Expository notes on the pattern storage capacity of perceptrons with constrained weight distributions

Jacob A. Zavatone-Veth*

May 22, 2025

Abstract

In these expository notes, we seek to obtain a better understanding of Zhong et al. [9]'s results on the storage capacity of perceptrons subjected to constraints on their weight distributions by recapitulating their results using alternative methods of imposing the constraint.

These notes were originally drafted in June 2022. Since then, they have been proofread and edited for clarity.

Contents

*Society of Fellows and Center for Brain Science, Harvard University
jzavatoneveth@fas.harvard.edu

1 Introduction

We consider a simple perceptron

$$y(\mathbf{x}) = \operatorname{sign}\left(\frac{1}{\sqrt{N}}\mathbf{w}\cdot\mathbf{x}\right), \tag{1}$$

and the usual task of memorizing P independent and identically distributed symmetric binary patterns $\{(\mathbf{x}^\mu, y^\mu)\}$.

Our goal is to determine the storage capacity of the perceptron subject to the constraint that the empirical weight distribution

$$F_{\mathbf{w}}(w) = \frac{1}{N}\sum_{j=1}^{N}\Theta(w - w_j) \tag{2}$$

tends to some desired distribution $F(w)$ in the thermodynamic limit $N \to \infty$. In a recent paper, Zhong et al. [9] approach this problem by enforcing a hard constraint that the Fourier transforms of the empirical and target densities (that is, the characteristic functions of the two distributions) coincide, which if the target distribution is continuous is in general only satisfiable in the thermodynamic limit. In these notes, we will explore alternative approaches to this problem, which recover Zhong et al. [9]'s result using softer constraints.

We will always include the standard spherical constraint on the weight vector, i.e., we demand that $\|\mathbf{w}\|_2^2 = N$. This implies that the second moment of the target distribution must be equal to unity in order for the constraints to be consistent, as

$$\int_{-\infty}^{\infty} dF_{\mathbf{w}}(w)\, w^2 = \frac{1}{N}\sum_{j=1}^{N} w_j^2 = 1 \tag{3}$$

under the spherical constraint.

We will not consider replica symmetry breaking; nor did Zhong et al. [9]. To precisely determine the capacity, RSB must be considered, as Zhong et al. [9] illustrate through the example of a bimodal Gaussian target distribution. When the separation between the modes is large relative to their width, the resulting model resembles a perceptron with binary weights. Following Krauth and Mézard [6], RSB affects the capacity of the binary perceptron, though only at the one-step level. We leave a full analysis of how RSB affects the capacity of perceptrons with weights constrained to follow more general distributions to future work.

These notes are far from self-contained: we take the results of the standard Gardner capacity calculation as given, and we reference results from Zhong et al. [9] without exposition or proof. For the details of the standard Gardner capacity calculation, we refer to her original papers [3, 4]. Our notation will be standard, and will in particular hew closely to that in our own work on the storage capacity problem for nonlinear networks [8]. Throughout, we write $\varphi(t) = \exp(-t^2/2)/\sqrt{2\pi}$ for the standard Gaussian density, and $H(z) = \int_z^\infty dt\, \varphi(t)$ for the corresponding tail distribution function.

2 Moment constraints

We first consider the possibility of constraining the first $K \ll N$ moments of the weight distribution, i.e., we fix

$$\frac{1}{N}\sum_{j=1}^{N} w_j^k = m_k \qquad (k = 1, 2, \ldots, K); \tag{4}$$

we of course have $m_2 = 1$ from the spherical constraint. This constraint is well-defined at any finite N, but it is of course possible that it may not be satisfiable. At the end of the computation, we will consider taking $K \to \infty$ after

taking $N \to \infty$. This approach has the advantage of not requiring any path integrals, which we will encounter in our subsequent approach.

The constrained Gardner volume is in this case

$$Z = \int d\mathbf{w} \left[\prod_{k=1}^{K} \delta \left(\sum_{j=1}^{N} w_j^k - N m_k \right) \right] \prod_{\mu=1}^{P} \Theta \left(\frac{y^\mu \mathbf{x}^\mu \cdot \mathbf{w}}{\sqrt{N}} - \kappa \right). \tag{5}$$

2.1 Evaluating the disorder average

As usual, we will proceed via the replica method. The replicated, averaged Gardner volume is

$$\mathbb{E}_{\mathbf{x},y} Z^n = \int \prod_{a=1}^{n} d\mathbf{w}^a \left[\prod_{a=1}^{n} \prod_{k=1}^{K} \delta \left(\sum_{j=1}^{N} (w_j^a)^k - N m_k \right) \right] \left[\mathbb{E}_{\mathbf{x},y} \prod_{a=1}^{n} \Theta \left(\frac{y \mathbf{x} \cdot \mathbf{w}^a}{\sqrt{N}} - \kappa \right) \right]^P. \tag{6}$$

The data average in the square brackets is identical to that in the standard Gardner capacity calculation. We thus recall the standard result that, in the thermodynamic limit,

$$\mathbb{E}_{\mathbf{x},y} \prod_{a=1}^{n} \Theta \left(\frac{y \mathbf{x} \cdot \mathbf{w}^a}{\sqrt{N}} - \kappa \right) \to e^{n G_1(\mathbf{Q})}, \tag{7}$$

for

$$Q^{ab} = \frac{1}{N} \mathbf{w}^a \cdot \mathbf{w}^b \tag{8}$$

the Edwards-Anderson order parameters and

$$G_1(\mathbf{Q}) = \mathbb{E}_{\mathbf{h} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})} \mathbb{E}_y \prod_{a=1}^{n} \Theta(y h^a - \kappa) \tag{9}$$

the usual perceptron energetic term. Enforcing the definitions of these order parameters and the moment constraints via Fourier representations of the Dirac distribution, we have

$$\begin{aligned} \mathbb{E}_{\mathbf{x},y} Z^n = \int \prod_{a,k} \frac{d\hat{R}_k^a}{4\pi i / N} \int \prod_{b<a} \frac{dQ^{ab}\, d\hat{Q}^{ab}}{2\pi i / N} \exp \left(N \sum_{b<a} Q^{ab} \hat{Q}^{ab} + \frac{N}{2} \sum_{a=1}^{n} \sum_{k=1}^{K} \hat{R}_k^a m_k + P n G_1(\mathbf{Q}) \right) \\ \times \left[\int \prod_{a=1}^{n} dw^a \, \exp \left(- \sum_{b<a} \hat{Q}^{ab} w^a w^b - \frac{1}{2} \sum_{a=1}^{n} \sum_{k=1}^{K} \hat{R}_k^a (w^a)^k \right) \right]^N \end{aligned} \tag{10}$$

