Hopfield Model in the Problem of Binary Minimization

Ya. M. Karandashev, B. V. Kryzhanovsky and A.B. Fonarev

Abstract—A problem of quadratic functional minimization in a discrete space is considered. It is shown that the transformation of a functional by modification of its matrix can significantly accelerate a procedure of a random search. As example we chose two well-known local optimization algorithms: Hopfield neural-network dynamics and Kernigan-Lin algorithm. The proposed method of functional transformation improves efficiency of both algorithms by many times.

I. INTRODUCTION

The goal of this paper is to improve the efficiency of a random search procedure used to solve binary minimization problems. In this class of problems, the solution is reduced to the minimization of the quadratic functional \( \mathcal{E}_i(S) \) constructed from a given \( N \times N \) matrix \( T \) in the \( N \) dimensional configuration space of states \( S = (s_1, s_2, ..., s_N) \) with discrete variables \( s_i = \pm 1 \), \( i = 1, 2, ..., N \).

A whole number of different NP-hard problems can be reduced to this minimization problem, e. g. graph partitioning, clustering, traveling salesman problem, scheduling theory etc.

Attempts are usually made to improve the efficiency of the random search procedure by modifying the dynamics of a descent over the landscape \([1–3]\) described by \( \mathcal{E}_i(S) \). In contrast to this approach, we propose not to change the dynamics of landscape descent but rather to transform the energy landscape itself so as to increase the radius of the attraction domain of the global minimum (and of other minima comparable in depth with the global one). We consider the simplest transformation, namely, the raising of \( T \) to the power \( k = 2, 3, ..., \). This approach is found to be fairly productive: due to the landscape transformation, the probability of finding the global minimum increases by several orders and the spectrum of found minima is strongly shifted toward the deep side.

The efficiency of the algorithm proposed is rigorously substantiated only for “random” matrices, whose elements are independent random variables. The application of the algorithm to matrices of other types is heuristic.

II. PROCEDURE

The standard statement of the binary minimization problem is as follows. Given an \( N \times N \) matrix \( T \), find an \( N \)-dimensional configuration vector \( S_m = (s_{m1}, s_{m2}, ..., s_{mN}) \), \( s_{mi} = \pm 1 \), \( i = 1, 2, ..., N \), that minimizes the energy functional \( \mathcal{E}_i(S) \):

\[
\mathcal{E}_i(S) = -\frac{1}{\sigma_T N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} T_{ij} s_i s_j
\]  

where \( \sigma_T \) is the standard deviation of the matrix elements \( T_{ij} \). Functional \((1)\) can be symmetrized. For this reason, without loss of generality, we assume that the matrix \( T_{ij} \) is symmetric and its diagonal elements are zero \((T_{ii} = 0)\).

The minimization procedure is based on the Hopfield model \([4]\), which is the core of most binary minimization algorithms. This is a one-dimensional system of \( N \) spins, whose interaction is defined by the energy functional \( \mathcal{E}_i(S) \).

The standard (asynchronous) dynamics of the model can be described as follows. The local field \( h_i = -\partial \mathcal{E}_i(S)/\partial s_i \) acting on the arbitrarily chosen \( i \)th spin is calculated as

\[
h_i = \frac{1}{\sigma_T N^2} \sum_{j=1}^{N} T_{ij} s_j
\]  

If \( h_i \neq 0 \), the state of the spin is updated according to the decision rule \( s_i = \text{sgn} h_i \). This procedure is sequentially applied to all the neurons until the network converges to a stable state \( S_m \). This dynamics is a descent over the energy landscape \( \mathcal{E}_i(S) \), which is a complete analogue of the coordinate-wise gradient descent in a real space. In section V we will give another more complicated and efficient dynamics; this dynamics makes it possible to go out of small local minima.

The random search procedure is described as follows. Given an arbitrarily initial state of the network, the nearest local minimum is found. This procedure is repeated until a minimum with an acceptable depth is found. The efficiency of the random search procedure is evaluated by the probability of finding the global minimum, by the rate of finding a minimum with a given depth, or by the mean depth of the minima found.

