Cybernetics and Systems Analysis, Vol. 35, No. 1, 1999

THE METHOD OF GLOBAL EQUILIBRIUM SEARCH

V. P. Shilo

UDC 519.854.33

The development and analysis of methods of solution of discrete optimization problems is now being given increasing attention all over the world. A method of adaptive random search for the solution of discrete global optimization problems is offered in this paper. This method is hereinafter called the global equilibrium search (GES). It is described as applied to problems of integer linear programming with Boolean variables (ILP BV). The method offered is conceptually related to the simulated annealing [1], which, despite successful application for solution of many complex optimization problems [10], has an asymptotic effectiveness lower [2] than that of even the conceptually trivial method of repeated local random search (RLMS). The GES method preserves all the advantages of the simulated-annealing method and, at the same time, has a higher asymptotic efficiency. The results of numerical experiments using this method allow us to speak about revival of the ideas of the Boltzmann optimization, which is, from our point of view, the most natural since it was borrowed from nature.

We will consider an ILP BV problem that is formulated as follows: it is necessary to minimize

$$f(x) = \sum_{j=1}^{n} c_j x_j \tag{1}$$

with constraints

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = 1, \dots, m, \tag{2}$$

$$0 \leqslant x_j \leqslant 1, \quad j = 1, \dots, n, \tag{3}$$

where

$$x_j \text{ are integers}, \; j = 1, \dots, n. \tag{4}$$

It is assumed that $c_j \geqslant 0$, $j = 1, \dots, n$.

Let a set S consist of admissible solutions of problem (1)–(4) and, for any $\mu \geqslant 0$,

$$Z(\mu) = \sum_{x \in S} \exp(-\mu f(x)).$$

We will define a random vector $\xi(\mu, \omega)$:

$$\pi(x, \mu) = P\{\xi(\mu, \omega) = x\} = \exp(-\mu f(x)) / Z(\mu), \quad x \in S. \tag{5}$$

Translated from Kibernetika i Sistemnyi Analiz, No. 1, pp. 74-81, January-February, 1999. Original article submitted September 22, 1998.

 1060-0396/99/3501-0068$22.00

Distribution (5) is known in statistical physics as the Boltzmann distribution. The annealing method is related to the same distribution: the stationary probability of being at the point x of a Markovian chain generated by this method is defined by expression (5).

Denote by S_j^1 and S_j^0 the sets consisting of admissible solutions of problem (1)–(4), whose jth component is equal, respectively, to 1 or 0:

$$S_j^1 = \{x : x \in S,\ x_j = 1\},\quad S_j^0 = \{x : x \in S,\ x_j = 0\}.$$

Let us define

$$Z_j^1(\mu) = \sum_{x \in S_j^1} \exp(-\mu f(x)),$$

$$Z_j^0(\mu) = \sum_{x \in S_j^0} \exp(-\mu f(x)),$$

$$p_j(\mu) = P\{\xi_j(\mu, \omega) = 1\},\quad j = 1, \ldots, n.$$

It is easy to see that

$$S = S_j^1 \cup S_j^0,\quad Z(\mu) = Z_j^1 + Z_j^0,\quad j = 1, \ldots, n,$$

$$p_j(\mu) = Z_j^1(\mu) / Z(\mu),\quad j = 1, \ldots, n,$$

$$\langle f \rangle_\mu = M f(\xi(\mu, \omega)) = -\frac{\partial}{\partial \mu} \ln Z(\mu) = \sum_{j=1}^{n} c_j\, p_j(\mu).$$

Let us calculate the derivative of the probability $p_j(\mu)$:

$$\frac{\partial p_j(\mu)}{\partial \mu} = \frac{\partial}{\partial \mu} \frac{Z_j^1(\mu)}{Z(\mu)} = \frac{Z(\mu)\dfrac{\partial Z_j^1(\mu)}{\partial \mu} - Z_j^1(\mu)\dfrac{\partial Z(\mu)}{\partial \mu}}{Z(\mu)^2} = p_j(\mu)\left(\langle f \rangle_\mu - \langle f \rangle_\mu^{x_j = 1}\right). \tag{6}$$

