Investigation of Topographical Stability of the Concave and Convex Self-Organizing Map Variant

Fabien Molle[1,2] and Jens Christian Claussen[2,⋆]

[1] Theoretical Physics, Chalmers Tekniska Högskola, Göteborg
[2] Institut für Theoretische Physik und Astrophysik
Leibnizstr. 15, 24098 Christian-Albrechts-Universität zu Kiel, Germany
claussen@theo-physik.uni-kiel.de
http://www.theo-physik.uni-kiel.de/~claussen/

Abstract. We investigate, by a systematic numerical study, the parameter dependence of the stability of the Kohonen Self-Organizing Map and the Zheng and Greenleaf concave and convex learning with respect to different input distributions, input and output dimensions.

Topical groups: Advances in Neural Network Learning Methods, Neural and hybrid architectures and learning algorithms, Self-organization.

Neural vector quantizers have become a widespreadly used tool to explore high-dimensional data sets by self-organized learning schemes. Compared to the vast literature on variants and applications that appeared the last two decades, the theoretical description proceeded more slowly. Even for the coining Self-Organizing Map (SOM) [1], still open questions remain, as a proper description of the dynamics for the case of dimension reduction and varying data dimensionality, or the question for what parameters stability of the algorithm can be guaranteed. This paper is devoted to the latter question. The stability criteria are especially interesting for modifications and variants, as the concave and convex learning [2], whose magnification behaviour has been discussed recently [3]. Especially for the variants, analytical progress becomes quite difficult, and in any case one will expect that the stability will depend on the input distribution to some —apart from special cases— unknown extent. As the invariant density in general is analytically unaccessible for input dimensions larger than one (see [4] for recent tractable cases), we expect a general theory not to be available immediately, and instead proceed with a systematic numerical exploration.

The Kohonen SOM, and the nonlinear variant of Zheng and Greenleaf. – The class of algorithms investigated here is defined by the learning rule, that for each stimulus $\mathbf{v} \in V$ each weight vector $\mathbf{w_r}$ is updated according to

$$\mathbf{w}_\mathbf{r}^{\mathsf{new}} = \mathbf{w}_\mathbf{r}^{\mathsf{old}} + \varepsilon \cdot g_{\mathbf{rs}} \cdot (\mathbf{v} - \mathbf{w}_\mathbf{r}^{\mathsf{old}})^K \tag{1}$$

⋆ Corresponding author.

S. Kollias et al. (Eds.): ICANN 2006, Part I, LNCS 4131, pp. 208–215, 2006.

($g_{\mathbf{rs}}$ being a gaussian function (width σ) of euclidian distance $|\mathbf{r}-\mathbf{s}|$ in the neural layer, thus describing the neural topology). Herein

$$|\mathbf{w_s} - \mathbf{v}| = \min_{\mathbf{r}\in R} |\mathbf{w_r} - \mathbf{v}| \tag{2}$$

determines for each stimulus $\mathbf{v}$ the *best-matching unit* or *winner* neuron.

The case $K = 1$ is the original SOM [1], corresponding to a linear or Hebbian learning rule. The generalization to K or $1/K$ taking integer values has been proposed by Zheng and Greenleaf [2], but arbitrary nonzero real values of K can be used [3], and the choice of $K \to 0$ has been shown (for the onedimensional case) to have an invariant density with the information-theoretically optimal value of the magnification exponent one [3], i.e., the neural density is proportional to the input density and hence can be used as a density estimator.

Convergence and stability. – It is well known that for the learning rate ε, one has to fulfill the Robbins-Munro conditions (see, e.g. [4]) to ensure convergence, with all other parameters fixed. However, practically it is necessary to use a large neighborhood width at the beginning, to have the network of weight vectors ordered in input space, and decrease this width in the course of time downto a small value that ensures topology preservation during further on-line learning. Thus the situation becomes more involved when additionally also σ is made time-dependent. Here we consider the strategy where the stability border in the (ε, σ) plane always is approached from small ε with σ fixed during this final phase. An ordered state has to be generated by preceding learning phases.

Measures for Topographical Stability. – To quantify the ordered state and the topology preservation, a variety of measures is used, e.g. the topographic product [5], the Zrehen measure [6], and the average quadratic reconstruction error. To detect instable behaviour, all measures should be suitable and give similar results. For an unstable and disordered map, also the total sum over all (squared) distances between adjacent weight vectors will increase significantly; so a threshholded increase will indicate instability as well. This indicator is used below; however, for the case of a large neighborhood (of network size), the weight vectors shrink to a small volume, thus influencing the results; however, this applies to a neighborhood widths larger than that commonly used for the pre-ordering.

