The Mix-Matrix Method in the Problem of Binary Quadratic Optimization

Iakov Karandashev and Boris Kryzhanovsky

Center of Optical Neural Technologies
Scientific Research Institute for System Analysis, Russian Academy of Sciences
Moscow, Russia
Yakov.Karandashev@phystech.edu, kryzhanov@mail.ru

Abstract. In the paper we deal with the NP-complete problem of minimization a quadratic form of N binary variables. The minimization approach based on extensive random search is considered. To increase the efficiency of the random-search algorithm, we vary the attraction area of the deepest minima of the functional by changing the matrix T it is based on. The new matrix M, called *mix-matrix*, is a mixture of T and T^2. We demonstrate that such a substitution brings about changes of the energy surface: deep minima displace very slightly in the space (the Hemming distance of the shift is of about $0.01*N$), they become still deeper and their attraction areas grow significantly. At the same time the probability of finding close to optimal solutions increases abruptly (by 2-3 orders of magnitude in case of a 2D Ising model of size 12×12 and in case of dense instances of size 500).

Keywords: quadratic optimization, binary optimization, combinatorial optimization, area of attraction, local search, random search, energy landscape transformation, mix-matrix.

1 Introduction

The goal of this paper is to improve the efficiency of a random search procedure used to solve binary minimization problems.

It is well known that there is no polynomial algorithm for solving this problem, i.e., it is impossible to find a global minimum in polynomial time (the problem is NP-hard). 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 $E(S)$. 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).

In previous work [4], we consider the simplest transformation, namely, the raising of T to the power $k = 2,3,...$. This approach was found to be fairly productive: due to the landscape transformation, the spectrum of found minima is strongly shifted

A.E.P. Villa et al. (Eds.): ICANN 2012, Part I, LNCS 7552, pp. 41–48, 2012.

towards the deep side and the probability of finding the global minimum increases by 10^3 times, but it turned out to be not reliable.

In present paper, we suggest to use a mix-matrix M, i.e., a mixture of T and T^2. We claim that this yields a more reliable approach.

The efficiency of the algorithm proposed is rigorously substantiated only for "random" matrices, whose elements generated as independent random variables. The application of the algorithm to matrices of other types is heuristic. Later on the experimental results will be given for matrices of two types: uniform matrices and matrices of 2D Ising model.

2 Problem Definition and Minimization 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, $s_{mi} = \pm 1$, $i = 1, 2, ..., N$, that minimizes the energy functional $E(S)$:

$$E(S) = -\frac{1}{\sigma_T N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} T_{ij} s_i s_j \tag{1}$$

where σ_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 (asynchronous) dynamics of Hopfield model. It can be described as follows. A start configuration S is chosen and then it is locally improved by flipping single unstable neurons. This procedure is sequentially applied to all the neurons until the network converges to a stable state S_m.

NP-complete problems are known to have a huge number of local minima. In order to find a global one we have to use the random search, i.e.: given an arbitrarily initial state of the network, the nearest local minimum is found and this procedure is repeated until a minimum with an acceptable depth is found.

3 Preliminaries

Before transforming the energy landscape, we repeat the basic relations (described in detail in [4-6]) 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 be the configuration corresponding to the global minimum $E_0 = E(S_0)$. We extract from T the term T_0 that is responsible for the formation of this minimum:

$$T = T_0 + T_1, \quad T_0 = r_0 \sigma_T S_0^+ S_0 \tag{2}$$

The 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 approximately $E_0 = -r_0$.

It was shown in [4-6] that any vector S_m is a minimizer of functional (1) if its weight r_m is larger than the critical value $r_c \approx 1.35/\sqrt{N}$. This assertion is concerned primarily with the point S_0, which by definition is a minimizer of functional (1) and satisfies the relations

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

The second necessary relation obtained in [6] is that, as the depth of minimum E_0 increases, its width increases as well and, accordingly, the probability of finding this minimum grows as $P(E_0) \sim \exp\left(-NE_c^2 / E_0^2\right)$.

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.

4 The Algorithm

In this section we describe the proposed minimization algorithm. The main idea underlying the algorithm is the transformation of energy landscape of the functional. The surface described by the quadratic form $E(S)$ can be transformed only by transforming the underlying matrix.

