Noname manuscript No.
(will be inserted by the editor)

Theoretical Justification of a Heuristic Subbox Selection Criterion for Interval Global Optimization

Vladik Kreinovich[1] **and Tibor Csendes**[2]

[1] Department of Computer Science, University of Texas at El Paso,
El Paso, TX 79968, USA
e-mail: `vladik@cs.utep.edu`

[2] Institute of Informatics, University of Szeged, Szeged, Hungary
e-mail: `csendes@inf.u-szeged.hu`

Received: February 1, 2001, accepted: ???

Abstract The most widely used guaranteed methods for global optimization are probably the interval-based branch-and-bound techniques. In these techniques, we start with a single box – the entire function domain – as a possible location of the global minimizer points, and then, in each step, subdivide some of the boxes, use interval computations to compute the enclosure $[\underline{F}(X), \overline{F}(X)] \supseteq f(X)$ of the range $f(X)$ of the objective function $f(x)$ on each new sub-box X, and, based on these computations, eliminate the boxes which cannot contain a global minimizer point. The computational efficiency of these methods strongly depends on which boxes we select for sub-division. Traditionally, for sub-division, the algorithms selected a box with the smallest value of $\underline{F}(X)$. Recently, it was shown that the algorithm converges much faster if we select, instead, a box with the largest possible value of the ratio $\dfrac{\widetilde{f} - \underline{F}(X)}{\overline{F}(X) - \underline{F}(X)}$, where $\widetilde{f}$ is a current upper bound on the actual global minimum. In this paper, we give a theoretical justification for this empirical criterion. Namely, we show that:

- first, this criterion is the only one that is *invariant* w.r.t. some reasonable symmetries; and
- second, that this criterion is *optimal* in some reasonable sense.

Key words Subbox selection, Global optimization, Theoretical justification.

1 Introduction

The most widely used guaranteed methods for global optimization (here we consider the minimizing of nonconvex functions) are interval-based branch-and-bound techniques; see, e.g., [7,13]. The main idea of these techniques is as follows. We start with a single box – the entire function domain $D \subseteq \mathbb{R}^n$ – as a possible location of a global minimizer point. In each step of the corresponding algorithm, we have one or several boxes X – which are chosen in such a way that all global minima are guaranteed to lie within one of these boxes. For each box X, we use interval computation techniques to compute the enclosure $[\underline{F}(X), \overline{F}(X)] \supseteq f(X)$ of the range $f(X)$ of the objective function $f(x) = f(x_1, \ldots, x_n)$. Here $\underline{F}(X)$ and $\overline{F}(X)$ are the lower and upper bound on the range, respectively.

For each point x within each of the selected boxes X, we have $f(x) \le \overline{F}(X)$; therefore,

$$\min_{x \in X} f(x) \le \overline{F}(X)$$

and hence,

$$\min_{x \in D} f(x) \le \min_{x \in X} f(x) \le \overline{F}(X).$$

Since this is true for every box X, we can therefore conclude that

$$\min_{x \in D} f(x) \le \widetilde{f} \stackrel{\text{def}}{=} \min_X \overline{F}(X).$$

If for some box X, we have $\underline{F}(X) > \widetilde{f}$, this means that a global minimizer point cannot be in this box, and so, this box can be safely deleted from the list of possible global minimizer.

After this deletion, we select a box X, divide it into two or more sub-boxes, and estimate the enclosure for all the new boxes. After this calculation, we recalculate $\widetilde{f} = \min\limits_X \overline{F}(X)$. Usually, the resulting new estimate is lower than the previous one, and this enables us to eliminates some of the old and/or new boxes. After this, the process starts anew.

The computational efficiency of these methods strongly depends on which boxes we select for sub-division. Traditionally, for sub-division, the algorithms selected a box with the smallest value of $\underline{F}(X)$ [13]. Recently, it was shown [5,6] that the algorithm converges much faster if we select, instead, a box with the largest possible value of the following *indicator* [2]:

$$I_{\text{new}}(X) = \frac{\widetilde{f} - \underline{F}(X)}{\overline{F}(X) - \underline{F}(X)}. \tag{1}$$

Similar formulas have also been successfully used in other contexts:

- the formula (1) is used to predict the computational cost of processing the box X – with the purpose of evenly distributing the load on a parallel computer [2];

– the same indicator was also used to improve the efficiency of multisection in interval optimization algorithms [3];
– and it was also applied to decide heuristically which subboxes can be deleted when due to the limited memory we have to give up solving optimization problems reliably [4];
– a similar formula
$$\frac{f(x)-m}{M-m},$$
where m is an estimate for a minimum of a function $f(x)$, and M is an estimate for its maximum, is used in fuzzy optimization to describe the degree to which the value x maximizes the objective function f [1,9,12].