In addition we use here an even more simple approach than the Zrehen measure (which counts the number of neurons that lie within a circle between each pair of neurons that are adjacent in the neural layer). For a mapping from d to d dimensions, we consider the determinant of the i vectors spanned by $\boldsymbol{w}_{\boldsymbol{r}+\mathbf{e}_i} - \boldsymbol{w}_{\boldsymbol{r}}$, with $\mathbf{e}_i$ being the i^{th} unit vector. The sign of this determinant, where $1 \le i \le d$, thus gives the orientation of the set of d vectors. Note that the 1-dimensional case just reads $\text{sgn}(w_{r+1} - w_r)$, which has been widely considered to detect the ordered state in the 1 to 1 dimensional case. Hence, we can define

$$\chi(\{\boldsymbol{w_r}\}) := \tag{3}$$
$$1 - \frac{1}{N}\left|\sum_{\boldsymbol{r}} \text{sgn}(\det((\boldsymbol{w}_{\boldsymbol{r}+\mathbf{e}_1} - \boldsymbol{w}_{\boldsymbol{r}}), \ldots (\boldsymbol{w}_{\boldsymbol{r}+\mathbf{e}_i} - \boldsymbol{w}_{\boldsymbol{r}}), \ldots (\boldsymbol{w}_{\boldsymbol{r}+\mathbf{e}_d} - \boldsymbol{w}_{\boldsymbol{r}})))\right| .$$

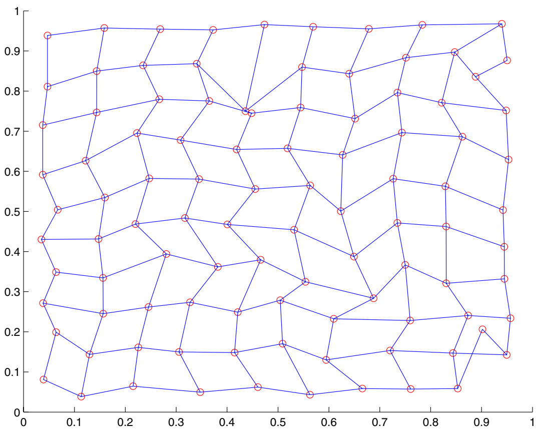

Fig. 1. Situation where one defect is detected by the crossproduct measure (3)

This evaluates the number of neurons N_+ (resp. N_-), where this sign is positive (resp. negative), hence the relative fraction of minority signs is given by $(1 - |N_+ - N_-|/N)$. A typical single defect is shown in Fig. 1. Due to its simplicity, this measure χ will be used in the remainder.

Modification of learning rules and data representation. – A classical result [7,8,9] states that the neural density for 1-D SOM (in the continuum limes) approaches not the input density itself, but a power of it, with exponent 2/3, the so-called magnification exponent. As pointed out by Linsker [10], the case of an exponent 1 would correspond to the case of maximal mutual information between input and output space. Different modifications of the winner mechanism or the learning rule, by additive or multiplicative terms, have been suggested and influence the magnification exponent [11,12,13,14]. Here we investigate the case of concave and convex learning [2,3], which defines a nonlinear generalization of the SOM.

Topographical Stability for the Self-Organizing Map. – Before investigating the case of concave and convex learning, the stability measures should be tested for the well-established SOM algorithm. Using the parameter path of Fig. 2, we first analyze the 2D → 2D case, for three input distributions: the homogeneous input density (equidistribution), an inhomogeneous input distribution $\sim \sin(\pi x_i)$ [14], and a varying-dimension dataset (Figs. 3, 4). The results are shown in Fig. 5.

Different input dimensions and varying intrinsic dimension. – As the input dimensionality is of pronounced influence on the maximal stable learning rate (Fig. 6), we also investigate an artificial dataset combining different dimensions: the box-plane-stick-loop dataset [15] (Fig. 3), or its 2D counterpart, the plane-stick-loop (Fig. 4). Here the crossproduct detection will become problematic where the input space is intrinsically 1D (stick and loop), thus the average distance criterion is used, and we restrict to the case $\sigma \leq 1$.

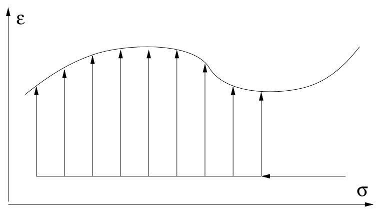

Fig. 2. Schematic diagram of the parameter path in (ε, σ) space. Starting with high values, σ is slowly decreased to the desired value, while the learning rate still is kept safely low. From there, at constant σ the learning rate is increased until instability is observed; giving an upper border to the stability area. – The same scheme is applied for the concave and convex learning, where the nonlinearity exponent is considered as a fixed parameter.

