Analytical Description of Boundary Effect in the 1D Kohonen Scheme for Constructing Adaptive Grids

A.V.Voytishek, D.S.Khmel
Institute of Computational Mathematics and Mathematical Geophysics of Siberian Branch of Russian Academy of Sciences
Prospect Akademika Lavrentyeva, 6, Novosibirsk, 630090 Russia
vav@osmf.sscc.ru

***Abstract*- Kohonen self-organizing maps (SOM) have a lot of fruitful applications. In the classical monograph *Kohonen T., Self-Organizing Maps (Third edition), New York: Springer-Verlag, 2001* the following fields of SOM applications are presented: signal processing, control theory, models of the biological brain function, experimental physics, chemistry and medicine, financial analysis, etc.**

One of the main applications of SOM is "automated" stochastic iterative numerical algorithm for constructing adaptive grids. Moreover, this algorithm can be treated as the most natural mathematical model of SOM.

The algorithm starts with introducing arbitrary initial positions of points (nodes) inside some domain. On every iterative step a sample value of the stochastic variable, which is distributed with respect to given probability distribution density (this density defines the demanded arrangement of nodes in the domain), is realized numerically. The closest ("winner") point to this sample value defines the learning coefficient (or neighborhood function) which influences on shift of every node. The special choice of the learning coefficient allows getting satisfactory arrangement of points after several iterations.

There are difficulties in analytical description of the Kohonen scheme. This description is rather useful for theoretical investigation of the self-organizing algorithm (convergence, estimating of errors, etc.). In monograph the mentioned above, T.Kohonen suggested the following "continuous" approach. Under some principal simplifications (1D case, simplified form of the learning coefficient, assumption about ordering of the initial distribution of nodes, uniform distribution for "attraction nodes") the recurrent formulas for node's shifts are replaced by the system of differential equations for "most probable" positions of nodes. This system has no general analytical solution, and only numerical experiments can be used for investigation of asymptotic positions of nodes.

In this paper we have proposed direct use of the formulas for node's shift to get analytical recurrent expressions for the most probable positions of nodes under the Kohonen's simplifications. We also showed that our approach helps to investigate some special effects of the self-organizing algorithm, in particular, the "boundary effect" which defines undesirable noticeable distances between the boundary nodes and the boundary of the domain. In addition we considered the possibilities for weakening of Kohonen's restrictions: in particular, we have constructed recurrent formulas for special practical learning coefficient.

Keywords- Kohonen Self-Organizing Maps; Recurrent Formulas; Boundary Effect; Learning Radius

I. INTRODUCTION

Self-organizing maps (SOM) have a lot of fruitful applications: signal processing, control theory, models of the biological brain function, experimental physics, chemistry and medicine, financial analysis, etc (see the monograph [1] of T. Kohonen and references in this monograph).

One of the main applications of SOM is "automated" stochastic iterative numerical algorithm for constructing adaptive grids [1]. Moreover, this algorithm can be treated as the most natural mathematical model of SOM. This model can be considered as a segment of vector Markov chain, where states are the groups of nodes $\{\boldsymbol{x}_1(t), \dots, \boldsymbol{x}_M(t)\}$ in some domain $X \subset R_X^d$.

There are many examples of computer calculations which prove the effectiveness of Kohonen scheme [1, 2]. This algorithm, however, induces some problems during its implementation. One of them is the "boundary effect" which defines undesirable noticeable distances between the boundary nodes and the boundary ∂X of the domain X [1, 2].

There are also difficulties in analytical description of the Kohonen scheme (this description is rather useful for theoretical investigation of the self-organizing algorithm– convergence, estimating of errors, etc.). Constructive approaches for this description were elaborated only for the simplest one-dimensional (1D) case [1, 3]. The Kohonen's "continuous" approach from the monograph [1] is presented in the Section III of the present paper. This approach leads to a system of differential equations for the "most probable" mean asymptotic positions of nodes $\{\boldsymbol{x}_1(t), \dots, \boldsymbol{x}_M(t)\}$. This system has no general analytical solution, and only numerical experiments can be used for investigation of asymptotic positions of nodes. Section IV demonstrates that recurrent formulas for the most probable mean positions of grid nodes for the first iterations of the algorithm can be derived for the simplified case considered by Kohonen [1]. As implementations of the algorithm in practice imply a comparatively small number of iterations, the use of such an analytical description seems to be more promising than the "asymptotic" approach from the monograph [1]. It is noted in Section V that our approach allows also investigating analytically the 1D boundary effect and modifications of 1D Kohonen scheme which lead to decrease of this effect. Finally, the main results of this work are shortly presented in the Section "Conclusions".

II. BASIC RANDOMIZED ALGORITHM

Let us formulate the problem as follows (see also [1, 2]). In a "physical" domain X (or on its surface ∂X), we have to

construct a grid $X^{(M)} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_M\}$ with the distribution of nodes corresponding to a specified density $f(\boldsymbol{x})$, $\boldsymbol{x} \in R_X^d$. The structure of this grid (order and structure of disposition of nodes) is determined by a "map" $Q^{(M)} = \{\boldsymbol{q}_1, \dots, \boldsymbol{q}_M\}$ and by a system of the "neurons" $E^{(M)} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_M\}$ (where $\boldsymbol{e}_i = (\boldsymbol{q}_i, \boldsymbol{x}_i)$) determining the correspondence between the grids $X^{(M)}$ and $Q^{(M)}$.

