Control and Cybernetics

vol. **33** (2004) No. 4

Hybrid search for optimum in a small implicitly defined region

by

A. Žilinskas

Institute of Mathematics and Informatics
4 Akademijos str., Vilnius, LT-08663, Lithuania
antanasz@ktl.mii.lt

Abstract: We consider optimization problems with a small implicitly defined feasible region, and with an objective function corrupted by irregularities, e.g. small noise added to the function values. Known mathematical programming methods with high convergence rate can not be applied to such problems. A hybrid technique is developed combining random search for the feasible region of a considered problem, and evolutionary search for the minimum over the found region. The solution results of two test problems and of a difficult real world problem are presented.

Keywords: random search, evolutionary optimization, implicit constrains, optimal design

1. Introduction

In a standard optimization problem

$$\min_{X \in A} f(X), \tag{1}$$

the feasible region $A \subset R^n$ is defined explicitly by equalities and/or inequalities, and the objective function is defined not only inside but also outside of the feasible region. However, in some applications there occur optimization problems with implicitly defined feasible regions and objective functions not defined outside of the feasible region. The problem is especially difficult in the cases where the feasible region A is small with respect to the known rectangular enclosure B. An example of a problem from the oil industry of such a type is considered in Žilinskas, Fraga, Mackute and Varoneckas (2004) and is briefly discussed below. Let us give a formal statement of the problem. We need to find the minimum value and a minimizer for the problem (1), where $f(X)$ is not defined for $X \notin A$,

$A \subset B \subset R^n$, $B = \{X : b_{i-} \leq x_i \leq b_{i+}\}$, and A itself is defined by an indicator function $I(X)$

$$I(X) = \begin{cases} 1, & X \in A, \\ 0, & X \notin A. \end{cases}$$

Since the region A is small with respect to B, even the finding of feasible points is a challenging task. If a feasible point is known and the objective function is smooth, the construction of a descent trajectory nevertheless is difficult because of difficulty to assess a qualitative measure of constrains violation. We aim to consider the problems whose objective functions are obtained by modelling packages including solution of auxiliary problems with modest accuracy acceptable from the engineering point of view. Therefore the non-differentiability of the objective function should be assumed as well as other irregularity features. For example, in the problems considered in Fraga and Žilinskas (2003) the objective is obtained by means of computer simulation causing discontinuous behavior and some small noise added to the function values.

The properties of the problem leave few possibilities to construct a rational search procedure not only for a minimum but also for feasible points. However, the following assumptions, generalizing properties of the problem, see Žilinskas, Fraga, Mackute and Varoneckas (2004), make the situation not so hopeless. First, we assume that the feasible region A is not disjoint. Second, we assume the possibility to construct enclosures for A. A θ-enclosure of A, denoted A_θ, is defined by the inclusions $A \subset A_\theta \subset B,\ 0 < \theta < 1$, and $A_\theta \subset A_\pi,\ \pi < \theta$. We assume, that $A = A_1$, and that the ratio $vol(A)/vol(B)$ is of the order of $10^{-4} - 10^{-6}$, where $vol(\cdot)$ denotes hypervolume. The hypervolume of A_0 constitutes about one percent of the hypervolume of B.

Examples of real world problems corresponding to the assumptions above are typical for optimal design of some technological process which are modelled by means of software packages allowing limited access to the implemented models. A problem related to processing raw hydrocarbon feed stock into oil and gas products is considered below. The optimal values of some of the physical parameters of the technological process should be found. For the given vector of variable parameters the objective, profit of the process, is calculated using a modelling package. The objective is calculated using the technological parameters, obtained via modelling the physical processes, and market data. The package returns the objective function value in cases where the design parameters are feasible, and the package returns 10^{20} in the cases where the design parameters are infeasible. The reasonable intervals of the variables are known, and they constitute the hyper-rectangle set B, however, only very small part of B is feasible. Some quantitative data of the problem are discussed below; for the details we refer to Žilinskas, Fraga, Mackute and Varoneckas (2004).

2. Random search for feasible region

Because of possible irregularities, especially multimodality, of $f(\cdot)$, a search technique for a global minimizer of $f(\cdot)$ needs a representation of the whole feasible region A. Information on A in the considered problem may be obtained only via points recognized as belonging to A. Therefore the aim is to construct a sample of points uniformly distributed over A. Such a sample may be obtained by selecting feasible points from the points generated randomly with uniform distribution over B. However, such a brute force approach is very inefficient.

