Published in: Neurocomputing 63, 125-137 (2004)

Magnification Control in Winner Relaxing Neural Gas

Jens Christian Claussen [a] and Thomas Villmann [b]

[a] *Christian-Albrechts-Universität zu Kiel, Institut für Theoretische Physik und Astrophysik, D–24098 Kiel, Leibnizstr.15, Germany*

[b] *Universität Leipzig, Klinik für Psychotherapie, D–04107 Leipzig, Karl–Tauchnitz–Str.25, Germany*

24 October 2003, final version 2 February, 2004

Abstract

An important goal in neural map learning, which can conveniently be accomplished by magnification control, is to achieve information optimal coding in the sense of information theory. In the present contribution we consider the winner relaxing approach for the neural gas network. Originally, winner relaxing learning is a slight modification of the self-organizing map learning rule that allows for adjustment of the magnification behavior by an *a priori* chosen control parameter. We transfer this approach to the neural gas algorithm. The magnification exponent can be calculated analytically for arbitrary dimension from a continuum theory, and the entropy of the resulting map is studied numerically conf irming the theoretical prediction. The influence of a diagonal term, which can be added without impacting the magnification, is studied numerically. This approach to maps of maximal mutual information is interesting for applications as the winner relaxing term only adds computational cost of same order and is easy to implement. In particular, it is not necessary to estimate the generally unknown data probability density as in other magnification control approaches.

Key words: Neural gas, Self-organizing maps, Magnification control, Vector quantization

1 Introduction

Neural maps are a widely ranging class of neural vector quantizers which are commonly used e.g. in data visualization, feature extraction, principle component analysis, image processing, and classification tasks. A well studied approach is the Neural Gas Network (NG) [21]. An important advantage of the NG is the adaptation dynamics, which minimizes a potential, in contrast to the self-organizing map (SOM) [17] frequently used in vector quantization problems.

In the present paper we consider a new control scheme for the *magnification* of the map [2,10,12,20,24,27]. Controlling the magnification factor is relevant for many applications in control theory or robotics, were (neural) vector quantizers are often used to determine the actual state of the system in a first step, which is an objective of the control task [31,27]. For instance, in [28] was demonstrated that the application of a magnification control scheme for the neural gas based classification system of position and movement state of a robot can reduce the crash probability. Another area of application is information-theoretically optimal coding of high-dimensional data as occur in satellite remote sensing image analysis of hyperspectral images [22,30] which is, in fact, the task of equiprobabilistic mapping [26]. Further applications can be found in medical visualization and classification tasks [32]. Generally, vector quantization according to an arbitrary l_p-norm can be related to the problem of magnification control as it is explained below.

2 The neural gas network

The NG maps data vectors $\mathbf{v}$ from a (possibly high-dimensional) data manifold $\mathcal{D} \subseteq \mathbb{R}^d$ onto a set A of neurons i, formally written as $\Psi_{\mathcal{D}\to A} : \mathcal{D} \to A$. Each neuron i is associated with a pointer $\mathbf{w}_i \in \mathbb{R}^d$ also called weight vector, or codebook vector. All weight vectors establish the set $\mathbf{W} = \{\mathbf{w}_i\}_{i\in A}$. The mapping description is a winner take all rule, i.e. a stimulus vector $\mathbf{v} \in \mathcal{D}$ is mapped onto the neuron $s \in A$ the pointer $\mathbf{w}_s$ of which is closest to the actually presented stimulus vector $\mathbf{v}$,

$$\Psi_{\mathcal{D}\to A} : \mathbf{v} \mapsto s(\mathbf{v}) = \underset{i\in A}{\operatorname{argmin}} \|\mathbf{v} - \mathbf{w}_i\| . \tag{1}$$

The neuron $s(\mathbf{v})$ is called *winner neuron*. The set

$$\Omega_i = \{\mathbf{v} \in \mathcal{D} | \Psi_{\mathcal{D}\to A}(\mathbf{v}) = i\} \tag{2}$$

is called (masked) receptive field of the neuron i.

During the adaptation process a sequence of data points $\mathbf{v} \in \mathcal{D}$ is presented to the map with respect to the stimuli distribution $P(\mathcal{D})$. Each time the currently most proximate neuron s according to (1) is determined, and the pointer $\mathbf{w}_s$ as well as all pointers $\mathbf{w}_i$ of neurons in the neighborhood of $\mathbf{w}_s$ are shifted towards $\mathbf{v}$, according to

$$\triangle \mathbf{w}_i = \epsilon h_\lambda (i, \mathbf{v}, \mathbf{W}) (\mathbf{v} - \mathbf{w}_i) . \tag{3}$$

The property of "being in the neighborhood of $\mathbf{w}_s$" is represented by a neighborhood function $h_\lambda(i, \mathbf{v}, \mathbf{W})$. The neighborhood function is defined as

$$h_\lambda(i, \mathbf{v}, \mathbf{W}) = \exp\left(-\frac{k_i(\mathbf{v}, \mathbf{W})}{\lambda}\right), \tag{4}$$

where $k_i(\mathbf{v}, \mathbf{W})$ is defined as the number of pointers $\mathbf{w}_j$ for which the relation $\|\mathbf{v} - \mathbf{w}_j\| \leq \|\mathbf{v} - \mathbf{w}_i\|$ is valid, i.e. $k_i(\mathbf{v}, \mathbf{W})$ is the winning rank [21]. In particular, for the winning neuron s we have $h_\lambda(s, \mathbf{v}, \mathbf{W}) = 1.0$. We remark that in contrast to the SOM the neighborhood function is evaluated in the input space. Moreover, the adaptation rule for the weight vectors in average follows a potential dynamics [21].

