Robustness of Hebbian and Anti–Hebbian Learning

Tibor Fomin [1] and András Lőrincz [2 3]
Department of Photophysics
Institute of Isotopes
The Hungarian Academy of Sciences
P.O. Box 77
Budapest
Konkoly–Thege út 29–33
H–1525 Hungary

February 10, 1995

[1]lepi@iserv.iki.kfki.hu
[2]lorincz@iserv.iki.kfki.hu
[3]To whom correspodence should be addressed

Abstract

Self–organizing neural networks with Hebbian and anti–Hebbian learning rules were found robust against variations in the parameters of neurons of the network, such as neural activities, learning rates and noisy inputs. Robustness was evaluated from the point of view of properties of soft competition for input correlations. Two models were studied: a neural network with presynaptic Hebbian learning and a similar network that models growing connections. Both networks were trained on pixel–discretized 2 dimensional images. If extended local objects of the 2 dimensional world placed at random positions are inputted to the network then neurons develop spatial filters with inhibitory connections between the neurons that correspond to the neighboring relation, i.e., the network builds up the internal representation of the geometry of the external world. Both networks were found robust against input noise and network parameters. A lower estimate for the change in all of the parameters and pixel noise that do not deteriorate performance is $\pm 10\%$, making the system attractive for analog hardware implementation. It was found that small amount of input noise and small variations in the network parameters can further improve the geometry representation.

Keywords:

- Hebbian learning
- anti–Hebbian learning
- robustness
- fault tolerance
- self–organization

Contents

1 Introduction

Fault tolerance of artificial neural networks (ANN) has been studied mostly for passive systems that do not react in any special way to compensate for the effect of internal failures (Rovnyak, Stirk, and Athale [12], Protzel, Palumbo, and Arras [10]). The other less extensively studied category is systems with active fault–tolerance that can reorganize their resources to counteract the fault effects. Studied examples describe adaptation or retraining after internal faults (Sejnowski and Rosenberg [15]). Other examples suggest prewired self–repair mechanism (Petsche and Dickinson [9]).

Here the fault tolerance of a self–organizing network equipped with Hebbian and anti–Hebbian (HAH) learning is studied. By anti–Hebbian (a–H) learning we mean self–organizing Hebbian (H) learning for inhibitory connections. Excitatory and inhibitory connections are called H and a–H connections, respectively. A HAH network has H connections between layers (e.g., between inputs and a neural layer) and a–H connections between the neurons of a layer. The H connections and the a–H connections are called feedforward and lateral connections, respectively. A family of HAH networks have been developed by Rubner and Tavan [14], Rubner and Schulten [13], Földiák [2] and Marshall [8]. The network used here is very similar to the one that was introduced by Földiák [2].

In a former study (Fomin and Lőrincz [3]) we have shown that HAH networks are very promising, since

- they can be used as building blocks for neural network and neural network hierarchy fabrications, and
- they tend to form a redundant representation, which is an inherently robust feature.

Another advantage of HAH networks against some other soft competitive networks, e.g., the network of Szepesvári, Balázs and Lőrincz [16] is that there is no need for a winner–take–all (WTA) subsystem. WTA needs precise prewiring of inhibitory connections. Here the question of tolerance in network parameters is the subject of investigation.

In the case of self–organized learning the definition of 'performance' is not straightforward as the network always does something. To make the point clear let us categorize ANN's into three distinct groups from the point of view

of means evaluating performance. The first group contains the networks working with supervised learning, e.g. backpropagation (BP) nets. These networks are taught for input–output pairs, or associations and the error is provided by an external observer, who knows the task exactly. Reinforcement learning (RL) (Barto, Sutton and Brouwer [1]) does not perform supervised learning but learning with a 'critic'. From our particular point of view, however, is belongs to the same class: there exist an external agent who can provide a reinforcement signal. In each case performance can be evaluated from the behaviour of the external agent, who can measure the errors (BP), or the average of the reinforcement signal (RL).

The second group from the point of view of performance evaluation contains the adaptive bidirectional associative memory (ABAM) (Kosko [7]). In this case input–output pairs are provided to the network and the network tries to learn these associations via Hebbian learning. The performance can be measured through the given examples of input–output relations, namely, how close the learnt associations are to the aimed ones. In this case one has the option of measuring performance by the given input–output relations. We consider this group separate, since there is no direct connection between the measure of performance, i.e., the errors and network training.

The last group contains, among others, the competitive learning (Grossberg [5]), the Kohonen network (Kohonen [6]), the ART network with preset vigilance factor, and the HAH networks. These networks always do something and there is no direct external error that could measure performance: one may try to measure convergence properties for a given set of inputs, or, one may try to derive the learning rules as the minimization of some error function. Nevertheless, there will be no direct information if the network can perform a specific task in a given environment or not. In other words: in the case of the first and second groups one either knows the aimed input–output relations or has feedback from the external world whether the learnt relation is correct or incorrect. This porperty is missing for this last group of networks. The lack of this property may be traced back to the need of viewing ANN's as black (input–output) boxes. The question that we cannot answer for this last ANN group is if the time evolving input–output mapping is satisfactory or not.

Our starting point for the derivation of a measure of performance for the HAH network is that HAH networks perform soft competition and thus HAH neurons search and compete for high order correlations and 'divide'

the input world between themselves. Thus we can measure performance by monitoring how well this division performed, i.e., how different the neural receptive fields are. In order to make that measurement a special pixel discretized 2 dimensional world example will be studied. As it will be shown for this example neurons develop a 'quasi–orthogonal–representation' (in the form of local spatial filters for the case of extended local objects as inputs) of the world and network performance may be measured by the orthogonality of the neural filter vectors.

The inhibitory connections can provide another measure of performance as these networks have another interesting property: the less orthogonal the receptive fields of two neurons are, i.e., the larger the overlap between their receptive fields is, the stronger inhibitory connections develop between them. The network thus develops strong inhibitory connections for neighbouring spatial filters and negligible inhibitory connections otherwise. In effect, the inhibitory connections develop an internal representation of the geometry of the external world. We can ask the question: how good that geometry representation is and we can try to measure the performance by the quality of the geometry representation.

It may be worth to emphasize that our aim is more than investigating the graceful degradation of a prewired ANN. We intend to study the degradation of HAH networks in the presence of working learning rules. In other words the goal is to study the graceful degradation of unsupervised learning for HAH networks.

