

# Supplemental Material

## Capacity-resolution trade-off in the optimal learning of multiple low-dimensional manifolds by attractor neural networks

Aldo Battista and Rémi Monasson

### Contents

<b>I. Numerical results: Support Vector Machine (SVM) learning</b>	2
A. Dataset preparation: environments and distributions of PF	2
B. SVM learning procedure implementation	2
C. Couplings obtained by SVM	2
D. Miscellaneous results on optimal stability and capacity	3
1. Comparison with Hebb rule	3
2. Heterogeneous distribution of positions	3
3. Dependence on $\phi_0$ and $D$	4
E. Monte Carlo simulations	6
1. Zero temperature scheme ( $T = 0$ )	6
2. Finite temperature scheme ( $T > 0$ )	7
3. Description of videos	8
<b>II. Theoretical results: Statistical physics of optimal CANN</b>	8
A. Gardner's framework for CANN	8
B. Case of a single location per map ( $p = 1$ )	9
C. Gaussian theory with quenched PF	10
1. Replica calculation	10
2. Computation of $\Xi$	12
3. Expression of log. volume and saddle-point equations close to the critical line	13
4. Large- $p$ behavior of the critical capacity	14
D. Comparison between Quenched PF theory and SVM	15
1. Values of couplings	15
2. Dependence on $\phi_0$	16
<b>References</b>	17

## I. NUMERICAL RESULTS: SUPPORT VECTOR MACHINE (SVM) LEARNING

### A. Dataset preparation: environments and distributions of PF

We define an environment as a  $D$ -dimensional torus with unitary volume in which each neuron  $i = 1 \dots N$  of the Continuous Attractor Neural Network (CANN) has a randomly located place field, *i.e.* a  $D$ -dimensional hyper-sphere of volume  $\phi_0 < 1$ , centered at position  $\mathbf{r}_i$ . We want to store in the network  $L = \alpha N$  maps that differ through random rearrangements of the place-field (PF) center positions,  $\mathbf{r}_i^\ell$ ,  $\ell = 1 \dots L$ . Each map  $\ell$  is approximated through a collection of  $p$  random positions  $\mathbf{r}^{\ell,\mu}$  in the environment; the maps become continuous in the large  $p$  limit. For every position  $\mathbf{r}^{\ell,\mu}$  we extract a pattern of activity of the network in the following way: all the neurons  $i$  whose place field overlap with the position are active ( $\sigma_i = 1$ ), the others are silent ( $\sigma_i = 0$ ). A sketch representation of how we construct the patterns to store is drawn in the Fig. 1(b) of the main text. We end up with a data-set of  $p \times L$  binary patterns  $\{\sigma_i^{\ell,\mu}\}$ , where  $i$  is the neuron index,  $\ell$  the environment index and  $\mu$  the index of the position in map  $\ell$ .

### B. SVM learning procedure implementation

After we have generated a data-set of activity patterns, we want to learn the connections  $W_{ij}$  of the CANN that maximize the stability  $\kappa$  at fixed  $\alpha$  and  $p$ . This choice of the weights will ensure the biggest basins of attraction in the pattern space, *i.e.* robustness against thermal noise. In order to do that we will implement SVM learning [1]. In practice, for each neuron  $i$ , we want to compute the connections  $W_{ij}$  from the other neurons  $j$  (with  $W_{ii} = 0$ , no self-connection), which are solution of the following primal constrained convex optimization problem

$$\begin{aligned} & \underset{\{W_{ij}\}}{\text{minimize}} && \frac{1}{2} \sum_{j(\neq i)} W_{ij}^2, \\ & \text{subject to} && (2\sigma_i^{\ell\mu} - 1) \sum_{j(\neq i)} W_{ij} \sigma_j^{\ell\mu} \geq 1, \quad \forall \ell, \mu. \end{aligned} \quad (1)$$

We have to solve  $N$  such problems to extract all the rows of the coupling matrix. The dual form of this problem is

$$\begin{aligned} & \underset{\{\lambda_{\ell,\mu}\}}{\text{maximize}} && \sum_{\ell=1}^L \sum_{\mu=1}^p \lambda_{\ell\mu} - \frac{1}{2} \sum_{\ell,m=1}^L \sum_{\mu,\nu=1}^p (2\sigma_i^{\ell\mu} - 1)(2\sigma_i^{m\nu} - 1) \lambda_{\ell\mu} \lambda_{m\nu} \sum_{j(\neq i)} \sigma_j^{\ell\mu} \sigma_j^{m\nu}, \\ & \text{subject to} && \lambda_{\ell,\mu} \geq 0, \quad \forall \ell, \mu, \end{aligned} \quad (2)$$

where the  $\lambda_{\ell\mu}$ 's are Lagrange multipliers enforcing the constraints in (1). This optimization problem can be solved using available numerical routines [2]. Once we obtain the  $\lambda_{\ell\mu}$ 's we can compute the connections through

$$W_{ij} = \sum_{\ell=1}^L \sum_{\mu=1}^p \lambda_{\ell\mu} (2\sigma_i^{\ell\mu} - 1) \sigma_j^{\ell\mu}. \quad (3)$$

We then normalize the rows of the couplings matrix to unity, *i.e.*  $\sum_{j(\neq i)} W_{ij}^2 = 1$ . Finally, the stability  $\kappa$  is computed through formula (3) of the main text. We have checked that the same values for  $\kappa$  are obtained with a standard package for SVM, LinearSVC [3].

As an illustration of the learning procedure, we show in Fig. 1 how the number of stored patterns (with positive stabilities) grows as a function of the number of iterations of the quadratic optimization algorithm solving (2), until all  $p$  prescribed patterns are stabilized.

### C. Couplings obtained by SVM

Hereafter, we report some qualitative features of the couplings obtained by SVM. As shown in Fig. 2 the couplings  $W_{ij}$  are correlated with the distances  $d_{ij}^\ell = |\mathbf{r}_i^\ell - \mathbf{r}_j^\ell|$  between the PF centers of the neurons  $i$  and  $j$  in the different maps  $\ell$ . Note that the dependence on distance is less marked as the number  $L$  of maps increases, due to the interferences between the maps.

In order to sustain a bump state with average activity  $\phi_0$ , couplings are excitatory at short distances, up to roughly the radius  $r_c$  of the PF, and inhibitory at longer ones. The sign of the couplings can be intuitively understood. Two

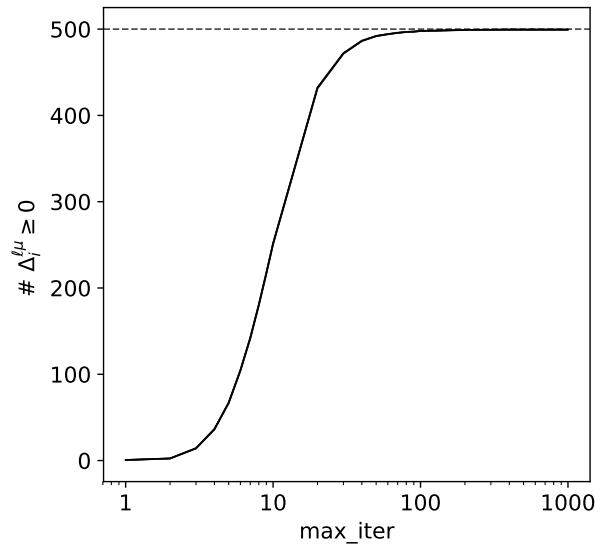


FIG. 1: Number of patterns with positive stabilities ( $y$ -axis) vs. number of iterations of the quadratic optimization solver ( $x$ -axis) for one map ( $L = 1$ ) with  $p = 500$  points stored by a network with  $N = 1000$  neurons. Parameter values:  $D = 2$ ,  $\phi_0 = .3$ .

neurons at short distances have largely overlapping PF: their activities are likely to be equal, and having a large coupling helps increasing the stability, see eqn (3) in main text. If the distance is bigger than  $r_c$ , the activities are likely to be different, hence inhibitory (negative) couplings will increase the stability.

Histograms of couplings in Fig. 2 (right) show that the amplitudes decay with  $N$ . In agreement with [6] we expect the average values and standard deviations to scale, respectively, as  $1/N$  and  $1/\sqrt{N}$ .

#### D. Miscellaneous results on optimal stability and capacity

##### 1. Comparison with Hebb rule

Here we show that, as it should be by construction, the stability obtained by SVM is always much higher than the one obtained by the Hebb rule (2) defined in the main text. In order to do that we consider the cases of an exponential kernel,

$$w(d) = a e^{-d/b} - 1, \quad (4)$$

and of a Gaussian kernel,

$$w(d) = a e^{-d^2/b} - 1. \quad (5)$$

We then optimize over  $a$  and  $b$ ; the value of the negative offset at large distance is arbitrary, since couplings are normalized row by row. Results for a typical sample are shown in Fig. 3. The stability for the best kernel  $w$  is always much lower (and negative in the examples considered here) than the optimal stability  $\kappa$  found with SVM.

##### 2. Heterogeneous distribution of positions

Throughout this work we have considered for simplicity that the  $p$  positions were drawn uniformly at random to produce statistically homogeneous maps, *i.e.* without preferred positions. It is straightforward to extend this setting to the case of heterogeneous densities of prescribed positions.

Figure 4 show the spatial distribution of stabilities for homogeneously scattered points (left) and a heterogeneous repartition on points, densely packed in a subregion (diagonal). In the latter case, the strong heterogeneity in the local distribution of stabilities will favor the location of the bump along the zones with a major density of positions. As a consequence, a  $1D$ -attractor is effectively built in the  $D = 2$ -dimensional space, see videos described in Section I.E.3.