The approximation of neurons is performed with a self-learning procedure, which is an iterative process based on consecutive formation of a learning set $\Xi^{(T)} = \{\boldsymbol{\xi}_0(1), \dots, \boldsymbol{\xi}_0(T)\}$ in the form of a sampling from the probability distribution of a random vector $\boldsymbol{\xi}$ having a density $f(\boldsymbol{x})$; here, T is the number of iterations and $\boldsymbol{\xi}_0(t) \in X$ (or $\boldsymbol{\xi}_0(t) \in \partial X$, $t = 1, \dots, T$). In addition, lateral relations between the neurons $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ are established at each iteration step by using special "learning coefficient" $\theta_{\boldsymbol{q}_m}(t, \boldsymbol{q}_i) \in [0,1)$. This procedure yields a sequence of grids $X^{(M)}(t) = \{\boldsymbol{x}_1(t), \dots, \boldsymbol{x}_M(t)\}$; in this case, the following relation has to be satisfied:

$$X^{(M)} \approx \tilde{X}^{(M)} = \lim_{t\to\infty} X^{(M)}(t) \approx X^{(M)}(T). \quad (1)$$

The approximate equality signs in Equation (1) mean not only that the distances $\rho(\boldsymbol{x}_i(\infty), \boldsymbol{x}_i)$, $\rho(\boldsymbol{x}_i(\infty), \boldsymbol{x}_i(T))$, and $\rho(\boldsymbol{x}_i(T), \boldsymbol{x}_i)$, where

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(x^{(1)} - y^{(1)})^2 + \dots + (x^{(d)} - y^{(d)})^2},$$
$$\boldsymbol{x} = (x^{(1)}, \dots, x^{(d)}), \boldsymbol{y} = (y^{(1)}, \dots, y^{(d)})$$

are small, but also that the required properties of the grid $X^{(M)}$ (e.g., the properties of rectangularity, the absence of the boundary effect, etc.) in implementation of the next iterative process are accurately reproduced.

Algorithm [1, 2]. 1. The initial positions of the grid nodes $\mathrm{X}^{(\mathrm{M})}(0) = \{\mathbf{x}_1(0), \dots, \mathbf{x}_\mathrm{M}(0)\}$ are established. 2.The following actions are performed at each iteration with the number $\mathrm{t} = 1, \dots, \mathrm{T}$:

(a) the next element $\boldsymbol{\xi}_0(\mathrm{t})$ of the sampling $\Xi^{(\mathrm{T})}$ is chosen;

(b) the distances $\rho\big(\boldsymbol{\xi}_0(\mathrm{t}),\ \mathbf{x}_\mathrm{i}(\mathrm{t}-1)\big)$ from the point $\boldsymbol{\xi}_0(\mathrm{t})$ to all nodes $\mathbf{x}_\mathrm{i}(\mathrm{t}-1)$ are calculated, and the nearest node $\mathbf{x}_\mathrm{m}(\mathrm{t}-1)$ to $\boldsymbol{\xi}_0(\mathrm{t})$ is chosen in accordance with the condition

$$\mathrm{m} = \arg\min_{\mathrm{i}=1,\dots,\mathrm{M}} \rho\big(\boldsymbol{\xi}_0(\mathrm{t}),\ \mathbf{x}_\mathrm{i}(\mathrm{t}-1)\big)$$

such a node $\mathbf{x}_\mathrm{m}(\mathrm{t}-1)$ is called a **series winner**;

(c) the positions of all nodes are corrected in accordance with the formula

$$\boldsymbol{x}_i(t) = \boldsymbol{x}_i(t-1) + \theta_{\boldsymbol{q}_m}(t, \boldsymbol{q}_i)\big(\boldsymbol{\xi}_0(t) - \boldsymbol{x}_i(t-1)\big) \quad (2)$$

for all $\mathrm{i} = 1, \dots, \mathrm{M}$; here $\theta_{\mathrm{q_m}}(\mathrm{t}, \mathrm{q_i})$ is a learning coefficient.

At each iteration of the algorithm, the grid nodes are shifted towards the random point $\boldsymbol{\xi}_0(t)$. Therefore, more and more nodes are accumulated in areas with high concentrations of sampling elements, resulting in grid refinement. It was demonstrated [2] that, as $T \to \infty$, the algorithm leads to satisfaction of the analog of the equidistribution principle, i.e., to obtaining of a required grid density determined by the function $f(\boldsymbol{x})$.

