The Adaptive-Subspace Self-Organizing Map (ASSOM)

Teuvo Kohonen, Samuel Kaski, Harri Lappalainen and Jarkko Salojärvi

Helsinki University of Technology
Neural Networks Research Centre
P.O.Box 2200, FIN-02015 HUT, Finland

Abstract

A special type of a Self-Organizing Map (SOM), Adaptive-Subspace SOM, learns to identify input patters invariantly of transformations acting on them. It is demonstrated that the ASSOM can form translation-, rotation- and scale-invariant filters, and that it can be used as a learning preprocessing stage of, for instance, a texture classifier. The ASSOM can also function as a translation-invariant novelty filter.

1 Learning in the ASSOM

A long-standing goal in our research has been to find out how certain invariant-feature filters may emerge in learning processes. This problem was recently solved by one of the authors [1, 2, 3]. The key insight was that when a system is supposed to recognize input patterns invariantly of certain transformations, these transformations are actually learned from *natural sequences* of such patterns that are produced from each other by the same transformations. If the sequences are relative short, one may think that a particular transformation predominates in them, and the successive patterns then belong to some linear subspace that corresponds to this transformation. Such signal subspaces can be learned by the architecture delineated in Fig. 1.

Each dotted line in Fig. 1 distinguishes a module, a processing unit in a special SOM array. The first-layer neurons are linear and they output the sums of dot products of x with the various synaptic input weight vectors. The second-layer neurons (Q) form quadratic functions of the first-layer neuron outputs. If the weight vectors of the linear layer are orthonormalized, the neurons

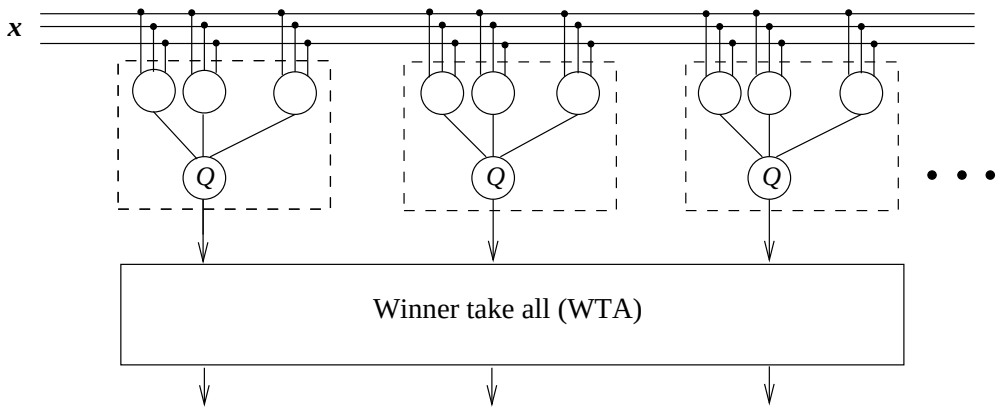


Figure 1: The ASSOM architecture.

of the output layer shall form sums of squares of their inputs. The circuit represented by Fig. 1 can then be shown to compare the input pattern x with the linear subspaces spanned by the weight vectors of the first layer. If the weight vectors can be defined in such a way that their linear combinations represent some transformation groups, then matching becomes invariant with respect to these groups. It will be shown below that such weight vectors emerge in an unsupervised learning process.

The outputs from the modules shall further be compared by a winner-take-all (WTA) function, which in Fig. 1 is shown as a separate operation. The WTA function specifies the "winner" module, indexed by c, as defined below; module c and its neighboring modules in the array will be updated in proportion to the so-called neighborhood function h_{ci} like in a usual SOM. In a physical network the WTA function may be integrated with the modules, for instance by their lateral interaction.

Let us denote the input weight vector h of module i by $b_h^{(i)}$. The $b_h^{(i)}$ of the same module are now assumed orthonormal; at least they can be orthonormalized easily. They can then be regarded as the *orthonormal basis vectors of some linear subspace* $\mathcal{L}^{(i)}$, or a set of vectors x, where every x can be expressed as a general linear combination of the $b_h^{(i)}$.