In this paper, we give a theoretical justification of formula for the empirical expression (1). Namely, we show that:

– first, this expression is the only one that is invariant w.r.t. some reasonable symmetries; and
– second, that this expression is *optimal* in some reasonable sense.

For that, we use *symmetry* techniques which have already been used in several similar problems:

– in [8] to find the optimal selection of a side to bisect;
– in [10] to select an optimal formula for the so-called "ε-inflation", and
– in [11] for several other computational problems.

2 Informal (Heuristic) Justification of the New Indicator Function

Before we start a formal analysis of this problem, let us give an informal (heuristic) justification of formula (1).

The main advance of a subdivision step is that we get a better estimate $\widetilde{f}$ for the global minimum. Since this estimate is the smallest of the upper bounds $\overline{F}(X)$, the only way to decrease this estimate is to get a new upper bound $\overline{F}(X)_{\rm new}$ smaller than the previous estimate, i.e., to get $\overline{F}(X)_{\rm new} < \widetilde{f}$.

The smaller the box, the more accurate range estimates; thus, subdivision usually does decrease the enclosure. We can crudely estimate this decrease by assuming that the midpoint $m = (\underline{F}(X) + \overline{F}(X))/2$ of the enclosure interval stays the same, while the half-width decreases to a q-th portion of its original value $w = (\overline{F}(X) - \underline{F}(X))/2$, for some $q < 1$. As a result, the new upper bound is equal to

$$\overline{F}(X)_{\rm new} = m + q \cdot w = \frac{\overline{F}(X) + \underline{F}(X)}{2} + q \cdot \frac{\overline{F}(X) - \underline{F}(X)}{2},$$

and the desired inequality $\overline{F}(X)_{\rm new} < \widetilde{f}$ is equivalent to

$$\frac{\overline{F}(X) + \underline{F}(X)}{2} + q \cdot \frac{\overline{F}(X) - \underline{F}(X)}{2} < \widetilde{f}.$$

Subtracting $\underline{F}(X)$ from both sides of this inequality, we get an equivalent formula

$$\frac{\overline{F}(X)-\underline{F}(X)}{2}+q\cdot\frac{\overline{F}(X)-\underline{F}(X)}{2}=\frac{1+q}{2}\cdot\left(\overline{F}(X)-\underline{F}(X)\right)<\widetilde{f}-\underline{F}(X),$$

which is equivalent to

$$\frac{1+q}{2}<\frac{\widetilde{f}-\underline{F}(X)}{\overline{F}(X)-\underline{F}(X)}=I_{\text{new}}(X),$$

i.e., to $q<2I_{\text{new}}(X)-1$. Thus, the larger $I_{\text{new}}(X)$, the more probable it is that an unknown compression ratio q will lead to a decrease in $\widetilde{f}$. Hence, to increase the possibility that the value $\widetilde{f}$ will decrease, we must select the box X for which the indicator $I_{\text{new}}(X)$ attains the largest possible value.

This explanation is a heuristic argument in favor of the formula (1), but it does not answer the question of whether this method is indeed the best. To answer that question, we need a more formal analysis of this problem.

3 First Result: Describing All Invariant Indicator Functions

3.1 Selecting an Indicator is Equivalent to Selecting a Function

We would like to assign, to every box X, a numerical value $I(X)$ (called the *indicator*) such that in every step, we will subdivide a subbox with the largest value of $I(X)$. Assume that for each box X, the only information available to determine this indicator consists of the enclosure $[\underline{F}(X),\overline{F}(X)]$ of the range $f(X)$ of the function $f(x)$ on this box, and of the upper bound $\widetilde{f}$ of the minimum of $f(x)$ on the whole domain. Thus, the desired indicator $I(X)$ is a function of the corresponding three real variables: $I(X)=I\left(\underline{F}(X),\overline{F}(X),\widetilde{f}\right)$.

