IMM

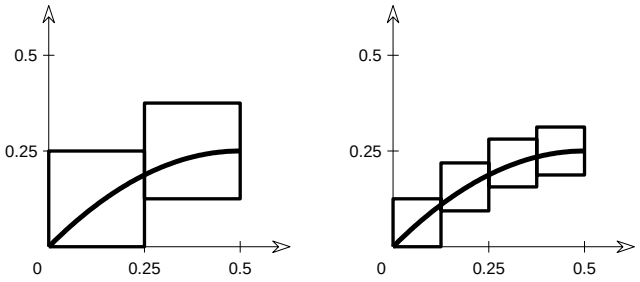
20.11.2002

Introduction to Interval Analysis

Ole Caprani

Kaj Madsen

Hans Bruun Nielsen



Abstract. Interval analysis was introduced by Ramon Moore in 1959 as a tool for automatic control of the errors in a computed result, that arise from input error, rounding errors during computation, and truncation errors from using a numerical approximation to the mathematical problem. An important application is the enclosure of ranges of functions, which enables us e.g. to solve equations with interval coefficients, to prove existence of certain solutions, and to solve global optimization problems. We discuss the basic concepts of interval analysis and present methods for interval solution of some numerical problems.

Contents

1. Introduction

In general there are three sources of error in connection with numerical computations: rounding errors, truncation errors and input errors. For estimation of the effect of such errors it may be advantageous to compute with *intervals* of numbers. In this introduction we give some simple examples. In the following chapters we give a more stringent presentation of interval calculation.

Example 1.1. First consider *rounding errors*. Assume that we work with 3 significant digits and normal rounding, and wish to compute

$$y \;=\; 1 - 0.531 + \frac{0.531^2}{2}$$

We get

$$0.469 + \frac{0.282}{2}\;, \qquad 0.469 + 0.141\;, \qquad 0.610$$

How large is the error in this computed result?
In this simple example it is easy to find the answer: all we have to do is to use a sufficient number of digits in the computation,

$$0.469 + \frac{0.281961}{2}\;, \qquad 0.469 + 0.1409805\;, \qquad 0.6099805$$

So the error in the first result is $0.610 - 0.6099805 \;=\; 0.0001950$. ∎

It is not always possible, however, to use this approach. There may be too many digits to keep track of or divisions may give infinite decimal fractions that need rounding. Instead we must estimate the error or keep track of it during computation. The latter is our choice when we calculate with intervals.

Example 1.2. With the problem from the first example, thw value of 0.531^2 is contained in the interval $[\,0.281\,,\,0.282\,]$, and we get

$$\begin{aligned} & 0.469 + \frac{0.531^2}{2} \\ \in\; & 0.469 + [\,0.281\,,\,0.282\,] \cdot \frac{1}{2} \\ \subseteq\; & 0.469 + [\,0.140\,,\,0.141\,] \;=\; [\,0.609\,,\,0.610\,] \end{aligned}$$

Thus, we know that the true value of y is between 0.609 and 0.610. During computation the end points of the intervals were rounded in the directions that ensured that the resulting interval contained the true result. By monitoring intervals in this special way we can keep track of rounding errors. ∎

Example 1.3. Next, we look at *truncation errors*. We wish to compute $y = e^{-0.531}$, and use the Taylor series with remainder term,

$$e^x \;=\; 1 + x + \frac{x^2}{2!}\,e^t\;, \quad \text{where } t \text{ is between 0 and } x$$

We are interested in the case $x < 0$, and since $e^t \in [\,0\,,\,1\,]$ for all $t < 0$, we get

$$e^x \;\in\; 1 + x + \frac{x^2}{2!}\,[\,0\,,\,1\,]\;. \tag{1.1}$$

With $x = -0.531$ we find

$$\begin{aligned} e^{-0.531} \;&\in\; 1 - 0.531 + \frac{0.531^2}{2}\,[\,0\,,\,1\,] \\ &\in\; 0.469 + [\,.140\,,\,0.141\,] \cdot [\,0\,,\,1\,] \\ &\in\; 0.469 + [\,0\,,\,0.141\,] \;=\; [\,0.469\,,\,0.610\,] \end{aligned}$$

Again we found an interval that is sure to contain the correct result. The computation shows that interval calculations allows us simultaneously to keep track of rounding errors and truncation errors. ∎

Example 1.4. Finally, to illustrate *input errors,* suppose that instead of $x = -0.531$ we only know that $x \in [\,-0.532\,,\,-0.531\,]$. We can estimate the effect of this uncertainty in indata: Equation (1.1) gives

$$\begin{aligned} e^x \;&\in\; 1 + [\,-0.532\,,\,-0.531\,] + \frac{[\,-0.532\,,\,-0.531\,]^2}{2} \cdot [\,0\,,\,1\,] \\ &\in\; [\,0.468\,,\,0.470\,] + \frac{[\,0.280\,,\,0.284\,]^2}{2} \cdot [\,0\,,\,1\,] \\ &\in\; [\,0.468\,,\,0.470\,] + [\,0\,,\,0.142\,] \;=\; [\,0.468\,,\,0.612\,] \end{aligned}$$

The computation shows that all three sources of error can be included, and the result is an interval that contains all values $\{e^x\}$ for x in the interval $[\,-0.532\,,\,-0.531\,]$. ∎

The results in the last two examples are not impressive with respect to accuracy. However, we can cope with that by increasing the number of digits to represent the interval end points, by increasing the number of terms in the Taylor series, and by using special techniques to compute mappings. These subjects are treated in the following chapters.

2. Interval Arithmetic

We start this chapter with a more formal description of intervals and then we specify what we mean by "calculating with intervals". The treatment falls in two parts. First we assume that the mathematically correct version of the four arithmetic operations is available. Next, we become realistic – we use floating point numbers to represent the interval end points and floating point arithmetic to do the calculations with these end points.

2.1. Real Intervals

A *real interval* A is the bounded, closed subset of the real numbers defined by

$$A \;=\; [\underline{a}, \overline{a}] \;=\; \{\, x \in \mathcal{R} \,|\, \underline{a} \le x \le \overline{a} \,\}\;,$$

where $\underline{a}, \overline{a} \in \mathcal{R}$ and $\underline{a} \le \overline{a}$. We use $\mathcal{I}(\mathcal{R})$ to denote the set of such intervals. A *singleton* A is a *degenerate interval*: $\underline{a} = \overline{a}$.

For later use we introduce four simple concepts, the *midpoint* $m(A)$, the *radius* $r(A)$, the upper bound $|A|$ and the *width* $w(A)$ of the interval A,

$$\begin{aligned} m(A) &= \tfrac{1}{2}(\underline{a} + \overline{a})\;, \\ r(A) &= \tfrac{1}{2}(\overline{a} - \underline{a})\;, \\ |A| &= \max_{a \in A} |a| \;=\; \max\{|\underline{a}|, |\overline{a}|\}\;, \\ w(A) &= \overline{a} - \underline{a} \;=\; 2r(A)\;. \end{aligned} \tag{2.1}$$

Note that if X is a singleton, then $m(X) = X$ and $r(X) = 0$.

2.2. Interval Vectors and Matrices

An *interval vector* V in $\mathcal{I}(\mathcal{R}^n)$ has n components (coordinates), each of which is an interval, $V_i \in \mathcal{I}(\mathcal{R}),\ i = 1, \ldots, n$.

The concepts defined in (2.1) have straightforward extensions to interval vectors. The first three are defined componentwise,

$$\begin{aligned} m(V)_i &= m(V_i)\ , \\ r(V)_i &= r(V_i)\ , \\ |V|_i &= |V_i|\ . \end{aligned} \tag{2.2}$$

Thus, each of these is a vector in $\mathcal{R}^n$. We shall further make use of the *width* and the *norm* of an interval vector. Both of these are scalars, $w(V), \|V\| \in \mathcal{R}_+$.

$$\begin{aligned} w(V) &= 2\|r(V)\|_\infty = \max_i\{w(V_i)\}\ , \\ \|V\| &= \| \, |V| \, \|_\infty = \max_i |V_i|\ . \end{aligned} \tag{2.3}$$

Example 2.1. Let $V = ([-3\,,\,1]\,,\,[3\,,\,5]) \subseteq \mathcal{I}(\mathcal{R}^2)$. Then $m(V) = (1,\ 4)$, $r(V) = (2,\ 1)$, $|V| = (3,\ 5)$, $w(V) = 4$, $\|V\| = 5$. ∎

An *interval matrix* M in $\mathcal{I}(\mathcal{R}^{m\times n})$ has m rows and n columns. Each element is an interval, $M_{ij} \in \mathcal{I}(\mathcal{R})$, $i = 1, \ldots, m$; $j = 1, \ldots n$, and again the concepts from (2.1) are defined elementwise, while the scalars from (2.3) generalize to

$$\begin{aligned} w(M) &= \max_{i,j}\{w(M_{i,j})\}\ , \\ \|M\| &= \| \, |M| \, \|_\infty = \max_i\{\sum_j |M_{ij}|\}\ . \end{aligned} \tag{2.4}$$

Example 2.2. Let $M = \begin{pmatrix} [4\,,\,6] & [0\,,\,2] \\ [-2\,,\,0] & [2\,,\,4] \end{pmatrix}$. Then $m(M) = \begin{pmatrix} 5 & 1 \\ -1 & 3 \end{pmatrix}$, $r(M) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $|M| = \begin{pmatrix} 6 & 2 \\ 2 & 4 \end{pmatrix}$, $w(M) = 2$, $\|M\| = 8$. ∎

Example 2.3. Let $Y \in \mathcal{I}(\mathcal{R}^m)$ denote the interval vector obtained as $Y = MV$, where $M \in \mathcal{I}(\mathcal{R}^{m\times n})$ and $V \in \mathcal{I}(\mathcal{R}^n)$. It can be shown that with the definitions in (2.3) and (2.4) it holds that

$$\|Y\| \ \leq\ \|M\| \cdot \|V\|\ .$$

We shall use this appraisal in chapters 9 and 10.

Note the equivalence with the infinity norm of real vectors and matrices. For $V \in \mathcal{R}^n$ and $M \in \mathcal{R}^{m\times n}$ the norms $\|V\|_\infty$ and $\|M\|_\infty$ are defined as in (2.3) and (2.4), respectively, and $\|MV\|_\infty \leq \|M\|_\infty \cdot \|V\|_\infty$. ∎

Now we are ready to specify what we mean by "calculating with intervals". The next section discusses arithmetic operations with simple intervals.

2.3. Real Interval Arithmetic

Let $\diamond$ denote one of the four arithmetic operators, $+$, $-$, $*$ or $/$. For $A, B \in \mathcal{I}(\mathcal{R})$ we define

$$A \diamond B \ =\ \{a \diamond b \,|\, a \in A \ \wedge\ b \in B\}\ , \tag{2.5}$$

with A/B undefined if $0 \in B$. The four arithmetic operations are continuous mappings of $\mathcal{R}^2$ onto $\mathcal{R}$ and A and B are bounded and closed. Therefore, also $A \diamond B$ is an interval; $A \diamond B \in \mathcal{I}(\mathcal{R})$. Thus, Equation (2.5) defines four operators on $\mathcal{I}(\mathcal{R})$. The definition ensures that the resulting interval $A \diamond B$ contains all possible outcomes from applying $\diamond$ with operands from A and B. The definition is not operational, however; it does not show how to find the end points of the interval $A \diamond B$, but it is easy to find maximum and minimum for the four arithmetic operations, and this leads to

$$\begin{aligned} A + B &= [\underline{a}+\underline{b},\, \overline{a}+\overline{b}] \\ A - B &= [\underline{a}-\overline{b},\, \overline{a}-\underline{b}] \\ A * B &= [\min(\underline{a}\,\underline{b}, \underline{a}\,\overline{b}, \overline{a}\,\underline{b}, \overline{a}\,\overline{b})\,,\, \max(\underline{a}\,\underline{b}, \underline{a}\,\overline{b}, \overline{a}\,\underline{b}, \overline{a}\,\overline{b}))\,] \\ A \,/\, B &= A * [1/\overline{b},\, 1/\underline{b}]\ . \end{aligned} \tag{2.6}$$

Example 2.4. $[0\,,\,1] + [1\,,\,2] = [1\,,\,3]$, $[1\,,\,2] - [1\,,\,2] = [-1\,,\,1]$, $[-3\,,\,2] * [1\,,\,2] = [-6\,,\,4]$, $[1\,,\,2]/[3\,,\,4] = [1/4\,,\,2/3]$. ∎

Identifying singletons with real numbers we see that the operations are extensions of the usual operations in $\mathcal{R}$. Therefore we permit ourselves to write e.g. $3 + [1\,,\,2]$ as short for $[3\,,\,3] + [1\,,\,2]$.

From the definition (2.5) it follows immediately that

$$A_1 \subseteq A_2 \ \text{ and } \ B_1 \subseteq B_2 \quad \Rightarrow \quad A_1 \diamond B_1 \subseteq A_2 \diamond B_2 \tag{2.7}$$

This property of the four interval operations is important. We say that the operations are *inclusion monotonic*.

2.4. Floating Point Interval Arithmetic

When performing interval arithmetic on a computer we have to represent intervals, and we choose to use floating point numbers for the end points. Let $\mathcal{F}$ denote the set of floating point numbers and define the floating point interval operations by

$$\begin{aligned} \mathsf{fl}(A \diamond B) \ = \ & \text{smallest interval with end points in } \mathcal{F}, \\ & \text{which contains } A \diamond B \ . \end{aligned} \tag{2.8}$$

Again, this definition is not operational. It depends e.g. on the floating point processor whether it is possible to find the two end points, but this problem is outside the scope of this introduction.

The definition (2.8) ensures that also floating point interval operations are inclusion monotonic.

In the following we describe a number of methods using real intervals and real interval arithmetic. Such interval computations are done with the help of floating point interval arithmetic. The resulting intervals are guaranteed by (2.8) to contain the intervals we would have found by the equivalent real operations. In the formulation of the methods we only need to use real intervals, and the practical computation leads to a wider interval. Thus, we have automated estimation of rounding errors, but – as indicated by the following example – we still have to be cautious when we formulate the algorithms.

Example 2.5. Following Ris (1972) it holds that

$$\sqrt{33} \in [\,5.744\,,\,5.745\,]\,, \qquad \sqrt{29} \in [\,5.385\,,\,5.386\,]\,,$$

and calculating with four significant digits gives

$$\sqrt{33} + \sqrt{29} \in [\,11.12\,,\,11.14\,]\,, \qquad \sqrt{33} - \sqrt{29} \in [\,0.3580\,,\,0.3600\,]\,.$$

The first result is quite good, but the other is terrible. It is the well known numerical problem about *cancellation* in connection with subtraction, and as usual the cure is to reformulate the calculation

$$\sqrt{33} - \sqrt{29} \ = \ \frac{4}{\sqrt{33} + \sqrt{29}} \ \in \ \frac{4}{[\,11.12\,,\,11.14\,]} \ \in \ [\,0.3590\,,\,0.3598\,]\,,$$

which is a much better result. ∎

Finally, consider the relation between ordinary floating point calculation and floating point interval arithmetic. If we make a floating point calculation, how does the result relate to the corresponding calculation with floating point interval arithmetic? To answer that question, remember that

$$\begin{aligned} \mathsf{fl}(a \diamond b) \ = \ & \text{one of the two numbers in } \mathcal{F}, \\ & \text{which are closest to } a \diamond b \ , \end{aligned} \tag{2.9}$$

unless $a \diamond b \in \mathcal{F}$, in which case $\mathsf{fl}(a \diamond b) = a \diamond b$. Combining this result with (2.8) we get

$$\mathsf{fl}(a \diamond b), \ a \diamond b \ \in \ \mathsf{fl}(A \diamond B) \quad \text{for} \ \ a \in A, \ b \in B \ .$$

Therefore, *the interval obtained by floating point interval arithmetic contains both the mathematically true result and the result we get by ordinary floating point calculation.*

This explains the result in the above example. If the floating point calculation is *unstable*, i.e. has a large rounding error, then the floating point interval arithmetic results in wide intervals. The opposite is unfortunately **not** the case, and this is probably the reason why some people say that interval arithmetic is too pessimistic.

3. Interval Algebra

In this chapter we look at the rules that are valid for $\mathcal{I}(\mathcal{R})$ with the four associated interval operations.

It is easy to prove that $+$ and $*$ are both *commutative* and *associative*,

For $A, B, C \in \mathcal{I}(\mathcal{R})$:

$$A+B \;=\; B+A\,, \quad (A+B)+C \;=\; A+(B+C)\,,$$
$$A*B \;=\; B*A\,, \quad (A*B)*C \;=\; A*(B*C)\,.$$

Moreover $[0\,,\,0]=0$ and $[1\,,\,1]=1$ are *neutral* with respect to addition and multiplication, respectively,

$$A+0 \;=\; 0+A \;=\; A\,, \qquad A*1 \;=\; 1*A \;=\; A\,.$$

A nondegenerate interval has no *inverse* with respect to $+$ and $*$ since either of these with a nondegenerate operand cannot result in the neutral element, which is a singleton.

The *distributive rule* is **not** valid in general.

Example 3.1. As a counter example consider

$$[1\,,\,2]*([1\,,\,2]-[1\,,\,2]) \;=\; [1\,,\,2]*[-1\,,\,1] \;=\; [-2\,,\,2]\,,$$
$$[1\,,\,2]*[1\,,\,2]-[1\,,\,2]*[1\,,\,2] \;=\; [1\,,\,4]-[1\,,\,4] \;=\; [-3\,,\,3]\,.$$

■

Instead the socalled *sub-distributivity* holds,

$$A(B+C) \;\subseteq\; AB+AC\,.$$

The proof of this is clear from the definition (2.5).

The lack of validity of the distributive rule often causes trouble when formulating interval calculations. It is so unusual not being allowed to multiply into a parenthesis.

The distributive rule is valid for special intervals, e.g.

$$t(A+B) \;=\; tA+tB \quad \text{for} \quad t\in\mathcal{R}\,.$$

A very detailed analysis of other special cases can be found in Ris (1972).

4. Interval Functions

In the introduction we saw the use of interval calculation to find upper and lower bounds for the range of the exponential function over a given domain. This application of intervals and interval operations is the subject of this chapter and more to come. It is of interest by itself to be able to estimate bounds of the range of a function, but the concepts presented in this chapter will prove useful later when we formulate methods for integration, root-finding and optimization.

4.1. Interval Hulls

Consider the function $f : \mathcal{R}^n \mapsto \mathcal{R}^m$ defined on a domain $A \subseteq \mathcal{I}(\mathcal{R}^n)$, cf. Section 2.2. The image of A is

$$f(A) \;=\; (f_1(A),\ldots,f_m(A)) \;=\; \{f(x_1,\ldots,x_n)\,|\,x_i\in A_i\}\,.$$

If f is continuous, then this set is closed and compact. It is obvious that $f(X)\subseteq f(A)$ for all $X\subseteq A$. Unfortunately the image is normally **not** an interval vector. This is illustrated by

Example 4.1. With $f : \mathcal{R}^2 \mapsto \mathcal{R}^2$ defined by $y_1 = x_1 + x_2$, $y_2 = x_1^2$, the image $f([0\,,\,1],[0\,,\,1])$ is the shaded region shown below.

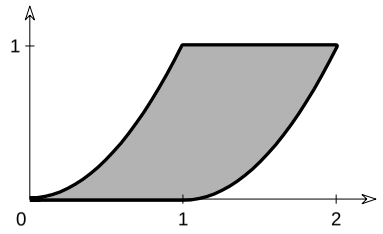


Figure 4.1. ■

Obviously, we cannot give such a figure as the result of a calculation. Therefore we enclose the image in an interval vector, and use this as the result. The smallest interval vector that encloses $f(X)$ is called the *interval hull* of f and denoted $\overline{f}$. Thus,

$$f(X) \;\subseteq\; \overline{f}(X) \quad \text{for} \quad X\subseteq A\,. \tag{4.1}$$

In the special case of a real function, i.e. $m=1$, the image is an interval in $\mathcal{R}$, and in this case (4.1) holds with equality.

If f is one of the four arithmetic operations, $f(a,b) = a \diamond b$ then the interval hull of f is the corresponding interval operation, $\overline{f}(A,B) = A \diamond B$.

