression, with cost function

$$E(\mathbf{w}; \mathbf{X}, \mathbf{y}) = \frac{1}{2}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2 \tag{1}$$

on weights $\mathbf{w} \in \mathbb{R}^d$, for covariates $\mathbf{X} \in \mathbb{R}^{p \times d}$ and targets $\mathbf{y} \in \mathbb{R}^p$. Or, consider the energy function of a ferromagnet with random defects in its lattice structure, which one might model by a system of spins $\mathbf{s} \in \{-1, +1\}^n$ with energy function defined by a sum of interactions of pairs of spins with couplings J_{ij}:

$$E(\mathbf{s}; \mathbf{J}) = -\sum_{i,j}^{n} J_{ij} s_i s_j. \tag{2}$$

In null or toy models, these fixed data are often taken to be random samples from some probability distribution, yielding an ensemble of systems with different realizations of these data. How should one characterize the behavior of this ensemble? Even evaluating the *average* behavior over random data is generally extremely challenging. In the 1970s, statistical physicists working on a class of systems known as spin glasses developed a heuristic procedure for computing these averages in the limit of large system size (Anderson 1988; Edwards and Anderson 1975). In the half-century since its introduction, this procedure, known as the *replica method* or *replica trick*, has proved enduringly useful in applications far removed from its physical roots. Here, we provide an introduction to the replica method through one of its statistical applications: the minimum and maximum eigenvalues of a random matrix.

2 Preliminaries

To avoid confusion, we briefly record our notational conventions, and warn the reader of impending abuses thereof. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we let $\mathbf{x} \cdot \mathbf{y} = \sum_{j=1}^{d} x_j y_j$ denote the Euclidean inner product, and $\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$ the corresponding norm. Let $\mathbb{S}^{d-1}(\sqrt{d}) = \{\mathbf{w} \in \mathbb{R}^d : \|\mathbf{w}\| = \sqrt{d}\}$ be the $(d-1)$-sphere of radius $\sqrt{d}$. We denote the $n \times n$ identity matrix by $\mathbf{I}_n$, and the n-dimensional vector with all elements equal to one by $\mathbf{1}_n$. We will freely interchange differentiation and integration in several places. As we will note later, this is the least of our worries in terms of mathematical rigor.

3 Wishart matrix eigenvalues and PCA of random Gaussian data

We introduce the replica method through one application to random matrix theory: determining the asymptotic statistics of the minimum and maximum eigenvalues of a Wishart random matrix. Our choice of Wishart eigenvalues as a motivating example was inspired in part by a review article by Advani, Lahiri, and Ganguli (2013), but our presentation will be distinct. For a statistician's take on this physics-style computation, see a series of blog posts by Mei (2019).

We will consider standard Wishart matrices $\mathbf{M} \in \mathbb{R}^{d \times d}$ with p degrees of freedom, which for general p can be defined as follows: Consider p independent and identically distributed samples $\mathbf{x}_\mu \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ from the standard d-dimensional Gaussian distribution. Then, their empirical uncentered covariance

$$\mathbf{M} = \frac{1}{p} \sum_{\mu=1}^{p} \mathbf{x}_\mu \mathbf{x}_\mu^\top \tag{3}$$

follows a standard Wishart distribution with p degrees of freedom:

$$\mathbf{M} \sim \mathcal{W}_d\left(\frac{1}{p}\mathbf{I}_d, p\right). \tag{4}$$

In general, the Wishart distribution is defined by choosing $\mathbf{x}_\mu$ to be independent and identically distributed samples from an anisotropic Gaussian; here we focus on the isotropic case for simplicity.

We will focus on the high-dimensional asymptotic limit $d, p \to \infty$ with $p/d \to \alpha \in (0, \infty)$. In keeping with physics terminology, we will refer to this as the *thermodynamic limit*. As α is the ratio of the number of samples to the number of dimensions, we will refer to it as the *data density*. The objective of studying this limit is to gain intuition for the behavior of sample covariance matrices of high-dimensional data, for which Wishart matrices are a null model.

Our goal is to compute the statistics of the minimum and maximum eigenvalues of $\mathbf{M}$, which we denote by $\lambda_{\min}(\mathbf{M})$ and $\lambda_{\max}(\mathbf{M})$, respectively. We will use the min-max characterization of these eigenvalues as Rayleigh quotients:

$$\lambda_{\min}(\mathbf{M}) = \min_{\mathbf{w}\in\mathbb{R}^d, \|\mathbf{w}\|=1} \mathbf{w}^\top \mathbf{M}\mathbf{w}, \qquad \lambda_{\max}(\mathbf{M}) = \max_{\mathbf{w}\in\mathbb{R}^d, \|\mathbf{w}\|=1} \mathbf{w}^\top \mathbf{M}\mathbf{w}. \tag{5}$$

As $\mathbf{M}$ is symmetric and positive semidefinite, we have

$$0 \leq \lambda_{\min}(\mathbf{M}) \leq \lambda_{\max}(\mathbf{M}). \tag{6}$$

As $\mathbf{M}$ has rank at most $\min\{d, p\}$, we have $\lambda_{\min}(\mathbf{M}) = 0$ if $p < d$, i.e., if $\alpha < 1$. If $\alpha > 1$, we expect $\mathbf{M}$ to be full rank with probability one, and for $\lambda_{\min}$ to be positive. In Appendix A, we make these intuition more precise, and prove additional simple bounds on $\lambda_{\min}$ and $\lambda_{\max}$.