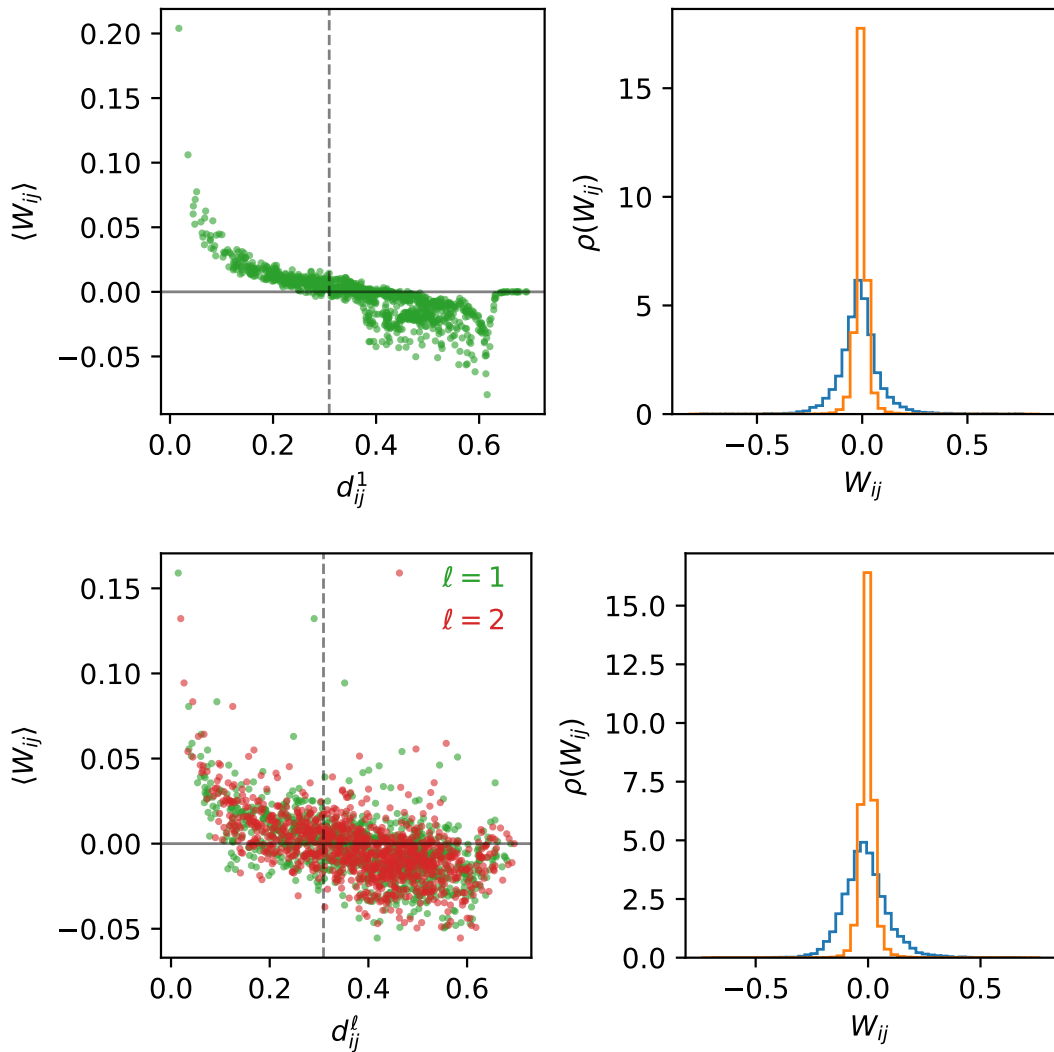


FIG. 2: Couplings obtained after training with SVM for  $L = 1$  (top) and  $L = 2$  (bottom) maps. Left: dependence of the average coupling with the distance between the corresponding neurons; the vertical line locates the radius  $r_c$  of the place fields. Averages were computed over 500 samples of the  $p$  positions per map at fixed PF centers;  $N = 1000$  neurons. Right: histograms of the couplings, for sizes  $N = 100$  (blue) and  $N = 1000$  (orange). Parameters:  $D = 2$ ,  $\phi_0 = .3$  and  $p \times L = N$ .

### 3. Dependence on $\phi_0$ and $D$

Figure 5(left) indicates that the behaviour of  $\alpha_c(p)$  vs.  $p$  changes from a  $\frac{1}{p}$ -scaling to a slower decay at a cross-over value  $p_{c.o.} \simeq \frac{1}{\phi_0}$ .

Figure 5(right) shows that the behaviour of  $\alpha_c(p)$  with  $\phi_0$ , obtained from SVM, is in qualitative agreement with equation (10) in the main text. In particular, we see that the critical capacity is largely independent of  $\phi_0$  in dimension  $D = 1$ , while it increases as the PF size  $\phi_0$  shrinks in dimensions  $D = 2$ , and even more so for  $D = 3$ . Notice that the agreement with the asymptotic result given in equation (10) of the main text is not perfect here, due to the moderate value of the number of points in simulations ( $p = 100$ ).

Last of all, Fig. 6 shows the spatial error of trained recurrent neural network and the optimal stability  $\kappa$  vs. the load  $\alpha$  for patterns generated in dimensions  $D = 1$  and  $3$ , completing the results shown for  $D = 2$  in the main text. We observe the faster decay of the critical capacity predicted by equation (10) in the main text with increasing values of  $D$ .

In these figures, as well as in the ones shown in the main text, the critical capacity  $\alpha_c(p)$  was estimated as follows from SVM data. We estimated the optimal stabilities  $\kappa$  (at fixed  $p$ ) for  $M$  different values of the load  $\alpha$ , with  $M$

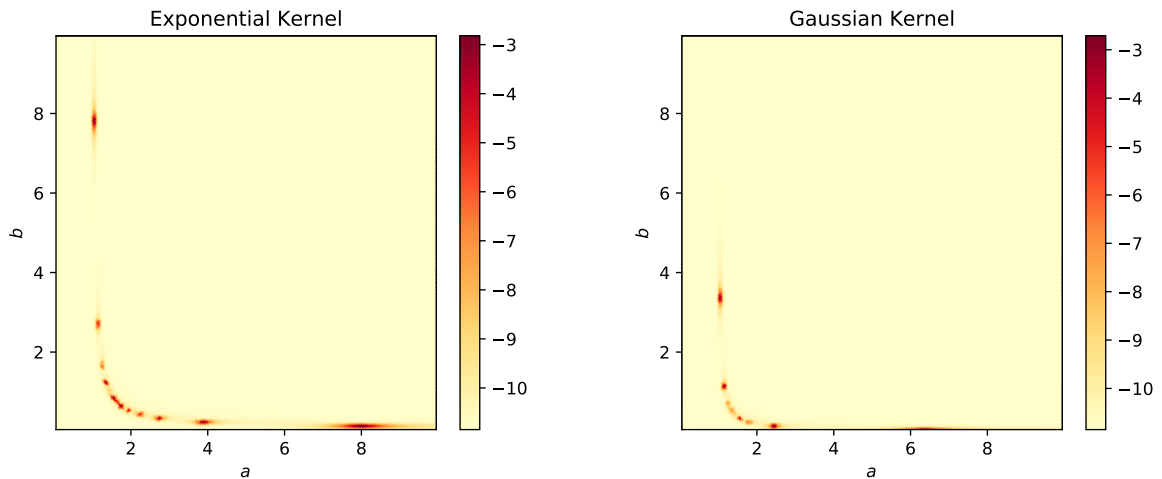


FIG. 3: Stabilities obtained with the Hebb rule with exponential (left) and Gaussian (right) kernels on a given representative sample. The kernel parameters  $a$  and  $b$  vary from 0 to 10 with a step of .01. Parameter values:  $N = 1000$ ,  $D = 2$ ,  $\phi_0 = .3$ ,  $\alpha = .1$  and  $p = 5$ . The optimal value of the stability on that sample obtained by SVM is  $\kappa \simeq .55$ .