4.2. Interval Extensions

A mapping $F : \mathcal{I}(\mathcal{R}^n) \mapsto \mathcal{I}(\mathcal{R}^m)$ is called an *interval function*. The interval hull of a function is an example of an interval function.

If we have a function f as above and an interval function F and it holds that

$$f(X_1,\ldots,X_n) \ \subseteq \ F(X_1,\ldots,X_n) \quad \text{for} \quad X_i \subseteq A_i \ , \tag{4.2}$$

then F is said to be an *interval extension* of f. In particular, the interval hull is an interval extension.

Example 4.2. We give three examples of interval extensions

$$f(x) = \sin x \ ; \ \ F(X) = [\,-1\,,\,1\,]$$
$$f(x) = x(1-x) \ ; \ \ F(X) = X * (1-X)$$
$$f(x_1,x_2) = (x_1{+}x_2,\ x_1^2) \ ; \ \ F(X_1,X_2) = (X_1{+}X_2,\ X_1{*}X_1)$$

The last function was also treated in Example 4.1. We find $F([\,0\,,\,1\,],[\,0\,,\,1\,]) = ([\,0\,,\,2\,],\ [\,0\,,\,1\,])$. This is the smallest interval that encloses the image, cf. Figure 4.1, so in this case $F(X) = \overline{f}(X)$.

Figure 4.2 illustrates the second example on the interval $X = [\,0\,,\,\frac{1}{2}\,]$. We give the graph of the function $f(x) = x(1-x)$ and the smallest box that contains all points (x,y) for which $x \in X$ and $y \in F(X) = [\,0\,,\,\frac{1}{2}\,]$. The interval hull is $\overline{f}(X) = [\,0\,,\,\frac{1}{4}\,]$.

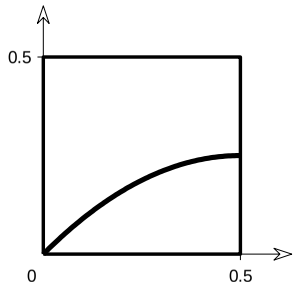


Figure 4.2.

■

4.3. Interval Extensions for Rational Functions

For rational functions there is a simple, general method for obtaining an interval extension: First write a rational expression defining the rational function. Next, simply replace the real operations by interval operations and the variables by intervals.

Example 4.3. We give two examples,

$$f(x) = \frac{1-x^2}{3+x^2} \ ; \ \ F(x) = (1 - X{*}X)/(3 + X{*}X)$$
$$f(x_1,x_2,x_3) = x_1 - x_2x_3 \ ; \ \ F(X_1,X_2,X_3) = X_1 - X_2{*}X_3$$

■

The reason why we can use this simple method to get an interval extension lies in the definition (2.5) of interval operations. We can also use the inclusion monotonicity to show that F is an interval extension: Let $x = (x_1,\ldots,x_n) \in (X_1,\ldots,X_n)$, then

$$F([\,x_1\,,\,x_1\,],\ldots,[\,x_n\,,\,x_n\,]) \ \subseteq \ F(X_1,\ldots,X_n) \ ,$$

and since $f(x_1,\ldots,x_n) = F([\,x_1\,,\,x_1\,],\ldots,[\,x_n\,,\,x_n\,])$, it follows that $f(x_1,\ldots,x_n) \in F(X_1,\ldots,X_n)$ and therefore

$$f(X_1,\ldots,X_n) \ \subseteq \ F(X_1,\ldots,X_n) \ .$$

By use of the algebraic rules for real operators a rational expression can be formulated in many ways. Note, however, that we cannot make the same reformulations of interval expressions, e.g. because of the invalidity of the distributive rule. This implies that different formulations of a rational function will lead to different interval extensions.

Example 4.4. Consider

$$\begin{aligned} f(x) &= 1 - x + x^2 - x^3 + x^4 - x^5 \\ &= (1-x^6)/(1+x) \\ &= (1-x)(1+x^2+x^4) \end{aligned}$$

on the interval $A = [\,2\,,\,3\,]$. It is easy to show that $f'(x) < 0$ for all x, and therefore

$$\overline{f}(A) \;=\; [\,f(3)\,,\,f(2)\,] \;=\; [\,-182\,,\,-21\,]\;.$$

The three formulations of f above give respectively $F_1(A) = [\,-252\,,\,49\,]$, $F_2(A) = [\,-242\tfrac{2}{3}\,,\,-15\tfrac{3}{4}\,]$ and $F_3(A) \;=\; [\,-182\,,\,-21\,]$, i.e., very different intervals. Later we shall see that among the many possible interval extensions some stand out as particularly good. ∎

With an eye on the example one could ask whether it is always possible to reformulate the rational function so that the interval extension is equal to the interval hull. Alas, this is **not** the case. As a counter example consider a rational function of one variable defined on an interval with rational end points, and with irrational upper and/or lower limit of the image, e.g. $x^3 - x$ on $[\,0\,,\,1\,]$. Then it is not possible to find the interval hull by a finite number of real operations on the end points. This shows that if we wish to compute the interval hull to arbitrary accuracy, then we need more tools than just the four interval operations. Such tools are described later.

In all implementations of interval arithmetic you have access to interval extensions of *extended rational functions*. This is the set of functions, where we can compute the exact interval hull, and consist of functions, where the simple arithmetic operators are combined with standard functions like $sin(x)$, e^x, etc.

5. The Set of Intervals as Metric Spaces

Not every interval extension is equally attractive. In some situations we wish that the extension is "close" to the interval hull, but to be meaningful, we must be able to measure the "distance" between two intervals. In other words, we have to organize the set of real intervals and the set of real interval vectors as metric spaces.

First, we have to define a distance function. For $A, B \in \mathcal{I}(\mathcal{R})$ we define the *distance* between the two intervals as

$$d(A,B) \;=\; \max\{\,|\underline{a} - \underline{b}|\,,\,|\overline{a} - \overline{b}|\,\}\;. \tag{5.1}$$

It is easy to show that d is a *metric* in $\mathcal{I}(\mathcal{R})$, i.e., it satisfies

$$\begin{aligned} d(A,B) &= 0 \quad \text{if and only if} \quad A = B \\ d(A,B) &\geq 0 \\ d(A,B) &= d(B,A) \qquad\qquad \text{symmetry} \\ d(A,B) &\leq d(A,C) + d(C,B) \qquad \text{insertion rule} \end{aligned} \tag{5.2}$$

The definition of d does not indicate how far an element in A can be from an element in B, as measured by the usual distance in $\mathcal{R}$. This is of interest, e.g. when A is a computed interval and B is the desired interval, and we know that $d(A,B) \leq \varepsilon$, where $\varepsilon > 0$. In that case, how far is an individual element in A from the elements in B?

In order enable an answer to such questions we reformulate the definition (5.1) to

Lemma 5.1. Let $A, B \in \mathcal{I}(\mathcal{R})$. Then

$$d(A,B) \;=\; \max\Big\{\, \max_{a\in A}\,\{\,\min_{b\in B}\;|a-b|\},\;\; \max_{b\in B}\,\{\,\min_{a\in A}\;|a-b|\}\Big\}$$

Proof. The first minimum can be simplified as follows

$$\min_{b\in B}\;|a-b| \;=\; \begin{cases} 0 & \text{if} \quad a \in B \\ |a - \overline{b}| & \text{if} \quad b < a \quad \text{for all } b \in B \\ |a - \underline{b}| & \text{if} \quad b > a \quad \text{for all } b \in B \end{cases}$$

and the maximum with respect to a is

$$\max_{a\in A}\{\min_{b\in B}|a-b|\} = \begin{cases} 0 & \text{if } A\subseteq B \\ |\overline{a}-\overline{b}| & \text{if } A\geq B \\ |\underline{a}-\underline{b}| & \text{if } A\leq B \\ \max\{|\underline{a}-\underline{b}|,\,|\overline{a}-\overline{b}|\} & \text{if } A\supseteq B \end{cases}$$

The notation $A\geq B$ (respectively $A\leq B$) means that there exists no $a\in A$ so that $a<b$ (respectively $a>b$) and $a\notin B$.

Similarly we get

$$\min_{a\in A}|a-b| = \begin{cases} 0 & \text{if } b\in A \\ |\overline{a}-b| & \text{if } a<b \text{ for all } a\in A \\ |\underline{a}-b| & \text{if } a>b \text{ for all } a\in A \end{cases}$$

and

$$\max_{b\in B}\{\min_{a\in A}|a-b|\} = \begin{cases} 0 & \text{if } B\subseteq A \\ |\overline{a}-\overline{b}| & \text{if } B\geq A \\ |\underline{a}-\underline{b}| & \text{if } B\leq A \\ \max\{|\underline{a}-\underline{b}|,\,|\overline{a}-\overline{b}|\} & \text{if } B\supseteq A \end{cases}$$

Finally, we combine all the possibilities, taking maximum and ignoring the cases that exclude each other. This leads to $\max\{\,|\underline{a}-\underline{b}|\,,\,|\overline{a}-\overline{b}|\,\}$, which is the definition (5.1) of the distance between A and B. □

We can use this lemma to answer one of the questions we raised at the beginning of this chapter:

Lemma 5.2. Let $A,B\in\mathcal{I}(\mathcal{R})$ and $\varepsilon>0$. Then the statement

1° $\quad d(A,B)\ \leq\ \varepsilon$

is equivalent with

2° $\quad(\alpha)$ for all $a\in A$ there exists $b\in B$ such that $|a-b|\leq\varepsilon$

$\quad(\beta)$ for all $b\in B$ there exists $a\in A$ such that $|a-b|\leq\varepsilon$

Proof. We need Lemma 5.1, and first look at 1° ⇒ 2°. Since $d(A,B)\leq\varepsilon$, then $\min_{b\in B}|a-b|\leq\varepsilon$ for all $a\in A$ and $\min_{a\in A}|a-b|\leq\varepsilon$ for all $b\in B$. Thus, 1° implies (α) and (β).

Next, to show that 2° ⇒ 1°, assumption (α) implies that for fixed a it holds that $\min_{b\in B}|a-b|\leq\varepsilon$, and this inequality is also true when we maximize over $a\in A$. Similarly, we use assumption (β) to ensure that the second number in the maximum expression in Lemma 5.1 is bounded by ε, and therefore $d(A,B)\leq\varepsilon$. □

For interval vectors $A=(A_1,\ldots,A_n)$ and $B=(B_1,\ldots,B_n)$ we define the distance as

$$d(A,B) = \max_i\, d(A_i,B_i)\ . \tag{5.3}$$

Thus, also $\mathcal{I}(\mathcal{R}^n)$ is organized as a metric space. It is simple to extend Lemma 5.2 to interval vectors:

Lemma 5.3. Let $A,B\in\mathcal{I}(\mathcal{R}^n)$ and $\varepsilon>0$. Then the statement

1° $\quad d(A,B)\ \leq\ \varepsilon$

is equivalent with

2° $\quad(\alpha)$ for all $a\in A$ there exists $b\in B$ such that $\|a-b\|_\infty\leq\varepsilon$

$\quad(\beta)$ for all $b\in B$ there exists $a\in A$ such that $\|a-b\|_\infty\leq\varepsilon$

Proof. We use Lemma 5.2 and definition (5.3). The assumption $d(A,B)\leq\varepsilon$ implies that $d(A_i,B_i)\leq\varepsilon$ for all i, and thus, for all $a_i\in A_i$ there exist $b_i\in B_i$ such that $|a_i-b_i|\leq\varepsilon$. Therefore, for all $a\in A$ there exists $b\in B$ such that $\|a-b\|_\infty\leq\varepsilon$. Implication (β) is proved the same way, and 2° ⇒ 1° is equally easy. □

5.1. Continuity of Interval Functions

Lemmas 5.2 and 5.3 quantify in terms of the interval elements, what is meant by saying that two intervals are close. Now, we will use the metric concepts to define limits and continuity in $\mathcal{I}(\mathcal{R}^n)$.

A sequence of intervals $\{X_{(s)}\}$ in $\mathcal{I}(\mathcal{R}^n)$ is said to *converge* to X if $d(X_{(s)}, X) \to 0$ for $s \to \infty$. From the definition (5.3) it follows that this means componentwise convergence, i.e. $d(X_{(s)i}, X_i) \to 0$, and from the definition (5.1) of the scalar distance this means that $\underline{x}_{(s)i} \to \underline{x}_i$ and $\overline{x}_{(s)i} \to \overline{x}_i$, i.e, the end points converge to the end points of the limit interval.

An interval function $F : \mathcal{I}(\mathcal{R}^n) \mapsto \mathcal{I}(\mathcal{R}^m)$ is *continuous* for $X \subseteq A \in \mathcal{I}(\mathcal{R}^n)$ if

$$\lim_{s\to\infty} F(X_{(s)}) \;=\; F(X) \quad \text{for} \quad \lim_{s\to\infty} X_{(s)} \;=\; X \;. \tag{5.4}$$

Constant interval functions and mappings of $\mathcal{I}(\mathcal{R}^n)$ into $\mathcal{I}(\mathcal{R}^m)$, where each coordinate in $X \in \mathcal{I}(\mathcal{R}^n)$ is either eliminated or duplicated are continuous. Since *composite mappings* in general metric spaces are continuous, we only need to show that the four interval operations are continuous in order to verify that rational interval functions are continuous. This is done when we have proved the following general theorem about the interval hull for a continuous function,

Theorem 5.1. If f is a continuous mapping of $\mathcal{R}^n$ into $\mathcal{R}^m$, defined on an interval $A \subseteq \mathcal{R}^n$, then the interval hull $\overline{f}$ is a continuous interval function from $\mathcal{I}(\mathcal{R}^n)$ into $\mathcal{I}(\mathcal{R}^m)$.

Proof. We have to show that the continuity condition

$$||f(u) - f(v)||_\infty < \varepsilon \quad \text{for} \;\; u, v \in A \text{ and } ||u - v||_\infty < \delta$$

implies that $d(\overline{f}(U), \overline{f}(V)) < \varepsilon$ for $U, V \subseteq A$ and $d(U, V) < \delta$.

First, we consider the images $f(U)$ and $f(V)$. They are compact, and the mappings of these under the projection of $\mathcal{R}^n$ onto the ith coordinate is also compact, which means that $Y_i = (f(U))_i$ and $Z_i = (f(V))_i$ are

intervals in $\mathcal{R}$. These intervals are what we get by taking the ith coordinate of the interval hull, $Y_i = (\overline{f}(U))_i$, $Z_i = (\overline{f}(V))_i$. Therefore, for $y_i \in Y_i$ there exists $u \in U$ such that $y_i = f(u)_i$. Since $d(U, V) < \delta$, Lemma 5.3 guarantees the existence of $v \in V$ with $||u - v||_\infty < \delta$ and therefore $||f(u) - f(v)||_\infty < \varepsilon$. The number z_i is an element in Z_i, and we have shown that corresponding to y_i there exists an element z_i in Z_i such that $|y_i - z_i| < \varepsilon$.

Starting with z_i and proceeding the same way we get the same inequality, and Lemma 5.2 shows that $d(Y_i, Z_i) < \varepsilon$ when $d(U, V) < \delta$. This holds for every i, and we can deduce that $\max_i d(Y_i, Z_i) = d(\overline{f}(U), \overline{f}(V)) < \varepsilon$. □

The four ordinary operations are continuous mappings of $\mathcal{R}^2$ into $\mathcal{R}$, and therefore the theorem implies that the four interval operations are continuous. Further, this is seen to imply the continuity of any interval function obtained by the method of Section 4.3. In particular, if $X_{(s)} \to x$ with $x \in \mathcal{R}^n$, then $d(F(X_{(s)}), F(x)) \to 0$, and since $f(x) = F(x)$ for these interval extensions, it follows that

$$d(F(X_{(s)}),\; f(x)) \;\to\; 0 \quad \text{for} \quad s \to \infty \;. \tag{5.5}$$

In words: if $X_{(s)}$ is a small argument interval containing x, then $F(X_{(s)})$ is close to the function value $f(x)$.

The first function in Example 4.2 shows that this property does **not** hold for all interval extensions, even if they are continuous. The extension of $sin(x)$ to $F(X) = [\,-1\,,\,1\,]$ is evidently continuous and it does **not** satisfy (5.5).

In the next chapter we shall see that with certain types of interval extensions it is possible to quantify the speed of convergence in (5.5). This is based on an estimate of $d(F(X),\, \overline{f}(X))$.

5.2. Lipschitz Continuity

Now we look at a presumption on interval functions which is stronger than continuity, viz. *Lipschitz continuity*. An ordinary function $f : \mathcal{R}^n \mapsto \mathcal{R}^m$ defined on an interval A is said to be Lipschitz continuous if

there exists a number $K>0$ such that

$$d(f(x),\, f(y)) \;\le\; K\, d(x,y) \quad \text{for} \quad x,y \in A\ . \tag{5.6}$$

This definition is transferred literally to interval functions: An interval function F from $A \in \mathcal{I}(\mathcal{R}^n)$ into $\mathcal{I}(\mathcal{R}^m)$ is said to be Lipschitz continuous if there exists a number $K>0$ such that

$$d(F(X),\, F(Y)) \;\le\; K\, d(X,Y) \quad \text{for} \quad X,Y \subseteq A\ . \tag{5.7}$$

Regarding the interval hull we have

Theorem 5.2. If the function $f : A \subseteq \mathcal{R}^n \mapsto \mathcal{R}^m$ is Lipschitz continuous, then the same is true for the interval hull $\overline{f} : A \in \mathcal{I}(\mathcal{R}^n) \mapsto \mathcal{I}(\mathcal{R}^m)$.

Proof. We have to consider $d(\overline{f}(X),\, \overline{f}(Y))$ for $X, Y \in A$, and as we did before, we look at the ith coordinate. Using Lemma 5.1 we get

$$\begin{aligned} d(\overline{f}(X)_i,\, \overline{f}(Y)_i) \;=\; \max\{ & \max_{x\in X}\{\min_{y\in Y} |f(x)_i - f(y)_i|\}\ , \\ & \max_{y\in Y}\{\min_{x\in X} |f(x)_i - f(y)_i|\}\}\ . \end{aligned}$$

Here we used that for every element $\eta \in \overline{f}(X)$ there is an $x \in X$ such that $\eta_i = f(x)_i$. Next, f being Lipschitz continuous implies that $|f(x)_i - f(y)_i| \le K\, |x_i - y_i|$. The factor K goes outside all the max$^{\text{es}}$ and min$^{\text{s}}$, and another application of Lemma 5.1 leads to

$$d(\overline{f}(X)_i,\, \overline{f}(Y)_i) \;\le\; K\, d(X_i, Y_i) \;\le\; K\, d(X,Y)\ .$$

This is true for all $i = 1, \ldots, m$ and the theorem follows. □

Excepting division by zero, the four arithmetic operations are Lipschitz continuous functions. Therefore, the corresponding interval operations are Lipschitz continuous, and through the same steps that were used to show that rational interval functions are continuous, we can show that they are also Lipschitz continuous. All we need is

Theorem 5.3. If F and G are Lipschitz continuous interval functions, then the composite function $F \circ G$ is Lipschitz continuous.

Proof. $d(F(G(X)),\, F(G(Y))) \;\le\; K_1\, d(G(X),\, G(Y)) \;\le$ $K_1 K_2\, d(X,Y)$ with $K_1, K_2 > 0$, and the theorem is proven. □

In conclusion, rational interval functions are Lipschitz continuous. In particular, this is true for the interval extensions from Section 4.3.

