Interval Computations
No 4, 1993

How To Use Monotonicity-Type Information To Get Better Estimates of the Range of Real-Valued Functions

Vyacheslav M. Nesterov

The problem of estimating the range of a real-valued function is considered in the case when additional information is available about the type of monotonicity of the function and that of the functions contained in its analytical representation.

Использование информации о характере монотонности для лучшего оценивания множества значений действительной функции

В. М. Нестеров

Рассматривается задача оценивания множества значений действительной функции при наличии дополнительной информации о характере монотонности рассматриваемой функции и о характере монотонности функций, входящих в ее аналитическую запись.

V. M. Nesterov

In this paper, we consider the problem of estimating the range of a real-valued function $f(x) : \mathbb{R}^n \to \mathbb{R}$ for x from an interval X. This is a typical problem of interval computations, and it is thus well analyzed in interval mathematics.

Let's formulate the problem in precise mathematical terms. Suppose that a function $f(x) : \mathbb{R}^n \to \mathbb{R}$ and a set X, $X \subseteq \mathbb{R}^n$, are given, such that the function $f(x)$ is defined for all $x \in X$. The *range* $\overline{f}(X)$ is defined as

$$\overline{f}(X) = \{z : \text{There is an } x \in X \text{ such that } z = f(x)\}.$$

(the range is also called a *united extension*). We want to estimate this range, i.e., we want to find a set $Y \supseteq \overline{f}(X)$. In other words, we want to find a set $Y \subseteq \mathbb{R}$ such that

$$x \in X \Rightarrow f(x) \in Y.$$

We are assuming that a set X is given to us in some algorithmic sense (e.g., this set must be somehow encoded as a computer data). Therefore, this set X must belong to a class Ω_1 of sets that can be thus encoded. The resulting estimate Y must also belong to some class of encodable sets. Since the encoding used for Y may be different from the encoding for X, in the general case, we require that Y belongs to a class Ω_2 that does not necessarily coincide with Ω_1.

The function f must also be given to us in some computer form. There are several different ways to describe functions for a computer. In this paper, we assume that f is defined as a superposition of finitely many functions that will be called *basic*. As basic functions, one can take simple ("elementary") functions. Usually, the set of such functions contains arithmetic operations $+, -, *, /$ and basic mathematical functions such as $\sin, \cos, \tan, \sqrt{\ }$.

In interval computations, as Ω_1, we take a class consisting of n-dimensional intervals for all n (here, n is the number of variables of the function f), and as Ω_2, we take the class of all one-dimensional intervals. Sometimes, larger classes Ω_1 and Ω_2 are considered; elements of these larger classes are called *generalized intervals.*

For a given function f, we want a method that for every $X \in \Omega_1$ generates an estimate $F(X)$ for a range $\overline{f}(X)$. In other words, we want a function $F(X) : \Omega_1 \to \Omega_2$. It is reasonably easy to construct some estimate F. For example, we can replace all basic functions contained in a representation of f by their united extensions (for elementary functions, united extensions are easily computable). The extension F obtained in this manner is usually called *natural*.

A natural estimate $F(X)$ can be several times larger than the range itself. Our goal is to find an estimate that will be closer to $\overline{f}(X)$ (in the sense that the difference $F(X)\backslash\overline{f}(X)$ is smaller), i.e., a method F that will give a more accurate estimation of the range $\overline{f}(X)$.

To improve accuracy, other extensions F have been proposed, e.g., a MV-form [2] and a centered form [1].

When we try to implement the natural extension, we may encounter difficulties of two types.

- First, for some basic functions, the range may not belong to a selected class Ω_i of subsets of the real line. The simplest example is the division function $d(x_1, x_2) = x_1/x_2$. If an interval X_2 contains 0, then the resulting range (united extension) is not an interval. There are two main ways to solve this problem:
  - First, we can use a *generalized* interval arithmetic in which classes Ω_1 and Ω_2 contain not only intervals, but also other sets (that are called generalized intervals) [5, 8] (for a different generalization see also [3, 4]).
  - Second, we can still use standard intervals, but in this case, for the operations for which the range is not an interval, we cannot use a united extension. Instead, we must use an interval that contains the range. This idea was proposed and used in [8, 9].
- The second problem is related to the fact that a natural estimate can be too large. It is well known (see, e.g., [1]) that if in the analytical expression for f, each variable occurs only once, then natural extension coincides with the range. But if some variable occurs several times, then the natural extension can be too large: a classical example is an expression $f(x) = x - x$. Possible methods of overcoming the extending estimate interval are discussed in [5–8].