III. PRELIMINARIES

Before transforming the energy landscape, we establish

Manuscript received January 29, 2012.

Ya. M. Karandashev and B. V. Kryzhanovsky are with the Scientific Research Institute of System Analysis RAS, Moscow, Russia

A.B. Fonarev is with CUNY City University of New York Department of Engineering and Science, 2800 Victory Blvd. SI, NY 10314
the basic relations associated with the depth of the global (local) minimum, which underlie the subsequent argument.

The first relation is a constraint on the depth of the minimum. Let $S_0 = (s_{01}, s_{02}, ..., s_{0N})$ be the configuration corresponding to the global minimum $E_0 = E_i(S_0)$. We extract from $T$ the term $T_0$ that is responsible for the formation of this minimum. To this end, $T$ is represented as

$$T = T_0 + T_1, \quad T_0 = r_0 \sigma_f \sum_{i=1}^N s_{0i} s_i$$

(3)

The statistical weight $r_0$ is found from the condition that the elements of $T_0$ and $T_1$ do not correlate. Calculating the covariance of the matrix elements and setting it equal to zero, we obtain

$$r_0 = -\frac{E_0 + \bar{T} \sigma}{1 - \sigma^2}, \quad \sigma = \frac{1}{N^2} \left[ \sum_{i=1}^N s_{0i} \right]^2 - N$$

(4)

where $\bar{T}$ is the mean of the elements of $T$ and $\delta$ is a variable with a zero mean and a small standard deviation $\sigma_\delta = \sqrt{2} / N$. For simplicity, we set $\bar{T} = 0$ and $\delta = 0$ (the generalization to other cases is obvious). Then (4) yields the relation

$$E_0 = -r_0$$

(5)

The variances of the elements of $T_0$ and $T_1$ are $\sigma_0^2 = r_0^2 \sigma_f^2$ and $\sigma_1^2 = \sigma_f^2 - \sigma_0^2$. Therefore, we have managed to present the random matrix $T$ as the sum of two independent random matrices $T_0$ and $T_1$. Moreover, (3) and (4) imply that $S_0 T_1 S_0^T = 0$, which suggests that the contribution of $T_1$ to $E_0$ is strictly zero; i.e., the minimum in $S_0$ is caused only by the contribution of $T_0$.

Following [5], we continue decomposition (3) and represent the matrix as a weighted sum of exterior products of random vectors:

$$T = \sigma_f \sum_{m=0}^\infty r_m S_m^T S_m, \quad \sum r_m^2 = 1.$$  

For this type of matrices, it was shown in [6] that any of the vectors $S_m$ present in the decomposition of $T$ is a minimizer of functional (1) if and only if its statistical weight $r_m$ is larger than the critical value

$$r_c = \frac{1}{2 \sqrt{N \alpha_c}}$$

(6)

where $\alpha_c \approx 0.138$ is the critical value of loading parameter.

This assertion is concerned primarily with the point $S_0$, which by definition is a minimizer of functional (1) and satisfies the relations

$$1 \geq r_0 \geq r_c, \quad E_c \geq E_0 \geq -1, \quad E_c = -r_c$$

(7)

The second necessary relation obtained in [7] is that, as the depth of $E_0$ increases, its width increases as well and, accordingly, the probability of finding this minimum grows as $P(E_0) \sim \exp\left( -NE_0^2 / E_0^2 \right)$.

As a result, we have established the following two relations: (a) for a larger statistical weight $r_0$ of the addition of $S_0$ in the original matrix $T$, the minimum $E_0$ is deeper and the probability of finding it is higher; and (b) $S_0$ can be a minimum only if $r_0 \geq r_c$; i.e., the depth of the minimum is always larger than the critical value $|E_c|$. These relations suggest the direction of improving the efficiency of the random search algorithm: the energy landscape (1) has to be transformed so as to increase the depth of the global minimum and, accordingly, to increase the probability of finding it.

