Small Digit Capacity Arithmetic for Problems of Discrete Optimization

M.V. Kryzhanovsky and M.Yu. Malsagov

Scientific-Research Institute for System Analysis of RAS
Vavilova Street 44/2, Moscow 119333 Russian Federation
iont.niisi@gmail.com

Abstract

A possibility of application of clipping procedure in the optimization task of quadratic functional $E = (\mathbf{x}, \hat{\mathbf{A}}\mathbf{x})$ has been researched. It is shown that a direct use of clipping procedure is not a particular advantage by acceleration of algorithm performance in the search for global minimum. A modification of clipping procedure with q parameter (number of gradations) is proposed. It is shown that with the increase of q the probability of coincidence of gradients $E(x)$ direction with its clipped analog $E_C(x) = (\mathbf{x}, \hat{\mathbf{C}}\mathbf{x})$ goes up to 1.

1 Introduction

The replacement of matrix $\hat{\mathbf{A}}$ by the clipped matrix $\hat{\mathbf{C}}$ was first investigated in patterns recognition tasks (Van Hemmen, 1987 and Kintzel 1995). Analytical estimations of neural network memory and its recognizing capacity were obtained. These investigations were continued in (Alieva, 2006, Kryzhanovsky, 2007, Widrow, 1960), which had the following primary results:

- Reduction in energy $E_C(x)$ of the clipped Hopfield network by the transition from one state to another is followed by reduction in energy $E(x)$ of initial network.
- Performance speed of algorithm, based on the use of the clipped matrix is 40 times faster than the performance speed of algorithm, based on the use of Hopfield network.
- The RAM requirements are being reduced by approximately as much times.

This was the basis for the use of clipping procedure by solving optimization tasks, offered in (Kryzhanovsky, 2007). In our work we have proposed a modified clipping algorithm, enabling to accelerate the search for global minimum.

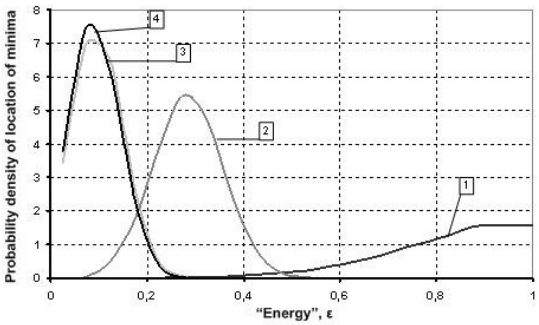

Figure 1: Comparison of two optimization methods: the method based on Hopfield network with the method, using clipping network.

2 Use of the traditional clipping procedure

The search for global minimum of quadratic functional $E = (\mathbf{x}, \hat{\mathbf{A}}\mathbf{x})$ in the discrete binary space lies in the repeated use of the standard Hopfield optimization procedure, i.e. carrying out a series of "descents" on the energy surface $E(x)$ from initial states $\{x_1\}$ to final states $\{x_4\}$, transitions $\{x_1\} \to \{x_4\}$. The deepest minimum is selected in the course of descents series.

By using the clipped functional the search process is divided into two phases, transitions $\{x_1\} \to \{x_2\} \to \{x_3\}$. The local minimums $\{x_2\}$ of functional $E_C(x)$, determined during the first phase, become the initial start minimums while minimizing functional $E(x)$ during the second phase.

In order to determine the efficiency of such approach, a computer modeling was carried out, during which matrix $\hat{\mathbf{A}}$ with the dimension 100 with random equal distribution of elements was generated, and its clipped analog – matrix $\hat{\mathbf{C}}$ was calculated. For each configuration of neural network a number of 200 000 starts was carried out from the states, specified by vector $\mathbf{x}$, which components were randomly generated. In the course of experiment the amount of calculations was determined and the energy of local minimum was recorded for each "descent". Figure 1

shows the results of comparison of two optimization methods – standard method and method using a clipping procedure. An "energy" $\varepsilon = (E_0 - E)/E_0$ is plotted along the axis of abscissas, where E_0 is the global minimum energy, E is the energy of the obtained local minimum of functional $E = (\mathbf{x}, \hat{\mathbf{A}}\mathbf{x})$. The density of probability of detecting a minimum is plotted along the axis of ordinates.

Curve 1 – shows distribution of initial start states $\{x_1\}$ by "energies" ε.