2 Hebbian and Anti–Hebbian Learning

We have performed our computations on a HAH network consisting of 24 neural output units (Fig. 1). In a HAH network each neuron receives inputs from an input vector $\mathbf{x}$ via excitatory input weights. Let us denote the weight connecting the ith neuron with the jth component of the input vector by q_{ij} $(i = 1, \ldots, 24;\ j = 1, \ldots, 400)$. The input vector x_j $(j = 1, \ldots, 400)$ represents a 20 by 20 pixel sized grid. The neurons are also connected with each other through the lateral inhibitory connections w_{ik} $(i, k = 1, \ldots, 24)$. Two models shall be studied here differing in the network relaxation equation: when an input is connected to the system the continuous output y_i of neural unit i develops through a relaxation process governed by one of the following non–linear differential equations:

$$\dot{y}_i = -y_i + f\left(\sum_j q_{ij} x_j - \sum_{\substack{k \\ i \neq k}} w_{ik} y_k - \theta_i\right) \tag{1}$$

$$\dot{y}_i = -y_i + f\left(\sum_j S_1\left(q_{ij}\right) x_j - \sum_{\substack{k \\ i \neq k}} S_2\left(w_{ik}\right) y_k - \theta_i\right) \tag{2}$$

In Eqs. (1) and (2) θ_i is a time–dependent offset term. There are two separate timescales in the network evolution. The first one is described by the decay time of Eqs. (1) and (2) where sustained inputs are assumed. The other timescale belongs to the tuning of the network parameters, such as θ_i. The nonlinearity of the neural units is represented by function f; in the computations we used a sigmoid function:

$$f(z) = \frac{1}{1 + e^{-\lambda z}} \tag{3}$$

Functions S_1 and S_2 denote similar sigmoid functions

$$S_l(z) = \frac{\delta_l}{1 + e^{-\lambda_l (z - \Theta_l)}} \qquad (l = 1, 2) \tag{4}$$

with multipliers δ_1 and δ_2, decay constants λ_1^{-1} and λ_2^{-1}, and thresholds Θ_1 and Θ_2, respectively (see Table I). The role of these functions is described below. The feedforward weights are initially random, selected from a uniform distribution on [0,1]. The lateral feedback weights and the offset terms were initially set to zero.

Connection weights are modified according to the following differential equations:

$$\varepsilon_q^{(n)} \dot{q}_{ij} = -q_{ij} + \alpha y_i x_j \qquad (n = 1, 2) \tag{5}$$

$$\varepsilon_w \dot{w}_{ik} = -w_{ik} + \beta y_i y_k \tag{6}$$

where dot denotes the time derivative, α and β limit the values of the connection strengths, ε^{-1} determines the learning rate as compared to the settling time of the network given by Eqs. (1) or (2). Indices of ε refer to the connection (lower indices) and to the equation (upper indices). The values of ε used set the 'relaxation time' of network parameters to about 10,000 iterations. In every example 150,000 iterations were computed (Table I).

Equation (1) is a HAH relaxation equation with connections that can become stronger via correlating activations at the end connected points, or fade if only one or none of the connection endpoints is active. Equation (2) describes a network with the modified (transformed) effect of excitatory or inhibitory connections. The sigmoid form of transformation in Eq. (2) was chosen as a model of an initially not connected network with growing connections: in the modelled system if two points, input j and neuron i are active simultaneously, then connection q_{ij} 'grows in length' towards neuron i. If at least one of the endpoints is inactive then the connection 'shortens'. The connection is established – becomes effective in the argument of function f – when the growing link reaches neuron i. This process is modelled by using a sharp sigmoid function S_1 and a threshold value for the sigmoid function S_1: the effect of a connection in Eq. (2) is small as long as the magnitude of the corresponding weight is small but can become strong very quickly. The magnitude of connection strength q_{ij} can further increase and can thus stabilize the connection, that, prevents quick disconnection. The effect of the connection on neural activity, however, quickly saturates. For all of the

connections the same threshold value was used modelling a situation when connection endpoints are initially at about the same distance from the respective neural units. Function S_2 in Eq. (2) models the growth of lateral inhibitory connections in the same way. Parameters for the growth of inhibitory connections are, however, such that the network is limited to work in the supralinear regime and not in the saturation regime of S_2. This point will be discussed later. The networks described by Eqs. (1) and (2) together with threshold learning rule (see below) and the connection learning rules shall be called the traditional model (TM) and the growth model (GM), respectively. Data on network parameters are given in Table I. To make GM a biologically plausible model spatial constraints should be introduced. This feature is attractive as we shall see that the number of effective connections is low.

The equation for the offset term is given as:

$$\dot{\theta}_i = \gamma(y_i - p) \tag{7}$$

Equation (7) increases the threshold if the output of a neuron is larger than p, the so called bit probability and decreases it if the neuron output is below that number, i.e. guides high activity neurons to become more selective and helps low activity neurons to seek for inputs they can respond to.

3 Spatial Filter Formation

The networks were inputted by input vectors of consecutively arranged lines of a pixel digitized 2 dimensional external world of the size of 20 by 20 pixels; thus the ith element of the input vector corresponds to point

$$\underline{r}_i(x,y) = \left(\mathrm{mod}_{20} i, \left\lfloor \frac{i}{20} \right\rfloor \right) \tag{8}$$

in the image, where mod and $\lfloor \rfloor$ denote modulo and floor operations, respectively. Pixel intensities could range between zero and one. The input vectors represented randomly positioned quasi–circular objects. One example is shown in Fig. 2. In the figure components of the input vectors were rearranged into the 20 by 20 form to guide the eye. The figure shows an 8 by 8 part of the image.

Bit probability (p in Eq. (7)) was set to 1/16, i.e. it was set somewhat above the inverse of the number of neurons. That favours, through Eq. (6), one or sometimes two neurons to fire (become high) at a time. The network produced circular local filters in a similar way to the competitive net of Szepesvári, Balázs and Lőrincz [16]. The left hand side of Fig. 3 shows the receptive field of one of the neurons for the TM network. The filter size corresponds to the average object size. It may be worth noting that the local filters represent the simplest correlations of the different input patterns that the competing neurons can detect. That is, the formation of the local filters is the result of the search for correlations and the competition for different correlations. In a former study (Fomin and Lőrincz [3]) it was shown, that

- the result is the same if input objects of different shapes were used provided that the bit probability p is left unchanged, and
- for higher bit probability values more than one neuron can fire simultaneously allowing the neurons to pick up specific correlations from the input samples and in that case spatial filters loose their local circular character.

One possible source of error that may be studied at this point is the deviation of the saturation value of function f. Equations (1) and (2) are changed to

$$f_i(z) = \frac{\sigma_i}{1 + e^{-\lambda z}} \tag{9}$$

with σ_i (saturation activity) varying from neuron to neuron. Figure 4 shows the 24 filters for the TM network with the σ_i values chosen randomly with the help of uniform distribution from the range of [1.0–1.0], [0.8–1.2], and [0.2–1.8], for the top, middle and bottom sets, respectively. The actual value of each neural saturation activity for the bottom set of receptive fields are given in Table II. For every set the run started with the same initial random excitatory weights. At every training step the same circular pattern was presented to the network at different positions. The position vector of the sample as generated randomly and the same sample order was kept during the courses of learning. Thus the neuron receptive fields can be compared for the different runs. As it may be seen from the figure a $\pm 20\%$ uniform distribution of neural activity saturation values gives rise to very little change in the neuron receptive fields. The $\pm 80\%$ uniform distribution of the saturation activity (that is, for values between 0.2 and 1.8) gives a more pronounced effect. Still, receptive fields are similar for neurons of activity values above 0.8.

Receptive fields for the GM network developed in a similar fashion. The receptive field of a typical neuron is shown in the right hand side of Fig. 3. The change of neural receptive fields is somewhat different for the GM network (Fig. 5). In the case of the $\pm 20\%$ distribution one neuron had a small receptive field, others show small deviations. In the $\pm 80\%$ distribution case the receptive fields of the weak neurons disappear under the gray scale resolution of the figure. Comparing GM receptive fields to TM receptive fields the former ones have sharper boundaries.