IV. TRANSFORMATION OF ENERGY LANDSCAPE

The surface described by the quadratic form $E_i(S)$ can be transformed only by transforming the underlying matrix. The matrix $M = (1 - z)T + zT^4$, where $T^4$ is obtained by raising $T$ to the power $k$ and setting the diagonal elements equal to zero, is plugged into (1). Changing the parameter $z$ from 0 to 1, we pass from the matrix $T$ to $M = T^4$. Accordingly, the landscape described by $E_i(S)$ is transformed into that described by $E_i(S)$:

$$E_i(S) = -\frac{1}{\sigma_M N^2} \sum_{i=1}^N \sum_{j=1}^N M_{ij} s_i s_j$$

(8)

where $\sigma_M$ is the standard deviation of the elements of $M$. Obviously, under the landscape transformation, the global minimum is shifted in space and its depth and the width of the attraction domain change as well. It will be shown below that, for $k \leq 5$, the above transformation leads to a considerable increase in the depth of the minimum, while its shift is relatively small.

Accordingly, we propose the following two-step minimization algorithm. At the first step, a descent over $E_i(S)$ is performed and a configuration $S_m$ is found that minimizes $E_i(S)$. The second step involves correction, namely, from the point $S_m$, we descend over $E_i(S)$ to the nearest minimum $S_n$ of $E_i(S)$. The descent over $E_i(S)$ is performed as described above: we calculate the local field of the $i$th spin $h^{(i)}_m = -\partial E_i(S)/\partial s_i$ and, if $h^{(i)}_m \neq 0$, the state of the spin is updated according to the decision rule $s_i = \text{sgn} h^{(i)}_m$. The algorithm is substantiated in the case of $k = 2$.

4.1. Let us show that the landscape transformation leads to a deeper minimum. Consider the energy $E_{20} = E_2(S_0)$ at the point $S_0$. Following (3), the matrix $M = T^2$ is represented...
as $M = T_0^2 + T_i^2 + (T_0 T_i + T_i T_0)$. In view of $S_0 T_i S'_0 = 0$ and $\sigma_\mu = \sqrt{N} \sigma^2$, we then derive from (8) that

$$E_{20} = -\sqrt{NE_0^2} + \frac{1}{\sigma_\mu^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (T^2)_{ij} s_{0i} s_{0j}$$  (9)

In the limit of $N \gg 1$, $E_{20}$ can be viewed as a normally distributed quantity with the mean $E_{20} = -\sqrt{NE_0^2}$ and the relatively small standard deviation $\sigma_\mu = (1 - r_0^2) / N$. Since $E_{20} / E_0 = r_0 \sqrt{N} \geq 1.33$, it should be expected that the minimum is deepened under the landscape transformation. The probability of this event is given by

$$\Pr\{E_{20} < E_0\} = \frac{1}{2} (1 + \text{erf} \gamma)$$  (10)

where

$$\gamma = \frac{E_0 - E_{20}}{\sqrt{2} \sigma_\mu} = \frac{r_0 N (r_0 \sqrt{N} - 1)}{\sqrt{2} (1 - r_0^2)}$$

The relation $r_0 \geq r_1$ implies $\gamma \geq 0.3 \sqrt{N}$. Therefore, as $N$ increases, $\Pr\{E_{20} < E_0\}$ tends asymptotically to unity. In other words, with an overwhelming probability, the landscape transformation yields a considerably deeper minimum ($E_{20} / E_0 \geq 1.33$). As a result, according to [8], the probability of finding the global minimum increases exponentially in $N$.

4.2. Let us estimate the shift of the minimum under the landscape transformation. The mean shift can be represented as $d = NP$, where $P = \Pr\{s_{0i} h^{(i)} < 0\}$ is the probability that the directions of the spin $s_{0i}$ and the local field $h^{(i)}$ do not coincide. The value $s_{0i} h^{(i)}$ can be represented as

$$s^{(0)}_{0i} h^{(i)} = \frac{1}{N^{3/2}} (N r_0^2 + R),$$  (11)

where $R$ is a normally distributed variable

$$R = \frac{1}{N \sigma_\mu^2} \sum_{i,j} (N r_0 \sigma T_i + T_i^2) s_{0i} s_{0j}$$

with zero mean and the standard deviation $\sigma_\mu = \sqrt{1 - r_0^2}$. In view of (11), $P$ is expressed in terms of the error function. Then we obtain

$$d = \frac{N}{2} (1 - \text{erf} \eta),$$  (12)

where

$$\eta = \frac{N r_0^2}{\sqrt{2} (1 - r_0^2)}$$

Evidently, the shift of the minimum is not large: in view of $r_0 \geq r_1$, it follows from (12) that $d \leq 0.04N$.