It is easy to see also that

$$\frac{\partial p_j(\mu)}{\partial \mu} = p_j(\mu)(1 - p_j(\mu))\left(\langle f \rangle_\mu^{x_j = 0} - \langle f \rangle_\mu^{x_j = 1}\right), \tag{7}$$

where the conditional expectations are defined as

$$\langle f \rangle_\mu^{x_j = 1} = M(f(\xi(\mu, \omega)) \mid x_j = 1) = \frac{\sum_{x \in S_j^1} f(x) \exp(-\mu f(x))}{Z_j^1(\mu)},$$

$$\langle f \rangle_\mu^{x_j = 0} = M(f(\xi(\mu, \omega)) \mid x_j = 0) = \frac{\sum_{x \in S_j^0} f(x) \exp(-\mu f(x))}{Z_j^0(\mu)}.$$

The following obvious equality was used earlier:

$$\langle f \rangle_\mu = p_j(\mu) \langle f \rangle_\mu^{x_j = 1} + (1 - p_j(\mu)) \langle f \rangle_\mu^{x_j = 0}, \quad j = 1, \ldots, n.$$

We will introduce the independent random variables $\eta_j(\mu, \omega)$, $j = 1, \ldots, n$:

$$P\{\eta_j(\mu, \omega) = 1\} = p_j(\mu), \quad P\{\eta_j(\mu, \omega) = 0\} = 1 - p_j(\mu)$$

and a random vector $\zeta(\mu, \omega) = \left(\zeta_0(\mu, \omega), \zeta_1(\mu, \omega), \ldots, \zeta_m(\mu, \omega)\right)$:

$$\zeta_0(\mu, \omega) = \sum_{j=1}^{n} c_j \eta_j(\mu, \omega), \quad \zeta_i(\mu, \omega) = \sum_{j=1}^{n} a_{ij} \eta_j(\mu, \omega), \quad i = 1, \ldots, m.$$

Under definite assumptions it is possible to show the asymptotic normality of the normalized random variables $\zeta_i(\mu, \omega)$, $i = 0, \ldots, m$. For the sake of simplicity, we formally assume normality of these quantities and, therefore, normality of the conditional distribution of the quantity $\zeta_0(\mu, \omega)$ under the conditions $\zeta_i(\mu, \omega) = b_i$, $i = 1, \ldots, m$. In this case we can calculate the conditional expectation [5]

$$M\left(\zeta_0(\mu, \omega) \mid \zeta_i(\mu, \omega) = b_i\right) = M\zeta_0(\mu, \omega) + \sum_{i=1}^{m} r_i\left(b_i - M\zeta_i(\mu, \omega)\right)$$

$$= \sum_{j=1}^{n} c_j p_j(\mu) + \sum_{i=1}^{m} r_i(\mu) \left(b_i - \sum_{j=1}^{n} a_{ij} p_j(\mu)\right) = \sum_{j=1}^{n} c_j p_j(\mu),$$

where $r_i(\mu)$, $i = 1, \ldots, m$, are the regression coefficients that are determined from the following systems of linear equations:

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{kj} a_{ij} p_j(\mu)(1 - p_j(\mu))\right) r_i(\mu)$$

$$= \sum_{j=1}^{n} c_j a_{ij} p_j(\mu)(1 - p_j(\mu)), \quad k = 1, \ldots, m. \tag{8}$$

It is natural to assume that

$$\langle f \rangle_\mu^{x_j = 0} - \langle f \rangle_\mu^{x_j = 1} = M\left(\zeta_0(\mu, \omega) \mid \zeta_i(\mu, \omega) = b_i, \ i = 1, \ldots, m, \ x_j = 0\right)$$

$$- M\left(\zeta_0(\mu, \omega) \mid \zeta_i(\mu, \omega) = b_i, \ i = 1, \ldots, m, \ x_j = 1\right).$$

Then we will obtain

$$\langle f \rangle_\mu^{x_j = 0} - \langle f \rangle_\mu^{x_j = 1} = \sum_{k=1, k \neq j}^{n} c_k p_k(\mu) + \sum_{i=1}^{m} r_1(\mu) \left(b_i - \sum_{k=1, k \neq j}^{n} a_{ik} p_k(\mu)\right)$$

$$- c_j - \sum_{k=1, k \neq j}^{n} c_k p_k(\mu) - \sum_{i=1}^{m} r_i(\mu) \left(b_i - a_{ij} - \sum_{k=1, k \neq j}^{n} a_{ik} p_k(\mu)\right) = \sum_{i=1}^{m} r_1(\mu) a_{ij} - c_j.$$

Therefore,

$$\frac{\partial p_j(\mu)}{\partial \mu} = p_j(\mu)(1 - p_j(\mu)) \left(\sum_{i=1}^{m} r_i(\mu)\, a_{ij} - c_j \right), \quad j = 1, \ldots, n. \tag{9}$$

Formula (9) defines a variant of a continuous analog of the interior-point method for the problem of linear programming with two-sided constraints (1)–(3) ([6, pp. 7–12], [13, pp. 139–150]).