Our study of the minimum and maximum eigenvalues is motivated by the fact that $\lambda_{\max}(\mathbf{M})$ has a natural interpretation in the context of unsupervised learning (Advani, Lahiri, and Ganguli 2013; Johnstone 2001). Concretely, consider *principal component analysis (PCA)*, which seeks the direction that maximises the projected variance:

$$\mathbf{w}_{\text{PCA}} = \operatorname*{arg\,max}_{\mathbf{w}\in\mathbb{R}^d, \|\mathbf{w}\|=1} \frac{1}{p}\sum_{\mu=1}^{p} (\mathbf{w}\cdot\mathbf{x}_\mu)^2. \tag{7}$$

Expanding the square, it is easy to see by comparison with (5) that $\mathbf{w}_{\text{PCA}}$ is the eigenvector of $\mathbf{M}$ corresponding to $\lambda_{\max}(\mathbf{M})$. Then, $\lambda_{\max}(\mathbf{M})$ is the variance captured by the first principal component. Therefore, the asymptotics of Wishart matrix eigenvalues provide a null model for the behavior of PCA in high dimensions (Johnstone 2001).

4 Statistical mechanics formulation

To employ the replica trick, we first need to frame our analysis as a statistical mechanics problem. We will first set up the problem for the minimum eigenvalue, and then extend that setup to allow us to compute the maximum eigenvalue.

We introduce a Gibbs distribution at inverse temperature $\beta > 0$ over vectors in $\mathbb{S}^{d-1}(\sqrt{d})$, with density

$$p(\mathbf{w}; \beta, \mathbf{M}) = \frac{1}{Z(\beta, \mathbf{M})} \exp[-\beta E(\mathbf{w}, \mathbf{M})]. \tag{8}$$

Here,

$$E(\mathbf{w};\mathbf{M}) = \frac{1}{2}\mathbf{w}^\top \mathbf{M}\mathbf{w} \tag{9}$$

is the energy function associated to the maximization problem (5), and the normalization factor

$$Z(\beta,\mathbf{M}) = \int_{\mathbb{S}^{d-1}(\sqrt{d})} d\mathbf{w}\, \exp[-\beta E(\mathbf{w};\mathbf{M})] \tag{10}$$

is the *partition function*. Here, we have scaled the energy (9) such that it is extensive (i.e., scales linearly) in the problem dimensionality d, and included a factor of $1/2$ for later notational convenience. From a physical perspective, we can view this Gibbs distribution as describing the equilibrium statistics of a system defined by a quadratic energy function with random interactions.[1]

As $\beta \to \infty$, the Gibbs distribution (8) will concentrate on the ground state of (9), which is the eigenvector of $\mathbf{M}$ corresponding to its minimum eigenvalue. We denote averages with respect to the Gibbs distribution (8) by $\langle\cdot\rangle_{\beta,\mathbf{M}}$. Then, recalling our definition of E in (9) and the Rayleigh quotient (5), we have

$$\lambda_{\min}(\mathbf{M}) = \lim_{\beta\to\infty} \frac{2}{d}\langle E\rangle_{\beta,\mathbf{M}} = \lim_{\beta\to\infty} \frac{\partial f(\beta,\mathbf{M})}{\partial\beta}, \tag{11}$$

where we have defined the *reduced free energy per site*

$$f(\beta,\mathbf{M}) = -\frac{2}{d}\log Z(\beta,\mathbf{M}). \tag{12}$$

We can also use this setup to compute the maximum eigenvalue if we allow for negative temperatures. We can see that the computation of the maximum eigenvalue is identical up to a sign, and that

$$\lambda_{\max}(\mathbf{M}) = -\lim_{\beta\to\infty} \frac{2}{d}\langle E\rangle_{-\beta,\mathbf{M}} = \lim_{\beta\to\infty} \frac{\partial f(-\beta,\mathbf{M})}{\partial\beta}. \tag{13}$$

On general grounds, we expect $f(\beta,\mathbf{M})$ to be *self-averaging* in the high-dimensional limit, or, in mathematical terms, for it to *concentrate*. Intuitively, this means that the physical properties of a given realization of the system, with a particular draw of the random matrix $\mathbf{M}$, will with high probability be the same as the average over $\mathbf{M}$. More precisely, we have

$$\lim_{d,p\to\infty} f(\beta,\mathbf{M}) = \lim_{d,p\to\infty} \mathbb{E}_{\mathbf{M}} f(\beta,\mathbf{M}) \tag{14}$$

with probability one over the distribution of $\mathbf{M}$. Importantly, the partition function Z itself does not concentrate. This is illustrated by the fact that the *annealed approximation* $\log Z(\beta,\mathbf{M}) \approx \log \mathbb{E}_{\mathbf{M}} Z(\beta,\mathbf{M})$ yields incorrect predictions for the minimum and maximum eigenvalues, which we show explicitly in Appendix B.

As they are given in terms of derivatives of the free energy, we similarly expect $\lambda_{\min}$ and $\lambda_{\max}$ to concentrate, i.e., to have

$$\lim_{d,p\to\infty} \lambda_{\min}(\mathbf{M}) = \lim_{d,p\to\infty} \mathbb{E}_{\mathbf{M}}\lambda_{\min}(\mathbf{M}) \quad \text{and} \quad \lim_{d,p\to\infty} \lambda_{\max}(\mathbf{M}) = \lim_{d,p\to\infty} \mathbb{E}_{\mathbf{M}}\lambda_{\max}(\mathbf{M}) \tag{15}$$

with probability one over the distribution of $\mathbf{M}$. For Wishart matrices, one can directly prove concentration of $\lambda_{\min}$ and $\lambda_{\max}$ rigorously (Geman 1980; Silverstein 1985).

[1] To be precise, this is a spherical spin glass model with a Wishart-distributed interaction matrix. For Gaussian-distributed interactions, this model was studied by Kosterlitz, Thouless, and Jones (1976).