The idea of our algorithm is to use the contracting enclosures of A. Let a sequence $0 = \theta_1 < \theta_2 < ... < \theta_r = 1$ be fixed. A sample of points uniformly distributed over B is generated. The points belonging to A_{θ_1} are selected, and their set is denoted by Z_1. The size, configuration and orientation of the set A_{θ_1} may be estimated by analyzing Z_1. We approximate A_{θ_1} by the minimal rectangular box B_1 oriented along the principal coordinates of the set of points Z_1, and containing all points of Z_1. Further the following three steps of the algorithm are repeated:

- generating random uniformly distributed points over B_i,
- selecting Z_i, i.e. the set of points in $A_{\theta_{i+1}}$,
- constructing the box B_{i+1} oriented along the principal coordinates of Z_i.

The number of reductions r, and the values of θ_i, $i = 1, ..., r$ should be chosen depending on the character of decrease of $vol(A_\theta)/vol(B)$, as well as on the shape and hypervolume of A. Frequently, these characteristics are not known and may only be guessed by relying on the experience and heuristic considerations.

The motivation of the algorithm may be described in the language of evolutionary computing, since the algorithm indeed was inspired by the paradigm of evolutionary computing, see Bäck (1996), Goldberg (1989), Micvhalewich (1996), Schwefel (1995). Assume that a phenotype characteristics of a population correspond to B and there are no advantages for specific phenotypes. Then a current generation of the population may be modeled by the uniform distribution of random points over B irrespective of the mechanism of reproduction. The change in environmental conditions is modeled by the sequence of values of the parameter θ increasing towards 1. A new generation adapts to the environmental changes, i.e. individuals of the new generation, whose phenotype does not belong to $A_\theta \subset B$, die. However, all individuals whose phenotype corresponds to A_θ equally fit to the new conditions. Therefore, a new generation may be modeled by the points uniformly distributed over A_θ.

In Žilinskas, Fraga, Mackute and Varoneckas (2004) a new generation is produced using an evolutionary algorithm where a vector of variables X is interpreted as a floating-point encoding of a chromosome. The chromosomes of the descendants are modeled in Žilinskas, Fraga, Mackute and Varoneckas (2004) by means of crossover. The survivable parents have higher selection probability than the non survivable parents. However, this algorithm is slow since a large population should be maintained to avoid clustering of descendants around few survivable parents representing only a small subset of A_θ. The algorithm proposed in the present paper is much faster, and works well with the populations of modest size. However, its efficiency crucially depends on the approximability of sets A_θ by rectangular hyper boxes.

3. Evolutionary search for minimum

To search for the minimum a global optimization algorithm should be applied because of the supposed irregularities of the objective function. The algorithms of black box global optimization normally assume a feasible region defined by the interval constrains, Torn and Žilinskas (1989). In the considered case the only information on A is contained in the sample of points Z_r generated by the algorithm described in the previous section. A rectangle box approximating A can be constructed. However, a large part of such a box would consist of infeasible points, implying inefficiency of many global optimization algorithms. The flexibility of the evolutionary approach allows for adapting the algorithms modeling the natural evolution to the situation of interest. Consider the sample Z_r as the initial generation Y_1 of the population. Besides the strict condition for all generations Y_i, $i = 1, ..., s$ of the evolutionary search to belong to A, the evolution is driven by the fitness criterion expressed via the objective function $f(X)$. While modeling the reproduction we are not considering the mechanism of crossover, but rely on the fact that the descendants are similar to their parents according to the general phenotype characteristics. Further, we take into account that some properties important to the survival are obtained by learning from the fittest individuals. The new generation is produced from the individuals of the current generation and their descendants by means of a selection procedure. The evolutionary search algorithm is defined by the following steps, where $\beta_1 + \beta_2 + \beta_3 = 1$, $\beta_i \geq 0$:

1. $\beta_1 N$ descendants are generated with the uniform distribution over $A \bigcap D_i$, where D_i is the smallest rectangular box oriented according to the principal axis of the set Y_i and containing all these points; these points model the strongly mutated descendants who do not learn from the fittest members of the generation.

2. $\beta_2 N$ points are generated in the neighborhoods of γ-fraction of best (with respect to the value of $f(\cdot)$) points; they model the descendants of the fittest parents, and the descendants who become similar to the fittest members of the generation via learning; a neighborhood of the point X is defined as a reduced copy of D_i with the center at X,

3. $\beta_3 N$ points are generated in the neighborhood of the best point; it is a special case of the rule 2 corresponding to the fittest parent,

4. $\beta_1 N$ worst points are replaced by the points generated according to rule 1 to model the dying out of the worst fitted (e.g. oldest) fraction of

the current generation and survival of arbitrary descendants randomly avoiding the filter of natural selection.