The *magnification* of the trained map reflects the relation between the data density $P(\mathcal{D})$ and the density ρ of the weight vectors [23]. For the NG the relation

$$\rho(\mathbf{w}) \propto P(\mathcal{D})^{\alpha_{\mathsf{NG}}} \tag{5}$$

with

$$\alpha_{\mathsf{NG}} = \frac{d}{d+2} \tag{6}$$

has been derived [21]. The exponent α_{NG} is called *magnification factor*. For the NG it depends on the *intrinsic* dimensionality d of the data which can be numerically determined by several methods [5,6,7,14,25]. For simplicity we further require that the (embedding) data dimension is the intrinsic one.

Generally, the information transfer is not independent of the magnification of the map [33]. It is known that for a vector quantizer (or a neural map in our context) with optimal information transfer the relation $\alpha = 1$ holds. Otherwise, a vector quantizer which minimizes the mean distortion error

$$E_\gamma = \int_{\mathcal{D}} \|\mathbf{w}_s - \mathbf{v}\|^\gamma P(\mathbf{v})\, d\mathbf{v} \tag{7}$$

has the magnification factor

$$\alpha = \frac{d}{d+\gamma} \tag{8}$$

with $\mathbf{v} \in \mathcal{D} \subseteq \mathbb{R}^d$, i.e. the magnification of a vector quantizer is directly related to the minimization of the description error according to a certain l_p-norm [33]. Hence, the NG minimizes the usual E_2 distortion error.

We now address the question how to extend the NG to achieve an *a priori* chosen optimization goal, i.e. an *a priori* chosen magnification factor.

3 Controlling the magnification in NG

For the SOM several methods exist to control the magnification of the map. The first approach to influence the magnification of a learning vector quantizer, proposed in [13] is called the *mechanism of conscience*. For this purpose a bias term is added in the winner rule (1):

$$\Psi_{\mathcal{D}\to A} : \mathbf{v} \mapsto s\left(\mathbf{v}\right) = \operatorname*{argmin}_{i\in A}\left(\left\|\mathbf{v}-\mathbf{w}_i\right\| - \gamma\left(\frac{1}{N}-p_i\right)\right) \tag{9}$$

where p_i is the actual winning probability of the neuron i and γ is a balance factor. Hence, the winner determination is influenced by this modification. The algorithm should converge such that the winning probabilities of all neurons are equalized. This is related to a maximization of the entropy and consequently the resulting magnification is equal to unity. However, as pointed out by [26], adding a conscience algorithm to the SOM does not equate to equiprobabilistic mapping, in general. Only for *very high dimensions*, a minimum distortion quantizer (such as the conscience algorithm) approaches an equiprobable quantizer ([26] - page 93). Further, an arbitrary magnification cannot be achieved by this mechanism. Moreover, numerical studies of the algorithm have shown instabilities [26]. To control the magnification, a local learning parameter was introduced [1] into the usual SOM-learning scheme. The now localized learning allows in principle an arbitrary magnification. Other authors proposed variants which lead more away from the original SOM by kernel methods [26] or statistical approaches [19].

For the NG a solution of the magnification control problem can be realized by introducing an adaptive *local learning* step size $\epsilon_{s(\mathbf{v})}$ [27] according to the above mentioned approach for SOM [1]. Then, the new *localized* learning rule reads as

$$\triangle\,\mathbf{w}_i = \epsilon_{s(\mathbf{v})} h_\lambda\left(i, \mathbf{v}, \mathbf{W}\right)\left(\mathbf{v}-\mathbf{w}_i\right) \tag{10}$$

with the local learning parameters $\epsilon_i = \epsilon\left(\mathbf{w}_i\right)$ depending on the stimulus density P at the position of the weight vectors $\mathbf{w}_i$ via

$$\left\langle \epsilon_i \right\rangle = \epsilon_0 P\left(\mathbf{w}_i\right)^m . \tag{11}$$

The brackets $\langle \ldots \rangle$ denote the average in time, and $s\left(\mathbf{v}\right)$ is the best–matching neuron with respect to (1). Note, that the local learning rate $\epsilon_{s(\mathbf{v})}$ of the winning neuron is applied in the adaptation step (10) for each neuron. This approach finally leads to the new magnification law

$$\alpha' = \alpha_{NG} \cdot (m+1) \tag{12}$$

which is a modification of the old one. Hence, the parameter m plays the role of a control parameter.