5 Introducing the replica trick: the moments of the partition function

With the setup of the previous section, our goal is to compute

$$\lim_{d,p\to\infty} \mathbb{E}_{\mathbf{M}} f(\beta,\mathbf{M}) = -\lim_{d,p\to\infty} \frac{2}{d}\mathbb{E}_{\mathbf{M}} \log Z(\beta,\mathbf{M}), \tag{16}$$

but it is easy to see that averaging through the logarithm is hard. The *replica trick* gets around this obstacle by exploiting the useful identity relating moments of an almost-surely positive random variable to the expectation of its logarithm, assuming that this limit exists:[2]

$$\mathbb{E}_{\mathbf{M}} \log Z(\beta,\mathbf{M}) = \lim_{n\to 0} \frac{1}{n} \log \mathbb{E}_{\mathbf{M}} Z(\beta,\mathbf{M})^n. \tag{17}$$

We now make the first of several leaps of faith: we evaluate the moments $\mathbb{E}_{\mathbf{M}} Z(\beta,\mathbf{M})^n$ for non-negative integer n, and assume that we can later safely analytically continue to $n \to 0$. Evaluation of the moments for non-negative integer n allows us to interpret $\mathbb{E}_{\mathbf{M}} Z(\beta,\mathbf{M})^n$ as the partition function for n copies of the original physical system. The existence and uniqueness of an analytic continuation is not guaranteed in general,[3] and determining the choice of analytic continuation is a rich topic for study. Indeed, determining the correct analytic continuation for the Sherrington-Kirkpatrick spin glass model is among the contributions to science for which Giorgio Parisi received the 2021 Nobel Prize in Physics.[4] However, in the present case it turns out that no such complications arise, and the results obtained here can be proven rigorously.

5.1 Integrating out the data

With this setup, our first task is to evaluate the moments $\mathbb{E}_{\mathbf{M}} Z(\beta,\mathbf{M})^n$ for n a non-negative integer. Introducing replicas indexed by $a = 1,\dots,n$, we have

$$\mathbb{E}_{\mathbf{M}} Z^n = \int_{\mathbb{S}^{d-1}(\sqrt{d})} d\mathbf{w}^1 \cdots \int_{\mathbb{S}^{d-1}(\sqrt{d})} d\mathbf{w}^n \, \mathbb{E}_{\mathbf{M}} \exp\left(-\frac{\beta}{2}\sum_{a=1}^{n} (\mathbf{w}^a)^\top \mathbf{M}\mathbf{w}^a\right); \tag{18}$$

we will henceforth not explicitly write the domain of integration over $\mathbf{w}^a$. Expanding the definition of $\mathbf{M}$ and using the fact that the columns of $\mathbf{X}$ are independent and identically distributed, we have

$$\begin{aligned}
\mathbb{E}_{\mathbf{M}} \exp\left(-\frac{\beta}{2}\sum_{a=1}^{n} (\mathbf{w}^a)^\top \mathbf{M}\mathbf{w}^a\right) &= \mathbb{E}_{\mathbf{X}} \exp\left(-\frac{\beta}{2p}\sum_{a=1}^{n} (\mathbf{w}^a)^\top \mathbf{X}\mathbf{X}^\top \mathbf{w}^a\right) && (19)\\
&= \mathbb{E}_{\mathbf{X}} \exp\left[-\frac{\beta}{2p}\sum_{a=1}^{n} (\mathbf{w}^a)^\top \left(\sum_{\mu=1}^{p} \mathbf{x}_\mu \mathbf{x}_\mu^\top\right) \mathbf{w}^a\right] && (20)\\
&= \left\{\mathbb{E}_{\mathbf{x}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_d)} \exp\left[-\frac{1}{2}\mathbf{x}^\top \left(\frac{\beta}{p}\sum_{a=1}^{n} \mathbf{w}^a (\mathbf{w}^a)^\top\right)\mathbf{x}\right]\right\}^p && (21)\\
&= \det\left(\mathbf{I}_d + \frac{\beta}{p}\sum_{a=1}^{n} \mathbf{w}^a(\mathbf{w}^a)^\top\right)^{-p/2} && (22)
\end{aligned}$$

[2]In the context of quantum information theory, the replica trick often refers to the related identity $\operatorname{tr}(\rho\log\rho) = \lim_{n\to 1} \partial \operatorname{tr}(\rho^n)/\partial n$ for ρ a positive definite density matrix. There has recently been significant interest in applications of this variant of the replica trick, under the memorable name of "replica wormholes" (Almheiri et al. 2021).

[3]The relevant technical tool is Carlson's theorem; see for instance van Hemmen and Palmer (1979) or Ogure and Kabashima (2004).

[4]See Parisi's citation here: https://www.nobelprize.org/prizes/physics/2021/parisi/facts/.

after evaluating the Gaussian integral over $\mathbf{x}$. Using the Weinstein–Aronszajn identity (Horn and Johnson 2012), we can re-write this as a n-dimensional determinant

$$\det\left(\mathbf{I}_d + \frac{\beta}{p}\sum_{a=1}^{n}\mathbf{w}^a(\mathbf{w}^a)^\top\right) = \det\left(\mathbf{I}_n + \frac{\beta}{\alpha}\mathbf{Q}\right), \tag{23}$$

where we have defined the $n \times n$ overlap matrix

$$Q^{ab} = \frac{1}{d}\mathbf{w}^a \cdot \mathbf{w}^b. \tag{24}$$

Here, we remind the reader that $\alpha = p/d$. We have therefore obtained

$$\mathbb{E}_{\mathbf{M}} Z^n = \int_{\mathbb{S}^{d-1}(\sqrt{d})\times\cdots\times\mathbb{S}^{d-1}(\sqrt{d})} \prod_{a=1}^{n} d\mathbf{w}^a \det\left(\mathbf{I}_n + \frac{\beta}{\alpha}\mathbf{Q}\right)^{-p/2}. \tag{25}$$

From a physical perspective, averaging over the random matrix $\mathbf{M}$ has coupled together the initially non-interacting copies of the system, because the integrals over different $\mathbf{w}^a$ cannot be evaluated independently. Importantly, however, these interactions occur only through the overlaps $\mathbf{Q}$.