5. The points generated according to the rules 2 and 3 and better than the median of criterion values for Y_i replace the worst points of the part of Y_i left after applying selection rule 4; i.e. the descendants survive if they are better than the worse half of the current population.

The rules 1–5 ensure the combination of diversity of the population with an elitist selection. In terms of optimization it is a combination of local and global search strategies.

4. Test examples

To illustrate the performance of the proposed algorithm we consider a problem with an analytically defined test function and an analytically defined feasible region. However, the formulas defining the problem are not used by the algorithm. Let the objective function and the enclosures of the feasible region be defined by the formulae

$$f(X) = -100(x_1+1)^2 + 121(x_2-1)^2 + 0.175617, \tag{2}$$
$$A_\theta = \{X :\ X \in B \subset R^2,\ 5|x_1+x_2| + |x_1-x_2+2| \le 2-1.8\theta\},$$

where $A = A_1$ and $B = \{X :\ -2 \le x_i \le 2,\ i = 1,2\}$. The minimum value is equal to 0 and it is attained at two points X_{01}=(-0.9473,0.9710) and X_{02}=(-1.0527,1.0290); the maximum value is larger than 0.459. The algorithm is implemented assuming B a unit hyper-cube, therefore B is re-scaled correspondingly. The minimum points in new scales are (0.263167, 0.742750), and (0.236827, 0.757259). The largest enclosure A_0 constitutes approximately 5% of B, while an estimate of the ratio $vol(A)/vol(B)$ is equal to 0.0005; it is obtained from 10^6 random trials with uniform distribution over B.

We assume that fifty uniformly distributed points can represent a not very complicated two-dimensional region with acceptable confidence level. This assumption implies the termination condition for generating the uniformly distributed points over the box B_i: it is terminated when K, the number of hits of $A_{\theta_{i+1}}$, reaches 50. We assume that the box B_i sufficiently well approximates A_{θ_i}; in fact the ratio of their volumes is about 2. Assuming $\theta = 0.25$, 0.5, 0.75, 1 we expect that the ratio of volumes of $A_{\theta_{i+1}}$ and B_i will be no less than 5%. If these assumptions are correct, then the number of trials at each stage should be on the average less than 1000, and the search of the feasible region should cost no more than 4000 calls of the subroutine implementing (2).

During the evolutionary search for the minimum we want to maintain information on the whole region, and search over prospective subregions. Therefore, the size of population at the second stage of the algorithm should be chosen larger than for the first stage where we search for the feasible region; we have fixed the size of population equal to $1.5 \cdot K = 75$. The following values of the parameters of evolutionary search have been chosen β_1=0.2, β_2=0.5, β_3=0.3, γ=0.2. We do not assume the properties of the problem, allowing to define a rigorous stopping criterion. In the real world problems a solution found is normally investigated by various methods, e.g. by means of graphical representation, and using more subtle local models. Therefore, it does not seems reasonable to search for a solution with a very high precision. The stopping criterion may be defined by choosing the maximal number of trials. We assume it is equal to 1000. The number of generations in the evolutionary search is expected to be four since several unsuccessful trials may be needed to generate a feasible point, especially when searching in the vicinities of points on the border of the feasible region.

The algorithm with the chosen parameters has been run 100 times. The average number of trials was N=3563, and N_1=905 of them were successful in the sense that the targeted region was hit. The average value of the obtained minimum evaluations was $f_0 = 5.3627 \cdot 10^{-4}$, the best value was $f_m = 1.5206 \cdot 10^{-5}$. The best point was (0.2631, 0.7428). The points representing A_1 and $A = A_4$ are shown in Fig. 1a. Fig. 1b illustrates the uniform distribution over A of the trial points of the first generation as well as the concentration of the trial points of the fourth generations in the vicinities of the two minimizers.