Here, we exclude the diagonal elements from $\mathbf{Q}$ and $\hat{\mathbf{Q}}$, and use $\hat{R}_2^a$ to enforce the spherical constraint. The entropic term is then

$$\begin{aligned} n G_2 = \sum_{b<a} Q^{ab} \hat{Q}^{ab} + \frac{1}{2} \sum_{a=1}^{n} \sum_{k=1}^{K} R_k^a m_k \\ + \log \int \prod_{a=1}^{n} dw^a \, \exp \left(- \sum_{b<a} \hat{Q}^{ab} w^a w^b - \frac{1}{2} \sum_{a=1}^{n} \sum_{k=1}^{K} \hat{R}_k^a (w^a)^k \right). \end{aligned} \tag{11}$$

2.2 Replica-symmetric saddle point equations

We make a replica-symmetric *Ansatz*

$$Q^{ab} = q \tag{12}$$
$$\hat{Q}^{ab} = -\hat{q} \tag{13}$$
$$\hat{R}_2^a = \hat{r}_2 + \hat{q} \tag{14}$$
$$\hat{R}_k^a = \hat{r}_k \qquad (k \neq 2). \tag{15}$$

The energetic term is identical to the standard Gardner calculation, and thus has $n \to 0$ limit

$$G_1^{\mathrm{RS}} = \mathbb{E}_t \log H\left(\frac{\kappa + t\sqrt{q}}{\sqrt{1-q}}\right), \tag{16}$$

where $t \sim \mathcal{N}(0,1)$. The entropic term simplifies to

$$G_2 = \frac{1}{2}(1-n)q\hat{q} - \frac{1}{2}\hat{q} + \frac{1}{2}\sum_{k=1}^{K}\hat{r}_k m_k + \frac{1}{n}\log\int\prod_{a=1}^{n}dw^a\,\exp\left(-\frac{1}{2}\sum_{a=1}^{n}\sum_{k=1}^{K}\hat{r}_k(w^a)^k - \frac{1}{2}\hat{q}\Big[\sum_a w^a\Big]^2\right) \tag{17}$$

which has $n \to 0$ limit

$$G_2 = \frac{1}{2}\sum_{k=1}^{K}\hat{r}_k m_k - \frac{1}{2}(1-q)\hat{q} + \mathbb{E}_{t\sim\mathcal{N}(0,1)}\log\int_{-\infty}^{\infty}dw\,\exp\left(-\frac{1}{2}\sum_{k=1}^{K}\hat{r}_k w^k + \sqrt{\hat{q}}tw\right). \tag{18}$$

We now can derive the saddle point equations. First, as in the standard perceptron calculation, the saddle point equation for q is

$$0 = \alpha\frac{\partial G_1^{\mathrm{RS}}}{\partial q} + \frac{\partial G_2^{\mathrm{RS}}}{\partial q} \tag{19}$$
$$= -\alpha\mathbb{E}_t\left[\left(\frac{t}{\sqrt{q(1-q)}} + \frac{\kappa + t\sqrt{q}}{(1-q)^{3/2}}\right)\frac{\varphi(z)}{H(z)}\Big|_{z=(\kappa+t\sqrt{q})/\sqrt{1-q}}\right] + \frac{1}{2}\hat{q}, \tag{20}$$

which yields

$$\hat{q} = \alpha\mathbb{E}_t\left[\left(\frac{t}{\sqrt{q(1-q)}} + \frac{\kappa + t\sqrt{q}}{(1-q)^{3/2}}\right)\frac{\varphi(z)}{H(z)}\Big|_{z=(\kappa+t\sqrt{q})/\sqrt{1-q}}\right]. \tag{21}$$

The saddle point equations for $\hat{r}_k$ are

$$\frac{\partial G_2}{\partial \hat{r}_k} = \frac{1}{2}m_k - \frac{1}{2}\mathbb{E}_t\frac{\int_{-\infty}^{\infty}dw\,w^k\exp\left(-\frac{1}{2}\sum_{k=1}^{K}\hat{r}_k w^k + \sqrt{\hat{q}}tw\right)}{\int_{-\infty}^{\infty}dw\,\exp\left(-\frac{1}{2}\sum_{k=1}^{K}\hat{r}_k w^k + \sqrt{\hat{q}}tw\right)}. \tag{22}$$

Parallelling Zhong et al. [9]'s analysis,

$$p(w|t) = \frac{\exp\left(-\frac{1}{2}\sum_{k=1}^{K}\hat{r}_k w^k + \sqrt{\hat{q}}tw\right)}{\int_{-\infty}^{\infty}dw\,\exp\left(-\frac{1}{2}\sum_{k=1}^{K}\hat{r}_k w^k + \sqrt{\hat{q}}tw\right)} \tag{23}$$

can be interpreted under appropriate conditions as a conditional density. In the present setup, it has a clear interpretation as a maximum-entropy distribution subject to moment constraints. Then, these saddle point equations are just the moment conditions

$$m_k = \int_{-\infty}^{\infty} p(w) w^k \tag{24}$$

where

$$p(w) = \mathbb{E}_t p(w|t) \tag{25}$$

is the resulting marginal density.

Using Stein's lemma and the second-moment condition

$$\int_{-\infty}^{\infty} p(w) w^2 = m_2 = 1, \tag{26}$$

the saddle point equation for $\hat{q}$ is

$$\frac{\partial G_2}{\partial \hat{q}} = \frac{1}{2}(q-1) + \frac{1}{2\sqrt{\hat{q}}} \mathbb{E}_t \left[t \frac{\int_{-\infty}^{\infty} dw\, w \exp\left(-\frac{1}{2} \sum_{k=1}^{K} \hat{r}_k w^k + \sqrt{\hat{q}} t w \right)}{\int_{-\infty}^{\infty} dw\, \exp\left(-\frac{1}{2} \sum_{k=1}^{K} \hat{r}_k w^k + \sqrt{\hat{q}} t w \right)} \right] \tag{27}$$

$$= \frac{1}{2}(q-1) + \frac{1}{2\sqrt{\hat{q}}} \mathbb{E}_t \left[\frac{\partial}{\partial t} \frac{\int_{-\infty}^{\infty} dw\, w \exp\left(-\frac{1}{2} \sum_{k=1}^{K} \hat{r}_k w^k + \sqrt{\hat{q}} t w \right)}{\int_{-\infty}^{\infty} dw\, \exp\left(-\frac{1}{2} \sum_{k=1}^{K} \hat{r}_k w^k + \sqrt{\hat{q}} t w \right)} \right] \tag{28}$$

$$= \frac{1}{2}(q-1) + \frac{1}{2} \mathbb{E}_t \left[\int_{-\infty}^{\infty} dw\, p(w|t) w^2 \right] - \frac{1}{2} \mathbb{E}_t \left[\left(\int_{-\infty}^{\infty} dw\, p(w|t) w \right)^2 \right] \tag{29}$$

$$= \frac{1}{2} q - \frac{1}{2} \mathbb{E}_t \left[\left(\int_{-\infty}^{\infty} dw\, p(w|t) w \right)^2 \right] \tag{30}$$

hence

$$q = \mathbb{E}_t \left[\left(\int_{-\infty}^{\infty} dw\, p(w|t) w \right)^2 \right]. \tag{31}$$