Curve 2 – characterizes the distribution of states $\{x_2\}$, obtained while optimizing of clipped functional $E_C(x)$, i.e. transitions $\{x_1\} \to \{x_2\}$.

Curve 3 – determines the distribution $\{x_2\} \to \{x_3\}$ after correction of states with the use of Hopfield network during the second phase.

Curve 4 – corresponds to the distribution by "energy" ε during transitions $\{x_1\} \to \{x_4\}$ using Hopfield network.

The figure demonstrates that the use of a clipped network does really shift the distribution of states by "energies" ε by 0.7, as compared to the initial $\{x_1\}$. Despite such a shift, the amount of calculations has been decreased insignificantly. The final distribution with the use of two-phase algorithm is close to the use of the standard Hopfield neural network. The probabilities of reaching the global minimum are also equal, amounting to ≈ 0.006. The modified algorithm of clipping has been offered in order to accelerate the performance speed of the two-phase algorithm of search for deeper minimums.

3 The use of a modified clipping procedure while searching for global minimum

The modification of clipping procedure is carried out as follows. Each element of interconnections matrix $\hat{\mathbf{A}}$ of Hopfield neural network is compared with an element of matrix $\hat{\mathbf{C}}$ by using the following formula:

$$\hat{C}_{ik} = \frac{1}{(q+1/2)} \operatorname{sgn}(\hat{A}_{ik}) \cdot round\left[(q+1/2) \cdot |\hat{A}_{ik}|\right], \quad (1)$$

where q is the number of gradations, more than zero. Function *sgn* forms the sign of a number and function *round* enables to round to the nearest whole number.

We will analyze correlation of gradients $\mathbf{H}$ and $\mathbf{h}$, i.e. initial functional $E(x)$ and clipped $E_C(x)$, which have the following form:

$$\mathbf{H} = \hat{\mathbf{A}} \cdot \mathbf{x}, \quad (2)$$

$$\mathbf{h} = \hat{\mathbf{C}} \cdot \mathbf{x} = \hat{\mathbf{A}} \cdot \mathbf{x} - \mu \cdot \hat{\mathbf{B}} \cdot \mathbf{x}. \quad (3)$$

Value $\mu = 1/(2q+1)$ characterizes the roughening level. The addend in (3) characterizes remainder, obtained from the roughening of elements $\hat{\mathbf{A}}$. $\hat{\mathbf{B}}$ is the matrix with an equal random distribution of elements within the range $[-1;1]$, and each vector component $(\hat{\mathbf{B}}, \mathbf{x})$ acts as a random value, having the Gaussian distribution $(N >> 1)$. Let us calculate the probability of coincidence of directions of P fields, i.e. coincidence by the sign of some components of vectors $\mathbf{H}$ and $\mathbf{h}$:

$$P = \Pr[H_i \cdot h_i > 0]. \quad (4)$$

The elements of initial matrix $\hat{\mathbf{A}}$ are equally distributed with a mean zero $\overline{A}$ and variance $\sigma^2(\hat{A})$.

Taking into account expressions (2) and (3) it may be shown that mathematical expectation and variance of value H_i and h_i are described by means of expressions:

$$\overline{H}_i = 0, \quad \sigma^2(H_i) = N\sigma^2(\hat{A}); \quad (5)$$

$$h_i = 0, \ \sigma^2(h_i) = (1-\mu^2) \cdot \sigma^2(H_i); \quad (6)$$

$$\overline{h_i \cdot H_i} = \sigma^2(h_i). \quad (7)$$

Taking into account expressions (5-7) the coefficient of ρ correlation of gradients H_i and h_i will be equal to:

$$\rho = \frac{\overline{H_i \cdot h_i} - \overline{H}_i \cdot \overline{h}_i}{\sigma(H_i)\sigma(h_i)} = 1 - \mu^2. \quad (8)$$

Minimum reached level $\rho = 0.889$ corresponds to $q = 1$ and $\mu = 1/3$.

The probability of coincidence of the fields $P = \Pr[H_i \cdot h_i > 0]$ directions in its turn is determined as:

$$P = \frac{1}{\pi\sqrt{1-\rho^2}} \int_0^\infty \int_0^\infty Exp\left\{ -\frac{[f(H_i, h_i)]}{2(1-\rho^2)} \right\} \frac{dH_i dh_i}{\sigma(H_i)\sigma(h_i)}, \quad (9)$$

where

$$f(H_i, h_i) = \left(\frac{H_i - \overline{H}_i}{\sigma(H_i)}\right)^2 - 2\rho\left(\frac{H_i - \overline{H}_i}{\sigma(H_i)}\right)\left(\frac{h_i - \overline{h}_i}{\sigma(h_i)}\right) + \left(\frac{h_i - \overline{h}_i}{\sigma(h_i)}\right)^2.$$

Calculation of this double integral is possible only numerically. The results of calculation are shown on figure 2.