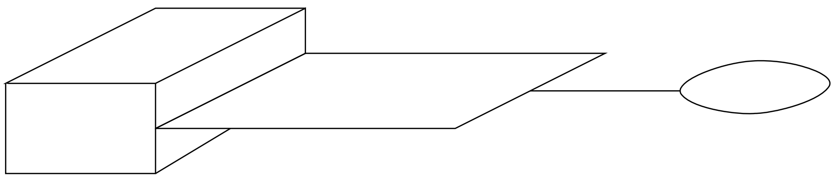

Fig. 3. Schematic view of the classical box-plane-stick-loop dataset. Its motivation is to combine locally different input data dimensions within one data set.

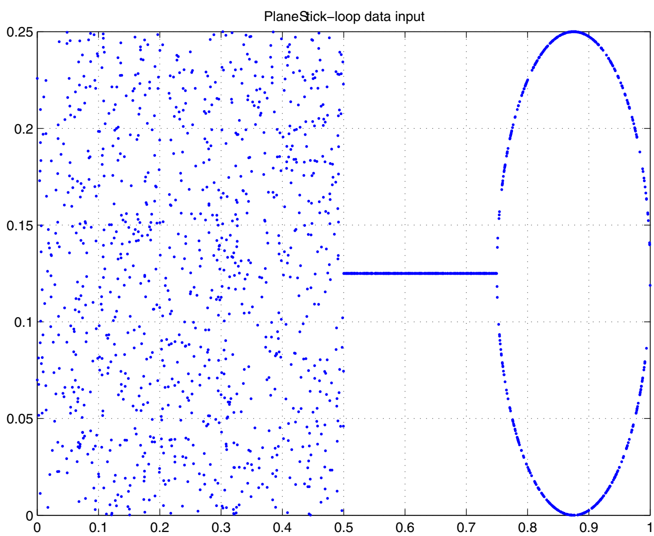

Fig. 4. Part of the input data for the 2D plane-stick-loop data set (Fig. 3)

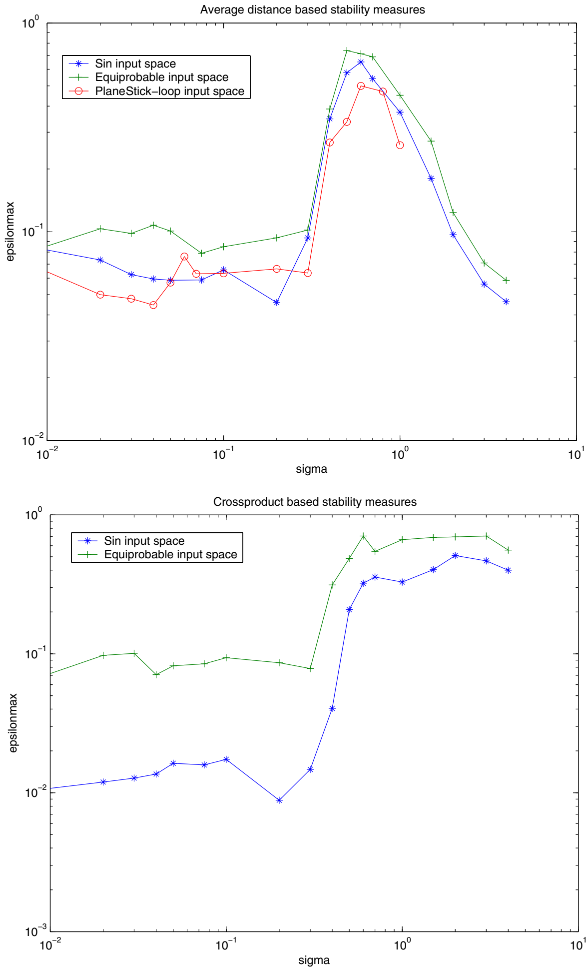

Fig. 5. Critical $\varepsilon_{\max}(\sigma)$ where (coming from small ε values, see Fig. 2) the SOM learning loses stability. Here a 2D array of $10{\times}10$ neurons was used with decay $\exp(-t/k)$ exponentially in time t, with k between 30000 and 60000 depending on ε_0 (for σ between 0.1 and 0.001, $k = 300000$). Top: Unstable ε detected from growth of the averaged distance of neurons; here a threshold of 15% was chosen. For large σ, this measure becomes less reliable due to shrinking of the network, i.e. $\forall_{\boldsymbol{r}} \boldsymbol{w}_{\boldsymbol{r}} \to \langle \boldsymbol{v} \rangle$. Bottom: Unstable ε detected from the crossproduct measure, eq. (3), with threshold of 1 defect per 100 iterations. The ε value depends on the data distribution, but the qualitative behaviour remains similar. In all cases, below a certain σ_{crit} of about 0.3, ε has to be decreased significantly. This independently reproduces [16], here we investigate also different ε values.