2.3 Extracting the critical capacity

We now consider the limit $q \uparrow 1$. We first consider the equation for $\hat{q}$ and the energetic term, whose limits follow the standard perceptron calculation. In this limit, we use the fact that

$$\frac{\varphi(z)}{H(z)} \sim \begin{cases} 0 & z \to -\infty \\ z & z \to +\infty \end{cases}, \tag{32}$$

which yields

$$\hat{q} \sim \frac{\alpha}{(1-q)^2} \int_{-\kappa}^{\infty} dt\, \varphi(t)(\kappa + t)^2. \tag{33}$$

We can then see that the correct scaling of $\hat{q}$ as $q \uparrow 1$ is

$$\hat{q} = \frac{\tilde{q}}{(1-q)^2} \tag{34}$$

for $\tilde{q}$ of order one. Then,

$$\tilde{q} = \alpha \int_{-\kappa}^{\infty} dt\, \varphi(t)(\kappa + t)^2 \tag{35}$$

at the $q \uparrow 1$ saddle point. This quantity is positive and of order one. Moreover, we can use the asymptotics of $H(z)$ to obtain

$$2(1-q)G_1^{\mathrm{RS}} \sim -\alpha \int_{-\kappa}^{\infty} dt\, \varphi(t)(\kappa + t)^2 \tag{36}$$

We can then see that the self-consistent scalings for the remaining Lagrange multipliers are

$$\hat{r}_k = \frac{\tilde{r}_k}{1-q}. \tag{37}$$

With this *Ansatz*, the entropic term becomes

$$2(1-q)G_2 = \sum_{k=1}^{K} \tilde{r}_k m_k - \tilde{q} + 2(1-q)\mathbb{E}_{t\sim\mathcal{N}(0,1)} \log \int_{-\infty}^{\infty} dw\, \exp\left[\frac{1}{1-q}\left(-\frac{1}{2}\sum_{k=1}^{K} \tilde{r}_k w^k + \sqrt{\tilde{q}}tw\right)\right]. \tag{38}$$

In the limit $q \uparrow 1$, we can evaluate the integral over w using Laplace's method, yielding

$$2(1-q)G_2 = \sum_{k=1}^{K} \tilde{r}_k m_k - \tilde{q} + \mathbb{E}_t \max_w \left(2\sqrt{\tilde{q}}tw - \sum_{k=1}^{K} \tilde{r}_k w^k\right). \tag{39}$$

At the maximum w_*, we have

$$\sum_{k=1}^{K} k\tilde{r}_k w_*^{k-1} = 2\sqrt{\tilde{q}}t. \tag{40}$$

The saddle point equation for $\tilde{q}$ is

$$0 = -1 + \mathbb{E}_t\left[\frac{\partial}{\partial\tilde{q}}\left(2\sqrt{\tilde{q}}tw_* - \sum_{k=1}^{K} \tilde{r}_k w_*^k\right)\right] \tag{41}$$

$$= -1 + \frac{1}{\sqrt{\tilde{q}}}\mathbb{E}_t[tw_*] + \mathbb{E}_t\left[\left(2\sqrt{\tilde{q}}t - \sum_{k=1}^{K} k\tilde{r}_k w_*^{k-1}\right)\frac{\partial w_*}{\partial\tilde{q}}\right] \tag{42}$$

$$= -1 + \frac{1}{\sqrt{\tilde{q}}}\mathbb{E}_t[tw_*], \tag{43}$$

where the term in the round brackets vanishes by the optimality condition. The saddle point equation for $\tilde{r}_k$ is

$$0 = m_k + \mathbb{E}_t\left[\frac{\partial}{\partial\tilde{r}_k}\left(2\sqrt{\tilde{q}}tw_* - \sum_{l=1}^{K} \tilde{r}_l w_*^l\right)\right] \tag{44}$$

$$= m_k - \mathbb{E}_t[w_*^k] + \mathbb{E}_t\left[\left(2\sqrt{\tilde{q}}t - \sum_{l=1}^{K} l\tilde{r}_l w_*^{l-1}\right)\frac{\partial w_*}{\partial\tilde{r}_k}\right] \tag{45}$$

$$= m_k - \mathbb{E}_t[w_*^k], \tag{46}$$

where the term in the round brackets once again vanishes by the optimality condition.

We are then left with the conditions

$$\mathbb{E}_t[w_*^k] = m_k \qquad (k = 1, \ldots, K) \tag{47}$$

and

$$\tilde{q} = \mathbb{E}_t[tw_*]^2. \tag{48}$$

This yields

$$\alpha_c = \mathbb{E}_t[tw_*]^2 \left[\int_{-\kappa}^{\infty} dt\, \varphi(t)(\kappa + t)^2\right]^{-1}, \tag{49}$$

which is precisely analogous to Zhong et al. [9]'s result. At this stage, we must solve the moment constraints to obtain $\tilde{r}_k$, and from that compute $\mathbb{E}_t[tw_*]$ and then the capacity.

If we now pass to the limit $K \to \infty$ in which we constrain an infinite number of moments, and the moments satisfy an appropriate Carleman condition to uniquely determine a measure [1], then the constraints $\mathbb{E}_t[w_*^k] = m_k$ imply the that the induced distribution of w_* must coincide with the desired distribution. We then recover Zhong et al. [9]'s result. However, this approach cannot be applied to those distributions that are not uniquely determined by their moments. Importantly, this class of distributions includes the lognormal [1], on which Zhong et al. [9] focus.

3 Penalizing the Cramér distance

We will now consider introducing a soft penalty on the distance between the probability distributions—one that is well-defined even at finite N—and then take the penalty to be of infinite strength after taking $N \to \infty$. For analytical convenience, we use the squared Cramér distance between the empirical and target distributions:

$$C(\mathbf{w}) = \int_{-\infty}^{\infty} dw\, [F_{\mathbf{w}}(w) - F(w)]^2, \tag{50}$$

i.e., the square of the L_2 distance between their CDFs [2, 7]. This distance measure is well-defined at any finite N. We then introduce the modified Gardner volume

$$Z = \int d\sigma_N(\mathbf{w})\, \exp[-N\beta C(\mathbf{w})] \prod_{\mu=1}^{P} \Theta\left(\frac{y^\mu \mathbf{x}^\mu \cdot \mathbf{w}}{\sqrt{N}} - \kappa\right), \tag{51}$$

where σ_N is the uniform probability measure on the N-sphere of radius $\sqrt{N}$. In the limit $\beta \to \infty$, this will give the capacity subject to the desired distribution constraint.

This approach will closely mirror that of Zhong et al. [9]; we will recover precisely the same saddle point equations.