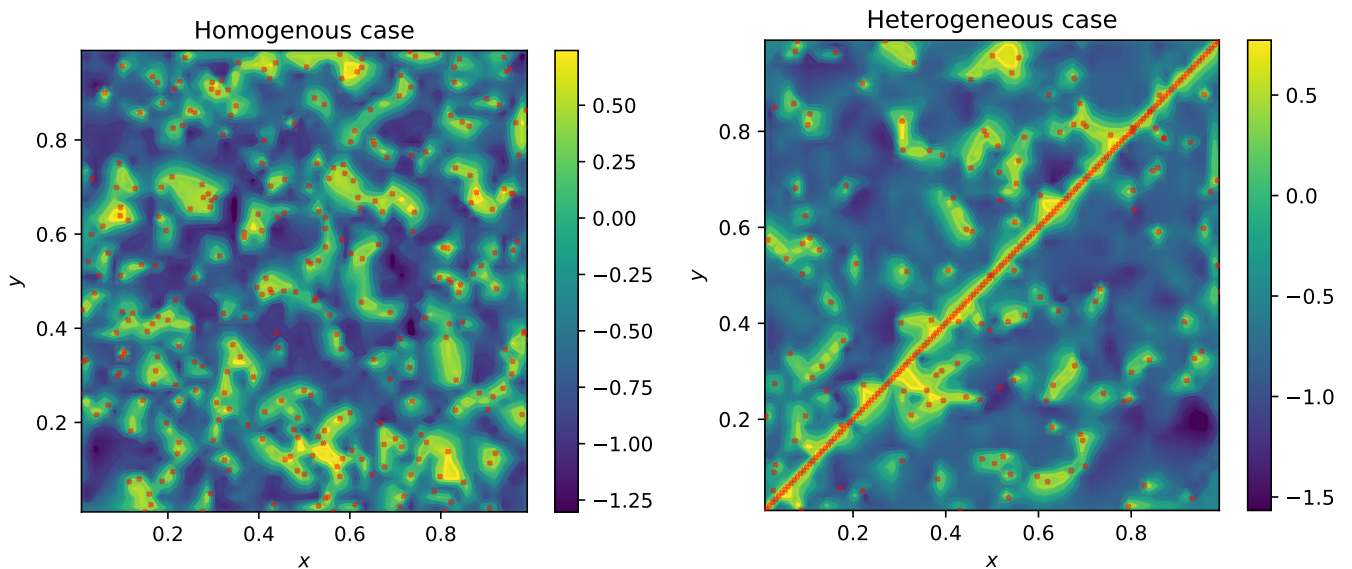


FIG. 4: Distribution of local stabilities after the learning of a map with SVMs. Left) Homogeneous case: the  $p$  positions of the data-set are drawn randomly. Right) Heterogeneous case: here the data-set has 150 positions on the diagonal of the maps and the other 150 positions are drawn at random. Here,  $D = 2$ ,  $\phi_0 = .3$ ,  $N = 1000$ ,  $p = 300$  and  $L = 1$ . We show the contour map made from 2500 realization of random positions, for which we evaluate the stabilities of the corresponding patterns. The overall network stabilities (minimal pattern stabilities) in the homogeneous and heterogeneous cases are, respectively,  $\kappa \simeq .44$  and  $\kappa \simeq .53$  for the samples considered here.

generally equal to 20. Then we fitted these points with the empirical function (depending on the parameters  $a, b, c$ )

$$\kappa = \frac{a}{\sqrt{\alpha}} + b\alpha + c, \quad (6)$$

and extrapolated from the fit the value of the load at which the fitted function vanished; this defined our estimate for  $\alpha_c(p)$ . Note that the small  $\alpha$  behaviour in equation (6) above can be justified analytically from Gardner's calculation.

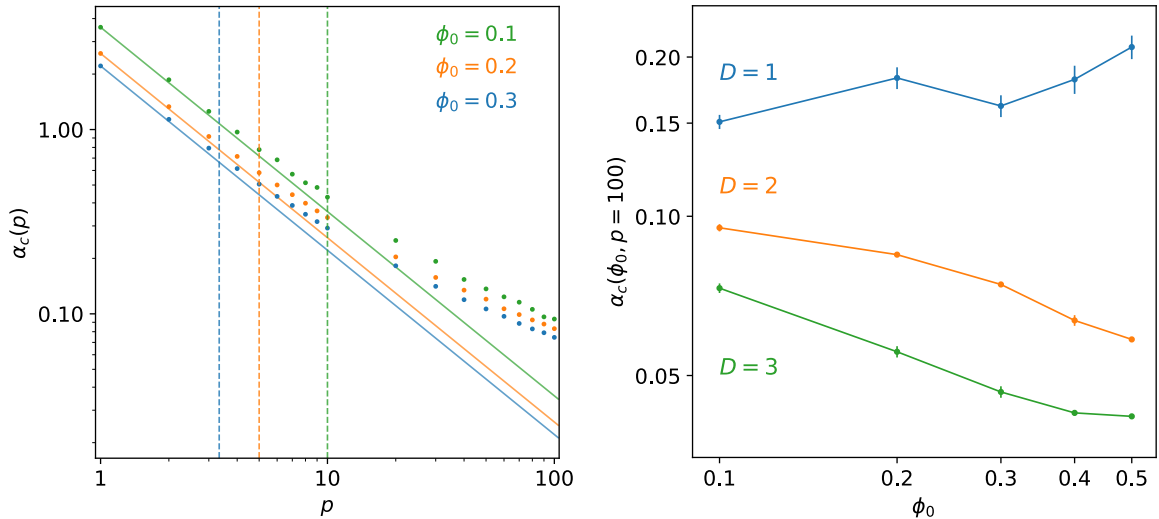


FIG. 5: (Left) Scaling cross-over of  $\alpha_c(p)$  vs.  $p$  for different values of  $\Phi_0$ . The vertical lines correspond to the values of  $p_{c.o.} \sim \frac{1}{\phi_0}$ . We use for this results  $D = 2$ ,  $N = 5000$ , and we have averaged over 50 different realization of the environments and different realizations of the  $p$  positions. (Right) Critical capacity obtained by SVM vs.  $\phi_0$  for different values of  $D$  in log-log scale. Parameter values:  $N = 5000$ ,  $p = 100$ , Samples=25.

### E. Monte Carlo simulations

Once the coupling matrix  $W_{ij}$  has been learned, we may perform Monte Carlo simulations to investigate the behavior of the network.

#### 1. Zero temperature scheme ( $T = 0$ )

In order to compute the spatial error  $\epsilon$  we consider that the dynamics of the system follows a sequential updating rule of the form

$$\sigma_i(t+1) = \Theta\left(\sum_{j(\neq i)} W_{ij} \sigma_j(t)\right), \quad (7)$$

where  $\Theta$  is the Heaviside step-function and at every time  $t+1$  we choose uniformly at random the neuron  $i$  to update. Starting from an initial activity configuration, we track the system dynamics for at most  $N^2$  MC steps ( $N$  sweeps), and retain the visited configuration with the minimum number of violated constraints, *i.e.* with the highest number of non-negative stabilities

$$\Delta_i = (2\sigma_i - 1) \sum_{j(\neq i)} W_{ij} \sigma_j \geq 0. \quad (8)$$

The choice of  $N$  sweeps as a maximal simulation time is empirical: we do not find that significantly better results are obtained by increasing this bound. Actually, the dynamics often ends up in a fixed point with  $\Delta_i > 0$  for all neurons  $i$  in much less sweeps.

We generate  $L$  environments and  $p$  points in each of them, learn the coupling matrix corresponding to these  $p \times L$  patterns. We then pick at random a position in one of the learned maps, and use that position to construct the initial activity configuration of the dynamics. After the dynamics described above is done we keep the final configuration and use it to decode the final position on that map, as the center of mass of PF (on the map) of the active neurons in the final configuration. The distance between this estimated position and the initial one (taking care of the periodic boundary conditions), after averaging over many initial positions (100 in the figures showed), defines the spatial error  $\epsilon$ .

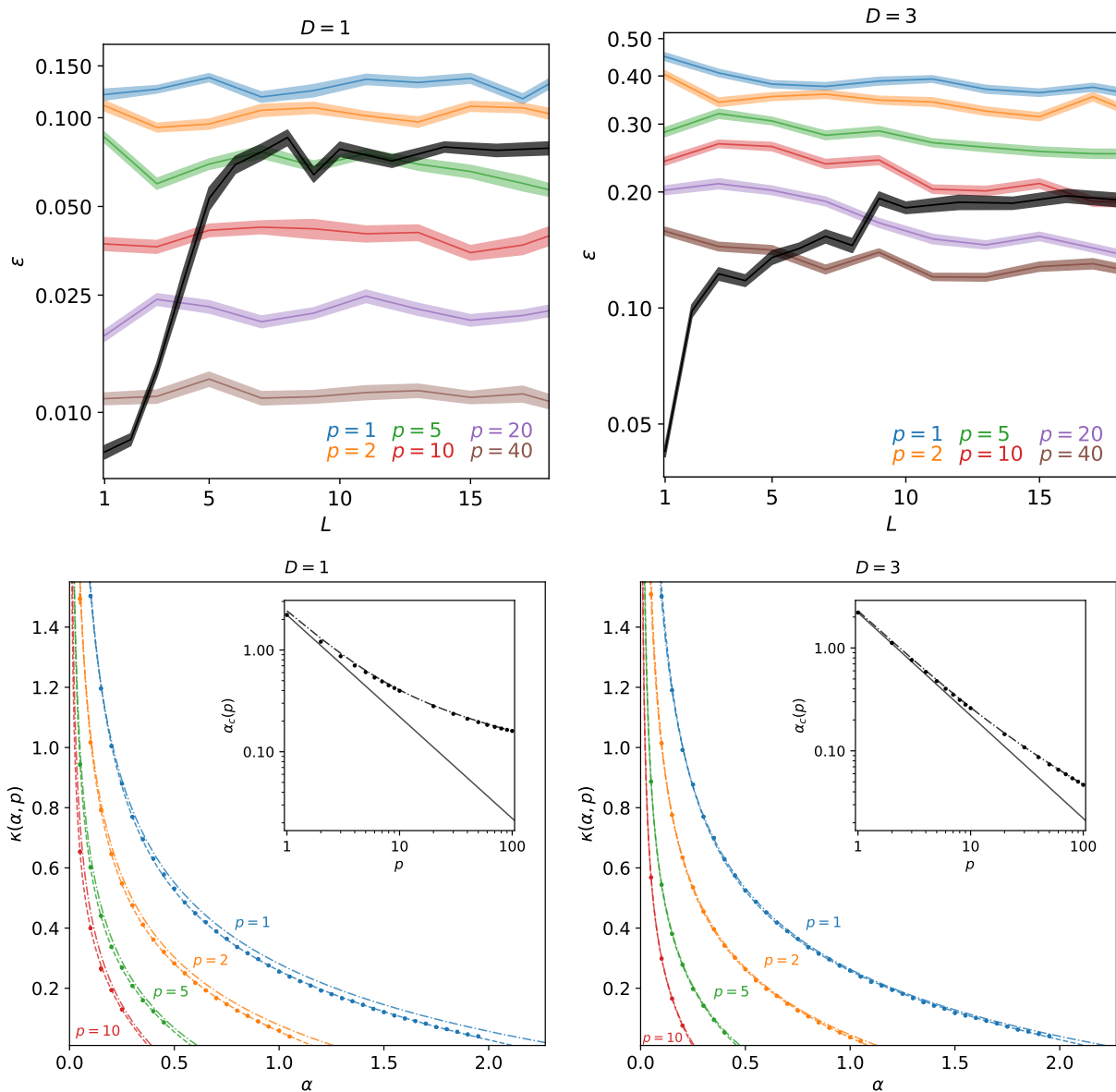


FIG. 6: Same as Fig. 2(a) and Fig. 3 in the main text for dimensions  $D = 1$  (left) and  $D = 3$  (right).

