Homeostatic Synaptic Scaling in Self-Organizing Maps

Thomas J. Sullivan and Virginia R. de Sa
University of California, San Diego
La Jolla, CA, USA
tom@sullivan.to, desa@cogsci.ucsd.edu

Abstract - *Various forms of the self-organizing map (SOM) have been proposed as models of cortical development [11, 2, 7]. Typically, these models use weight normalization to contain the weight growth associated with Hebbian learning. A more plausible mechanism for controlling the Hebbian process has recently emerged. Turrigiano and others [14] have shown that neurons in the cortex actively maintain an average firing rate by scaling their incoming weights. In this work, it is shown that this type of homeostatic synaptic scaling can replace the common stop-gap measure of standard weight normalization. Organized maps still form and the output neurons are able to maintain an unsaturated firing rate, even in the face of large-scale cell proliferation or die-off. In addition, it is shown that in some cases synaptic scaling leads to a better representation of the input probability distribution.*

Key words - self-organizing map, homeostasis, weight normalization

1 Introduction

The self-organizing map (SOM), in its various forms, has been a useful model of cortical development [11, 2, 7, 10]. Sirosh and Miikkulainen [11] showed the simultaneous development of receptive field properties and lateral interactions in a realistic model. The usefulness of the developed lateral connections was shown by Choe and Miikkulainen [2] for contour integration and segmentation. It is this lateral connectivity that ensures the neighboring neurons come to respond to similar stimuli and form a good map.

In these models, Hebbian learning is used to strengthen associations between stimuli and winning neurons. This type of associative learning has been well-documented in the experimental literature [1, 6], but our understanding has remained incomplete. It is well known that the most straight-forward implementations of Hebbian learning lead to unconstrained weight growth. To counteract this problem, typical SOM algorithms use weight normalization: after each learning iteration all the weights are divided by the sum of each neuron's incoming weights. It has been argued that this type of weight normalization is biologically plausible. For example, a neuron might have a finite resource necessary for maintaining incoming synapses. This might keep an upper limit on the total summed size of the incoming synapses. While this sounds within the realm of biological possibility, and is obviously helpful in keeping Hebbian learning in check, little evidence from the experimental literature is available for support.

A more plausible mechanism for controlling the Hebbian process has recently emerged.

Turrigiano and others [14, 9, 13] have shown that neurons in the cortex actively maintain an average firing rate by scaling their incoming weights. The mechanism has been examined in cultures and in other experiments using in-vivo visual deprivation. It has been shown that the incoming synapses are altered by a multiplicative factor, which presumably preserves the relative strengths of the synapses. The underlying mechanisms are not yet known, but there is ongoing research looking at intracellular chemical factors such as calcium and brain-derived neurotrophic factor (BDNF) [14]. The level of these factors are related to firing rates, so integrating them over time could lead to an estimate of average firing rate and produce a chemical signal for synaptic change. Another interesting finding is that a neuron with high average firing rate will decrease the strength of incoming excitatory synapses, but increase the strength of incoming inhibitory neurons [9]. While many details are yet to be worked out, the homeostatic synaptic scaling mechanism is an interesting candidate to constrain weight growth.

2 Architecture with Homeostatic Synaptic Scaling

The SOM model is trained with a series of episodes in which randomly selected input vectors are presented. At each step, the input vector, $\vec{x}$, is first multiplied by the feedforward weights, W_{FF}. In order to get the self-organizing map effect, this feedforward activity is then multiplied by a set of lateral connections, W_{lat}, as in:

$$\vec{y} = f[W_{lat}(W_{FF}\vec{x})]$$

Here $f[a] = max(0, a)$ and W_{lat} is preset to a Mexican hat shape. In this version the lateral connections are not updated with learning. After the output activity is set, the feedforward weights are updated with a Hebbian learning rule:

$$\tilde{w}_{ij}^t = w_{ij}^{t-1} + \alpha x_j^t y_i^t$$

α is the Hebbian learning rate, x_j is the presynaptic activity and y_i is the postsynaptic activity.