3.1 Evaluating the disorder average

The replicated, averaged Gardner volume expands as

$$\mathbb{E}_{\mathbf{x},y} Z^n = \int \prod_{a=1}^{n} d\sigma_N(\mathbf{w}^a) \exp\left(-N\beta \sum_{a=1}^{n} C(\mathbf{w}^a)\right) \left[\mathbb{E}_{\mathbf{x},y} \prod_{a=1}^{n} \Theta\left(\frac{y\mathbf{x} \cdot \mathbf{w}^a}{\sqrt{N}} - \kappa\right)\right]^P. \tag{52}$$

As above, the data average in the square brackets is identical to that in the standard Gardner capacity calculation. Then, the definitions of the Edwards-Anderson order parameters via Fourier representations of the Dirac distribution, we have

$$\mathbb{E}_{\mathbf{x},y} Z^n = \frac{1}{\omega_{N-1}^n} \int \frac{d\mathbf{Q}\, d\hat{\mathbf{Q}}}{(4\pi i/N)^{n(n+1)/2}} \exp\left(\frac{N}{2}\operatorname{tr}(\mathbf{Q}\hat{\mathbf{Q}}) + PnG_1(\mathbf{Q})\right) \times \int \prod_{a=1}^{n} d\mathbf{w}^a \exp\left(-\frac{1}{2}\sum_{a,b=1}^{n} \hat{Q}^{ab}\mathbf{w}^a \cdot \mathbf{w}^b - N\beta \sum_{a=1}^{n} C(\mathbf{w}^a)\right) \tag{53}$$

where the integral over $\hat{\mathbf{Q}}$ is taken over all $n \times n$ imaginary symmetric matrices, and that over $\mathbf{Q}$ is taken over all $n \times n$ symmetric matrices with diagonal elements equal to one. Here,

$$\omega_{N-1} = \int_{\|\mathbf{w}\|_2 = \sqrt{N}} d\mathbf{w} \; = \frac{2\pi^{N/2}}{\Gamma(N/2)} N^{(N-1)/2} \tag{54}$$

is the normalizing factor in σ_N.

To proceed further, we must deal with the fact that the constraint term couples different elements of the weight vectors. Via a functional Hubbard-Stratonovich transformation with auxiliary fields $\phi^a(x)$, we may write

$$\exp\left(-N\beta\sum_{a=1}^{n} C(\mathbf{w}^a)\right) = \frac{1}{\mathcal{Z}^n} \int \prod_{a=1}^{n} \mathcal{D}\phi^a \exp\left(-\frac{N}{2\beta}\sum_{a=1}^{n}\int_{-\infty}^{\infty} dx\, \phi^a(x)^2 - iN\sum_{a=1}^{n}\int_{-\infty}^{\infty} dx\, \phi^a(x)[F_{\mathbf{w}^a}(x) - F(x)]\right), \tag{55}$$

where

$$\mathcal{Z} \equiv \int \mathcal{D}\phi \exp\left(-\frac{N}{2\beta}\int_{-\infty}^{\infty} dx\, \phi(x)^2\right) \tag{56}$$

is the partition function of the auxiliary Gaussian field. Substituting in the definition of $F_{\mathbf{w}^a}(x)$ and evaluating the resulting integrals against indicator functions, this yields

$$\exp\left(-N\beta\sum_{a=1}^{n} C(\mathbf{w}^a)\right) = \frac{1}{\mathcal{Z}^n} \int \prod_{a=1}^{n} \mathcal{D}\phi^a \exp\left(-\frac{N}{2\beta}\sum_{a=1}^{n}\int_{-\infty}^{\infty} dx\, \phi^a(x)^2 + iN\sum_{a=1}^{n}\int_{-\infty}^{\infty} dx\, \phi^a(x)F(x)\right) \times \prod_{j=1}^{N} \exp\left(-i\sum_{a=1}^{n}\int_{w_j^a}^{\infty} dx\, \phi^a(x)\right). \tag{57}$$

We can now factorize the weight integrals across dimensions, yielding

$$\int \prod_{a=1}^{n} d\mathbf{w}^a \exp\left(-\frac{1}{2}\sum_{a,b=1}^{n} \hat{Q}^{ab}\mathbf{w}^a \cdot \mathbf{w}^b - N\beta\sum_{a=1}^{n} C(\mathbf{w}^a)\right) \tag{58}$$

$$= \frac{1}{\mathcal{Z}^n} \int \prod_{a=1}^{n} \mathcal{D}\phi^a \exp\left(-\frac{N}{2\beta}\sum_{a=1}^{n}\int_{-\infty}^{\infty} dx\, \phi^a(x)^2 + iN\sum_{a=1}^{n}\int_{-\infty}^{\infty} dx\, \phi^a(x)F(x)\right) \times \left[\int \prod_{a} dw^a \exp\left(-\frac{1}{2}\sum_{a,b=1}^{n} \hat{Q}^{ab}w^a w^b - i\sum_{a=1}^{n}\int_{w^a}^{\infty} dx\, \phi^a(x)\right)\right]^N. \tag{59}$$

Thus, defining

$$
\begin{aligned}
nG_2(\mathbf{Q},\hat{\mathbf{Q}},\boldsymbol{\phi}) \equiv \frac{1}{2}\operatorname{tr}(\mathbf{Q}\hat{\mathbf{Q}}) \\
&- \frac{1}{2\beta}\sum_{a=1}^{n}\int_{-\infty}^{\infty} dx\,\phi^a(x)^2 + i\sum_{a=1}^{n}\int_{-\infty}^{\infty} dx\,\phi^a(x)F(x) \\
&+ \log\int\prod_{a=1}^{n} dw^a\,\exp\left(-\frac{1}{2}\sum_{a,b=1}^{n}\hat{Q}^{ab}w^a w^b - i\sum_{a=1}^{n}\int_{w^a}^{\infty} dx\,\phi^a(x)\right),
\end{aligned}
\tag{60}
$$

we can write the averaged replicated partition function as

$$
\mathbb{E}_{\mathbf{x},y}Z^n = \frac{1}{\omega_{N-1}^n \mathcal{Z}^n (4\pi i/N)^{n(n+1)/2}}\int d\mathbf{Q}\,d\hat{\mathbf{Q}}\int \mathcal{D}\boldsymbol{\phi}\,\exp\Big(Nn[\alpha G_1(\mathbf{Q}) + G_2(\mathbf{Q},\hat{\mathbf{Q}},\boldsymbol{\phi})]\Big). \tag{61}
$$

This form is suitable for saddle point evaluation.

3.2 Replica-symmetric saddle point equations

We now make a replica-symmetric *Ansatz*

$$
Q^{ab} = (1-q)\delta_{ab} + q \tag{62}
$$

$$
\hat{Q}^{ab} = \hat{z}\delta_{ab} - \hat{q} \tag{63}
$$

$$
\phi^a(x) = i\phi(x), \tag{64}
$$

where our variable definitions are made such that $\hat{q}$ will turn out to be real and positive and $\phi(x)$ will be real at the saddle point.