4 Geometry Representations

The HAH network works by searching and competing for correlations. The inhibitory connections responsible for competition can reveal properties of the correlations. Neurons of very different receptive fields do not compete strongly with each other, while neurons with somewhat similar receptive fields exhibit strong inhibitory connections. That is, strong anti–Hebbian connections develop between neurons of overlapping receptive fields. This way inhibitory connections reflect neighbouring relations, just like in the work of Szepesvári, Balázs and Lőrincz [16] where a winner–take–all network equipped with Hebbian learning was used in order to develop interneuron connections.

Geometry connections for the TM network are shown in Fig. 6. Runs were made with neural saturation activity values chosen from an uniform distributions over [1.0–1.0], [0.8–1.2], [0.55–1.45] and [0.2–1.8], for the upper right, upper left, lower right and lower left figures, respectively. The last three examples are tests of robustness of network performance, now on the geometry connections. Figures were made in the following fashion: first, excitatory connections of connection strengths below the 0.1 value of the maximal connection strength of all of the neurons were set to zero:

$$\hat{q}_{ij} = \max(0, q_{ij} - 0.1 \cdot q_{\max}) \tag{10}$$

where $q_{\max} = \max_{ij} q_{ij}$ with i and j being the neuron and pixel indices, respectively. Then position vectors of receptive field centers $C_i(C_x^{(i)}, C_y^{(i)})$ were computed:

$$C_x^{(i)} = \sum_j \hat{q}_{ij} \bmod {}_{20} j \tag{11}$$

$$C_y^{(i)} = \sum_j \hat{q}_{ij} \left\lfloor \frac{j}{20} \right\rfloor \tag{12}$$

Radii of receptive fields were determined as:

$$r_i = \sqrt{\frac{\sum_j \hat{q}_{ij} \left(\left(\mathrm{mod}_{20} j - C_x^{(i)} \right)^2 + \left(\left\lfloor \frac{j}{20} \right\rfloor - C_y^{(i)} \right)^2 \right)}{\sum_j \hat{q}_{ij}}} \tag{13}$$

(Expression (13) gives the correct radius for a narrow ring and underestimates the $1/e$ width of a Gaussian function by about 11%.) Then circles in Fig. 6 are drawn around centers C_i with radii r_i. The connection between the centers of the circles represent the strength of the inhibitory connections. These connections were thresholded in a similar fashion to the excitatory connections. The linewidth is porportional to the thresholded connection strengths. One expects a grid representation of the 2 dimensional pixel discretized world with equal connections strength between the neurons separated by the same distance. As it may be seen from the figure the network tries to build a close packed configuration, but the geometry representation is somewhat distorted even for the case of identical saturation values (upper left figure). The network shows reasonable (i.e., similar) performance for saturation activity values chosen from a uniform distribution over the range of [0.8–1.2]. Larger variations led to worse results. The resistivity of the network against the dynamic range of neural saturation value is, however, high. The sensitivity of performance of the two network models against network parameters shall be tested below.

5 Further Results of Network Tests

The following network properties were monitored for the different tests:

1. Neurons were ordered according to the magnitude of the radius of their receptive fields and the radii were plotted to monitor how narrow the distribution is. That property reflects the quality of discretization.

2. The a–H weights were ordered according to their strength and the a–H weights were plotted in this order to demonstrate that only a small number of connections plays a role.

3. The overlap between pairs of receptive fields was plotted vs. the a–H weights between the corresponding neurons to monitor the correlation between these quantities. The overlap between the ith and the jth neuron was defined as:

$$O(i,j) = \sum_k q_{ik} q_{jk} \tag{14}$$

4. The distance between the centers of the receptive fields was plotted as a function of the a–H weight strengths to get an insight into the representation of the geometry.

The receptive fields and the geometrical connections may be drawn in several ways for the GM network. Two version are shown in Fig. 7 that differ in connection strengths. In the left hand side of Fig. 7 connection strengths between neurons i and k are taken according to their effect in the neural relaxation equation (Eq. (2)) and the strength drawn correspond to $S_2(w_{ik})$. In the right hand side of Fig. 7 connection strengths are taken according to the actual magnitude of w_{ik}. The distinction is meaningful, the network was developed to improve the representation of geometry being important for self–organizing unsupervised neurocontrol (Fomin, Szepesvári and Lőrincz [4]). The geometry representation looks more uniform in the right hand side of Fig. 7. This is the result that the network is used in the supralinear region of the sigmoid function and that sharpens the range of connection strengths. In the figures below the value of w_{ik} will be used.

Test results for the TM and GM networks with identical neural saturation values are shown in Figs. 8 and 9, respectively. According to the figures the discretization of the GM model is superior to the TM model: the range of receptive field sizes spans a narrower region. It is, however, almost certain, that the best network parameters were not found. It is reasonable to say that TM and GM networks exhibit similar performance. The ordered a–H weight graph is close to zero for up to 75% of the whole range (upper right graphs) indicating that a small portion of inhibitory connection is used by the networks. This effect would be more pronounced for larger nets since it is a consequence of the need for inhibition only between neighbours, i.e., the number of effective inhibitory connections is assumed to scale with the number of neurons in a linear fashion. The last 25% of the upper right graphs differ: the TM model inhibits steeper curve indicating a wider distribution of inhibitory connections. This property is underlined by the lower left figures: strong a–H weights are clustered into a smaller region for the GM network. That means a better geometry representation for the GM network as the dependent variable of the figure is the distance between centers of receptive fields. The distance between receptive fields and the inhibitory connection strengths are strongly clustered. One may try to construct the triangles of the geometry of Fig. 6. Another important property is the overlap between receptive fields vs. the inhibitory connection strength between the same receptive fields. Now the TM network produces a linear relationship in clear contrast with the nonlinear "curve" produced by the GM model. The distribution is, again, narrower for the GM model.

Tests were run for

- noisy inputs and
- for distributions of learning parameters α, β, and γ.

The noise was additive from a uniform distributions between positive and negative values. The inputs were thresholded to positive values. It was found that small noise first improved performance of TM networks, and up to 10% noise (uniform distribution of $\pm 10\%$ of additive noise of the maximal pixel value) performance was not altered. At higher values ($\pm 20\%$) performance decreased considerably. Similar performance was found for the GM network with a stronger drop in performance at noise values higher than $\pm 10\%$.

In the tests of the learning parameters it was found that both networks were fairly insensitive to the threshold learning rate distribution. Values chosen randomly from a uniform distribution $\pm 80\%$ around the average still gave reasonable results. The connection learning rates are more sensitive points of both networks. In general it may be stated that the GM network is less resistive to these parameters. Four examples are given at around the edge of sharp deterioration of performance in Fig. 10. The figure plots a few examples for different distributions of a–H learning rates. Numbers are as before, e.g. the number of $\pm 10\%$ means that the learning rate values were chosen from a uniform distribution of $\pm 10\%$ around the average value. The upper figures show the geometry representation for the TM network. Distribution values are $\pm 45\%$ (left hand side) and $\pm 80\%$ (right hand side). The lower figures show geometry representation for $\pm 20\%$ (left hand side) and $\pm 45\%$ (right hand side) for the GM network. Performance changes in a similar fashion for the H connections. It is again the GM model that shows higher sensitivity to the width of the distribution of the H learning rate.