It is this type of Hebbian learning rule that would normally give problems. Since each update is positive, there is nothing to limit the growth of the weights. Normally, a weight normalization is used that is based on the sum of the magnitude of the weights coming into each neuron. In our case, we will normalize the weights with a value based on the recent activity of the neuron:

$$w_{ij}^t = \frac{\tilde{w}_{ij}^t}{ActivityNorm_i^t}, \quad ActivityNorm_i^t = 1 + \beta_N(\frac{A_{avg,i}^{t-1} - A_{target}}{A_{target}})$$

Here, A_{target} is the internally defined preferred activity level for the neurons, $A_{avg,i}$ is the average of recent activity for neuron i, and β_N is the homeostatic learning rate. This model assumes that some intracellular chemical, such as calcium for example, is integrated over the recent past and is related to the average firing rate. If the target level of this chemical is exceeded, the incoming synapse sizes will be decreased multiplicatively. If the target is not achieved, the synapses will be increased. Since the underlying relevant chemicals and their dynamics are not yet known, we instead use the average firing rate and firing rate target

value directly in computing *ActivityNorm*. In the model, each neuron keeps track of its average output firing rate, $A_{avg,i}^t$, with a running average over the recent past. This is a local computation, in the sense that each neuron keeps track of its own average firing rate. If this average, or difference from the target, is expressed as an internal level of some chemical all the synapses would conceivably have access to that information. Using the $A_{avg,i}^t$ and A_{target} values directly avoids modelling the concentrations of unknown chemicals, but as more details become available through experiments, the model can become more explicit.

3 Simulation Results

Self-organizing maps were simulated using the synaptic scaling described with the previous equations. In other work, we show how to find the range of effective learning rate parameters, [12] so only values in this range are used here.

3.1 Homeostasis and Map Formation

In order to verify proper formation of a map, a network was created with 150 inputs and 15 outputs. The input vectors are specified by a 1-D gaussian shape (standard deviation σ of 15 units). The input gaussian is centered on one of the input units, selected uniformly at random. Plots of typical network behavior are shown in Figure 1. In the left-most plot, the average firing rate of the output neurons is shown. As the simulation progresses the average neuron firing rate approaches A_{target}. For each input, if we view the most active output neuron as the winner, then we can keep track of the neurons' winning percentages. The center plot shows these winning percentages for all the output neurons. It can be seen that they approach a roughly equal probability of winning. The right-most plot shows an input-output map that has formed after training has ended. To obtain this plot, every possible input was presented to the network, one at a time. The winning output neuron (the one with the highest output rate) was then recorded for each input. The input number is shown on the x axis, and the corresponding winning output is plotted. This is a good map since similar inputs (inputs whose center is located on neighboring input units) correspond to nearby winning output units.
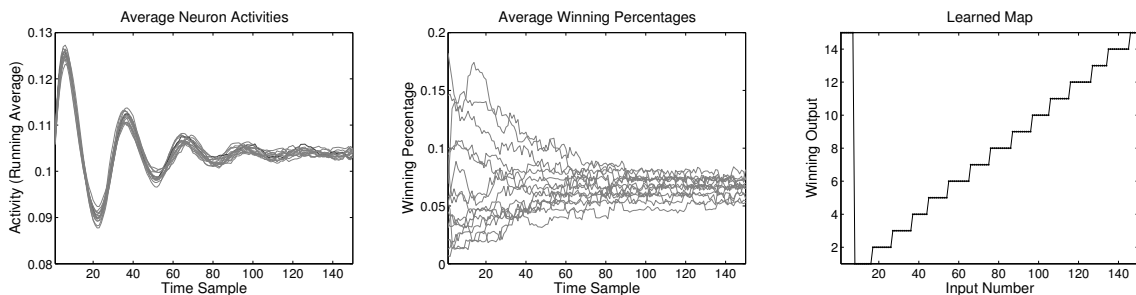

Figure 1: Typical Behavior. 150 inputs, 15 outputs, ring topology, $\alpha = 8.3\text{x}10^{-4}$, $\beta_N = 3.3\text{x}10^{-4}$, and $A_{target} = 0.1$ Hz *(Left)*. The average neuron activities are driven to the same value. A running average of the firing rate is shown. *(Middle)*. The winning percentages of each of the neurons is shown. They converge to a roughly equal winning frequency. *(Right)*. A smooth input-output map if formed. For each possible input, the output winner (the neuron with the maximum firing rate) is plotted. Both the inputs and outputs are arranged in a ring configuration to eliminate edge effects (so Output 1 is next to Output 15, for example).