5.3. Width of an Interval

The final basic metric concept to consider is the *width* of an interval (vector). This was defined in Chapter 2,

$$w(A) \;=\; \begin{cases} \overline{a} - \underline{a} & \text{if } A \in \mathcal{I}(\mathcal{R})\ , \\ \max_i\{w(A_i)\} & \text{if } A \in \mathcal{I}(\mathcal{R}^n)\ . \end{cases}$$

For $A \subseteq B$ there is the following important relation between the width and the distance defined in Lemma 5.1,

$$\tfrac{1}{2}(w(B) - w(A)) \;\le\; d(A,B) \;\le\; w(B) - w(A)\ . \tag{5.8}$$

Further, for $A, B, C, D \in \mathcal{I}(\mathcal{R})$, $a \in \mathcal{R}$ we have the following relations for the width (the norm $\|A\|$ is defined in (2.3))

$$\begin{aligned} w(A+B) \;&=\; w(A) + w(B)\ , \\ w(A-B) \;&=\; w(A) + w(B)\ , \\ w(a\,A) \;&=\; \|a\|\, w(A)\ , \\ \max\{\|B\|w(A), \|A\|w(B)\} \;&\le\; w(A * B) \\ &\le\; \|B\|w(A) + \|A\|w(B)\ , \\ w(A/B) \;&\le\; \frac{1}{\underline{b}\,\overline{b}}\left(\|B\|w(A) + \|A\|w(B)\right)\ , \end{aligned} \tag{5.9}$$

and for the distance

$$
\begin{aligned}
d(A+C,B+C) &= d(A,B)\ , \\
d(A+B,C+D) &\le d(A,C)+d(B,D)\ , \\
d(A-B,C-D) &\le d(A,C)+d(B,D)\ , \\
d(A*B,C*D) &\le \|A\|d(B,C)+\|C\|d(A,D)\ , \\
d(1/A,1/B) &\le \frac{1}{\|A\|\,\|B\|}\,d(A,B)\ .
\end{aligned}
\tag{5.10}
$$

We omit the proof of these relations. It is purely technical and rather tedious.

Having introduced the width we can prove an important property of the interval hull of a Lipschitz continuous function,

Theorem 5.4. Let $\overline{f}$ be the interval hull of a Lipschitz continuous function on the interval A. Then, for $X \subseteq A$ there exists $K>0$ such that

$$w(\overline{f}(X)) \;\le\; K\,w(X)\ .$$

Proof. There exist $u, v \in X$ so that

$$
\begin{aligned}
w(\overline{f}(X)) &= \max_i w(\overline{f}(X)_i) \\
&= \max_i |f(u)_i - f(v)_i| \\
&\le K \max_i |u_i - v_i| \;\le\; K\,w(X)\ .
\end{aligned}
$$

Again we exploited that all elements in the ith projection of the hull are mappings of elements in X, and in particular this is true for the end points of $\overline{f}(X)$. □

Not all Lipschitz continuous interval functions satisfy the inequality in Theorem 5.4. This implies that images of singletons are singletons, and $F(X) = [\,-1\,,\,1\,]$ is a simple counter example.

6. Linear and Quadratic Interval Extensions

In this chapter we look at different types of interval extensions, and assess their quality. We start with an example to introduce the concepts that we shall work with.

Example 6.1. The rational function $f(x) = x(1-x)$ has an interval extension $F(X) = X(1-X)$. We concentrate on the intervals

$$X \;=\; [\,\tfrac14 - r\,,\,\tfrac14 + r\,] \qquad \text{for} \quad 0 \le r \le \tfrac14\ ,$$

and from (5.5) we know that $F(X) \to f(\frac14) = \frac{3}{16}$ for $r \to 0$. How fast is the convergence? To find the answer, we make the following calculation,

$$
\begin{aligned}
F([\,\tfrac14 - r\,,\,\tfrac14 + r\,]) &= [\,\tfrac14 - r\,,\,\tfrac14 + r\,] * (1 - [\,\tfrac14 - r\,,\,\tfrac14 + r\,]) \\
&= [\,\tfrac{3}{16} + r^2 - r\,,\,\tfrac{3}{16} + r^2 - r\,] \\
&= \tfrac{3}{16} + r^2 + [\,-r\,,\,r\,]
\end{aligned}
$$

The function f is increasing on X, so it is easy to find its image,

$$\overline{f}(X) \;=\; [\,f(\tfrac14 - r)\,,\,f(\tfrac14 + r)\,] \;=\; \tfrac{3}{16} - r^2 + \tfrac12\,[\,-r\,,\,r\,]\ ,$$

and the distance between $F(X)$ and $\overline{f}(X)$ is

$$d(F(X), \overline{f}(X)) \;=\; \tfrac12 r + 2r^2 \;=\; \mathsf{O}(r)\ .$$

We shall see that this is generally valid: Interval extensions obtained simply by replacing real arguments by their intervals have "errors" that go to zero linearly with the width of the intervals.

By means of the *mean value theorem* we can get a better approximation to $\overline{f}$. For $x \in X$ the theorem says that

$$f(x) \;=\; f(\tfrac14) + (x - \tfrac14) f'(\xi)\ ,$$

for some ξ between x and $\frac14$. Since $f'(\xi) = 1 - 2\xi$, we see that

$$f(x) \;\in\; F_M(X) \;\equiv\; f(\tfrac14) + (X - \tfrac14)(1 - 2X)\ ,$$

i.e., F_M is an interval extension of f on the interval X. As regards the difference between $F_M(X)$ and $\overline{f}(X)$, we find

$$
\begin{aligned}
F_M([\,\tfrac14 - r\,,\,\tfrac14 + r\,]) &= \tfrac{3}{16} + [\,-r\,,\,r\,] * [\,\tfrac12 - 2r\,,\,\tfrac12 + 2r\,] \\
&= \tfrac{3}{16} + \tfrac12[\,-r\,,\,r\,] + 2[\,-r^2\,,\,r^2\,]\ ,
\end{aligned}
$$

and $d(F_M(X), \overline{f}(X)) = 3r^2$, which is better than $d(F(X), \overline{f}(X))$ for small values of r. We shall see that in general $F(X)$ has an "error" $\mathsf{O}(w(X))$, while the "error" with $F_M(X)$ is $\mathsf{O}(w(X)^2)$. ∎

6.1. Linear Interval Extensions

If F is an interval extension of f on an interval $A \subseteq \mathcal{I}(\mathcal{R}^n)$ and there exists a $K > 0$, independent of X, so that

$$d(F(X), \overline{f}(X)) \;\leq\; K\, w(X) \quad \text{for all } X \subseteq A\ , \tag{6.1}$$

then we say that F is a *linear interval extension*. It is obvious that a linear interval extension satisfies the convergence criterion (5.5) for $w(X) \to 0$.

The first, simple result about linear interval extensions is

Theorem 6.1. If F is a linear interval extension of a Lipschitz continuous function f, then there exists a constant $K > 0$ such that

$$w(F(X)) \;\leq\; K\, w(X)\ .$$

Proof. We make use of $\overline{f}(X) \subseteq F(X)$, Equations (5.8) and (6.1), and Theorem 5.4,

$$\begin{aligned} w(F(X)) \;&\leq\; 2d(F(X), \overline{f}(X)) + w(\overline{f}(X)) \\ &\leq\; 2K_1 w(X) + K_2 w(X) \;=\; K\, w(X)\ . \end{aligned}$$

□

Next, we look at composite functions,

Theorem 6.2. If F and G are linear interval extensions of Lipschitz continuous mappings $f : A \subseteq \mathcal{R}^n \mapsto \mathcal{R}^m$ and $g : F(A) \subseteq \mathcal{R}^m \mapsto \mathcal{R}^p$, respectively, then $F \circ G$ is a linear interval extension of $f \circ g$.

Proof. Let $X \subseteq A$ and use the insertion rule to get

$$\begin{aligned} &d(G(F(X)), \overline{f \circ g}(X)) \;\leq \\ &\quad d(G(F(X)), \overline{g}(F(X))) + d(\overline{g}(F(X)), \overline{g}(\overline{f}(X))) \\ &\quad + d(\overline{g}(\overline{f}(X)), g(f(x))) + d(g(f(x)), \overline{f \circ g}(X))\ , \end{aligned}$$

where x is an arbitrary point in X. We now estimate each of the four terms,

$$d(G(F(X)), \overline{g}(F(X))) \;\leq\; K_1 w(F(X)) \;\leq\; K_1 K_2 w(X)\ .$$

Here we used that G is a linear interval extension and that F satisfies the assumptions of Theorem 6.1. In the second term we use the Lipschitz continuity of $\overline{g}$ and the linearity of F,

$$d(\overline{g}(F(X)), \overline{g}(\overline{f}(X))) \;\leq\; K_3 d(F(X), \overline{f}(X)) \;\leq\; K_3 K_4 w(X)\ .$$

In the third term we make use of $g(f(x)) \in \overline{g}(\overline{f}(x))$, $w(g(f(x))) = 0$, the relation (5.8) between w and d, and Theorem 5.4,

$$\begin{aligned} d(\overline{g}(\overline{f}(X)), g(f(x))) \;&\leq\; w(\overline{g}(\overline{f}(X))) \\ &\leq\; K_5 w(\overline{f}(X)) \;\leq\; K_5 K_6 w(X)\ . \end{aligned}$$

Finally, $\overline{f \circ g}$ is Lipschitz continuous and we apply the same technique to the last term,

$$d(g(f(x)), \overline{f \circ g}(X)) \;\leq\; w(\overline{f \circ g}(X)) \;\leq\; K_7 w(X)\ .$$

Combining these estimates we see that $F \circ G$ satisfies (6.1), and the proof is finished. □

Constants in the form of singletons and the four interval operations are linear interval extensions. In both cases this is because we work with interval hulls. Also mappings that duplicate variables or eliminate them are linear interval extensions. This implies that the rational interval functions treated in Section 4.3 are linear interval extensions of extended rational functions.

In some cases the definition (6.1) of a linear interval extension may be reformulated to:

There exists $E \in \mathcal{I}(\mathcal{R}^n)$ with $0 \in E$ and $w(E) \leq K\, w(X)$ such that

$$F(X) \;=\; \overline{f}(X) + E \;. \tag{6.2}$$

In words, every point in $F(X)$ is the sum of two vectors, one of which is in the interval hull and the other has limited length.

6.2. Mean Value Forms

Until now we have seen one general method for constructing interval extensions, viz. interval extensions for rational functions. Now, we shall see how the mean value theorem can be used to construct interval extensions for a much wider class of functions, the differentiable functions.

Consider $f : \mathcal{R}^n \mapsto \mathcal{R}$, which is differentiable in $A \subseteq \mathcal{R}^n$. For such a function the mean value theorem says

$$f(x) \;=\; f(\hat{x}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\xi)\,(x_i - \hat{x}_i) \;,$$

with $\hat{x}, x \in A$ and ξ is a point on the line segment between $\hat{x}$ and x. Let F' be an interval extension of the gradient

$$f' \;=\; \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right) \;,$$

and let $X \subseteq A$ be an interval that contains both $\hat{x}$ and x. Then

$$f(x) \;\in\; f(\hat{x}) + F'(\hat{x}) \cdot (X - \hat{x}) \;,$$

where the multiplication in the last term is the inner product of two interval vectors. In this reformulation we got rid of ξ, and the expression only involves intervals and interval operations. In order for this to be an interval extension of f we must accompany the interval X with a specification of the choice of $\hat{x}$.

If we choose $\hat{x} = m(X)$, the *midpoint* of X, we say that the resulting interval function

$$F_M(X) \;=\; f(m(X)) + F'(X) \cdot (X - m(X)) \tag{6.3}$$

is a *mean value form.*

Example 6.2. For $X = [\,-r\,,\, r\,]$ with $r > 0$ a mean value form of the function *cos* is

$$F_M(X) \;=\; \cos(0) + [\,-1\,,\, 1\,] \cdot (X - 0) \;=\; 1 + [\,-1\,,\, 1\,] * X \;,$$

since $\cos'(\xi) \in [\,-1\,,\, 1\,]$, and this interval constant is therefore an interval extension of *cos'*. F_M is clearly a linear interval extension. ∎

The mean value form is not constructed solely by the four interval operations, but involves getting the midpoint etc. Therefore, it is not obvious that F_M is inclusion monotonic, but the next theorem guarantees that F_M has this property provided that F' does.

Theorem 6.3. If F' is an inclusion monotonic interval extension of f' then F_M is an inclusion monotonic interval extension of f.

Proof. Let $X, Y \subseteq \mathcal{I}(\mathcal{R}^n)$ with $X \subseteq Y$. We have to show that $F_M(X) \subseteq F_M(Y)$. To ease the notation we introduce $x = m(X)$ and $y = m(Y)$, and start by using the mean value theorem on f and the inclusion monotonicity of F',

$$\begin{aligned} F_M(X) \;&=\; f(x) + F'(X)(X - x) \\ &=\; f(y) + f'(\xi)(x - y) + F'(X)(X - x) \\ &\subseteq\; f(y) + f'(\xi)(x - y) + F'(Y)(X - x) \end{aligned}$$

with $\xi \in Y$. We rewrite the current result to

$$F_M(X) \;\subseteq\; f(y) + \sum_{i=1}^{n} \big(f_i'(\xi)(x_i - y_i) + F_i'(Y)(X_i - x_i)\big) \;,$$

and look at the ith contribution to the sum, using that $f_i'(\xi) \in F_i'(Y)$, $x_i = m(X_i)$, and $y_i = m(Y_i)$:

$$\begin{aligned} &f_i'(\xi)(x_i - y_i) + F_i'(Y)(X_i - x_i) \\ &\quad =\; f_i'(\xi)(x_i - y_i) + |F_i'(Y)|\, w(X_i)\,[\,-\tfrac{1}{2}\,,\, \tfrac{1}{2}\,] \;. \end{aligned}$$

It is easily seen that

$$|f_i'(\xi)(x_i - y_i)| \;\leq\; |F_i'(Y)|(w(Y_i) - w(X_i))/2 \;,$$

and therefore

$$\begin{aligned} & f_i'(\xi)(x_i - y_i) + |F_i'(Y)|\, w(X_i)\,[-\tfrac{1}{2}\,,\,\tfrac{1}{2}\,] \\ & \quad \subseteq\ |F_i'(Y)|(w(Y_i) - w(X_i) + w(X_i))\,[-\tfrac{1}{2}\,,\,\tfrac{1}{2}\,] \\ & \quad =\ |F_i'(Y)|\, w(Y_i)\,[-\tfrac{1}{2}\,,\,\tfrac{1}{2}\,]\ =\ F_i'(Y)(Y_i - y_i)\ . \end{aligned}$$

Now the theorem follows immediately. □

If we use a good interval extension for f' we can get better results than illustrated in Example 6.2. This is the subject of the next section, but before that we give two examples.

Example 6.3. As in the previous example we consider $f(x) = \cos(x)$ on $X = [-r\,,\,r]$, $r > 0$. The derivative is $\cos' = -\sin$, and we can get an interval extension for $\cos'$ by using a mean value form for $-\sin$, i.e.

$$\begin{aligned} -\sin(x)\ &=\ -\sin(0) - (x-0)\sin'(\xi) \\ &=\ -x\ \cos(x)\ \in\ X * [-1\,,\,1]\ . \end{aligned}$$

Therefore, $\cos(x) \in 1 + X * X * [-1\,,\,1]$.
Comparing this with the Taylor expansion of $\cos(x)$ around $\hat{x} = 0$, $\cos(x) = 1 - \frac{1}{2}\,x^2 + \mathsf{O}(x^4)$, we see that we have obtained a quadratic approximation to the interval hull of cos on the interval X. ■

Example 6.4. As in Example 4.2 we consider $f(x) = x(1-x)$ on the interval $X = [\,0\,,\,\frac{1}{2}\,]$. Since $f'(x) = 1 - 2x$, we can use $F'(X) = 1 - 2X = [\,0\,,\,1\,]$, and with $m(X) = \frac{1}{4}$, $f(\frac{1}{4}) = \frac{3}{16}$ we get

$$F_M(X)\ =\ \frac{3}{16} + [\,0\,,\,1\,][\,0\,,\,\tfrac{1}{2}\,]\ =\ [\,\tfrac{-1}{16}\,,\,\tfrac{7}{16}\,]\ .$$

Actually, $F'(X) = \overline{f'}(X)$, and the idea from the previous example would give us the same interval extension of f'. In Example 4.2 we found $F(X) = [\,0\,,\,\frac{1}{2}\,]$, and $F_M(X)$ is seen to be a better approximation to the interval hull $\overline{f}(X) = [\,0\,,\,\frac{1}{4}\,]$. ■

Obviously, we can extend the results to functions $f : \mathcal{R}^n \mapsto \mathcal{R}^m$ by using the mean value theorem separately on each coordinate function.

Finally, it should be mentioned that when we use the mean value form in practice on a computer, then the real vector $f(m(X))$ should be represented by an interval vector with bounds on the rounding errors connected with the evaluation of f at the point $m(X)$, i.e. $f(m(X))$ should be represented by $F([\,m(X)\,,\,m(X)\,])$.

6.3. Quadratic Interval Extensions

If F is an interval extension of f on an interval $A \subseteq \mathcal{I}(\mathcal{R}^n)$ and there exists a $K > 0$, independent of X, so that

$$d(F(X),\, \overline{f}(X))\ \leq\ K\, w(X)^2 \quad \text{for all } X \subseteq A\ , \tag{6.4}$$

then we say that F is a *quadratic interval extension*. This is similar to the definition in Section 6.1. The advantage of a quadratic interval extension is when $w(X) \ll 1$, in which case F is almost indistinguishable from the interval hull.

For quadratic interval extensions of Lipschitz continuous functions we have a theorem similar to Theorem 6.1,

Theorem 6.4. If F is a quadratic interval extension of a Lipschitz continuous function f, then there exists a constant $K > 0$ such that

$$w(F(X))\ \leq\ K\, w(X)\ .$$

Proof. Similar to the proof of Theorem 6.1. □

The most important example of a quadratic interval extension is via the mean value form with F' chosen as a linear interval extension.

Theorem 6.5. Let $f : A \in \mathcal{I}(\mathcal{R}^n) \mapsto \mathcal{R}$ be differentiable and assume that the derivatives are Lipschitz continuous. Further, let F' be a linear interval extension of f'. Then

$$F_M(X)\ =\ f(m(X)) + F'(X)(X - m(X))$$

is a quadratic interval extension of f.

Proof. We apply (6.2) to the interval extension F',

$$F'(X) \;=\; \overline{f'}(X) + E\,,$$

where $E \subseteq \mathcal{I}(\mathcal{R}^n)$ with $w(E) \le K_1 w(X)$ and $0 \in E$, implying that $|E| \le w(E) \le K_1 w(X)$. Now we can write

$$F_M(X) \;=\; f(m(X)) + (\overline{f}'(X) + E)(X - m(X))\,,$$

showing that every element $y \in F_M(X)$ has the form

$$y \;=\; f(m(X)) + (f'(u) + e)(v - m(X))$$

with $u, v \in X$ and $e \in E$. A corresponding point in the interval hull $\overline{f}(X)$ is $\overline{y} = f(v)$, and from the mean value theorem we know that there exists a $\xi \in X$ such that

$$\overline{y} \;=\; f(m(X)) + f'(\xi)(v - m(X))\,.$$

We proceed as follows

$$\begin{aligned} d(F_M(X), \overline{f}(X)) \;&\le\; |y - \overline{y}| \\ &=\; \big|\big(f'(u) - f'(\xi) + e\big)\big(v - m(X)\big)\big| \\ &\le\; \sum_{i=1}^{n} \big(|(f'(u)_i - f'(\xi)_i| + |e_i|\big)|v_i - m(X)| \\ &\le\; \sum_{i=1}^{n} \big(K_2 w(X) + K_1 w(X)\big)\tfrac{1}{2}\, w(X) \\ &=\; \frac{n}{2}(K_1 + K_2) w(X)^2\,. \end{aligned}$$

In the third reformulation we used the Lipschitz continuity of f' and the bound on E. Thus, (6.4) is satisfied, and the proof is finished. □

7. Accurate Computation of Interval Hulls

In the previous chapters we have seen some generally applicable interval extensions that can be used to find approximations to the interval hull. Under certain conditions we are guaranteed that the error

$$d(F(X), \overline{f}(X)) \;\le\; K_\alpha w(X)^\alpha\,, \tag{7.1}$$

where $\alpha = 1$ for the rational interval extensions of Section 4.3 and linear interval extensions discussed in Section 6.1, while $\alpha = 2$ for the quadratic interval extensions of Section 6.3. It is possible to construct interval extensions with larger α-values, but it really does not help: we cannot guarantee an arbitrarily good approximation if $w(X) \ge 1$.

On the other hand, if $w(X) \ll 1$, then (7.1) shows that $F(X)$ is very close to $\overline{f}(X)$, and this suggests an idea: Divide the interval X into subintervals, each of which has smaller width than X, compute an interval extension on each subinterval, and gather the information about $F(X)$ from all subintervals. Let us try the idea on a simple problem.