Let's introduce the mix-matrix M:

$$M = \frac{1-z}{\sigma_T} T + \frac{z}{\sigma_{2T}} T^2 \tag{3}$$

(where T^2 is obtained by raising T to the second power and setting the diagonal elements equal to zero, σ_T and σ_{2T} are the standard deviations of matrices T and T^2 respectively) and substitute it into (1). Changing the parameter z from 0 to 1, we pass from the matrix T to T^2. Accordingly, the landscape described by $E(S)$ is transformed into that described by:

$$E_z(S) = -\frac{1}{\sigma_M N^2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} M_{ij} s_i s_j \tag{4}$$

where σ_M is the standard deviation of M_{ij}. 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.

Now we can propose the following minimization algorithm.

Firstly, we choose a value z, construct the mix-matrix (3) and accordingly the functional $E_z(S)$. Then we start the minimization procedure consisting of two steps:

- At the first step, a descent over $E_z(S)$ is performed and a configuration S_{zm} is found that minimizes $E_z(S)$.
- The second step involves correction, namely, from the point S_{zm}, we descend over $E(S)$ to the nearest minimum S_m of $E(S)$.

The descent over landscape is performed as described above (see Section 2).

In previous work [4] we consider the simplest transformation, namely, when $M = T^k$, $k = 2,3,4,5$. It was shown that the optimal value of power is $k = 3$. In this case the probability of finding global minima increases by 3 orders of magnitude for the most part (about 70% of instances). But sometimes (the rest 30%) it may decrease up to zero due to vanishing a minimum near S_0.

The proposed in present paper mix-matrix (3), i.e., a mixture of T and T^2, yields a more reliable approach.

We will show that at $z \approx 0.5$ the proposed transformation leads to significant increase of the global minimum depth, while the shift from the minimum is smaller ($1-2\%$ of N) than in case $M = T^3$ (3% of N).

5 Correctness of the Algorithm

5.1 The Deepening of the Minima

Let us show that the landscape transformation leads to a deeper minimum. Consider the energy $E_{z0} = E_z(S_0)$ at the point S_0. Following (2), the mix-matrix M is represented as

$$M = (1-z)\frac{T_0 + T_1}{\sigma_T} + z\frac{T_0^2 + T_0T_1 + T_1T_0 + T_1^2}{\sigma_{2T}}$$

In view of $S_0 T_1 S_0^+ = 0$ and $\sigma_M^2 = (1-z)^2 + z^2$, we then derive from (4) that in the limit of $N >> 1$, E_{z0} can be viewed as a normally distributed quantity with the mean value $\overline{E}_{z0}$ and the relatively small noise R of standard deviation $\sigma_R = 1/N$. The ratio:

$$\frac{\overline{E}_{z0}}{E_0} = \frac{(1-z) + z\sqrt{N}r_0}{\sqrt{(1-z)^2 + z^2}} \tag{5}$$

shows how many times the average value of the modified functional at point S_0 more than the initial functional value at the same point. Taking into account $\sqrt{N}r_0 \approx 1.35$, it is obvious that at any value of z expression (5) is larger than unit, hence when $N >> 1$ one can be sure that the minimum becomes deeper. Fig. 1 confirms this. The largest deepening ($E_{z0} \approx 1.6E_0$) is observed at $z \approx 0.6$.

5.2 The Shift of the Minima

Under the landscape transformation the mean shift can be represented as

$$d = N \cdot P,$$

where $P = \Pr\{s_{0i}h_i^{(z)} < 0 \mid s_{0i}h_i > 0\}$ is the probability that the directions of the spin s_{0i} and the local field $h_i^{(z)}$ do not coincide. Omitting the unnecessary constants, the value $s_{0i}h_i^{(z)}$ can be represented as

$$h_i^{(z)}s_{i0} = (1-z)Nr_0 + zN^{3/2}r_0^2 + H$$

where

$$H = \sum_{i=1}^{N}\left(\frac{(1-z)T_1}{\sigma_T} + \frac{z(Nr_0\sigma_T T_1 + T_1^2)}{\sigma_{2T}}\right)_{ij} s_{0i}s_{0j}$$

Therefore, P is expressed in terms of the error function:

$$P = \frac{1}{2\Phi(\gamma)\sqrt{2\pi}}\int_0^{\infty} dx\, e^{-\frac{1}{2}(x-\gamma)^2}\left(1-\Phi\left(\alpha\sigma x\right)\right) \tag{6}$$

where $\Phi(\cdot)$ is the probability integral and

$$\gamma = \sqrt{N}r_0/\sigma \approx 1.9, \quad \sigma = \sqrt{\frac{1}{\sigma_T^2 N}\sum_{i=1}^{N}h_i^2} \approx 0.7, \quad \alpha = \frac{1-z}{z} + \sqrt{N}r_0 \ .$$

Note that at $z = 0$ the functional $E_{z=0}(S)$ coincides with initial $E(S)$ and therefore the shift is absent, this agrees ($d = N \cdot P = 0$) with (7).