3.2   Synapse Proliferation

Homeostatic mechanisms that maintain a steady output firing rate may play a particularly important role during development. As many neurons and synapses are added and pruned away, the total amount of input drive to a neuron will change dramatically. In order to avoid having a saturated firing rate, neurons must regulate themselves.

In order to test the ability of homeostatic synaptic scaling to withstand dramatic changes in network architecture, we created a simulation in which the number of input neurons doubled after learning had begun. The network started with 75 input neurons. After the average neuron activities settled to a constant value, 75 more inputs were added, as shown on the left side of Figure 2. After the neuron activities settled again, the 75 added inputs were taken away. This is a simple example meant to simulate the large scale neuron proliferation and die-off seen during cortical development.

The effect on the average output activities is shown in the middle of Figure 2. When the number of inputs was changed, the average firing rate of the output neurons changed. The firing rates quickly returned to the target value in the network with the homeostatic mechanism. Furthermore, the continuity (smoothness) of the map was unaffected. The network using standard weight normalization has its average output firing rate permanently changed by the additional (or subtracted) neurons. Real neurons with output nonlinearities could have their outputs saturate with such changes in input drive. This highlights the important role that homeostatic synaptic scaling may play during development. Additionally, during normal functioning in a hierarchical system such as the visual cortex, one area's output is another's input. For the benefit of higher areas, it may be important for neurons to maintain a consistent firing rate level.
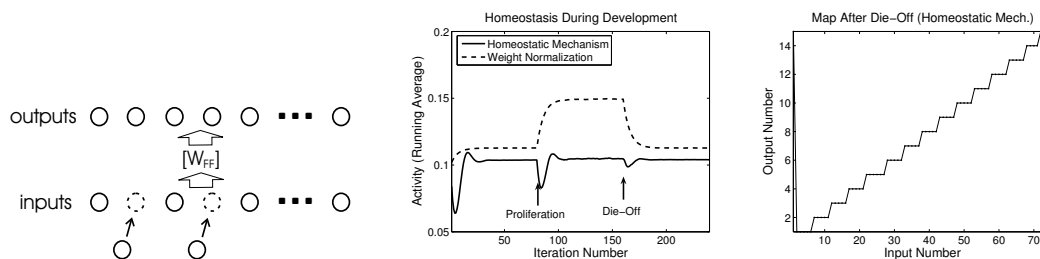

Figure 2: Addition of New Inputs. *(Left).* Ten additional inputs were added halfway through learning. *(Middle).* The average neuron activity is disturbed when new inputs are added or taken away, but the average quickly recovers if the homeostatic mechanism is used. *(Right).* For both cases, a smooth input-output map is preserved after neuron proliferation and die-off.

3.3   Probability Representation

Computer simulations were run that compare networks using homeostatic scaling with networks using standard weight normalization. Networks of neurons that maintain their own average firing rate avoid the problem of dead units that don't respond to any input and overactive neurons. Intuitively, it seems that networks of this type might get more out of their neurons and thus increase their information transfer. This thought is reminiscent of networks that explicitly try to obtain maximum information, such as the works of DeSieno [4] using a Conscience mechanism and Linsker [8] using information maximization. In both

of these examples, though, global information is needed at the synapses. Perhaps the homeostatic mechanism, or some variant, can approximate information maximization within a biologically-realistic framework. This idea is tested here by comparing network performance with inputs drawn from several probability distributions.