However, in real applications one has to estimate the generally unknown data distribution P. Usually this is done by estimation of the volume of the receptive fields and the firing rates [1,29]. This may lead to numerical instabilities of the control mechanism [15,26,31]. Therefore, an alternative control mechanism is demanded.

Recently, a new approach for magnification control of the SOM was introduced [8,9] which avoids the P-estimation problem. The respective approach is a generalization of a modification of the usual SOM [16]. It is called (generalized) Winner Relaxing SOM (WRSOM) [8,9]. In winner relaxing SOM an additional term occurs in weight vector update for the winning neuron, implementing a relaxing behavior. The relaxing force is a weighted sum of the difference between the weight vectors and the input according to their distance rank. The relaxing term was originally introduced in [16] to obtain a learning dynamic for SOM according to an average reconstruction error including the effect of shifting Voronoi borders.

It was shown that the generalized winner relaxing mechanism applied in WRSOM can be used for magnification control in SOM, too [8]. Thereby, the winner relaxing approach provides a magnification control scheme for SOM which is *independent* of the shape of the data distribution only depending on parameters of the winner relaxing term.

3.1 The winner relaxing neural gas

We now transfer the generalized winner relaxing approach for SOM to the NG and consider its influence on the magnification. In complete analogy to the WRSOM we add a general winner relaxing term $R(\xi, \kappa)$ to the usual NG-learning dynamic (3). Then the weight update reads as

$$\triangle \mathbf{w}_i = \epsilon h_\lambda (i, \mathbf{v}, \mathbf{W}) (\mathbf{v} - \mathbf{w}_i) + R(\xi, \kappa), \tag{13}$$

whereby the winner relaxing term is defined as

$$R(\xi, \kappa) = (\xi + \kappa)(\mathbf{v} - \mathbf{w}_i)\,\delta_{is} - \kappa \delta_{is} \sum_j h_\lambda (j, \mathbf{v}, \mathbf{W}) (\mathbf{v} - \mathbf{w}_j) \tag{14}$$

depending on the additional weighting parameters ξ and κ. We refer to this algorithm as the *winner relaxing NG* (WRNG). The original winner relaxing term described in [16] is obtained for the special parameter choice $\xi = 0, \kappa = \frac{1}{2}$. Note, that the relaxing term only contributes to the winner weight vector update as in the original approach.

3.2 Derivation of the Magnification for WRNG

We now derive a relation between the densities ρ and P in analogy to [21] for the winner relaxing learning (13). The procedure is very similar as in [21,27]. The average change $\langle \triangle \mathbf{w}_i \rangle$ for the winner relaxing NG learning rule (13) is

$$\begin{aligned}\langle \triangle \mathbf{w}_i \rangle = \int P(\mathbf{v})\, h_\lambda(i, \mathbf{v}, \mathbf{W})(\mathbf{v} - \mathbf{w}_i) + (\xi + \kappa) \cdot \delta_{is} \cdot (\mathbf{v} - \mathbf{w}_i) \\ -\delta_{is} \kappa \sum_j h_\lambda(j, \mathbf{v}, \mathbf{W})(\mathbf{v} - \mathbf{w}_j)\, d\mathbf{v}.\end{aligned} \tag{15}$$

We now consider the equilibrium state, i.e. $\langle \triangle \mathbf{w}_i \rangle = 0$.

For this purpose, we first separate the integral (15) into

$$\langle \triangle \mathbf{w}_i \rangle = I_1 + I_2 + I_3 \tag{16}$$

with

$$I_1 = \int P(\mathbf{v})\, h_\lambda(i, \mathbf{v}, \mathbf{W})(\mathbf{v} - \mathbf{w}_i)\, d\mathbf{v}, \tag{17}$$

$$I_2 = \int P(\mathbf{v})(\xi + \kappa) \cdot \delta_{is} \cdot (\mathbf{v} - \mathbf{w}_i)\, d\mathbf{v} \tag{18}$$

and

$$I_3 = -\int P(\mathbf{v})\, \delta_{is} \kappa \sum_j h_\lambda(j, \mathbf{v}, \mathbf{W})(\mathbf{v} - \mathbf{w}_j)\, d\mathbf{v} \tag{19}$$

The integral I_1 is the usual one according to the NG dynamics whereas I_2, I_3 are related to the winner relaxing scheme. In the following we treat each integral in a separate manner. Thereby we always assume a continuum approach, i.e. the index i becomes continuous. Hence, for a given input $\mathbf{v}$ one can find an optimal $\mathbf{w}_s$ fulfilling even $\mathbf{w}_s = \mathbf{v}$ [24].

Doing so, the I_2-integral vanishes in the (first order) continuum limit because the integration over δ_{is} only contributes for $\mathbf{w}_s$, but in this case $(\mathbf{v} - \mathbf{w}_s) = 0$ holds.