5.2 Introducing the order parameters

To allow us to integrate over the vectors $\{\mathbf{w}^a\}$, we will now employ a physicist's favorite trick: multiplying by one.[5] In less obscure terms, we want to convert the integral over the vectors $\{\mathbf{w}^a\}$ into an integral over the overlaps $\mathbf{Q}$, which can be accomplished by inserting Dirac δ-distributions to enforce the definition of each element and then integrating over $\mathbf{Q}$. As $\mathbf{w}^a \in \mathbb{S}^{d-1}(\sqrt{d})$, the diagonal elements of the overlap matrix are identically equal to one, i.e., $Q^{aa} = 1$. We then write

$$1 = \int d\mathbf{Q} \prod_{a,b=1}^{n} \delta\left(Q^{ab} - \frac{1}{d}\mathbf{w}^a \cdot \mathbf{w}^b\right), \tag{26}$$

where the integral is taken over all $n \times n$ real symmetric positive-semidefinite matrices with diagonal elements equal to unity. Of course, some of these constraints are redundant, but that is not an issue. Then, we have

$$\mathbb{E}_{\mathbf{M}} Z^n = \int d\mathbf{Q} \det\left(\mathbf{I}_n + \frac{\beta}{\alpha}\mathbf{Q}\right)^{-p/2} \int \prod_{a=1}^{n} d\mathbf{w}^a \prod_{a,b=1}^{n} \delta\left(Q^{ab} - \frac{1}{d}\mathbf{w}^a \cdot \mathbf{w}^b\right). \tag{27}$$

We observe that, since we're integrating over $\mathbf{Q}$ with $Q^{aa} = 1$, the δ-distributions automatically enforce the constraint that $\mathbf{w}^a \in \mathbb{S}^{d-1}(\sqrt{d})$, so we can extend the integrals over $\mathbf{w}^a$ to all of $\mathbb{R}^d$.

To make further progress, we use the Fourier representation of the δ-distribution,

$$\delta(x) = \int_{-i\infty}^{+i\infty} \frac{ds}{2\pi i} \exp(sx), \tag{28}$$

to write

$$\prod_{a,b=1}^{n} \delta\left(Q^{ab} - \frac{1}{d}\mathbf{w}^a \cdot \mathbf{w}^b\right) = \frac{1}{(4\pi i/d)^{n^2}} \int d\hat{\mathbf{Q}} \exp\left(\frac{d}{2}\operatorname{tr}(\mathbf{Q}\hat{\mathbf{Q}}) - \frac{1}{2}\sum_{a,b}\hat{Q}^{ab}\mathbf{w}^a \cdot \mathbf{w}^b\right), \tag{29}$$

[5] Particularly in the context of quantum mechanics, this is usually known as "inserting a resolution of the identity."

where the integral is taken over $n \times n$ imaginary symmetric matrices $\hat{\mathbf{Q}}$. In the physics literature, the variables $\hat{Q}^{ab}$ are often referred to as *Lagrange multipliers*, as they enforce the definitions of Q^{ab}. We note that the particular scaling we have chosen for the Lagrange multipliers will make subsequent steps easier. We can now factor and evaluate the integrals over the dimensions of the vectors $\{\mathbf{w}^a\}$:

$$\int \prod_{a=1}^{n} d\mathbf{w}^a \prod_{a,b=1}^{n} \delta\left(Q^{ab} - \frac{1}{d}\mathbf{w}^a \cdot \mathbf{w}^b\right) \tag{30}$$

$$= \frac{1}{(4\pi i/d)^{n^2}} \int d\hat{\mathbf{Q}} \exp\left(\frac{d}{2}\operatorname{tr}(\mathbf{Q}\hat{\mathbf{Q}})\right) \int \prod_{a=1}^{n} d\mathbf{w}^a \exp\left(-\frac{1}{2}\sum_{a,b}\hat{Q}^{ab}\mathbf{w}^a \cdot \mathbf{w}^b\right) \tag{31}$$

$$= \frac{1}{(4\pi i/d)^{n^2}} \int d\hat{\mathbf{Q}} \exp\left(\frac{d}{2}\operatorname{tr}(\mathbf{Q}\hat{\mathbf{Q}})\right) \left[\int \prod_{a=1}^{n} dw^a \exp\left(-\frac{1}{2}\sum_{a,b}\hat{Q}^{ab}w^a w^b\right)\right]^d \tag{32}$$

$$= \frac{(2\pi)^{nd/2}}{(4\pi i/d)^{n^2}} \int d\hat{\mathbf{Q}} \exp\left(\frac{d}{2}\operatorname{tr}(\mathbf{Q}\hat{\mathbf{Q}})\right) \det(\hat{\mathbf{Q}})^{-d/2}, \tag{33}$$

where we have used the fact that the Gaussian integral makes sense for almost all $\hat{\mathbf{Q}}$ despite the fact it is an imaginary matrix.

5.3 Saddle-point evaluation

Collecting the results of the previous subsection, we have obtained the expression

$$\mathbb{E}_{\mathbf{M}} Z^n = \frac{1}{(4\pi i/d)^{n^2}} \int d\mathbf{Q}\, d\hat{\mathbf{Q}} \exp\left(-\frac{d}{2} G(\mathbf{Q}, \hat{\mathbf{Q}})\right), \tag{34}$$

where

$$G(\mathbf{Q}, \hat{\mathbf{Q}}) = \alpha \log\det\left(\mathbf{I}_n + \frac{\beta}{\alpha}\mathbf{Q}\right) - \operatorname{tr}(\mathbf{Q}\hat{\mathbf{Q}}) + \log\det(\hat{\mathbf{Q}}) - n\log 2\pi. \tag{35}$$

At this point, it may appear that we have simply exchanged an intractable integral over the weights $\mathbf{w}^a$ for an intractable integral over the matrices $\mathbf{Q}$ and $\hat{\mathbf{Q}}$. To get around this obstacle, we now take a second jump, and assume that we can interchange the limit $n \to 0$ with the limit $d, p \to \infty$. That is, we assert that

$$\lim_{d,p\to\infty} \mathbb{E}_{\mathbf{M}} f(\beta, \mathbf{M}) = -\lim_{n\to 0}\lim_{d,p\to\infty} \frac{2}{nd} \log \mathbb{E}_{\mathbf{M}} Z(\beta, \mathbf{M})^n. \tag{36}$$

In general, this integral of limits is not justified; see for instance van Hemmen and Palmer (1979) and Ogure and Kabashima (2004) for some discussion of these issues.

