SAMPLING THROUGH ALGORITHMIC DIFFUSION IN NON-CONVEX PERCEPTRON PROBLEMS

Elizaveta Demyanenko[1*] **Davide Straziota**[1*] **Carlo Baldassi**[1] **Carlo Lucibello**[1]
[1]Bocconi University, Milano, Italy
{elizaveta.demyanenko,davide.straziota}@phd.unibocconi.it

ABSTRACT

We analyze the problem of sampling from the solution space of simple yet non-convex neural network models by employing a denoising diffusion process known as Algorithmic Stochastic Localization, where the score function is provided by Approximate Message Passing. We introduce a formalism based on the replica method to characterize the process in the infinite-size limit in terms of a few order parameters, and, in particular, we provide criteria for the feasibility of sampling.

We show that, in the case of the spherical perceptron problem with negative stability, approximate uniform sampling is achievable across the entire replica symmetric region of the phase diagram. In contrast, for the binary perceptron, uniform sampling via diffusion invariably fails due to the overlap gap property exhibited by the typical set of solutions. We discuss the first steps in defining alternative measures that can be efficiently sampled.

1 INTRODUCTION

A fundamental challenge in modern machine learning is the ability to efficiently sample from complex probability distributions. This problem is central to many areas, from generative modeling, where methods such as variational autoencoders or autoregressive transoformers aim to approximate real-world data distributions, to Bayesian learning, where one seeks to explore posterior distributions over model parameters.

Diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020) have emerged in the last years as a prominent generative model family, able to capture highly complex distributions in several domains while maintaining low sample complexity.

In this paper, we consider a setting where the target distribution is known up to a normalization function and our goal is fair sampling from the distribution. We will adopt a diffusion setting where the score function is produced algorithmically rather than learned from data.

The diffusion flavor that we use is that of Stochastic Localization (SL) (Eldan, 2013). This probabilistic technique defines a continuous-time stochastic process of conditional distributions, progressively concentrating probability mass on relevant regions of the configuration space. SL has been shown to be equivalent to a denoising diffusion process (Montanari, 2023; El Alaoui & Montanari, 2022). In Grenioux et al. (2024), the authors propose a technique for effectively sampling from multimodal distributions, making use of the Markov Chain Monte Carlo method. In El Alaoui et al. (2022) the SL process was employed to construct an algorithm uniformly sampling the solutions of the SK model at high temperature, using a drift term (related to the score) given by approximate message passing (Donoho et al., 2009).

In this work, inspired by Ghio et al. (2024), we leverage SL in combination to AMP to sample from the solution space of perceptron problems. In particular, we devise an analytic framework to assess the performance of SL in the large system size limit. Notice that AMP is conjecturally optimal among polynomial algorithms in this denoising setting.

We first analyze SL-based sampling for spherical weights, then extend our investigation to the more challenging problem of binary perceptron solutions. We find that efficient sampling is feasible in

[*]Equal contribution

a large region of the hyperparameters' space of the spherical model. In the binary case instead, uniform sampling from the solution space is always unfeasible. This result was expected since the typical solutions are isolated and therefore hard to find (Huang & Kabashima, 2014; Baldassi et al., 2015). Nevertheless, this limitation motivates the exploration of other statistical measures that may allow for efficient sampling.

The remainder of this paper is structured as follows: Section 2 introduces the Stochastic Localization process. Section 3 presents an asymptotic analysis of SL using the replica formalism. Section 4 discusses perceptron models and details the algorithmic implementation of SL-based sampling. The end of this section discusses alternative measures for sampling binary perceptron solutions. In Section 5, we summarize and conclude the results.

2 Stochastic Localization for Sampling

In this section, we introduce the key components of our sampling algorithm based on Stochastic Localization. We begin by presenting the fundamental idea behind SL and providing its mathematical formulation.

Given a probability measure μ over $\mathbb{R}^N$, referred to as the *target measure*, our objective is to generate samples from μ. We assume μ to be a graphical model with a possibly hard-to-compute normalization factor, and write it as

$$\mu(\boldsymbol{W}) = \frac{\psi(\boldsymbol{W})}{Z}, \tag{1}$$

where the partition function Z is the hard to compute normalization and $\psi(\boldsymbol{W})$ is easy to evaluate.

Instead of directly sampling from the target measure, SL entails constructing a sequence of *time-dependent tilted measures*, denoted as $\mu_{\mathbf{h},t}$, which evolves according to a stochastic process. Specifically, these measures take the form

$$\mu_{\mathbf{h},t}(d\boldsymbol{W}) := \frac{\psi(d\boldsymbol{W})\, e^{\langle \mathbf{h}, \boldsymbol{W}\rangle - \frac{t}{2}\|\boldsymbol{W}\|^2}}{Z_{\mathbf{h},t}}, \tag{2}$$

where $t > 0$ represents time, $\mathbf{h} \in \mathbb{R}^N$ is an external field vector, and $Z_{\mathbf{h},t}$ is a normalization constant ensuring that $\mu_{\mathbf{h},t}$ remains a probability measure:

$$Z_{\mathbf{h},t} = \int \psi\left(d\boldsymbol{W}\right) e^{\langle \mathbf{h}(t), \boldsymbol{W}\rangle - \frac{t}{2}\|\boldsymbol{W}\|^2}. \tag{3}$$

The field vector $\mathbf{h}(t)$ evolves dynamically according to the stochastic differential equation:

$$\mathrm{d}\mathbf{h}(t) = \mathbf{m}(\mathbf{h}(t), t)\,\mathrm{d}t + \mathrm{d}\mathbf{B}(t), \qquad \mathbf{h}(0) = 0, \tag{4}$$

where $(\mathbf{B}(t))_{t\geq 0}$ is a standard Brownian motion, and the drift term $\mathbf{m}$ is given by the expectation of $\boldsymbol{W}$ under the tilted measure:

$$\mathbf{m}(\mathbf{h}, t) = \int_{\mathbb{R}^N} \boldsymbol{W}\, \mu_{\mathbf{h},t}(d\boldsymbol{W}). \tag{5}$$

The key advantage of considering the sequence of tilted measures $(\mu_{\boldsymbol{h}(t),t})_{t\geq 0}$ instead of directly sampling from μ lies in its probabilistic properties. Under mild conditions, the measure $\mu_{\mathbf{h}(t),t}$ localizes to a sample $\boldsymbol{W}^*$ from the target measure μ (Eldan, 2013):

$$\mu_{\mathbf{h}(t),t} \to \delta_{\boldsymbol{W}^*}, \quad \boldsymbol{W}^* \sim \mu, \quad \text{a.s. as } t \to \infty.$$

A crucial aspect of the practical implementation of SL is the possibility of efficiently computing the drift term $\mathbf{m}(\mathbf{h}, t)$ using Approximate Message Passing (AMP) (Donoho & Tanner, 2009). In the following sections, we introduce the replica formalism used to analyze the asymptotic properties of SL and evaluate its effectiveness when sampling from the solution space of perceptron models.

3 Asymptotic Analysis of SL with the Replica Formalism

In this section, we investigate the asymptotic behavior of the SL sampling process in the system size limit $N \to \infty$. To analyze the feasibility of sampling from the target measure μ, we leverage the well-established replica method (Mézard et al., 1986).

Given the target measure μ, and the evolving tilted measure $\mu_{\mathbf{h},t}$ defined in Equation 2, which we denote as μ_t for short, our primary quantity of interest is the alignment between a sample $\mathbf{W}_t$ drawn from μ_t at time t and a reference sample $\mathbf{W}^*$ drawn from μ. The overlap is expressed as:

$$q(t) = \frac{1}{N}\mathbb{E}\left[\langle \boldsymbol{W}^*, \boldsymbol{W}_t\rangle_t\right]. \tag{6}$$

By analyzing the behavior of q as a function of time, we can assess whether the SL process successfully converges to samples from μ.

The main object of study in the replica analysis is the partition function associated with the tilted measure, $Z_{X,\mathbf{h},t}$. One can show that (see Montanari (2023) for a rigorous derivation)

$$\mathbf{h}(t) = t\boldsymbol{W}^* + \sqrt{t}\mathbf{g}, \qquad (\boldsymbol{W}^*, \mathbf{g}) \sim \mu \otimes \mathcal{N}(0, I_N). \tag{7}$$

Our goal is to evaluate its asymptotic expectation over the patterns of the partition function defined in equation 3:

$$\phi_t = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\log Z_{\mathbf{h}(t),t}, \tag{8}$$

where the expectation is over the field defined by equation 7 and any other source of disorder, e.g. in the factor ψ. To compute this limit, we employ the replica method twice, introducing s replicas of the target measure weights and n replicas of the tilted measure weights:[1]

$$\phi_t = \lim_{N\to\infty}\frac{1}{N}\lim_{s\to 0}\lim_{n\to 0}\frac{1}{n}\partial_n \mathbb{E}\int \prod_{\alpha=1}^{s}\psi\left(d\boldsymbol{W}_\alpha^*\right)\prod_{a=1}^{n}\psi\left(d\boldsymbol{W}_a\right)e^{\langle \mathbf{h}(t),\boldsymbol{W}_a\rangle - \frac{t}{2}\|\boldsymbol{W}_a\|^2}. \tag{9}$$

For integer values of n and s, before taking the zero limit, we can interpret the argument of the expectation as the partition function of a system of $n+s$ interacting replicas. The structure of this computation is closely related to the so-called teacher-student setting in replica theory, as introduced in the work of Franz and Parisi (Franz & Parisi, 1997). In this setting, the sample $\boldsymbol{W}$ (student) attempts to infer the reference weight vector $\boldsymbol{W}^*$ (teacher).