Let $\hat{x}^{(i)}$ be the *orthogonal projection* of x on $\mathcal{L}^{(i)}$, or

$$\hat{x}^{(i)} = \sum_h b_h^{(i)\mathrm{T}} x \ . \tag{1}$$

For an arbitrary x that need not belong to $\mathcal{L}^{(i)}$ one can define its *distance* d from $\mathcal{L}^{(i)}$, defined by

$$d^2 = d^2(x, \mathcal{L}^{(i)}) = \|x - \hat{x}^{(i)}\|^2 \ . \tag{2}$$

The *vectorial projection error* is the residual

$$\tilde{x}^{(i)} = x - \hat{x}^{(i)} \ . \tag{3}$$

For an arbitrary x, its *minimum projection error* can be defined as the distance of x *from the closest subspace* $\mathcal{L}^{(i)}$, and the "winner subspace" with index c is defined by

$$\|\tilde{x}^{(c)}\| = \min_i \{\|\tilde{x}^{(i)}\|\} \ , \text{ or} \tag{4}$$

$$\|\hat{x}^{(c)}\| = \max_i \{\|\hat{x}^{(i)}\|\} \ . \tag{5}$$

Our goal is to let all the modules of Fig. 1 approximate x by its different projections, and always select the module that produces the best approximation over the array. The objective function that defines the *average expected spatially weighted normalized squared projection error* is

$$E_1 = \int \sum_i h_{ci} \frac{\|\tilde{x}^{(i)}\|^2}{\|x\|^2} p(x) dx \ , \tag{6}$$

where h_{ci} is the neighborhood function that defines the interaction of modules c and i like in a usual SOM, and c is the index of the winner subspace. Notice that c is a function of x and all the basis vectors $b_h^{(i)}$.

Minimization of (6), i.e., selection of the basis vectors $b_h^{(i)}$ for all subspaces $\mathcal{L}^{(i)}$ such that the *average expected distance of* x *from the closest subspace is minimized*, is a rather complicated process [1, 2, 3]. Some extra problems are caused by the stability of the recursion by which E_1 is minimized. Without quoting all the details it may be mentioned that if the Robbins-Monro stochastic approximation process [4] is used, the optimal values of the $b_h^{(i)}$ are obtained in the recursion [5]

$$b_h^{(i)}(t+1) = b_h^{(i)}(t) + \lambda(t) h_{ci}(t) \frac{x(t) x^{\mathrm{T}}(t)}{\|x(t)\|^2} b_h^{(i)}(t) \ . \tag{7}$$

Consider now an "episode" $\mathcal{S}$ that consist of a finite set of successive sampling times t_p; denote $\mathcal{S} = \{t_p\}$. The set of samples $X = \{x(t_p)|t_p \in \mathcal{S}\}$ has to be recognized as one class, such that any member of X and even an arbitrary linear combination of the $x(t_p), t_p \in \mathcal{S}$ shall be decoded by the same module of Fig. 1 (subspace $\mathcal{L}^{(i)}$). In learning, the vector set X, defined as the Cartesian product of the $x(t_p), t_p \in \mathcal{S}$, must be taken as one batch, instead of optimizing the error using single patterns $x(t_p)$ one at a time. The error minimization problem will now be modified by defining the new objective function in terms of the *average expected spatially weighted normalized squared projection error over the episodes*:

$$E_2 = \int \sum_{t_p \in \mathcal{S}} \sum_i h_{c_r i} \frac{\|\tilde{x}^{(i)}(t_p)\|^2}{\|x(t_p)\|^2} p(X) dX \ . \tag{8}$$

Here $p(X)$ is the joint probability density for the samples $x(t_p), t_p \in \mathcal{S}$ that produce the Cartesian product set X, and dX is a shorthand notation meaning a volume differential in the Cartesian product space of the $x(t_p)$. The winning subspace c_r is now defined to be the subspace for which the sum of the projections, $\sum_{t_p \in \mathcal{S}} \|\hat{x}^{(i)}(t_p)\|^2$, is maximal.