III. "CONTINUOUS" KOHONEN'S FORMALISM

Sequence (2) is random. Attempts to obtain analytical expressions even for the simplest characteristics, such as mathematical expectations $\mathbf{E}\boldsymbol{x}_i$, face significant difficulties. This fact was noted in [1, Section 3.5]. Simplifications that allowed analytical approaches to study the algorithm to be developed were also proposed there. Let us briefly describe these simplifications. Let $X = R$ (i.e., a 1D case is considered), and also $\rho(x, y) = |x - y|$. In addition, we assume that learning coefficient is equal to

$$\theta_{\boldsymbol{q}_m}(t, \boldsymbol{q}_i) \equiv \theta(t) \text{ for } i \in \{\max(1, m-1), m, \min(M, m+1)\},$$
$$\theta_{\boldsymbol{q}_m}(t, \boldsymbol{q}_i) = 0 \text{ otherwise}, \quad (3)$$

which means that no more than three nodes (the winner node and the nearest nodes "on the left" and "on the right" in terms of their numbers) are involved into the learning process, and the corresponding learning coefficient is independent of the position of the winner node. Further, Kohonen [1] proposed to replace the Relation (2) by a "continuous analog":

$$d\tilde{x}_i(\tilde{t})/\,d\tilde{t} = \theta(\tilde{t})(\xi - \tilde{x}_i(\tilde{t})) \text{ for } i \in \{\max(1, m-1), m, \min(M, m+1)\}$$
$$d\tilde{x}_i(\tilde{t})/\,d\tilde{t} = 0 \text{ otherwise.} \quad (4)$$

Further, Kohonen [1] formulated a statement that, for arbitrary initial positions of the points $\{\tilde{x}_i(0)\}$ and for $\theta(\tilde{t}) \equiv \mathrm{const}$ there exists a time instant $T > 0$ for Process (4), such that

$$\tilde{x}_1(\tilde{t}) \le \tilde{x}_2(\tilde{t}) \le \dots \le \tilde{x}_M(\tilde{t}) \quad (5)$$

for all $\tilde{t} > T$, i.e., the points $\{\tilde{x}_i(\tilde{t})\}$ become ordered, beginning from the time T, and this order persists with increasing $\tilde{t}$; in addition, the distribution density of the points $\{\tilde{x}_i(\tilde{t})\}$ in the limit approximates the monotonic function of the distribution density $f(x)$ of the random quantity ξ.

In this paper, we do not discuss the process of ordering of the points $\{\tilde{x}_i(\tilde{t})\}$ in detail, assuming that the following relation is valid already at the initial time:

$$-\infty < a \le x_1(0) \le x_2(0) \dots \le x_M(0) \le b \le \infty. \quad (6)$$

It was further noted in [1] that Relation (5) is valid for all $\tilde{t} > 0$ if Condition (6) is satisfied. Assuming that the distribution of the random quantity ξ is concentrated on the segment $[a, b]$ (this means that $f(x) > 0$ for $x \in [a, b]$ and $f(x) = 0$ for $x \notin [a, b]$) we can introduce the intervals

$$\begin{cases} S_i(\tilde{t}) = \left[\frac{1}{2}\left(\tilde{x}_{i-2}(\tilde{t}) + \tilde{x}_{i-1}(\tilde{t})\right),\ \frac{1}{2}\left(\tilde{x}_{i+1}(\tilde{t}) + \tilde{x}_{i+2}(\tilde{t})\right)\right], & 3 \le i \le M-2 \\ S_1(\tilde{t}) = \left[a,\ \frac{1}{2}\left(\tilde{x}_2(\tilde{t}) + \tilde{x}_3(\tilde{t})\right)\right], \\ S_2(\tilde{t}) = \left[a,\ \frac{1}{2}\left(\tilde{x}_3(\tilde{t}) + \tilde{x}_4(\tilde{t})\right)\right], \\ S_{M-1}(\tilde{t}) = \left[\frac{1}{2}\left(\tilde{x}_{M-3}(\tilde{t}) + \tilde{x}_{M-2}(\tilde{t})\right),\ b\right], \\ S_M(\tilde{t}) = \left[\frac{1}{2}\left(\tilde{x}_{M-2}(\tilde{t}) + \tilde{x}_{M-1}(\tilde{t})\right),\ b\right]. \end{cases} \quad (7)$$

for fixed chosen values of $\{\tilde{x}_i(\tilde{t})\}$ and consider conditional (at fixed $\{\tilde{x}_j(\tilde{t})\}$) mathematical expectations equal (with allowance for the Equation (4)) to

$$\langle \dot{\tilde{x}}_i(\tilde{t}) \rangle \stackrel{\text{def}}{=} \mathbf{E}\left(\dot{\tilde{x}}_i(\tilde{t}) \mid \{\tilde{x}_i(\tilde{t})\}\right) = \theta(\tilde{t})\left(\mathbf{E}\left(\xi(\tilde{t}) \middle| \xi(\tilde{t}) \in S_i(\tilde{t})\right) - \tilde{x}_i(\tilde{t})\right)\mathbf{P}(\xi(\tilde{t}) \in S_i, \quad (8)$$

where $\mathbf{P}(\xi(\tilde{t}) \in S_i(\tilde{t}))$ is the probability of the situation where $\xi(\tilde{t})$ belongs to $S_i(\tilde{t})$. The calculation of the two last factors in Relation (8) presents a certain problem. Kohonen noted [1] that this problem can be definitely resolved for the uniform distribution density

$$f(x) \equiv 1/(b-a) \quad \text{for} \quad x \in [a,b] \quad \text{and} \quad f(x) = 0 \quad \text{for} \quad x \notin [a,b]. \quad (9)$$