Since our networks do not have a winner-take-all output, there is no obvious winner for each input. For sake of comparison, we will define the winner as the neuron with the highest output rate. In this way we can find the probability of winning for each output neuron, and then calculate the entropy of this vector of probabilities. For our network with population-coded outputs, it may not be completely fair to measure information in this way, but for a
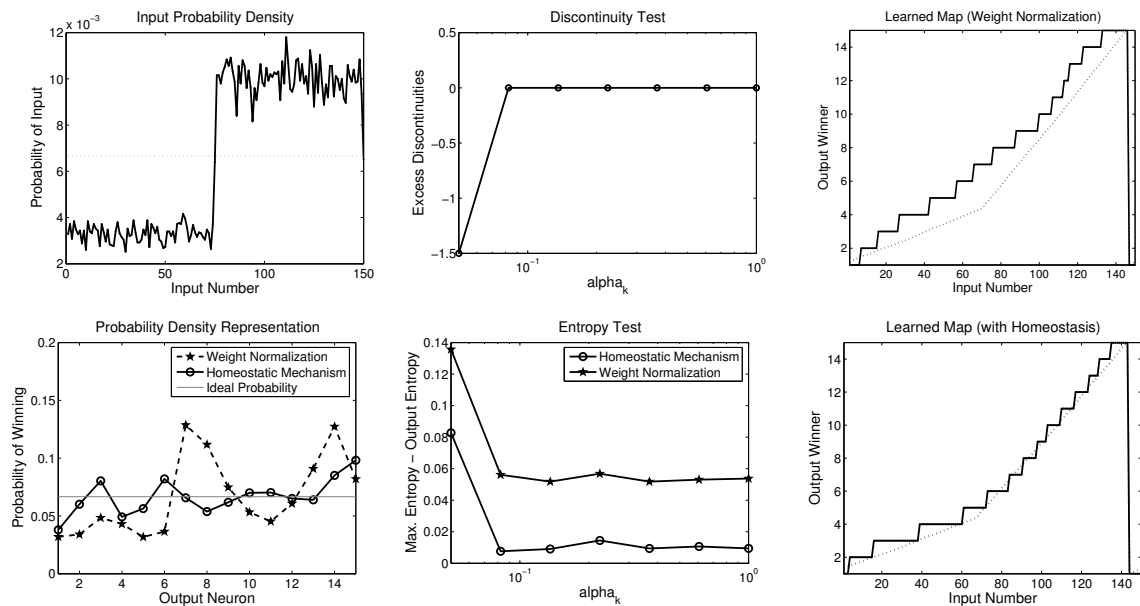

Figure 3: Performance Comparison on 150x15 Map with Step Input Distribution. *(Top Left)*. The center of the input patterns were drawn from this step distribution in which half of the potential inputs were three times more likely than the others. The input patterns were gaussian shapes with a standard deviation σ width equal to 15 input units. *(Bottom Left)*. For one simulation for each case (standard weight normalization (WN) and the homeostatic mechanism (HM)), the actual output winning probabilities are shown. The weight normalization case has large errors at the steps in the input probability densities. *(Center)*. Two measures of map quality were compared for WN and HM. The Hebbian learning rate, α_k was varied over a wide range. Each point is the average of five separate computer simulations. *(Top Center)*. The number of discontinuities in a given an input-output map (as in the bottom panels) were counted and subtracted from the number of output units. A smooth map that utilizes all the outputs will have a Discontinuity Test score of 0. Both WN and HM have the same perfect performance. *(Bottom Center)*. For a large number of randomly chosen inputs, the output unit with the highest activation was called the winner. The entropy of the output unit winning probabilities was computed and subtracted from the highest possible entropy (all winning an equal number of times). The best value is zero. The homeostatic mechanism had entropy that was closer to maximum (and thus higher information content) in all cases. *(Right)*. For the simulation whose results are depicted on the middle right panel, the final input-output maps are shown with *(Top Right)* WN and *(Bottom Right)* HM. For reference, the cumulative distribution of the input is plotted with a dashed line. The network with the homeostatic mechanism has learned a mapping that better matches this distribution, and thus increases the output entropy.

first approximation it might be helpful.

The first example, whose results are shown in Figure 3, uses networks with 150 inputs and 15 outputs. The learning rates and target output rates were set the same as above. The inputs are drawn from a step function with half the inputs being unlikely, while the other half are more likely (as seen in the top right panel). In this case, the network with homeostatic scaling has consistently better performance as measured by output entropy (see top right and middle right panels). The top right panel shows the probability of winning for each output neuron for one simulation. This was obtained by tested the trained network on a representative set of inputs. The difference in performance between the two network types is due to the shape of the input-output map that forms (bottom panels).