Minimization of (8) defines the basis vectors $b_h^{(i)}$ and a set of analyzers that are *optimally invariant to the transformations that occur in the input signal patterns.*

The Robbins-Monro stochastic approximation is applicable to the minimization of E_2, too, when the gradient step is made to consist of the whole episode $\mathcal{S}$. The learning phase is then described by the following equation:

$$b_h^{(i)}(t+1) = b_h^{(i)}(t) + \lambda(t) h_{c_r i} \sum_{t_p \in \mathcal{S}(t)} \frac{x(t_p) x^{\mathrm{T}}(t_p)}{\|x(t_p)\|^2} b_h^{(i)}(t) \ . \tag{9}$$

When $\lambda(t)$ is small, (9) is equivalent with the following learning process in which the basis vectors are formed by a product of elementary projection operators, each one corresponding to one pattern $x(t_p), t_p \in \mathcal{S}(t)$:

$$b_h^{(i)}(t+1) = \prod_{t_p \in \mathcal{S}(t)} \left[I + \alpha(t_p) h_{c_r i} \frac{x(t_p) x^{\mathrm{T}}(t_p)}{\|\hat{x}^{(i)}(t_p)\| \, \|x(t_p)\|} \right] b_h^{(i)}(t) \ . \tag{10}$$

The special learning-rate factor $\lambda = \alpha(t_p)\|x(t_p)\| / \|\hat{x}^{(i)}(t_p)\|$ in (10) has been chosen for stability reasons.

There are several other minor details in the process that improve the algorithm [3, 5].

2 Experiments

We have produced various ASSOM filters for very different kinds of input data [1, 2]. Here a simple demonstration, illustrating the basic idea, is shown.

Over the input field we generated patterns consisting of colored noise (white noise, low-pass filtered by a second-order Butterworth filter with cut-off frequency of 0.6 times the Nyquist frequency of the sampling lattice). The input episodes for learning were formed by taking samples from this data field (Fig. 2). The mean of the samples was always subtracted from the pattern vector.

In the *translation-invariant filter* experiment, the episodes were formed by shifting the receptive field randomly into five nearby locations, the average shift thereby being ± 2 pixels in both dimensions. Fig. 3 shows the basis vectors b_{i1} and b_{i2}, similar to Gabor filters, in a gray scale at each array point of a two-dimensional ASSOM. One should notice that the spatial frequencies of the basis vectors of the same unit are the same, but the b_{i1} and b_{i2} are mutually 90 degrees out of phase. (The absolute phase of b_{i1} can be zero or 180 degrees, though.)

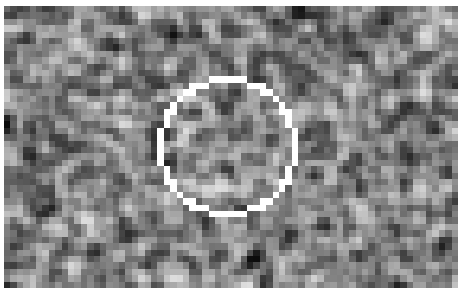
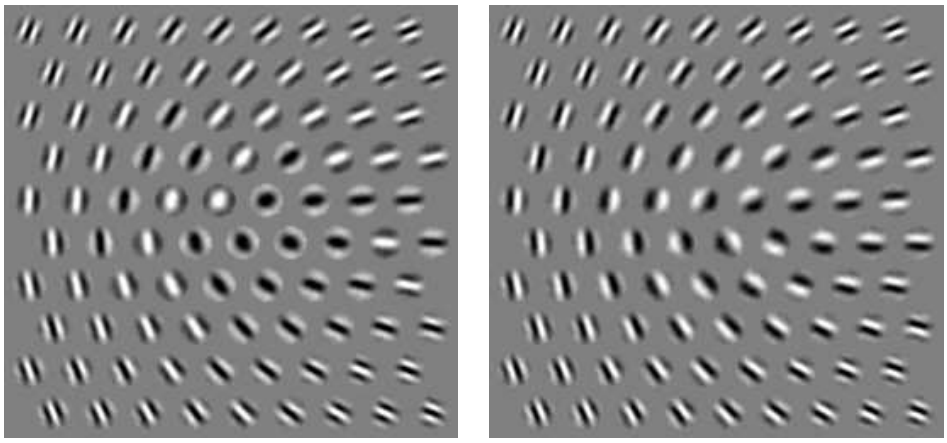


Figure 2: Colored noise (second-order Butterworth-filtered white noise with cut-off frequency of 0.6 times the Nyquist frequency of the lattice) used as input data. The receptive field is demarcated by the white circle.