To proceed, we assume the Replica Symmetric (RS) ansatz (Mezard & Montanari, 2009) and the Bayesian optimality conditions (Nishimori, 1980; Iba, 1999), which simplifies the analysis by assuming that all replicas are statistically equivalent. Under this hypothesis, the overlap of interest, at time t, is given by:

$$q_0(t) = \frac{1}{N}\mathbb{E}\left[\langle \boldsymbol{W}^*, \boldsymbol{W}_a\rangle\right] = \mathbb{E}\left[\langle \boldsymbol{W}_a, \boldsymbol{W}_b\rangle\right], \qquad \forall a, b \in [n] \text{ and } b \neq a. \tag{10}$$

with $\boldsymbol{W}_a$ samples from μ_t and q_0 an order parameter that enters the replica calculation as we will see. The next section presents the explicit evaluation of this overlap and discusses the implications of the RS assumption.

4 Application to Perceptron Models

In this section, we first present the perception model (Engel & Van den Broeck, 2001), and then we introduce the SL sampling algorithm, specifically designed for this simple one-layer neural network.

4.1 The Perceptron Model

The perceptron is the simplest instance of a one-layer neural network performing binary classification. Given a dataset $X = \{\mathbf{X}^\mu, y^\mu\}_{\mu=1}^M$, where $\mathbf{X}^\mu \in \mathbb{R}^N$ denotes the input patterns and $y^\mu \in \{-1, +1\}$ represents the corresponding binary labels, classification problem is resolved by learning a weight vector $\boldsymbol{W} \in \mathbb{R}^N$ that minimizes a given loss function.

A common approach to training the perceptron is through margin-based optimization. We thus consider as target measure the flat one over probability of satisfying the constraint

$$\psi(\boldsymbol{W}) = \prod_\mu \Theta\left(\kappa - y^\mu \frac{\langle \mathbf{X}^\mu, \boldsymbol{W}\rangle}{\sqrt{N}}\right), \tag{11}$$

[1] We use $\log Z = \lim_{n\to 0} n^{-1}\partial_n Z$ and $Z^{-1} = \lim_{s\to 0} Z^{s-1}$.

where $\Theta(\cdot)$ is the Heaviside step function, $\Theta(x) = 1$ if $x > 0$ and 0 otherwise, and $\kappa \in \mathbb{R}$ is the margin parameter, which controls the level of tolerance to classification errors. For $\kappa = 0$ this loss simply counts the classification errors, whereas for $\kappa < 0$ small-margin errors are admissible, leading to a non-convex solution space.

In this work, we focus on a prototypical setting where both the input features and the labels are randomly drawn from uniform distributions: $y^\mu \sim \mathrm{Unif}(\{-1,+1\})$ and $\mathbf{X}^\mu \sim \mathrm{Unif}_N(\{-1,+1\})$.

This choice corresponds to a widely studied random perceptron model, which serves as a fundamental benchmark for analyzing the statistical properties of learning in high-dimensional settings (Baldassi et al., 2023; 2020).

In this paper, we consider two perceptron models: the binary and the spherical perceptron. In the binary perception, the weights $\boldsymbol{W}$ are binary variables in $\{-1,+1\}^N$, while in the spherical perception, they belong to $\mathbb{R}^N$, but are normalized such that $\|\boldsymbol{W}\|^2 = N$.

4.2 Algorithmic Implementation

In this section we present the Stochastic Localization sampling scheme used to explore the solution space of the perceptron problems. The SL algorithm is formally described in Algorithm 2. To estimate the expected values of the weights under the tilted measure, we employ the AMP algorithm, whose formulation is detailed in Algorithm 1.

The AMP framework relies on the definition of two key functions (ϕ_{in} and ϕ_{out}), commonly referred to as the input and output channels, given by:

$$\phi_{\text{out}}(y,\omega,V) = \log\left(\frac{1}{\sqrt{2\pi V}}\int dz\Theta(z)\exp\left(-\frac{(z-y\omega)^2}{2V}\right)\right), \tag{12}$$

$$\phi_{\text{in}}(A,B) = \log\left(\int_{\mathbb{R}^2} dW \exp\left(BW - \frac{1}{2}W^2 A\right)\right), \tag{13}$$

$$g_{\text{out}}(\omega,y,V) = \partial_\omega \phi_{\text{out}}(y,\omega,V). \tag{14}$$

For brevity, we have omitted the dataset indices from the notation. In the spherical perceptron, the weight vector $\boldsymbol{W}$ is constrained to lie on the N-dimensional sphere, enforcing $\|\boldsymbol{W}\|^2 = N$ dynamically during the AMP iterations. This constraint is enforced as

$$A_i' := A_i + \Delta A_i, \quad \forall i \in \{1,\dots,N\}, \qquad \sum_{i=1}^N \partial_{A_i}\phi_{\text{in}}(A_i,B_i)\Big|_{A_i=A_i'} = N. \tag{15}$$

For the binary perceptron model, the AMP equations remain structurally similar to those described above, with the primary distinction lying in the definition of ϕ_{in}.

4.3 Spherical Non-Convex Perceptron

In this section, we present the asymptotic analysis of the perceptron sampling scheme, focusing on the spherical perceptron model. As discussed in Section 3, a key quantity for characterizing the sampling behavior is the free entropy, which allows us to determine the overlap q_0. Using the replica approach, the free entropy at time t can be expressed as a function of the order parameters:

$$\phi_t(q_0,\hat{q}_0,r_0,\hat{r}_0) = \frac{1}{2}(1-\hat{q}_0 q_0) + \hat{r}_0 r_0 + \frac{1}{2}t q_0 + G_S + \alpha G_E, \tag{16}$$

$$G_S = -\frac{1}{2}\log(\hat{q}_0+1) - \frac{1}{2(1+\hat{q}_0)}\left(\hat{q}_0 + 2\hat{r}_0\frac{(\hat{q}_0-\hat{r}_0)}{(\hat{r}_0+1)} + \frac{(\hat{q}_0-\hat{r}_0)^2}{(\hat{r}_0+1)}\left(\frac{\hat{r}_0}{\hat{r}_0+1}+1\right)\right), \tag{17}$$

$$G_E = \int DzD\gamma \frac{H\left(-\frac{\gamma\sqrt{r_0}+z\sqrt{q_0-r_0}}{\sqrt{1-q_0}}\right)}{H\left(-\frac{\gamma\sqrt{r_0}}{\sqrt{1-r_0}}\right)}\log H\left(-\frac{z\sqrt{q_0-r_0}+\gamma\sqrt{r_0}}{\sqrt{1-q_0}}\right). \tag{18}$$

where $r_0 = \langle \mathbf{W}_a^*, \mathbf{W}_b^* \rangle$ (for $a \neq b$) represents the overlap between two distinct teacher weight assignments, and $\hat{q}_0$, $\hat{r}_0$ are the corresponding dual overlap parameters. The optimization of ϕ_t is

performed through the saddle point method, enabling us to determine the optimal value of the order parameters.

Algorithm 1
Approximate Message Passing (AMP) on the spherical perceptron

Input: Data: $\boldsymbol{F} \in \mathbb{R}^{N\times M}$, $\boldsymbol{y} \in \{-1,1\}^N$, parameters: $\boldsymbol{A}_{\text{ext}} \in \mathbb{R}^N$, $\boldsymbol{B}_{\text{ext}} \in \mathbb{R}^N$, ϵ, T

Initialize: $\boldsymbol{W}^0 \coloneqq \bar{0}_N$, $\boldsymbol{\Delta}^0 \coloneqq \bar{0}_N$, $\boldsymbol{g}^0_{\text{out}} \coloneqq \bar{0}_M$, $\boldsymbol{A} \coloneqq \bar{0}_N$, $\boldsymbol{B} \coloneqq \bar{0}_N$, $\Delta_{\text{iter}} = 0$, $t = 1$

repeat

Updating mean and variance estimates ω_μ, V_μ

$$V^t_\mu \leftarrow \sum_i F^2_{\mu i}\Delta^{t-1}_i \tag{19}$$

$$\omega^t_\mu \leftarrow \sum_i F_{\mu i} W^{t-1}_i - V^t_\mu g^{t-1}_{\text{out},\mu} \tag{20}$$

Updating estimates A_i, B_i, $g_{\text{out},\mu}$

$$g^t_{\text{out},\mu} \leftarrow g_{\text{out}}(\omega^t_\mu, y_\mu, V^t_\mu) \tag{21}$$

$$A^t_i \leftarrow -\sum_\mu F^2_{\mu i}\partial_\omega g_{\text{out}}(\omega^t_\mu, y_\mu, V^t_\mu) \tag{22}$$

$$B^t_i \leftarrow W^{t-1}_i A^t_i + \sum_\mu F_{\mu i} g^t_{\text{out},\mu} \tag{23}$$

$$\tag{24}$$

Enforce spherical constraint

$$\boldsymbol{A}^t \leftarrow (\boldsymbol{A}^t + \Delta\boldsymbol{A}^t) \tag{25}$$