If we take the thermodynamic limit $d, p \to \infty$ for fixed n, we can evaluate the integral over $\mathbf{Q}$ and $\hat{\mathbf{Q}}$ using the *method of steepest descent*. Roughly speaking, this allows us to approximate an integral with a sharply peaked integrand simply by evaluating the integrand at its maximal value.[6] This yields

$$\lim_{d\to\infty} -\frac{2}{d} \log \mathbb{E}_{\mathbf{M}} Z^n = \operatorname*{extr}_{\mathbf{Q},\hat{\mathbf{Q}}} G(\mathbf{Q}, \hat{\mathbf{Q}}), \tag{37}$$

[6] More formally, the method of steepest descent is a generalization of Laplace's method for asymptotic approximation of univariate real integrals to high-dimensional complex integrals. For a reasonably accessible introduction, see the Wikipedia page.

where the notation $\operatorname{extr}_{\mathbf{Q},\hat{\mathbf{Q}}}$ means that we should evaluate G at a saddle point, where

$$\frac{\partial G(\mathbf{Q},\hat{\mathbf{Q}})}{\partial \mathbf{Q}} = \mathbf{0} \quad \text{and} \quad \frac{\partial G(\mathbf{Q},\hat{\mathbf{Q}})}{\partial \hat{\mathbf{Q}}} = \mathbf{0}. \tag{38}$$

6 Solving the saddle-point equations under a replica-symmetric *Ansatz*

Using the result of the previous section, we have found that the disorder-averaged reduced free energy is given as

$$f = \lim_{n\to 0} \frac{1}{n} \operatorname*{extr}_{\mathbf{Q},\hat{\mathbf{Q}}} G(\mathbf{Q},\hat{\mathbf{Q}}) \tag{39}$$

in the thermodynamic limit. Our remaining task is to solve the saddle-point equations and analytically continue to the $n \to 0$ limit. To do so, we suppose that $\mathbf{Q}$ and $\hat{\mathbf{Q}}$ are of a special form. Heuristically, since we started out with identical copies of the system, it is reasonable to suppose that the overlaps between any two copies should be the same, hence we make the *Ansatz*

$$Q^{ab} = \begin{cases} 1 & a = b \\ q & a \neq b \end{cases} \tag{40}$$

or, in matrix form,

$$\mathbf{Q} = (1-q)\mathbf{I}_n + q\mathbf{1}_n\mathbf{1}_n^\top. \tag{41}$$

We assume that $\hat{\mathbf{Q}}$ is of a similar form:

$$\hat{\mathbf{Q}} = (\hat{Q}-\hat{q})\mathbf{I}_n + \hat{q}\mathbf{1}_n\mathbf{1}_n^\top. \tag{42}$$

This is known as the *replica-symmetric*, or RS, *Ansatz*. In simple cases—such as the problem at hand—the RS *Ansatz* gives the correct analytic continuation to $n \to 0$, but it does not do so in general. For more details, see Mézard, Parisi, and Virasoro (1987) or Engel and van den Broeck (2001). Under the RS *Ansatz*, we have

$$\frac{1}{n}\operatorname{tr}(\mathbf{Q}\hat{\mathbf{Q}}) = \hat{Q} - (1-n)q\hat{q} \tag{43}$$

and, by the matrix determinant lemma,

$$\det(\hat{\mathbf{Q}}) = (\hat{Q}-\hat{q})^n\left(1 + \frac{n\hat{q}}{\hat{Q}-\hat{q}}\right) \tag{44}$$

and

$$\det\left(\mathbf{I}_n + \frac{\beta}{\alpha}\mathbf{Q}\right) = \det\left[\left(1+\frac{\beta}{\alpha}(1-q)\right)\mathbf{I}_n + \frac{\beta q}{\alpha}\mathbf{1}_n\mathbf{1}_n^\top\right] \tag{45}$$

$$= \left(1+\frac{\beta}{\alpha}(1-q)\right)^n\left(1+n\frac{\beta q}{\alpha}\frac{1}{1+\beta(1-q)/\alpha}\right). \tag{46}$$

Thus,

$$f = \lim_{n\to 0}\operatorname*{extr}_{q,\hat{Q},\hat{q}}\left\{\alpha\log\left(1+\frac{\beta}{\alpha}(1-q)\right) + \frac{\alpha}{n}\log\left(1+n\frac{\beta q}{\alpha}\frac{1}{1+\beta(1-q)/\alpha}\right)\right.$$
$$\left. -\hat{Q} + (1-n)q\hat{q} + \log(\hat{Q}-\hat{q}) + \frac{1}{n}\log\left(1+\frac{n\hat{q}}{\hat{Q}-\hat{q}}\right)\right\} - \log(2\pi). \tag{47}$$

We now exchange the limit $n \to 0$ with the extremization over the replica-symmetric order parameters. In this simple setting, it is easy to explicitly justify the implied interchange of limits by first extremizing at finite n and then taking the limit $n \to 0$. This yields

$$f = \operatorname*{extr}_{q,\hat{Q},\hat{q}} \left\{ \alpha \log\left(1 + \frac{\beta}{\alpha}(1-q)\right) + \frac{\beta q}{1+\beta(1-q)/\alpha} - \hat{Q} + q\hat{q} + \log(\hat{Q}-\hat{q}) + \frac{\hat{q}}{\hat{Q}-\hat{q}} \right\} - \log(2\pi). \tag{48}$$

Extremizing with respect to $\hat{Q}$ and $\hat{q}$ gives

$$\hat{Q} - \hat{q} = \frac{1}{1-q} \qquad \text{and} \qquad \hat{q} = -\frac{q}{(1-q)^2}. \tag{49}$$

This yields

$$f = \operatorname*{extr}_{q} \left\{ \alpha \log[1+\omega(1-q)] + \alpha \frac{\omega q}{1+\omega(1-q)} - \log(1-q) - \frac{1}{1-q} \right\} - \log(2\pi), \tag{50}$$

where we have defined $\omega \equiv \beta/\alpha$ for brevity.