This corresponds to uniformly distributed positions of nodes in the computational domain $Q = X = [a,b]$; here, $Q^{(M)} = \{x_i = (i-1)h;\ i = 1, \dots, M;\ h = (b-a)/(M-1)$. In this case, the following relations are valid:

$$\begin{cases} \mathbf{E}\left(\xi(\tilde{t}) \middle| \xi(\tilde{t}) \in S_i(\tilde{t})\right) = \frac{1}{4}\left(\tilde{x}_{i-2}(\tilde{t}) + \tilde{x}_{i-1}(\tilde{t}) + \tilde{x}_{i+1}(\tilde{t}) + \tilde{x}_{i+2}(\tilde{t})\right), & 3 \le i \le M-2 \\ \mathbf{E}\left(\xi(\tilde{t}) \middle| \xi(\tilde{t}) \in S_1(\tilde{t})\right) = \frac{1}{4}\left(2a + \tilde{x}_2(\tilde{t}) + \tilde{x}_3(\tilde{t})\right), \\ \mathbf{E}\left(\xi(\tilde{t}) \middle| \xi(\tilde{t}) \in S_2(\tilde{t})\right) = \frac{1}{4}\left(2a + \tilde{x}_3(\tilde{t}) + \tilde{x}_4(\tilde{t})\right), \\ \mathbf{E}\left(\xi(\tilde{t}) \middle| \xi(\tilde{t}) \in S_{M-1}(\tilde{t})\right) = \frac{1}{4}(\tilde{x}_{M-3}(\tilde{t}) + \tilde{x}_{M-2}(\tilde{t}) + 2b), \\ \mathbf{E}\left(\xi(\tilde{t}) \middle| \xi(\tilde{t}) \in S_M(\tilde{t})\right) = \frac{1}{4}(\tilde{x}_{M-2}(\tilde{t}) + \tilde{x}_{M-1}(\tilde{t}) + 2b). \end{cases} \quad (10)$$

$$\begin{cases} \mathbf{P}(\xi(\tilde{t}) \in S_i(\tilde{t})) = \frac{1}{2(b-a)}\left(\tilde{x}_{i+1}(\tilde{t}) + \tilde{x}_{i+2}(\tilde{t}) - \tilde{x}_{i-1}(\tilde{t}) - \tilde{x}_{i-2}(\tilde{t})\right), & 3 \le i \le M-2 \\ \mathbf{P}(\xi(\tilde{t}) \in S_1(\tilde{t})) = \frac{1}{2(b-a)}(\tilde{x}_2(\tilde{t}) + \tilde{x}_3(\tilde{t}) - 2a), \\ \mathbf{P}\left(\xi(\tilde{t}) \in S_2(\tilde{t})\right) = \frac{1}{2(b-a)}(\tilde{x}_3(\tilde{t}) + \tilde{x}_4(\tilde{t}) - 2a), \\ \mathbf{P}\left(\xi(\tilde{t}) \in S_{M-1}(\tilde{t})\right) = \frac{1}{2(b-a)}\left(2b - \tilde{x}_{M-3}(\tilde{t}) - \tilde{x}_{M-2}(\tilde{t})\right), \\ \mathbf{P}\left(\xi(\tilde{t}) \in S_M(\tilde{t})\right) = \frac{1}{2(b-a)}\left(2b - \tilde{x}_{M-2}(\tilde{t}) - \tilde{x}_{M-1}(\tilde{t})\right). \end{cases} \quad (11)$$

Based on these relations, Kohonen [1] formulated a statement that, for an arbitrary set of initial values (6), the most probable mean values $\hat{x}_i(\tilde{t}) = \mathbf{E}\tilde{x}_i(\tilde{t})$ satisfy the system of differential equations

$$\frac{d\boldsymbol{z}}{dt} = \Pi(\boldsymbol{z})(F\boldsymbol{z} + \boldsymbol{h}) \quad (12)$$

where $\boldsymbol{z} = [\hat{x}_1(\tilde{t}), \hat{x}_2(\tilde{t}), \dots, \hat{x}_M(\tilde{t})]^{\mathbf{T}}$ (here, $\mathbf{T}$ is the sign of transposition), matrix F is equal to

$$\frac{\theta(\tilde{t})}{4}\begin{pmatrix} -4 & 1 & 1 & 0 & 0 & 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & -4 & 1 & 1 & 0 & 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & -4 & 1 & 1 & 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 1 & 1 & -4 & 1 & 1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & 1 & 1 & -4 & 1 & 1 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & 0 & 1 & 1 & -4 & 1 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & 0 & 0 & 1 & 1 & -4 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & 0 & 0 & 0 & 1 & 1 & -4 \end{pmatrix}$$

$\boldsymbol{h} = \frac{\theta(\tilde{t})}{4}[a, a, 0, 0, \dots, 0, b, b]^{\mathrm{T}}$ and $\Pi(\boldsymbol{z})$ is diagonal matrix with the values on the diagonal having the Form (11) (with accuracy to replacement of $\tilde{x}_i(\tilde{t})$ by $\hat{x}_i(\tilde{t})$. Relation (12) is a system of ordinary differential equations of the first order with constant coefficients.