The energetic term G_1 is the same as in the standard Gardner calculation, while the entropic term simplifies to

$$
\begin{aligned}
G_2^{\mathrm{RS}} = \frac{1}{2}(\hat{z}-\hat{q}) - \frac{1}{2}(n-1)q\hat{q} \\
&+ \frac{1}{2\beta}\int_{-\infty}^{\infty} dx\,\phi(x)^2 - \int_{-\infty}^{\infty} dx\,\phi(x)F(x) \\
&+ \frac{1}{n}\log\int\prod_{a=1}^{n} dw^a\,\exp\left(-\frac{1}{2}\sum_{a=1}^{n}\hat{z}(w^a)^2 + \frac{1}{2}\hat{q}\left(\sum_{a=1}^{n} w^a\right)^2 + \sum_{a=1}^{n}\int_{w^a}^{\infty} dx\,\phi(x)\right)
\end{aligned}
\tag{65}
$$

which, following standard manipulations from the usual Gardner calculation can be written in the limit $n \to 0$ after a Hubbard-Stratonovich transformation as

$$
\begin{aligned}
G_2^{\mathrm{RS}} = \frac{1}{2}\hat{z} - \frac{1}{2}(1-q)\hat{q} \\
&+ \frac{1}{2\beta}\int_{-\infty}^{\infty} dx\,\phi(x)^2 - \int_{-\infty}^{\infty} dx\,\phi(x)F(x) \\
&+ \mathbb{E}_{t\sim\mathcal{N}(0,1)}\log\int_{-\infty}^{\infty} dw\,\exp\left(-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_{w}^{\infty} dx\,\phi(x)\right).
\end{aligned}
\tag{66}
$$

The saddle point equation for q is as in the standard perceptron calculation, yielding

$$
\hat{q} = \alpha\mathbb{E}_t\left[\left(\frac{t}{\sqrt{q(1-q)}} + \frac{\kappa + t\sqrt{q}}{(1-q)^{3/2}}\right)\frac{\varphi(z)}{H(z)}\bigg|_{z=(\kappa+t\sqrt{q})/\sqrt{1-q}}\right]. \tag{67}
$$

The saddle point equation for $\phi(x)$ is

$$0 = \frac{\delta G_2^{\mathrm{RS}}}{\delta \phi(x)} \tag{68}$$

$$= \frac{1}{\beta}\phi(x) - F(x) + \mathbb{E}_t \left[\frac{\int_{-\infty}^{\infty} dw \, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} du \, \phi(u)] \Theta(x-w)}{\int_{-\infty}^{\infty} dw \, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} du \, \phi(u)]} \right]. \tag{69}$$

which yields

$$F(x) = \frac{1}{\beta}\phi(x) + \mathbb{E}_t \left[\frac{\int_{-\infty}^{x} dw \, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} du \, \phi(u)]}{\int_{-\infty}^{\infty} dw \, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} du \, \phi(u)]} \right]. \tag{70}$$

As noted by Zhong et al. [9],

$$p(w|t) \equiv \frac{\exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx \, \phi(x)]}{\int_{-\infty}^{\infty} dw \, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx \, \phi(x)]} \tag{71}$$

has the interpretation of a probability density over w conditioned on t for any fixed $\hat{z}$, $\hat{q}$, and $\phi(x)$. This interpretation holds provided that $\hat{z}$ and $\phi(x)$ are real and that $\hat{q}$ is positive. Then, letting

$$p(w) \equiv \mathbb{E}_t p(w|t) \tag{72}$$

be the resulting marginal distribution,

$$\mathbb{E}_t \int_{-\infty}^{x} dw \, p(w|t) = \int_{-\infty}^{x} dw \, \mathbb{E}_t p(w|t) = \int_{-\infty}^{x} dw \, p(w) \equiv P(x) \tag{73}$$

has the interpretation of a CDF. We can then see that the sensible scaling for $\phi(x)$ as $\beta \to \infty$ is for it to remain order one (such that $\phi(x)/\beta \to 0$), which yields a limiting condition that can be interpreted as a pointwise equality of CDFs:

$$F(x) = P(x). \tag{74}$$

The saddle point equation for $\hat{z}$ is

$$0 = \frac{\partial G_2^{\mathrm{RS}}}{\partial \hat{z}} \tag{75}$$

$$= \frac{1}{2} - \frac{1}{2}\mathbb{E}_t \left[\frac{\int_{-\infty}^{\infty} dw \, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx \, \phi(x)] w^2}{\int_{-\infty}^{\infty} dw \, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx \, \phi(x)]} \right], \tag{76}$$

which gives the second-moment constraint

$$\int_{-\infty}^{\infty} dw \, p(w) w^2 = 1. \tag{77}$$

The saddle point equation for $\hat{q}$ is

$$0 = \frac{\partial G_2^{\mathrm{RS}}}{\partial \hat{q}} \tag{78}$$

$$= \frac{1}{2}(q-1) + \frac{1}{2}\frac{1}{\sqrt{\hat{q}}}\mathbb{E}_t \left[t \frac{\int_{-\infty}^{\infty} dw \, w \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx \, \phi(x)]}{\int_{-\infty}^{\infty} dw \, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx \, \phi(x)]} \right] \tag{79}$$

By Stein's lemma,

$$\mathbb{E}_t\left[t\frac{\int_{-\infty}^{\infty} dw\, w \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx\, \phi(x)]}{\int_{-\infty}^{\infty} dw\, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx\, \phi(x)]}\right] \tag{80}$$

$$= \mathbb{E}_t\left[\frac{\partial}{\partial t}\frac{\int_{-\infty}^{\infty} dw\, w \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx\, \phi(x)]}{\int_{-\infty}^{\infty} dw\, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx\, \phi(x)]}\right] \tag{81}$$

$$= \sqrt{\hat{q}}\mathbb{E}_t\left[\frac{\int_{-\infty}^{\infty} dw\, w^2 \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx\, \phi(x)]}{\int_{-\infty}^{\infty} dw\, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx\, \phi(x)]}\right]$$

$$- \sqrt{\hat{q}}\mathbb{E}_t\left[\left(\frac{\int_{-\infty}^{\infty} dw\, w \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx\, \phi(x)]}{\int_{-\infty}^{\infty} dw\, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx\, \phi(x)]}\right)^2\right] \tag{82}$$

$$= \sqrt{\hat{q}}\int_{-\infty}^{\infty} dw\, p(w)w^2 - \sqrt{q}\mathbb{E}_t\left[\left(\int_{-\infty}^{\infty} dw\, p(w|t)w\right)^2\right] \tag{83}$$