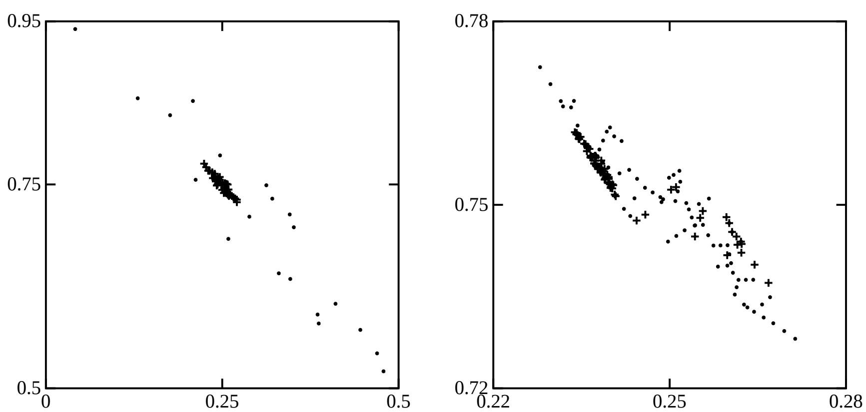

Figure 1. a)The points representing A_1 and A for the problem (2) denoted by (.) and (+) correspondingly; b)The trial points of the first (.) and the fourth (+) generations of evolutionary search for the problem (2).

The accuracy of the algorithm can be enhanced by means of the increasing K – number of points representing the feasible region; let us mention that the increasing of K increases also the size of population in evolutionary search equal to $1.5 \cdot K$. For example, by increasing K two times (to $K = 100$) we have obtained in 100 runs of the algorithm the average value of minimum evaluations equal to $f_0 = 4.16 \cdot 10^{-4}$, and the best found value equal to $f_m = 7.65 \cdot 10^{-5}$. The average number of trials as well as the number of successful trials is increased approximately two times: to N=6771 and to N_1=1765 correspondingly.

The sequence of values of the parameters θ should be chosen in such a way that the ratio of volumes of $A_{\theta_{i+1}}$ and B_i is not too small, e.g. larger than 0.05. However, the further increase of this ratio normally does not significantly enhance accuracy. Table 1 presents the minimization results for different sequences $\theta_i = i/m,\ i = 1, ..., m$.

Table 1. Influence of $\theta_i,\ i = 1, ..., m$, on the minimization results ($K = 50$)

m	N	N_1	f_0	f_m
3	3568	824	$6.28 \cdot 10^{-4}$	$4.79 \cdot 10^{-5}$
4	3563	905	$5.36 \cdot 10^{-4}$	$1.52 \cdot 10^{-5}$
5	3225	967	$5.21 \cdot 10^{-4}$	$4.51 \cdot 10^{-6}$
6	3205	1045	$5.32 \cdot 10^{-4}$	$4.70 \cdot 10^{-6}$

The parameters of evolutionary search $\beta_1,\ \beta_2,\ \beta_3$ should be chosen taking into account the irregularities of the objective function. The values of the parameters 0.2, 0.5, 0.3 above were chosen assuming irregularities of the objective function similar to those in Fraga and Žilinskas (2003), Žilinskas, Fraga, Mackute and Varoneckas (2004), i.e. we have assumed discontinuities of the objective function and small noise (about 1%) in its values. If it is known that the objective function is smooth and unimodal, then the values of $\beta_1,\ \beta_2$ can be reduced to make the search more local. For example, the test function (2) is smooth, contains no noise in function values, and has two equal local minima. Assuming these properties known, we may recommend to increase locality of the evolutionary search. The results of Table 2 show that the increase of the locality parameter β_3 improves the accuracy of the algorithm. If the essential irregularities of the objective function are expected then, to the contrary, the values of $\beta_1,\ \beta_2$ should be increased.

Table 2. Influence of β_1 and β_2 on the minimization results ($K = 50$, $\beta_3 = 1 - \beta_1 - \beta_2$)

β_1	β_2	N	N_1	f_0	f_m
0.2	0.5	3563	905	$5.36 \cdot 10^{-4}$	$1.52 \cdot 10^{-5}$
0.1	0.4	3319	888	$4.56 \cdot 10^{-4}$	$1.46 \cdot 10^{-5}$
0	0	3303	870	$3.32 \cdot 10^{-4}$	$1.02 \cdot 10^{-5}$

To test the algorithm for the case of higher dimensionality the problem (2) is generalized as follows

$$f(X) = -100(x_1 + 1)^2 + 121 \sum_{i=2}^{5} (x_i - 1)^2 + 0.1756 \tag{3}$$

$$A_\theta = \{X: X \in B \subset R^5, 5|x_1 + x_2| + |x_1 - x_2 + 2| \le 2 - 1.8\theta,$$
$$-2 \le x_i \le 2, i = 3, 4, 5\},$$

where $A = A_1$ and $B = \{X: -2 \le x_i \le 2, i = 1, ..., 5\}$. We re-scale the problem to the unit hyper cube. The minimum value of $f(x)$, $x \in A$, is equal to zero. There are two minimum points; two first coordinates of minimum points are equal to those of problem (2), and the other ones are equal to 0.75. Maximum of $f(\cdot)$ is bigger than 3215. The feasible region of the problem (3) models a thin slice of the hyper cube B. Such a type of feasible region may be expected in an industrial problem considered below. The feasible region of (3) is oriented along some axis but this property can not bias the testing results since that property is not used by the algorithm at all.