## 2. Finite temperature scheme ( $T > 0$ )

In order to show the diffusion of the activity bump within a map and the transitions between maps, we implement a noisy dynamical scheme, where neuronal states are updated stochastically according to the probabilities

$$\text{Prob}(\sigma_i(t+1) | \{\sigma_j(t)\}) = \frac{1}{1 + \exp\left[-\frac{1}{T}(2\sigma_i(t+1) - 1) \sum_{j(\neq i)} W_{ij} \sigma_j(t)\right]}. \quad (9)$$

The bump of activity may form and sustain itself when  $T$  is comparable to, or smaller than the stability  $\kappa$  of the network.

### 3. Description of videos

We illustrate the dynamical properties of the model with three examples:

- First, we consider a network with  $N = 1000$  neurons, in which we store one map ( $L = 1$ ) in dimension  $D = 2$  and with average activity  $\phi_0 = .3$ . We consider than the case in which the learning is done on a small number of points,  $p = 30$ , resulting in a large value of the stability,  $\kappa = 1.7$ . And then the case in which  $p = 300$  is higher, and the stability is smaller:  $\kappa = .6$ . Our noise parameter  $T$  is set to  $.8$  to allow the bump to form in both cases. In the large  $\kappa$  case, the bump gets stuck very quickly in one of the  $p$  training positions, depending on the initial configuration, see attached videos `LargeKappaL1.mp4` and `LargeKappaL1Bis.mp4`. In the small  $\kappa$  case, the bump diffuses on the map, see attached video `SmallKappaL1.mp4`. For larger  $p$ , the bump can easily travel through the environment, with a large diffusion coefficient; in contrast, in the small  $p$  case, the stability landscape is very rough and the bump is stucked close to the stored positions.
- In the second example we consider the case of  $L = 2$  maps and  $p = 150$  points. The other parameters have the same values as in the first example, *e.g.* the stability is fixed to  $\kappa = .5$ . In the video `SmallKappaL2.mp4` we see that, as  $\kappa$  is small, the bump diffuses in one maps and sporadically jumps to the other map.
- The third example corresponds to the heterogeneous distribution of positions shown in Fig. 4, right. The video `SmallKappaL1Hetero.mkv` was obtained with the same parameters as in `SmallKappaL1.mp4`, but with 150 out of the  $p = 300$  positions drawn on the diagonal of the map; the stability of the network was  $\kappa = .5$ .

## II. THEORETICAL RESULTS: STATISTICAL PHYSICS OF OPTIMAL CANN

### A. Gardner's framework for CANN

Here we are going to extend the Gardner theory for capacity of the perceptron (SVM with linear kernel and hard margin) [4] to the case of continuous attractors. The training set consist of  $p \times L$  binary patterns  $\{\sigma_i^{\ell\mu}\}$  constructed by drawing randomly  $p$  positions in each of the  $L$  environments so that the resulting patterns are spatially correlated. The stability of the  $i$  component of the pattern that correspond to position  $\mu$  in the environment  $\ell$  is given by

$$\Delta_i^{\ell\mu} = (2\sigma_i^{\ell\mu} - 1) \sum_{j(\neq i)} W_{ij} \sigma_j^{\ell\mu}. \quad (10)$$

The training set is said to be stored if all the patterns have stabilities larger than some threshold  $\kappa \geq 0$ .

The volume in the space of couplings that corresponds to admissible solutions of the storage problem, is

$$Z = \int \prod_{i \neq j}^N dW_{ij} \prod_i \delta\left(\sum_{j(\neq i)} W_{ij}^2 - 1\right) \prod_{i,\ell,\mu} \theta\left((2\sigma_i^{\ell\mu} - 1) \sum_{j(\neq i)} W_{ij} \sigma_j^{\ell\mu} - \kappa\right) \quad (11)$$

and is equal to the product of the  $N$  single-site volumes  $Z_i$ , with  $i = 1 \dots N$ . So we may focus for example on the volume associated with  $i = 1$ :

$$Z_1 = \int \prod_{j=2}^N dW_j \delta\left(\sum_{j \geq 2} W_j^2 - 1\right) \prod_{\ell,\mu} \theta\left((2\sigma_1^{\ell\mu} - 1) \sum_{j \geq 2} W_j \sigma_j^{\ell\mu} - \kappa\right), \quad (12)$$

where  $W_j \equiv W_{1j}$ . Using the replica method [5], we compute the average of  $\log Z_1$  over the patterns. Introducing integral representations of the Heaviside function and exploiting the statistical independence of the different maps, we write the average of the  $n^{\text{th}}$  power of the volume,

$$\langle Z_1^n \rangle = \int \prod_{j,a} dW_{ja} \prod_a \delta\left(\sum_j W_{ja}^2 - 1\right) \chi(\mathbf{W})^{\alpha N} \quad (13)$$

with  $a = 1, \dots, n$  is the replica index, and

$$\chi(\mathbf{W}) = \int \prod_{\mu=1}^p d\hat{\mathbf{r}}_\mu \int \prod_{j=1}^N d\mathbf{r}_j \int_{\kappa}^{\infty} \prod_{\mu,a} dt_{\mu a} \int_{-\infty}^{\infty} \prod_{\mu,a} \frac{d\hat{t}_{\mu a}}{2\pi} e^{i \sum_{\mu,a} \hat{t}_{\mu a} t_{\mu a}} \prod_j e^{-i \sum_{\mu,a} \hat{t}_{\mu a} (2\sigma_1^\mu - 1) W_{ja} \sigma_j^\mu}, \quad (14)$$



where  $\hat{\mathbf{r}}_\mu$  denotes the  $p$  prescribed locations in the environment, and  $\mathbf{r}_j$  the  $N$  PF of the neurons in the map. We first carry out explicitly the integrals over the PF with indices  $j = 2, 3, \dots, N$ , leaving the integrals over  $\mathbf{r}_1$  and all  $\hat{\mathbf{r}}_\mu$  in  $\chi(\mathbf{W})$ . We introduce the order parameters

$$m^a = \phi_0 \sum_{j \geq 2} W_{ja} \quad (15)$$

and

$$q^{ab} = \sum_{j \geq 2} W_{ja} W_{jb} , \quad (16)$$

and rewrite

$$\begin{aligned} \langle Z_1^n \rangle &= \int \prod_{j,a} dW_{j,a} \int \prod_a \frac{d\hat{u}^a}{4\pi} e^{\sum_a \frac{\hat{u}^a}{2} (1 - \sum_j W_{ja}^2)} \int \prod_a \frac{d\hat{m}^a dm^a}{2\pi} e^{\sum_a \hat{m}^a (m^a - \sum_j W_{ja})} \\ &\quad \times \int \prod_{a \leq b} \frac{d\hat{q}^{ab} dq^{ab}}{2\pi} e^{\sum_{a \leq b} \hat{q}^{ab} (q^{ab} - \sum_j W_{ja} W_{jb})} \chi(\mathbf{W})^{\alpha N} \end{aligned} \quad (17)$$

where we have used the integral representation of the Dirac-delta function. Let  $\Phi(\mathbf{r})$  be the indicator function of the place field centered in  $\mathbf{O}$ :  $\Phi = 1$  if  $|\mathbf{r}| < r_c$ , where  $r_c$  is the radius of the PF (with  $\int d\mathbf{r} \Phi(\mathbf{r}) = \phi_0$ ), and 0 otherwise. Let  $\Gamma(\mathbf{r}) = \int d\mathbf{r}' \Phi(\mathbf{r}') \Phi(\mathbf{r} - \mathbf{r}')$  the correlation function of  $\Phi$ . Given  $p$  points  $\hat{\mathbf{r}}_\mu$ ,  $\mu = 1, \dots, p$  drawn uniformly at random in space, we define the  $p \times p$  Euclidean random matrix with entries

$$\mathbf{\Gamma}_{\mu,\nu}(\hat{\mathcal{R}} \equiv \{\hat{\mathbf{r}}_\mu\}) = \Gamma(\mathbf{r}_\mu - \mathbf{r}_\nu) - \phi_0^2 . \quad (18)$$