The formula (6) describes a monotone increase of the minimum shift with growing z in view of enlarging functional transformation. This corresponds to a common sense and is proved by experiment (see fig. 2).

Expressions (5)–(6) 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 (5) is larger for a larger

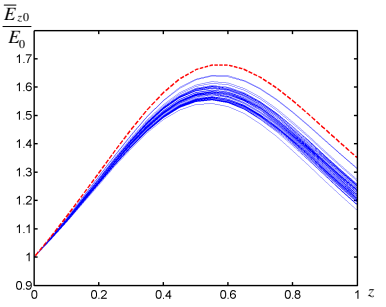

Fig. 1. The decrease of energy in the point S_0 (global minimum) due to energy landscape transformation (mix-matrix with T^2). The dashed line is theoretical (5). Other lines are experimental for 50 random instances with uniform matrices.

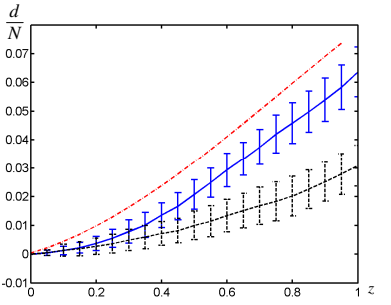

Fig. 2. The shift (in bits) of the global minimum as a function of z (mix with T^2). The curves with error bars were obtained by experiment for two types of matrices: matrix with uniformly distributed elements (solid line) and 2D Ising matrices (dashed lines). The dash-dot line is theoretical (6).

initial depth $|E_0| \approx r_0$. This means that the spectrum of minima found by the algorithm shifts considerably toward the global minimum. The spatial minima displacements caused by the transformation are relatively small: it follows from (6) that the smallest shifts are expected for the deepest minima.

6 Results

The efficiency of the two-step descent algorithm was verified for z ranging from 0 to 1 for matrices of size $N = 100 - 500$ of two types:

- full matrices with random elements uniformly distributed within $(-1;1)$;
- sparse matrices of 2-dimensional Ising model [2].

During numerical experiments we built a mix-matrix for different values of z from 0 to 1 equally spaced with $\Delta z = 0.05$. The results were averaged over 50 random instances of each size and type.

Each experiment included $N_{runs} = 10^6$ runs. Each run resulted in a local minimum. We chose two parameters to trace: the mean energy E_{mean} of the minima found and the probability of finding a minimum in the interval of energy close to the global one $E \in [-1;-0.99]$, where -1 corresponds to E_0.

In experiments, we try to use not only the mix-matrix (3) but also analogous mix with T^3. The numerical results are shown in figs. 3-4.

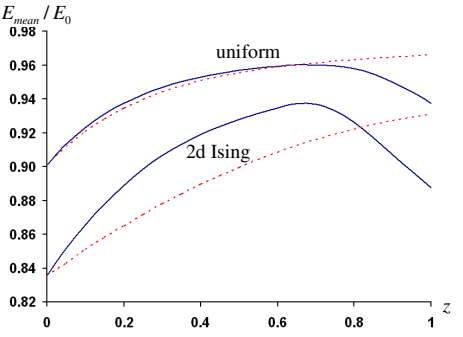

Fig. 3. The mean value E_{mean} of energy of local minima found with the proposed two-step algorithm. The solid lines are for mix-matrices with T^2. The dashed lines are for mix-matrices with T^3. The curves are drawn for two types of matrices: uniform matrices (on top) and 2d Ising matrices. The value of E_{mean} is divided by the energy of global minimum E_0 and does not depend on the problem dimension N.