In these terms, the choice of an indicator $I(X)$ is reformulated as the choice of the corresponding function $I(a,b,c)$ – which is defined for all $a<b$. Examples:

- the traditional indicator $\underline{F}(X)\to\min$ (or, equivalently, $-\underline{F}(X)\to\max$) is described by a function $I(a,b,c)=-a$;
- the new indicator (1) is described by the function

$$I_{\text{new}}(a,b,c)=\frac{c-a}{b-a}. \tag{2}$$

Let us describe natural requirements for the desired function I.

3.2 Scale Invariance

In real-life problems, the objective function $f(x)$ often represents the value of some physical quantity (e.g., voltage) for the situation described by the parameters $x = (x_1, \ldots, x_n)$. If we change the unit in which we measure the corresponding physical quantity (e.g., use centimeters instead of meters), all numerical values will be multiplied by a constant λ (> 0), so instead of the original function $f(x)$, we get a new "re-scaled" function $\lambda \cdot f(x)$. In many practical cases the corresponding values of $\underline{F}(X)$, $\overline{F}(X)$, and $\widetilde{f}$ will be multiplied by the same constant λ: $\underline{F}(X) \to \lambda \cdot \underline{F}(X)$, $\overline{F}(X) \to \lambda \cdot \overline{F}(X)$, and $\widetilde{f} \to \lambda \cdot \widetilde{f}$.

It is natural to require that the indicator used for selecting a subbox should not depend on the choice of units, i.e., that

$$I\left(\lambda \cdot \underline{F}(X), \lambda \cdot \overline{F}(X), \lambda \cdot \widetilde{f}\right) = I\left(\underline{F}(X), \overline{F}(X), \widetilde{f}\right)$$

for all λ, $\underline{F}(X)$, $\overline{F}(X)$, and $\widetilde{f}$. In other words, the desired function I must be *scale-invariant* in the sense that

$$I(\lambda \cdot a, \lambda \cdot b, \lambda \cdot c) = I(a, b, c) \tag{3}$$

for all possible values $\lambda > 0$, a, b, and c.

3.3 Shift Invariance

For some physical quantities (e.g., for voltage), we can not only change the measuring unit, we can also change the starting point. In this case, a constant s will be added to all numerical values of $f(x)$: $f(x) \to f(x) + s$. The corresponding values of $\underline{F}(X)$, $\overline{F}(X)$, and $\widetilde{f}$ will also be increased by the same constant s for many range estimates: $\underline{F}(X) \to \underline{F}(X) + s$, $\overline{F}(X) \to \overline{F}(X) + s$, and $\widetilde{f} \to \widetilde{f} + s$.

It is natural to require that the indicator used for selecting a subbox should not depend on the choice of the starting point, i.e., that

$$I\left(\underline{F}(X) + s, \overline{F}(X) + s, \widetilde{f} + s\right) = I\left(\underline{F}(X), \overline{F}(X), \widetilde{f}\right)$$

for all s, $\underline{F}(X)$, $\overline{F}(X)$, and $\widetilde{f}$. In other words, the desired function I must be *shift-invariant* in the sense that

$$I(a + s, b + s, c + s) = I(a, b, c) \tag{4}$$

for all possible values s, a, b, and c.

Although the shift- and scale invariance are attractive properties, these are not absolutely necessary: it can also be satisfactory when the precedence of the subintervals remains the same, or even when only those subintervals with the maximal indicator value does not change.

3.4 Equivalent Indicators

It is easy to check that the function $I_{\rm new}(a,b,c)$ corresponding to the new indicator (1) is both scale- and shift-invariant. The following result shows that this function is, in essence, the *only* invariant indicator:

Theorem 1. *If a function I is scale- and shift-invariant, then $I(a,b,c) = G(I_{\rm new}(a,b,c))$ for some function $G : \mathbb{R} \to \mathbb{R}$.*

In other words, if $I(X)$ is an invariant indicator, then, for every box X, we have $I(X) = G(I_{\rm new}(X))$. Thus, if we know the value of $I_{\rm new}(X)$ for all boxes X, we will be able to determine the values of $I(X)$ for all the boxes, and thus, select the box to be subdivided. In this sense, the use of an arbitrary invariant indicator is equivalent to the use of the indicator (1).