We can rewrite  $\chi$  as

$$\begin{aligned} \chi(\mathbf{W}) &= \int \prod_\mu d\hat{\mathbf{r}}_\mu \int d\mathbf{r}_1 \int_\kappa \prod_{\mu,a} \frac{dt_{\mu a}}{\sqrt{2\pi}} \int_{-\infty}^\infty \prod_{\mu,a} \frac{d\hat{t}_{\mu a}}{\sqrt{2\pi}} e^{-\frac{1}{2} \sum_{\mu,\nu,a,b} q^{ab} \mathbf{\Gamma}_{\mu,\nu}(\hat{\mathcal{R}}) \hat{t}_{\mu a} \hat{t}_{\nu b}} \\ &\quad e^{-i \sum_{\mu,a} m^a \hat{t}_{\mu a} \Phi(\mathbf{r}_1 - \hat{\mathbf{r}}_\mu) + i \sum_{\mu,a} \hat{t}_{\mu a} t_{\mu a}} . \end{aligned} \quad (19)$$

Due to translation invariance, the integral over  $\mathbf{r}_1$  is irrelevant, and we can set  $\mathbf{r}_1 = \mathbf{O}$ . We can now make the RS Ansatz (expected to be valid since the domain of suitable couplings is convex) on the structure of the order parameters and their conjugate variables, and, after standard manipulation, we write the  $n^{\text{th}}$  power of the volume in the small  $n$  limit as

$$\begin{aligned} \frac{\langle Z_1^n \rangle - 1}{nN} &\simeq \frac{1}{2\epsilon} \left\{ 1 - \alpha \int \prod_\mu d\hat{\mathbf{r}}_\mu \int \prod_\mu \frac{dz_\mu}{\sqrt{2\pi}} \frac{\exp\left(-\frac{1}{2} \sum_{\mu,\nu} z_\mu \mathbf{\Gamma}(\hat{\mathcal{R}})_{\mu,\nu}^{-1} z_\nu\right)}{\sqrt{\det \mathbf{\Gamma}(\hat{\mathcal{R}})}} \right. \\ &\quad \left. \times \min_{\{t_\mu \geq \kappa + w \phi_0\}} \sum_{\mu,\nu} [t_\mu - (z_\mu + 2m \Phi(\mathbf{r}_1 - \hat{\mathbf{r}}_\mu))] \mathbf{\Gamma}(\hat{\mathcal{R}})_{\mu,\nu}^{-1} [t_\nu - (z_\nu + 2m \Phi(\mathbf{r}_1 - \hat{\mathbf{r}}_\nu))] \right\} + O\left(\left|\log \frac{1}{\epsilon}\right|\right) , \end{aligned} \quad (20)$$

where we have computed the integrals over the order parameters and related conjugate variables thanks to the saddle-point method. Since we are interested in the critical capacity we have also restricted our analysis to the case  $\epsilon = 1 - q \ll 1$ , in which the space of solutions reduces to the optimal coupling matrix. We finally obtain the expression for the critical capacity  $\alpha_c(\kappa, p) = \max_m \alpha_c(m; \kappa, p)$ , where  $\alpha_c(m; \kappa, p)$  is the load  $\alpha$  cancelling the terms inside the curly brackets in (20).

## B. Case of a single location per map ( $p = 1$ )

We now show that, when a single pattern is present in each map ( $p = 1$ ), the equations above are equivalent to the celebrated Gardner calculation in the case of biased patterns [4].

For  $p = 1$ , the Euclidean random matrix  $\mathcal{C}$  reduces to the scalar

$$\mathcal{C}_{1,1} = \phi_0(1 - \phi_0) \equiv \frac{1 - M^2}{4} , \quad (21)$$

where  $M$  is the average activity of the binary pattern in  $\pm 1$  notations, *i.e.* under the change of variable  $\Phi(\mathbf{r}_1 - \hat{\mathbf{r}}_1) = \{0, 1\} \rightarrow \xi = \{-1, +1\}$ . The convex optimization problem to be solved in (20) thus amounts to compute

$$F(z_1, v, \kappa) = \min_{\{t_1 \geq \kappa\}} \left[ \frac{4}{1 - M^2} (t - (z_1 + v\xi))^2 \right], \quad (22)$$

where  $v = m\phi_0$  and the Gaussian variable  $z_1$  in (20) has zero mean and variance  $\mathcal{C}_{1,1}$ . The minimum over  $t_1$  in (22) can easily be determined, with the result

$$F(z_1, v, \kappa) = \begin{cases} \frac{4}{1 - M^2} (\kappa - (z_1 + v\xi))^2 & \text{if } \kappa \geq z_1 + v\xi \\ 0 & \text{otherwise.} \end{cases} \quad (23)$$

As  $\hat{\mathbf{r}}_1$  is drawn uniformly at random,  $\xi$  is a random binary variable:

$$\xi = \begin{cases} +1 & \text{with probability } \frac{1 + M}{2}, \\ -1 & \text{with probability } \frac{1 - M}{2}. \end{cases} \quad (24)$$

We get, with the normalized Gaussian variable  $z = z_1 \times 2/\sqrt{1 - M^2}$  and the measure  $Dz = dz/\sqrt{2\pi} \exp(-z^2/2)$ ,

$$\frac{1}{\alpha_c(v; \kappa, p = 1)} = \frac{1 + M}{2} \int_{\frac{2vM - 2\kappa}{\sqrt{1 - M^2}}}^{\infty} Dz \left( \frac{2\kappa - 2vM}{\sqrt{1 - M^2}} + z \right)^2 + \frac{1 - M}{2} \int_{\frac{-2\kappa - 2vM}{\sqrt{1 - M^2}}}^{\infty} Dz \left( \frac{2\kappa + 2vM}{\sqrt{1 - M^2}} + z \right)^2, \quad (25)$$

where  $v$  is chosen in order to maximize  $\alpha_c(v; \kappa, p = 1)$ :

$$\frac{1 + M}{2} \int_{\frac{2vM - 2\kappa}{\sqrt{1 - M^2}}}^{\infty} Dz \left( \frac{2\kappa - 2vM}{\sqrt{1 - M^2}} + z \right) = \frac{1 - M}{2} \int_{\frac{-2\kappa - 2vM}{\sqrt{1 - M^2}}}^{\infty} Dz \left( \frac{2\kappa + 2vM}{\sqrt{1 - M^2}} + z \right). \quad (26)$$

These equations coincide with the results of [4] up to the change  $\kappa \rightarrow 2\kappa$  due to the fact that the neuron activities take here values 0,1 and not  $\pm 1$ .

### C. Gaussian theory with quenched PF

In order to compute the critical capacity  $\alpha_c(\kappa)$  with the approach of Section II A we have to solve a  $p$ -dimensional constrained quadratic optimization problem, depending on  $p$  correlated Gaussian random variables, see (20), and then average over  $p$  random positions. This task becomes quickly intractable in practice as  $p$  increases. In this section, following closely [6], we present an alternative approximate approach that allows us to reach arbitrarily large values of  $p$ . While this calculation is approximate, it is argued that it becomes exact in the large  $p$  limit. A potentially interesting feature of this approach is that it holds at fixed PF, instead of averaging over them as in Section II A, and could be applied to specific situations, *e.g.* sets of PF measured in experiments.

#### 1. Replica calculation

Starting from the replicated volume  $\langle Z_1^n \rangle$  in (13), we now perform first the average over each one of the  $p$  locations in  $\chi(\mathbf{W})$  in (14) as follows

$$\int d\hat{\mathbf{r}}_\mu \exp \left( -i \sum_a \hat{t}_{\mu a} (2\sigma_1^{\ell, \mu} - 1) \sum_{j \geq 2} W_{ja} \sigma_j^{\ell, \mu} \right) = \exp \left( -i \sum_a m_\ell^a \hat{t}_{\mu a} - \frac{1}{2} \sum_{a,b} \hat{t}_{\mu a} (q_\ell^{ab} - m_\ell^a m_\ell^b) \hat{t}_{\mu b} + O(\hat{t}^3) \right) \quad (27)$$

where we have reintroduced the map index  $\ell$  to underline that the PF are kept fixed here. The order parameters in the formula above are

$$m_\ell^a = \sum_{j \geq 2} W_{ja} \left( 2\Gamma(|\mathbf{r}_j^\ell - \mathbf{r}_1^\ell|) - \phi_0 \right) \quad (28)$$

and

$$q_\ell^{ab} = \sum_{j,k \geq 2} W_{ja} W_{kb} \Gamma(|\mathbf{r}_j^\ell - \mathbf{r}_k^\ell|) . \quad (29)$$

We simplify the calculation with two approximations:

- We truncate the expansion in powers of  $\hat{t}$  in (27) to the second order, and omit all higher order terms. This amounts to approximate the distribution of couplings  $W_{ij}$  (at fixed PF) by a Gaussian. This approximation is valid only if the couplings fluctuate weakly around their means, which is the case in the large- $p$  limit, see Section II.C.4.
- We also neglect the dependence of the order parameters  $m_\ell$  and  $q_\ell$  above on the map  $\ell$ . The histogram of the overlaps  $q_\ell$  measured by SVM are shown in Fig. 7. As can be seen from the figure, the distribution of overlaps is not concentrated in the large- $N$  limit at fixed  $p$ . Therefore, while  $m_\ell^a = m^a$  and  $q_\ell^{ab} = q^{ab}$  is a valid Ansatz for the saddle-point equations of the log. partition function (due to the statistical equivalence between the maps), we expect Gaussian fluctuations to be relevant even in the infinite- $N$  limit. However, as  $p$  increases, these fluctuations are smaller and smaller, and are asymptotically negligible. The order parameters then reduce to, after summation over the maps  $\ell = 1 \dots L$ ,

$$m^a \equiv \frac{1}{L} \sum_{\ell=1}^M m_\ell^a = \sum_{j \geq 2} W_{ja} \left( 2 \mathcal{C}_{1j}(\{\mathbf{r}_j^\ell\}) - \phi_0 \right) \quad (30)$$

and

$$q^{ab} \equiv \frac{1}{L} \sum_{\ell=1}^M q_\ell^{ab} = \sum_{j,k \geq 2} W_{ja} W_{kb} \mathcal{C}_{jk}(\{\mathbf{r}_j^\ell\}) . \quad (31)$$

The  $N \times N$  multi-space Euclidean random matrix  $\mathcal{C}$  appearing in the expressions above is defined in equation (8) of the main text. In the following, we denote by  $\rho(\lambda)$  the density of eigenvalues  $\lambda$  of  $\mathcal{C}$ . This density is self-averaging when the PF are randomly drawn in the large  $L, N$  double limit. Its resolvent, defined as

$$g(U) = \int d\lambda \frac{\rho(\lambda)}{\lambda + U} , \quad (32)$$

where the integral runs over the support of  $\rho$ , is solution of the implicit equation (9) of the main text [7].