Updating marginals W_i and Δ_i

$$W^t_i \leftarrow \partial_{B_i}\phi_{\text{in}}\left(A^t_i + (A^t_{\text{ext}})_i\ , B^t_i + (B^t_{\text{ext}})_i\right) \tag{26}$$

$$\Delta^t_i \leftarrow \partial^2_{B_i}\phi_{\text{in}}\left(A^t_i + (A^t_{\text{ext}})_i\ , (B^t_{\text{ext}})_i\right) \tag{27}$$

$$\tag{28}$$

$\Delta_{\text{iter}} = \|\boldsymbol{W}^t - \boldsymbol{W}^{t-1}\|^2_2/N$

$t \leftarrow t+1$

until $t = T$ or $\Delta_{\text{iter}} < \epsilon$
return $\boldsymbol{W}, \boldsymbol{\Delta}$

Algorithm 2
Sampling a solution to the perceptron problem stochastic localization

Input: Data: $\boldsymbol{F} \in \mathbb{R}^{N\times M}$, $\boldsymbol{y} \in \{-1,1\}^N$, parameters: ϵ, L, δ, T

Initialize the problem:
$\boldsymbol{A}^0_{ext} = \bar{0}_N$, $\boldsymbol{B}^0_{ext} = \bar{0}_N$, $t = 0$, $iter = 1$

while $l \leq L$ **do**

AMP updates of the marginals:

$$\boldsymbol{W}^l, \boldsymbol{\Delta}^l = \text{AMP}(\boldsymbol{F}, \boldsymbol{y}, \boldsymbol{A}^{l-1}_{\text{ext}}, \boldsymbol{B}^{l-1}_{\text{ext}},\ \epsilon,\ T)$$

Update the fields using the Euler-Maruyama scheme:

$$t^l \leftarrow t^{l-1} + \delta \tag{29}$$

$$(A^l_{\text{ext}})_i \leftarrow t \tag{30}$$

$$(B^l_{\text{ext}})_i \leftarrow (B^{l-1}_{\text{ext}})_i + W^{l-1}_i\delta + \sqrt{\delta}\, Z_i \tag{31}$$

$$\text{with } \boldsymbol{Z} \sim \mathcal{N}(0, \mathbb{I}_N), \forall i \in \{1,\dots,N\} \tag{32}$$

$$l \leftarrow l+1 \tag{33}$$

end while

$$\boldsymbol{W}^L_i = \text{sign}\,(\boldsymbol{W}^L_i), \forall i \in \{1\dots N\}$$

return a solution sample $\boldsymbol{W}^L$

As shown in Ghio et al. (2024), the behavior of the free entropy function $\phi(q_0)$ provides direct insight into whether the SL sampling scheme (Algorithm 2) can recover samples from the target measure asymptotically.

The right panel of Fig. 1 shows the free entropy $\phi(q_0)$ as a function of the overlap q_0 for different values of t[2], with parameters $\alpha = 278$ and $\kappa = -2.5$. Initially, $\phi(q_0)$ exhibits a single global maximum. However, a second local maximum emerges as t increases, eventually becoming the global optimizer of ϕ. This transition marks the onset of multimodality in the free entropy landscape. The ability of SL to correctly sample from the target measure hinges on the unimodality of $\phi(q_0)$. If at any time t, the free entropy remains unimodal, the algorithm successfully samples from the target measure. Conversely, once $\phi(q_0)$ becomes multimodal, the tilted measure loses information about the target measure, making sampling unreliable.

[2]In the teacher-student scenario, the overlaps of the teacher weights are fixed in advance and depends only on t and α. Instead, $\hat{q}_0$ can be derived implicitly as a function of q_0.This idea implies that ϕ_t can be seen as a function of the only q_0.

In general, sampling algorithms rely on two types of initialization strategies: *Uninformed Initialization:* No prior knowledge of the target measure is provided. *Informed Initialization:* Some prior information about the target measure is utilized.

The left panel of Fig. 1 presents the phase diagram for the SL scheme for a fixed margin κ, delineating the different sampling regimes. Fig. 2 presents the phase transition line for SL applied to the

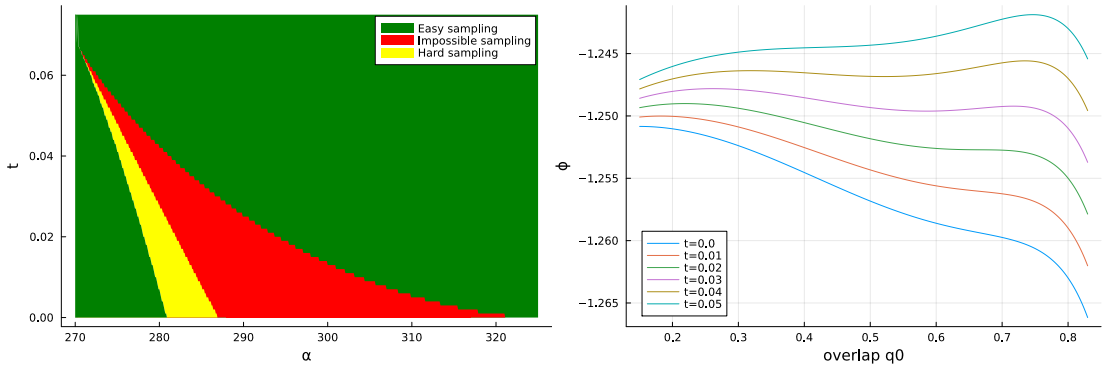

Figure 1: **Left:** Phase diagram of the SL scheme for the spherical perceptron with $\kappa = -2.5$ over the choices of constraint density α and time t. **Green region:** $\phi(q_0)$ has a single optimizer, meaning Algorithm 2 correctly samples from the target measure. **Yellow region:** $\phi(q_0)$ has two optimizers, but the global maximum corresponds to the smaller overlap q_0. The SL algorithm still samples correctly. **Red region:** $\phi(q_0)$ has two optimizers, but the global maximum corresponds to a larger overlap q_0. In this case, Algorithm 2 fails to sample from the target measure. **Right:** Free entropy function $\phi(q_0)$ for different values of t. Initially, $\phi(q_0)$ has a single maximum, but as t increases, a second maximum appears, eventually dominating.

spherical perceptron, in comparison with other phase transitions reported in Baldassi et al. (2023). Notably, the SL transition coincides with both the Dynamical Approximate Threshold (DAT) and the dynamical 1-step Replica Symmetry Breaking (1RSB) transition.

This result highlights a fundamental limitation: the SL algorithm fails to sample from the target measure precisely when Replica Symmetry Breaking (RSB) occurs. In statistical physics, RSB is required when the solution space fragments into disconnected clusters. In other words, SL ceases to function as soon as ergodicity is broken.

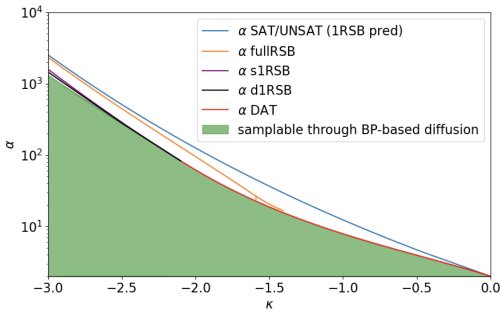

Figure 2: Phase transition delineating the samplable and non-samplable regions for SL in the spherical perceptron (red line). The blue line represents the dynamical transition of the spherical perceptron, as discussed in Baldassi et al. (2023).

To further investigate the uniformity of the SL sampling scheme, we analyze the stability distribution $P(\Delta) = P(\{\Delta^1, \dots, \Delta^M\})$, where the stability vector is defined as $\Delta^\mu = y^\mu \frac{\langle \mathbf{X}^\mu, \boldsymbol{W} \rangle}{\sqrt{N}}$ for $\mu \in [M]$.

Our results suggest that whenever SL successfully samples from the target measure, the obtained samples are uniformly distributed over the solution space. A detailed stability analysis is provided in the Supplementary Material.

4.4 Binary Perceptron

We analyze the samplability of the binary perceptron measure. We first examine the flat measure over solution, which fails to be samplable by SL at any α. Then we consider a tilted measure introducing a potential function.

4.4.1 Flat Measure

Considering the flat measure over solutions, and performing the replica computation, we observe that the term G_E remains identical to those derived for the spherical perceptron. However, the entropic term undergoes a structural modification:

$$G_S = \sum_{W^*=\pm 1} \int Dz\, D\gamma\, \frac{e^{\gamma\sqrt{\hat{r}_0}W^*}}{2\cosh(\gamma\sqrt{\hat{r}_0})} \log\left(2e^{\frac{1}{2}(\hat{q}_d-\hat{q}_0)} \cosh\left(z\sqrt{\hat{q}_0-\hat{r}_0}+\gamma\sqrt{\hat{r}_0}+(\hat{q}_0-\hat{r}_0)W^*\right)\right). \tag{34}$$

Following the methodology outlined in Section 4.3, we conclude that sampling solutions from the flat binary perceptron measure using the SL algorithm, in the limit $N \to \infty$, is intrinsically infeasible. This result is not surprising, as the asymptotic analysis performed via the replica method is known to accurately predict the behavior of the AMP algorithm for large system sizes (Rangan, 2011). As demonstrated in Gamarnik et al. (2022), due to the inherent stability of AMP, the exact evaluation of the expected values in equation 5 becomes impossible whenever the solution space is fragmented, as is the case in the frozen-1RSB phase and overlap gap property is observed (Gamarnik, 2021). Thus, the failure mode of the SL sampling scheme can be directly attributed to the limitations of AMP.