6 Conclusions

As it was discussed in the introduction it is not an easy task to define the term 'performance' for self–organizing competitive networks, since self–organizing networks always do something. Network performance can be evaluated from a specific porperty of these networks, that neurons are competing for correlations and force the others to find different correlations. It was concluded that HAH networks should provide a representation of the geometry if inputted by pixel vectors of a pixel discretized 2 dimensional world. Networks were discussed from this point of view: how good the representation of geometry is. Both HAH networks, the traditional model (TM) network and the growth model (GM) network were found robust against network parameters. The range, where network parameters can be changed without significant deterioration of performance exceed $\pm 5\%$ for all parameters. In many cases, e.g. for the threshold learning rates, the range could be much larger. Small noise content improved performance for both networks. Both TM and GM networks could tolerate a large range of input noise. TM network was more resistive to learning parameters than GM network. It was found that small noise superimposed on input samples improved performance, indicating that these networks with constant learning rates may land in local minima. The introduction of time dependent learning rates – just like in the case of the Kohonen network – may improve performance.

GM network gave similar geometry representation to TM network. This is an important point and could lead to an attractive family of networks for image processing. In particular it may be worth noting that the GM network models growing connections. The number of efficient excitatory and inhibitory connections was found to be much smaller than the square of the number of neurons (for inhibitory connections) or the product of the number of inputs and the number of neurons (excitatory connections). In fact, one may expect that the number of connections scales in a linear fashion with the number of neurons (for the inhibitory connections) and the number of excitatory connection per neurons does not grow with growing neuron number for the examples treated in this work. This experience leads to the question if there is a need for all the connection possibilities for these networks. It is a challenging question, since the network developed with the help of Hebbian learning is similar to a cellular neural network (CNN) (Roska and Chua [11]). To make the question sharper we ask if CNN can tolerate similar

variations in network properties. The self–organizing nature of the HAH network makes us to think that similar resistivity against variations of network parameters for CNN networks cannot be expected. Then the question is if CNN models could be replaced by locally connected HAH networks. Looking at the results one may conclude that it may be possible by building spatial constraints into the GM model, i.e., assuming a small number of connections around each neurons that can grow towards the active neurons. The development of HAH networks along this route may lead to an alternative modelling of visual systems.

The conclusion we should like to draw is the following: from the point of view of analog hardware implementation (electronic or optical) HAH nets are very promising: HAH nets are noise and fault tolerant.

References

[1] A.G. Barto, R.S. Sutton, and P.S. Brouwer. Associative search network: A reinforcement learning associative memory. *Biological Cybernetics*, (40):201–211, 1981.

[2] P. Földiák. Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics*, (64):165–170, 1990.

[3] T. Fomin and A. Lőrincz. Position determination and feature extraction with Hebbian and anti-Hebbian networks. Submitted to Neural Networks, April 1993.

[4] T. Fomin, Cs. Szepesvári, and A. Lőrincz. Self-organizing neurocontrol. In *Proceedings of IEEE International Conference on Neural Networks. IEEE World Coongress on Computational Intelligence*, pages 2777–2780, Orlando, Florida, 1994. IEEE Publication.

[5] S.A. Grossberg. Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, (11):23–63, 1987.

[6] T. Kohonen. *Self-Organization and Associative Memory*. Springer Verlag, Berlin, 1984.

[7] B. Kosko. Adaptive bidirectional associative memories. *Applied Optics*, (26):4947–4960, 1987.

[8] J. Marshall. A self-organizing scale-sensitive neural network. In *Proceedings of the International Joint Conference on Neural Networks*, pages 649–654, San Diego, California, 1990. IEEE Publication.

[9] T. Petsche and B.W. Dickinson. Trellis, codes, receptive fields, and fault-tolerant self-organizing neural networks. *IEEE Transactions on Neural Networks*, (2):154–166, 1990.

[10] P.W. Protzel, D.L. Palumbo, and M.K. Arras. Performance and fault-tolerance of neural networks for optimization. *IEEE Transactions on Neural Networks*, (4):600–614, 1993.

[11] T. Roska and L.O. Chua. The CNN universal machine: An analogic array computer. *IEEE Transactions on Circuits and Systems*, (40):163–173, 1993.

[12] S.M. Rovnyak, C.W. Stirk, and R.A. Athale. Optical implementation of winner take all networks: Noise considerations. *Proceedings of SPIE*, (882):137–142, 1988.

[13] J. Rubner and K. Schulten. Development of feature detectors by self-organization. *Biological Cybernetics*, (62):193–199, 1990.

[14] J. Rubner and P. Tavan. A self-organizing network for principal component analysis. *Europhysics Letters*, (10):693–698, 1989.

[15] T.J. Sejnowski and C.R. Rosenberg. JHU. Technical Report EECS–86/01, John Hopkins University, 1986.

[16] Cs. Szepesvári, L. Balázs, and A. Lőrincz. Topology learning solved by extended objects: a neural network model. *Neural Computation*, (6):439–456, 1994.

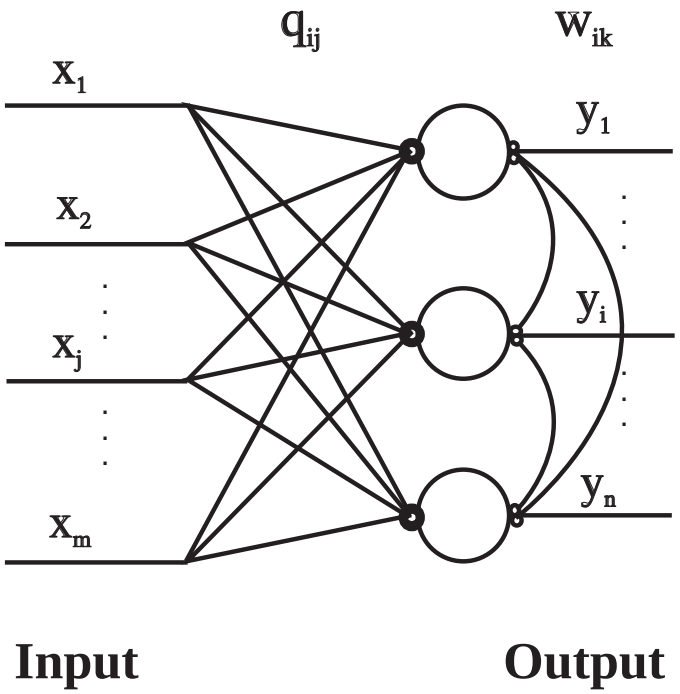

Figure 1: Hebbian and anti–Hebbian network. The inputs are connected to the neurons through the q_{ij} excitatory Hebbian connections. Neurons inhibit each other through the w_{ik} inhibitory anti–Hebbian connections. Excitatory and inhibitory connections are denoted by full and open circles, respectively.