(a) (b)

Figure 3: The ASSOM that has formed Gabor-type filters: (a) The b_{i1}, (b) The b_{i2}.

The episodes for the *rotation filters* were formed by rotating the input field at random five times in the range of zero to 60 degrees, the rotation center coinciding with the center of the receptive field. Fig. 4 shows the rotation filters thereby formed at the ASSOM units; clearly they are sensitive to azimuthal optic flow.

Scale-invariant filters were formed by zooming the input pattern field, with the center of the receptive field coinciding with the zooming center. The filters thereby formed, shown in Fig. 5, have clearly become sensitive to radial optic flow, corresponding to approaching or withdrawing objects.

The ASSOM can be used as a learning preprocessing stage in pattern recognition. To demonstrate this we used the outputs of an ASSOM to classify different textures with a k-nearest neighbor texture classifier. In Fig. 6 it is shown how the accuracy of the preprocessing stage increases as the feature extractor gradually adapts to the data.

The ASSOM can also form a *translation-invariant novelty filter*. When the novelty filter has learned to represent certain image types, the contribution of any of these types in a new image can be removed, thus revealing the novel aspects in the input. The original novelty filter, introduced by one of the authors [7], was not transformation-invariant, whereas the ASSOM is not sensitive to, for instance, translation of the input image.

In a demonstration of translation-invariant novelty filtering the ASSOM first learned to represent Butterworth lowpass-filtered (cut-off at 20% of the Nyquist frequency; cf. Fig. 7) white noise. The

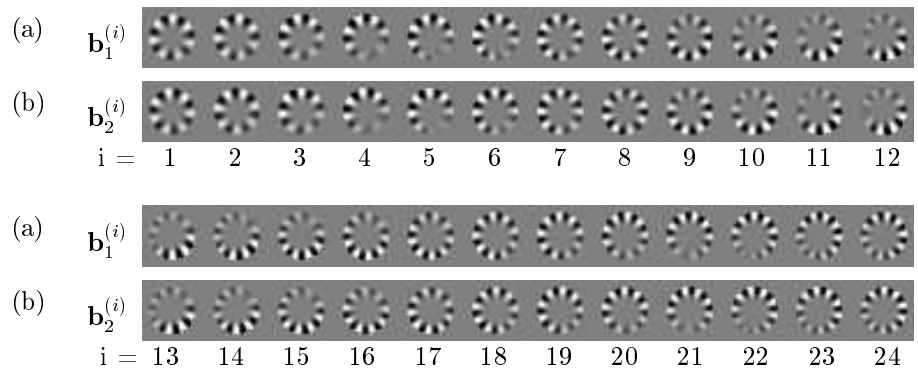


Figure 4: One-dimensional rotation-invariant ASSOM. (a) Cosine-type "azimuthal wavelets" (b_{i1}), (b) Sine-type "azimuthal wavelets" (b_{i2}). Notice that the linear array has been shown in two parts.

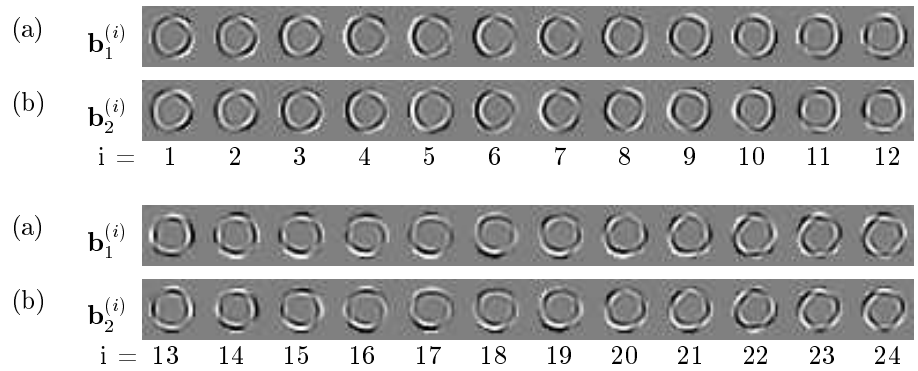


Figure 5: One-dimensional zoom-invariant ASSOM. (a) Cosine-type "radial wavelets" (b_{i1}), (b) Sine-type "radial wavelets" (b_{i2}). Notice that the linear array has been shown in two parts.