Thus, using the fact that

$$\mathbb{E}_t\left[\int_{-\infty}^{\infty} dw\, p_t(w)w^2\right] = 1 \tag{84}$$

at the saddle point, we have the condition

$$q = \mathbb{E}_t\left[\left(\int_{-\infty}^{\infty} dw\, p(w|t)w\right)^2\right] \tag{85}$$

$$= \mathbb{E}_t\left[\mathbb{E}[w|t]^2\right]. \tag{86}$$

Combining these results, we have the conditions

$$F(x) = \mathbb{E}_t\left[\int_{-\infty}^{x} dw\, p(w|t)\right] \tag{87}$$

$$1 = \mathbb{E}_t\left[\int_{-\infty}^{\infty} dw\, p(w|t)w^2\right] \tag{88}$$

$$q = \mathbb{E}_t\left[\left(\int_{-\infty}^{\infty} dw\, p(w|t)w\right)^2\right] \tag{89}$$

$$\hat{q} = -\alpha\mathbb{E}_t\left[\left(\frac{t}{\sqrt{q(1-q)}} + \frac{\kappa + t\sqrt{q}}{(1-q)^{3/2}}\right)\frac{\varphi(z)}{H(z)}\bigg|_{z=(\kappa+t\sqrt{q})/\sqrt{1-q}}\right] \tag{90}$$

with the definition

$$p(w|t) \equiv \frac{\exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx\, \phi(x)]}{\int_{-\infty}^{\infty} dw\, \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw + \int_w^{\infty} dx\, \phi(x)]}. \tag{91}$$

We must solve these equations for q, $\hat{q}$, $\hat{z}$, and $\phi(x)$.

Except for the fact that the distribution constraint is expressed in terms of CDFs, these saddle point equations are identical to those found by Zhong et al. [9] after a field redefinition, which we will further discuss below. Of course, if $F(x)$ is sufficiently regular, we may recover the condition on PDFs used by Zhong et al. [9] by differentiating:

$$f(x) = p(x). \tag{92}$$

This requires that $F(x)$ is absolutely continuous with respect to Lebesgue measure. More broadly, it seems possible that this construction using a functional Hubbard-Stratonovich transformation, which in the $\beta \to \infty$ limit should correspond to simply introducing a Lagrange multipler field that constrains the CDFs to coincide pointwise, requires $F(x)$ to be absolutely continuous with respect to Lebesgue measure. We will leave a detailed examination of this issue to furture work.

We now remark that Zhong et al. [9] make a clever choice of fields that simplifies their study of the $q \uparrow 1$ limit. Let us assume that $F(x)$ has a well-behaved density $f(x)$, and recall that its second moment must be unity for consistency with the spherical constraint, i.e., that

$$\int_{-\infty}^{\infty} dx\, f(x) x^2 = 1. \tag{93}$$

Then, integrating by parts under the assumption that $\int_w^\infty dx\, \phi(x)$ vanishes as $w \to \infty$, we may write

$$\frac{1}{2}\hat{z} - \int_{-\infty}^{\infty} dw\, \phi(w) F(w) = \int_{-\infty}^{\infty} dw\, f(w) \left(\frac{1}{2}\hat{z} w^2 + \int_w^\infty dx\, \phi(x) \right). \tag{94}$$

In the zero-temperature limit $\beta \to \infty$, we may then write the entropic term as

$$G_2^{\mathrm{RS}} = -\frac{1}{2}(1-q)\hat{q} + \int_{-\infty}^{\infty} dw\, f(w) \lambda(w) + \mathbb{E}_{t \sim \mathcal{N}(0,1)} \log \int_{-\infty}^{\infty} dw\, \exp\left(\sqrt{\hat{q}} t w - \lambda(w) \right) \tag{95}$$

where we have defined the new field

$$\lambda(w) \equiv \frac{1}{2}\hat{z} w^2 + \int_w^\infty dx\, \phi(x), \tag{96}$$

which packages together all dependence on $\hat{z}$ and ϕ. The saddle point equation for $\lambda(x)$ is

$$0 = \frac{\delta G_2^{\mathrm{RS}}}{\delta \lambda(x)} = f(x) - \mathbb{E}_t \left[\frac{\exp\left(\sqrt{\hat{q}} t x - \lambda(x) \right)}{\int_{-\infty}^{\infty} dw\, \exp\left(\sqrt{\hat{q}} t w - \lambda(w) \right)} \right], \tag{97}$$

which yields the density constraint

$$f(x) = \mathbb{E}_t \left[\frac{\exp\left(\sqrt{\hat{q}} t x - \lambda(x) \right)}{\int_{-\infty}^{\infty} dw\, \exp\left(\sqrt{\hat{q}} t w - \lambda(w) \right)} \right]. \tag{98}$$

3.3 Extracting the critical capacity

We now consider the limit $q \uparrow 1$. As in the standard perceptron calculation and our analysis using moment constraints, we have

$$\hat{q} = \frac{\tilde{q}}{(1-q)^2} \tag{99}$$

for

$$\tilde{q} = \alpha \int_{-\kappa}^{\infty} dt\, \varphi(t) (\kappa + t)^2 \tag{100}$$

at the $q \uparrow 1$ saddle point. We also have the asymptotic

$$2(1-q) G_1^{\mathrm{RS}} \sim -\alpha \int_{-\kappa}^{\infty} dt\, \varphi(t) (\kappa + t)^2. \tag{101}$$

We now consider the entropic term with Zhong et al. [9]'s field choices. Examining the zero-temperature entropic term G_2^{RS}, we can see that the self-consistent scaling of $\lambda(x)$ is

$$\lambda(x) = \frac{\tilde{\lambda}(x)}{1-q} \tag{102}$$

for $\tilde{\lambda}(x)$ real and of order one. The entropic term then becomes

$$\begin{aligned} 2(1-q)G_2^{\mathrm{RS}} \sim -\tilde{q} + 2\int_{-\infty}^{\infty} dw\, f(w)\tilde{\lambda}(w) \\ + 2(1-q)\mathbb{E}_t \log \int_{-\infty}^{\infty} dw\, \exp\left[\frac{1}{1-q}\left(\sqrt{\tilde{q}}tw - \tilde{\lambda}(w)\right)\right]. \end{aligned} \tag{103}$$