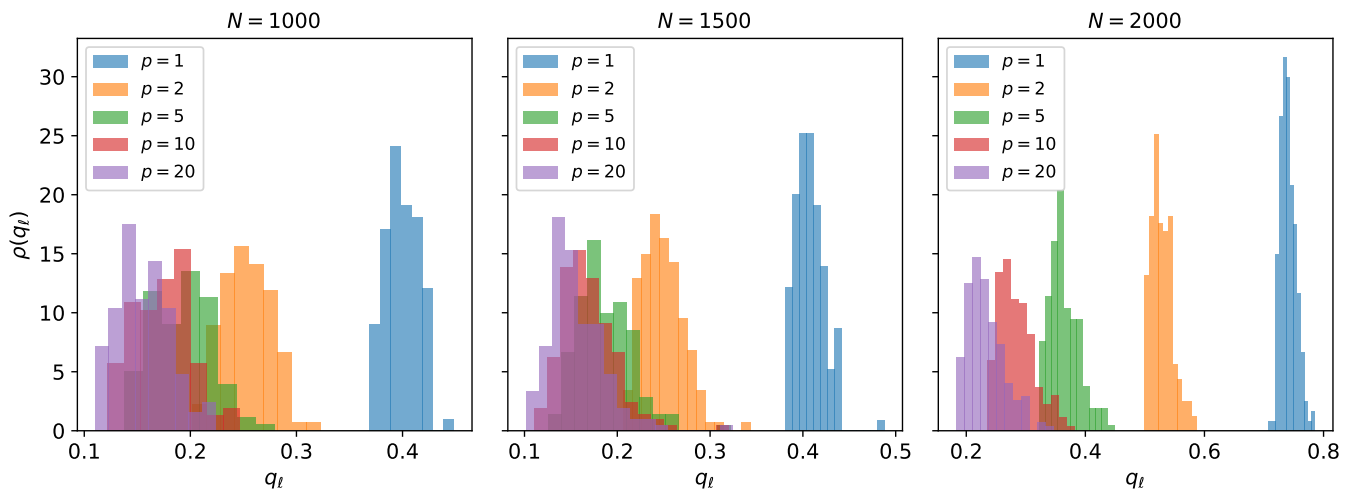


FIG. 7: Distributions of the overlaps  $q_\ell$  for different values of  $N$  and  $p$ . It is clear that the histograms are roughly Gaussian. We use for this results  $D = 2$ ,  $\phi_0 = .3$ ,  $\alpha = .1$ , and we have averaged over 500 realization of the  $p$  positions at fixed PF.

Within the RS Ansatz, the overlap matrix  $q^{ab}$  is fully characterized by its diagonal and off-diagonal elements that we denote by, respectively,  $s$  and  $q$ :

$$s = \sum_{i,j \geq 2} \langle \mathbf{C}_{ij} [W_{1i} W_{1j}] \rangle \quad , \quad q = \sum_{i,j \geq 2} \langle \mathbf{C}_{ij} [W_{1i}] [W_{1j}] \rangle . \quad (33)$$

where, as before, the brackets denotes the average over the random patterns, and the square parenthesis stand for the average over all couplings satisfying the inequalities (1).

Following closely [6], we obtain the expression of the average logarithm of the volume,

$$\begin{aligned} \frac{\langle \log Z_1 \rangle}{N} = & -\frac{1}{2}q\hat{q} + s\hat{s} + m\hat{m} + \hat{u} - \frac{1}{2} \int d\lambda \rho(\lambda) \left[ \log(2\hat{u} + (2\hat{s} - \hat{q})\lambda) + \frac{\hat{q}\lambda}{2\hat{u} + (2\hat{s} - \hat{q})\lambda} \right] \\ & + \frac{\hat{m}^2 \Xi}{2(2\hat{s} - \hat{q})} + \alpha p \int Dz \log H \left( \frac{z\sqrt{q - m^2} - m + \kappa}{\sqrt{s - q}} \right) \end{aligned} \quad (34)$$

where  $Dz$  denotes the Gaussian measure,  $H(x) = \int_x^\infty Dz = \frac{1}{2} \operatorname{erfc}(\frac{x}{\sqrt{2}})$ , and the  $\hat{\cdot}$  Lagrange parameters enforce the definitions of the order parameters ( $\hat{u}$  enforces the normalization condition over the rows of the  $W$  matrix). The quantity  $\Xi$  is a function of the argument

$$U = \frac{2\hat{u}}{2\hat{s} - \hat{q}} , \quad (35)$$

and is defined as

$$\Xi(U) = \sum_{j,k \geq 2} H_j \left( U \mathbf{Id} + \mathbf{C} \right)_{jk}^{-1} H_k \quad \text{with} \quad H_j = 2\mathbf{C}_{1j} - \phi_0 , \quad (36)$$

and  $\mathbf{Id}$  is the identity matrix. In the above equation, the inverse is intended over the  $N - 1$ -dimensional restriction of the matrix  $U \mathbf{Id} + \mathbf{C}$  to entries  $j, k \geq 2$ .

## 2. Computation of $\Xi$

Expanding the terms in  $\Xi(U)$  in eqn. (36) above, we write  $\Xi(U) = \Xi_1(U) + \Xi_2(U) + \Xi_3(U)$  with

$$\Xi_1(U) = 4 \sum_{j,k \geq 2} \mathbf{c}_{1j} \left( U \mathbf{Id} + \mathbf{C} \right)_{jk}^{-1} \mathbf{c}_{1k} , \quad (37)$$

$$\Xi_2(U) = \phi_0^2 \sum_{j,k \geq 2} \left( U \mathbf{Id} + \mathbf{C} \right)_{jk}^{-1} , \quad (38)$$

$$\Xi_3(U) = -4\phi_0 \sum_{j,k \geq 2} \mathbf{c}_{1j} \left( U \mathbf{Id} + \mathbf{C} \right)_{jk}^{-1} . \quad (39)$$

Computation of  $\Xi_1$ : Consider the  $N \times N$  matrix  $\mathbf{C}^{(N)}$ , with entries  $\mathbf{C}_{ij}$  for  $i, j$  comprised between 1 and  $N$ . Let us also define  $\mathbf{Id}^{(N)}$  the identity matrix in dimension  $N$ , while  $\mathbf{Id}$  above referred to the identity matrix in dimension  $N - 1$ . Using block-matrix inversion formulas, we write that

$$\left( U \mathbf{Id}^{(N)} + \mathbf{C}^{(N)} \right)_{11}^{-1} = \frac{1}{U + \mathbf{c}_{11} - \sum_{j,k \geq 2} \mathbf{c}_{1j} \left( U \mathbf{Id} + \mathbf{C} \right)_{jk}^{-1} \mathbf{c}_{1k}} \quad (40)$$

The left hand side of the equation above is equal, in the large- $N$  limit, to the resolvent  $g(U)$  of  $\mathbf{C}$  defined in (32). Using  $\mathbf{c}_{11} = \Gamma(0) = \phi_0$  and the definition of  $\Xi_1(U)$ , we obtain

$$\Xi_1(U) = 4 \left( U + \phi_0 - \frac{1}{g(U)} \right) . \quad (41)$$

Computation of  $\Xi_2$ : Let  $|v_+\rangle$  be the normalized vector with  $N$  identical components,  $(v_+)_i = \frac{1}{\sqrt{N}}$ . We have

$$\Xi_2(U) = N \phi_0^2 \langle v_+ | \left( U \mathbf{Id} + \mathbf{C} \right)^{-1} | v_+ \rangle . \quad (42)$$

For large  $N$ ,  $|v_+\rangle$  is the top eigenvector of  $\mathbf{C}$ , with (extensive) eigenvalue  $\lambda_+ = N \int d\mathbf{r} \Gamma(\mathbf{r}) = N \phi_0^2$ . Hence,

$$\Xi_2(U) = N \phi_0^2 \times \frac{1}{U + N \phi_0^2} \rightarrow 1 , \quad (43)$$

in the large- $N$  limit (since  $U$  remains bounded, see below).