Example 7.1. As in Example 4.2 we look at the function $f(x) = x(1-x)$ on the interval $X = [\,0\,,\,\frac{1}{2}\,]$, with the interval extension $F(X) = X(1-X)$. In Figure 7.1 we show the result obtained when we split the interval into two and four equal parts. Comparing with Figure 4.2 we see that $F(X)$ is reduced from $[\,0\,,\,\frac{1}{2}\,]$ to $[\,0\,,\,\frac{3}{8}\,]$ and $[\,0\,,\,\frac{5}{16}\,]$, respectively, which are closer to the range of $f(X)$.

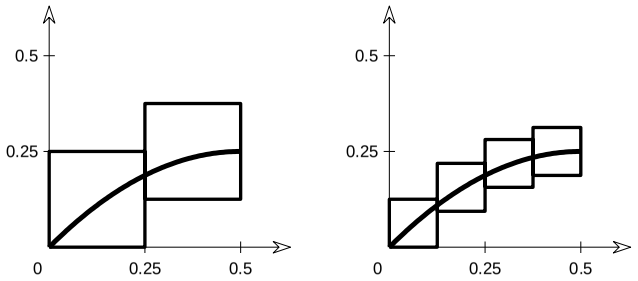


Figure 7.1. *Splitting the interval* ■

We return to this problem later. First, however, we formalize the approach.

Theorem 7.1. Let the function f be defined in $X \subseteq \mathcal{I}(\mathcal{R}^n)$, and let $X_{(1)}, \ldots, X_{(p)}$ be a subdivision of X, i.e.,

$$X_{(j)} \subseteq X \quad \text{and} \quad \bigcup_{j=1}^{p} X_{(j)} \;=\; X \;.$$

Further, let F be an interval extension of f on X. Then

$$\overline{f}(X) \;\subseteq\; \bigcup_{j=1}^{p} F(X_{(j)}) \;.$$

Proof. In each subinterval it clearly holds that $\overline{f}(X_{(j)}) \subseteq F(X_{(j)})$, and the theorem follows by taking the union of both sides of the inclusion. □

In other words, given an interval extension of f and a subdivision of the domain X, we get a new interval function

$$F_p(X) \;=\; \bigcup_{j=1}^{p} F(X_{(j)}) \;. \tag{7.2}$$

This is an interval extension of f, provided that we define the union of two intervals as the smallest interval that contains the union. This interval extension is the subject of

Theorem 7.2. Let F be an interval extension of $A \subseteq \mathcal{I}(\mathcal{R}^n)$ and assume that there exists a constant $K > 0$ such that

$$d(F(X), \overline{f}(X)) \;\leq\; K\, w(X)^\alpha \quad \text{for all} \quad X \subseteq A$$

with $\alpha = 1$ or $\alpha = 2$. Then

$$d(F(X), \overline{f}(X)) \;\leq\; K \max_j \{ w(X_{(j)})^\alpha \} \;.$$

Proof. We can find two subintervals with indices l and u so that

$$F_p(X) \;=\; \bigcup_{j=1}^{p} F(X_{(j)}) \;=\; F(X_{(l)}) \cup F(X_{(u)}) \;,$$

and by using the definition of the distance in Lemma 5.1 and the construction of F_p we find

$$\begin{aligned} d(F(X), \overline{f}(X)) &\leq \max\{ d(F(X_{(l)}), \overline{f}(X_{(l)})),\; d(F(X_{(u)}), \overline{f}(X_{(u)})) \} \\ &\leq K \max\{ w(X_{(l)})^\alpha,\; w(X_{(u)})^\alpha \} \\ &\leq K \max_{j=1}^{p} \{ w(X_{(j)})^\alpha \} \;. \end{aligned}$$

□

As a special case consider equidistant subdivision in each coordinate direction, i.e. $w(X_{(j)}) = \frac{1}{q}\, w(X)$, where p, q, and the dimensionality n are related through $p = q \cdot n$. Then the last theorem tells us that

$$d(F(X), \overline{f}(X)) \;\leq\; K\, w(X)^\alpha (n/p)^\alpha \;. \tag{7.3}$$

The right-hand side converges to zero, i.e. $F_p(X) \to \overline{f}(X)$ for $p \to \infty$. This means that now we have a method for computing the interval hull to arbitrary accuracy. In practice, of course, when carried out in floating point on a computer, rounding errors limit the obtainable accuracy. We refer to Section 2.4, and ignore this limitation in what follows.

Example 7.2. As in the previous example we look at

$$f(x) \;=\; x(1-x) \quad \text{on} \quad X \;=\; [\,0\,,\, \tfrac{1}{2}\,] \quad \text{with} \quad \overline{f}(X) \;=\; [\,0\,,\, \tfrac{1}{4}\,] \;.$$

We consider the two interval extensions

$$\begin{aligned} F(X) &\;=\; X(1-X) \;, \\ F_M(X) &\;=\; f(m(X)) + (1-2X)(X - m(X)) \;, \end{aligned}$$

and use equidistant subdivision,

$$X_{(j)} \;=\; [\,\frac{j-1}{p}\,,\, \frac{j}{p}\,] \;, \quad j = 1, \ldots, p \;.$$

We illustrated the choice $F(X)$ in Figures 4.4 and 7.1 for $p = 1$ and $p = 2$. Below, we give results obtained by successive doubling of p.

p	$d(F_p(X), \overline{f}(X))$	$d(F_{Mp}(X), \overline{f}(X))$
1	0.250000	0.187500
2	0.125000	0.046875
4	0.062500	0.011719
8	0.031250	0.002930
16	0.015625	0.000732
32	0.007813	0.000183
64	0.003906	0.000046
128	0.001953	0.000011
⋮	⋮	⋮

As expected, F_p shows linear convergence and F_{Mp} shows quadratic convergence: Each doubling of p reduces the error with a factor 2 and 4, respectively. ■

The proof of Theorem 7.2 suggests a way of reducing the number of times needed to recompute the interval extension on subintervals. The error satisfies

$$d(F(X), \overline{f}(X)) \;\le\; K \max\{w(X_{(l)})^\alpha,\, w(X_{(u)})^\alpha\} \;,$$

and to reduce the distance between $F_p(X)$ and $\overline{f}(X)$ it may not be necessary to refine all subintervals, but only $X_{(l)}$ and $X_{(u)}$, the subintervals that currently define the end points of $F_p(X)$. This idea has been used by Skelboe (1974) to develop an adaptive version, where the subdivision of X is monitored by F.

We want to stop the process when the current approximation is sufficiently close to the interval hull. It is surprisingly simple to get a simple estimate of the error: Because $\overline{f}(X_{(l)}) \subseteq F(X_{(l)})$ and $\overline{f}(X_{(u)}) \subseteq F(X_{(u)})$ we find that

$$\begin{aligned} d(F_p(X), \overline{f}(X)) \;&\le\; \max\{d(F(X_{(l)}), \overline{f}(X_{(l)})),\, d(F(X_{(u)}), \overline{f}(X_{(u)}))\} \\ &\le\; \max\{w(F(X_{(l)})),\, w(F(X_u))\} \;. \end{aligned}$$

Therefore, if the simple *stopping criterion*

$$\max\{w(F(X_{(l)})),\, w(F(X_u))\} \;\le\; \delta$$

is satisfied for the desired value of δ, then $d(F_p(X), \overline{f}(X)) \le \delta$.

If the mean value form F_M is used, then we not only have the advantage of quadratic convergence, but we also know F', i.e. we have information about the partial derivatives. Suppose e.g., that $F'(X_{(j)})_i > 0$, then f is monotonic increasing on $X_{(j)}$ in the direction of x_i, and if we replace the ith coordinate in $X_{(j)}$ with the lower and upper bound, respectively, then the interval hull over $X_{(j)}$ is contained in the union of F_M applied to the two reduced-dimension intervals. Thus, monotony in a coordinate direction gives faster convergence in that direction, but we still have to use nondegenerate intervals when extrema are attained in interior points. Moore (1975) and Skelboe (1977) report considerable gain from using such information about the gradient.

We give more details in Chapter 11, where we focus on finding the lower bound of $\overline{f}(X)$.

8. Interval Integration

Given a continuous function $f : \mathcal{R} \mapsto \mathcal{R}$ on the interval $A = [\, a\,,\, b\,]$. We want to compute the integral

$$T \;=\; \int_a^b f(x)\, dx \;.$$

This problem arises in many connections, and in some cases it is easy: If we can find a function φ such that $\varphi'(x) = f(x)$, then T is simply evaluated as $T = \varphi(b) - \varphi(a)$. If we cannot find such a function – or if an implementation of it is not available and it is very complicated to develop, then we can resort to numerical quadrature, and get an approximation $\widetilde{T}$ to the integral. The approximation will be affected by both rounding errors and truncation errors, see Chapter 1. It may also happen that the given f has input error, i.e. it represents some theoretical function $\widehat{f}$ in the sense that $f(x) = \widehat{f}(x) + e(x)$, and we are given bounds on e. What is the effect of all these errors, i.e. what can we say about $|T - \widetilde{T}|$?

8.1. Basic Method

It can come as no surprise that we offer interval analysis as a means of answering all these questions. The mathematical background is the additive property of integration and the mean value theorem for integrals. If $a \le c \le b$, then

$$\int_a^b f(x)\, dx \;=\; \int_a^c f(x)\, dx + \int_c^b f(x)\, dx \;,$$

and there exists a $\xi \in X = [\,\underline{x}\,,\, \overline{x}\,]$ such that

$$\int_{\underline{x}}^{\overline{x}} f(x)\, dx \;=\; w(X)\, f(\xi) \;.$$

Therefore, if we split the interval A into p subintervals with break points $\{a_j\}$,

$$a = a_0 < a_1 < \ldots < a_{p-1} < a_p = b \;, \tag{8.1}$$

then

$$T \;=\; \sum_{i=1}^{p} \int_{a_{i-1}}^{a_i} f(x)\, dx \;=\; \sum_{i=1}^{p} w(A_i)\, f(\xi_i) \;,$$

where $A_i = [\, a_{i-1}\,,\, a_i\,]$ and $\xi_i \in A_i$.

Now, let F be an interval extension of f on A. Then

$$T_i \;\equiv\; w(A_i) f(\xi_i) \;\subseteq\; R_i \;\equiv\; w(A_i) F(A_i) \;,$$

and we have derived an interval inclusion of T,

$$T \;\subseteq\; R \;\equiv\; \sum_{i=1}^{p} R_i \;. \tag{8.2}$$

The width of the interval R measures how well we have determined the integral. If we take the midpoint to represent T, then we have the following bound on the *truncation error*

$$|T - m(R)| \;\le\; \tfrac{1}{2} w(R) \;. \tag{8.3}$$

Example 8.1. Consider $T = \int_0^1 \frac{1}{1+x}\, dx$ with the solution $T = \ln 2 \simeq 0.69314718$. The integrand is monotonic decreasing in the integration domain $A = [\,0\,,\,1]$, so we can use $F(X) = [\, \frac{1}{1+\overline{x}}\,,\, \frac{1}{1+\underline{x}}\,]$. If we divide A into p subintervals of equal width, then the break points are $a_j = j/p$, $j = 0, 1, \ldots, p$, $w(A_i) = 1/p$, and we find

$$\begin{aligned} R_i \;&=\; \frac{1}{p}\,[\, \frac{p}{p+i}\,,\, \frac{p}{p+i-1}\,] \;, \\ R \;&=\; [\, \frac{1}{p} + \frac{1}{p+1} + \cdots + \frac{1}{2p}\,,\, \frac{1}{p-1} + \frac{1}{p} + \cdots + \frac{1}{2p-1}\,] \;. \end{aligned}$$

Thus, $|T - m(R)| \;\le\; \frac{1}{2}\, w(R) \;=\; \frac{1}{2}\big(\frac{1}{p-1} - \frac{1}{2p}\big) \;=\; \frac{1}{4p}$, showing that $|T - m(R)| \le 0.005$ if $p \ge 50$. ∎

In this simple example we used the interval hull to represent the variation of f on each subinterval. In general, assume that F is a linear interval extension of f. Then Theorem 6.1 tells us that

$$w(F(A_i)) \;\le\; K\, w(A_i)$$

for some constant $K>0$, and by means of (5.9) and (8.2) we find

$$\begin{aligned} w(R) \;&=\; \textstyle\sum_i w(R_i) \;=\; \sum_i w(A_i)\cdot w(F(A_i)) \\ &\le\; \textstyle\sum_i w(A_i)\cdot K w(A_i) \;=\; K\cdot\sum_i w(A_i)^2 \;. \end{aligned} \tag{8.4}$$

With equidistant break points each $w(A_i)=(b-a)/p$, and (8.4) leads to

$$w(R) \;\le\; K\cdot p\cdot\left(\frac{b-a}{p}\right)^2 \;=\; \frac{K_1}{p}\;. \tag{8.5}$$

Thus, with a good interval extension of f we can expect that a doubling of p reduces the truncation error by a factor 2.

Suppose that we can ignore input and rounding errors, then we could simply increase p (e.g. successively double it) until $w(R)$ is sufficiently small. Often, however, a considerable amount of computation can be saved by focusing on those parts of the range, where $F(X)$ has the greatest width. The framework of such an *adaptive algorithm* could be the following *"divide-and-conquer"* algorithm, which stops when we can guarantee that $|T-m(R)|\le\delta$ with the value of δ chosen by the user,

```
ok := false
while not ok
  R := Σ_i R_i
  if w(R) ≤ 2δ then ok := true
  else
    j := argmax{ w(R_i) }
    Split A_j into two subintervals of equal width
  end
end
```
(8.6)

8.2. Interval Simpson

The method described above behaves similar to the simple methods for "ordinary" numerical integration known as the left and right *Riemann formulas*,

$$\begin{aligned} \widetilde{T}_{(L)} \;&=\; \frac{b-a}{p}\big(f(a_0)+f(a_1)+\cdots+f(a_{p-1})\big)\;, \\ \widetilde{T}_{(R)} \;&=\; \frac{b-a}{p}\big(f(a_1)+\cdots+f(a_{p-1})+f(a_p)\big)\;. \end{aligned}$$

If the integrand is differentiable, we can get accurate approximations with fewer function evaluations. As an example consider *Simpson's formula*. If f is four times continuously differentiable, then

$$T_i \;=\; \frac{w(A_i)}{6}\big(f(\underline{a}_i)+4f(m(A_i))+f(\overline{a}_i)\big) - \frac{w(A_i)^5}{2880}\,f^{(4)}(\xi_i)\;,$$

where $\xi_i\in A_i$. If we have interval extensions F and G of f and $f^{(4)}$, then we can use these to get

$$T \;\subseteq\; S \;\equiv\; \sum_{i=1}^{p} S_i\;, \tag{8.7}$$

with $S_i \;=\; \dfrac{w(A_i)}{6}\big(F(\underline{a}_i)+4F(m(A_i))+F(\overline{a}_i)\big) - \dfrac{w(A_i)^5}{2880}\,G(A_i)\;.$

Note that the calls of F with singleton arguments return intervals that cater for effects of rounding errors and uncertainty of the integrand. Ignoring these, and assuming that G is a linear interval extension of $f^{(4)}$, we get

$$\begin{aligned} w(S_i) \;&=\; \frac{1}{2880}\,w(A_i)^5\,w(G(A_i)) \\ &\le\; \frac{1}{2880}\,w(A_i)^5\,K_G\,w(A_i) \;=\; K_2\,w(A_i)^6\;, \end{aligned}$$

and similar to the derivation of (8.5) we see that in the case of equidistant break points we get $w(S)\le K_3/p^5$, showing that we can expect that a doubling of p reduces the truncation error by a factor 32.

Example 8.2. For the problem of the previous example we can use $G(X) = 24/(1+X)^5$, and below we give the results for equidistant break points. Note that with p subintervals we use $2p+1$ singleton calls of F and p calls of G.

p	$m(S)$	$r(S)$
1	0.6901 47569 44444	4.04e-003
2	0.6930 85397 35741	1.26e-004
4	0.6931 46100 32219	3.94e-006
8	0.6931 47163 08433	1.23e-007
16	0.6931 47180 28433	3.85e-009

Each time p is doubled, the radius $r(S)$ is reduced by the factor 32 predicted above. ∎

We cannot use (8.7) on a problem like $T = \int_0^1 \sqrt{x}\,dx$ because the fourth derivative of the integrand is singular at the left hand end of the range. We could use (8.2), improved by (8.6), but we shall assume that such a singularity occurs only at one and/or the other end, and suggest the following *hybrid method*, where the Riemann approach is used in the two "boundary intervals" and the Simpson approach is used in the interior,

$$T \subseteq H \equiv R_1 + R_p + \sum_{i=2}^{p-1} S_i \,. \tag{8.8}$$

We can combine this with the divide-and-conquer approach of (8.6), with special handling of the cases where the critical subinterval is a boundary interval, $j=1$ or $j=p$. In the first case let $[\,a_0\,,\,a_1\,]$ be the current first interval with length $\Delta = a_1 - a_0$. We split this into the new boundary interval $[\,a_0\,,\,a_0+\Delta/32\,]$ and the interior interval $[\,a_0+\Delta/32\,,\,a_1\,]$. Similar if the critical subinterval is at the other end.

In the implementation it is exploited that $F(\overline{a}_i) = F(\underline{a}_{i+1})$ and when an interior interval is split, then the midpoint is an end point of both new subintervals, so we only have to evaluate F at the two new midpoints and G on the two new subintervals. Also the special handling of the boundary intervals calls for 4 evaluations of an interval function. The initial layout has $p=3$ (one interior interval) and break points $a_{0:3} = (a,\ a+\Delta,\ b-\Delta,\ b)$, where $\Delta = (b-a)/32$, and the first H needs 6 evaluations of an interval function.

Example 8.3. For $f(x) = \sqrt{x} = x^{1/2}$ we can use $G(X) = -\frac{15}{16}\,X^{-7/2}$, and with the stopping criterion $r(H) \le 10^{-3}$ we found the result

$$H \;=\; [\,\texttt{0.6663 33615}\,,\,\texttt{0.6685 63604}\,]$$

obtained with 40 interval function evaluations. The figure shows the points $(x, F(x))$ that were used in Simpson's formula. As expected, they concentrate close to the singularity at $x=0$.

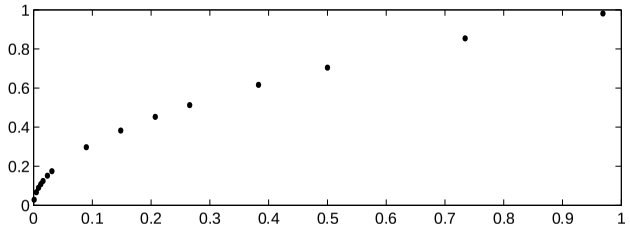


Figure 8.1. ∎

Example 8.4. For comparison, we have also used the MATLAB6 function `quad` on the squareroot problem. As our interval algorithm, this is an adaptive method based on Simpson's formula and bisection of "critical" subintervals. The algorithm stops when two consecutive approximations differ less than δ, the desired accuracy. In the table below we give results for the number of subintervals and the errors for different choices of the desired accuracy.

The two algorithms use about the same number of subintervals, but while `quad` uses $2p+1$ evaluations of a floating point implementation of f, the interval algorithm uses $4p-6$ evaluations of interval functions, which is more than twice as costly. On the other hand, the interval algorithm returns the result to the desired accuracy, except for the last example. We used the "safety valve" $p \le 1000$, while `quad` may proceed until the number of function evaluations exceeds 10 000. In all cases (except for the first, where $p=2$ is sufficient) the error on the `quad` result is up to almost 10 times the desired accuracy.

The true value of the integral, $T = \frac{2}{3}$ is in an interval of width $2^{-53} \simeq$ `1.11e-16` between two floating point numbers. With that in mind, both methods achieve impressive accuracy.