We now pay attention to the I_3-integral: The continuum assumption made above allows a turn over from sum $\sum_j h_\lambda(\mathbf{w}_j, \mathbf{v}, \mathbf{W})(\mathbf{v} - \mathbf{w}_j)$ to the integral form $\int h_\lambda(\mathbf{w}, \mathbf{v}, \mathbf{W})(\mathbf{v} - \mathbf{w})\, d\mathbf{w}$ in (19). The further treatment is in complete analogy to the derivation of the magnification in the usual NG [21]. Let $\mathbf{r}$ be the difference vector

$$\mathbf{r} = \mathbf{v} - \mathbf{w}_i \tag{20}$$

The winning rank $k_i(\mathbf{v}, \mathbf{W})$ only depends on $\mathbf{r}$, therefore we introduce the new variable

$$\mathbf{x}(\mathbf{r}) = \hat{\mathbf{r}} \cdot \mathbf{k}_i(\mathbf{r})^{\frac{1}{d}} \tag{21}$$

which can be assumed as monotonously increasing with $\|\mathbf{r}\|$. Thus, the inverse $\mathbf{r}(\mathbf{x})$ exists and we can rewrite the I_3-integral (19) into

$$I_3 = \int P(\mathbf{v})\, \delta_{is} \kappa \left[\int h_\lambda(\mathbf{x}) \cdot \mathbf{r}(\mathbf{x}) \cdot \mathbf{J}(\mathbf{x})\, d\mathbf{x}\right] d\mathbf{v} \tag{22}$$

with the $d \times d$–Jacobian–matrix

$$\mathbf{J}(\mathbf{x}) = \det\left(\frac{\partial r_k}{\partial x_l}\right). \tag{23}$$

I_3 only contributes to $\langle \triangle \mathbf{w}_i \rangle$ for the winning weight (realized by δ_{is}), i.e., for $\mathbf{w}_i = \mathbf{w}_s$ which is equal to $\mathbf{v}$ according to the continuum approach. Hence, the integration over $\mathbf{v}$ yields

$$I_3 = \kappa P(\mathbf{w}_i) \cdot \int h_\lambda(\mathbf{x}) \cdot \mathbf{r}(\mathbf{x}) \cdot \mathbf{J}(\mathbf{x})\, d\mathbf{x} \tag{24}$$

If $h_\lambda(\mathbf{k}_i(\mathbf{r}))$ rapidly decreases to zero with increasing $\mathbf{r}$, we can replace the quantities $\mathbf{r}(\mathbf{x})$, $\mathbf{J}(\mathbf{x})$ by the first terms of their respective Taylor expansions around the point $\mathbf{x} = 0$ neglecting higher derivatives. We obtain

$$\mathbf{x}(\mathbf{r}) = \mathbf{r}(\tau_d \rho(\mathbf{w}_i))^{\frac{1}{d}} \left(1 + \frac{\mathbf{r} \cdot \partial_{\mathbf{r}} \rho(\mathbf{w}_i)}{d \cdot \rho(\mathbf{w}_i)} + \mathcal{O}\left(\mathbf{r}^2\right)\right) \tag{25}$$

which corresponds to

$$\mathbf{r}(\mathbf{x}) = \mathbf{x}(\tau_d \rho(\mathbf{w}_i))^{-\frac{1}{d}} \left(1 - (\tau_d \rho(\mathbf{w}_i))^{-\frac{1}{d}} \cdot \frac{\mathbf{x} \cdot \partial_{\mathbf{r}} \rho(\mathbf{w}_i)}{d \cdot \rho(\mathbf{w}_i)} + \mathcal{O}\left(\mathbf{x}^2\right)\right) \tag{26}$$

with $\tau_d = \pi^{\frac{d}{2}} / \Gamma\left(\frac{d}{2} + 1\right)$ as the volume of a d–dimensional unit sphere [21]. Further,

$$\mathbf{J}(\mathbf{x}) = \left(\mathbf{J}(0) + x_k \frac{\partial \mathbf{J}}{\partial x_k} + \ldots\right) \tag{27}$$

$$= (\tau_d \cdot \rho)^{-1} \left(1 - (\tau_d \cdot \rho)^{-\frac{1}{d}} \left(1 + \frac{1}{d}\right) \cdot \mathbf{x} \cdot \frac{\partial_{\mathbf{r}} \rho}{\rho}\right) + \mathcal{O}\left(x^2\right) \tag{28}$$

and, hence,

$$\left.\frac{\partial \mathbf{J}}{\partial \mathbf{x}}\right|_{\mathbf{x}=0} = -\left(\tau_d \cdot \rho\right)^{-\left(1+\frac{1}{d}\right)} \frac{\partial_{\mathbf{r}} \rho}{\rho}. \tag{29}$$