The minimum eigenvalues are then given by the $\beta \to \infty$ limit of

$$\frac{\partial f}{\partial \beta} = \frac{1+\omega(1-q)^2}{[1+\omega(1-q)]^2}, \tag{51}$$

where this expression should be evaluated at the extremizing value of q. The extremum condition is

$$0 = \frac{\alpha \omega^2 q}{[1+\omega(1-q)]^2} - \frac{q}{(1-q)^2}. \tag{52}$$

This clearly yields a trivial solution $q = 0$. If $\alpha \neq 1$, the remaining non-trivial solutions are determined by a quadratic equation, and are given as

$$q_\pm = 1 + \frac{1}{\omega} \frac{1}{1 \pm \sqrt{\alpha}}. \tag{53}$$

If $\alpha = 1$, there is a single solution in addition to $q = 0$,

$$q_1 = 1 + \frac{1}{2\omega}, \tag{54}$$

which coincides with q_+ at $\alpha = 1$, hence we need not consider it separately, and can simply extend q_+ to all $\alpha > 0$.

These candidate solutions yield energies

$$\left.\frac{\partial f}{\partial \beta}\right|_{q=0} = \frac{1}{1+\omega} \tag{55}$$

and

$$\left.\frac{\partial f}{\partial \beta}\right|_{q=q_\pm} = \left(1 \pm \frac{1}{\sqrt{\alpha}}\right)^2 + \frac{1}{\alpha\omega}, \tag{56}$$

and, at large $|\omega|$,

$$\left. f \right|_{q=0} = \alpha \log(1+\omega) + \mathcal{O}(1) \tag{57}$$

and

$$f\Big|_{q=q_\pm} = (1 \pm \sqrt{\alpha})^2 \omega + \mathcal{O}(1). \tag{58}$$

For $\omega > 0$, the solution q_+ is always greater than one, and is therefore unphysical because $q \leq 1$. Similarly, q_- is greater than one (and therefore unphysical) for $\alpha < 1$. This leaves us with $q = 0$ as the only physical solution for $\alpha \leq 1$, and $q = 0$ and $q = q_-$ as our candidates for $\alpha > 1$. From experience, we expect (somewhat counter-intuitively) the solution which maximizes f to be physical (Mézard, Parisi, and Virasoro 1987), hence $q = q_-$ for all $\alpha > 1$. Therefore, we have

$$\lambda_{\text{min}} = \begin{cases} 0, & \text{if } \alpha \leq 1 \\ \left(1 - \frac{1}{\sqrt{\alpha}}\right)^2, & \text{if } \alpha > 1. \end{cases} \tag{59}$$

We now consider the maximum eigenvalue. In this case, we must simply swap the sign of ω. This immediately means that $q = 0$ is unphysical, as it yields as value of $\partial f / \partial \beta$ that tends to zero from below as $\omega \to -\infty$ through negative values of ω. Similarly, q_- is unphysical for $\alpha > 1$. In this case, the physical solution is given by maximizing $-f$, hence $q = q_+$ for all $\alpha > 0$. Therefore, we conclude that

$$\lambda_{\text{max}} = \left(1 + \frac{1}{\sqrt{\alpha}}\right)^2. \tag{60}$$

In general, one would at this point have to check whether the replica-symmetric result is correct, or if replica symmetry breaking (RSB) must be taken into account. In some cases, one can appeal to general theorems that prove that replica symmetry should be unbroken (Barbier, Panchenko, and Sáenz 2021), while in other cases one must perform direct analytical or numerical checks (Mézard, Parisi, and Virasoro 1987; Parisi, Urbani, and Zamponi 2020). Thanks to Parisi, there is a generally accepted systematic procedure for obtaining the correct free energy in systems where RSB occurs. However, we will not address this rich subject in the present notes, as in this case no RSB occurs. Instead, we direct the interested reader to books by Mézard, Parisi, and Virasoro (1987) and Parisi, Urbani, and Zamponi (2020), and to references therein.

7 Conclusions and further reading

We have found that the minimum and maximum eigenvalues of a Wishart matrix (3) should tend almost surely in the thermodynamic limit to

$$\lambda_{\text{min}} = \begin{cases} 0, & \text{if } \alpha \leq 1 \\ \left(1 - \frac{1}{\sqrt{\alpha}}\right)^2, & \text{if } \alpha > 1. \end{cases} \tag{61}$$

and

$$\lambda_{\text{max}} = \left(1 + \frac{1}{\sqrt{\alpha}}\right)^2, \tag{62}$$

respectively. Though we have obtained these results through non-rigorous heuristics, they can be proved rigorously (Geman 1980; Johnstone 2001; Silverstein 1985).

From the perspective of PCA, these results represent a sort of illusory structure in high-dimensional data, as, though the covariance matrix is on average equal to the identity, the fluctuations in its elements are large enough that its eigenspectrum will not be flat on average. The full limiting eigenspectrum of a matrix of the form (3) is known as the Marchenko–Pastur distribution. This could be computed using the replica method, but the calculation is more complicated; we direct the interested reader to a review by Advani, Lahiri, and Ganguli (2013), a series of blog posts by Mei (2019), and a generalized version of this computation from our own work (Zavatone-Veth and Pehlevan 2022).

These notes have illustrated the recipe for a replica theory calculation, which has the following rough steps:

Step 0. Set up the calculation of interest as a statistical mechanics problem. Roughly speaking, this means expressing quantities of interest as derivatives of the free energy associated with some partition function, which itself is written as an integral over a high-dimensional state space.

Step 1. Write the integer moments of the partition function as an average over an integer number of replicas, and evaluate the data average.

Step 2. Determine the overlaps that describe the resulting interactions between replicas. Introduce these overlaps as integration variables using Fourier representations of the δ-distribution, and integrate out the original state variables. If all has gone well, one will at this point have an integral that can be evaluated using the method of steepest descent.