Based on formula (9) calculations it is possible to approximately evaluate the probability of coincidence of gradients using formula (10), correct when the values of correlation ρ coefficient close to 1:

$$P(\mu) = \frac{1}{2} - \frac{1}{\pi}\arcsin\rho\,. \qquad (10)$$

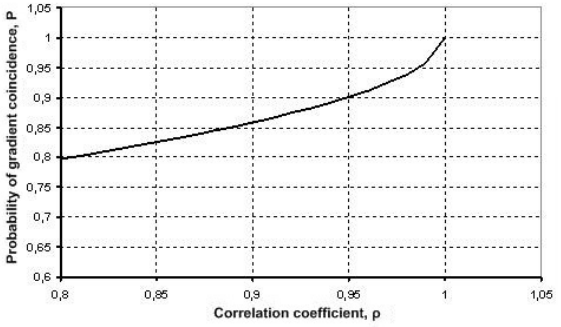

Figure 2: The probability of gradients coincidence of the clipped functional and initial functional depending on correlation coefficient.

From the graph, formulas (9) and (10), shown on figure 2, it follows that it is highly probable that the local gradients of the initial functional and clipped functional coincide $(P \geq 0.894)$. With the increase of the number of q gradations (decrease of μ) this probability grows and tends to 1. Since the optimization process lies in the consequent overturn of all N spins of Hopfield model, it becomes apparent that with the increase of the task dimension $(N >> 1)$ the probability of the fact that in the process of optimization of functional the energy E will be increased, asymptotically tends to zero.

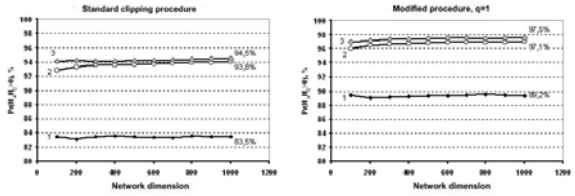

Figure 3: The probability of gradients coincidence using traditional (to the left) and modified (to the right) clipping procedure.

Computer modeling was carried out in order to verify the obtained results. Figure 3 represents experimental data on probability of gradients **H** and **h** coincidences in case of using standard clipping procedure (to the left) and its modification when $q = 1$ (to the right). For this purpose a point x_1 (curve 1) has been selected in a random manner, from which Hopfield network with interconnections matrix $\hat{\mathbf{C}}$ converged in the nearest local minimum x_2 (curve 2). The point of minimum x_3 is obtained after correction of states has been made by the standard Hopfield network (curve 3). In these points, x_1, x_2, x_3 a correspondence of vectors **H** and **h** directions was determined. Dimension of network for which the calculations have been carried out, was plotted along the axis of ordinates.

The obtained results of experiments precisely correspond to the theory. This way experimentally measured probability of coincidence of gradients for the modified method when $q = 1$ makes $P = 0.892$ (to the right, curve 1), the estimated value at the same time equals to $P = 0.896$.

The comparison of experimental data on these figures shows an advantage of the modified clipping method.

In randomly chosen start points the probability of coincidence for the modified method equals to $P = 0.892$, while for the initial one $P = 0.835$.

In minimum points of the clipped network the probability to the left equals to $P = 0.938$, to the right – $P = 0.971$.

Accordingly in minimum points of standard network to the left we have $P = 0.945$, and to the right – $P = 0.975$.

The obtained results do not depend on network dimension.

Figure 4 shows the correlation between "energies" of minimums of clipped network and standard Hopfield network for various values of q parameter. For this purpose a point x_0 has been randomly chosen, and from this point the Hopfield network with interconnections matrix $\hat{\mathbf{C}}(q)$ has been converged to the nearest local minimum x_m. The values of functional $E(x_m)$ and clipped functional $E_C(x_m)$ as well as further indicated values of "energies" ε were calculated.

It is shown that with the increase of q parameter the domain of proportional dependence is expanding, its boundary becomes closer to the point of coordinates origin. This means that with the increase of q parameter the area of correspondence "the deeper is the clipped minimum – the deeper is the minimum of standard network" closer approaches to the global minimum.