	Interval		quad	
δ	p	$\lvert m(R)-T\rvert$	p	$\lvert\widetilde{S}-T\rvert$
1e-02	6	2.71e-04	2	8.91e-03
1e-04	12	4.54e-05	8	3.94e-04
1e-06	23	7.07e-07	18	6.34e-06
1e-08	52	4.48e-09	48	3.58e-08
1e-10	120	2.86e-11	110	5.63e-10
1e-12	296	1.24e-13	280	3.34e-12
1e-14	805	6.66e-16	700	4.90e-14
1e-16	1000	4.44e-16	1740	7.77e-16

■

The interval version of Simpson's method needs a good interval extension of $f^{(4)}(x)$, or the use of *automatic differentiation*, see e.g. Neumaier (2001) or Corliss *et al* (2002). Chapter 8 in Moore (1966) discusses further interval integration methods based on Taylor expansions of the integrand. Also here, automatic differentiation is useful, and he gives a short introduction in his Chapter 11.

9. Roots of Functions

Looking for a root of a function $f : A \subseteq \mathcal{R}^n \mapsto \mathcal{R}^m$ means that we wish to find a point $x^* \in A$ such that $f(x^*) = 0$. Variants of the problem is to find all roots in A, etc, etc.

9.1. Basic Method

When we use interval methods to find a root we need an algorithm that can tell us whether a given interval $X \subseteq \mathcal{I}(\mathcal{R}^n)$ contains no roots and an algorithm that can give us intervals that contain roots and are as narrow as we wish.

The first algorithm is provided by the simple

Theorem 9.1. If F is an interval extension for $f : A \subseteq \mathcal{I}(\mathcal{R}^n) \mapsto \mathcal{R}^m$ and $X \subseteq A$ with $0 \notin F(X)$, then f has no roots in X.

Proof. Straightforward, since $0 \notin F(X) \;\Rightarrow\; 0 \notin \overline{f}(X)$. □

If we combine this result with the subdivision scheme of Chapter 7, we have the framework for a simple algorithm. Divide A into subintervals, $A = \bigcup_{j=1}^{p} X_{(j)}$, and on each subinterval check whether $0 \in F(X_{(j)})$. If not, then this subinterval is discarded, and the remaining $X_{(j)}$ are further subdivided. If F is a good approximation to $\overline{f}$ and the set of roots in A for f is benign (e.g. finite), then the method results in a collection of arbitrarily narrow intervals that may contain roots, and the remainder of A is guaranteed to contain no roots.

Example 9.1. Let $A = [-1\,,\,2]$, $f(x) = x - x^2$ and $F(X) = X(1-X)$. We get $F(A) = [-2\,,\,4]$, so there may be roots in A. We split the interval into $X_{(1)} = [-1\,,\,\frac{1}{2}]$, $X_{(2)} = [\frac{1}{2}\,,\,2]$, and find that 0 is contained in both $F(X_{(1)})$ and $F(X_{(2)})$. So we split again and get the results tabulated below .

j	1	2	3	4
$X_{(j)}$	$[-1\,,\,-\frac{1}{4}\,]$	$[-\frac{1}{4}\,,\,\frac{1}{2}\,]$	$[\,\frac{1}{2}\,,\,\frac{5}{4}\,]$	$[\,\frac{5}{4}\,,\,1]$
$F(X_{(j)})$	$[-2\,,\,\frac{-5}{16}\,]$	$[\,\frac{-5}{16}\,,\,\frac{5}{8}\,]$	$[\,\frac{-5}{16}\,,\,\frac{5}{8}\,]$	$[-2\,,\,\frac{-5}{16}\,]$

Thus, after the computation of 7 interval extensions we can discard $X_{(1)}$ and $X_{(4)}$, while both of the two neighbouring intervals $X_{(2)}$ and $X_{(3)}$ may contain roots. Further use of the algorithm will isolate the two roots of f, 0 and 1. ∎

Example 9.2. The method also works for multiple roots. Consider e.g. $f(x) = x^2 - 4x + 4 = (x-2)^2$ on $A = [-4\,,\,8]$ with the interval extension $F(X) = X * (X-4) + 4$. We get

j	1	2	3
$X_{(j)}$	$[-4\,,\,0]$	$[\,0\,,\,4]$	$[\,4\,,\,8]$
$F(X_{(j)})$	$[\,4\,,\,36\,]$	$[-12\,,\,4\,]$	$[\,4\,,\,36\,]$

We can discard $X_{(1)}$ and $X_{(3)}$ and subdivide the interval $X_{(2)}$, which contains the double root. ∎

In the first example the problem is simple and the interval extension is quite good. For a more complicated problem and/or a worse approximation to $\overline{f}(X)$ many subdivisions may be needed before we are able to discard root free subintervals.

Moore and Jones (1977) describe how the method works in multiple dimensions. They halve consecutively in each coordinate direction and keep a list of intervals that still have not been discarded.

In the case $n = m = 1$ the method can be extended to prove existence of simple roots: If f is continuous and F applied to the two neighbouring intervals of $X_{(j)}$ has opposite sign, then $X_{(j)}$ contains at least one root, see Figure 9.1 below.

The existence test can be made more efficient by just examining F at points – singletons. The end points of $F([\,x\,,\,x\,])$ reflect the rounding errors in the computation of $\mathsf{fl}(f(x))$, but except for that we do not have to distinguish between F and $\overline{f}$. Nickel (1967) used this to construct an interval version of the classical bisection algorithm.

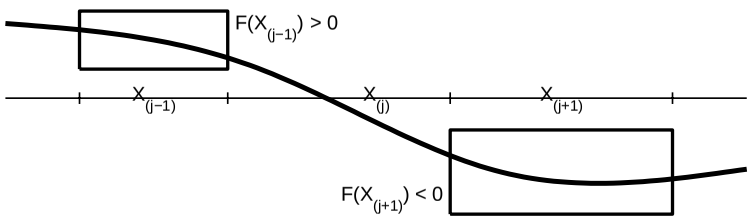


Figure 9.1. *Test of existence of root*

Example 9.3. The existence test only works when $X_{(j)}$ contains an odd number of roots. The problem from Example 9.2 has two roots (viz. the double root 2) in $X_{(2)}$, and therefore $F(X_{(1)}) * F(X_{(3)}) > 0$. ∎

Until now we have presented variants of interval versions of bisection. This method is robust but rather slow. In the next section we present a faster interval algorithm, but before that we present a *hybrid method.* Let $\widetilde{x}$ denote an (approximate) root, computed in ordinary floating point by means of any root finding algorithm, and let δ denote an estimate of the error. This means that we expect that there is a root x^* such that $|x^* - \widetilde{x}| \leq \delta$. In other words, define the interval $[\,a\,,\,b\,] = \widetilde{x} + [-\delta\,,\,\delta]$, and if $F([\,a\,,\,a\,]) * F([\,b\,,\,b\,]) < 0$, then we know that $[\,a\,,\,b\,]$ contains at least one root. If the test is not satisfied, we can extend the interval to $[\,a\,,\,b\,] = \widetilde{x} + 2[-\delta\,,\,\delta\,]$, etc.

9.2. Interval Version of Newton's Method

In this section we discuss the case $n = m = 1$. Consider an interval X that contains a root x^* of f, where f is continuously differentiable. Then the mean value theorem tells us that there exists a $\xi \in X$ such that

$$0 \;=\; f(x^*) \;=\; f(x) + (x^* - x) f'(\xi)\;.$$

Assuming that $f'(\xi) \neq 0$, this leads to

$$x^* \;=\; x - \frac{f(x)}{f'(\xi)}\;.$$

Now, let F' be an interval extension of f' and assume that $F'(X)$ does not contain 0. Then

$$x^* \in N(x,X) \equiv x - \frac{f(x)}{F'(X)} , \tag{9.1}$$

where we have introduced the *Newton operator* $N(x,X)$. The root is also contained in X, and therefore it must lie in the intersection between X and $N(x,X)$,

$$x^* \in X \cap N(x,X) .$$

Thus, we can formulate the interval version of *Newton's method*: Start with $X_{(0)}$ containing x^* and compute a nested sequence of intervals $X_{(1)}$, $X_{(2)}$, ... by the formula

$$X_{(s+1)} = X_{(s)} \cap N(x_{(s)}, X_{(s)}) \quad \text{with} \quad x_{(s)} \in X_{(s)} , \quad s = 0,1,\ldots . \tag{9.2}$$

The term "nested" is used because each new interval is contained in the previous, $X_{(0)} \supseteq X_{(1)} \supseteq \cdots$. This implies that the widths decrease, $w(X_{(0)}) \geq w(X_{(1)}) \geq \ldots$, and since all the $X_{(s)}$ are contained in $X_{(0)}$, they are bounded. Therefore, there exists a limit X^* containing x^*.

Example 9.4. Consider the function $f(x) = x^2 - 2$ with $f'(x) = 2x$ and $X_{(0)} = [1\,,\,2]$. We choose $x_{(s)} = m(X_{(s)})$ and $F'(X) = 2X$, so that the Newton operator is

$$N(x_{(s)}, X_{(s)}) = m(X_{(s)}) - (m(X_{(s)})^2 - 2)/(2X_{(s)}) .$$

The first step with (9.2) is

$$N(x_{(0)}, X_{(0)}) = \tfrac{3}{2} - \tfrac{1}{4}/[2\,,\,4] = [\,\tfrac{22}{16}\,,\,\tfrac{23}{16}\,] ,$$
$$X_{(1)} = [1\,,\,2] \cap [\,\tfrac{22}{16}\,,\,\tfrac{23}{16}\,] = [\,\tfrac{22}{16}\,,\,\tfrac{23}{16}\,] .$$

Below we show results obtained by the MATLAB Toolbox INTLAB, and displayed with `format long`.

s	$X_{(s)}$
1	[1.3749 99999 99998 , 1.4375 00000 00001]
2	[1.4140 62499 99998 , 1.4144 17613 63637]
3	[1.4142 13559 29452 , 1.4142 13565 94719]
4	[1.4142 13562 37309 , 1.4142 13562 37310]

■

The condition that $F'(X)$ does not contain 0 is necessary for the application of the Newton operator, and it has an interesting implication,

Lemma 9.1. If $f : \mathcal{R} \mapsto \mathcal{R}$ is continuously differentiable on the interval X and $0 \notin F'(X)$, then X either contains a simple root x^* or no roots.

Proof. The empty part is obvious, so assume that there is either a multiple root or more than one root in X.

A root x^* of f of muliplicity q satisfies

$$f(x^*) = f'(x^*) = \cdots = f^{(q-1)}(x^*) = 0 .$$

However, the condition on $F'(X)$ implies that $f'(x) \neq 0$ for all $x \in X$. Therefore, X cannot contain a root of multiplicity $q \geq 2$.

Next, consider two (or more) separate, simple roots $x^*, y^* \in X$, $y^* \neq x^*$. They must be separated by a point $z \in X$, where f has a local extremum, i.e. $f'(z) = 0$. But the assumption on F' excludes the existence of such a point. □

In Example 9.4 we had a very fast convergence towards the simple root $x^* = \sqrt{2}$. This exemplified that under the conditions of convergence the limit X^* is the singleton $X^* = x^*$, irrespective of how we choose the points $x_{(s)} \in X_{(s)}$,

Theorem 9.2. If x^* is a root of f in $X_{(0)}$ and $F'(X_{(0)})$ does not contain 0, then the sequence $X_{(1)}, X_{(2)}, \ldots$ defined by (9.2) converges to x^*.

Proof. From (9.2) and the definition (9.1) of the Newton operator we see that

$$\begin{aligned} w(X_{(s+1)}) &\le w(N(x_{(s)}, X_{(s)})) = w(f(x_{(s)}) / F'(X_{(s)})) \\ &= |f(x_{(s)})|\, w(1/F'(X_{(s)})) \\ &\le |f(x_{(s)})|\, w(1/F'(X_{(0)})) = K\,|f(x_{(s)})| \,, \end{aligned} \tag{9.3}$$

where $K = w(1/F'(X_{(0)}))$ is independent of $X_{(s)}$. Since x^* is contained in each $X_{(s)}$, it is possible to choose $\{x_{(s)}\}$ so that $x_{(s)} \to x^*$, in which case (9.3) shows that $w(X_{(s)}) \to 0 \;\Leftrightarrow\; X_{(s)} \to x^*$.

We have to prove that the convergence is obtained for all choices of $x_{(s)} \in X_{(s)}$, and according to (9.3) this is ensured if we can show that it is only possible to choose a finite number of $x_{(s)}$ for which $|f(x_{(s)})| > \varepsilon > 0$. Such a point satisfies

$$\left|\frac{f(x_{(s)})}{F'(X_{(s)})}\right| \ge \frac{|f(x_{(s)})|}{|F'(X_{(s)})|} \ge \frac{|f(x_{(s)})|}{|F'(X_{(0)})|} \,,$$

and $X_{(s+1)} = X_{(s)} \cap \left(x_{(s)} - \dfrac{f(x_{(s)})}{F'(X_{(s)})}\right)$ shows that the distance from $x_{(s)}$ to $X_{(s+1)}$ is at least $\varepsilon/|F'(X_{(0)})|$, and since $x_{(s)} \in X_{(s)}$, this implies that $w(X_{(s+1)}) \le w(X_{(s)}) - \varepsilon/|F'(X_{(0)})|$. Now, $w(X_{(0)})$ is finite, so we can only subtract the positive number a finite number of times, and this concludes the proof. □

The Newton operator can help us answering the question raised in Lemma 9.1 about the existence or non-existence of a roots in a given interval,

Theorem 9.3. If X is an interval with $0 \notin F'(X)$ and there exists an $x \in X$ such that $X \cap N(x, X) = \emptyset$, then f has no root in X.

Proof. By contradiction. If X contained a root, then it would also be contained in $N(x, X)$ and in the intersection. Since $X \cap N(x, X)$ is empty, there can be no root in X. □

Theorem 9.4. If f is continuously differentiable, X is an interval with $0 \notin F'(X)$, and there exists an $x \in X$ such that $N(x, X) \subseteq X$, then f has a root in X.

Proof. This proof is more difficult; it uses Schauder's theorem about existence of fixed points.

Consider the point x and another point $y \in X$. According to the mean value theorem there exists a ξ between x and y so that

$$f(x) = f(y) + f'(\xi)(x - y) \,. \tag{9.4}$$

Since $f'(\xi) \in F(X)$, it is nonzero, and we define the function g,

$$g(y) = y - \frac{f(y)}{f'(\xi)} \,. \tag{9.5}$$

This is continuous, and from (9.4) and the assumption on $N(x, X)$ we get

$$g(y) = x - \frac{f(x)}{f'(\xi)} \in x - \frac{f(x)}{F'(X)} = N(x, X) \subseteq X \,.$$

Thus, g is a continuous mapping of X into itself. According to Schauder's theorem g has a fixed point in X, $y^* = g(y^*)$, and (9.5) shows that this is a root of f. □

Now we are ready to sketch an algorithm that examines an interval A for simple roots:

1. Split A into subintervals $A_{(1)}, \ldots, A_{(p)}$, where each subinterval satisfies either $0 \notin F'(A_{(j)})$ or $w(A_{(j)}) \le \delta$. The second type consists of sufficiently narrow intervals around local extrema and multiple roots.

2. Let $X_{(0)} = A_{(j)}$ be a subinterval with $0 \notin F'(X_{(0)})$. If $X_{(0)} \cap N(x_{(0)}, X_{(0)}) = \emptyset$, then we can discard this subinterval, otherwise proceed with (9.2).

If any of the iterates $X_{(s)}$ satisfies $N(x_{(s)}, X_{(s)}) \subseteq X_{(s)}$, then we are guaranteed that there is a root in $X_{(s)}$, and therefore in $X_{(0)}$. Depending on the choice of $x_{(s)}$, some iteration steps may be needed before the non-existence or the guaranteed existence of a root is revealed.

From the definition (9.1) of the Newton operator and the assumption $0 \notin F'(X)$ it follows that all points in the interval $N(x, X)$ are either to the left or to the right of the point x. This has the important implication

$$w(X \cap N(m(X), X)) \;<\; \tfrac{1}{2} w(X) \; . \tag{9.6}$$

Thus, if we take $x_{(s)}$ as the midpoint of $X_{(s)}$, we have a guaranteed reduction of the interval width. If there is a root in the interval, then we achieve faster final convergence:

Theorem 9.5. If there is a root in $X \subseteq X_{(0)}$, F' is a linear interval extension of f', and $0 \notin F'(X)$, then there exists a constant $K > 0$ such that

$$w((N(x, X)) \;\leq\; K\, w(X)^2 \qquad \text{for all} \quad x \in X \; .$$

Proof. From (9.3) we know that $|w(N(x, X))| \leq K_1 |f(x)|$, and from the mean value theorem and the linearity of F' we get

$$\begin{aligned} |f(x)| \;&=\; |f(x^*) + f'(\xi)(x - x^*) \\ &\leq\; |f'(\xi)|\, w(X) \;\leq\; |F'(X)|\, w(X) \;\leq K_2\, w(X) \cdot w(X) \; , \end{aligned}$$

and the theorem follows. □

This theorem shows that under the given assumptions the sequence (9.2) satisfies $w(X_{(s+1)}) \leq K\, w(X_{(s)})^2$, i.e. it has *quadratic convergence*, as illustrated in Example 9.4.

The method can be made more efficient if we realize that $x_{(s)}$ can be chosen freely in $X_{(s)}$. The estimate (9.3) shows that the width of $X_{(s+1)}$ is specially small if we choose $x_{(s)}$ close to the root, and to reduce the amount of interval computation we can use an ordinary floating point algorithm to find an approximation $\widetilde{x}$ to x^*, and use $x_{(s)} = \widetilde{x}$. The ensuing computation must be done in interval arithmetic.

Example 9.5. It is easily seen that

$$f(x) \;=\; x^3 + x^2 + 3x + 1$$

has a root between -1 and 0. With $X_{(0)} = [-1\,,\, 0]$, $F'X) = 3 + X(2 + 3X)$ and $x_{(s)} = m(X_{(s)})$ we get the following results from (9.2),

s	$X_{(s)}$	$w(X_{(s)})$
1	[-0.4375 00000 00001 , -0.1249 99999 99998]	3.13e-01
2	[-0.3793 69941 54677 , -0.3453 56565 80493]	3.40e-02
3	[-0.3611 36165 68634 , -0.3610 69134 33279]	6.70e-05
4	[-0.3611 03080 55118 , -0.3611 03080 50612]	4.51e-11
5	[-0.3611 03080 52865 , -0.3611 03080 52864]	5.55e-17

Again, we note the quadratic convergence. This is most clearly seen in the column with $w(X_{(s)})$. Rounding errors have dominating influence in the last step.
Three steps with the ordinary floating point Newton algorithm from the starting point -0.5 gives $\widetilde{x} = -0.3611\,03080\,52865$, and using the interval Newton operator $N(\widetilde{x}, X_{(0)})$ we get

$$\begin{aligned} X_{(1)} \;&=\; \widetilde{x} - \frac{f(\widetilde{x})}{F'(X_{(0)}} \\ &=\; [\texttt{-0.3611 03080 52865}\,,\, \texttt{-0.3611 03080 52864}] \\ w(X_{(1)}) \;&=\; \texttt{3.0e-15} \end{aligned}$$

Thus, three floating point steps plus one interval step gives a better result than four interval steps. ■

9.3. Multidimensional Newton's Method

In this section we consider continuously differentiable functions $f : \mathcal{R}^n \mapsto \mathcal{R}^n$, and seek $x^* \in \mathcal{R}^n$ such that $f_i(x^*) = 0$, $i = 1, \ldots, n$. This is a system of n (nonlinear) equations in n unknowns, the components $x_1^*, \ldots, x_n^*$ of x^*.