Several more networks were tested using different input distributions. For these simulations, networks of 750 inputs and 75 outputs were used. Typical results for each distribution are shown in Figure 4. For each row, the type of input distribution is shown on the left. The middle panel shows the probability of winning for each output neuron. The dark line represents the network with the homeostatic mechanism, and the dashed line gives the network with standard weight normalization. A faint line gives the ideal output winning probability, which is 1 divided by the number of output units. To the right of the plotted probabilities are the resulting entropy measures. These numbers are the difference between the entropy of the output probabilities and the maximum entropy for this network, making zero the best possible value.

In three out of four cases, the network with the homeostatic mechanism had better performance. The network with standard weight normalization was slightly better in the second case. Interestingly, this ramp-like distribution caused the homeostatic mechanism to converge to a state in which some neurons rarely won (had the most activation). These output neurons received enough activation from neighboring winners that their target activity goal was achieved. In other words, all neurons had similar average activities, but some neurons rarely "won". Also interesting were the results of the last distribution. This was the same input step distribution used in the example above. As the network size increased, performance gets worse for both networks. This is again due to the algorithm optimizing for average firing rate, not average winning percentage. The discrepancy between these measures, especially in the regions of low probability, should be interesting grounds for future investigation.

4 Conclusions

In this work, we have proposed a way to go beyond the standard weight normalization. This long-used measure has worked to counteract the unconstrained growth of Hebbian learning, but there is little evidence from experiments that justifies its use. Homeostatic synaptic scaling, on the other hand, has been seen recently in biological experiments. It has been shown here that using homeostatic synaptic scaling in place of standard weight normalization still leads to proper organized map formation. In addition, the neurons are able to maintain their average firing rates at a set point. This helps them from becoming saturated as neurons are added or taken away, as happens during development. Finally, it was shown that synaptic scaling in some cases leads to a better representation of the input probability distribution compared with weight normalization. This observation suggests the intriguing possibility that this homeostatic mechanism helps drive the network to a state of increasing
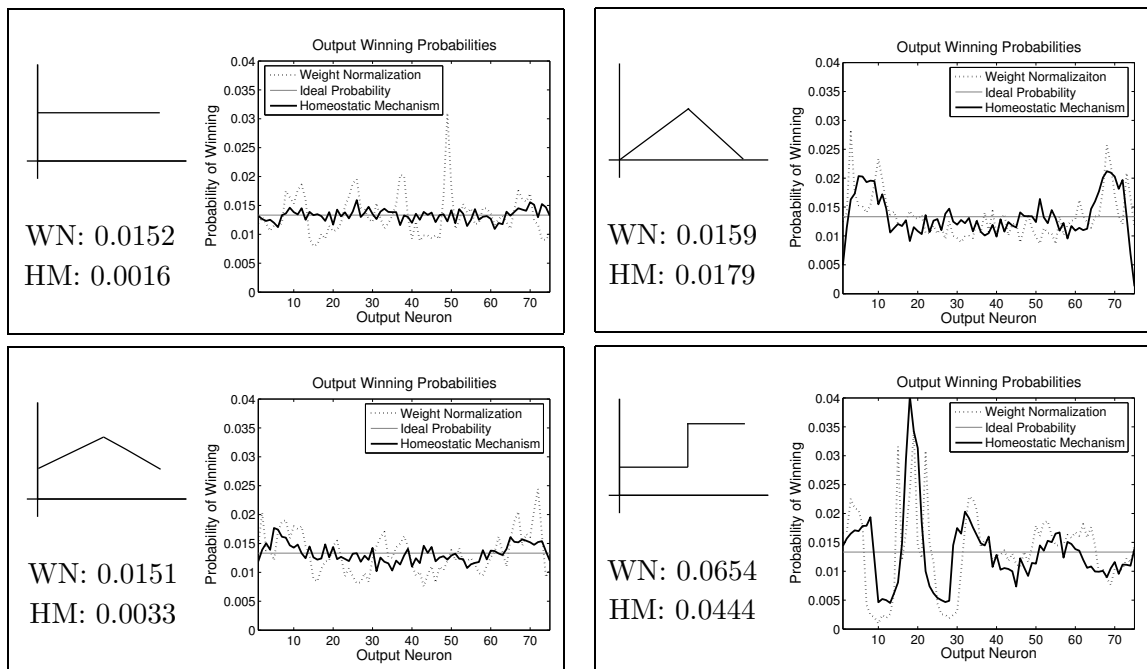
Figure 4: Examples with Various Input Density Distributions. Each panel shows the results of simulations with the two networks (Standard Weight Normalization (WN) and the Homeostatic Mechanism (HM)) using a different input distribution. Within each panel, a diagram shows the input distribution in the upper left. To the right, the probability of winning for each output neuron is plotted. The dark line represents HM, and the dashed line shows WN. A faint line gives the ideal output probability, which is 1 divided by the number of output units. The two numbers presented on the bottom left are the resulting entropy measures. These numbers are the difference between the entropy of the output probabilities and the maximum entropy for this network, making zero the best possible value. In three of the four examples, HM performs better than WN.