On figure 5 the distribution of states by energy for various numbers of gradations, obtained after the first phase of optimization (q clipping procedure) is represented. It is shown that with the increase of q value, the distribution is shifted to the left side, therefore, bringing us closer to the global minimum and reducing the way, which needs to be done by standard Hopfield network on the 2nd phase. It should

be noted that the modification of clipping procedure has not changed the probability of discovering a global minimum.

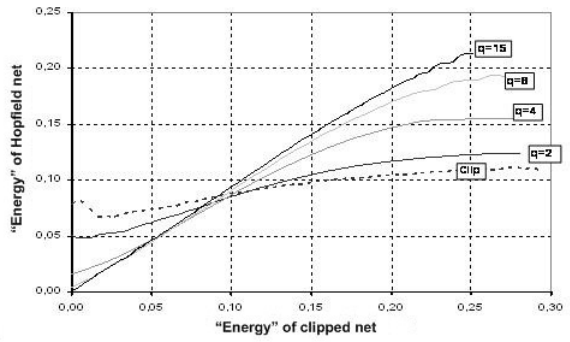

Figure 4: Correlation between "energies" of minimums of clipped network and standard Hopfield network for various values of q parameter.

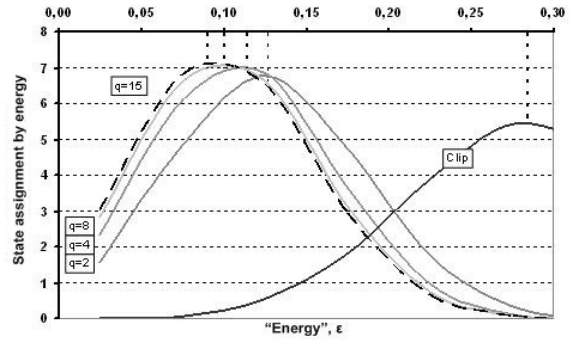

Figure 5: Distribution of states by energy for different amounts of gradations.

4 The use of a modified algorithm of clipping

The principal amount of computations by calculations of Hopfield neural network falls on calculation of gradients, i.e. on matrix-vector product. For high dimensional matrixes $N \sim 10^3 - 10^4$ the 10^{th} byte number representation is used for obtaining a required calculation accuracy. The representation of interconnection matrix of neural network by means of the numbers of shortened capacity enables to accelerate computation process. If the addition of two 10-byte numbers requires the first cycle of processor time, in order to add two ten-dimensional vectors, which components are 1-byte numbers, the time required will be less than 1 cycle. In addition to it, the load of operands from memory to processor registers also requires time, comparable to the time of operation's fulfillment. Therefore, the "efficient" time used for the fulfillment of 1-byte operation is ~15 times less than the time used for fulfillment of 10-byte operation.

Idea of modification lies in the application of whole numbers in one- and two-byte representation for matrix elements. The algorithm of descent over the "energy surface" was developed according to it, and contains three stages, when local minimum received at the current stage represented the initial approximation for the next stage. The first and second stages used clipping modification with the number of gradations $q = 2 \div 64$ at the first (one-byte operations) and $q \geq 255$ at the second stage (two-byte operations). The standard Hopfield network was used for the correction of states during the third stage.

The acceleration of computation process depends on the number of $q^{(1)}$ gradations, chosen during the first stage. The second stage enable to choose q number as maximum number $q \geq 2048$.

For determining optimum value $q^{(1)}$ during the first stage and the acceleration level of θ algorithm, a computational experiment was carried out. O each phase of functional optimization a number of iterations, proportional to the amount of calculations, was defined by hitting the local minimum. In such a case a number of steps at the time of "descent" was recorded by using the initial Hopfield network.

The computational experiment's results are shown on figure 6 and in table 1. The θ acceleration of algorithm performance is determined by formula (11).

$$\theta = \frac{15 I^{(H)}}{I^{(1)} + 2I^{(2)} + 10 I^{(HL)}}, \qquad (11)$$

where $I^{(H)}$ – is the number of iterations, using the standard Hopfield method;

$I^{(1)}$ – is the number of iterations at the first stage (one-byte arithmetic);

$I^{(2)}$ – is the number of iterations at the second stage (two-byte calculations);

$I^{(HL)}$ – is the number of iterations at the stage of algorithm correction using the standard Hopfield method;

Figure 6 shows correlation between the number of iterations, expended at different stages of the modified clipping algorithm with regard to the number of iterations, expended by standard method.