According to similar arguments as before we have chosen $\theta = 0.25, 0.5, 0.75, 1$, and β_1=0.2, β_2=0.5, β_3=0.3, $\gamma = 0.2$. However, to represent a feasible set in the five dimensional space more points are needed; the number of 500 hits of a targeted region and the maximal number of 10000 trials were chosen as the termination criterion for the search for A_i, $i = 1, ..., 4$. Therefore the maximal number of trials at the first stage of the algorithm can not exceed 40000. The size of the population in evolutionary search was chosen as equal N=750, and the maximal number of trials for evolutionary search was chosen as equal to 10000. The algorithm with the chosen parameters has been run 30 times. The best value found in 30 runs was 0.07224, at the point (0.2669, 0.7368, 0.7482, 0.7485, 0.7520). The average of found estimates for the minimum was 0.07725. The average number of trials was 40219 with 11833 of them successful.

The test results show that the proposed algorithm can find an acceptable approximation of the minimum and of the minimizer in the problems with small implicitly defined feasible regions, and the search terminates after an acceptable number of evaluations of an objective function.

5. Application to an industrial problem

The problem of black box optimization is typical for process engineering in the case where the physical and economical properties are modeled by software packages allowing for a limited access to the implemented models. For example, the only output of the package is either the objective function value or an indication that the input variables are infeasible. An optimization problem of such a type is related to industrial processing of raw hydrocarbon feed stock into oil and gas production, McCarthy, Fraga and Ponton (1998), Žilinskas, Fraga, Mackute and Varoneckas (2004). The design and optimization of such a process is difficult due to a combination of features of the process and the models used, both for modeling the physical behavior of the process and for deriving cost estimates of a given configuration and set of operating conditions, Žilinskas, Fraga, Mackute and Varoneckas (2004). Input data is the flowsheet of the process and its physical parameters, as well as the requested quality parameters of the products and their market prices. The design variables are some of the physical parameters of the process. For the given vector of variable parameters the objective function value is calculated using the modeling package, see Jacaranda, Fraga, Steffens, Bogle and Hind (2000). The objective is the profit of the process (expressed in 10^8 of USD), which is calculated using the parameters obtained via modeling the physical processes, and market data. The package returns the objective function value (profit with minus sign) for the design parameters guaranteeing requested quality of the products. The package returns 10^{20} in the case where either the process is physically impossible or product quality is not satisfactory. Reasonable intervals of the variables are known, and they constitute the hyper-rectangle set B. However, a large part of B is not feasible since the parameter combinations are not compatible with physical feasibility of the desired process. An even larger part of B is not feasible because of not acceptable product quality.

The corresponding optimization problem has five variables; the former hyper-rectangle B is re-scaled to a unit hypercube in R^5. The hypervolume of the feasible region A is approximately equal to 10^{-4}, as estimated in Žilinskas, Fraga, Mackute and Varoneckas (2004). The set A is expected to constitute a thin slice of the hypercube. The enclosures A_θ can be obtained by giving the model's input parameter θ a value smaller than 1, where θ=1 means quality of products satisfying the market conditions. The decrease of the volume of A_θ is nearly linear for $0 \leq \theta \leq 0.8$ but it is much faster when θ approaches 1, see Žilinskas, Fraga, Mackute and Varoneckas (2004). Therefore, we choose a sequence of θ values more densely close to 1: θ= 0.3, 0.6, 0.8, 0.9, 0.97, 1. The number of hits of the targeted region equal to 500, and maximal number of 10000 trials were chosen as the termination criteria for the search for A_i, $i = 1, ..., 6$. The structure of the evolving population is the same as in the testing examples: β_1=0.2, β_2=0.5, β_3=0.3, γ=0.2. The size of the population was 750, and the total number of calls of the modeling subroutine in the evolutionary search was restricted by 10000. The algorithm with the chosen parameters has been run 30 times. The best value found in 30 runs was -3.246, at the point (0.8027, 0.5088, 0.0000, 0.8519, 0.5428). The average of the found estimates of the minimum was -3.225. The average number of trials was 44521 with 10963 of them successful. The average cpu time of minimization was 1030 sec. using a 700 MHz Pentium computer with 256MB of RAM.