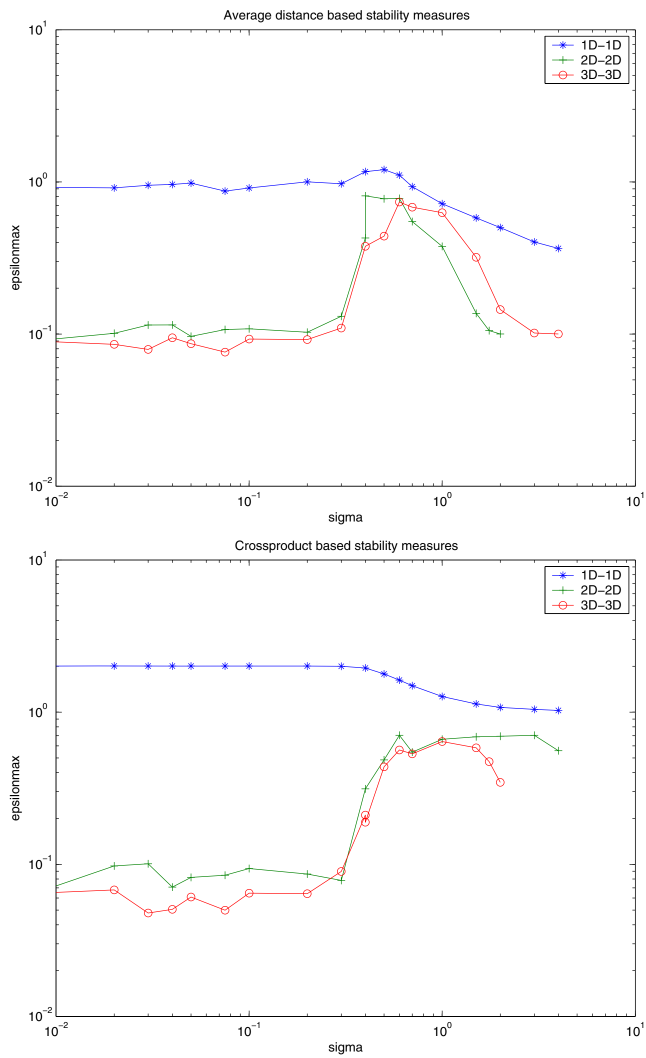

Fig. 6. Stability border dependence on input dimension (1D, 2D, 3D). The known 1D case is included for comparison. Top: Using the average length criterion (for $\sigma > 1$, the result can be misleading due to total shrinking of the network, see text). Bottom: Using the crossproduct detection for defects, similar results are obtained; for large σ instabilities are detected earlier.

Concave and convex learning: Stability of nonlinear learning. – The simulation results are given in Fig. 7: Clearly, a strong influence of the nonlinearity parameter K is observed. Especially one has to take care when decreasing σ, because for