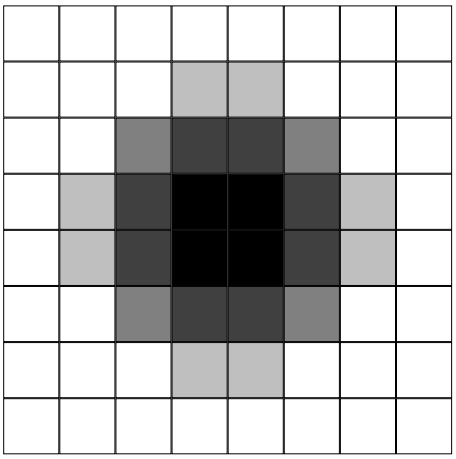

Figure 2: Example of input vectors. Pixel intensities could range between zero and one. Components of the input vector are rearranged into the 2 dimensional arrangement to guide the eye. 8 by 8 portion of the image is shown. Gray scale ranges from white to black are: [0.0–0.2], [0.2–0.4], [0.4–0.6], [0.6–0.8] and [0.8–1.0].

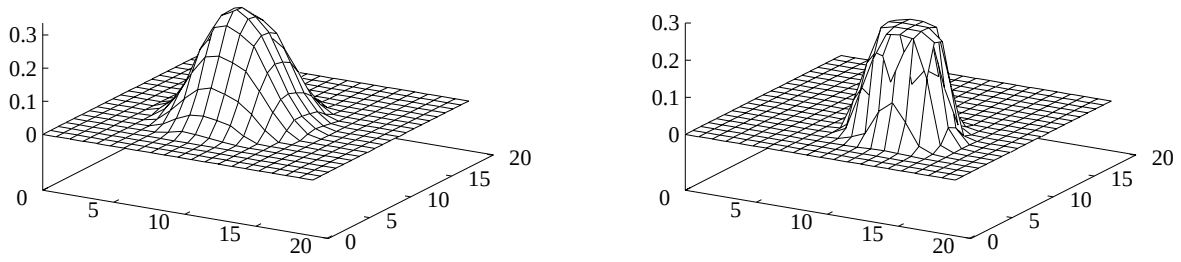

Figure 3: Typical receptive fields for the traditional (left hand side) and the growth model (right hand side) networks. The height of a pixel on the graph is proportional to the strength of the excitatory Hebbian connection for the TM network and to its sigmoid transformed value for the GM network.

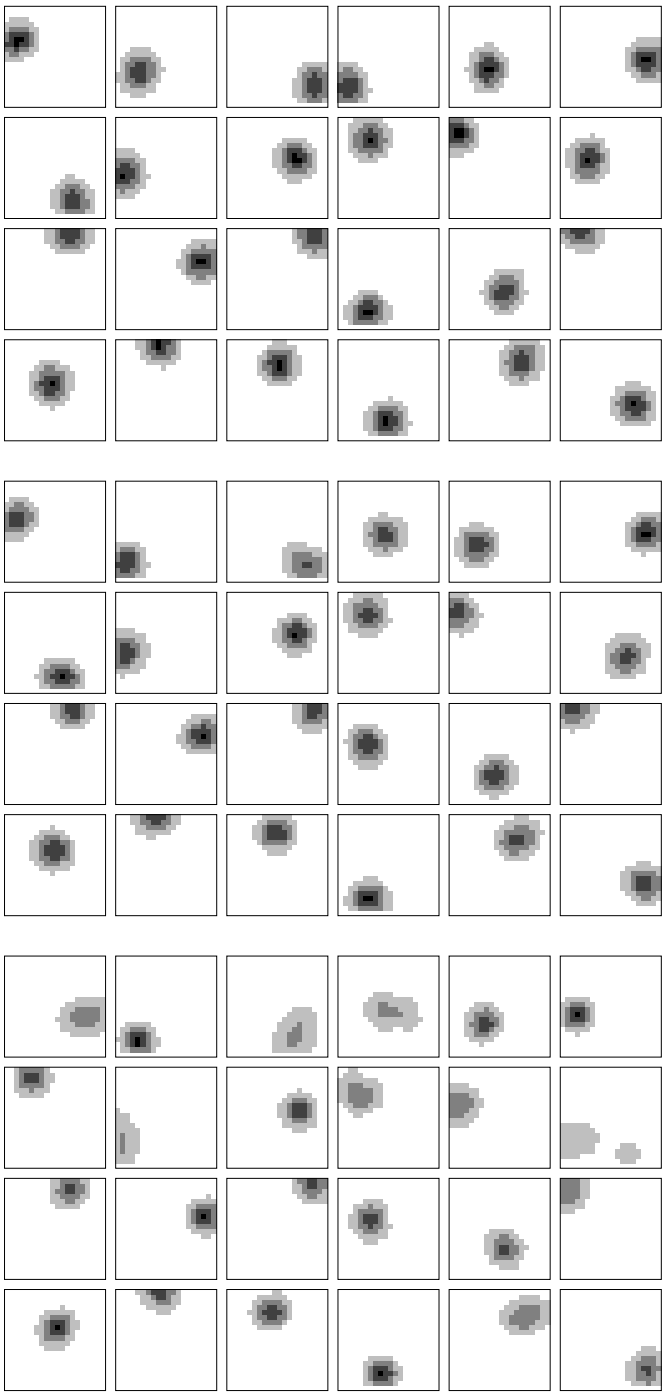

Figure 4: Traditional model (TM) receptive fields for distributions of neural saturation activities. Saturation activities were chosen with the help of uniform distributions in the range of [1.0–1.0], [0.8–1.2], and [0.2–1.8], for the top, middle and bottom sets, respectively. The figures are plotted with the help a five grade gray scale. White to black regions cover the [0.0–0.05], [0.05–0.3], [0.3–0.6], [0.6–0.9] and [0.9–1.0] portions of the maximal excitatory connection value.