Therefore, the integral in equation (24) can be rewritten as

$$\begin{aligned} I_3 = {} & \epsilon' \kappa P \left(\tau_d \cdot \rho\right)^{-\frac{1}{d}} \int_{\mathcal{D}} h_\lambda(\mathbf{x}) \cdot \mathbf{x} \cdot \\ & \cdot \left(\left(\tau_d \cdot \rho\right)^{-1} - \left(1 + \frac{1}{d}\right) \left(\tau_d \cdot \rho\right)^{-\left(1+\frac{1}{d}\right)} \cdot \mathbf{x} \cdot \frac{\partial_{\mathbf{r}} \rho}{\rho} + \ldots\right) \\ & \cdot \left(1 - \left(\tau_d \cdot \rho\right)^{-\frac{1}{d}} \cdot \mathbf{x} \cdot \frac{\partial_{\mathbf{r}} \rho}{d \cdot \rho} + \ldots\right) \, d\mathbf{x} \end{aligned} \tag{30}$$

The integral terms in (30) of odd order in $\mathbf{x}$ vanish because of the rotational symmetry of $h_\lambda(\mathbf{x})$. Then (24) yields, neglecting terms in higher order in $\mathbf{x}$,

$$I_3 = \epsilon' \kappa P \frac{d+2}{d} \frac{\partial_{\mathbf{r}} \rho}{\rho} \tag{31}$$

with

$$\epsilon' = \frac{\epsilon_0}{\left(\tau_d \cdot \rho\right)^{\frac{2+d}{d}}} \int_{\mathcal{D}} h_\lambda(\mathbf{x}) \cdot \|\mathbf{x}\|^2 \; d\mathbf{x}. \tag{32}$$

It remains to consider the I_1-integral. As mentioned above, it is identical to the averaged adaptation of the usual NG. Hence, the treatment can be taken from there and we get

$$I_1 = \epsilon' \left(\partial_{\mathbf{r}} P - P \cdot \frac{d+2}{d} \cdot \frac{\partial_{\mathbf{r}} \rho}{\rho}\right) \tag{33}$$

as an equivalent equation [21].

Taking together (33) and (31), the stationary solution of (13) is given by

$$\langle \triangle \mathbf{w}_i \rangle = 0 = \partial_{\mathbf{r}} P - P \cdot \frac{d+2}{d} \cdot \frac{\partial_{\mathbf{r}} \rho}{\rho} + P \kappa \frac{d+2}{d} \frac{\partial_{\mathbf{r}} \rho}{\rho} \tag{34}$$

This differential equation roughly has the same form as the one for the usual Neural Gas (33). Its solution is given by

$$\rho \propto P^{\alpha_{\mathsf{WRNG}}} \tag{35}$$

with the exponent

$$\alpha_{\mathsf{WRNG}} = \frac{1}{1-\kappa}\frac{d}{d+2} \tag{36}$$

being the magnification factor. Hence, the magnification factor of the WRNG can be described also in terms of the magnification of the usual neural gas

$$\alpha_{\mathsf{WRNG}} = \frac{1}{1-\kappa}\alpha_{\mathsf{NG}} \tag{37}$$

Note, that the parameter ξ of the winner relaxing term $R(\xi, \kappa)$ does not influence the magnification.

3.3 Discussion of the theoretical result and comparison with winner relaxing SOM

Two direct observations can be immediately made: Firstly, the magnification exponent appears to be independent of the additional diagonal term (controlled by ξ) for the winner which is in agreement with the WRSOM result [8]. Therefore $\xi = 0$ again is the usual setting in WRNG for magnification control. Secondly, by adjusting κ appropriately, the magnification exponent can be adjusted, e.g. to the most interesting case of maximum mutual information [18,33]. Maximum mutual information, which corresponds to optimal information transfer, is obtained when magnification equals the unit [3,4]. Hence, we have for this case the optimum value

$$\kappa_{\mathsf{opt}} = \frac{2}{d+2}. \tag{38}$$

If the same stability borders $|\kappa| = 1$ of the WRSOM also are valid here, one can expect to increase the NG exponent by positive values of κ, or to lower the NG exponent by a factor 1/2 for $\kappa = -1$. In contrast to the Winner Enhancing SOM, where the relaxing term has to be inverted ($\kappa < 0$) to increase the magnification exponent, for the neural gas positive values of κ are required to increase the magnification exponent. However, the magnification factor still remains dependent on the generally unknown (intrinsic) dimension of the data. If this dimension is known, the parameter κ can be set *a priori* to obtain a neural gas of maximal mutual information. In this approach it is not necessary to keep track of the local reconstruction errors and firing rate for each neuron to adjust a local learning rate. Possibilities for estimating the intrinsic dimension are the well-known Grassberger-Procaccia-analysis [14] or the neural network approach using again a NG [5].