It was noted [1] that it is rather difficult to find the general analytical solution for this system. Taking into account that the matrix $\Pi(\boldsymbol{z})$ is a diagonal, positively determined matrix, it is possible to demonstrate that a fundamental solution exists at $d\boldsymbol{z}/dt = 0$ and has the form $\boldsymbol{z}_0 = -F^{-1}\boldsymbol{h}$ (the proof of the latter fact is rather cumbersome); there are also some ideas on the existence of an asymptotic solution of System (12) [1].

Mentioning the difficulties in finding analytical expressions for the most probable mean values of $\{\hat{x}_i(\tilde{t})\}$, Kohonen proposed to study the asymptotic positions of nodes numerically. In particular, he has proved numerically the existence of the boundary effect (noticeable distances between the boundary nodes and the segment boundaries) and noted this effect becomes less pronounced with increasing M.

IV. RECURRENT FORMULAS

Direct averaging of the Equation (2) yields the equality [3]:

$$\mathbf{E}_{(t)}\boldsymbol{x}_i(t) = \mathbf{E}_{(t-1)}\boldsymbol{x}_i(t-1) + \mathbf{E}_{(t-1,\boldsymbol{\xi}(t))}\theta_{\boldsymbol{q}_m}(t, \boldsymbol{q}_i)\big(\boldsymbol{\xi}(t) - \boldsymbol{x}_i(t-1);\tag{13}$$

here, the index "$(t-1)$" means taking the mathematical expectation over the distribution of the set of grids $\{X^{(M)}(0), X^{(M)}(1), \dots, X^{(M)}(t-1)\}$; addition of $\boldsymbol{\xi}(t)$ in this index means additional averaging over the distribution with the density $f(\boldsymbol{x})$ at the t-th step of the algorithm (see Section 1).

The calculation of the second term in the right-hand side of Equation (13) is particularly difficult. Nevertheless, using the simplifications of Kohonen's formalism (see Section 2):

1) 1D case: $[a, b] = [0,1]$,
2) simplified form of the learning Coefficient (3),
3) assumption about ordering of the initial distribution of Nodes (6),
4) uniform distribution for "Attraction Nodes"(9),
5) replacement of conditional values by "most probable mean" values,

we can obtain an analog of the Equation (8), which allows consecutive obtaining of exact values of the most probable mean positions of nodes ($t = 1,2, \dots$),

$$\langle x_i(t)\rangle = \langle x_i(t-1)\rangle + \theta(t)\Big(\mathbf{E}_i^{(\xi)}(t-1) - \langle x_i(t-1)\rangle\Big)\mathbf{P}_i^{(\xi)}(t-1),\tag{14}$$

where

$$\mathbf{P}_i^{(\xi)}(t-1) = s_i^{(+)}(t-1) - s_i^{(-)}(t-1),$$
$$\mathbf{E}_i^{(\xi)}(t-1) = \frac{s_i^{(+)}(t-1) - s_i^{(-)}(t-1)}{2}\tag{15}$$

and $s_i^{(-)}(t-1)$ and $s_i^{(+)}(t-1)$ are the left and right ends of the segment $S_i(t-1)$ determined by the Equation (7). Note that the formulas in System (15) are analogs of Relations (10) and (11) for the case with $a = 0$ and $b = 1$. The accuracy of the resultant mean positions of nodes was verified through Monte Carlo computations in the Paper [3]. Examples of these results are shown on the Figures 1 and 2. We can fix the convergence of Monte Carlo calculations to the Values (13)-(15) when $n \to \infty$ (here n is the number of realizations of the main algorithm). We can also note that expensive Monte Carlo calculations can be replaced by formulas of the (13)-(15) type.

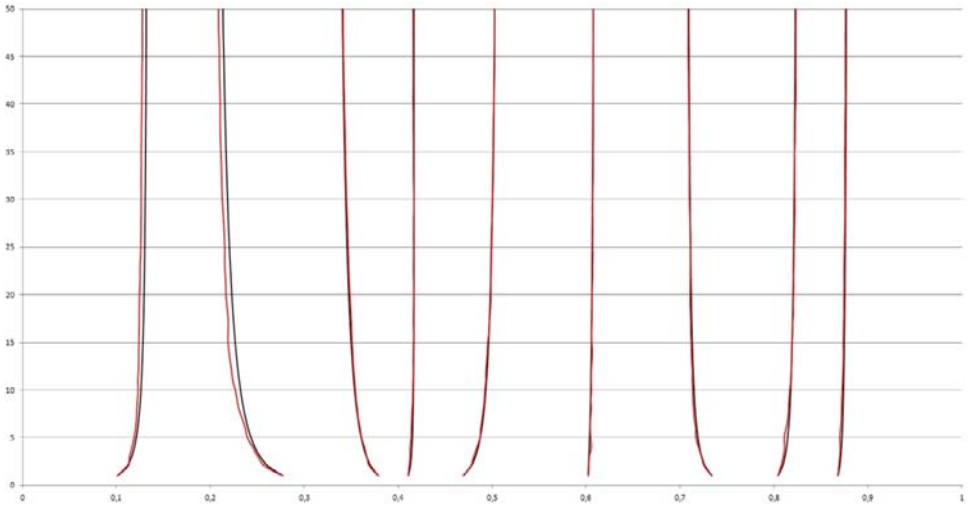

Figure 1 Iterative nodes moving: *grey* – formulas (13)-(15), *red* – Monte Carlo computations ($n = 100$)