Computation of  $\Xi_3$ : As  $\mathbf{C}_{jk}$  with  $j, k \geq 2$  does not depend on the locations  $\mathbf{r}_1^\ell$  of the place fields associated to neuron  $i = 1$  in the different maps  $\ell$ , we may substitute  $\mathbf{C}_{1j}$  in eqn. (39) with its average over those positions, equal to  $\phi_0^2$ . We obtain

$$\Xi_3(U) = -4 \phi_0^3 \sum_{j,k \geq 2} \left( U \mathbf{Id} + \mathbf{C} \right)_{jk}^{-1} = -4 \phi_0 , \quad (44)$$

in the large- $N$  limit, see calculation of  $\Xi_2(U)$  above.

Expression of  $\Xi$ : Gathering the three terms above, we obtain

$$\Xi(U) = 1 + 4U - \frac{4}{g(U)} . \quad (45)$$

### 3. Expression of log. volume and saddle-point equations close to the critical line

As  $\alpha$  reaches its maximal value (at fixed  $\kappa$ ), the set of couplings satisfying the inequalities (1) shrink to a single solution, and we expect  $s, q$  to reach the same value according to (33). We therefore look for an asymptotic expression for  $\frac{1}{N^2} \langle \log Z \rangle$  in (34) when

$$\epsilon = s - q , \quad (46)$$

is very small and positive. In this regime, we expect the conjugated Lagrange parameters to diverge as inverse powers of  $\epsilon$ . More precisely, calling

$$\hat{\epsilon} = 2\hat{s} - \hat{q} , \quad (47)$$

we assume that

$$\hat{\epsilon} = \frac{V}{\epsilon} , \quad \hat{q} = \frac{T}{\epsilon^2} . \quad (48)$$

as  $\epsilon \rightarrow 0$ . To the leading order, we obtain

$$\frac{1}{N} \langle \log Z_1 \rangle = \frac{F(\alpha)}{2\epsilon} + O(|\log \epsilon|) , \quad (49)$$

where  $F(\alpha)$  is the extremum over  $m, q, U, V, T$  of

$$F(\alpha; m, q, U, V, T) = V \left( q + U - \frac{m^2}{\Xi(U)} \right) + T \left( 1 - \frac{1}{V} \int d\lambda \rho(\lambda) \frac{\lambda}{\lambda + U} \right) - \alpha p(q - m^2) \int_x^\infty \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} (z - x)^2 \quad (50)$$

with  $x = \frac{m - \kappa}{\sqrt{q - m^2}}$  and  $U$  defined in (35). This equation is equivalent to equation (7) of the main text. Note that, in order to obtain (50), the saddle point equation over  $\hat{m}$  in (34) was derived and solved for  $\hat{m}$  explicitly. Extremizing over  $U, T, V$ , we obtain

$$V = \int d\lambda \rho(\lambda) \frac{\lambda}{\lambda + U} , \quad (51)$$

$$T = - \left( q + U - \frac{m^2}{\Xi(U)} \right) \int d\lambda \rho(\lambda) \frac{\lambda}{\lambda + U} , \quad (52)$$

$$1 + \frac{m^2}{\Xi(U)^2} \frac{d\Xi}{dU} = \left( q + U - \frac{m^2}{\Xi(U)} \right) \int d\lambda \rho(\lambda) \frac{\lambda}{(\lambda + U)^2} . \quad (53)$$

Note that the derivative of  $\Xi$  with respect to  $U$  can be easily computed from the derivative of  $g$  with respect to  $U$  according to eqn (45). Following the implicit equation over  $g$  in equation (9) in the main text, we find

$$\frac{dg}{dU}(U) = \frac{1}{\sum_{\mathbf{k} \neq \mathbf{0}} \frac{\alpha \hat{\Gamma}(\mathbf{k})}{(\alpha + g \hat{\Gamma}(\mathbf{k}))^2} - \frac{1}{g^2}}. \quad (54)$$

We may now write the saddle-point equations over  $q$  and  $m$ , which give, after some elementary manipulation,

$$\alpha p H(x) = \int d\lambda \rho(\lambda) \frac{\lambda}{\lambda + U}, \quad (55)$$

$$\frac{m}{m - \kappa} \left( \frac{1}{\Xi(U)} - 1 \right) = \frac{1}{\sqrt{2\pi} x e^{x^2/2} H(x)} - 1. \quad (56)$$

The three coupled equations (53,55,56) allows one, in principle, to compute  $q, m, U$  and, and therefore  $T, V$  and  $F(\alpha)$ . In addition, the optimization of  $\langle \log Z \rangle$  in (49) over  $\epsilon$  immediately gives  $F(\alpha) = 0$ , hence, a fourth equation to determine the critical value of  $\alpha$  at fixed  $\kappa$ . This last equation read, after simplification according to eqn (56),

$$\frac{U}{\kappa} = m \left( \frac{1}{\Xi(U)} - 1 \right). \quad (57)$$

#### 4. Large- $p$ behavior of the critical capacity

We now focus on the maximal capacity, obtained when  $\kappa \rightarrow 0$ . According to (57),  $U$  vanishes, and equations (55,56) as a well as the implicit equation (9) of the main text on the resolvent  $g$  give a set of two coupled equations for  $x$  and the resolvent  $g$ :

$$\frac{1}{g} = \sum_{\mathbf{k} \neq \mathbf{0}} \frac{\hat{\Gamma}(\mathbf{k})}{1 + g p H(x) \hat{\Gamma}(\mathbf{k})}, \quad (58)$$

$$1 - \frac{4}{g} = x \sqrt{2\pi} H(x) e^{x^2/2}. \quad (59)$$

from which the capacity can be computed as a function of the number  $p$  of points,

$$\alpha_c(p) = \frac{1}{p H(x)}. \quad (60)$$

In practice, we can choose  $x$  at will, compute  $g$  from (59), then  $p$  from (58), and, finally,  $\alpha_c$  from (60).

Remark that equation (58) can be rewritten as

$$p H(x) = G\left(g p H(x)\right) \quad \text{with} \quad G(y) = \sum_{\mathbf{k} \neq \mathbf{0}} \frac{y \hat{\Gamma}(\mathbf{k})}{1 + y \hat{\Gamma}(\mathbf{k})}. \quad (61)$$

According to dimensional analysis, the large momentum scaling of the Fourier coefficients is given by

$$\hat{\Gamma}(\mathbf{k}) \sim \frac{\phi_0^2}{\left(k \phi_0^{\frac{1}{D}}\right)^{D+1}} = \frac{\phi_0^{1-\frac{1}{D}}}{k^{D+1}}, \quad (62)$$

where  $k = |\mathbf{k}|$  and  $D$  is the dimension of the physical space. We deduce that, for large arguments  $y$ ,

$$G(y) \sim A_1(D) \phi_0^{\frac{D-1}{D+1}} y^{\frac{D}{D+1}} \quad \text{with} \quad A_1(D) = \int \frac{d^D \mathbf{u}}{|\mathbf{u}|^{D+1} + 1}. \quad (63)$$

In addition, using the asymptotic expansion of the erfc function, we have

$$x \sqrt{2\pi} H(x) e^{x^2/2} \simeq 1 - \frac{1}{x^2} \quad (64)$$

for large  $x$ . Combining these expressions allows us to obtain the asymptotic relation between  $x$  and  $y$ ,

$$y^{\frac{1}{D+1}} = 4 A_1(D) \phi_0^{\frac{D-1}{D+1}} x^2 . \quad (65)$$

and, to the leading order in  $p$ ,

$$x \simeq \sqrt{2 \log p} - \left(D + \frac{1}{2}\right) \frac{\log \log p}{\sqrt{2 \log p}} . \quad (66)$$

We then deduce the asymptotic scaling of the critical capacity given by equation (10) of the main text, with

$$A(D) = \frac{1}{8^D A_1(D)^{D+1}} . \quad (67)$$

The scaling for  $x$  in eqn. (66) entails the following relation between the order parameters  $q$  and  $m$  in the large- $p$  regime,

$$\frac{q}{m^2} - 1 \sim \frac{1}{2 \log p} . \quad (68)$$

To interpret the consequences of the equation above, we consider a set of replicated couplings,  $\{W_{ia}\}$ . For any random position  $\mathbf{r}$  in map  $\ell$  defining the pattern  $\sigma$ , we define the rescaled and centered random variable

$$Y(\mathbf{r}|\{W_{ia}\}) = \frac{1}{m_\ell^a} \left( (2\sigma_i^\ell - 1) \sum_{j \geq 2} W_{ja} \sigma_j^\ell - m_\ell^a \right) . \quad (69)$$

By definition of the order parameter  $m$ , the average value of  $Y$  vanishes:

$$\langle Y(\mathbf{r}|\{W_{ia}\}) \rangle_{\mathbf{r}} = 0 . \quad (70)$$

Equation (68) implies that the variance of  $Y$  is

$$\langle Y(\mathbf{r}|\{W_{ia}\})^2 \rangle_{\mathbf{r}} \simeq \frac{1}{2 \log p} , \quad (71)$$

as  $p$  gets large and the load takes its maximal value (critical capacity). In other words, the standard deviation of  $Y$  scales as  $(\log p)^{-1/2}$  for large  $p$ . We thus expect that the  $k^{\text{th}}$  cumulant of  $Y$  will scale as  $(\log p)^{-k/2}$ . Under this assumption, the distribution of the stability  $t$  has mean value  $m$  and fluctuations of the order of  $\Delta t = m/\sqrt{\log p}$ . These fluctuations are negligible in the large- $p$  limit, since resolution of the saddle-point equation (53) shows that

$$m \simeq \frac{D}{4} - \frac{D^2}{256 (\log p)^3} + o\left(\frac{1}{(\log p)^3}\right) \quad (72)$$

at the critical point. Hence,  $\Delta t \sim (\log p)^{-1/2}$  is smaller and smaller as  $p$  increases, and the distribution of  $t$  is well approximated by a Gaussian in the large- $p$  limit. The Gaussian approximation obtained by discarding all powers of  $\hat{t}$  of order  $\geq 3$  in eqn (27) in our quenched PF theory is therefore expected to be exact in this limit.