It is shown that at the first stage $I^{(1)} / I^{(H)} \approx 1$. This correlation does not depend on the choice of $q^{(1)}$, number of iterations at the second and third stages depends on the value of gradations amount at the first stage: $I^{(2)} / I^{(H)} \approx 0.3$, and $I^{(HL)} / I^{(H)} \approx 0.05$. The number of iterations at the second stage does not depend on the amount of gradations.

The speed of algorithm performance depends on the choice of $q^{(1)}$ parameter, used during the first stage.

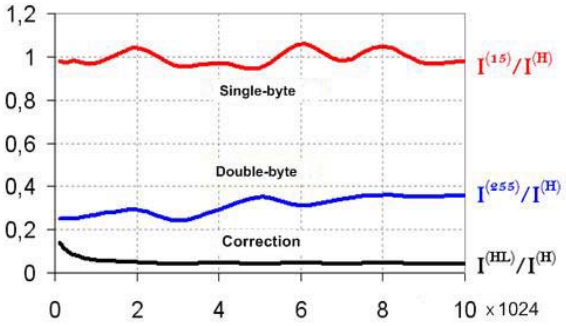

Figure 6: Number of steps at each stage of algorithm depending on dimension of $\hat{\mathbf{A}}$ matrix of neural interconnections.

With the increase of gradations number at the second stage, the number of steps at the stage of correction of minimum position by the initial neural network is shortened to 1. Therefore during the second stage $q^{(2)} \geq 2048$ should be chosen and the third stage may not be required.

Table 1: Algorithm accelerations as compared to the standard Hopfield method depending on q parameter values.

Network dimension N = 256

Number of gradations, q	**Algorithm acceleration, θ**
2	3,3
4	4,3
8	7,9
12	10,3
15	11,2
32	6,4
64	3

Table 1 contains the results of algorithm acceleration as compared to the standard Hopfield method depending on $q^{(1)}$ parameter values. The results were obtained on the dimension matrix $N = 256$, and $q^{(2)} = 255$ (the worst option). It is shown that the optimum of function $\theta(q)$ more than 10 times is reached by $q^{(1)} = 10 \div 15$. For calculation of algorithm acceleration value by formula (11) the data shown on figure 6 are used.

5 Conclusions

The probability of discovering global minimum does not depend on the number of gradations and is the same as by using Hopfield network.

1. The modification of clipping procedure of neural interconnections matrix enables to accelerate the computation process by more than 10 times.
2. The optimal values of gradations number at each stage of algorithm performance have been found: $q^{(1)} = 12$, $q^{(2)} \geq 2048$. The results have been confirmed for dimension matrixes $N = 64 \div 10240$.

Acknowledgments

The work was carried out supported by grant 9-07-00159-a from the Russian Foundation for Fundamental Research and in part supported by the Russian Federation President Grant for Government Support of Leading Scientific School (356-2008-9).

Reference

Alieva D. I. and V. M. Kryzhanovsky (2006). “Small size Hopfield model with clipped interconnections” Iskusstvenny Intellekt, №3, pp. 240-248.

Kintzel W. and M. Opper (1995). “Models of Neural Networks I.” Physics of Neural Networks, Eds. E. Domany, J.L. van Hemmen, K. Schulten, Springer, p.170.

Kryzhanovsky B. V. and V. M. Kryzhanovsky (2007). “On using clipping procedure in optimization tasks” IX All-Russian scientific and technical conference «Neuroinformatics-2007», MEPHI, vol.1, pp.197-205.

Kryzhanovsky B. V. and V. M. Kryzhanovsky, A. L. Mikhaelyan (2007). “The use of clipping procedure in binary minimization of quadratic functional” DAN-2007, vol. 413, №6, pp.730-733.

Kryzhanovsky V. M. and D. I. Simkina (2007). “Small size clipped Hopfield models” Vestnik kompyuternykh i informatsionnykh tekhnologiy, №10, pp.27-31.

Van Hemmen, J. L. (1987). “Nonlinear neural networks near saturation” Physical Review A, V36, pp.1959-1962.

Widrow, B. and M. E. Jr. Hoff (1960). “Adaptive switching circuits” IRE Western Electric Show and Convention Record, Part 4, pp.96-104.