In the following text, we will consider the case of functions of one variable f, for which both classes Ω_1 and Ω_2 coincide with the the class of all the intervals $\mathbb{IR}$. We will also assume that the term T_f that describes a given function f is of the type $g(p, q)$, where g is a basic function of two variables, and p and q are some functions given by subterms of the term T_f. In [5–8], it was observed that additional information on monotonicity of p, q, and g, enables us to improve the accuracy of estimating the range $\overline{f}(X)$. This idea can be formalized in several different ways:

- by introducing several different interval extensions for each real-valued function;
- by defining an arithmetic of *directed* intervals, etc.

In [7], the case where g is one of the four arithmetical operations is considered. For example, for addition $g(p, q) = p + q$, two different interval extensions are constructed: $I + J$ and $I +^- J$, where $I, J \in \mathbb{IR}$. The interval function "+" is a usual interval addition, and the function "$+^-$" is defined by the following expression: if $I = [a_1, b_1]$ and $J = [a_2, b_2]$, then

$$I +^- J = \begin{cases} [a_1 + b_2, a_2 + b_1], & \text{if } a_1 + b_2 \le a_2 + b_1 \\ [a_2 + b_1, a_1 + b_2], & \text{if } a_1 + b_2 > a_2 + b_1. \end{cases}$$

In [7], it is proved that if p and q are continuous monotone functions on the interval X, and the function $f(p, q) = p + q$ is also continuous and monotone on X, then for an arbitrary interval $Y \subseteq X$, we have the following expression:

$$\overline{(p+q)}(Y) = \begin{cases} \overline{p}(Y) + \overline{q}(Y), & \text{if p and q are of the same type of monotonicity,} \\ \overline{p}(Y) +^- \overline{q}(Y), & \text{if the types of monotonicity of } p \text{ and } q \text{ are different.} \end{cases} \qquad (1)$$

Here, by the "type of monotonicity" we mean "increasing" or "decreasing".

Similar results for other arithmetic functions are given in [7].

These results are of somewhat limited practical value, because they are applicable only in the case when the resulting function (for addition, it is $p + q$) is itself monotonic. In this paper, we will show that a similar result is true when we require a weaker condition instead of monotonicity. We will also generalize these results to the case when other (non-arithmetic) basic functions are used.

To make our exposition clear, let us give a simple graphical illustration.

Assume that a function $f = g(p(x), q(x))$ if defined for all x from a given interval $X \in \mathbb{IR}$. Assume that for functions p and q, we have already constructed the interval estimates $P(X)$ and $Q(X)$ that coincide with the ranges (since both p and q are monotonic, this can be easily done: $\overline{p}([a, b]) = [p(a), p(b)]$ if p is increasing, and $\overline{p}([a, b]) = [p(b), p(a)]$ if p is decreasing). So,

$$x \in X \Rightarrow \big(p(x) \in P(X)\big) \;\&\; \big(q(x) \in Q(X)\big).$$

These estimates are precise in the following sense: for all $y_1 \in P(X), y_2 \in Q(X)$ there exist $x_1, x_2 \in X$ such that $p(x_1) = y_1$ and $q(x_2) = y_2$.

On the coordinate plane $\mathbb{R}^2$, we fix the rectangle $P(X) \times Q(X)$. On the same plane, when $x \in X$, the points $(p(x), q(x))$ form a curve. On this curve, we mark the direction corresponding to the increase of x. Using the same plane and denoting the coordinate system on it by (t_1, t_2) we can illustrate the behavior of the function $g(t_1, t_2)$ by drawing additional curves $g(t_1, t_2) = \text{const}$. On these additional curves, we will place small arrows pointing to the direction in which the value of g decreases.

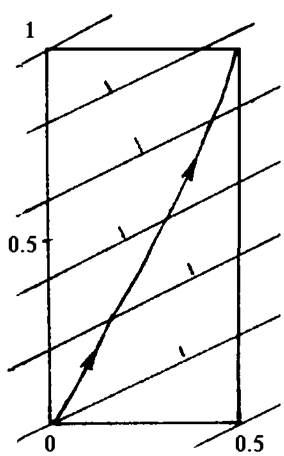

Fig.1

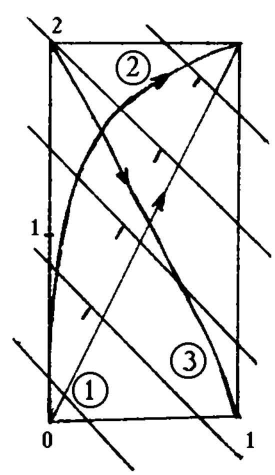

Fig.2

Example 1 (Figure 1). Let $p(x) = x$; $q(x) = \tan(\pi x/2)$; $g(t_1, t_2) = t_1 - 2t_2$, $X = [0, 1/2]$.