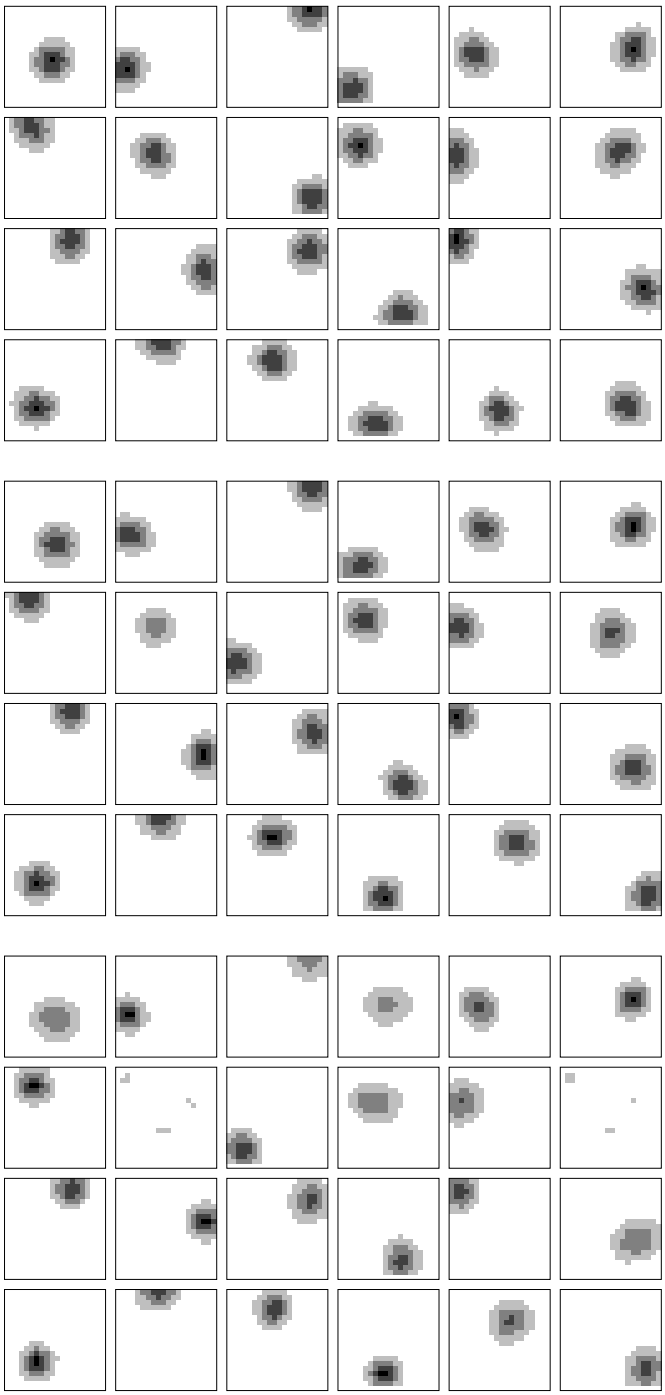

Figure 5: Growth model (GM) receptive fields for distributions of neural saturation activities. Saturation activities were chosen with the help of uniform distributions in the range of [1.0–1.0], [0.8–1.2], and [0.2–1.8], for the top, middle and bottom sets, respectively. The figures are plotted with the help a five grade gray scale. White to black regions cover the [0.0–0.05], [0.05–0.3], [0.3–0.6], [0.6–0.9] and [0.9–1.0] portions of the maximal excitatory connection value.

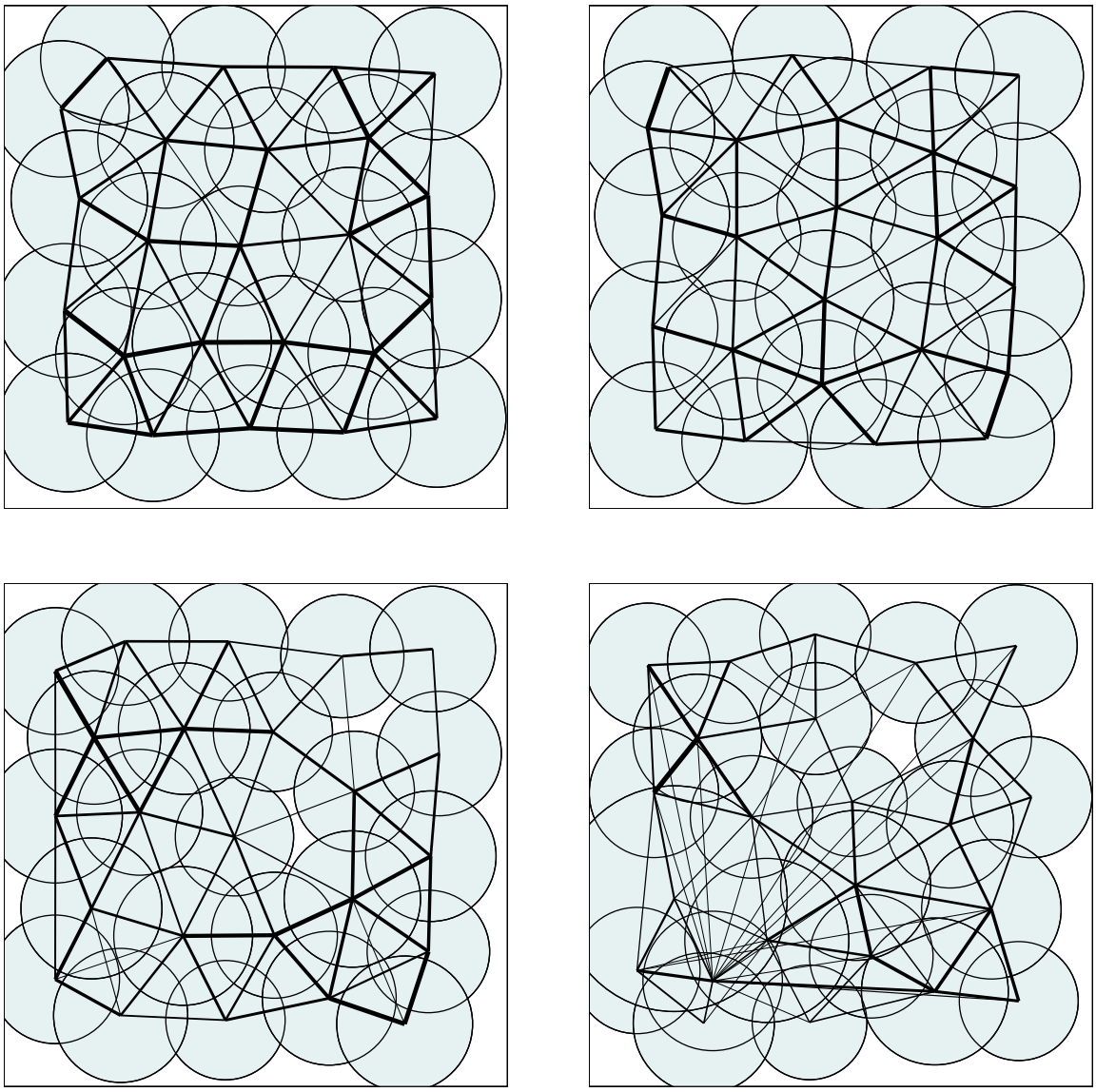

Figure 6: Geometry representation with the TM network. Circles represent the receptive fields. Thickness of connections between the circles shows the thresholded inhibitory connection strengths. Saturation activities were chosen with the help of uniform distributions from the range of [1.0–1.0], [0.8–1.2], [0.55–1.45] and [0.2–1.8], for the upper left, upper right, lower left and lower right figures, respectively.

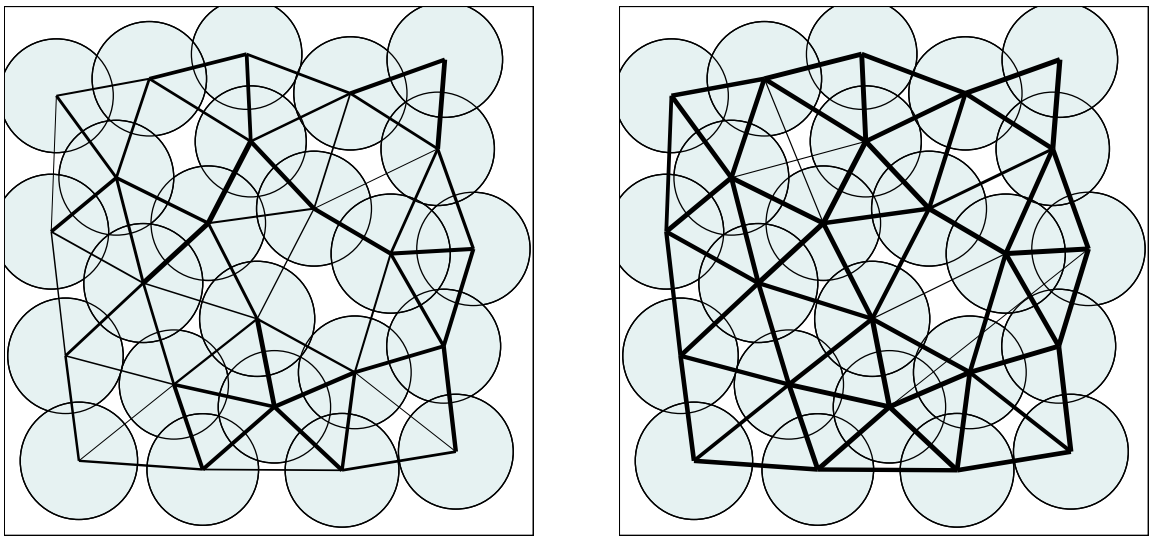

Figure 7: Geometry representation with the GM network. Connection strengths on the left hand side and right hand side figures correspond to sigmoid transformed connection strengths and to the connection strengths without the transformation, respectively.