Expressions (9)–(12) suggest the following conclusions.

With a high probability, the landscape transformation leads to deeper minima and, as a result, to a higher probability of finding them. Moreover, the depth increase $E_{20} / E_0 = r_0 \sqrt{N}$ is larger for a larger initial depth $|E_0| \approx r_0$. In other words, deep minima become even deeper and the probability of finding them increases, while shallow minima become shallower (or disappear at all) and the probability of finding them is reduced. This means that the spectrum of minima found by the algorithm shifts considerably toward the global minimum, and the probability of finding the latter increases considerably. The shift caused by the transformation is relatively small: it follows from (12) that the smallest shifts are expected for the deepest minima.

These conclusions were confirmed by numerical experiments. Fig. 1 shows the spectral densities of minima of $E_1(S)$ and $E_2(S)$: $N(E)dE$ is the number of local minima found on the energy interval $[E, E + dE]$ with respect to the total number of minima found. The squares mark the energies $E_1(S_0)$ and $E_2(S_0)$ in the configuration $S_0$ corresponding to the global minimum of $E_1(S)$.

Fig. 1. Spectral density of minima of $E_1(S)$ and $E_2(S)$: $N(E)dE$ is the number of local minima found on the energy interval $[E, E + dE]$ with respect to the total number of minima found. The squares mark the energies $E_1(S_0)$ and $E_2(S_0)$ in the configuration $S_0$ corresponding to the global minimum of $E_1(S)$.

It should be emphasized that the proposed two-step descent algorithm with modified matrix of coupling may use any dynamics of descent, not only Hopfield dynamics. Then let us consider another more efficient dynamics.
V. KERNIGHAN-LIN ALGORITHM

Now let us discuss the local search algorithm based on procedure proposed by Kernighan and Lin [1]. In [9] it was adapted to our problem. This local search algorithm is like Hopfield dynamics. However it is more efficient due to capability of going out of small cavities in the energy landscape. Here we use this algorithm in form as it was proposed in [9].

Let us first define a few terms. Consider a configuration and one of its spins \( s_i \). We call gain of that spin the decrease in the configuration’s energy when \( s_i \) is flipped (the gain is thus minus the change in the total energy). A spin is said stable (resp. unstable) if its gain is negative (resp. positive). Spin \( s_i \) is said to be more unstable than spin \( s_j \) if gain of \( s_i \) is greater than the gain of \( s_j \).

Kernighan-Lin local search algorithm proceeds as follows.

1. Choose a spin that will be the “seed” of the growing cluster. This is done by taking any of the strictly unstable spins of the current configuration; if there are none, choose a spin at random anywhere. Virtually flip this spin and mark it so it cannot be flipped again at any time during the cluster’s growth. Compute the new gains of all the other spins of this modified configuration.
2. Add to the cluster the spin with the highest gain, be it positive or negative. Update the configuration and the gains and again mark this spin so it will not be considered for flipping during the growth of this cluster.
3. Return to step 2 unless there are no more spins to add to the cluster.
4. If the best cluster encountered during the growth process has a strictly positive gain, flip it.
5. Return to step 1 unless all clusters found had negative gains.

At the end of this search, all spins have negative gains, so we guarantee that the configuration is at least one-spin-flip optimal. One of the important features of this Kernighan-Lin-like algorithm is that the size of the cluster is not limited (except by \( N \)). It can be very large in practice, especially when the original configuration is random.

VI. HELPFUL HINTS

Finally, we formulate the minimization algorithm proposed.