If instead of the function $f = g(p(x), q(x))$ with one variable $x \in [0, 1/2]$, we would have a function $\tilde{f}(x_1, x_2) = g(p(x_1), q(x_2))$ with two variables $x_1 \in X_1 = [0, 1/2]$ and $x_2 \in [0, 1]$, then the range of possible values of the function $\tilde{f}(x_1, x_2)$ would coincide with the range of values of a function g

with arguments included in the rectangle $[0, 1/2] \times [0, 1]$. This range contains all the values of $f(x)$, therefore, it can be taken as an estimate for the range $\overline{f}(X)$. Moreover, this estimate coincides with the natural estimate.

Since we have one variable, and not two, we cannot take arbitrary pairs (t_1, t_2), only the pairs for which $t_1 = p(x)$ and $t_2 = q(x)$ for some $x \in X$.

In other words, we have to consider only those points of the rectangle that are located on the curve $(p(x), q(x))$. If we do not take into consideration that x in p and q is the same variable, then we obtain the estimate $[-2, 1/2]$ for $\overline{f}(X)$. The actual range is $\overline{f}(X) = [-3/2, 0]$.

The above-mentioned result of [7] can be illustrated on this graph (Figure 2). Here, $g(t_1, t_2) = t_1 + t_2$. As an example, let's take the following p, q and X:

$$
\begin{aligned}
&1)\ \ X_1 = [0,1],\ p_1 = x^2,\ q_1 = 2\sqrt{x};\\
&2)\ \ X_2 = [0,1],\ p_2 = x,\ q_2 = 2x;\\
&3)\ \ X_3 = [1,2],\ p_3 = \log_2 x,\ q_3 = 4 - 2x.
\end{aligned}
$$

In all these cases, all three functions p, q, and $p + q$ are continuous and monotone on the corresponding segments. In the first and second cases, p and q have the same type of monotonicity; in the third case, p increases, while q decreases.

According to formula (1), the range of each function $f_i = p_i + q_i$ on an arbitrary subinterval $I \subseteq X_i$ can be computed as follows:

$$
\begin{aligned}
&1)\ \ F_i(I) = P_i(I) + Q_i(I),\ i = 1, 2;\\
&2)\ \ F_3(I) = P_3(I) +^{-} Q_3(I).
\end{aligned}
$$

It is easy to see that the monotonicity conditions on the functions p, q, and f can be substantially weakened. Let us give an example of a situation where none of the functions mentioned are monotone, but estimates performed using formula (1) remain true (Figure 3).

Let us denote by x_A, x_B, and x_C, the values of x that correspond to the points A, B and C that we marked on the curve. The endpoints of the interval X are denoted by x_1 and x_2. Let $x_1 < x_A < x_B < x_C < x_2$. Then, $p(x_1) < p(x_2)$ and $p(x_B) > p(x_C)$ (so p is not monotone), and $q(x_1) < q(x_2)$ and $q(x_A) > q(x_B)$ (so q is not monotone). Moreover, $f(x_A) > f(x_B)$, so f is not monotone either. All these three functions are not monotone. This

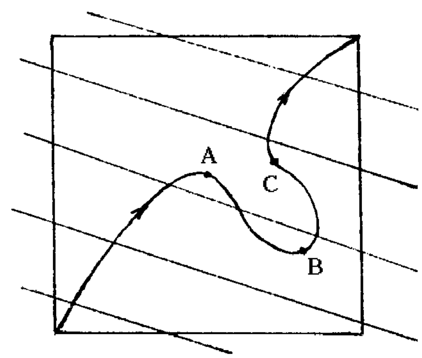

Fig.3

examples show that the monotonicity condition can be replaced by a weaker condition, namely, by the following.

Definition. *We say that a function* $f : \mathbb{R} \to \mathbb{R}$ *satisfies a predomination condition for an interval* $I = [x_1, x_2]$ *(and denote it by* $f \in Z(I)$*) if* f *is continuous on* I *and either* $f(x_1) \le f(x) \le f(x_2)$ *for all* $x \in I$*, or* $f(x_1) \ge f(x) \ge f(x_2)$ *for all* $x \in I$.

In the first case, we will write that $f \in Z_{\nearrow}(I)$, and in the second case that $f \in Z_{\searrow}(I)$.