In the limit $q \uparrow 1$, assuming that $\tilde{\lambda}(x)$ is sufficiently well-behaved, the integral over w can be evaluated by the method of steepest descent. This yields

$$2(1-q)\mathbb{E}_t \log \int_{-\infty}^{\infty} dw\, \exp\left[\frac{1}{1-q}\exp\left(\sqrt{\tilde{q}}tw - \tilde{\lambda}(w)\right)\right] \sim 2\mathbb{E}_t\left[\sqrt{\tilde{q}}tw_* - \tilde{\lambda}(w_*)\right], \tag{104}$$

where w_* is determined by the condition that it maximizes $\sqrt{\tilde{q}}tw - \tilde{\lambda}(w)$, which gives the first-order condition

$$\tilde{\lambda}'(w_*) = \sqrt{\tilde{q}}t. \tag{105}$$

We can now easily obtain the saddle point equation for $\tilde{q}$, which is

$$0 = -1 + \frac{1}{\sqrt{\tilde{q}}}\mathbb{E}_t[tw_*] + 2\mathbb{E}_t\left[\left(\sqrt{\tilde{q}}t - \tilde{\lambda}'(w_*)\right)\frac{\partial w_*}{\partial \tilde{q}}\right] \tag{106}$$

$$= -1 + \frac{1}{\sqrt{\tilde{q}}}\mathbb{E}_t[tw_*], \tag{107}$$

where the term in the round brackets vanishes by the first-order optimality condition for w_*. This yields

$$\tilde{q} = \mathbb{E}_t[tw_*]^2. \tag{108}$$

Similarly,

$$0 = 2f(w) - 2\mathbb{E}_t\delta(w_* - w) \tag{109}$$

$$= 2f(w) - 2\varphi(t_*(w))\left|t_*'(w)\right|, \tag{110}$$

where $t_*(w)$ is determined by inverting the relation $w_*(t)$. Integrating this result to obtain a condition on CDFs, we have

$$F(w) = \int_{-\infty}^{w} dx\, \varphi(t_*(x))\left|t_*'(x)\right| = \Phi(t). \tag{111}$$

This recovers Zhong et al. [9]'s result.

4 Comparison to the capacity calculation for a perceptron with a nonuniform weight prior

It is instructive to compare the above computation with a penalty on the Cramér distance to that for a perceptron with some fixed prior density $\rho(w)$ over the elements of the weight vector, in addition to the spherical constraint. In this case, the modified Gardner volume is

$$Z = \int d\sigma_N(\mathbf{w}) \left[\prod_{j=1}^{N}\rho(w_j)\right]\prod_{\mu=1}^{P}\Theta\left(\frac{y^\mu \mathbf{x}^\mu\cdot\mathbf{w}}{\sqrt{N}} - \kappa\right). \tag{112}$$

4.1 Evaluating the disorder average

We again proceed by using the fact that the training examples are independent and identically distributed, which yields

$$\mathbb{E}_{\mathbf{x},y} Z^n = \int \prod_{a=1}^{n} d\sigma_N(\mathbf{w}^a) \left[\prod_{a=1}^{n} \prod_{j=1}^{N} \rho(w_j^a) \right] \left[\mathbb{E}_{\mathbf{x},y} \prod_{a=1}^{n} \Theta\left(\frac{y\mathbf{x} \cdot \mathbf{w}^a}{\sqrt{N}} - \kappa \right) \right]^P . \tag{113}$$

As before, the data average in the square brackets is identical to that in the standard Gardner capacity calculation. Then, enforcing the definitions of the Edwards-Anderson order parameters via Fourier representations of the Dirac distribution, we have

$$\begin{aligned} \mathbb{E}_{\mathbf{x},y} Z^n = \frac{1}{\omega_{N-1}^n} \int \frac{d\mathbf{Q}\, d\hat{\mathbf{Q}}}{(4\pi i/N)^{n(n+1)/2}} \exp\left(\frac{N}{2} \operatorname{tr}(\mathbf{Q}\hat{\mathbf{Q}}) + PnG_1(\mathbf{Q}) \right) \\ \times \left[\int \prod_{a=1}^{n} dw^a \rho(w^a) \, \exp\left(-\frac{1}{2} \sum_{a,b=1}^{n} \hat{Q}^{ab} w^a w^b \right) \right]^N , \end{aligned} \tag{114}$$

where the integral over $\hat{\mathbf{Q}}$ is taken over all $n \times n$ imaginary symmetric matrices, and that over $\mathbf{Q}$ is taken over all $n \times n$ symmetric matrices with diagonal elements equal to one. The entropic term is then

$$nG_2(\mathbf{Q}, \hat{\mathbf{Q}}) \equiv \frac{1}{2} \operatorname{tr}(\mathbf{Q}\hat{\mathbf{Q}}) + \log \int \prod_{a=1}^{n} dw^a \rho(w^a) \, \exp\left(-\frac{1}{2} \sum_{a,b=1}^{n} \hat{Q}^{ab} w^a w^b \right) , \tag{115}$$

and we can write the averaged replicated partition function as

$$\mathbb{E}_{\mathbf{x},y} Z^n = \frac{1}{\omega_{N-1}^n (4\pi i/N)^{n(n+1)/2}} \int d\mathbf{Q}\, d\hat{\mathbf{Q}} \, \exp\left(Nn[\alpha G_1(\mathbf{Q}) + G_2(\mathbf{Q}, \hat{\mathbf{Q}})] \right) . \tag{116}$$

4.2 Replica-symmetric saddle point equations

We again make a replica-symmetric *Ansatz*

$$Q^{ab} = (1-q)\delta_{ab} + q \tag{117}$$

$$\hat{Q}^{ab} = \hat{z}\delta_{ab} - \hat{q} \tag{118}$$

where our variable definitions are made such that $\hat{q}$ will turn out to be real and positive at the saddle point. Again, the energetic term is identical to the standard Gardner calculation. Following standard manipulations from the usual Gardner calculation, the energetic term can be written in the limit $n \to 0$ after a Hubbard-Stratonovich transformation as

$$G_2^{\mathrm{RS}} = \frac{1}{2}\hat{z} - \frac{1}{2}(1-q)\hat{q} + \mathbb{E}_t \log \int_{-\infty}^{\infty} dw \, \rho(w) \exp\left(-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}} t w \right) . \tag{119}$$

The weight integral is identical to the previous setting with the replacement of the dynamical field-dependent term $\exp\left(\int_w^\infty dx \, \phi(x) \right)$ with the fixed density $\rho(w)$. Therefore, if we define the function

$$p(w|t) \equiv \frac{\rho(w) \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw]}{\int_{-\infty}^{\infty} dw \, \rho(w) \exp[-\frac{1}{2}\hat{z}w^2 + \sqrt{\hat{q}}tw]} , \tag{120}$$

which again has the interpretation of a conditional probability density under appropriate constraints on ρ, $\hat{z}$, and $\hat{q}$, we will obtain the nearly-identical saddle point equations

$$1 = \mathbb{E}_t \left[\int_{-\infty}^{\infty} dw\, p(w|t) w^2 \right] \tag{121}$$

$$q = \mathbb{E}_t \left[\left(\int_{-\infty}^{\infty} dw\, p(w|t) w \right)^2 \right] \tag{122}$$

$$\hat{q} = -\alpha \mathbb{E}_t \left[\left(\frac{t}{\sqrt{q(1-q)}} + \frac{\kappa + t\sqrt{q}}{(1-q)^{3/2}} \right) \frac{\varphi(z)}{H(z)} \bigg|_{z=(\kappa+t\sqrt{q})/\sqrt{1-q}} \right] \tag{123}$$

which we now must solve for q, $\hat{q}$, and $\hat{z}$.