However, one has to be cautious when transferring the $\lambda \to 0$ result obtained above (which would require to increase the number of neurons as well) to a realistic situation where a decrease of λ with time will be limited to a final finite value to avoid the stability problems found in [15]. If the neighborhood length in SOM is kept small but fixed for the limit of fine discretization, the neighborhood function of the second but one winner will again be of order 1 (as for the winner). For the NG however the neighborhood is defined by the rank list. As the winner is not present in the $I_2 + I_3$ integral, all terms share the factor $\mathrm{e}^{-\lambda}$ by $h_\lambda(k) = \mathrm{e}^{-\lambda} h_\lambda(k-1)$ which indicates that in the discretized algorithm κ has to be rescaled by $\mathrm{e}^{+\lambda}$ to agree with the continuum theory. [1]

4 Numerical results

A numerical study shows how the winner-relaxing mechanism is able to control the magnification for optimization of the mutual information of a map generated by the WRNG. Using a standard setup as in [15] of $N = 50$ Neurons and 10^7 training steps with a probability density $P(x_1 \ldots x_d) = \prod_i \sin(\pi x_i)$, with fixed $\lambda = 1.5$ and ϵ decaying from 0.5 to 0.05, the entropy of the resulting map computed for an input dimension of 1, 2 and 3 is plotted in Fig. 1. Thereby, the entropy is computed using the winning probabilty p_i of the neurons:

$$H = -\sum_{i=1}^{N} p_i \ln(p_i) \tag{39}$$

The entropy shows a dimension-dependent maximum approximately at $\kappa = \frac{2}{d+2}\mathrm{e}^{\lambda}$. The scaling of the position of the entropy maximum with input dimension is in agreement with the continuum theory, as well as the prediction of the opposite sign of κ that has to be taken to increase mutual information. Our numerical investigation indicates that the above discussed prefactor, in fact, has to be taken in account for finite λ and a finite number of neurons. We obtain, within a broad range around the optimal κ the entropy is close to the maximum $\sum_{i=1}^{N} P_i \ln(P_i) = \ln(N)$ given by information theory.

In a second numerical study we investigate the influence of the additional diagonal term (controlled by ξ) for the winner. Already for the WRSOM the magnification exponent is independent of this diagonal term [8]. In the respective derviation (I_2-integral (18)) only first order approximations were

[1] In particular, for a finite λ the maximum coefficient h_λ that contributes to the $I_2 + I_3$ integral is given by the prefactor of the second but one winner, which is given by e^{λ}.

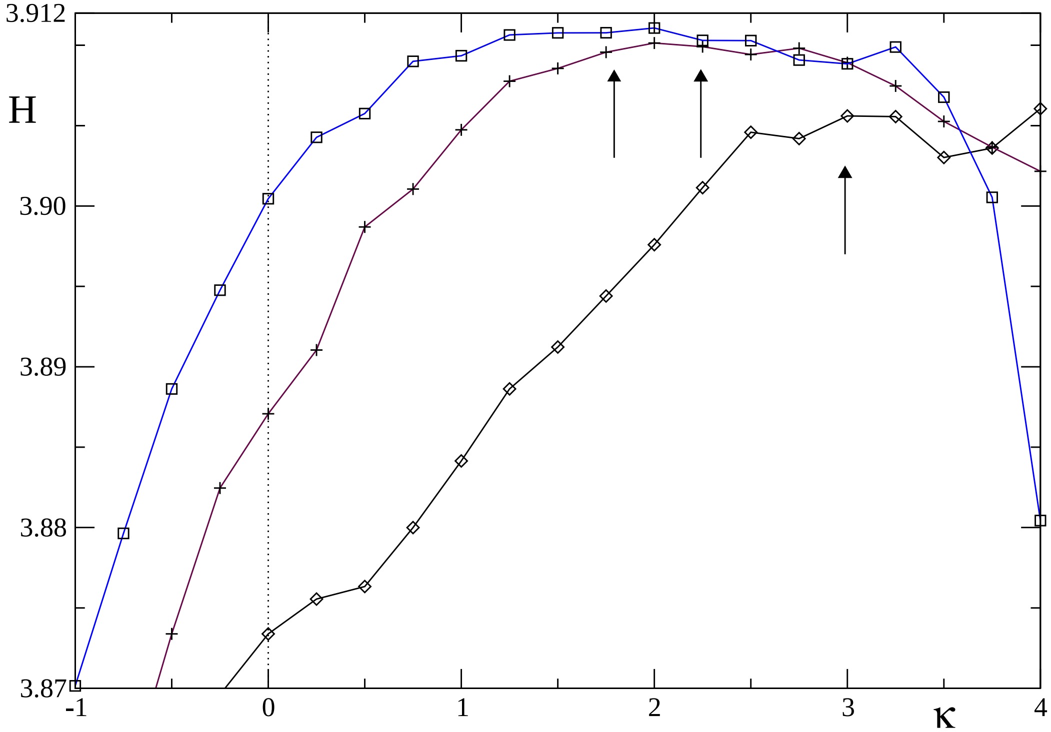

Fig. 1. Plot of the entropy H according (39) curves for varying values of κ for one- ($\diamond$), two- ($+$), and three-dimensional ($\square$) data. The entropy has the maximum $\ln(50) \simeq 3.912$ if the magnification equals unity [33]. The arrows indicate the rescaled κ_{opt}-values for the respective data dimensions.

used. Otherwise, I_2 may contribute in higher orders. To verify that the contribution of an additionally added diagonal term is marginal, the entropy was calculated both for $\xi = 0$ and $(\kappa + \xi) = 0$ [11]. However, no influence on the entropy was found for the choice $\kappa + \xi = 0$ instead of $\xi = 0$. (Fig 2). More pronounced is the influence of the diagonal term on stability; according to the larger prefactor no stable behavior has been found for $|\xi| \geq 1$, therefore $\xi = 0$ is the recommended setting.