For a function $f(t_1, t_2)$, we can (in the general case) construct the following four interval extensions F_i, $i = 1, \ldots, 4$:

$$\begin{aligned}
F_1([a,b],[c,d]) &= [f(a,c), f(b,d)];\\
F_2([a,b],[c,d]) &= [f(b,c), f(a,d)];\\
F_3([a,b],[c,d]) &= [f(a,d), f(b,c)];\\
F_4([a,b],[c,d]) &= [f(b,d), f(a,c)].
\end{aligned}$$

For an arbitrary function f, and for arbitrary intervals $[a,b]$ and $[c,d]$, only two of these extensions make sense, because for the remaining two the right end is smaller than the left end (so they do not define any reasonable interval).

Proposition *Assume that we are given an interval* $X \in \mathbb{R}$ *and two functions* $p(x)$ *and* $q(x)$ *from* X *to* $\mathbb{R}$*. Assume also that we know the ranges* $\overline{p}(X) = Y_1$ *and* $\overline{q}(X) = Y_2$ *of* p *and* q*. Let the function* $f(x)$ *be specified as* $f(x) = g(p(x), q(x))$*, where* g *is a function of two variables. Then, for a united* *extension $\overline{f}(X)$, we have the following expressions:*

$$
\begin{aligned}
p \in Z_{\nearrow}(X) \,\&\, q \in Z_{\nearrow}(X) \,\&\, f \in Z_{\nearrow}(X) &\Rightarrow \overline{f}(X) = G_1(Y_1, Y_2);\\
p \in Z_{\nearrow}(X) \,\&\, q \in Z_{\nearrow}(X) \,\&\, f \in Z_{\searrow}(X) &\Rightarrow \overline{f}(X) = G_4(Y_1, Y_2);\\
p \in Z_{\nearrow}(X) \,\&\, q \in Z_{\searrow}(X) \,\&\, f \in Z_{\nearrow}(X) &\Rightarrow \overline{f}(X) = G_3(Y_1, Y_2);\\
p \in Z_{\nearrow}(X) \,\&\, q \in Z_{\searrow}(X) \,\&\, f \in Z_{\searrow}(X) &\Rightarrow \overline{f}(X) = G_2(Y_1, Y_2);\\
p \in Z_{\searrow}(X) \,\&\, q \in Z_{\nearrow}(X) \,\&\, f \in Z_{\nearrow}(X) &\Rightarrow \overline{f}(X) = G_2(Y_1, Y_2);\\
p \in Z_{\searrow}(X) \,\&\, q \in Z_{\nearrow}(X) \,\&\, f \in Z_{\searrow}(X) &\Rightarrow \overline{f}(X) = G_3(Y_1, Y_2);\\
p \in Z_{\searrow}(X) \,\&\, q \in Z_{\searrow}(X) \,\&\, f \in Z_{\nearrow}(X) &\Rightarrow \overline{f}(X) = G_4(Y_1, Y_2);\\
p \in Z_{\searrow}(X) \,\&\, q \in Z_{\searrow}(X) \,\&\, f \in Z_{\searrow}(X) &\Rightarrow \overline{f}(X) = G_1(Y_1, Y_2);
\end{aligned}
$$

Figure 4 illustrates all the eight cases (a-h).

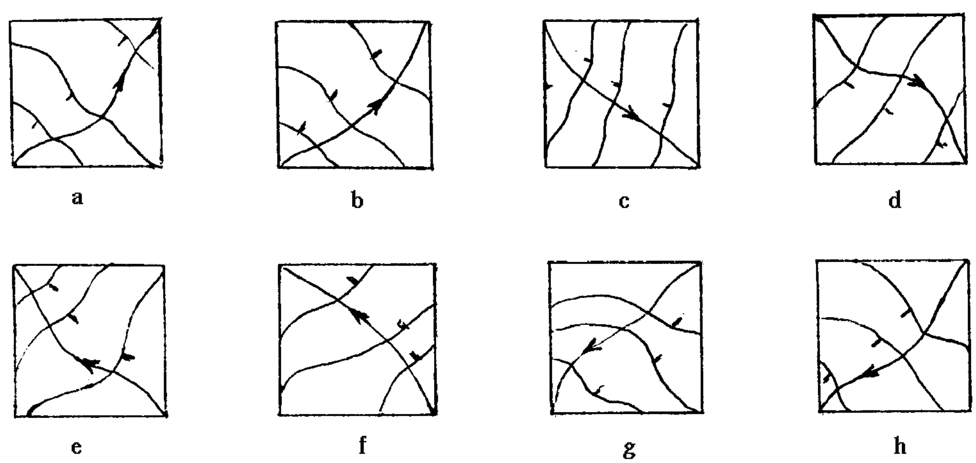

Fig.4

Example 2. Let $f(x) = \sqrt{\left|(x+6)^2 - (\frac{x^3}{3} - \frac{x}{3} + 2)^2\right|}$, $X = [-2, 2]$. The set of base functions includes $x + y$, $x - y$, $x * y$, x^2, x^3, $\sqrt{|x^2 - y^2|}, \ldots$. Let us denote $p(x) = x + 6$, $q(x) = \frac{x^3}{3} - \frac{x}{3} + 2$, $g(t_1, t_2) = \sqrt{|t_1^2 - t_2^2|}$. Using the standard interval arithmetic one can calculate:

$$\overline{p}(X) = [-2, 2] + 6 = [4, 8]; \quad (2)$$

$$\overline{q}(X) \subseteq \tfrac{[-2,2]^3}{3} - \tfrac{[-2,2]}{3} + 2 = \left[-\tfrac{4}{3}, \tfrac{16}{3}\right]. \quad (3)$$

Using standard technique one can obtain the estimate

$$\overline{f}(X) \subseteq \sqrt{\left|[4, 8]^2 - [-\tfrac{4}{3}, \tfrac{16}{3}]^2\right|} = [0, 8].$$

This estimate is an overshoot. Let us show that we can use our Proposition to get a better estimate. Indeed, the function q can be represented as $q(x) = h(r(x), t(x))$, where $r(x) = \frac{x^3}{3}$, $s(x) = -\frac{x}{3} + 2$, and $h(t_1, t_2) = t_1 + t_2$.

It is easy to see that the function q satisfies the conditions of the Proposition. Using the Proposition, one can thus find the exact range of q:

$$\overline{q}(X) = H_3\Big(\big[-\tfrac{8}{3}, \tfrac{8}{3}\big], \big[\tfrac{4}{3}, \tfrac{8}{3}\big]\Big) = \Big[h\big(-\tfrac{8}{3}, \tfrac{8}{3}\big), h\big(\tfrac{8}{3}, \tfrac{4}{3}\big)\Big] = [0, 4].$$

The function f also satisfies the conditions of the Proposition. One can therefore find the desired range of f.

$$\overline{f}(X) = G_1\big([4, 8], [0, 4]\big) = \big[g(4, 0), g(8, 4)\big] = \Big[4, \sqrt{48}\Big].$$

It is necessary to note that in this example, the function $q(x)$ is not monotonic on the interval $[-2, 2]$ therefore, it does not satisfy the conditions of Markov's theorem [7].

This proposition can be generalized in two directions:

- First, one can consider lists of basic functions that contain functions of more than two variables. This generalization is pretty straightforward. It should be noted, however, for a basic function of n variables, the number of possible extensions is equal to 2^n (and thus grows fast with n).
- Second, this Proposition can be generalized to the case when the resulting function f has several variables. This generalization will be described in a future paper.

References

[1] Alefeld, G. and Herzberger, J. *Introduction to interval computations.* Academic Press, New York etc., 1983.

[2] Caprani, O. and Madsen, K. *Mean value forms in interval analysis.* Computing **25** (2) (1980), pp. 147–154.

[3] Dimitrova, N. and Markov, S. *On the interval-arithmetic presentation of the range of a class of monotone functions of many variables.* In: Kaucher, E., Markov, S. M., and Meyer, G. (eds) "Computer Arithmetic, Validated Computation and Mathematical Modeling." Academic Press, 1991, pp. 213–228.

[4] Dimitrova, N and Markov, S. M. *Uber die intervall-arithmetische Berechnung des Wertebereichs einer Funktion mit Anwendungen,* Freiburger Intervall-Berichte, Univ. Freiburg, 81/4 (1981), pp. 1–22.

[5] Markov, S. M. *Extended interval arithmetic involving infinite intervals.* Mathematica Balkanica **6** (3) (1992), pp. 269–304.

[6] Markov, S. M. *On an arithmetic for directed intervals and its application: the real arithmetic case.* (To be published.)

[7] Markov, S. M. *On the presentation of ranges of monotone functions using interval arithmetic.* Interval Computations 4 (6) (1992), pp. 19–31.

[8] Nesterov, V. M. *On a generalization of interval analysis and its application to estimation of the range of values of a function.* In: "Mathematical methods for the construction and analysis of algorithms", Nauka, Leningrad, 1990, pp. 109–124 (in Russian).

[9] Nesterov, V. M. *Estimating a range of values of functions using extended interval arithmetic.* Interval Computations 4 (6) (1992), pp. 47–53.

St. Petersburg Institute
for Informatics and Automation,
St.Petersburg, 195256, Box 52
Russia