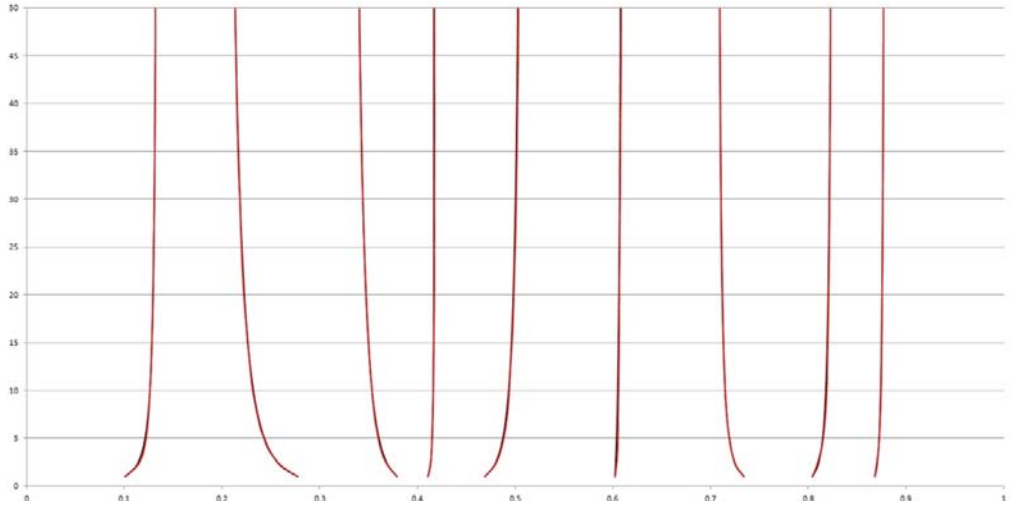

Figure 2 Iterative nodes moving: *grey* – formulas (13)-(15), *red* – Monte Carlo computations ($n = 1000$)

Let's analyze what Kohonen's restrictions *A*-*E* for deriving formulas of the (13)-(15) type can be weakened. Transition to multidimensional (in particular, 2D) case (restriction *A*) seems to be rather difficult because of appearance of additional geometrical requirements for adaptive grids. It is not difficult to get the analytical formulas for various probability 1D-distributions (for which $f(x) \neq \text{const}$ and $[a, b] \neq [0,1]$) – see Restrictions *A* and *D*, but this formulas will be not so universal. The analytical investigation of the ordering Procedure (5) (Restriction *C*) has its own difficulties, the special mathematical approaches are needed here (may be induction on number of grid nodes M). It is also possible to consider another form of learning coefficient (Restriction *B* and Formula (3)) – see next Section V of this paper.

V. LEARNING RADIUS AS THE MEASURE OF BOUNDARY EFFECT

In fundamental works [1, 2] it was shown that for effective realization of learning algorithm (see Section 1) it is expedient to choose the learning coefficient as a multiplication of two functions:

$$\theta_{\boldsymbol{q}_m}(t, \boldsymbol{q}_i) = \delta(t) \times \tau_{\boldsymbol{q}_m}(t, \boldsymbol{q}_i). \quad (16)$$

The function $\delta(t)$ influences on *learning step*, which defines the changing values of nodes positions. Numerical experiments [2] lead to choose this function according the form

$$\delta(t) = t^{-0.2}\omega(t),\ t = 1, \dots, T, \text{ where } \omega(t) = 1 - e^{5(t-T)/T}, \quad (17)$$

The function $\tau_{\boldsymbol{q}_m}(t, \boldsymbol{q}_i)$ influences on *learning radius* $r(t)$ [2]:

$$\tau_{\boldsymbol{q}_m}(t, \boldsymbol{q}_i) = s^{\left(\frac{d(\boldsymbol{q}_m, \boldsymbol{q}_i)}{r(t)}\right)^2}, \quad (18)$$

where $s \in (0,1)$ is a fixed small constant; for our computations we chose $s = 10^{-5}$.

There are no difficulties to construct a formula of the (13)-(15) type for the learning coefficient of the Form (16)-(18). This formula becomes more cumbersome: it includes sum with not three but $(M + 1)$ addendums

$$\langle x_i(t)\rangle = \langle x_i(t-1)\rangle + t^{-0.2} \times \left(1 - e^{\frac{5(t-T)}{T}}\right) \times \left(\frac{\langle x_1(t-1)\rangle + \langle x_2(t-1)\rangle}{2} \times s^{\left(\frac{i-1}{r(t)(M+1)}\right)^2} \times \right.$$

$$\times \frac{2\langle x_1(t-1)\rangle + \langle x_2(t-1)\rangle - 4\langle x_i(t-1)\rangle}{4} + \frac{2 - \langle x_{M-1}(t-1)\rangle - \langle x_M(t-1)\rangle}{2} \times$$

$$\times s^{\left(\frac{M-i}{r(t)(M+1)}\right)^2} \times \frac{\langle x_{M-1}(t-1)\rangle + 2\langle x_M(t-1)\rangle + 1 - 4\langle x_i(t-1)\rangle}{4} +$$