It is also natural to require that if the two boxes differ only in the value of $\underline{F}(X)$, then the box with the smaller value of $\underline{F}(X)$ will be more likely to be sub-divided. The decrease in $\underline{F}(X)$ increases the ration $I_{\rm new}$; thus, we require that if the ratio $I_{\rm new}(X)$ increases, the value $I(X) = G(I_{\rm new}(X))$ should also increase. In other words, we require that the function $G(t)$ be strictly increasing. For such functions, the selection $I(X) \to \max$ is equivalent to the selection $I_{\rm new}(X) \to \max$.

Proof of Theorem 1. Since the function $I(a,b,c)$ is shift-invariant, we can take $s = -a$ and conclude that

$$I(a,b,c) = I(a+(-a), b+(-a), c+(-a)) = I(0, b-a, c-a). \qquad (5)$$

Since we assumed that $a < b$, we have $b - a > 0$. We can now use scale-invariance with $\lambda = 1/(b-a)$, and conclude that

$$I(0, b-a, c-a) = I\left(0, 1, \frac{c-a}{b-a}\right). \qquad (6)$$

Combining (5) and (6), we conclude that $I(a,b,c) = G(I_{\rm new}(a,b,c))$, where $G(t) = I(0,1,t)$. The theorem is proven. Q.E.D.

4 Second Result: Describing All Optimal Indicator Functions

4.1 What is an Optimality Criterion?

When we say that some *optimality criterion* is given, we mean that, given two different indicator functions I and I', we can decide whether the first or the second one is better, or whether these indicator functions are equivalent w.r.t. the given criterion. In mathematical terms, this means that we have a *pre-ordering relation* $\preceq$ on the set of all possible indicator functions.

4.2 We Want to Solve an Ambitious Problem: Enumerate all Indicator Functions That Are Optimal Relative to some Natural Criteria

One way to approach the problem of choosing the "best" indicator function I is to select *one* optimality criterion, and to find a function that is the best with respect to this criterion. The main drawback of this approach is that there can be different optimality criteria, and they can lead to different optimal solutions. It is, therefore, desirable not only to describe a function that is optimal relative to some criterion, but to describe *all* functions that can be optimal relative to different natural criteria[1]. In this paper, we are planning to implement exactly this more ambitious task.

4.3 Examples of Optimality Criteria

Pre-ordering is the general formulation of optimization problems in general, not only of the problem of choosing an indicator function I. In general optimization theory, in which we are comparing arbitrary *alternatives* a', a'', ..., from a given set A, the most frequent case of such a pre-ordering is when a *numerical criterion* is used, i.e., when a function $J : A \to R$ is given for which $a' \preceq a''$ iff $J(a') \leq J(a'')$.

Several natural numerical criteria can be proposed for choosing a function J. For example, we can take, as a criterion, the *average* number of iterations that lead to determining all global minima with a given relative accuracy (average in the sense of some natural probability measure on the set of all problems).

Alternatively, we can fix a class of problems, and take the largest number of iterations for problems of this class as the desired (numerical) optimality criterion.

Many other criteria of this type can be (and have actually been) proposed. For such "worst-case" optimality criteria, it often happens that there are several different alternatives that perform equally well in the worst case, but whose performance differ drastically in the average cases. In this case, it makes sense, among all the alternatives with the optimal worst-case behavior, to choose the one for which the average behavior is the best possible. This very natural idea leads to the optimality criterion that is not described by one numerical optimality criterion $J(a)$: in this case, we need *two* functions: $J_1(a)$ describes the worst-case behavior, $J_2(a)$ describes the average-case behavior, and $a \preceq b$ iff either $J_1(a) < J_1(b)$, or $J_1(a) = J_1(b)$ and $J_2(a) \leq J_2(b)$.

[1] In this phrase, the word "natural" is used informally. We basically want to say that from the purely mathematical viewpoint, there can be weird ("unnatural") optimality criteria. In our text, we will only consider criteria that satisfy some requirements that we would, from the common sense viewpoint, consider reasonable and natural.