Step 3. Evaluate the saddle-point equations under a replica-symmetric *Ansatz*.

Step 4. Test the validity of the replica-symmetric predictions. If replica symmetry is unbroken, then one can declare victory. If it is broken, then one must consider replica symmetry breaking effects.

Though the replica method is a non-rigorous heuristic, it has proved to be a powerful method for generating conjectures in high-dimensional statistics. Some of these conjectures have been proven rigorously,[7] while others still defy rigorous treatment.[8] For a more detailed treatment of the replica method and further applications, we direct the reader to the classic monograph by Mézard, Parisi, and Virasoro (1987), and to more recent books by Engel and van den Broeck (2001), Mézard and Montanari (2009), Nishimori (2001), and Parisi, Urbani, and Zamponi (2020). For an accessible general introduction to random matrix theory, see the recent text by Potters and Bouchaud (2020). Finally, some recent applications of the replica method to machine learning theory are summarized in a review by Bahri et al. (2020).

Acknowledgements

These notes were prepared for a supplementary lecture I gave as part of the Fall 2022 iteration of Harvard Applied Math 226: Theory of Neural Computation, led by Cengiz Pehlevan. I thank Alexander Atanasov and Benjamin S. Ruben for helpful comments on previous versions.

[7]For instance, Parisi's formula for the free energy of the Sherrington-Kirkpatrick (SK) model was proved by Talagrand (2006), and the ultrametric structure of the overlaps in the SK model was proved by Panchenko (2013).

[8]Krauth and Mézard (1989)'s result for the capacity of the Ising perceptron has yet to be rigorously proved; the best partial result is due to Ding and Sun (2018).

A Elementary bounds on the limiting eigenvalues

In this appendix, we derive some elementary bounds on the limiting eigenvalues, which provide useful sanity checks of our replica-theory results. Using the definition (3) of a Wishart matrix in terms of Gaussian random vectors, we can write the Rayleigh quotients (5) as

$$\lambda_{\min}(\mathbf{M}) = \min_{\mathbf{w}\in\mathbb{R}^d, \|\mathbf{w}\|=1} \frac{1}{p}\sum_{\mu=1}^{p}(\mathbf{w}\cdot\mathbf{x}_\mu)^2 \quad \text{and} \quad \lambda_{\max}(\mathbf{M}) = \max_{\mathbf{w}\in\mathbb{R}^d, \|\mathbf{w}\|=1} \frac{1}{p}\sum_{\mu=1}^{p}(\mathbf{w}\cdot\mathbf{x}_\mu)^2. \tag{63}$$

We can then see that $\lambda_{\min}(\mathbf{M}) = 0$ if and only if the minimizing vector $\mathbf{w}$ is orthogonal to all $\mathbf{x}_\mu$, which is obviously possible if $p < d$. As a set of Gaussian random vectors is with probability one in general position (Vershynin 2018), we expect $\mathbf{M}$ to be full-rank if $p \geq d$, and for $\lambda_{\min}$ to be positive.

We can obtain elementary bounds on the eigenvalues by evaluating the Rayleigh quotients for fixed test vectors. Choosing a test vector $\mathbf{w} = (1, 0, \dots, 0)^\top$, we have

$$\lambda_{\min}(\mathbf{M}) \leq \frac{1}{p}\sum_{\mu=1}^{p}(x_{\mu,1})^2 \quad \text{and} \quad \lambda_{\max}(\mathbf{M}) \geq \frac{1}{p}\sum_{\mu=1}^{p}(x_{\mu,1})^2. \tag{64}$$

As $p \to \infty$, the strong law of large numbers implies that

$$\lim_{p\to\infty} \frac{1}{p}\sum_{\mu=1}^{p}(x_{\mu,1})^2 = 1 \tag{65}$$

with probability one (Vershynin 2018). We therefore have that

$$\lim_{p\to\infty} \lambda_{\min}(\mathbf{M}) \leq 1 \quad \text{and} \quad \lim_{p\to\infty} \lambda_{\max}(\mathbf{M}) \geq 1 \tag{66}$$

with probability one. This holds for any d, including in the simultaneous limit $d, p \to \infty$, as the left-hand-side of the bounds does not depend on d.

B Failure of the annealed approximation

In this appendix, we demonstrate the failure of the annealed approximation

$$f \approx f_{\text{annealed}} = -\lim_{d,p\to\infty} \frac{2}{d} \log \mathbb{E}_{\mathbf{M}} Z(\beta, \mathbf{M}) \tag{67}$$

to accurately predict the minimum and maximum eigenvalues. Evaluating the general formula for the moments of the partition function after integrating out the data from (25) at $n = 1$, we have

$$\mathbb{E}_{\mathbf{M}} Z = \int_{\mathbb{S}^{d-1}(\sqrt{d})} d\mathbf{w} \left(1 + \frac{\beta}{\alpha}\right)^{-p/2} \tag{68}$$

$$= \text{vol}[\mathbb{S}^{d-1}(\sqrt{d})] \left(1 + \frac{\beta}{\alpha}\right)^{-p/2} \tag{69}$$

using the fact that the single overlap $Q^{11} = 1$. This yields

$$f_{\text{annealed}} = \alpha \log\left(1 + \frac{\beta}{\alpha}\right) + C; \tag{70}$$

the constant $C = -\lim_{d\to\infty} \frac{2}{d} \log \mathrm{vol}[\mathbb{S}^{d-1}(\sqrt{d})]$ can be computed explicitly, but we will not do so as it does not affect the predicted eigenvalues. Then, the annealed approximation predicts

$$\lambda_{\min} \approx \lim_{\beta\to\infty} \frac{\partial f_{\text{annealed}}(\beta)}{\partial \beta} \tag{71}$$

$$= \lim_{\beta\to\infty} \frac{1}{1+\beta/\alpha} \tag{72}$$

$$= 0 \tag{73}$$

and, similarly, $\lambda_{\max} \approx 0$. Comparing these results with (61) and (62), we see that the annealed approximation for $\lambda_{\min}$ is correct if $\alpha \leq 1$ and incorrect otherwise, while the approximation for the maximum eigenvalue is always incorrect.