$$+ \sum_{j=2}^{M-1} \left(\frac{\langle x_{j+1}(t-1)\rangle - \langle x_{j-1}(t-1)\rangle}{2} \times s^{\left(\frac{j-i}{r(t)(M+1)}\right)^2} \times \right.$$

$$\left. \times \frac{\langle x_{j-1}(t-1)\rangle + 2\langle x_j(t-1)\rangle + \langle x_{j+1}(t-1)\rangle - 4\langle x_i(t-1)\rangle}{4}, \right. \quad (19)$$

where $\langle x_0(t-1)\rangle = 0$ and $\langle x_{M+1}(t-1)\rangle = 1$. The multiplier $t^{-0.2} \times \left(1 - e^{\frac{5(t-T)}{T}}\right)$ from the first line of this

formula equals to the function $\delta(t)$ of the Form (17). The first two terms (first-third lines of the Formula (17)) define the contribution to the shift of the node's mean $\langle x_i(t)\rangle$ with respect to boundary nodes – the first and the M-th ones (existence of this terms in many respects defines the boundary effect). In formation of this expressions it is considered that the left boundary of the first interval $S_1(t)$ is equal to the point $x = 0$ (for "internal" numbers $j = 2, \dots, M-1$ the left boundaries of the intervals $S_j(t)$ are equal to the values $\langle x_{j-1}(t-1)\rangle$) and the right boundary for the M-th interval $S_M(t)$ is equal to $x = 1$ (for "internal" numbers the right boundaries of the intervals $S_j(t)$ are equal to the values $\langle x_j(t-1)\rangle$). In every one of $(M+1)$ terms the first multiplier of the form $(\langle x_{j+1}(t-1)\rangle - \langle x_{j-1}(t-1)\rangle)/2$ is equal to probability for belonging of the "attraction point" $\xi(t)$ (uniformly distributed in the interval $(0,1)$ to the corresponding interval $S_j(t)$. The second multiplier $s^{\left(\frac{j-i}{r(t)(M+1)}\right)^2}$ is equal to the function $\tau_{\boldsymbol{q}_m}(t, \boldsymbol{q}_i)$ of the Form (18); here $d(\boldsymbol{q}_j, \boldsymbol{q}_i) = |j-i|/(M+1)$. At last, the third multiplier of the form $(\langle x_{j-1}(t-1)\rangle + 2\langle x_j(t-1)\rangle + \langle x_{j+1}(t-1)\rangle - 4\langle x_i(t-1)\rangle)/4$ is the analog of the bracket $\left(\mathbf{E}_i^{(\xi)}(t-1) - \langle x_i(t-1)\rangle\right)$ from the Expressions (14), (15).

For the Formulas (16)-(19) it is also possible to provide the same numerical experiments as in [3] – see, for example, Figures 3 and 4.

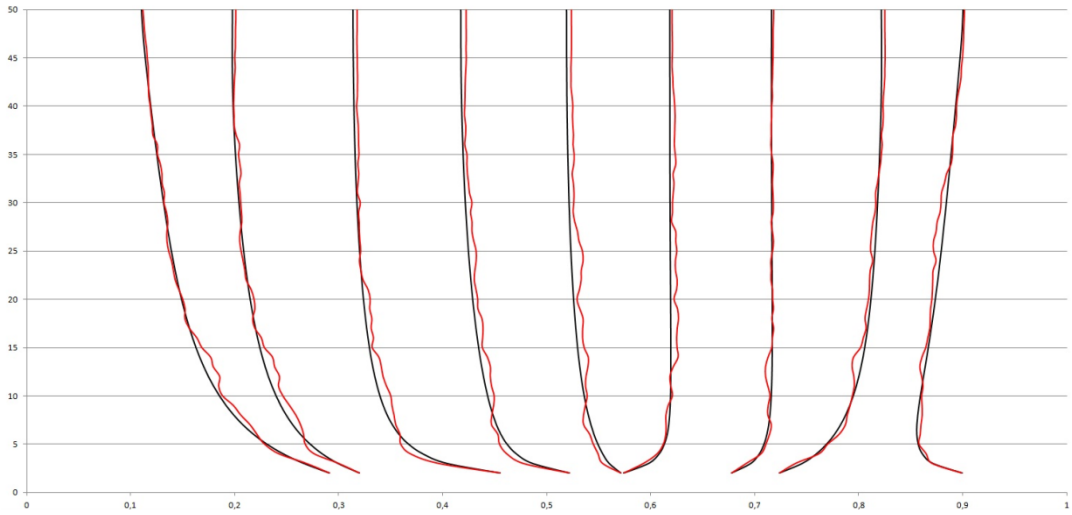

Figure 3 Iterative nodes moving: *grey* – formulas (16)-(19), *red* – Monte Carlo computations ($n = 100, r(t) \equiv 6$)