The preliminary phase consists of the following steps. The original matrix \( T \) is symmetrized (if it is initially not symmetric) and its diagonal elements are set to zero. The matrix is raised to the \( k \)th power and the diagonal elements in the resulting matrix \( M = T^k \) are set to zero. The functional \( E_i(S) \) is constructed from \( M \) according to (8).

After the preliminary phase, the random search procedure based on the two-step descent algorithm is executed. Specifically, at the first step, a descent over \( E_i(S) \) is performed from a random initial configuration to the nearest local minimum \( S_{m}^{(a)} \) of \( E_i(S) \). The second step involves correction: from the point \( S_{m}^{(i)} \) on \( E_i(S) \), we descend to the nearest local minimum \( S_{m}^{(s)} \) of \( E_i(S) \), which is, as a rule, located near \( S_{m}^{(a)} \). Both steps of the descent are performed according to Hopfield or Kernighan-Lin dynamics.

VII. OBTAINED RESULTS

The efficiency of the two-step descent algorithm was verified for \( k \) ranging from 2 to 7 for matrices of size \( N = 100, 196 \) from the two-dimensional Ising model [2].

For comparison we tested our algorithm with both dynamics Hopfield and Kernighan-Lin. The numerical results are shown in table 1. The first column contains the data for the standard algorithm (descent over \( E_i(S) \)), while the subsequent columns present the data for the new algorithm based on a descent over \( E_i(S) \) for \( k = 2\ldots5 \).

Three characteristics are used to estimate the efficiency of the algorithms:

i. The first line gives the probability of hitting the energy range \( E_m \in [-1, -0.99] \), which is close to the global minimum (the energy of the global minimum is assumed to be \(-1\)).

ii. The second line presents the probability of hitting the global minimum.

iii. The third line lists the mean energy of minima \( \bar{E}_m \).

It can be seen that the results for \( k = 5 \) differ little from those for \( k = 3 \). For \( k > 5 \), the results are worse. Therefore, a further increase in \( k \) is not reasonable.

![Fig. 2. The vertical axis represents the probability density of finding a minimum of \( E_i(S) \) by various algorithms: (1) the standard algorithm and (2–5) the two-step descent algorithm with \( k = 2\ldots5 \). The energy of the global minimum is \(-1\). \( N = 100 \), Hopfield dynamics.](image)
TABLE I
EFFICIENCY COMPARISON OF THE PRESENT ALGORITHM AND THE STANDARD RANDOM SEARCH ALGORITHM

<table>
<thead>
<tr>
<th></th>
<th>N=100</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dynamics</td>
<td>$E_1$</td>
<td>$E_2$</td>
<td>$E_3$</td>
<td>$E_4$</td>
<td>$E_5$</td>
<td></td>
</tr>
<tr>
<td>Probability of hitting the energy range [-1, -0.99]</td>
<td>H</td>
<td>2.2×10^{-5}</td>
<td>2.3×10^{-3}</td>
<td>1.1%</td>
<td>5.6×10^{-3}</td>
<td>2.5%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>KL</td>
<td>5.0%</td>
<td>8.2%</td>
<td>16.2%</td>
<td>9.0%</td>
<td>17.4%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>2.4×10^{-6}</td>
<td>3.8×10^{-3}</td>
<td>2.1×10^{-3}</td>
<td>1.2×10^{-3}</td>
<td>4.1×10^{-3}</td>
<td></td>
</tr>
<tr>
<td>Probability of hitting the global minimum</td>
<td>KL</td>
<td>1.4%</td>
<td>2.8%</td>
<td>6.0%</td>
<td>3.9%</td>
<td>5.9%</td>
<td></td>
</tr>
<tr>
<td>Mean value of the minimum energy $\overline{E_m}$</td>
<td>H</td>
<td>-0.84</td>
<td>-0.89</td>
<td>-0.94</td>
<td>-0.90</td>
<td>-0.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>KL</td>
<td>-0.95</td>
<td>-0.96</td>
<td>-0.97</td>
<td>-0.96</td>
<td>-0.97</td>
<td></td>
</tr>
</tbody>
</table>