Since, on the strength of (8), the relations below are fulfilled,

$$\sum_{j=1}^{n} a_{kj} \frac{\partial p_j(\mu)}{\partial \mu} = \sum_{j=1}^{n} a_{kj}\, p_j(\mu)(1 - p_j(\mu)) \left(\sum_{i=1}^{m} r_i(\mu)\, a_{ij} - c_j \right)$$

$$= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{kj}\, a_{ij}\, p_j(\mu)(1 - p_j(\mu)) \right) r_i(\mu)$$

$$- \sum_{j=1}^{n} c_j\, a_{kj}\, p_j(\mu)(1 - p_j(\mu)) = 0, \; k = 1, \ldots, m,$$

if limitations (2) are fulfilled with $\mu = 0$

$$\sum_{j=1}^{n} a_{ij}\, p_j(0) = b_i, \quad i = 1, \ldots, m,$$

they will be fulfilled for any $\mu > 0$ and $p_j(\mu)$, $j = 1, \ldots, n$, obtained after integration of system (9).

Let us calculate the derivative of the mean value of the objective function in the annealing method:

$$\frac{\partial \langle f \rangle_\mu}{\partial \mu} = \frac{\partial}{\partial \mu} \left(\frac{\sum_{x \in S} f(x) \exp(-\mu f(x))}{Z(\mu)} \right)$$

$$= - \frac{\sum_{x \in S} f^2(x) \exp(-\mu f(x))}{Z(\mu)} + \left(\frac{\sum_{x \in S} f(x) \exp(-\mu f(x))}{Z(\mu)} \right)^2 .$$

It is equal to the variance of the values of the objective function with the negative sign.

On the other hand,

$$\frac{\partial \langle f \rangle_\mu}{\partial \mu} = \frac{\partial}{\partial \mu} \left(\sum_{j=1}^{n} c_j\, p_j(\mu) \right) = \sum_{j=1}^{n} c_j\, p_j(\mu)(1 - p_j(\mu)) \left(\sum_{i=1}^{m} r_i(\mu)\, a_{ij} - c_j \right)$$

$$= - \left(\sum_{j=1}^{n} c_j^2\, p_j(\mu)(1 - p_j(\mu)) + \sum_{i=1}^{m} r_i(\mu) \left(\sum_{j=1}^{n} c_j\, a_{ij}\, p_j(\mu)(1 - p_j(\mu)) \right) \right)$$

$$= D^2(\zeta_0(\mu, \omega) \mid \zeta_i(\mu, \omega) = b_i, \; i = 1, \ldots, m).$$

The penultimate expression with the minus sign represents the conditional variance of the random normally distributed quantity $\zeta_0(\mu, \omega)$.

Thus, it has been established that under conditions providing asymptotic normality of the random variables $\zeta_i(\mu, \omega)$, $i = 0, \ldots, m$, the annealing method is similar to the interior-point method, which is one of the most effective methods in solving linear programming problems (as was noted in [7], the Karmarkar methods [8] and Dikin–Barnes method [9] are quite similar to them).

It is easy to obtain an integral solution of systems (6)–(7):

$$p_j(\mu) = p_j(0) \exp \left(\int_0^{\mu} \left(\langle f \rangle_{\mu} - \langle f \rangle_t^{x_j = 1} \right) dt \right) \tag{10}$$

and

$$p_j(\mu) = \frac{1}{1 + \dfrac{1 - p_j(0)}{p_j(0)} \exp \left(- \int_0^{\mu} \left(\langle f \rangle_t^{x_j = 0} - \langle f \rangle_t^{x_j = 1} \right) dt \right)} . \tag{11}$$

Using (9), we obtain the following formula:

$$p_j(\mu) = \frac{1}{1 + \dfrac{1 - p_j(0)}{p_j(0)} \exp \left(- \int_0^{\mu} \left(\sum_{i=1}^{m} r_i(t)\, a_{ij} - c_j \right) dt \right)}$$

$$= \frac{1}{1 + \dfrac{1 - p_j(0)}{p_j(0)} \exp \left(\mu c_j - \sum_{i=1}^{m} a_{ij} \int_0^{\mu} r_i(t) dt \right)} . \tag{12}$$