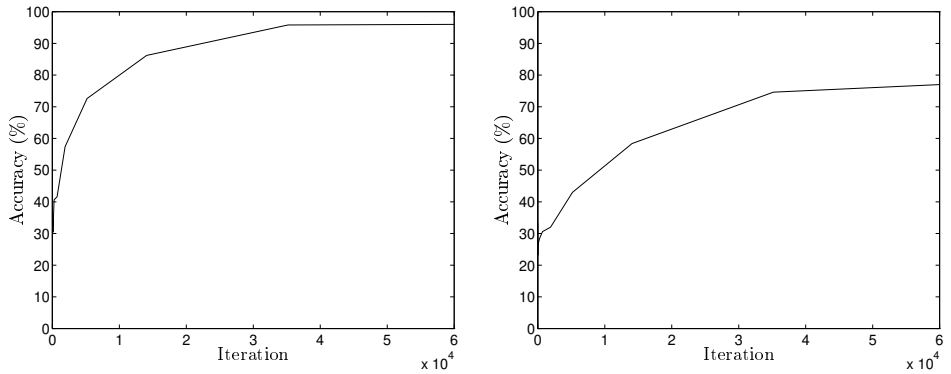


Figure 6: Learning curves of a k-nearest-neighbor texture classifier that uses ASSOM as a preprocessing stage. The 4x5-unit ASSOM was first taught using five different Brodatz textures [6], and then independent samples from these textures were classified, one hundred samples per texture. The learning curve on the left was obtained when the input patterns were translated during learning (max ± 25 pixels). On the right, the patterns were rotated (max 60 degrees).

Figure 7: Sample images used in the novelty filter experiment. On the left, samples of the colored noise used in teaching, in the middle samples of the novel part of the images, and on the right the sum of these two (the input to the novelty filter).

Figure 8: The familiar and novel portions of the input, as distinguished by the ASSOM. On the left, the projection of the input on the winning ASSOM subspace for the sample episode (compare with the leftmost texture in Fig. 7), on the right the result when the projection was subtracted from the input (compare with the texture in the middle in Fig. 7).

dimension of the subspace of the ASSOM units was 20, and the size of the ASSOM was 4 by 5 modules. During the episodes of length 6, the image was translated by a maximum of ± 10 pixels in both dimensions. The diameter of the circle-shaped receptive field was 14-pixels.

After the ASSOM had adapted to represent the input texture type, higher-frequency signal (Butterworth bandpass-filtered white noise having cut-off frequencies at 36 % and 74 %; cf. Fig. 7) was added to the inputs.

The resulting projections on the winning subspace are shown on the left in Fig. 8. The images that remain when the projections are removed are shown on the right. The results correspond fairly closely to the original images shown in Fig. 7. The ASSOM is thus capable of functioning as a *translation-invariant novelty filter*.

References

[1] T. Kohonen. *Self-Organizing Maps*. Springer Series in Information Sciences, vol. 30. Springer, Heidelberg, 1995.

[2] T. Kohonen. Emergence of invariant-feature detectors in self-organization. In Palaniswami, M. et al. (eds.), *Computational Intelligence, A Dynamic System Perspective*, pp. 17–31. IEEE Press, New York, 1995.

[3] T. Kohonen. Emergence of invariant-feature detectors in the adaptive-subspace self-organizing map. *Biological Cybernetics*, 75:281-291, 1996.

[4] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statist.*, 22:400–407, 1951.

[5] T. Kohonen, S. Kaski, and H. Lappalainen. Self-organized formation of various invariant-feature filters in the adaptive-subspace SOM. To appear in *Neural Computation*, 1997.

[6] P. Brodatz. *Textures - A Photographic Album for Artists and Designers*. New York : Reinhold, 1968.

[7] T. Kohonen. *Self-Organization and Associative Memory*. Springer Series in Information Sciences, Springer, Heidelberg, 1984.