information transfer.

The output entropy was measured using the probability of each output neuron having the highest activation. This may not be the most natural way to measure information transfer in this network, since a population code is used as the output. Indeed, since the neurons' goal is to maintain a useful average firing rate, information transfer may not be the most important measure of performance. These issues will be addressed in future work. The algorithm will also be tested more extensively with two-dimensional input and output spaces. An interesting challenge is how to integrate this mechanism into existing models of cortical development like the LISSOM [11] and if it will lead to increased performance in practical applications [3]. How this homeostatic mechanism interacts with other synaptic modifications like short-term adaptation [5] and long-term depression (LTD) [6] will also be interesting avenues for investigation.

References

[1] T. V. P. Bliss and T. Lomo. Long-lasting potentiation of synaptic transmission in the dentate area of the anesthetized rabbit following stimulation of the perforant path. *Nature Neuroscience*, 232:331, 1973.

[2] Yoonsuck Choe and Risto Miikkulainen. Contour integration and segmentation with self-organized lateral connections. *Biological Cybernetics*, 90:75–88, 2004.

[3] Yoonsuck Choe, Joseph Sirosh, and Risto Miikkulainen. Laterally-interconnected self-organizing maps in hand-written digit recognition. In David Touretzky, Michael Mozer, and Mark Hasselmo, editors, *Advances in Neural Information Processing Systems 8*, volume 8, Cambridge, MA, 1996. MIT Press.

[4] Duane DeSieno. Adding a conscience to competitive learning. In *Proceedings of the IEEE International Conference on Neural Networks*, volume 1, pages 117–124, San Diego, 1988.

[5] Anatoli Gorchetchnikov. Introduction of threshold self-adjustment improves the convergence in feature-detective neural nets. *Neurocomputing*, 32-33:385–390, 2000.

[6] Guo-qiang Bi and Mu-ming Poo. Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience*, 24:139–166, 2001.

[7] Teuvo Kohonen. *Self-Organizing Maps*. Springer, third edition, 2001.

[8] Ralph Linsker. How to generate ordered maps by maximizing the mutual information between input and output signals. *Neural Computation*, 1(3):402–411, 1989.

[9] Arianna Maffei, Sacha B. Nelson, and Gina G. Turrigiano. Selective reconfiguration of layer 4 visual cortical circuitry by visual deprivation. *Nature Neuroscience*, 7(12):1353–1359, December 2004.

[10] K. Obermayer, G.G. Blasdel, and K. Schulten. Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps. *Physical Review A*, 45(10):7568, 1992.

[11] Joseph Sirosh and Risto Miikkulainen. Topographic receptive fields and patterned lateral interation in a self-organizing model of the primary visual cortex. *Neural Computation*, 9(3):577–594, 1997.

[12] Thomas J. Sullivan and Virginia R. de Sa. A self-organizing map with homeostatic synaptic scaling. *Computational Neuroscience Meeting*, 2005.

[13] Gina G. Turrigiano, Kenneth R. Leslie, Niraj S. Desai, Lana C. Rutherford, and Sacha B. Nelson. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391:892–896, 26 February 1998.

[14] Gina G. Turrigiano and Sacha B. Nelson. Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience*, 5:97–107, February 2004.