In the binary perceptron case a persistent double-peak structure is observed, but, unlike the spherical case, increasing the time parameter t does not eliminate the secondary peak (see right panel of Fig. 3). Notably, the global optimizer corresponds to an overlap value of $q_0 = 1$, suggesting the presence of a spinodal transition.

The fundamental difference between these two models lies in the statistical properties of their solution spaces. For sufficiently small values of α the negative spherical perceptron can be fully described within the RS ansatz, implying a connected solution space. In contrast, the binary perceptron is characterized by a frozen 1RSB phase, where the solution space consists of disconnected clusters. This structural fragmentation directly impacts the effectiveness of the SL sampling scheme, ultimately rendering it unsuitable for this problem.

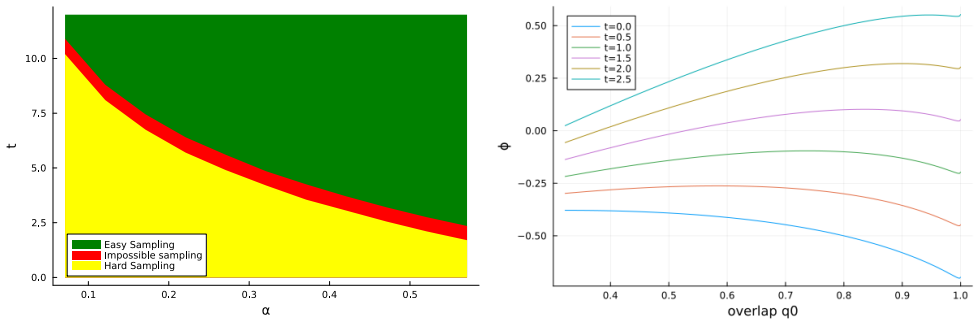

Figure 3: Left: Phase transition of the SL algorithm for the binary perceptron problem. The unfeasible phase is represented by all the parameter domains. Right: The free entropy of the SL for the binary perceptron problem as a function of the overlap q. Double peak behavior is exhibited for all the values of t with the global optimizer corresponding to the overlap $q_0 = 1$.

4.4.2 CROSS-ENTROPY POTENTIAL

Cross-entropy loss Due to the negative result for the uniform measure presented in the last section, we consider a different measure where interpolation is not strictly enforced but it is obtained in the large β and γ limit of the following factor which contains the binary cross-entropy:

$$\psi(\boldsymbol{W}) = \prod_{\mu} e^{-\frac{\beta}{\gamma}\log\left(1+e^{-2\gamma y^{\mu}\frac{\langle \mathbf{X}^{\mu},\boldsymbol{W}\rangle}{\sqrt{N}}}\right)}. \tag{35}$$

As studied in Baldassi et al. (2020) this loss allows for targeting dense regions of solutions (the high local entropy regions) which is controlled by the $\gamma > 0$ parameter. One may note that in the limit of $\beta, \gamma \to \infty$ the model is forcing the constraints $\frac{\langle \mathbf{X}^{\mu},\boldsymbol{W}\rangle}{\sqrt{N}} > 0, \forall \mu \in [M]$ to be strictly satisfied, thus the solution to such a problem corresponds to a binary perceptron one. Our results indicate the possibility of sampling from this target measure with the SL scheme for a sufficiently low density constraint α. The derivation of the free entropy for this model as well as the experimental results can be found in the Supplementary Material. This measure while sampleable is not entirely satisfying, since the necessity to set β large enough to comply with the constraints greatly reduces the support of the measure.

Tilted binary perceptron measure As a logical continuation of the previous experiments, we now keep the cross-entropy potential, but also use an hard constraint to enforce interpolation:

$$\psi(\boldsymbol{W}) = \prod_{\mu} \Theta\left(\kappa - y^{\mu}\frac{\langle \mathbf{X}^{\mu},\boldsymbol{W}\rangle}{\sqrt{N}}\right) e^{-\frac{\beta}{\gamma}\log\left(1+e^{-2\gamma\frac{\langle \mathbf{X}^{\mu},\boldsymbol{W}\rangle}{\sqrt{N}}}\right)}. \tag{36}$$

Despite introducing this tunable potential, our results indicate that similarly to the binary perceptron case the multimodality of the free entropy as a function of q_0 persists over all the parameter regimes, rendering the SL sampling impossible.

5 CONCLUSION

In this work, we investigated the feasibility of sampling solutions to perceptron problems using SL. We analyzed both the spherical and binary perceptron models, leveraging statistical physics techniques to characterize the structure of their solution spaces and assess the effectiveness of the SL scheme. For the spherical perceptron, we demonstrated that SL is capable of sampling from the target measure as long as the free entropy landscape remains unimodal along the trajectory. The scheme is shown to be successful in the spherical perceptron case for the constraint density parameter α being within the replica symmetric regime. In the binary case, we show that the uniform measure is always unsamplable instead. On the other hand, measure not strictly enforcing interpolation but containing a crossentropy potential can be sampled. It is left for future work to devise an efficiently samplable measure that enforces all constraints for the binary perceptron model.

REFERENCES

Carlo Baldassi, Alessandro Ingrosso, Carlo Lucibello, Luca Saglietti, and Riccardo Zecchina. Subdominant dense clusters allow for simple learning and high computational performance in neural networks with discrete synapses. *Phys. Rev. Lett.*, 115:128101, Sep 2015. doi: 10.1103/PhysRevLett.115.128101. URL https://link.aps.org/doi/10.1103/PhysRevLett.115.128101.

Carlo Baldassi, Fabrizio Pittorino, and Riccardo Zecchina. Shaping the learning landscape in neural networks around wide flat minima. *Proceedings of the National Academy of Sciences*, 117(1): 161–170, 2020.

Carlo Baldassi, Enrico M Malatesta, Gabriele Perugini, and Riccardo Zecchina. Typical and atypical solutions in nonconvex neural networks with discrete and continuous weights. *Physical Review E*, 108(2):024310, 2023.

David Donoho and Jared Tanner. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1906): 4273–4293, 2009.

David L. Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009. doi: 10.1073/pnas.0909892106. URL https://www.pnas.org/doi/abs/10.1073/pnas.0909892106.

Ahmed El Alaoui and Andrea Montanari. An information-theoretic view of stochastic localization. *IEEE Transactions on Information Theory*, 68(11):7423–7426, 2022.

Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Sampling from the sherrington-kirkpatrick gibbs measure via algorithmic stochastic localization. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 323–334. IEEE, 2022.

Ronen Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geometric and Functional Analysis*, 23(2):532–569, 2013.

A. Engel and C. Van den Broeck. *Statistical Mechanics of Learning*. Cambridge University Press, 2001.

Silvio Franz and Giorgio Parisi. Phase diagram of coupled glassy systems: A mean-field study. *Phys. Rev. Lett.*, 79:2486–2489, Sep 1997. doi: 10.1103/PhysRevLett.79.2486. URL https://link.aps.org/doi/10.1103/PhysRevLett.79.2486.

David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41):e2108492118, 2021. doi: 10.1073/pnas.2108492118. URL https://www.pnas.org/doi/abs/10.1073/pnas.2108492118.

David Gamarnik, Eren C. Kızıldağ, Will Perkins, and Changji Xu. Algorithms and barriers in the symmetric binary perceptron model. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 576–587, 2022. doi: 10.1109/FOCS54457.2022.00061.

Davide Ghio, Yatin Dandi, Florent Krzakala, and Lenka Zdeborová. Sampling with flows, diffusion, and autoregressive neural networks from a spin-glass perspective. *Proceedings of the National Academy of Sciences*, 121(27):e2311810121, 2024.

Louis Grenioux, Maxence Noble, Marylou Gabrié, and Alain Oliviero Durmus. Stochastic localization via iterative posterior sampling, 2024. URL https://arxiv.org/abs/2402.10758.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Haiping Huang and Yoshiyuki Kabashima. Origin of the computational hardness for learning with binary synapses. *Physical Review E*, 90(5):052813, 2014.

Yukito Iba. The nishimori line and bayesian statistics. *Journal of Physics A: Mathematical and General*, 32(21):3875–3888, January 1999. ISSN 1361-6447. doi: 10.1088/0305-4470/32/21/302. URL http://dx.doi.org/10.1088/0305-4470/32/21/302.

Marc Mezard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, Inc., USA, 2009. ISBN 019857083X.

Marc Mézard, Giorgio Parisi, and M. A. Virasoro. Spin glass theory and beyond: An introduction to the replica method and its applications. 1986. URL https://api.semanticscholar.org/CorpusID:116804781.

Andrea Montanari. Sampling, diffusions, and stochastic localization. *arXiv preprint arXiv:2305.10690*, 2023.

H Nishimori. Exact results and critical properties of the ising model with competing interactions. *Journal of Physics C: Solid State Physics*, 13(21):4071, jul 1980. doi: 10.1088/0022-3719/13/21/012. URL https://dx.doi.org/10.1088/0022-3719/13/21/012.

Sundeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 2168–2172. IEEE, 2011.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. PMLR, 2015.