The projections of points representing the first and the last generations of the evolving population illustrate the rationality of the search strategy starting with uniform covering of the region and concentrating the search in prospective subregion while the search progresses; see Fig. 2. The candidate solutions found in different runs are sufficiently close to the best known approximation to the global minimum.

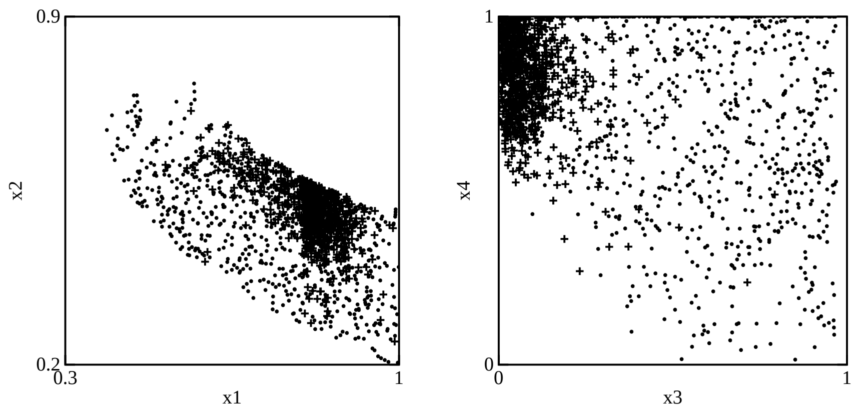

Figure 2. An illustration of the evolutionary search in the problem of optimal design; the projections of the trial points of the first (.) and of the fourth (+) generations on the plane $x_1 - x_2$ and $x_3 - x_4$.

6. Conclusions

A new method is proposed to solve the optimization problems with small implicitly defined regions. The testing results demonstrate the applicability of the developed algorithm to solve the problems with very small feasible regions in acceptable time. The algorithm is successfully applied to solve a difficult real world problem.

The global minima for the test problems have been found with precision acceptable for applications, however, in the cases where some special properties of $f(\cdot)$ and/or A are known, more efficient special search methods can be developed for searching in the region A represented by its points.

The proposed method gives not only an estimate of the global minimum but also enables to draw some conclusions about the feasible region. Consider, for example, the feasible region A of the industrial problem. The optimization results support the hypothesis that A is a thin slice of the hypercube B. First, some projections of A almost coincide with the projections of B as it is illustrated by Fig. 2 where the projections of the points uniformly distributed over A are denoted by (.). Second, we have calculated the eigenvalues of this set of points; they are equal to (0.097, 0.066, 0.030, 0.013, 0.0055), i.e. the largest eigenvalue is almost 20 times larger than the smallest eigenvalue.

References

BÄCK, T. (1996) *Evolution and Optimum Seeking.* Oxford University Press, Oxford.

FRAGA, E.S., STEFFENS, M.A., BOGLE, I.D.L. and HIND, A.K. (2000) An objected oriented framework for process synthesis and simulation. In: Malone, M.F., Trainham, J.A. and Carnahan, B., eds., *Foundations of Computer-Aided Process Design*, AIChE Symposium Series **95**, 446-449.

FRAGA, E.S. and ŽILINSKAS, A. (2003) Evaluation of hybrid optimization methods for the optimal design of heat integrated distillation sequences. *Advances in Engineering Software* **34**, 73-86.

GOLDBERG, D. (1989) *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison Wesley, Reading, MA.

MCCARTHY, E., FRAGA, E.S. and PONTON, J.W. (1998) An automated procedure for multicomponent product separation synthesis. *Computers & Chemical Engineering* **22** (7), 77-84.

MICHALEWICH, Z. (1996) *Genetic Algorithms + Data Structures = Evolutionary Programs.* Springer, New York.

SCHWEFEL, H.-P. (1995) *Genetic Algorithms + Data Structures = Evolutionary Programs.* J. Wiley, New York.

TORN, A. AND ZILINSKAS, A. (1989) *Global Optimization. Lecture Notes in Computer Science* **350**, 1-250.

ZILINSKAS, A., FRAGA, E.S., MACKUTE, A. and VARONECKAS, A. (2004) Adaptive search for optimum in a problem of oil stabilization process. In: Parmee, I., ed., *Adaptive Computing in Design and Manufacture VI.* Springer, 87-98.