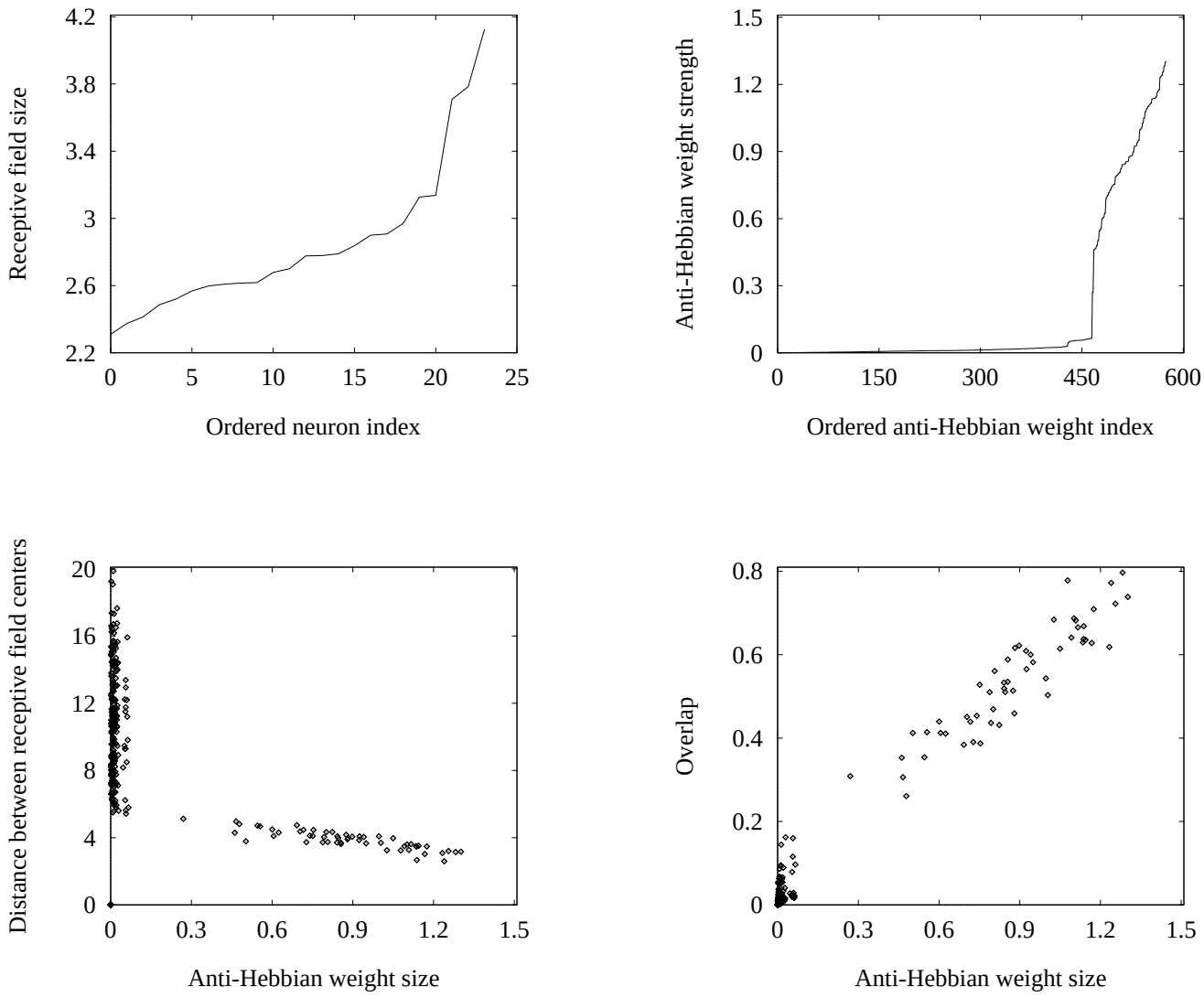

Figure 8: Properties of the traditional model: in the upper subfigures both the receptive fields and the anti–Hebbian weights are ordered according to their magnitudes and their values are plotted in this order. The lower figures show the correlations between a–H weights and receptive field overlaps (left hand side) and between a–H weights and the distance between receptive fields (right hand side).