Again, we start with the mean value theorem, but this time in its multidimensional version. Let $x^* \in X \subseteq \mathcal{I}(\mathcal{R}^n)$ be a root for x. There exist vectors $\xi_{[1]}, \ldots, \xi_{[n]} \in X$ such that

$$0 \;=\; f(x^*) \;=\; f(x) + J(x,x^*)(x^* - x) \;,$$

where $J \in \mathcal{R}^{n\times n}$ is the *Jacobian* with the elements

$$\big(J(x,x^*)\big)_{i,j} \;=\; \frac{\partial f_i}{\partial x_j}(\xi_{[i]}) \;.$$

Assume that the Jacobian is nonsingular for all $x \in X$. Then the inverse of this matrix exists, and

$$x^* \;=\; x - \big(J(x,x^*)\big)^{-1} f(x) \;\in\; x - V(X)f(x) \;, \tag{9.7}$$

where V is an interval matrix containing all the possible J^{-1},

$$\{\big(J(x,x^*)\big)^{-1},\; x \in X\} \;\subseteq\; V(X) \;.$$

Thus, from $X_{(0)}$ containing the root x^* we can construct a nested sequence of intervals, that all contain x^*,

$$\begin{aligned} &X_{(s+1)} \;=\; X_{(s)} \cap N(x_{(s)}, X_{(s)}) \;, \qquad s = 0,1,\ldots \\ &\text{with}\;\; N(x,X) \;=\; x - V(X)f(x), \quad x_{(s)} \in X_{(s)} \;. \end{aligned} \tag{9.8}$$

Example 9.6. Consider the function $f : \mathcal{R}^2 \mapsto \mathcal{R}^2$ defined by

$$f(x) \;=\; \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} \;=\; \begin{pmatrix} x_1^2 + x_2^2 - 1 \\ x_1 - x_2 \end{pmatrix}$$

on the interval $X_{(0)} = ([\,\frac{1}{2}\,,\,1\,],\,[\,\frac{1}{2}\,,\,1\,])$. The Jacobian is

$$J(x,x^*) \;=\; \begin{pmatrix} 2\xi_{11} & 2\xi_{12} \\ 1 & -1 \end{pmatrix}, \quad \xi_{[1]} = (\xi_{11},\; \xi_{12}) \;.$$

This is nonsingular if $\xi_{11}+\xi_{12} \neq 0$, which is satisfied for all $\xi_{[1]} \in X_{(0)}$. The inverse is

$$\big(J(x,x^*)\big)^{-1} \;=\; \frac{1}{\xi_{11}+\xi_{12}} \begin{pmatrix} \frac{1}{2} & \xi_{12} \\ \frac{1}{2} & -\xi_{11} \end{pmatrix} \;,$$

and we can use $V(X) \;=\; \frac{1}{X_1+X_2} \begin{pmatrix} \frac{1}{2} & X_2 \\ \frac{1}{2} & -X_1 \end{pmatrix}$. If we choose the midpoints $x_{(s)} = m(X_{(s)})$ in (9.8), we get the following results

s	$X_{(s),1} = X_{(s),2}$	$w(X_{(s)})$
1	[0.6874 99999 99998 , 0.7187 50000 00001]	3.12e-02
2	[0.7070 31249 99999 , 0.7072 08806 81819]	1.78e-04
3	[0.7071 06779 64726 , 0.7071 06782 97359]	3.33e-09
4	[0.7071 06781 18654 , 0.7071 06781 18655]	1.11e-16
5	[0.7071 06781 18654 , 0.7071 06781 18655]	1.11e-16

We see a fast (quadratic) convergence to the solution, $x_1^* = x_2^* = \sqrt{0.5}$. The width of $X_{(4)}$ is half of the *machine accuracy*, and we cannot get closer to x^* because of rounding errors. ∎

Example 9.7. Now let $f : \mathcal{R}^2 \mapsto \mathcal{R}^2$ be defined by

$$f(x) \;=\; \begin{pmatrix} x_1^2 + x_2^2 - 5 \\ x_1 x_2 - 2 \end{pmatrix}$$

on the interval $X_{(0)} = ([\,1.6\,,\,2\,],\,[\,1\,,\,1.4\,])$. It has the root $x^* = (2,\;1)$. The Jacobian is $J(x,x^*) \;=\; \begin{pmatrix} 2\xi_{11} & -2\xi_{12} \\ -\xi_{22} & \xi_{21} \end{pmatrix}$. It is nonsingular in $X_{(0)}$ and we can use

$$V(X) \;=\; \frac{1}{2(X_1X_1 - X_2X_2)} \begin{pmatrix} X_1 & -2X_2 \\ -X_2 & 2X_1 \end{pmatrix} \;,$$

If we take $x_{(s)} = m(X_{(s)})$ in (9.8), we get the sequence (displayed with fewer digits than before)

s	$X_{(s),1}$	$X_{(s),2}$	$w(X_{(s)})$
1	[1.9386 66666 , 2]	[1 , 1.061 33334]	6.13e-02
2	[1.9984 51819 , 2]	[1 , 1.0015 48181]	1.55e-03
3	[1.9999 99001 , 2]	[1 , 1.0000 00998]	9.98e-07
4	[1.9999 99999 , 2]	[1 , 1.0000 00001]	4.15e-13
5	[1.99999 99999 , 2]	[1 , 1.0000 00001]	2.22e-16
6	[2 , 2]	[1 , 1]	0

The only new feature is that now the components of the root can be represented in floating point without error.

If, instead we take $x_{(0)} = [\,2\,,\,1.4\,]$, we get

$$N(x_{(0)}, X_{(0)}) \;=\; \begin{pmatrix} [\texttt{0.6666 66666}\,,\,\texttt{3.6106 66667}] \\ [\texttt{-1.106 66667}\,,\,\texttt{2.0933 33334}] \end{pmatrix} \;\supseteq\; X_{(0)} \;.$$

Therefore, the sequence defined by (9.8) gets stuck at $X_{(0)}$, and it shows that the choice of $x_{(0)} \in X_{(0)}$ can have great importance. ■

The last example showed that the sequence defined by (9.8) does not necessarily converge to a singleton. A possible cause is that $V(X)$ may contain singular matrices, in which case the interval $V(X_{(s)})f(x_{(s)})$ can contain the zero vector even if $f(x_{(s)}) \neq 0$, and $x_{(s)} \in N(x_{(s)}, X_{(s)}) \subseteq X_{(s+1)}$ with $x_{(s)} \neq x^*$. If we choose $x_{(s+1)} = x_{(s)}$, both $x_{(s)}$ and x^* are contained in $X^* = \lim_{s\to\infty} X_{(s)}$, and therefore $w(X^*) > 0$. However, the following theorem is valid,

Theorem 9.6. Let x^* be a root of f in $X_{(0)}$ and assume that V is inclusion monotonic on $X_{(0)}$. If the $\{x_{(s)}\}$ are chosen so that $\lim\limits_{s\to\infty} x_{(s)} = x^*$ then $\lim\limits_{s\to\infty} X_{(s)} = x^*$.

Proof. $X_{(s+1)} \subseteq x_{(s)} - V(X_{(s)})f(x_{(s)}) \subseteq x_{(s)} - V(X_{(0)})f(x_{(s)})$, and the theorem follows from $f(x_{(s)}) \to 0$. □

This theorem has an important practical implication: We can use an ordinary numerical method to compute the sequence $\{x_{(s)}\}$ and combine it with (9.8) to bound the error. The method can e.g. be the ordinary Newton method,

$$x_{(s+1)} \;=\; x_{(s)} - \big(J(x_{(s)}, x_{(s)})\big)^{-1} f(x_{(s)}) \;, \tag{9.9}$$

where $J(x_{(s)}, x_{(s)})$ is the Jacobian evaluated at $x_{(s)}$. If $x_{(s+1)}$ computed by (9.9) is not contained in $X_{(s+1)}$ as computed by (9.8), then we just replace it by the closest point in $X_{(s+1)}$. If the Newton process converges, so does the sequence $\{x_{(s)}\}$, and the theorem can be applied. As in the one-dimensional case a considerable amount of interval computation can be saved if we just use (9.8) to estimate the error of the results from the ordinary floating point algorithm.

A serious problem with the multidimensional Newton method is how to find V ? In the above examples it was derived from analytical expressions for $\big(J(x, x^*)\big)^{-1}$, but to automate this approach we would need a symbolic language like e.g. MAPLE. Alternatively we could use an interval extension of $J(x, x^*)$ and invert this numerically, as discussed in Chapter 10.

Finally, Theorems 9.3 and 9.4 about the existence of a root generalize to the multidimensional case.

Theorem 9.7. If X is an interval in which $V(X)$ exists and we can find an $x \in X$ so that $X \cap N(x, X) = \emptyset$, then there is no root in X.

Proof. Follows from (9.7). □

Theorem 9.8. If f is continuously differentiable, X is an interval in which $V(X)$ exists, and there exists an $x \in X$ such that $N(x, X) \subseteq X$, then f has a root in X.

Proof. Similar to the proof of Theorem 9.4: For the given x and any $y \in X$ there exists a nonsingular matrix $J(x, y)$ such that

$$f(x) \;=\; f(y) + J(x, y)(x - y) \;.$$

Next, we define $g(y) = y - \big(J(x, y)\big)^{-1} f(y)$. This is continuous in y, $\big(J(x, y)\big)^{-1} \in V(X)$, and we see that

$$g(y) \;=\; x - \big(J(x, y)\big)^{-1} f(x) \;\in\; N(x, X) \;\subseteq\; X \;.$$

The remaining part of the proof is as in the proof of Theorem 9.4. □

9.4. Krawczyk's Version of Newton's Method

Krawczyk (1969) presented a version of Newton's method that avoids the inversion of interval matrices. This method shares the nice properties of Newton's method, but may have slightly slower convergence.

As before, we start by looking at an interval X containing a root x^* of f. Let $x \in X$ and let H be an arbitrary matrix in $\mathcal{R}^{n\times n}$. Then

$$\begin{aligned} x^* &= x - H\,f(x) + (x^* - x) - H\big(f(x^*) - f(x)\big) \\ &= x - H\,f(x) + (I - H\,J(x, x^*))(x^* - x) \\ &\in K(x, X) \equiv x - H\,f(x) + (I - H\,J(X))(X - x)\,. \end{aligned} \tag{9.10}$$

Here, $I = \mathrm{diag}(1, \ldots, 1)$ is the unit matrix, and we introduced the so-called *Krawczyk operator* $K(x, X)$. If we start with an interval $X_{(0)}$ containing a root x^*, then the formula

$$\begin{aligned} &X_{(s+1)} = X_{(s)} \cap K(x_{(s)}, X_{(s)}) \\ &\text{with } x_{(s)} \in X_{(s)} \qquad s = 0, 1, \ldots \end{aligned} \tag{9.11}$$

generates a nested sequence of intervals, all of which contain the root x^*. For this sequence we have the following equivalents to Theorems 9.7 and 9.8,

Theorem 9.9. If there exists an x in the interval X and a matrix $H \in \mathcal{R}^{n\times n}$ such that $X \cap K(x, X) = \emptyset$, then there is no root in X.

Proof. Similar to the proof of Theorem 9.3. □

Theorem 9.10. If there exists an x in the interval X and a nonsingular matrix $H \in \mathcal{R}^{n\times n}$ such that $K(x, X) \subseteq X$, then there is a root in X.

Proof. As in the proof of Theorem 9.4 we look at fixed points. For any $y \in X$ we see that

$$\begin{aligned} y - Hf(y) &= x - Hf(x) + y - x - H(f(y) - f(x)) \\ &= x - Hf(x) + (I - HJ(y, x))(y - x) \\ &\in x - Hf(x) + (I - HJ(X))(X - x) \\ &= K(x, X) \subseteq X\,. \end{aligned}$$

Thus, the continuous function $y - Hf(y)$ maps X into itself, and there exists a fixed point y^*: $y^* = y^* - Hf(y^*)$. Because H is nonsingular, this implies that $f(y^*) = 0$, i.e. there is a root for f in X. □

It follows from (9.11) that if $x_{(s)}$ and $H_{(s)}$ can be chosen so that $K(x_{(s)}, X_{(s)})$ is narrow, then we get fast convergence. More precisely, from (9.11) we find

$$\begin{aligned} w(X_{(s)}) &\le w(K(x_{(s)}, X_{(s)})) \\ &= w((I - H_{(s)}J(X_{(s)}))(X_{(s)} - x_{(s)})) \\ &\le \|I - H_{(s)}J(X_{(s)})\|\, w(X_{(s)}) \equiv \varrho_s w(X_{(s)})\,. \end{aligned}$$

The norm $\|\cdot\|$ of an interval matrix was defined in (2.4). We want to make ϱ_s as small as possible. This is achieved by taking $H_{(s)} = \big(m(J(X_{(s)}))\big)^{-1}$, because then $H_{(s)}J(X_{(s)})$ is centered around I. The choice

$$\begin{aligned} x_{(s)} &= m(X_{(s)}) \\ H_{(s)} &= \begin{cases} \text{an approximation to } \big(m(J(X_{(s)}))\big)^{-1} & \text{if } \varrho_s \le \varrho_{s-1} \\ H_{(s-1)} & \text{if } \varrho_s > \varrho_{s-1} \end{cases} \end{aligned} \tag{9.12}$$

is the subject of

Theorem 9.11. Let the Krawczyk operator use the choice (9.12). If $K(x_{(0)}, X_{(0)}) \subseteq X_{(0)}$ and $\varrho_0 < 1$, then $x^* \in X_{(s)}$, $\lim_{s\to\infty} X_{(s)} = x^*$ and $w(X_{(s)}) \le \varrho_0^s\, w(X_{(0)})$.

Proof. Follows immediately from $w(X_{(s)}) \le \varrho_{s-1} w(X_{(s-1)})$, and the sequence $\{\varrho_s\}$ is monotonic decreasing because of the inclusion monotonicity of $J(X)$ and the choice of $H_{(s)}$ in (9.12). □

Thus, if $G_{(0)} = I - H_{(0)} J(X_{(0)})$ satisfies $\|G_{(0)}\| < 1$, then we are ensured at least linear convergence. Under this condition we may also simplify the process without loosing the linear convergence: Replace (9.11) by

$$X_{(s+1)} \;=\; X_{(s)} \cap \big(x_{(s)} - H_{(0)} f(x_{(s)}) + G_{(0)} (X_{(s)} - x_{(s)}) \big) \;. \qquad (9.13)$$

Then we avoid the computation of $J(X_{(s)})$ and the inversion of its midpoint in each step, but typically this simplification increases the number of steps needed to obtain a satisfactory narrow interval.

Example 9.8. We seek a root for $f(x) = \begin{pmatrix} x_1^2 + x_2^2 - 1 \\ x_1^2 - x_2 \end{pmatrix}$, and assume that we know the approximate root $x_{(0)} = (0.8,\ 0.62)$, e.g. found by a floating point method. We can use Krawczyk's method both to test whether there is a root close to $x_{(0)}$, say in $X_{(0)} = ([0.7\,,\,0.9\,],\,[0.5\,,\,0.7\,])$, and to get a better approximation if this is satisfied.

It is trivial to see that $J(X) = \begin{pmatrix} 2X_1 & 2X_2 \\ 2X_1 & -1 \end{pmatrix}$, and with

$$H \;=\; \begin{pmatrix} 0.279 & 0.346 \\ 0.446 & -0.446 \end{pmatrix} \;\simeq\; \big(J(x_{(0)}) \big)^{-1}$$

we get $G_0 = \begin{pmatrix} [-0.1250\,,\,0.1250\,] & [-0.0446\,,\,0.0670\,] \\ [-0.1784\,,\,0.1784\,] & [-0.0704\,,\,0.1080\,] \end{pmatrix}$,
$\varrho_0 = 0.2864 < 1$, and

$$K(x_{(0)}, X_{(0)}) \;=\; \big([0.7657\,,\,0.8064\,],\,[0.5872\,,\,0.6481\,] \big) \;\subseteq\; X_{(0)} \;.$$

This shows that there is a root in $X_{(0)}$ and we can use (9.13) to find it. In the table below we give the results from this algorithm with $x_{(s)} = m(X_{(s)})$ and the stopping criterion $w(X_{(s)}) \le 10^{-6}$.

s	$X_{(s)1}$	$X_{(s)2}$	$w(X_{(s)})$
1	[0.7657323 , 0.8063645]	[0.5872375 , 0.6480857]	6.08e-02
2	[0.7815712 , 0.7907271]	[0.6111227 , 0.6249431]	1.38e-02
3	[0.7851161 , 0.7871866]	[0.6164709 , 0.6195970]	3.13e-03
4	[0.7859172 , 0.7863856]	[0.6176805 , 0.6183875]	7.07e-04
5	[0.7860984 , 0.7862044]	[0.6179540 , 0.6181140]	1.60e-04
6	[0.7861394 , 0.7861634]	[0.6180159 , 0.6180521]	3.62e-05
7	[0.7861486 , 0.7861541]	[0.6180298 , 0.6180381]	8.18e-06
8	[0.7861507 , 0.7861520]	[0.6180330 , 0.6180350]	1.85e-06
9	[0.7861512 , 0.7861516]	[0.6180337 , 0.6180342]	4.18e-07

The true Krawczyk method (9.11) with the choices (9.12) converges considerably faster; quadratically.

s	$X_{(s)1}$	$X_{(s)2}$	$w(X_{(s)})$
1	[0.7657323 , 0.8063645]	[0.5872375 , 0.6480857]	6.08e-02
2	[0.7850995 , 0.7872034]	[0.6164672 , 0.6196009]	3.13e-03
3	[0.7861485 , 0.7861542]	[0.6180298 , 0.6180382]	8.35e-06
4	[0.7861513 , 0.7861514]	[0.6180339 , 0.6180340]	5.93e-11

■

Example 9.9. The one-dimensional version of the Krawczyk operator is

$$K(x, X) \;=\; x - H f(x) + (1 - H F'(X))(X - x)$$

with $H \simeq 1/f'(x)$ as a good choice. Newton's method (9.2) can be used only if $0 \notin F'(X_{(s)})$, but this version of Krawczyk's method only demands that $f'(x_{(s)}) \neq 0$.
The 1D version of (9.13) is closely related to the ordinary root finding algorithm known as the "saw-tooth method", $x_{(s+1)} = x_{(s)} - G f(x_{(s)})$, where the constant G is (an approximation to) $1/f'(x_{(0)})$. ■

Finally, the remark in Example 9.9 generalizes: with Krawczyk's method we do not demand that all matrices in $J(X_{(s)})$ be invertible. The test of existence in Theorem 9.10 only demands that $H_{(s)}$ is nonsingular, and the test in Theorem 9.9 does not even require this. Therefore, Krawczyk's method can be used to search regions in $\mathcal{R}^n$ for roots. Moore and Jones (1977) used this combined with interval subdivision to get an efficient, global algorithm for root finding.

10. Linear Systems of Equations

This is probably the most common numerical problem, both in its own right and as part of an iterative process as we saw in the previous chapter. In matrix-vector notation the problem is: Given the coefficient matrix $A \in \mathcal{R}^{n\times n}$ and right-hand side vector $b \in \mathcal{R}^n$, find $x \in \mathcal{R}^n$ such that

$$Ax \;=\; b\,. \tag{10.1}$$

We assume that A is nonsingular, in which case the solution x exists and is unique. Now, suppose that the elements in A and b are uncertain, but we know bounds for each of them. We can use these bounds as end points of intervals, and replace (10.1) with

$$AX \;=\; b\,, \tag{10.2}$$

where A and b are the interval matrix and vector as described above, and we seek $X \in \mathcal{I}(\mathcal{R}^n)$ that contains the solution to every problem encompassed in (10.2). We assume that A is *regular*, meaning that every real matrix contained in A is nonsingular.

Example 10.1. Consider the system

$$\begin{aligned} [2\,,\,3]X_1 + [-1\,,\,2]X_2 &= [3\,,\,4] \\ [1\,,\,3]X_1 + [4\,,\,6]X_2 &= [2\,,\,4] \end{aligned}$$

A vector $x = (x_1, x_2) \in X$ must satisfy both equations. The first equation can be reformulated to

$$[2\,,\,3]x_1 \;=\; [3\,,\,4] - [-1\,,\,2]x_2\,,$$

which has the solution

$$x_1 \;=\; \begin{cases} [\,\frac{3}{2} - x_2\,,\, 2 + \frac{1}{2}x_2\,] & \text{for} \quad x_2 \ge \frac{3}{2} \\ [\,1 - \frac{2}{3}x_2\,,\, 2 + \frac{1}{2}x_2\,] & \text{for} \quad 0 \le x_2 < \frac{3}{2} \\ [\,1 + \frac{1}{3}x_2\,,\, 2 - x_2\,] & \text{for} \quad -3 \le x_2 < 0 \\ [\,\frac{3}{2} + \frac{1}{2}x_2\,,\, 2 - x_2\,] & \text{for} \quad x_2 < -3 \end{cases}$$

The set of points (x_1, x_2) given by this expression are marked by horizontal hatching in Figure 10.1. Similarly, we find that the points satisfying the second equation are given by

$$x_2 \;=\; \begin{cases} [\,\frac{1}{2} - \frac{3}{4}x_1\,,\, \frac{2}{3} - \frac{1}{6}x_1\,] & \text{for} \quad x_1 \ge 4 \\ [\,\frac{1}{2} - \frac{3}{4}x_1\,,\, 1 - \frac{1}{4}x_1\,] & \text{for} \quad \frac{2}{3} \le x_1 < 4 \\ [\,\frac{1}{3} - \frac{1}{2}x_1\,,\, 1 - \frac{1}{4}x_1\,] & \text{for} \quad 0 \le x_1 < \frac{2}{3} \\ [\,\frac{1}{3} - \frac{1}{6}x_1\,,\, 1 - \frac{3}{4}x_1\,] & \text{for} \quad x_1 < 0 \end{cases}$$

These points lie in the vertically hatched domain. The set of solutions to the system is the intersection of the two domains, the star-shaped, doubly hatched region.