Let us describe a schematic diagram of the method of global equilibrium search for problem (1)–(4). We set the numbers K and $\mu_0 < \mu_1 < \ldots < \mu_K$. Let the set $\widetilde{S}$ be some subset of solutions of problem (1)–(4) obtained by the GES method and

$$\widetilde{S}_j^1 = \left\{ x : x \in \widetilde{S},\ x_j = 1 \right\}, \quad \left\{ \widetilde{S}_j^0 = \{ x : x \in \widetilde{S},\ x_j = 0 \right\}.$$

We define the following quantities for $k = 0, \ldots, K$:

$$\widetilde{Z}_k = \sum_{x \in \widetilde{S}} \exp\{-\mu_k f(x)\}, \quad \widetilde{F}_k = \sum_{x \in \widetilde{S}} f(x) \exp\{-\mu_k f(x)\}, \quad \widetilde{E}_k = \frac{\widetilde{F}_k}{\widetilde{Z}_k},$$

$$\widetilde{Z}_{kj}^1 = \sum_{x \in \widetilde{S}_j^1} \exp\{-\mu_k f(x)\}, \quad j = 1, \ldots, n,$$

$$\widetilde{F}_{kj}^1 = \sum_{x \in \widetilde{S}_j^1} f(x) \exp\{-\mu_k f(x)\}, \quad j = 1, \ldots, n,$$

$$\widetilde{E}_{kj}^1 = \frac{\widetilde{F}_{kj}^1}{\widetilde{Z}_{kj}^1}, \quad j = 1, \ldots, n, \quad \widetilde{Z}_{kj}^0 = \sum_{x \in \widetilde{S}_j^0} \exp\{-\mu_k f(x)\}, \quad j = 1, \ldots, n,$$

$$\widetilde{F}_{kj}^0 = \sum_{x \in \widetilde{S}_j^0} f(x) \exp\{-\mu_k f(x)\}, \quad j = 1, \ldots, n, \quad \widetilde{E}_{kj}^0 = \frac{\widetilde{F}_{kj}^0}{\widetilde{Z}_{kj}^0}, \quad j = 1, \ldots, n,$$

$$\tilde{p}_{kj} = \frac{1}{1 + \frac{1-\tilde{p}_{0j}}{\tilde{p}_{0j}} \exp\left\{ -0.5 \sum_{i=0}^{k-1} \left(\tilde{E}_{ij}^{0} + \tilde{E}_{i+1\,j}^{0} - \tilde{E}_{ij}^{1} - \tilde{E}_{i+1\,j}^{1} \right) (\mu_{i+1} - \mu_i) \right\}},$$

$$j = 1, \ldots, n. \tag{13}$$

Expression (13) is obtained by application of the trapezoid rule to the integral in formula (11). Let us present the schematic diagram of the GES method.

```
procedure global equilibrium search
  repeat
    repeat for k = 0 to K − 1
      determination of the number of tosses n_k in random search
        repeat for t = 1 to n_k
          random search with the probabilities p̃_kj, j = 1, ..., n, — the solution
            x is obtained
          if (x ∉ S̃) S̃ = S̃ ∪ x
          calculation of the quantities Ẽ¹_kj, Ẽ⁰_kj, k = 1, ..., K, j = 1, ..., n,
      calculation of the probabilities p̃_k+1 j, j = 1, ..., n, by formula (13)
    determination of the set S̃
  until (stopping criterion is satisfied)
```

Let us consider this scheme in more detail.

The Probabilities of Random Search. The probabilities $p_j(0)$, $j = 1, \ldots, n$, can be obtained for $\mu = 0$ with the help of statistical simulation in the following way. The equiprobable permutations of the indices of the problem variables are generated a desired number of times. Then, to solve initial problem (1)–(4), the algorithm of sequential assignment of units using the obtained permutation is applied. The calculated frequency of a single value of the variable x_j can be used as a good estimate for the quantity $p_j(0)$, $j = 1, \ldots, n$. If $\mu_0 = 0$, then $\tilde{p}_{0j} = p_j(0)$, $j = 1, \ldots, n$. For the values of μ that are not very large, we can calculate $p_j(\mu)$ through a numerical integration of expression (12) and by putting $\mu_0 = \mu$, $\tilde{p}_{0j} = p_j(\mu)$. For setting μ_k, the choice of $\alpha > 1$ and $\mu_k = \alpha^k \mu_0$ and the condition where, at the "temperature" μ_k (see [1]), the probability $\tilde{p}_{kj}$ would approximately be equal to the best obtained solution seems to be natural.