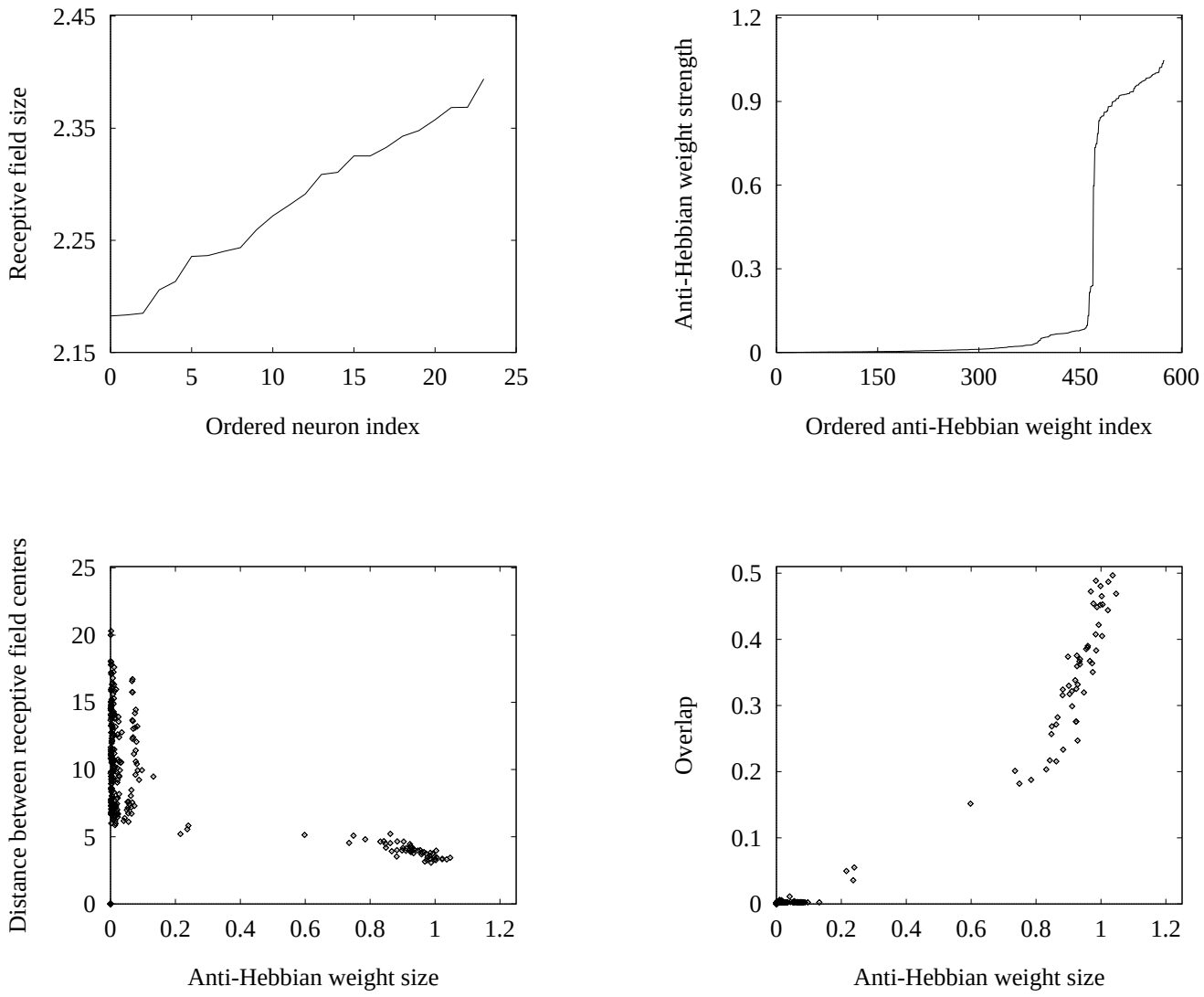

Figure 9: Properties of the growth model: in the upper subfigures both the receptive fields and the anti–Hebbian weights are ordered according to their magnitudes and their values are plotted in this order. The lower figures show the correlations between a–H weights and receptive field overlaps (left hand side) and between a–H weights and the distance between receptive fields (right hand side).

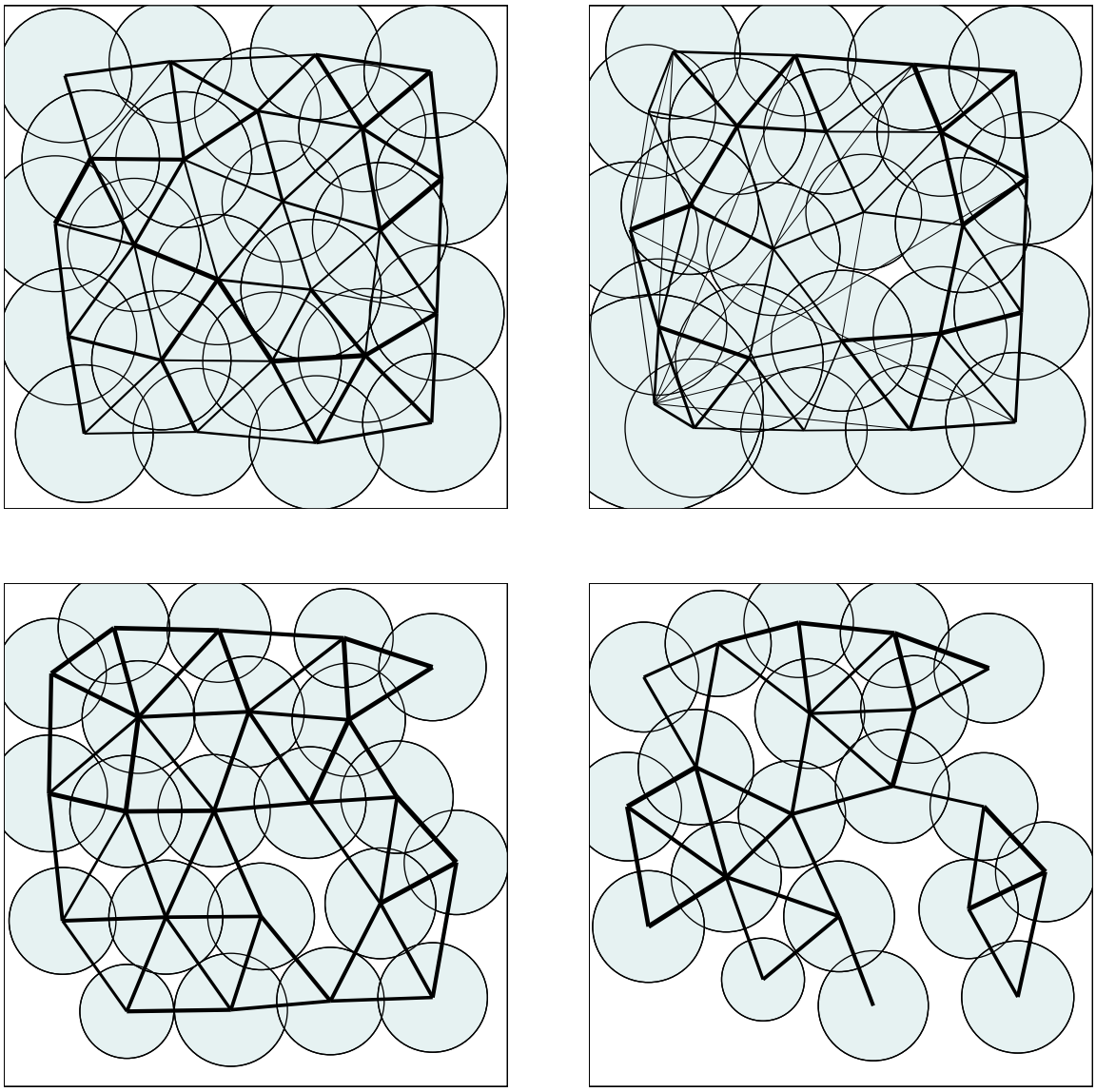

Figure 10: Sensitivity to the distribution of a–H learning rates. The upper figures show the geometry representation for the TM network. Distribution values are $\pm 45\%$ (left hand side) and $\pm 80\%$ (right hand side). The lower figures show geometry representation for $\pm 20\%$ (left hand side) and $\pm 45\%$ (right hand side) for the GM network. ($\pm A\%$ means that the learning rate values were chosen from a uniform distribution of $\pm A\%$ around the average value).

Number of neurons	24
Dimension of the input vector	400
Bit probality (p)	0.0625
α_1	10.0
α_2	10.0
β_1	100.0
β_2	100.0
γ	100.0
$\varepsilon_q^{(1)}$	10^4
$\varepsilon_q^{(2)}$	$2.5 \cdot 10^3$
ε_w	$2.5 \cdot 10^3$
δ_1	0.3
δ_2	2.0
λ_1	50.0
λ_2	10.0
λ	10.0
Θ_1	0.15
Θ_2	1.0

Table 1: Parameters of the networks

0.66	1.74	0.61	0.69	1.13	1.62
1.71	0.58	1.32	0.77	0.80	0.51
1.34	1.79	1.09	1.19	1.17	0.75
1.60	1.17	1.41	1.77	0.93	1.04

Table 2: Neural activity saturation values choosen randomly from the uniform distribution between 0.2 and 1.8. Saturation values are arranged in accord with Figure 4.