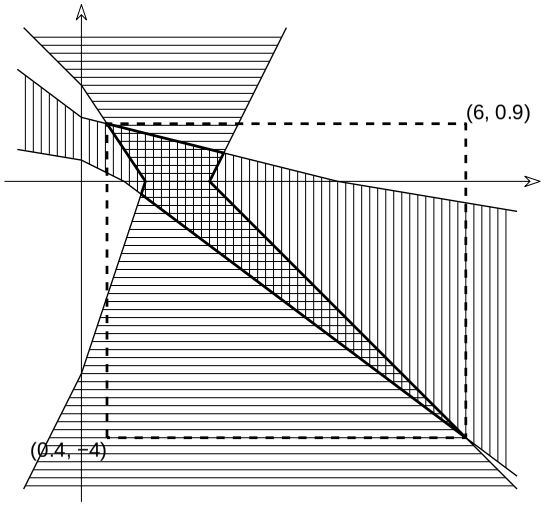


Figure 10.1.

As discussed in Example 4.1, we cannot give such a figure as the result, but have to make do with the interval hull, marked by the dotted rectangle in the figure. Thus, the system has the solution

$$X \;=\; \begin{pmatrix} [0.4\,,\,6] \\ [-4\,,\,0.9] \end{pmatrix}\,. \qquad \blacksquare$$

10.1. Gaussian Elimination

The methods used for ordinary floating point solution of linear systems of equations (at least for small and medium sized problems, $n \lesssim 200$) are variants of Gaussian elimination, possibly with pivoting. To give the background for interval versions we start with a quite detailed description.

Step 1. Transform $Ax=b$ to $Ux=y$, where U is upper triangular. The nonzero elements of U overwrite the upper triangle of A and y overwrites b

```
for k = 1,...,n-1
  p_k = argmax_{i=k,...,n} |A_{i,k}|
  if p_k > k then swap rows k and p_k in A and b
  for i = k+1,...,n
    ℓ_{i,k} = A_{i,k} / A_{k,k}
    A_{i,j} := A_{i,j} - ℓ_{i,k} A_{k,j},   j = k+1,...,n
    b_i := b_i - ℓ_{i,k} b_k
  end
end
```

(10.3)

Step 2. Solve $Ux=y$ by back substitution.

```
for i = n,n-1,...,1
  x_i = (b_i - Σ_{j=i+1}^{n} A_{i,j} x_j) / A_{i,i}
end
```

The matrix $A \in \mathcal{R}^{n\times n}$ is nonsingular iff all the pivots $U_{k,k}$ (stored in $A_{k,k}$) are nonzero.

For the interval system (10.2) the simplest idea is to replace all variables and operations in the Gauss algorithm by their interval equivalent. Unfortunately, this approach often gives very pessimistic results – excessively wide intervals X_j – and too often it breaks down because of the current pivot $A_{k,k}$ containing 0. P. Wongwises (1975) made a very detailed study and a further discussion can be found in Nickel (1977). The essential problem is that some of the matrix elements take part in up to $n-1$ of the transformations $A_{i,j} := A_{i,j} - \ell_{i,k}A_{k,j}$, and each time the relative width of the interval can grow.

Example 10.2. For the problem of Example 10.1,

$$\begin{pmatrix} [2\,,\,3] & [-1\,,\,2] \\ [1\,,\,3] & [4\,,\,6] \end{pmatrix} X \;=\; \begin{pmatrix} [3\,,\,4] \\ [2\,,\,4] \end{pmatrix} ,$$

Step 1 of the interval version of (10.3) without pivoting gives

$$\ell_{2,1} = [\,\tfrac{1}{3}\,,\,\tfrac{3}{2}\,]\,, \quad U = \begin{pmatrix} [2\,,\,3] & [-1\,,\,2] \\ 0 & [1\,,\,7.5] \end{pmatrix}, \quad y = \begin{pmatrix} [3\,,\,4] \\ [-4\,,\,3] \end{pmatrix},$$

and by back substitution in $UX = y$ we get

$$X \;\subseteq\; \begin{pmatrix} [-1.5\,,\,6] \\ [-4\,,\,3] \end{pmatrix} .$$

By comparison with Example 10.1 we see that for this problem we get the correct value for $\overline{x}_1$ and $\underline{x}_2$, but in both intervals the other end point is too pessimistic. ∎

Example 10.3. To demonstrate the over-pessimism about singularity of the matrix, we need a problem of order $n \geq 3$. Consider the problem (10.2) with

$$A \;=\; \begin{pmatrix} [2\,,\,4] & [0\,,\,2] & [-1\,,\,1] \\ 1 & [3\,,\,4] & [-.5\,,\,1.5] \\ [4\,,\,6] & [3\,,\,5] & [8\,,\,10] \end{pmatrix} , \qquad b \;=\; \begin{pmatrix} [8\,,\,10] \\ [10\,,\,12] \\ [8\,,\,10] \end{pmatrix} .$$

By means of the interval version of (10.3) without pivoting we find

$$\ell_{2,1} = [\,.25\,,\,.5\,]\,, \quad \ell_{3,1} = [\,1\,,\,3\,]\,, \quad \ell_{3,2} = [\,-1.5\,,\,2.5\,]\,,$$

$$U = \begin{pmatrix} [2\,,\,4] & [0\,,\,2] & [-1\,,\,1] \\ 0 & [2\,,\,4] & [-1\,,\,2] \\ 0 & 0 & [0\,,\,16] \end{pmatrix} , \qquad \widetilde{b} = \begin{pmatrix} [5\,,\,10] \\ [-47\,,\,17] \\ [-64.5\,,\,27.5] \end{pmatrix} .$$

Since $0 \in U_{3,3}$ we cannot proceed. We will now show, that the zero **can not** be obtained with any of the matrices in A. It follows from (10.3) and the above matrices, that

$$\begin{aligned} U_{3,3} \;&=\; A_{3,3} - \ell_{3,1}A_{1,3} - \ell_{3,2}U_{2,3} \\ &=\; [8\,,\,10] - [1\,,\,3] * [-1\,,\,1] - [-1.5\,,\,2.5] * [-1\,,\,2] \\ &=\; [8\,,\,10] - [-3\,,\,3] - [-3\,,\,5]\,. \end{aligned}$$

To get the zero in $U_{3,3}$ we have to combine the worst case in the two subtrahends. They occur for $A_{1,3}=1$ and $U_{2,3}=2$, but this end point of $U_{2,3}$ is achieved only if $A_{1,3}=-1$, i.e. the other end of the interval $A_{1,3}$. By means of the method described in Section 10.2 below we found that

$$X \;=\; \begin{pmatrix} [-2\,,\,10.0500] \\ [0.1200\,,\,5.7000] \\ [-10.1000\,,\,-0.0258] \end{pmatrix} .$$

∎

Example 10.4. We look at the same problem as in the previous example, except that we change $A_{3,3}$ to $[\,8.5\,,\,10\,]$. The only change in the reduction (Step 1 in (10.3)) is that now $U_{3,3} = [\,0.5\,,\,16\,]$, and proceeding with back substitution we find

$$X \;\subseteq\; \begin{pmatrix} [\,-142\,,\,123.75\,] \\ [\,-44.5\,,\,99\,] \\ [\,-94\,,\,34\,] \end{pmatrix} .$$

The set of all solutions is now contained in

$$X \;=\; \begin{pmatrix} [\,-1.8947\,,\,9.3914\,] \\ [\,0.2037\,,\,5.2632\,] \\ [\,-8.7827\,,\,-0.0258\,] \end{pmatrix} .$$

The bounds obtained by the simple replacement of floating point variables and operations by their interval equivalent are more than 10 times wider than the true interval. ∎

In order to improve the accuracy we may try *Krawczyk's method* from Section 9.4. A solution $x \in X$ to (10.2) is a root of the function $f(x) = Ax - b$, and the Krawczyk operator (9.10) takes the form

$$K(x,X) \;=\; x - H(Ax-b) + (I - HA)(X - x) \;,$$

where $H = \big(m(A)\big)^{-1}$ is the optimal choice for the matrix H. If $\|I - HA\| < 1$ and we know an initial $X_{(0)}$ that contains a solution, then the method should work.

Example 10.5. For the problem of Examples 10.1 and 10.2 we get

$$H \;=\; \begin{pmatrix} 2.5 & 0.5 \\ 2 & 5 \end{pmatrix}^{-1} \;=\; \tfrac{1}{23}\begin{pmatrix} 10 & -11 \\ -4 & 5 \end{pmatrix} ,$$
$$I - HA \;=\; \tfrac{1}{23}\begin{pmatrix} [\,-6\,,\,6\,] & [\,-16\,,\,16\,] \\ [\,-7\,,\,7\,] & [\,-11\,,\,11\,] \end{pmatrix}, \quad \|I - HA\| \;=\; \tfrac{22}{23} \;<\; 1 \;,$$

so the necessary condition for convergence of Krawczyk's method is satisfied. However, with the interval found in Example 10.2, $X_{(0)} = ([\,-1.5\,,\,6\,],[\,-4\,,\,3\,])$ we have not been able to find an $x \in X_{(0)}$ that works. In all the cases tried we get $K(x, X_{(0)}) \supset X_{(0)}$, so Krawczyk's method is stuck at $X_{(0)}$. ∎

Example 10.6. For the problem of Example 10.4 we find $\|I - HA\| = 1.261 > 1$, so we cannot use Krawczyk's method. If we replace A and b by

$$\widehat{A} = m(A) + 0.1\,r(A) = \begin{pmatrix} [\,2.9\,,\,3.1\,] & [\,0.9\,,\,1.1\,] & [\,-0.1\,,\,0.1\,] \\ 1 & [\,3.45\,,\,3.55\,] & [\,0.4\,,\,0.6\,] \\ [\,4.9\,,\,5.1\,] & [\,3.9\,,\,4.1\,] & [\,9.175\,,\,9.325\,] \end{pmatrix},$$
$$\widehat{b} = m(b) + 0.1\,r(b) = \big([\,8.9\,,\,9.1\,],\ [\,10.9\,,\,11.1\,],\ [\,8.9\,,\,9.1\,]\big) \;,$$

then the method works. We get $\|I - HA\| = 0.1261 < 1$, and with

$$X_{(0)} \;=\; \begin{pmatrix} [\,1.7415\,,\,2.4394\,] \\ [\,2.4978\,,\,3.0292\,] \\ [\,-1.6906\,,\,-1.0396\,] \end{pmatrix} ,$$

obtained by the interval version of (10.3), and $x_{(0)} = m(X_{(0)})$ we get

$$K(x_{(0)}, X_{(0)}) \;=\; \begin{pmatrix} [\,1.7429\,,\,2.4319\,] \\ [\,2.5204\,,\,2.9552\,] \\ [\,-1.5919\,,\,-1.0867\,] \end{pmatrix} ,$$
$$X_{(1)} \;=\; X_{(0)} \cap K(x_{(0)}, X_{(0)}) \;=\; \begin{pmatrix} [\,1.7429\,,\,2.4319\,] \\ [\,2.5204\,,\,2.9552\,] \\ [\,-1.5919\,,\,-1.0867\,] \end{pmatrix} ,$$

and proceeding like this with the stopping criterion

$$d(X_{(s)}, X_{(s-1)}) \;\le\; 10^{-12} \;, \tag{10.4}$$

we stop with

$$X_{(12)} \;=\; \begin{pmatrix} [\,1.7519\,,\,2.4229\,] \\ [\,2.5271\,,\,2.9484\,] \\ [\,-1.5854\,,\,-1.0931\,] \end{pmatrix} .$$

This is better than $X_{(0)}$, but still quite far from the true solution, $\widehat{X} = ([\,1.7878\,,\,2.3937\,],\ [\,2.5479\,,\,2.9311\,],\ [\,-1.5783\,,\,-1.1314\,])$.

In the Table below we give results for the problems given by

$$\widehat{A} = m(A) + \gamma r(A) \;, \qquad \widehat{b} = m(b) + \gamma r(b) \;,$$

for decreasing values of γ. For each problem the vector $X_{(0)}$ is the result from the interval version of (10.3), we use $x_{(s)} = m(X_{(s)})$, and *its* is the number of iteration steps before (10.4) is satisfied.

γ	$\|I - H\widehat{A}\|$	*its*	$d(X_{its}, \widehat{X})$
10^{-1}	1.26e-01	12	3.83e-02
10^{-2}	1.26e-02	6	3.71e-04
10^{-3}	1.26e-03	4	3.70e-06
10^{-4}	1.26e-04	3	3.70e-08
10^{-5}	1.26e-05	3	3.70e-10
10^{-6}	1.26e-06	2	3.71e-12
10^{-7}	1.26e-07	2	3.86e-14
10^{-8}	1.26e-08	2	2.66e-15
10^{-9}	1.26e-09	2	2.66e-15

Note how each reduction of the width of the matrix and right-hand side by a factor 10 reduces the distance between the Krawczyk solution and the true solution by a factor 100. The last three results for the distance are increasingly affected by rounding errors. ■

Compared with the discussion in Section 9.4 the results in the last two examples are somewhat disappointing. However, we introduced Krawczyk's method for problems, where the truncation was the dominating error source. The initial error was sufficiently small for the method to converge, and it was reduced with the width of the interval. Now, we apply the method on problems with input errors, that stay fixed throughout the iteration. Example 10.6 illustrates that the method performs well for problems with small input errors. This was also the conclusion of Wongwises (1975), who used the method to compute error bounds for computed solutions to problems of the form (10.1), i.e. she used the method to estimate the effect of rounding errors.

10.2. Sign-Accord Algorithm

This algorithm was presented by Rohn (1989) and is currently considered as the best method for solving systems of linear interval equations. Our presentation is based on Madsen and Toft (1994). First, let $\mathcal{X} = \mathcal{X}(A, b)$ denote the set of solutions to (10.2), i.e.

$$\mathcal{X} \;=\; \{x \in \mathcal{R}^n \;|\; \breve{A}x = \breve{b} \;\text{ for real } \breve{A} \in A,\; \breve{b} \in b\}\;. \tag{10.5}$$

Figure 10.1 illustrates that this set is **not** convex. This property also follows from the following, surprisingly simple characterization of the solution set,

Theorem 10.1. Let $A \in \mathcal{I}(\mathcal{R}^{n\times n})$, $b \in \mathcal{I}(\mathcal{R}^n)$. The solution set of $AX = b$ is

$$\mathcal{X} \;=\; \{x \in \mathcal{R}^n \;|\; |m(A)x - m(b)| \le r(A)|x| + r(b)\}\;. \tag{10.6}$$

Proof. See Oettli and Prager (1964). □

The nonconvexity follows from the factor $|x|$. Suppose that $u, v \in \mathcal{X}$, i.e. each of them satisfy (10.6), but it does **not** follow that any linear combination of them as e.g. $x = u+v$ satisfies it.

Among all the elements in $\mathcal{X}$ the *corner solutions* have special interest. The subset $\mathcal{X}_{\text{Corn}}$ consists of the vectors $\breve{A}^{-1}\breve{b}$, where all elements in $\breve{A}$ and $\breve{b}$ are chosen as one of the end points of the corresponding interval element in A and b, respectively. The number of corner solutions is 2^{n^2+n}.

Example 10.7. For the problem of Example 10.1 we have computed 250 elements of $\mathcal{X}$. 64 of these are the corner solutions, marked by '+" in Figure 10.2 below, and the remaining 186 correspond to random choices of $\breve{A}_{i,j}$ in the interval A and similar for $\breve{b}$.

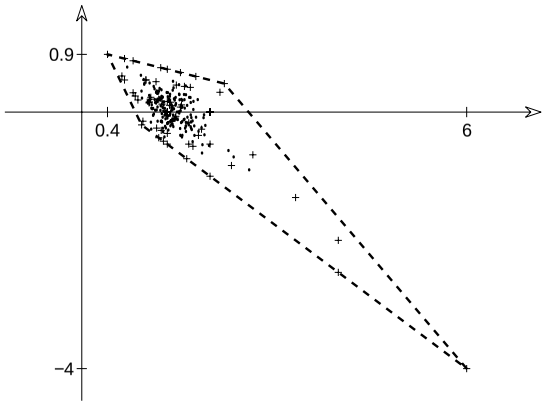


Figure 10.2.

The dotted line borders $\mathrm{Conv}(\mathcal{X})$, the *convex hull* of $\mathcal{X}$. Note that each corner of this polygon is a corner solution of the system. ■

The observation in this example is true in general:

Theorem 10.2. Let $A \in \mathcal{I}(\mathcal{R}^{n\times n})$ be regular and let $b \in \mathcal{I}(\mathcal{R}^n)$. Then

$$\mathrm{Conv}(\mathcal{X}_{\mathrm{Corn}}) \;=\; \mathrm{Conv}(\mathcal{X}) \;. \tag{10.7}$$

Proof. See Toft (1992). □

The solution X to the system (10.2) is an interval vector, and as already mentioned, X is the interval hull of $\mathcal{X}$, $X = \overline{\mathcal{X}}$.

Theorem 10.3. Let $A \in \mathcal{I}(\mathcal{R}^{n\times n})$ be regular and let $b \in \mathcal{I}(\mathcal{R}^n)$. The solution to $AX = b$ is

$$X \;=\; \overline{\mathcal{X}}_{\mathrm{Corn}} \;. \tag{10.8}$$

Proof. The inclusions

$$\mathcal{X} \;\subseteq\; \mathrm{Conv}(\mathcal{X}) \;\subseteq\; \overline{\mathcal{X}}$$

are obvious, and (10.8) follows from $X = \overline{\mathcal{X}}$ and (10.7). □

This theorem was originally proved in Beeck (1972). It shows that the problem is finite: we can find X via the computation of the corner solutions. Each of these is found by solving a real system of order n, and the complexity of this approach is $\mathsf{O}(n^3\, 2^{n^2+n})$, which is prohibitive even for small values of n,

n	2	5	10
$n^3\, 2^{n^2+n}$	$5.12 \cdot 10^2$	$1.34 \cdot 10^{11}$	$1.30 \cdot 10^{36}$

We can do much better. The points where the inequality in (10.6) turns into equality are the socalled *extreme solutions*,

$$\mathcal{X}_{\mathrm{Extr}} \;=\; \big\{x \in \mathcal{R}^n \;\big|\; |m(A)x - m(b)| = r(A)|x| + r(b)\big\} \;. \tag{10.9}$$

In connection with (10.14) below we show that $\mathcal{X}_{\mathrm{Extr}} \subseteq \mathcal{X}_{\mathrm{Corn}}$. We introduce the *sign vector*

$$y \;=\; \mathrm{sgn}\big(m(A)x - m(b)\big) \;.$$

This is an element in $\mathcal{S}^n$,

$$\mathcal{S}^n \;=\; \big\{y \in \mathcal{R}^n \;\big|\; y_i \in \{1,-1\},\; i = 1,\ldots,n\big\} \;.$$

Now, (10.9) can be reformulated to

$$\mathcal{X}_{\mathrm{Extr}} \;=\; \big\{x \in \mathcal{R}^n \;\big|\; m(A)x - m(b) = D_y(r(A)|x| + r(b)) \text{ for some } y \in \mathcal{S}^n\big\} \;, \tag{10.10}$$

where we have introduced the diagonal matrix

$$D_y \;=\; \mathrm{diag}(y_1,\ldots,y_n) \;. \tag{10.11}$$

The next two theorems of Rohn (1989) show that we can reduce the problem from finding the 2^{n^2+n} elements in $\mathcal{X}_{\mathrm{Corn}}$ to finding the 2^n elements in $\mathcal{X}_{\mathrm{Extr}}$.