A SPHERICAL PERCEPTRON

A.1 STABILITY DISTRIBUTION OF THE SPHERICAL PERCEPTRON SOLUTION

In this section, we revise the computation of the stabilities distribution in the spherical perceptron problem. Given an N-dimensional weight vector $\boldsymbol{W}$, representing a solution of the perceptron problem with a margin κ, we consider a stability distribution $P(\Delta) = P(\{\Delta^1, \dots, \Delta^M\})$, where Δ represents the stabilities vector and $\Delta^\mu = y^\mu \frac{\langle \boldsymbol{W}, \boldsymbol{X^\mu} \rangle}{\sqrt{N}},\ 1 \le \mu \le M$.

Since averaging over $\boldsymbol{X^\mu}$ renders all the coordinates of $P(\Delta)$ equivalent, we will consider without loss of generality a stability distribution w.r.t the first input, i.e. $P(\Delta^1)$.

We define the probability of the stability to be equal Δ, averaging over all the weight vectors representing the solutions to the problem and the inputs $\boldsymbol{X}$

$$P(\Delta) = \langle\!\langle \frac{\int d\mu(\boldsymbol{W}) \prod_{\mu=1}^{M} \Theta\left(\frac{y^\mu}{\sqrt{N}} \langle \boldsymbol{W}, \boldsymbol{X^\mu} \rangle - \kappa\right) \delta\left(\frac{y^1}{\sqrt{N}} \langle \boldsymbol{W}, \boldsymbol{X^1} \rangle - \Delta\right)}{d\mu(\boldsymbol{W}) \prod_{\mu=1}^{M} \Theta\left(\frac{y^\mu}{\sqrt{N}} \langle \boldsymbol{W}, \boldsymbol{X^\mu} \rangle - \kappa\right)} \rangle\!\rangle.$$

Now we introduce replicas rewriting the denominator as $Z^{-1} = \lim_{n\to 0} Z^{n-1}$, subsequently taking the $n \to 0$ limit

$$P(\Delta) = \lim_{n\to 0} \langle\!\langle \int \prod_a d\mu(\boldsymbol{W}^a) \prod_{\mu,a} \Theta\left(\frac{y^\mu}{\sqrt{N}} \langle \boldsymbol{W}^a, \boldsymbol{X^\mu} \rangle - \kappa\right) \delta\left(\frac{y^1}{\sqrt{N}} \langle \boldsymbol{W}^1, \boldsymbol{X^1} \rangle - \Delta\right) \rangle\!\rangle.$$

Introducing the $\delta-$ and $\Theta-$ integrals we obtain

$$\begin{aligned} P(\Delta) = \Theta(\Delta - \kappa) \lim_{n\to 0} \int \prod_a d\mu(\boldsymbol{W}^a) \int \frac{d\hat{\lambda}_1^1}{2\pi} \int_\kappa^\infty \prod_{a>1} \frac{d\lambda_1^a}{2\pi} \int d\hat{\lambda}_1^a \\ \times \int_\kappa^\infty \prod_{a,\mu>1} \frac{d\lambda_\mu^a}{2\pi} \int \prod_{a,\mu>1} d\hat{\lambda}_\mu^a \exp\left(i\Delta\hat{\lambda}_1^1 + i\sum_{a>1} \lambda_1^a \hat{\lambda}_1^a + i \sum_{\mu>1,a} \lambda_\mu^a \hat{\lambda}_\mu^a\right) \\ \times \langle\!\langle \exp\left(-\frac{i}{\sqrt{N}} \sum_{a,\mu} \hat{\lambda}_\mu^a \langle \boldsymbol{W}^a, \boldsymbol{X^\mu} \rangle\right) \rangle\!\rangle. \end{aligned}$$

Averaging over $\boldsymbol{X^\mu}$ and introducing the following order parameters

$$q_{ab} \equiv \frac{1}{N} \sum_i \boldsymbol{W}_i^a \boldsymbol{W}_i^b, \qquad k_a \equiv \frac{1}{N} \sum_i (\boldsymbol{W}_i^a)^2,$$

$$\begin{aligned} P(\Delta) = \Theta(\Delta - \kappa) \lim_{n\to 0} \int \prod_a \frac{d\hat{k}^a}{4\pi} \int \prod_{a<b} \frac{dq^{ab} d\hat{q}^{ab}}{2\pi/N} \int \frac{d\hat{\lambda}_1^1}{2\pi} \int_\kappa^\infty \prod_{a>1} \frac{d\lambda_1^a}{2\pi} \int d\hat{\lambda}_1^a \\ \times \exp\left(i\Delta\hat{\lambda}_1^1 + i\sum_{a>1} \lambda_1^a \hat{\lambda}_1^a - \frac{1}{2}\sum_a (\hat{\lambda}_1^a)^2 - \frac{1}{2}\sum_{a,b} \hat{\lambda}_1^a \hat{\lambda}_1^b q^{ab}\right) \\ \times \exp\left(\frac{iN}{2}\sum_a \hat{k}^a + iN \sum_{a<b} q^{ab}\hat{q}^{ab} + NG_S(\hat{k}^2, \hat{q}^{ab}) + (\alpha N - 1) G_E(q^{ab})\right). \end{aligned}$$

where the entropic and energetic parts are respectively

$$\begin{aligned} G_S(\hat{k}^a, \hat{q}^{ab}) &= \ln \int \prod_a \frac{d\boldsymbol{W}^a}{\sqrt{2\pi e}} \exp\left(-\frac{i}{2}\sum_a \hat{k}^a (\boldsymbol{W}^a)^2 - i\sum_{a<b} \hat{q}^{ab} \boldsymbol{W}^a \boldsymbol{W}^b\right), \\ G_E(q^{ab}) &= \ln \int_k^\infty \prod_a \lambda^a \int \prod_a \frac{d\hat{\lambda}_a}{2\pi} \prod_a \exp(i\sum_a \lambda^a \hat{\lambda}^a - \frac{1}{2}\sum_a (\hat{\lambda}^a)^2 - \frac{1}{2}\sum_{a,b} \hat{\lambda}^a \hat{\lambda}^b q^{ab}). \end{aligned}$$

Imposing replica symmetry assumption

$$q^{ab} = q \qquad \hat{q}^{ab} = \hat{q} \qquad \hat{k}^a = \hat{k},$$

And taking the $n \to 0$ limit we get

$$P(\Delta) = \Theta(\Delta - \kappa) \frac{1}{\sqrt{2\pi(1-q)}} \int Dt \exp\left(-\frac{(\Delta - \sqrt{q}t)^2}{2(1-q)}\right)\left[H\left(\frac{\kappa - \sqrt{q}t}{\sqrt{1-q}}\right)\right]^{-1}.$$

Where $H(x) = \int_x^\infty \frac{dt}{\sqrt{2\pi}} \exp(-t^2/2)$. The distribution of the stabilities is reported in Fig.4 for $\kappa = -2.1$ and several values of $\alpha \in [5, 20, 80]$.

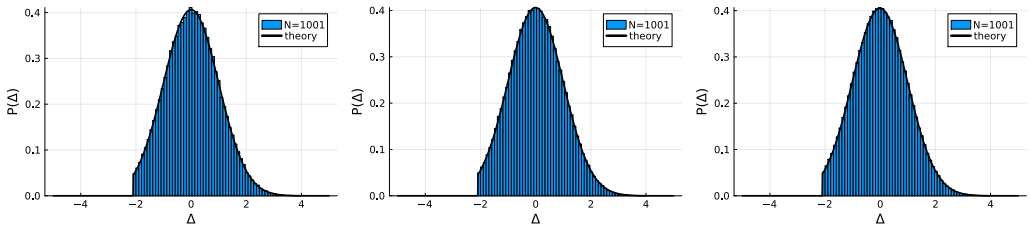

Figure 4: Distribution of the stabilities $P(\Delta) = P(\{\Delta_1 \dots \Delta_{\lceil \alpha N \rceil}\})$ of a negative spherical perceptron where $\Delta^\mu = y^\mu \frac{\langle \boldsymbol{W}, \boldsymbol{X}^\mu \rangle}{\sqrt{N}}$. For number of variables $N = 1001$, $\kappa = -2.1$ and $\alpha = 5$ (Left), $\alpha = 20$ (Center) and $\alpha = 80$ (Right). Empirical distribution of the stabilities of the SL samples (blue) coincides for different parameter regimes with the predicted distribution (black solid line).