We could further specify the described optimality criterion and end up with *one* natural criterion. However, as we have already mentioned, the goal of this paper is not to find *one* indicator function that is optimal relative to some criterion, but to describe *all* indicator functions that are optimal relative to some natural optimality criteria. In view of this goal, in the following text, we will not specify the criterion, but, vice versa, we will describe a very general class of *natural* optimality criteria.

So, let us formulate what "natural" means.

4.4 What Optimality Criteria are Natural?

We have already mentioned that the objective function $f(x)$ often represents the value of some measured quantity, and that the numerical value of a measured quantity changes if we select a new measuring unit and/or a new starting point. It is natural to require that the relative quality of two indicator functions does not depend on the choice of units or on the choice of the starting point.

How does replacing a unit change the indicator function I? If we replace a unit by one that is λ times smaller, then the quantity that was initially described by a value $f(x)$ will be described by a new value $\lambda \cdot f(x)$. Thus, the new value of the indicator $I(X)$ is equal to $I\left(\lambda \cdot \underline{F}(X), \lambda \cdot \overline{F}(X), \lambda \cdot \widetilde{f}\right)$. This value is equivalent to applying a new function $a, b, c \to I(\lambda \cdot a, \lambda \cdot b, \lambda \cdot c)$ to the old values $\underline{F}(X)$, $\overline{F}(X)$, and $\widetilde{f}$. We will denote this new, "re-scaled" indicator function by $S_\lambda(I)$.

Similarly, if we replace a starting point by a new one that is smaller by a constant s, then the quantity that was initially described by a value $f(x)$ will be described by a new value $f(x) + s$. Thus, the new value of the indicator $I(X)$ is equal to $I(\underline{F}(X) + s, \overline{F}(X) + s, \widetilde{f} + s)$. This value is equivalent to applying a new function $a, b, c \to I(a+s, b+s, c+s)$ to the old values $\underline{F}(X)$, $\overline{F}(X)$, and $\widetilde{f}$. We will denote this new, "shifted" ("translated") indicator function by $T_s(I)$.

In these terms, the above requirement is that if I is better than I', then:

- the "re-scaled" I (i.e., the function $S_\lambda(I)$) should be better than the "re-scaled" I' (i.e., than $S_\lambda(I')$); and
- the "shifted" I (i.e., the function $T_s(I)$) should be better than the "shifted" I' (i.e., than $T_s(I')$).

There is one more reasonable requirement for a criterion, that is related with the following idea: If the criterion does not select a single optimal indicator function, i.e., if it considers several different functions equally good, then we can always use some other criterion to help select between these "equally good" ones, thus designing a two-step criterion. If this new criterion still does not select a unique indicator function, we can continue this process until we arrive at a combination multi-step criterion for which there is only one optimal indicator function. Therefore, we can always assume that our criterion is *final* in this sense.

4.5 Definitions and the Main Result

Definition 1. *By an optimality criterion, we mean a pre-ordering (i.e., a transitive reflexive relation) $\preceq$ on the set A of all possible functions of three variables. An optimality criterion $\preceq$ is called:*

- *scale-invariant if for all I, I', and $\lambda > 0$, $I \preceq I'$ implies $S_\lambda(I) \preceq S_\lambda(I')$.*
- *shift-invariant if for all I, I', and s, $I \preceq I'$ implies $T_s(I) \preceq T_s(I')$.*
- *final if there exists one and only one indicator function I that is preferable to all the others, i.e., for which $I' \preceq I$ for all $I' \neq I$.*

Theorem 2.

- *If an indicator function I is optimal w.r.t. some scale- and shift-invariant final optimality criterion, then this function is equal to $G(I_{\text{new}})$ for some function $G : \mathbb{R} \to \mathbb{R}$.*
- *For every function $G : \mathbb{R} \to \mathbb{R}$, there exists a scale- and shift-invariant final optimality criterion for which the only optimal indicator function is $G(I_{\text{new}})$.*

Comment. In other words, if the optimality criterion satisfies the above-described natural properties, then the optimal indicator function is equivalent to (1).

Proof of Theorem 2.

We have already shown, in Theorem 1, that for every function G, as defined there, the indicator function $G(I_{\text{new}})$ is a scale- and shift-invariant, and that all scale- and shift-invariant functions have this form.