5 Conclusions

We introduced a winner-relaxing term in neural gas algorithm to obtain a winner-relaxing neural gas with the possibility of magnification control. The winner relaxing scheme is adopted from winner-relaxing SOM. The new controlling scheme offers a method which is independent on the explicit knowledge of the generally unknown data distribution which is an advantage in comparison to the earlier presented neural gas with localized learning for magnification control. In particular, we avoid the difficult determination of the data probability density by estimation of the volume of the receptive fields of the neuron and the firing rate. Numerical simulations show the abilities of the proposed algorithm.

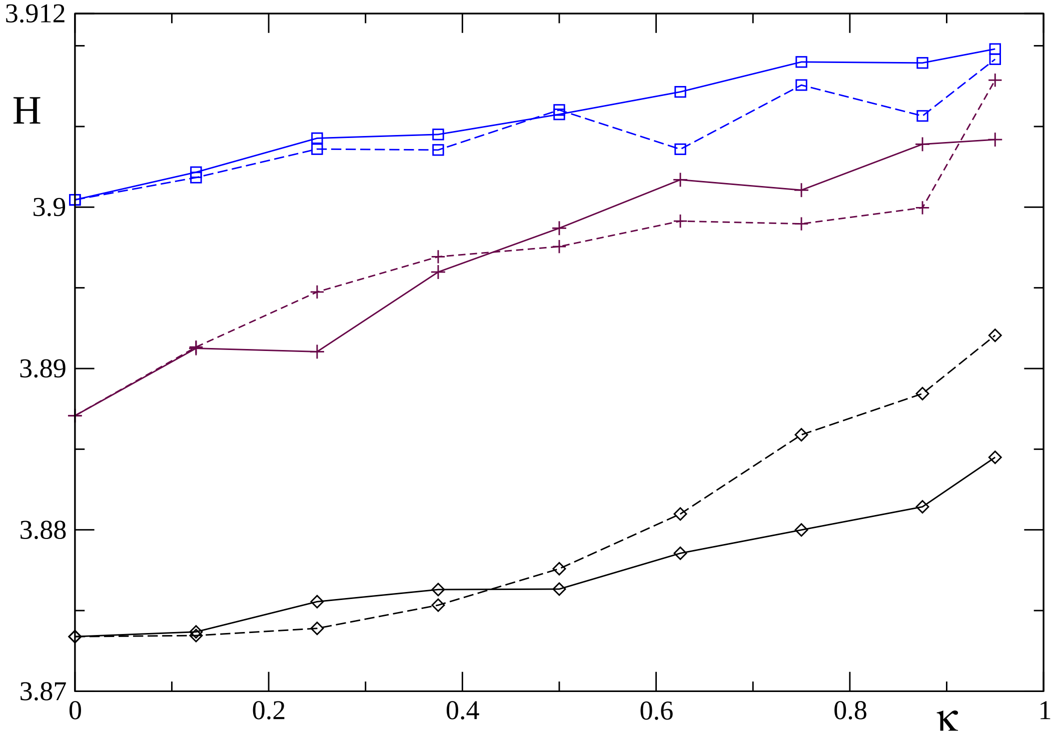

Fig. 2. Comparison of the entropies curves for different κ-values for $\xi = 0$ (straight) and $(\xi + \kappa) = 0$ (dashed) with respect to one- ($\diamond$), two- ($+$), and three-dimensional ($\square$) data.

Acknowledgements: The authors want to thank Th. Martinetz and H. Ritter for detailed comments and intensive discussions.

References

[1] H.-U. Bauer, R. Der, and M. Herrmann, Controlling the magnification factor of self–organizing feature maps, *Neural Computation* **8** (1996) 757–771.

[2] C. M. Bishop, M. Svensen, and C. K. I. Williams. Magnification factors for the SOM and GTM algorithms. In *Proceedings of WSOM'97, Workshop on Self-Organizing Maps, Espoo, Finland, June 4-6*, (Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland, 1997), 333–338.

[3] R. Brause. Optimal information distribution and performance in neighbourhood-conserving maps for robot control, *Int. J. Computers and Artificial Intelligence* **11** (1992) 173–199.

[4] R. W. Brause, An approximation network with maximal transinformation, in M. Marinaro and P. G. Morasso, editors, *Proc. ICANN'94, International Conference on Artificial Neural Networks*, volume I, (London, UK, 1994. Springer),701–704.

[5] J. Bruske and G. Sommer, Dynamic cell structure learns perfectly topology preserving map, *Neural Computation* **7** (1995) 845–65.

[6] F. Camastra and A. Colla, Data dimensionality estimation methods: a survey, *Pattern Recognition* **36** (2003) 2945–2954.