A.2 Replica computation for Spherical Perceptron

In this section, the main steps of the replica computation are presented for the spherical perceptron sampling, where $\mathbf{X}^\mu \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_n)$. The distribution $P(\mathbf{W}_\alpha^*) = P(\mathbf{W}_a) = \mathcal{N}(0,1)$. Thus, the expected value of the partition function is:

$$\mathbb{E}\tilde{Z}_t = \mathbb{E}_{X,\mathbf{g}} \int \prod_{\alpha=1}^{s} P(\mathrm{d}\mathbf{W}_\alpha^*) \prod_{a=1}^{n} P(\mathrm{d}\mathbf{W}_a) \prod_{\mu\alpha} \Theta\left(\sum_i \frac{W_{\alpha i}^* X_i^\mu}{\sqrt{N}}\right) \prod_{\mu a} \Theta\left(\sum_i \frac{W_{ai} X_i^\mu}{\sqrt{N}}\right) \times$$
$$\times\, e^{t \sum_{ai} W_{1i}^* W_{ai} + \sqrt{t} \sum_{ai} g_i W_{ai} - \frac{1}{2} t \sum_{ai} W_{ai}^2}$$

by defining $\lambda_\alpha^\mu = \sum_i \frac{W_{\alpha i}^* X_i^\mu}{\sqrt{N}}$ and $u_a^\mu = \sum_i \frac{W_a X_i^\mu}{\sqrt{N}}$, and making explicit the change of variable we get

$$\mathbb{E}\tilde{Z}_t = \mathbb{E}_{X,\mathbf{g}} \int \prod_{\alpha=1}^{s} \mathrm{d}\mathbf{W}_\alpha^* \prod_{a=1}^{n} \mathrm{d}\mathbf{W}_a \prod_{\mu\alpha} \frac{d\lambda_\alpha^\mu d\hat{\lambda}_\alpha^\mu}{2\pi} \prod_{\mu a} \frac{du_a^\mu d\hat{u}_a^\mu}{2\pi} \prod_{\mu\alpha} \Theta(\lambda_\alpha^\mu) \prod_{\mu a} \Theta(u_a^\mu)$$
$$\times\, e^{-i \sum_{\mu\alpha} \hat{\lambda}_\alpha^\mu \lambda_\alpha^\mu - i \sum_{\mu a} \hat{u}_a^\mu u_a^\mu + i \sum_{\mu\alpha i} \hat{\lambda}_\alpha^\mu \frac{W_{\alpha i}^* X_i^\mu}{\sqrt{N}} + i \sum_{\mu a i} \hat{u}_a^\mu \frac{W_{ai} X_i^\mu}{\sqrt{N}} + t \sum_{ai} W_{1i}^* W_{ai} + \sqrt{t} \sum_{ai} g_i W_{ai} - \frac{1}{2} t \sum_{ai} W_{ai}^2}$$

Taking the average over $\mathbf{X}^\mu$ and $\mathbf{g}$, one obtains

$$\mathbb{E}\tilde{Z}_t = \int \prod_{\alpha=1}^{s} \mathrm{d}\mathbf{W}_\alpha^* \prod_{a=1}^{n} \mathrm{d}\mathbf{W}_a \prod_{\mu\alpha} \frac{d\lambda_\alpha^\mu d\hat{\lambda}_\alpha^\mu}{2\pi} \prod_{\mu a} \frac{du_a^\mu d\hat{u}_a^\mu}{2\pi} \prod_{\mu\alpha} \Theta(\lambda_\alpha^\mu) \prod_{\mu a} \Theta(u_a^\mu)$$
$$\times\, e^{-i \sum_{\mu\alpha} \hat{\lambda}_\alpha^\mu \lambda_\alpha^\mu - i \sum_{\mu a} \hat{u}_a^\mu u_a^\mu + t \sum_{ai} W_{1i}^* W_{ai} + \frac{1}{2} t \sum_i \sum_{ab} W_{ai} W_{bi} - \frac{1}{2} t \sum_{ai} W_{ai}^2}$$
$$\times\, e^{-\frac{1}{2N} \sum_{\mu i} (\sum_{\alpha\beta} \hat{\lambda}_\alpha^\mu \hat{\lambda}_\beta^\mu W_{\alpha i}^* W_{\beta i}^* + \sum_{ab} \hat{u}_a^\mu \hat{u}_b^\mu W_{ai} W_{bi} + 2 \sum_{\alpha a} \hat{\lambda}_\alpha^\mu \hat{u}_a^\mu W_{\alpha i}^* W_{ai})}.$$

We can now define the overlaps as

$$q_{ab} = \frac{1}{N} \sum_i W_{ai} W_{bi} \qquad r_{\alpha\beta} = \frac{1}{N} \sum_i W_{\alpha i}^* W_{\beta i}^* \qquad p_{\alpha a} = \frac{1}{N} \sum_i W_{\alpha i}^* W_{ai}$$

By considering the overlaps and their corresponding Lagrange multipliers, $\hat{p}, \hat{q}, \hat{r}$, we can proceed in the computation as follows

$$
\begin{aligned}
\mathbb{E}\tilde{Z}_t = & \int \prod_{\alpha=1}^{s} \mathrm{d}\mathbf{W}_\alpha^* \prod_{a=1}^{n} \mathrm{d}\mathbf{W}_a \prod_{\alpha\leq\beta} dr_{\alpha\beta} d\hat{r}_{\alpha\beta} \prod_{a\leq b} dq_{ab} d\hat{q}_{ab} \prod_{\alpha a} dp_{\alpha a} d\hat{p}_{\alpha a} \\
& \times e^{-N\frac{1}{2}\sum_{\alpha\beta}\hat{r}_{\alpha\beta} r_{\alpha\beta} - \frac{1}{2}N\sum_{ab}\hat{q}_{ab} q_{ab} - N\sum_{\alpha a}\hat{p}_{\alpha a} p_{\alpha a}} \\
& \times e^{+\frac{1}{2}\sum_{\alpha\beta}\hat{r}_{\alpha\beta}\sum_i W_{\alpha i}^* W_{\beta i}^* + \frac{1}{2}\sum_{ab}\hat{q}_{ab}\sum_i W_{ai} W_{bi} + \sum_{\alpha a}\hat{p}_{\alpha a}\sum_i W_{\alpha i}^* W_{ai}} \\
& \times \prod_{\mu\alpha} \frac{d\lambda_\alpha^\mu d\hat{\lambda}_\alpha^\mu}{2\pi} \prod_{\mu a} \frac{du_a^\mu d\hat{u}_a^\mu}{2\pi} \prod_{\mu\alpha} \Theta(\lambda_\alpha^\mu) \prod_{\mu a} \Theta(u_a^\mu) \\
& \times e^{-i\sum_{\mu\alpha}\hat{\lambda}_\alpha^\mu \lambda_\alpha^\mu - i\sum_{\mu a}\hat{u}_a^\mu u_a^\mu + tN\sum_a p_{1a} + \frac{1}{2}tN\sum_{ab} q_{ab} - \frac{t}{2}N\sum_a q_{aa}} \\
& \times e^{-\frac{1}{2}\sum_\mu(\sum_{\alpha\beta}\hat{\lambda}_\alpha^\mu \hat{\lambda}_\beta^\mu r_{\alpha\beta} + \sum_{ab}\hat{u}_a^\mu \hat{u}_b^\mu q_{ab} + 2\sum_{\alpha a}\hat{\lambda}_\alpha^\mu \hat{u}_a^\mu p_{\alpha a})}
\end{aligned}
$$

Now everything is factorized over the sites i and the factors μ, so we obtain an expression amenable to saddle point evaluation

$$
\mathbb{E}\tilde{Z}_t = \int \prod_{\alpha\leq\beta} dr_{\alpha\beta} d\hat{r}_{\alpha\beta} \prod_{a\leq b} dq_{ab} d\hat{q}_{ab} \prod_{\alpha a} dp_{\alpha a} d\hat{p}_{\alpha a}\, e^{N\phi}
$$

where

$$
\phi = G_I + G_S + \alpha G_E
$$

$$
G_I = -\frac{1}{2}\sum_{\alpha\beta}\hat{r}_{\alpha\beta} r_{\alpha\beta} - \frac{1}{2}\sum_{ab}\hat{q}_{ab} q_{ab} - \sum_{\alpha a}\hat{p}_{\alpha a} p_{\alpha a} + t\sum_a p_{1a} + \frac{1}{2}t\sum_{ab} q_{ab} - \frac{t}{2}\sum_a q_{aa}
$$

$$
G_S = \log \int \prod_{\alpha=1}^{s} P(\mathrm{d}W_\alpha^*) \prod_{a=1}^{n} P(\mathrm{d}W_a)\, e^{+\frac{1}{2}\sum_{\alpha\beta}\hat{r}_{\alpha\beta} W_\alpha^* W_\beta^* + \frac{1}{2}\sum_{ab}\hat{q}_{ab} W_a W_b + \sum_{\alpha a}\hat{p}_{\alpha a} W_\alpha^* W_a}
$$

$$
\begin{aligned}
G_E = \log \int & \prod_{\alpha} \frac{d\lambda_\alpha d\hat{\lambda}_\alpha}{2\pi} \prod_{a} \frac{du_a d\hat{u}_a}{2\pi} \prod_{\alpha} \Theta(\lambda_\alpha) \prod_{a} \Theta(u_a) \\
& \times e^{-i\sum_\alpha \hat{\lambda}_\alpha \lambda_\alpha - i\sum_a \hat{u}_a u_a - \frac{1}{2}\sum_{\alpha\beta}\hat{\lambda}_\alpha \hat{\lambda}_\beta r_{\alpha\beta} - \frac{1}{2}\sum_{ab}\hat{u}_a \hat{u}_b q_{ab} - \sum_{\alpha a}\hat{\lambda}_\alpha \hat{u}_a p_{\alpha a}}
\end{aligned}
$$

A.2.1 Replica Symmetric Ansatz

To compute the expression of the free entropy, the RS ansatz is made:

$$
q_{ab} = \begin{cases} q_d & \text{if } a = b \\ q_0 & \text{if } a \neq b \end{cases} \qquad r_{\alpha\beta} = \begin{cases} r_0 & \alpha \neq \beta \\ r_d & \alpha = \beta \end{cases} \qquad p_{\alpha a} = \begin{cases} p^\star & \alpha = 1 \\ p & \alpha \neq 1 \end{cases}
$$

Plugging it in ϕ expression and carrying over the $n \to 0$ and $s \to 0$ limit. One can compute the analytic expression for: G_E, G_S, G_I.