If an improving solution is not found in the temperature cycle in k, then it is advisable to increase n_k; if an improving solution is obtained, then it is expedient not to change (to reduce) n_k.

Definition of the Set $\tilde{S}$. If an improving solution is found in the temperature cycle, then it is advisable to leave in $\tilde{S}$ only the improving solution or some best solutions. But if an improving solution is not obtained, then preservation of $\tilde{S}$ can be more preferable. To store the obtained solutions (the set $\tilde{S}$), the hashing technique can be used [11].

Let us point out that it is sometimes expedient to use for "acceleration" of the GES method a fast and effective approximate method, e.g., the method of the drop vector [12]. Then, before the first calculation of the temperature cycle, the set $\tilde{S}$ will consist of a solution obtained by the "acceleration" algorithm.

Efficiency of the Algorithm. Clearly, the GES method is more effective for solution of the knapsack problem than the annealing method. This method produces an exact solution in a finite number of steps. The results of test calculations testify to the high performance of this method (this is discussed in more detail in a special publication). For example, out of 400 problems described in [4], the GES method has found exact solutions of all the problems at given stopping criteria, while the RTS method (the reactive taboo method) has found exact solutions of 259 problems, and 317 with a 1% accuracy. The results of solution of problems of a large dimension (500×500) are even more impressive. Improving solutions are obtained for all the problems at a value of the stopping criterion 1020 times less than in [4].

All the problems offered at the All-Union Competition RANETs-85 were solved in 15 min on a Pentium 166 MMX computer, with the solutions obtained in all the problems being not worse than the declared one. Moreover, improving solutions were obtained for many problems. In other words, it was not possible, based on the totality of the solutions presented to this competition, to distinguish this algorithm from the exact one.

Let us point out that not only the computational performance but also the command property is of primary importance for the GES method. By the command property will be meant the capability of the method to effectively use the solutions obtained by other algorithms. For example, the command figure is equal to zero for the classical local search method. The command properties allow us (this is discussed in more detail in subsequent papers) to use the GES method rather effectively in the case of parallel computations, which is especially important for optimization on the Internet with the use of several computers.

The knapsack problem is selected in this paper as a convenient model. From our point of view, any optimization problem that satisfies certain requirements for the structure of optima can be effectively solved within the framework of the approach proposed in the present paper.

REFERENCES

1. S. Kirkpatrick, C. D. Gelatti, and M. P. Vecchi, "Optimization by simulated annealing," Science, **220**, 671-680 (1983).
2. A. G. Ferreira and J. Zerovnik, "Bounding the probability of success of stochastic methods for global optimization," Comp. Math. Applic., **25**, 1-8 (1993).
3. M. Ohlsson, C. Peterson, and B. Soderberg, "Neural networks for optimization problems with inequality constraints: the knapsack problem," Neural Computation, No. 5, 331-339 (1993).
4. R. Battiti and G. Tecchiolli, Local Search with Memory, Benchmarking RTS Preprint, Univ. di Trento (1994).
5. H. Cramer, Mathematical Methods of Statistics [Russian translation], Mir, Moscow (1975).
6. I. I. Dikin and V. I. Zorkal'tsev, Iterative Solution of the Problems of Mathematical Programming [in Russian], Nauka, Novosibirsk (1980).
7. N. Z. Shor and S. I. Stetsenko, Square-law Extremum Problems and Nondifferentiable Optimization [in Russian], Naukova Dumka, Kiev (1989).
8. N. Karmarkar, "A new polynomial time algorithm for linear programming," Combinatorica, **4**, No. 4, 373-395 (1984).
9. E. R. Barnes, "A variation on Karmarkar's algorithm for solving linear programming problems," Math. Progr., **36**, No. 2, 174-182 (1986).
10. E. Aarts and J. Korst, Simulated Annealing and Boltzmann Machines, John Wiley and Sons, New York (1989).
11. D. Knuth, The Art of Computer Programming, Vol. 3, Sorting and Searching [Russian translation], Mir, Moscow (1978).
12. I. V. Sergienko, Mathematical Models and Methods of Solving Discrete Optimization Problems [in Russian], Naukova Dumka, Kiev (1988).
13. V. I. Zorkal'tsev, The Method of Least Squares [in Russian], Nauka, Novosibirsk (1995).