[7] F. Camastra and A. Vinciarelli, Intrinsic dimension estimation od data: an approach based Grassberger-Procaccia's algorithm, *Neural Processing Letters* **14** (2001) 27–34.

[8] J. C. Claussen, Generalized winner relaxing Kohonen feature maps, *e-print cond-mat* (2002), (http://arXiv.org/abs/cond-mat/0208414).

[9] J. C. Claussen, Winner relaxing and winner enhancing Kohonen maps: maximal mutual information from enhancing the winner, *Complexity* **8(4)** (2003) 15–22.

[10] J. C. Claussen and H. G. Schuster, Asymptotic level density of the elastic net self-organizing feature map, in J. Dorronsoro, editor, *Proc. International Conf. on Artificial Neural Networks (ICANN)* (Springer, Berlin 2002), Lecture Notes in Computer Science 2415, 939–944.

[11] J. C. Claussen and T. Villmann, Magnification control in neural gas by winner relaxing learning: Independence of a diagonal term, in O. Kaynak, editor, *Proc. International Conference on Artificial Neural Networks (ICANN'2003)* (Istanbul, 2003), 58–61.

[12] D. Dersch and P. Tavan, Asymptotic level density in topological feature maps, *IEEE Trans. on Neural Networks* **6** (1995) 230–236.

[13] D. DeSieno, Adding a conscience to competitive learning, in *Proc. ICNN'88, Int. Conf. on Neural Networks* (1988).

[14] P. Grassberger and I. Procaccia, Maesuring the strangeness of strange attractors, *Physica D* **9** (1983) 189–208.

[15] M. Herrmann and T. Villmann, Vector quantization by optimal neural gas, in W. Gerstener, A. Germond, M. Hasler, and J.-D. Nicoud, editors, *Artificial Neural Networks – Proceedings of International Conference on Artificial Neural Networks (ICANN'97) Lausanne*, (Springer, Berlin 1997), Lecture Notes in Computer Science 1327, 625–630.

[16] T. Kohonen, Self-Organizing Maps: Optimization approaches, in T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, editors, *Artificial Neural Networks*, volume II (1991), 981–990.

[17] T. Kohonen, *Self-Organizing Maps*, (Springer, 1995).

[18] R. Linsker, Towards an organizing principle for a layered perceptual network, In D. Z. Anderson, editor, *Neural Information Processing Systems*, (Amer. Inst. Phys., New York, NY, 1987), pages 485–494.

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

[20] S. P. Luttrell, Code vector density in topographic mappings: scalar case, *IEEE Trans. on Neural Networks* **2** (1991) 427–436.

[21] 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.

[22] E. Merényi and A. Jain, Forbidden magnification? ii., In M. Verleysen, editor, *European Symposium on Artificial Neural Networks 2004*, (d-side publications, 2004).

[23] H. Ritter, Asymptotic level density for a class of vector quantization processes, *IEEE Trans. on Neural Networks* **2** (1991) 173–175.

[24] H. Ritter, T. Martinetz, and K. Schulten, *Neural Computation and Self-Organizing Maps: An Introduction*, (Addison-Wesley, Reading, MA, 1992).

[25] F. Takens. On the numerical determination of the dimension of an attractor, in B. Braaksma, H. Broer, and F. Takens, editors, *Dynamical Systems and Bifurcations*, (Springer, Berlin, 1985), Lecture Notes in Mathematics No. 1125, 99–106.

[26] M. M. van Hulle, *Faithful Representations and Topographic Maps From Distortion- to Information-based Self-organization*, (J. Wiley & Sons, Inc., 2000).

[27] T. Villmann, Controlling strategies for the magnification factor in the neural gas network, *Neural Network World* **10** (2000) 739–750.

[28] T. Villmann and A. Heinze, Application of magnification control for the neural gas network in a sensorimotor architecture for robot navigation, in H.-M. Groß, K. Debes, and H.-J. Böhme, editors, *Proceedings of Selbstorganisation Von Adaptivem Verfahren (SOAVE'2000) Ilmenau*, Fortschrittsberichte des VDI (VDI-Verlag Düsseldorf, 2000), 125–134.

[29] T. Villmann and M. Herrmann, Magnification control in neural maps, in *Proc. Of European Symposium on Artificial Neural Networks (ESANN'98)*, (D facto publications. Brussels, Belgium, 1998), 191–196.

[30] T. Villmann, E. Merényi, and B. Hammer, Neural maps in remote sensing image analysis, *Neural Networks* **16** (2003) 389–403.

[31] T. Villmann, S. Schünemann, and B. Michaelis, Two approaches to control the magnification factor in the topology representing neural network, *GMD Report* 59 (1999) 44–55.

[32] A. Wismüller, F. Vietze, D. Dersch, J. Behrends, K. Hahn, and H. Ritter, The deformable feature map - a novel neurocomputing algorithm for adaptive plasticity in pattern analysis, *Neurocomputing* 48 (2002) 107–139.

[33] P. L. Zador, Asymptotic quantization error of continuous signals and the quantization dimension, *IEEE Transaction on Information Theory* **28** (1982) 149–159.