References

Advani, Madhu, Subhaneil Lahiri, and Surya Ganguli (2013). "Statistical mechanics of complex neural systems and high dimensional data". In: *Journal of Statistical Mechanics: Theory and Experiment* 2013.03, P03014. DOI: 10.1088/1742-5468/2013/03/P03014.

Almheiri, Ahmed, Thomas Hartman, Juan Maldacena, Edgar Shaghoulian, and Amirhossein Tajdini (July 2021). "The entropy of Hawking radiation". In: *Rev. Mod. Phys.* 93 (3), p. 035002. DOI: 10.1103/RevModPhys.93.035002.

Anderson, Philip W (1988). "Spin glass III: Theory raises its head". In: *Physics Today* 41.6, p. 9. DOI: 10.1063/1.2811440.

Bahri, Yasaman, Jonathan Kadmon, Jeffrey Pennington, Sam S. Schoenholz, Jascha Sohl-Dickstein, and Surya Ganguli (2020). "Statistical Mechanics of Deep Learning". In: *Annual Review of Condensed Matter Physics* 11.1, pp. 501–528. DOI: 10.1146/annurev-conmatphys-031119-050745.

Barbier, Jean, Dmitry Panchenko, and Manuel Sáenz (Dec. 2021). "Strong replica symmetry for high-dimensional disordered log-concave Gibbs measures". In: *Information and Inference: A Journal of the IMA*. iaab027. ISSN: 2049-8772. DOI: 10.1093/imaiai/iaab027.

Ding, Jian and Nike Sun (2018). "Capacity lower bound for the Ising perceptron". In: *arXiv*. DOI: 10.48550/ARXIV.1809.07742.

Edwards, Samuel F and Philip W Anderson (May 1975). "Theory of spin glasses". In: *Journal of Physics F: Metal Physics* 5.5, pp. 965–974. DOI: 10.1088/0305-4608/5/5/017.

Engel, Andreas and Christian van den Broeck (2001). *Statistical Mechanics of Learning*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139164542.

Geman, Stuart (1980). "A Limit Theorem for the Norm of Random Matrices". In: *The Annals of Probability* 8.2, pp. 252–261. DOI: 10.1214/aop/1176994775.

Horn, Roger A and Charles R Johnson (2012). *Matrix Analysis*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511810817.

Johnstone, Iain M. (2001). "On the distribution of the largest eigenvalue in principal components analysis". In: *The Annals of Statistics* 29.2, pp. 295–327. DOI: 10.1214/aos/1009210544.

Kosterlitz, J. M., D. J. Thouless, and Raymund C. Jones (May 1976). "Spherical Model of a Spin-Glass". In: *Phys. Rev. Lett.* 36 (20), pp. 1217–1220. DOI: 10.1103/PhysRevLett.36.1217.

Krauth, Werner and Marc Mézard (1989). "Storage capacity of memory networks with binary couplings". In: *Journal de Physique* 50.20, pp. 3057–3066. DOI: 10.1051/jphys:0198900500200305700.

Mei, Song (2019). *Replica method and random matrices*. online blog post. URL: https://meisong541.github.io/jekyll/update/2019/08/04/Replica_method_1.html and https://meisong541.github.io/jekyll/update/2019/09/12/Replica_method_2.html.

Mézard, Marc and Andrea Montanari (Jan. 2009). *Information, Physics, and Computation*. Oxford University Press. ISBN: 9780198570837. DOI: 10.1093/acprof:oso/9780198570837.001.0001.

Mézard, Marc, Giorgio Parisi, and Miguel Angel Virasoro (1987). *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*. World Scientific Publishing Company. DOI: https://doi.org/10.1142/0271.

Nishimori, Hidetoshi (July 2001). *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. Oxford University Press. ISBN: 9780198509417. DOI: 10.1093/acprof:oso/9780198509417.001.0001.

Ogure, Kenzo and Yoshiyuki Kabashima (May 2004). "Exact Analytic Continuation with Respect to the Replica Number in the Discrete Random Energy Model of Finite System Size". In: *Progress of Theoretical Physics* 111.5, pp. 661–688. ISSN: 0033-068X. DOI: 10.1143/PTP.111.661.

Panchenko, Dmitry (2013). "The Parisi ultrametricity conjecture". In: *Annals of Mathematics* 177.1, pp. 383–393. ISSN: 0003-486X. DOI: 10.4007/annals.2013.177.1.8.

Parisi, Giorgio, Pierfrancesco Urbani, and Francesco Zamponi (2020). *Theory of Simple Glasses: Exact Solutions in Infinite Dimensions*. Cambridge University Press. DOI: 10.1017/9781108120494.

Potters, Marc and Jean-Philippe Bouchaud (2020). *A First Course in Random Matrix Theory: for Physicists, Engineers and Data Scientists*. Cambridge University Press. DOI: 10.1017/9781108768900.

Silverstein, Jack W. (1985). "The Smallest Eigenvalue of a Large Dimensional Wishart Matrix". In: *The Annals of Probability* 13.4, pp. 1364–1368. DOI: 10.1214/aop/1176992819.

Talagrand, Michel (2006). "The Parisi Formula". In: *Annals of Mathematics* 163.1, pp. 221–263. ISSN: 0003486X. DOI: annals.2006.163.221.

van Hemmen, J L and R G Palmer (1979). "The replica method and solvable spin glass model". In: *Journal of Physics A: Mathematical and General* 12.4, pp. 563–580. DOI: 10.1088/0305-4470/12/4/016.

Vershynin, Roman (2018). *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge University Press. DOI: 10.1017/9781108231596.

Zavatone-Veth, Jacob A. and Cengiz Pehlevan (2022). "Replica method for eigenvalues of real Wishart product matrices". In: *arXiv*. DOI: 10.48550/ARXIV.2209.10499.