The parameters to optimize in this context are 12, for speeding up the computation of the saddle points, it is possible to notice that Nishimori conditions, and Bayesian Optimality of the order parameters (Iba, 1999; Nishimori, 1980) apply.

$$
\begin{aligned}
\mathbb{E}_{1,W_2}[f(W_1, W_2)] &= \int dh dW_1 dW_2\, f(W_1, W_2) P(W_1|h) P(W_2|h) P(h) \\
\mathbb{E}_{W_*,W_1}[f(W_*, W_1)] &= \int dW_* dW_1\, f(W_*, W_1) P(W_1, W_*) \\
&= \int dh dW_* dW_1\, f(W_*, W_1) P(W_1|h) P(h|W_*) P(W_*) \\
&= \int dh dW_* dW_1\, f(W_*, W_1) P(W_1|h) P(W_*|h) P(h)
\end{aligned}
$$

The Nishimori line is defined by the following equality.

$$
\mathbb{E}_{W_*,W_1}[f(W_*, W_1)] = \mathbb{E}_{W_1,W_2}[f(W_1, W_2)].
$$

We can use this result to understand order parameters relations, in particular using previous reasoning we can notice that

$$\mathbb{E}_{W^*_\alpha, W_a}[W^*_{\alpha i} W_{ai}] = \mathbb{E}_{W_a, W_b}[W_{ai} W_{bi}].$$

leading to

$$p_1 = q_0$$

This equality at first sight might seem unreasonable, but it hides a simple concept: in the Bayes optimal scenario, the overlap of student and teacher weight equals the student-student overlap. Since all the students' weights are independent and identically distributed, we could expect that the performance of students is identical among themselves.

Using the same idea it is possible to show

$$\mathbb{E}_{W^*_\alpha, W^*_\beta}[W^*_{\alpha i} W^*_{\beta i}] = \mathbb{E}_{W_1, W^*_\beta}[W_{1i} W^*_{\beta i}].$$

leading to

$$p_0 = r_0$$

A similar intuition concerning the overlap of students' weights can be applied to the overlap among teacher weights. Furthermore, it is useful to remember that the weights lie on a sphere, providing us with another equality:

$$r_d = q_d$$

Similar reasoning can be applied for the "dual" order parameters, i.e.

$$\hat{r}_d = \hat{q}_d$$
$$\hat{p}_1 = \hat{q}_0$$
$$\hat{p}_0 = \hat{r}_0$$

Let's first look at the entropic term, by using the RS ansatz we can rewrite the entropic term as:

$$G_S = \log \int \prod_{\alpha=1}^{s} P(\mathrm{d}W^*_\alpha) \prod_{a=1}^{n} P(\mathrm{d}W_a)\, e^{+\frac{1}{2}(\hat{r}_d - \hat{r}_0)\sum_\alpha W^{*2}_\alpha + \frac{1}{2}(\hat{r}_0 - \hat{p}_0)(\sum_\alpha W^*_\alpha)^2}$$
$$\times\, e^{+\frac{1}{2}(\hat{q}_d - \hat{q}_0)\sum_a W^2_a + \frac{1}{2}(\hat{q}_0 - \hat{p}_0)(\sum_a W_a)^2 + (\hat{p_1} - \hat{p}_0) W^*_1 \sum_a W_a + \frac{\hat{p_0}}{2}(\sum_\alpha W^*_\alpha + \sum_a W_a)^2}$$

Substituting the Nishimori conditions in the previous expression

$$= \log \int \prod_{\alpha=1}^{s} P(\mathrm{d}W^*_\alpha) \prod_{a=1}^{n} P(\mathrm{d}W_a)\, e^{+\frac{1}{2}(\hat{r}_d - \hat{r}_0)\sum_\alpha W^{*2}_\alpha}$$
$$\times\, e^{+\frac{1}{2}(\hat{q}_d - \hat{q}_0)\sum_a W^2_a + \frac{1}{2}(\hat{q}_0 - \hat{r}_0)(\sum_a W_a)^2 + (\hat{q}_0 - \hat{r}_0) W^*_1 \sum_a W_a + \frac{\hat{r_0}}{2}(\sum_\alpha W^*_\alpha + \sum_a W_a)^2}$$

Using the Hubbard-Stratonovich transformation we can rewrite the entropic term as

$$G_S = \log \int \prod_{\alpha=1}^{s} P(\mathrm{d}W^*_\alpha) \prod_{a=1}^{n} P(\mathrm{d}W_a) Dz D\gamma\, e^{+\frac{1}{2}(\hat{r}_d - \hat{r}_0)\sum_\alpha W^{*2}_\alpha}$$
$$\times\, e^{+\frac{1}{2}(\hat{q}_d - \hat{q}_0)\sum_a W^2_a + z\sqrt{\hat{q}_0 - \hat{r}_0}\sum_a W_a + (\hat{q}_0 - \hat{r}_0) W^*_1 \sum_a W_a + \gamma\sqrt{\hat{r}_0}(\sum_\alpha W^*_\alpha + \sum_a W_a)}$$

which can be rewritten as

$$G_S = \log \int P(\mathrm{d}W^*_1) Dz D\gamma\, e^{+\frac{1}{2}(\hat{r}_d - \hat{r}_0) W^{*2}_1 + \gamma\sqrt{\hat{p_0}} W^*_1} [\mathcal{Z}(W^*_1, z, \gamma)]^n [\mathcal{Z}^*(\eta, \gamma)]^{s-1}$$

where

$$\mathcal{Z}(W^*_1, z, \gamma) = \int P(\mathrm{d}w) e^{+\frac{1}{2}(\hat{q}_d - \hat{q}_0) W^2 + \left(z\sqrt{\hat{q}_0 - \hat{r}_0} + \gamma\sqrt{\hat{r}_0} + (\hat{q}_0 - \hat{r}_0) W^*_1\right) W} =$$
$$= \frac{e^{-\frac{1}{2}\frac{\left(z\sqrt{\hat{q}_0 - \hat{r}_0} + \gamma\sqrt{\hat{r}_0} + (\hat{q}_0 - \hat{r}_0) W^*_1\right)^2}{(\hat{q}_d - \hat{q}_0)}}}{\sqrt{\hat{q}_0 - \hat{q}_d}}$$

$$\mathcal{Z}^*(z,\gamma) = \int P(\mathrm{d}W^*) e^{+\frac{1}{2}(\hat{r}_d-\hat{r}_0)W^{*2}+\gamma\sqrt{\hat{r}_0}W^*} =$$
$$= \frac{e^{-\frac{1}{2}\frac{\gamma^2\hat{r}_o}{\hat{r}_d-\hat{r}_0}}}{\sqrt{\hat{r}_0-\hat{r}_d}}$$

$$\mathcal{A}(W_1^*,\gamma) = e^{+\frac{1}{2}(\hat{r}_d-\hat{r}_0)W_1^{*2}+\gamma\sqrt{\hat{p}_0}W_1^*}.$$

Finally, by taking the limits for $n \to 0$ and $s \to 0$ in the same order we get

$$G_S = \int P(\mathrm{d}W_1^*) Dz D\gamma \, \frac{\mathcal{A}(W_1^*,\gamma)}{\mathcal{Z}^*(\gamma)} \log \mathcal{Z}(W_1^*, z, \gamma)$$

It is possible to compute the result of the above integration explicitly, obtaining the result reported in the main text.

Let's now focus our attention on the energetic term

$$G_E = \log \int \prod_\alpha \frac{d\lambda_\alpha d\hat{\lambda}_\alpha}{2\pi} \prod_a \frac{du_a d\hat{u}_a}{2\pi} \prod_\alpha \Theta(\lambda_\alpha) \prod_a \Theta(u_a)$$
$$\times e^{-i\sum_\alpha \hat{\lambda}_\alpha\lambda_\alpha - i\sum_a \hat{u}_a u_a - \frac{1}{2}\sum_{\alpha\beta}\hat{\lambda}_\alpha\hat{\lambda}_\beta r_{\alpha\beta} - \frac{1}{2}\sum_{ab}\hat{u}_a\hat{u}_b q_{ab} - \sum_{\alpha a}\hat{\lambda}_\alpha\hat{u}_a p_{\alpha a}}$$

After inserting the Nishimori conditions and making some manipulations

$$G_E = \log \int \prod_\alpha \frac{d\lambda_\alpha d\hat{\lambda}_\alpha}{2\pi} \prod_a \frac{du_a d\hat{u}_a}{2\pi} \prod_\alpha \Theta(\lambda_\alpha) \prod_a \Theta(u_a)$$
$$\times e^{-i\sum_\alpha \hat{\lambda}_\alpha\lambda_\alpha - i\sum_a \hat{u}_a u_a - \frac{1}{2}(r_d-r_0)\sum_\alpha \hat{\lambda}_\alpha^2 + \frac{q_0-r_0}{2}\hat{\lambda}_1^2}$$
$$\times e^{-\frac{1}{2}(q_d-q_0)\sum_\alpha \hat{u}_\alpha^2 - \frac{r_0}{2}(\sum_\alpha \hat{\lambda}_\alpha + \sum_a \hat{u}_a)^2 - \frac{q_0-r_0}{2}\left(\hat{\lambda}_1 + \sum_a \hat{u}_a\right)^2}$$