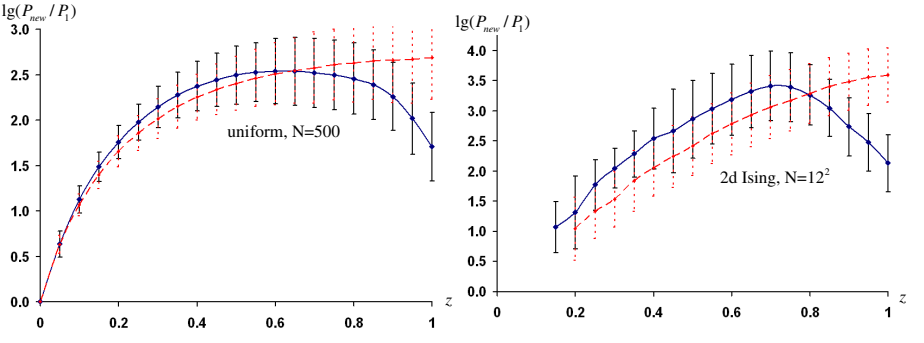

Fig. 4. The common logarithm of the ratio of probabilities of hitting the energy interval $E \in [-1, -0.99]$. The solid lines are for mix-matrices with T^2. The dashed lines are for mix-matrices with T^3. In the left panel the results for uniform matrices of $N = 500$ ($P_1 \approx 3 \cdot 10^{-5}$). In the right panel the results for 2d Ising matrices of $N = 144$ (. $P_1 \approx 2.6 \cdot 10^{-7}$). Note, that when z is too small, the algorithm does not find the global minimum in some instances, so the points are missed.

Fig. 4 shows how many times increases the probability of finding minima with energy differed from the global one less than 1%. For demonstration purpose we chose the maximal possible problem dimensions, which we can cope with. For 2D Ising matrices the probability of finding minima of energy $E \in [-1; -0.99]$ is not greater than $P_1 = 3 \cdot 10^{-7}$ for $N = 12 \times 12$. For uniform matrices the maximal dimension is $N = 500$ (the probability $P_1 = 3 \cdot 10^{-5}$). The probability obtained with the proposed algorithm was denoted by P_{new}. As we can see from fig. 4, the difference between P_{new} and P_1 turned out to be enormous – approximately 3 orders of magnitude.

An interesting fact is that for uniform matrices the T^2 and T^3 curves almost coincide (see. fig. 3-4), and they start to diverge when $z > 0.7$ only. For Ising matrices we have another picture: mix with T^2 prevails over mix with T^3 up to $z \approx 0.8$ and after that vice versa .

It can be also seen from fig. 4 that with increasing z the dispersion rises, and this can lead to the instability of the algorithm, i.e., the transformation may change the search procedure for the worse in some cases.

7 Conclusion

In the paper a new approach for the problem of binary quadratic optimization was described. It allows improving the random search technique for finding the optimal and suboptimal solutions.

The key parameter of the algorithm is the number $z \in (0,1)$. Varying this parameter, we can change the depth and the attraction area of the global and other minima. We showed theoretically that the best choice of this parameter is $z \approx 0.6$.

The experiment showed a good correspondence to the theory. Indeed, at $z = 0.7$ we significantly improved the random search and succeeded in decreasing the value $(E_0 - E_{mean})/E_0$ (difference between the mean energy of found minima and global one) by half. Due to the proposed method the probability of finding suboptimal solutions with energy differed from the optimum less than 1% increases by 2.5 orders of magnitude for (dense) uniform matrices of dimension $N = 500$ and by more than 3 orders for (sparse) matrices of Ising model of dimension $N = 12 \times 12$.

Acknowledgment. The work was supported by the program of the Presidium of the Russian Academy of Sciences (project 2.15) and in part by the Russian Basic Research Foundation (grant 12-07-00295).

References

1. Kernighan, B.W., Lin, S.: An Efficient Heuristic Procedure for Partitioning Graphs. Bell System Tech. Journal 49, 291–307 (1970)
2. Hartmann, A.K., Rieger, H.: New Optimization Algorithms in Physics. Wiley, Weinheim (2004)
3. Dang, C., Ma, W., Liang, J.: A deterministic annealing algorithm for approximating a solution of the min-bisection problem. Neural Networks 22(1), 58–66 (2009)
4. Karandashev, Y.M., Kryzhanovsky, B.V.: Transformation of Energy Landscape in the Problem of Binary Minimization. Doklady Mathematics 80(3), 927–931 (2009)
5. Amit, D.J., Gutfreund, H., Sompolinsky, H.: Spin-glass models of neural networks. Phys. Rev. A 32, 1007–1018 (1985); Annals of Physics 173, 30–67 (1987)
6. Kryzhanovsky, B.V., Kryzhanovsky, V.M.: The shape of a local minimum and the probability of its detection in random search. LNEE, vol. 24, pp. 51–61 (2009)