4.3 Extracting the critical capacity

We now consider the limit $q \uparrow 1$. We first consider the equation for $\hat{q}$ and the energetic term, whose limits follow the standard perceptron calculation. As before, the self-consistent scaling of $\hat{q}$ as $q \uparrow 1$ is

$$\hat{q} = \frac{\tilde{q}}{(1-q)^2} \tag{124}$$

for $\tilde{q}$ of order one. Then,

$$\tilde{q} = \alpha \int_{-\kappa}^{\infty} dt\, \varphi(t) (\kappa + t)^2 \tag{125}$$

at the $q \uparrow 1$ saddle point. This quantity is positive and of order one.

Examining the entropic term, we can see that the self-consistent scaling is to have

$$\hat{z} = \frac{\tilde{z}}{1-q} \tag{126}$$

for $\tilde{z}$ real and of order one. The entropic term then becomes

$$2(1-q) G_2^{\mathrm{RS}} \sim \tilde{z} - \tilde{q} + 2(1-q) \mathbb{E}_t \log \int_{-\infty}^{\infty} dw\, \rho(w) \exp \left[\frac{1}{1-q} \left(-\frac{1}{2} \tilde{z} w^2 + \sqrt{\tilde{q}} t w \right) \right]. \tag{127}$$

If $\rho(w)$ is smooth and of non-compact support, we can easily evaluate the integrals over w using the method of steepest descent,

$$w = \frac{\sqrt{\tilde{q}}}{\tilde{z}} t. \tag{128}$$

and

$$2(1-q) G_2^{\mathrm{RS}} \sim \tilde{z} - \tilde{q} + 2\mathbb{E}_t \left[-\frac{1}{2} \tilde{z} w^2 + \sqrt{\tilde{q}} t w \right]_{w=\sqrt{\tilde{q}} t/\tilde{z}} \tag{129}$$

$$\sim \tilde{z} - \tilde{q} + \frac{\tilde{q}}{\tilde{z}}, \tag{130}$$

where the contribution of $2(1-q)\mathbb{E}_t \log \rho(\sqrt{\tilde{q}} t/\tilde{z})$ can be neglected. The saddle point equation for $\tilde{q}$ then yields

$$\tilde{z} = 1, \tag{131}$$

and that for $\tilde{z}$ then yields

$$\tilde{q} = 1. \tag{132}$$

Then, we find that the critical capacity is

$$\frac{1}{\alpha} = \int_{-\kappa}^{\infty} dt\, \varphi(t)(\kappa + t)^2. \tag{133}$$

which recovers the classic Gardner result.

However, if $\rho(w)$ is not sufficiently smooth or is of compact support, then this saddle point evaluation is no longer valid, and the capacity is modified. For example, consider the case in which the weights are sign-constrained:

$$2(1-q)G_2^{\mathrm{RS}} \sim \tilde{z} - \tilde{q} + 2(1-q)\mathbb{E}_t \log \int_0^{\infty} dw\, \exp\left[\frac{1}{1-q}\left(-\frac{1}{2}\tilde{z}w^2 + \sqrt{\tilde{q}}tw\right)\right]. \tag{134}$$

If $t > 0$, then the integral is dominated by $w = \sqrt{\tilde{q}}t/\tilde{z}$. If $t < 0$, then it is dominated by $t = 0$. This leads to

$$2(1-q)G_2^{\mathrm{RS}} \sim \tilde{z} - \tilde{q} + 2\int_0^{\infty} dt\, \varphi(t)\left(-\frac{1}{2}\tilde{z}w^2 + \sqrt{\tilde{q}}tw\right)_{w=\sqrt{\tilde{q}}t/\tilde{z}} \tag{135}$$

$$\sim \tilde{z} - \tilde{q} + \frac{1}{2}\frac{\tilde{q}}{\tilde{z}}. \tag{136}$$

This yields

$$\tilde{z} = \tilde{q} = \frac{1}{2} \tag{137}$$

hence we have

$$\frac{1}{\alpha} = 2\int_{-\kappa}^{\infty} dt\, \varphi(t)(\kappa + t)^2 \tag{138}$$

meaning that the sign constraint reduces the capacity by a factor of 2.

As another possibility, consider the case in which the prior constrains a fraction f of the weights to be non-negative on average. We write this in the form

$$\rho(w) = f\mathbf{1}_{w\geq 0} + (1-f). \tag{139}$$

Combining the arguments above, we find in this case that

$$2(1-q)G_2^{\mathrm{RS}} \sim \tilde{z} - \tilde{q} + \left(1 - \frac{f}{2}\right)\frac{\tilde{q}}{\tilde{z}} \tag{140}$$

which leads to

$$\tilde{z} = \tilde{q} = 1 - \frac{f}{2}. \tag{141}$$

This gives

$$\alpha = \left(1 - \frac{f}{2}\right)\left[\int_{-\kappa}^{\infty} dt\, \varphi(t)(\kappa + t)^2\right]^{-1} \tag{142}$$

which results in a zero-margin critical capacity of

$$\alpha(\kappa = 0) = 2 - f. \tag{143}$$

This recovers the result of Kanter and Eisenstein [5].

References

[1] N. I. Akhiezer, *The classical moment problem and some related questions in analysis* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2020), https://epubs.siam.org/doi/abs/10.1137/1.9781611976397.

[2] M. G. Bellemare, I. Danihelka, W. Dabney, S. Mohamed, B. Lakshminarayanan, S. Hoyer, and R. Munos, "The Cramer distance as a solution to biased Wasserstein gradients", arXiv, 10.48550/ARXIV.1705.10743 (2017), https://arxiv.org/abs/1705.10743.

[3] E. Gardner, "The space of interactions in neural network models", Journal of Physics A: Mathematical and General **21**, 257 (1988).

[4] E. Gardner and B. Derrida, "Optimal storage properties of neural network models", Journal of Physics A: Mathematical and General **21**, 271 (1988).

[5] I. Kanter and E. Eisenstein, "On the capacity per synapse", Journal of Physics A: Mathematical and General **23**, L935 (1990), https://dx.doi.org/10.1088/0305-4470/23/17/016.

[6] W. Krauth and M. Mézard, "Storage capacity of memory networks with binary couplings", J. Phys. France **50**, 3057–3066 (1989), https://doi.org/10.1051/jphys:0198900500200305700.

[7] G. J. Székely, "E-statistics: the energy of statistical samples", Bowling Green State University, Department of Mathematics and Statistics Technical Report **3**, 1–18 (2003).

[8] J. A. Zavatone-Veth and C. Pehlevan, "Activation function dependence of the storage capacity of treelike neural networks", Physical Review E **103**, L020301 (2021), https://link.aps.org/doi/10.1103/PhysRevE.103.L020301.

[9] W. Zhong, B. Sorscher, D. Lee, and H. Sompolinsky, "A theory of weight distribution-constrained learning", in Advances in Neural Information Processing Systems, Vol. 35, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (2022), pp. 14113–14127, https://proceedings.neurips.cc/paper_files/paper/2022/hash/5b2db6dfda4d7362b2101b2d12dac029-Abstract-Conference.html.