Through Hubbard Stratonovich substitution, it gets

$$G_E = \log \int Dz\, D\gamma \prod_\alpha \frac{d\lambda_\alpha d\hat{\lambda}_\alpha}{2\pi} \prod_a \frac{du_a d\hat{u}_a}{2\pi} \prod_\alpha \Theta(\lambda_\alpha) \prod_a \Theta(u_a)$$
$$\times e^{-i\sum_\alpha \hat{\lambda}_\alpha(\lambda_\alpha + \gamma\sqrt{p_0}) - i\sum_a \hat{u}_a(u_a + \gamma\sqrt{p_0} + z\sqrt{q_0-r_0}) - \frac{1}{2}(r_d-r_0)\sum_\alpha \hat{\lambda}_\alpha^2 + \frac{q_0-r_0}{2}\hat{\lambda}_1^2}$$
$$\times e^{-\frac{1}{2}(q_d-q_0)\sum_\alpha \hat{u}_\alpha^2 - iz\sqrt{q_0-r_0}\hat{\lambda}_1}$$

Finally, taking the limit for $n \to 0$ and $s \to 0$ we get

$$G_E = \int Dz D\gamma \frac{H\left(-\frac{\gamma\sqrt{p_0}+\rho\sqrt{q_0-r_0}}{\sqrt{r_d-q_0}}\right)}{H\left(-\frac{\gamma\sqrt{p_0}}{\sqrt{r_d-r_0}}\right)} \log H\left(-\frac{z\sqrt{q_0-r_0}+\gamma\sqrt{p_0}}{\sqrt{q_d-q_0}}\right)$$

Regarding the Interaction term

$$nG_I = -\frac{1}{2}\sum_{\alpha\beta}\hat{r}_{\alpha\beta}r_{\alpha\beta} - \frac{1}{2}\sum_{ab}\hat{q}_{ab}q_{ab} - \sum_{\alpha a}\hat{p}_{\alpha a}p_{\alpha a} + t\sum_a p_{1a} + \frac{1}{2}t\sum_{ab}q_{ab} - \frac{t}{2}\sum_a q_{aa}$$

by taking the limits $n \to 0$ and $s \to 0$, one is able to get

$$G_I = -\frac{1}{2}\hat{q}_d q_d - \frac{1}{2}\hat{q}_0 q_0 + \hat{r}_0 r_0 + \frac{1}{2}t q_0$$

B Binary Perceptron

B.1 Replica computation for Binary Perceptron Flat measure

The replica computation for the binary perceptron involves the identical to A.2 steps adjusted to the fact that in this case $X_i^\mu \sim \frac{\delta_1+\delta_{-1}}{2}$. The same also holds for the weights distribution $P(W_i^*) = P(W_i) = \frac{\delta_1+\delta_{-1}}{2}$. One can notice that this change does not affect the specific form of the energetic

term G_E and of the interaction term G_I, but only the one of the entropic term G_S. By following the same steps as before one obtains

$$G_S = \int P(\mathrm{d}W_1^*) Dz D\gamma \, \frac{\mathcal{A}(W_1^*, \gamma)}{\mathcal{Z}^*(\gamma)} \log \mathcal{Z}(W_1^*, z, \gamma)$$

where

$$\begin{aligned}
\mathcal{Z}(W_1^*, z, \gamma) &= \int P(\mathrm{d}w) e^{+\frac{1}{2}(\hat{q}_d - \hat{q}_0) w^2 + \left(z\sqrt{\hat{q}_0 - \hat{r}_0} + \gamma\sqrt{\hat{r}_0} + (\hat{q}_0 - \hat{r}_0) W_1^*\right) w} \times \\
&= e^{+\frac{1}{2}(\hat{q}_d - \hat{q}_0)} \cosh\left(z\sqrt{\hat{q}_0 - \hat{r}_0} + \gamma\sqrt{\hat{r}_0} + (\hat{q}_0 - \hat{r}_0) W_1^*\right) \\
\mathcal{Z}^*(z, \gamma) &= \int P(\mathrm{d}w^*) e^{+\frac{1}{2}(\hat{r}_d - \hat{r}_0) w^{*2} + \gamma\sqrt{\hat{r}_0} w^*} \\
&= e^{+\frac{1}{2}(\hat{r}_d - \hat{r}_0)} \cosh(\gamma\sqrt{\hat{r}_0}) \\
\mathcal{A}(W_1^*, \gamma) &= e^{+\frac{1}{2}(\hat{r}_d - \hat{r}_0) W_1^{*2} + \gamma\sqrt{\hat{p}_0} W_1^*}
\end{aligned}$$

B.2 Experimental results for sampling from the cross-entropy loss measure

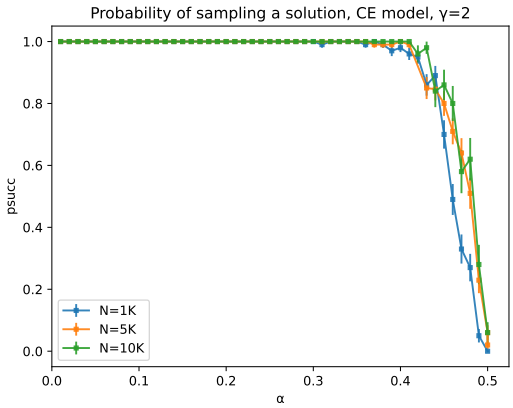

Figure 5: The probability of sampling a solution satisfying the binary perceptron constraints with the SL algorithm as a function of the constraint density α from a measure with the cross-entropy potential with $\beta = 10$ and $\gamma = 2$. The points represent the mean (with the corresponding standard deviation) obtained using the parallel 100 runs of the experiment.

B.3 Replica computation for the cross-entropy loss measure

In this appendix, we present the replica computation for the cross-entropy loss measure given in equation 35. It is important to note that the entropic term, G_S, remains unchanged and coincides with the one reported in Appendix B.1. The only term affected by the presence of the potential is the energetic term, G_E, which we compute as

$$\begin{aligned}
G_E = \log \int \prod_\alpha \frac{d\lambda_\alpha d\hat{\lambda}_\alpha}{2\pi} \prod_a \frac{du_a d\hat{u}_a}{2\pi} \prod_\alpha e^{-\beta V(\lambda_\alpha)} \prod_a e^{-\beta V(u_a)} \\
\times e^{-i\sum_\alpha \hat{\lambda}_\alpha \lambda_\alpha - i\sum_a \hat{u}_a u_a - \frac{1}{2}\sum_{\alpha\beta} \hat{\lambda}_\alpha \hat{\lambda}_\beta r_{\alpha\beta} - \frac{1}{2}\sum_{ab} \hat{u}_a \hat{u}_b q_{ab} - \sum_{\alpha a} \hat{\lambda}_\alpha \hat{u}_a p_{\alpha a}},
\end{aligned}$$

where $V(x) = \frac{1}{\gamma}\log\left(1+e^{-\gamma(x)}\right)$. Following the same steps as in the computation for the spherical perceptron, we obtain:

$$G_E = \int Dz D\gamma \frac{\hat{H}\left(-\frac{\gamma\sqrt{r_0}+z\sqrt{q_0-r_0}}{\sqrt{1-q_0}}\right)}{\hat{H}\left(-\frac{\gamma\sqrt{r_0}}{\sqrt{1-r_0}}\right)} \log \hat{H}\left(-\frac{z\sqrt{q_0-r_0}+\gamma\sqrt{r_0}}{\sqrt{1-q_0}}\right),$$

where

$$\hat{H}(x) = \int_{-\infty}^{\infty} Dy\, e^{-\beta V(y-x)}.$$

B.4 Replica computation for Binary perception tunable measure.

In this appendix, we present the replica computation for the tunable measure given in equation 36. The computation follows the steps performed for the cross-entropy loss measure, adding a Θ-function:

$$G_E = \log \int \prod_{\alpha} \frac{d\lambda_\alpha d\hat{\lambda}_\alpha}{2\pi} \prod_{a} \frac{du_a d\hat{u}_a}{2\pi} \prod_{\alpha} \Theta\left(\lambda_\alpha\right) e^{-\beta V(\lambda_\alpha)} \prod_{a} \Theta\left(u_a\right) e^{-\beta V(u_a)}$$
$$\times\, e^{-i\sum_\alpha \hat{\lambda}_\alpha \lambda_\alpha - i\sum_a \hat{u}_a u_a - \frac{1}{2}\sum_{\alpha\beta}\hat{\lambda}_\alpha\hat{\lambda}_\beta r_{\alpha\beta} - \frac{1}{2}\sum_{ab}\hat{u}_a\hat{u}_b q_{ab} - \sum_{\alpha a}\hat{\lambda}_\alpha \hat{u}_a p_{\alpha a}},$$

where $V(x) = \frac{1}{\gamma}\log\left(1+e^{-\gamma(x)}\right)$. Following the same steps as in the computation for the spherical perceptron, we obtain:

$$G_E = \int Dz D\gamma \frac{\tilde{H}\left(-\frac{\gamma\sqrt{r_0}+z\sqrt{q_0-r_0}}{\sqrt{1-q_0}}\right)}{\tilde{H}\left(-\frac{\gamma\sqrt{r_0}}{\sqrt{1-r_0}}\right)} \log \tilde{H}\left(-\frac{z\sqrt{q_0-r_0}+\gamma\sqrt{r_0}}{\sqrt{1-q_0}}\right),$$

where

$$\tilde{H}(x) = \int_{x}^{\infty} Dy\, e^{-\beta V(y-x)}$$

is a generalized version of the standard complementary error function.