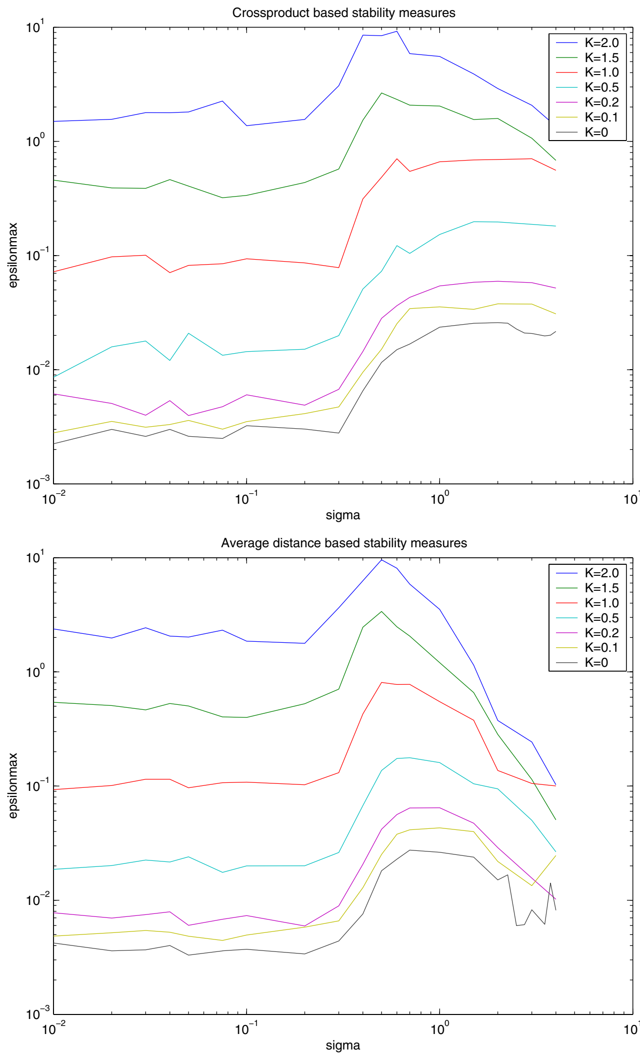

Fig. 7. Critical $\varepsilon_{\max}(\sigma)$ for different values of the nonlinearity parameter from $K = 2.0$ (top) to $K = 0$ (bottom). $K = 1$ corresponds to the SOM case.

too large ϵ the network becomes instable. For $K < 1$, much smaller values of ϵ are possible, thus considerably longer learning phases have to be taken into account compared to original SOM. For $K > 1$ the stability range becomes larger.

Discussion. – We have defined a standardized testbed for the stability analysis of SOM vector quantizers with serial pattern presentation, and compared the SOM with the recently introduced variants of concave and convex learning. The stability regions for different input distribution and dimension are of the same shape, thus qualitatively similar, but not coinciding exactly. The neighborhood width, but unfortunately also the input distribution affect the maximal stable learning rate. For the concave and convex learning, the exponent steering the nonlinear learning also crucially influences the learning rate. In all cases, a plateau for $\sigma \ll 1$ is found where the learning rate must be quite low compared to the intermediate range $0.3 \leq \sigma \leq 1$. As a too safe choice of the learning rate simply increases computational cost, an accurate knowledge of the stability range of neural vector quantizers is of direct relevance in many applications.

References

1. T. Kohonen. Self-Organized Formation of Toplogically Correct Feature Maps. Biological Cybernetics **43** (1982) 59-69
2. Y. Zheng and J. F. Greenleaf. The effect of concave and convex weight adjustments on self-organizing maps. IEEE Transactions on Neural Networks, **7** (1996) 87-96
3. T. Villmann and J. C. Claussen, Investigation of Magnification Control in Self-Organizing Maps and Neural Gas. Neural Computation **18** (2006) 449-469
4. T. Kohonen. Comparison of SOM point densities based on different criteria. Neural Computation **11** (1999) 2081-2095
5. T. Villmann, R. Der, M. Herrmann, and T. Martinetz. Topology Preservation in Self-Organizing Feature Maps: Exact Definition and Measurement. IEEE Transactions on Neural Networks, **8** (1997) 256-266
6. S. Zrehen. Analyzing Kohonen maps with geometry. In Gielen, S. and Kappen, B., editors, Proc. ICANN'93, p. 609-612, Springer, London (1993)
7. H. Ritter and K. Schulten. On the Stationary State of Kohonen's Self-Organizing Sensory Mapping. Biological Cybernetics **54** (1986) 99-106
8. H. Ritter and K. Schulten. Convergence Properties of Kohonen's Topology Conserving Maps: Fluctuations, Stability and Dimension Selection. Biological Cybernetics **60** (1988) 59-71
9. H. Ritter, T. Martinetz, and K. Schulten. Neural Computation and Self-Organizing Maps: An Introduction. Addison-Wesley (1992)
10. R. Linsker. How to generate maps by maximizing the mutual information between input and output signals. Neural Computation **1** (1989) 402-411
11. T. Heskes. Energy functions for self-organizing maps. In E. Oja and S. Kaski, editors, *Kohonen Maps*, p. 303-316. Elsevier, Amsterdam (1999)
12. E. Erwin, K. Obermayer, and K. Schulten. Self-organizing maps: Ordering, convergence properties and energy functions. Biol. Cyb. **67** (1992) 47–55
13. J. C. Claussen, Generalized Winner Relaxing Kohonen Feature Maps, Neural Computation **17** (2005) 996-1009
14. J. C. Claussen and T. Villmann, Magnification Control in Winner Relaxing Neural Gas, Neurocomputing **63** (2005) 125-137 2005.
15. T. M. Martinetz, S. G. Berkovich, and K. J. Schulten. 'Neural-gas' network for vector quantization and its application to time-series prediction. *IEEE Trans. on Neural Networks*, **4** (1993) 558–569
16. T. Villmann. PhD thesis, Leipzig, Verlag Harri Deutsch, Frankfurt (1998)