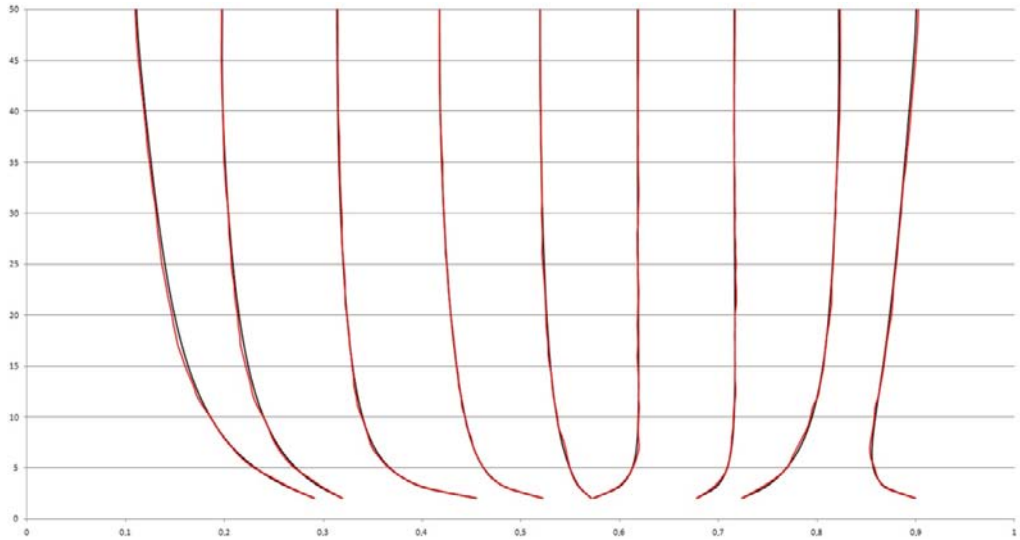

Figure 4 Iterative nodes moving: *grey* – formulas (16)-(19), *red* – Monte Carlo computations ($n = 1000, r(t) \equiv 6$)

Note that Figures 1-4 illustrate the boundary effect. This effect can be essentially decreased by special choice of the learning radius $r(t)$ from the Equation (18). For example, for computations presented on the Figures 3, 4 $r(t) \equiv 6$, and the boundary effect is evident. But just for $r(t) \equiv 4$ the boundary effect disappears (see Figure 5). More Monte Carlo (not analytical) computations concerning the boundary effect are presented in [4].

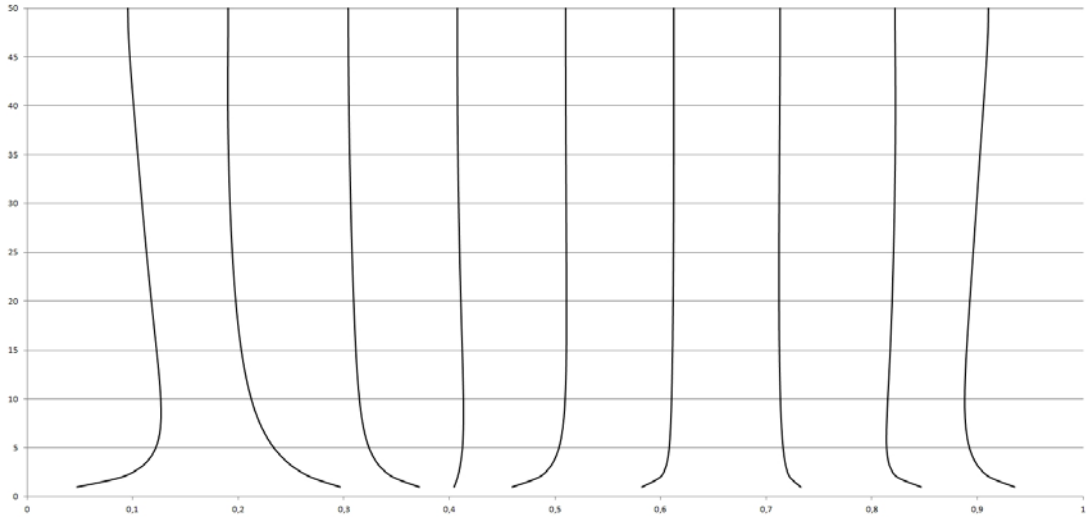

Figure 5 Iterative nodes moving: formulas (16)-(19) ($n = 1000,\ r(t) \equiv 4$); boundary effect is absent

Note that in [2] it is declared that it is expedient to use the learning radius of the following form:

$$r(t) = r(T) + \omega(t) \times \left(r(1)s^{t/T} - r(T)\right) \times t^{-1/4},$$

the function $\omega(t)$ and the constant s are from Relations (17), (18).

VI. CONCLUSIONS

In this paper we've proposed an analytical approach to description for application of the 1D Kohonen self-learning scheme and used it for investigation of boundary effect. The approach allows avoiding laborious Monte Carlo computations. We've proved that the boundary effect strongly depends on the learning radius.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 13-01-00032).

REFERENCES

[1] Kohonen, T., *Self-Organizing Maps* (Third edition), New York: Springer-Verlag, 2001.

[2] Nechaeva, O.I., Neural-Network Models, Algorithms, and Software Package for Constructing Adaptive Grids, Cand. Sci. (Phys.-Math.) Dissertation, Novosibirsk, 2007.

[3] Voytishek A.V. and Khmel D.S. Analytical description for application the 1D Kohonen scheme for constructing adaptive grids // *Numerical Analysis and Applications.* 2011. Vol. 4, no. 2. Pp. 105-113.

[4] Efimov A.I. *Discrete-Stochastic Numerical Methods of Constructing Adaptive Grids.* Master Thesis, Novosibirsk^ Novosibirsk State University, 2009.