## D. Comparison between Quenched PF theory and SVM

### 1. Values of couplings

Figure 8 compares how the couplings  $W_{ij}$  depend on the size of the network,  $N$ , and of the distances between the PF of neurons  $i, j$  in the maps. We generally find that the couplings  $W_{ij}$  obtained by SVM and the ‘thermal’ averages  $[W_{ij}]$  predicted by the quenched PF theory for a fixed set of PF centers are in excellent agreement, see equations (23) and (24) in [6] for details on the calculation of the average couplings and associated standard deviations.

- Both sets of couplings have mean values scaling as  $\frac{1}{N}$  and standard deviations scaling as  $\frac{1}{\sqrt{N}}$ .
- The sign of couplings depend on the distance between their PF centers in the maps. We get excitatory couplings for distances up to the radius of the PFs ( $r_c = \sqrt{\frac{\phi_0}{\pi}}$  in  $D = 2$ ), and inhibitory interactions for larger distances.

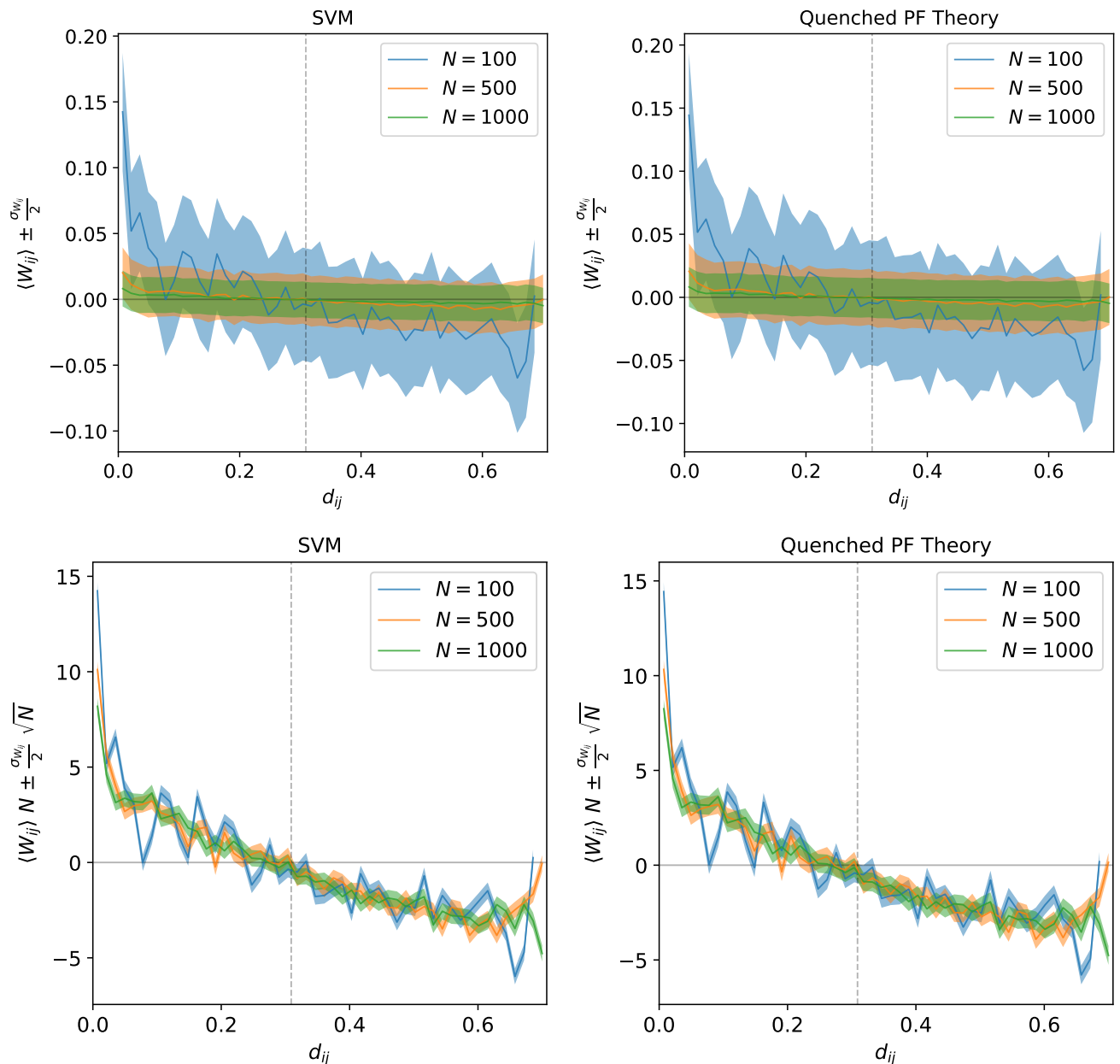


FIG. 8: Comparison of couplings obtained with SVM (left) and with the Quenched PF Theory (right). Top: Dependence of couplings on  $N$ . Bottom: Dependence of the couplings on distance; The vertical line locates the radius  $r_c$  of the PF. These results were obtained for  $D = 2$ ,  $\phi_0 = .3$ ,  $\alpha = .1$ ,  $p = 5$ ; we have averaged over 100 different realizations of the  $p$  positions at fixed PF centers for the SVM results. Space was divided in 50 bins with values ranging from 0 to  $\sqrt{2}/2$  (the maximal distance achievable in unit square with periodic boundary conditions). Couplings were then put in the corresponding bins for all maps, and the averages and standard deviations were plotted as functions of the bin centers. Average couplings and associated standard deviations with quenched PF theory were computed with (23) and (24) of [6], with the substitution  $\alpha_c \rightarrow p \alpha_c$  as the number of patterns is here  $p \times L$ .

## 2. Dependence on $\phi_0$

In Fig. 9 we show that the value of  $p$  such that the results obtained with the quenched PF theory and with SVMs match increases as  $\phi_0$  decrease.



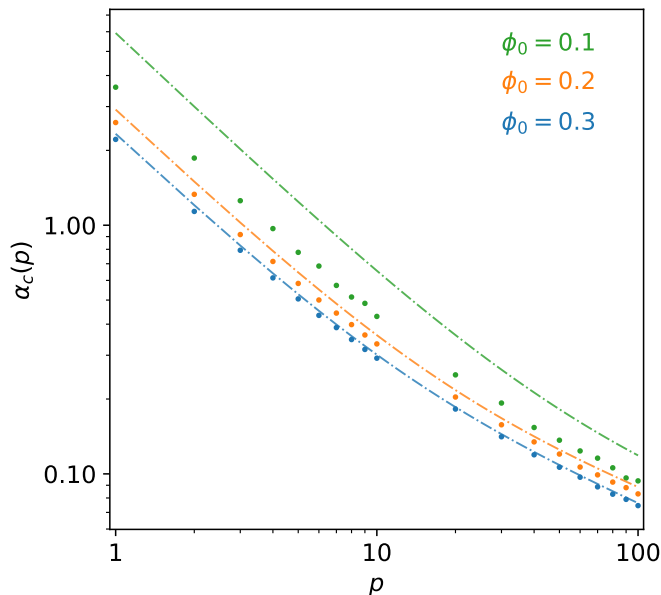


FIG. 9: Scaling cross-over of  $\alpha_c(p)$  vs.  $p$  for different values of  $\phi_0$ . Quenched PF Theory (dashed-dotted lines) gets closer to SVM (scatter plots) as  $p$  increase, the value of  $p$  for which Quenched PF Theory and SVM matches increase as  $\phi_0$  decrease. We use for this results  $D = 2$ ,  $N = 5000$ , and we have averaged over 50 different realization of the environments and different realizations of the  $p$  positions.

An analysis of equations (58,59), valid in the small  $\phi_0$  limit, indicate that this minimal value of  $p$  scales as

$$p_{match}(\phi_0) \sim \frac{e^{1/(8\phi_0)}}{\phi_0^{3/2}}, \quad (73)$$

and becomes very large as  $\phi_0$  is small. Realistic values for  $\phi_0$  are reported in the experimental literature to range between .2 and .3.

- 
- [1] Scholkopf, Bernhard, and Alexander J. Smola. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press, 2001.
  - [2] Diamond, Steven, and Stephen Boyd. "CVXPY: A Python-embedded modeling language for convex optimization." The Journal of Machine Learning Research 17.1 (2016): 2909-2913.
  - [3] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." Journal of machine learning research 12.Oct (2011): 2825-2830.
  - [4] Gardner, Elizabeth. "The space of interactions in neural network models." Journal of physics A: Mathematical and general 21.1 (1988): 257.
  - [5] Mézard, Marc, Giorgio Parisi, and Miguel Virasoro. Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications. Vol. 9. World Scientific Publishing Company, 1987.
  - [6] Monasson, Rémi. "Storage of spatially correlated patterns in autoassociative memories." Journal de Physique I 3.5 (1993): 1141-1152.
  - [7] A. Battista, R. Monasson. "On the spectrum of multi-space Euclidean random matrices". In preparation (2019)
  - [8] Livan, Giacomo, Marcel Novaes, and Pierpaolo Vivo. Introduction to random matrices: theory and practice. Springer International Publishing, 2018.