1. To prove the first part of Theorem 2, we will show that the optimal indicator function I_{opt} is scale-invariant and shift-invariant, i.e., that $S_\lambda(I_{\text{opt}}) = T_s(I_{\text{opt}}) = I_{\text{opt}}$ for all λ and s. Then, the result will follow from Theorem 1.

Indeed, let T be one of these transformations. Let us first show that each of these transformations is invertible, i.e., that the inverse transformation T^{-1} exists. Indeed:

- if $T = S_\lambda$, then $T^{-1} = S_{1/\lambda}$;
- if $T = T_s$, then $T^{-1} = T_{-s}$.

Now, from the optimality of I_{opt}, we conclude that for every $I' \in A$, $T^{-1}(I') \preceq I_{\text{opt}}$. From the invariance of the optimality criterion, we can now conclude that $I' \preceq T(I_{\text{opt}})$. This is true for all $I' \in A$ and therefore, the indicator function $T(I_{\text{opt}})$ is optimal. But since the criterion is final, there is only one optimal indicator function; hence, $T(I_{\text{opt}}) = I_{\text{opt}}$. So, the optimal indicator function is indeed invariant and hence, due to Theorem 1, it coincides with $G(I_{\text{new}})$. The first part is proven.

2. Let us now prove the second part of Theorem 2. Let G be fixed, and let $I_0 = G(I_{\rm new})$ be the corresponding indicator function. We will then define the optimality criterion as follows: $I \preceq I'$ iff I' is equal to this I_0.

Since the indicator function I_0 is scale-invariant and shift-invariant, thus the defined optimality criterion is also scale- and shift-invariant. It is also clearly final.

The indicator function I_0 is clearly optimal w.r.t. this invariant and final optimality criterion. Q.E.D.

Acknowledgments. This work was partially supported by NASA under cooperative agreement NCC5-209, by the United Space Alliance, grant No. NAS 9-20000 (PWO C0C67713A6), by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-95-1-0518, and by the Grants FKFP 0449/99, OTKA T 017241.

References

1. R. E. Bellman and L. A. Zadeh, "Decision-making in a fuzzy environment," *Management Sci.*, 1970, Vol. 17, pp. B141–B164.
2. L.G. Casado and I. García, "New load balancing criterion for parallel interval global optimization algorithm", *Proc. of the 16th IASTED International Conference*, Garmisch-Partenkirchen, Germany, February 1998, pp. 321–323.
3. L.G. Casado, I. García, and T. Csendes, "A new multisection technique in interval methods for global optimization", *Computing*, 65(2000) 263-269.
4. L.G. Casado, I. García, and T. Csendes, "A heuristic rejection criterion in interval global optimization algorithms", submitted for publication, http://www.inf.u-szeged.hu/~csendes/b2.ps.gz
5. L.G. Casado, J.A. Martínez, and I. García, "Experiments with a new selection criterion in a fast interval optimization algorithm", submitted for publication.
6. T. Csendes, "Convergence properties of interval global optimization algorithms with a new class of interval selection criteria", submitted for publication, http://www.inf.u-szeged.hu/~csendes/pfconv.ps.gz
7. R. Hammer, M. Hocks, U. Kulisch, and D. Ratz, *Numerical toolbox for verified computing. I. Basic numerical problems*, Springer-Verlag, Heidelberg, 1993.
8. R. B. Kearfott and Vladik Kreinovich, "Where to Bisect a Box? A Theoretical Explanation of the Experimental Results", In: G. Alefeld and R. A. Trejo (eds.), *Interval Computations and its Applications to Reasoning Under Uncertainty, Knowledge Representation, and Control Theory. Proceedings of MEXICON'98, Workshop on Interval Computations, 4th World Congress on Expert Systems*, México City, México, 1998.
9. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, NJ, 1995.
10. V. Kreinovich, S. A. Starks, and G. Mayer, "On a Theoretical Justification of The Choice of Epsilon-Inflation in PASCAL-XSC", *Reliable Computing*, 1997, Vol. 3, No. 4, pp. 437–452.
11. H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.

12. H. T. Nguyen and E. A. Walker, *First Course in Fuzzy Logic*, CRC Press, Boca Raton, FL, 1999.
13. H. Ratschek and J. Rokne, *New computer methods for global optimization*, Ellis Horwood, Chichester, 1988.