|          | N=196  |               |       |       |       |       |       |
|          | dynamics | $E_1$ | $E_2$ | $E_3$ | $E_4$ | $E_5$ |
| Probability of hitting the energy range [-1, -0.99] | H | - | 9.3×10^{-6} | 3.3×10^{-4} | 5.2×10^{-5} | 1.7×10^{-3} |
|          | KL     | 2.3×10^{-3} | 5.4×10^{-3} | 2.1% | 6.0×10^{-3} | 2.1% |
|          | H      | - | 1.0×10^{-7} | 2.3×10^{-6} | 2.0×10^{-8} | 8.3×10^{-9} |
| Probability of hitting the global minimum | KL | 4.7×10^{-5} | 2.2×10^{-4} | 4.3×10^{-4} | 3.6×10^{-4} | 6.7×10^{-4} |
| Mean value of the minimum energy $\overline{E_m}$ | H | -0.84 | -0.90 | -0.94 | -0.90 | -0.95 |
|          | KL     | -0.94 | -0.95 | -0.97 | -0.95 | -0.97 |

Firstly let us look into results when using Hopfield dynamics. Whereas the standard algorithm finds the global minimum with probability close to zero, the probability of finding the global minimum with the new two-step algorithm differs noticeably from zero. It can be seen from the Table at $k = 3$ and $N=100$ the probability of finding the global minimum is 875 times higher than in the standard algorithm. At $N=196$ the standard algorithm can not at all find the global minimum in an acceptable running time.

Fig. 2 displays the probability density $P(E)$, where $P(E)dE$ is the probability of finding a minimum of $E_i(S)$ in the energy range $[E, E + dE]$. It can be seen that, as $k$ increases, the spectrum of found minima is strongly shifted toward the global minimum. Fig. 3 shows the density probability $P_i(E)$ of finding a minimum normalized by the density probability $P_i(E)$ provided by the standard algorithm. Analogous charts for the case of Kernighan-Lin dynamics are shown in fig. 4 and 5 ($N=196$).

It should be noticed that the efficiency of our algorithm grows exponentially in $N$. Indeed, as follows from the table, when using $E_i(S)$ the probability of finding the global minimum increases by only 4.3 times at $N=100$ and increases by almost 10 times at $N=196$.

The optimal value $k = 3$ is explained by different causes. The main one is that the minimum at $S_0$ becomes deeper when the matrix is squared. However, it is no longer the global minimum of $E_i(S)$ (see Fig. 1), since new deeper minima (so called chimeras) are formed [6]. As the matrix is raised to the subsequent powers, these chimeras become even deeper and the probability of hitting them becomes much higher than for $S_0$. Moreover, at some step, the minimum at $S_0$ becomes relatively shallow. As a result, it either disappears or the probability of finding it becomes extremely low.
Fig. 4. The vertical axis represents the probability density: (1) the standard algorithm and (2–5) the two-step descent algorithm. The energy of the global minimum is \(-1\). \(N = 196\), Kernighan-Lin dynamics.

Fig. 5. The ratio \(P_2(E)/P_1(E)\) of the probability densities: \(P_2(E)\) corresponds to the two-step descent algorithm with \(k = 2, 3, 4, 5\), and \(P_1(E)\) corresponds to the standard algorithm. \(N = 196\), Kernighan-Lin dynamics.

VIII. CONCLUSION

The proposed method of functional modification proved to be highly productive. We have seen, in such algorithms as Hopfield neural-network dynamics and Kernighan-Lin bipartition graph algorithm, that this method works regardless of type of a local search algorithm used.

A comparison shows that the efficiency of the minimization algorithm is improved substantially due to the landscape transformation. Specifically, for a descent over \(E_1(S)\) with Hopfield dynamics, the probability of hitting the global minimum increases by more than 800 times and reaches a reasonable value of \(\sim 0.2\%\). The difference between \(E_m\) and the global minimum energy is reduced by three times. The probability of hitting the energy range \(E_m \in [-1, -0.99]\), which is close to the global minimum, increases by more than 500 times.

Moreover, owing to the Kernighan-Lin algorithm, it was confirmed that the efficiency of our algorithm grows as the problem dimension \(N\) increases.

REFERENCES