Theorem 10.4. Let $A \in \mathcal{I}(\mathcal{R}^{n\times n})$ be regular and let $b \in \mathcal{I}(\mathcal{R}^n)$. For each $y \in \mathcal{S}^n$ the nonlinear equation in (10.10) has exactly one extreme solution $x = x_y$, and

$$\mathrm{Conv}(\mathcal{X}_{\mathrm{Extr}}) \;=\; \mathrm{Conv}\big\{x_y \;\big|\; y \in \mathcal{S}^n\big\} \;=\; \mathrm{Conv}(\mathcal{X}) \;. \tag{10.12}$$

Theorem 10.5. Let $A \in \mathcal{I}(\mathcal{R}^{n\times n})$ be regular and let $b \in \mathcal{I}(\mathcal{R}^n)$. Then the set $\mathcal{X}_{\rm Extr}$ of extreme solutions for $Ax = b$ satisfies

$$\overline{\mathcal{X}}_{\rm Extr} \;=\; \overline{\mathcal{X}}\;. \tag{10.13}$$

Example 10.8. For the problem of Examples 10.1, 10.2, 10.5 and 10.7 the extreme solutions are

y	$(1,1)$	$(1,-1)$	$(-1,1)$	$(-1,-1)$
x_y	$(\frac{20}{9},\ \frac{4}{9})$	$(6,\ -4)$	$(0.4,\ 0.8)$	$(\frac{14}{15},\ -0.2)$

The $\{x_y\}$ are the corner points of the polygon in Figure 10.2. ∎

The reduction in the number of points is not for free. Each point in $\mathcal{X}_{\rm Extr}$ is the solution to a nonlinear system of equations. From (10.10) we see that for a given $y \in \mathcal{S}^n$ we have to solve

$$m(A)x - m(b) \;=\; D_y\big(r(A)|x| + r(b)\big)\;.$$

Introducing $z \in \mathcal{S}^n$ defined by $z = \mathrm{sgn}(x)$ (i.e. $|x| = D_z x$) we see that this system is equivalent with

$$\begin{aligned} &A_{xy}x \;=\; b_y \\ &\text{where}\;\; A_{xy} = m(A) - D_y r(A) D_x,\;\; b_y = m(b) + D_y r(b)\;. \end{aligned} \tag{10.14}$$

It is seen that all elements in A_{xy} and b_y are at an end point of the corresponding element in A and b, respectively, so the solution is a corner solution, and we have verified (10.9). Further, if the solution x satisfies $\mathrm{sgn}(x) = z$, then $x = x_y$.

Now, we are ready to formulate the algorithm for computing the solution to $AX = b$.

$$\begin{aligned} &X := \emptyset \\ &\textbf{for}\ \text{each}\ y \in \mathcal{S}^n\ \textbf{do} \\ &\quad \text{find the extreme solution}\ x_y \\ &\quad X := \overline{X \cup \{x_y\}} \end{aligned} \tag{10.15}$$

The extreme solution is found by the *sign-accord algorithm* due to Rohn (1989),

$$\begin{aligned} &\text{given}\ y \in \mathcal{S}^n \\ &\text{choose initial}\ z \in \mathcal{S}^n;\quad fd := \textbf{false} \\ &\textbf{repeat} \\ &\quad \text{solve}\ A_{xy}\,x = b_y \\ &\quad \textbf{if}\quad \text{all}\ z_i x_i \geq 0\ \textbf{then}\ \ x_y := x;\quad fd := \textbf{true} \\ &\quad \textbf{else} \\ &\qquad k := \mathrm{argmin}_j\{z_j x_j < 0\} \\ &\qquad z_k := -z_k \\ &\quad \textbf{end} \\ &\textbf{until}\ \ fd \end{aligned} \tag{10.16}$$

The stopping criterion is that the sign of x is in accordance with z. Rohn (1989) showed that the algorithm stops after at most 2^n iteration steps, and recommends the initial choice $z = \mathrm{sgn}\big(m(A)^{-1}b_y\big)$.

In the best case the first x is in sign-accordance with the initial z, so the complexity of algorithm (10.15) is between $\mathsf{O}(n^3 2^n)$ and $\mathsf{O}(n^3 2^{2n})$. We refer to Madsen and Toft (1994) about further enhancements of the algorithm, aspects of floating point arithmetic, and how the algorithm can be successfully implemented on a parallel computer.

11. Global Optimization

Many technical and economic problems can be formulated as mathematical programming problems, i.e. as the minimization of a continuous nonlinear function $f : \mathcal{D} \mapsto \mathcal{R}$, where $\mathcal{D} \subseteq \mathcal{R}^n$. Often, the function f has several local minima, but we are only interested in the smallest one

$$f^* \;=\; \inf\{f(x) \,|\, x \in \mathcal{D}\} \;. \tag{11.1}$$

f^* is the *global minimum* and the problem of finding f^* and the points $X^* = \{x \in \mathcal{D} \mid f(x) = f^*\}$ where it is attained is called *global optimization*. Methods for global optimization can be divided into two classes, deterministic methods and stochastic methods, see e.g. Pintér (1996). We shall present a deterministic method based on interval analysis.

11.1. Basic Method

The problem of finding the global minimum is closely related to the subject of Chapter 7: f^* is equal to the lower bound of the interval hull $\overline{f}(\mathcal{D})$, and we shall expand the ideas outlined at the end of Chapter 7, combined with relevant algorithms from Chapters 8 and 9. The presentation is based on Madsen (1991) and proof of convergence may be found in Moore and Ratschek (1988).

The algorithm is a branch and bound type method. At any stage of the algorithm we have the *candidate set* C. This is a finite set of subregions $C_{(j)} \subseteq \mathcal{D}$ with the set X^* of minimizers of f contained in the union of $\{C_{(j)}\}$. The aim is to reduce the candidate set, and to do that, let F be an interval extension of f and note that

$$\min_j\{L(C_{(j)})\} \;\le\; f^* \;\le\; \min_j\{U(C_{(j)})\} \;\equiv\; \tau \;, \tag{11.2}$$

where the functions L and U are the bounds of the interval returned by F,

$$\begin{aligned} L(C_{(j)}) \;&=\; \text{lower bound of } F(C_{(j)}) \;, \\ U(C_{(j)}) \;&=\; \text{upper bound of } F(C_{(j)}) \;. \end{aligned} \tag{11.3}$$

Therefore, if

$$L(C_{(j)}) > \tau \;, \tag{11.4}$$

then $C_{(j)}$ can be discarded from the candidate set. Otherwise, we can get sharper bounds by splitting $C_{(j)}$ into two, smaller subregions, cf. the algorithms in Chapters 7 – 9. If $n = 1$, then the splitting is done by simple bisection, and in multiple dimensions we bisect in the direction of the largest component of the radius of $C_{(j)}$.

Now we are ready to sketch the basic algorithm. We split the candidate set C into a work set S and a result set R, so that X^* is always contained in the union of S and R. If

$$w(F(S_{(j)})) \;\le\; \delta \;, \tag{11.5}$$

then the element $S_{(j)}$ is moved from S to R.

Algorithm *GlobalOptimization1* (11.6)

$S_{(1)} := \mathcal{D}$; $S := \{S_{(1)}\}$; $\tau := U(S_{(1)})$;
$R := \emptyset$
while $S \neq \emptyset$
  $i := \text{argmin}_j\{L(S_{(j)})\}$
  $X := S_{(i)}$; remove $S_{(i)}$ from S
  split X into $X_{(1)}$ and $X_{(2)}$
  $\hat{\tau} := \min\{U(X_{(1)}), U(X_{(2)})\}$
  if $\hat{\tau} < \tau$ **then**
    $\tau := \hat{\tau}$;
    use (11.4) to reduce S
  end
  for $k = 1, 2$
    if $L(X_{(k)}) \le \tau$ **then**
      if $w(F(X_{(k)})) \le \delta$ **then** $R := R \cup X_{(k)}$
      else $S := S \cup X_{(k)}$
    end
  end
end

Example 11.1. Consider the function $f(x) = (1-x^2)\cos(5x)$ on $\mathcal{D} = [\,0\,,\,2\,]$ with the interval extension $F(X) = (1 - X^2)\cos(5X)$.
In Figure 11.1 we give selected results from the iteration with Algorithm (11.6). k denotes the number of repeats.

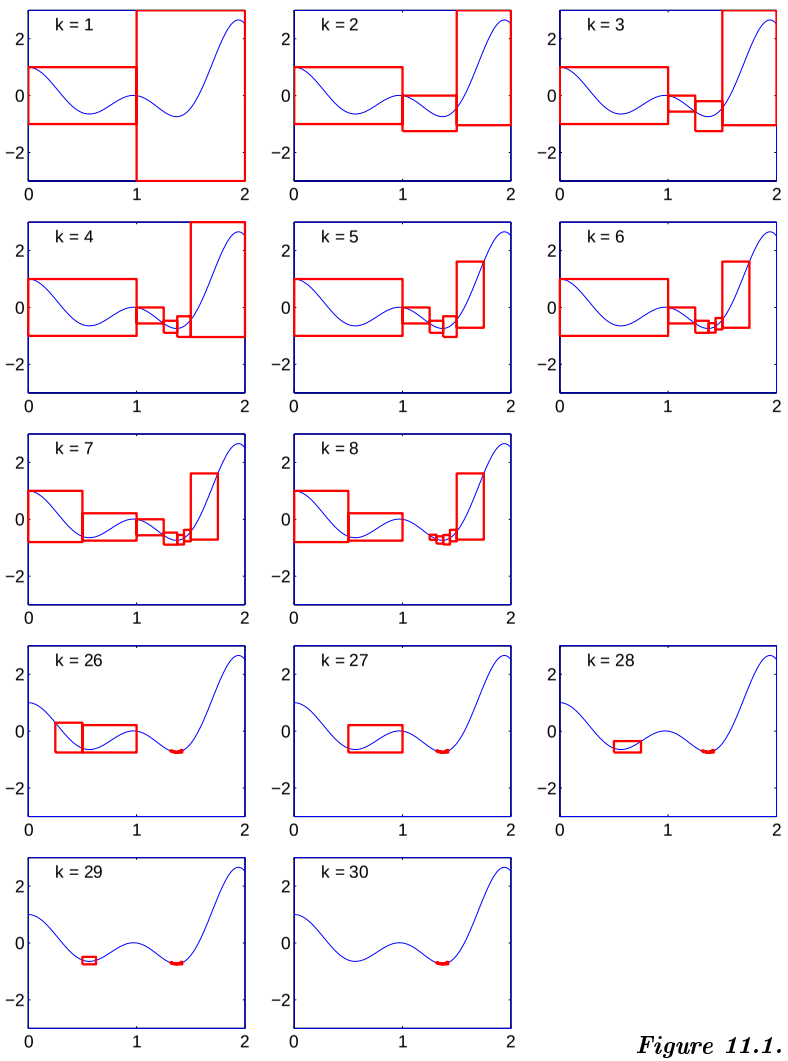


Figure 11.1.

The starting point is $S = \mathcal{D}$ with $F(\mathcal{D}) = [-3\,,\,3]$. For $k = 5$ the interval $[1.75\,,\,2]$ satisfies (11.4) and is discarded. For $k = 9, \ldots, 25$ computation is focused on the region around the true minimum, but we still keep the local minimum around 0.6 in the candidate set. Then the lower bound around the global minimum has become so large that the candidates around the local minimum are split and discarded.
With $\delta = 10^{-2}$ it takes 20 further iteration steps before the candidate set is empty. The result list holds 24 neighbouring intervals spanning the interval $X = [1.3457\,,\,1.3926]$, which contains X^*. The relation (11.2) takes the form $-0.7444 \le f^* \le -0.7352$. ∎

11.2. Use of Gradient Information

Suppose that we not only have access to the interval extension F, but also to and interval extension F' of the *gradient* $f' = \left(\dfrac{\partial f}{\partial x_1}, \ldots, \dfrac{\partial f}{\partial x_n}\right)$. This information can be used in two ways,

1. Monotonicity. If $0 \notin \big(F'(S_{(s)})\big)_j$, then f is monotone in the component x_j on the subdomain $S_{(s)}$. Therefore, we can reduce the jth component of the interval vector $S_{(s)}$ to its lower (respectively upper) bound if $\big(F'(S_{(s)})\big)_j > 0$ (respectively $\big(F'(S_{(s)})\big)_j < 0$). Furthermore, if this reduced candidate is interior to the original domain $\mathcal{D}$, then we can discard it from S because an interior minimum is a stationary point, i.e. the gradient is zero.

2. Stationary points. As just mentioned, the gradient at an interior minimizer is zero. This means that an interior minimizer is a root of the equation $f'(x) = 0$, and we can use one of the methods from Chapter 9 to locate this minimizer. Especially, if we also have an implementation of the Jacobian of f', i.e. the Hessian f'' (or use automatic differentiation to compute it), then we can use Krawczyk's version of Newton's method (9.11) to try and find the minimizer. There are three possible results for the interval $X = S_{(i)}$:

i) Krawczyk's method shows that X contains no root. If X is interior to $\mathcal{D}$, then we can discard $(S_{(i)})$ from the candidate set, otherwise we can reduce it to the union of its non-interior edges.

ii) X contains a root, and Krawczyk's method is used to find it with high accuracy. $S_{(i)}$ can be replaced by the resulting interval.

iii) Krawczyk's method stalls at X, maybe because the interval is too wide. Use simple splitting to reduce the width.

Now, we can outline the algorithm. The updating of the threshold τ and the sets S and R is performed as in Algorithm 11.6.

Algorithm *GlobalOptimization2* (11.7)

$S_{(1)} := \mathcal{D}$; $S := \{S_{(1)}\}$; $\tau := U(S_{(1)})$;
$R := \emptyset$
while $S \neq \emptyset$
  $i := \text{argmin}_j\{L(S_{(j)})\}$
  $X := S_{(i)}$; remove $S_{(i)}$ from S
  if Monotone(X) **then**
    X is reduced to R_X, see 1. above
  elseif Newton(X) works **then**
    X is reduced to R_X, see i) and ii) above
  else
    split: $R\_X := \{X_{(1)}, X_{(2)}\}$
      with $X = X_{(1)} \cup X_{(2)}$
  end
  use R_X to update τ and S
  with all elements in R_X
    discard or store in S or R
  end
end

Example 11.2. We have applied this algorithm to the same problem as in Example 11.1 and got the performance illustrated in Figure 11.2. A star signifies an element in the Result set R.
The first three iteration steps are identical with the results from Algorithm 11.6, and for $k=4$ the "best" subregion is $X = [1.25, 1.5]$, and Krawczyk's method gives us the global minimizer with associated threshold $\tau = -0.73979$, which also discards the candidate $[1, 1.25]$. The remaining iteration steps successively discard the remaining elements in the candidate set. The result list holds one element,

$$x^* \in X = [1.36864016023728, 1.36864016023729],$$

and the minimum is

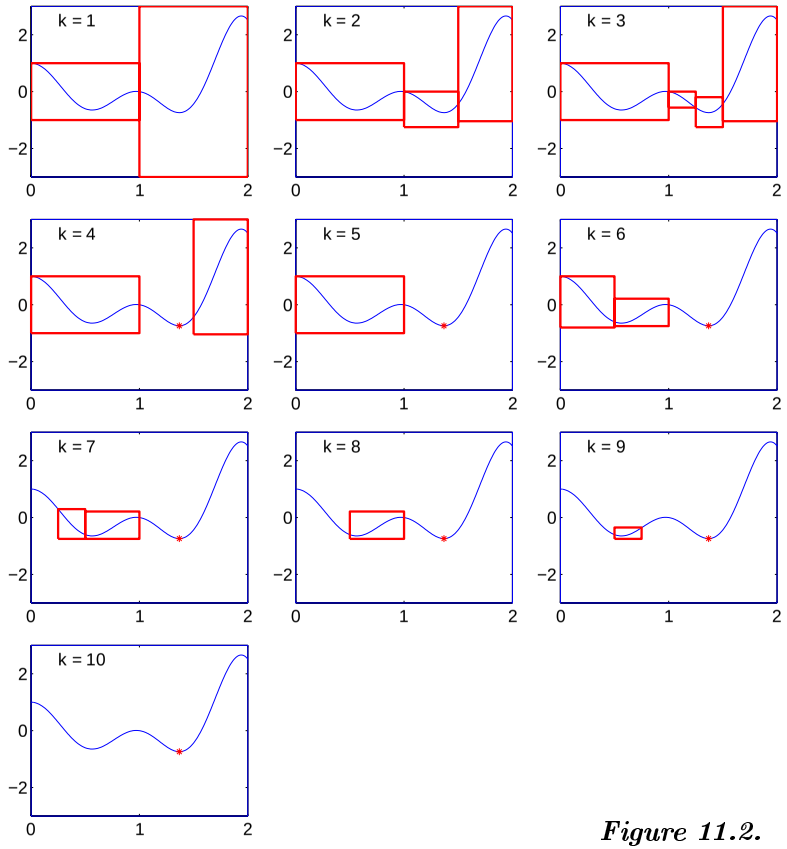


Figure 11.2.

$$f^* \in F(X) \quad \text{with} \quad m(F(X)) = -0.7397955,\ r(F(X)) = \texttt{1.55e-15}\ .$$

Thus, Algorithm 11.7 needs fewer iterations and gives a much more accurate result than Algorithm 11.6. ■

As many other problems global optimization suffers from the *"curse of dimensionality"*, i.e. the computational work grows fast with the dimension n. Therefore it is appropriate to use parallel calculation in order to decrease real-time. Madsen (1991) describes a parallel version of Algorithm (11.7).

REFERENCES

[1] H. Beeck (1972), *Über die Struktur und Abschätzungen der Lösungsmenge von linearen Gleichungssystem mit Intervalkoeffizienten*, Computing **10**, pp 231 – 244.

[2] G. Corliss, C. Faure, A. Griewank, L. Hascoët and U. Naumann (2002) (editors), *Automatic differentiation of algorithms*, Springer, New York.

[3] R. Krawczyk (1969), *Newton-Algorithmen zur Bestimmung von Nulstellen mit Fehlerschranken*, Computing **4**, pp 187 – 201.

[4] K. Madsen (1991), *Parallel algorithms for global optimization*, Report 91-07, IMM. DTU.

[5] K. Madsen and O. Toft (1994), *A parallel method for linear interval equations*, Interval Computations **3**, pp 81 – 105.

[6] R.E. Moore (1966), *Interval Analysis*. Prentice-Hall, Englewood Cliffs, N.J.

[7] R.E. Moore (1976), *On computing the range of values of a rational function of n variables over a bounded region*, Computing **16**, pp 1 – 15.

[8] R.E. Moore and S.T. Jones (1977), *Safe starting regions for iterative methods*, SIAM J.Numer. Anal. **14**, pp 1051 – 1065.

[9] R.E. Moore and H. Ratschek (1988), *Inclusion functions and global optimization II*, Math. Progr. **41**(3), A series pp 341 – 356.

[10] A. Neumaier (1990), *Interval methods for systems of equations*, Cambridge University Press.

[11] A. Neumaier (2001), *Introduction to Numerical Analysis*, Cambridge University Press, Cambridge.

[12] K. Nickel (1967), *Die Vollautomatische Berechnung einer Einfachen Nullstelle von $F(T) = 0$ einschliesslich einer Fehlerabschätzung*, Computing **2**, pp 232 – 245.

[13] K. Nickel (1977), *Interval-Analysis*. pp 193 – 226 in D. Jacobs (ed), Proceedings of the Conference on the State of the Art in Numerical Analysis held at the University of York, April 12th – 15th, 1976.

[14] W. Oettli and W. Prager (1964), *Compatibility of approximate solution of linear equations with given error bounds for coefficients and right-hand sides*, Numer. Math. **6**, pp 405 – 409.

[15] J.D. Pintér (1996), *Continuous global optimization: a brief review*, Optima **52**, pp 1 – 8. Available at http://plato.la.asu.edu/gom.html.

[16] F.N. Ris (1972), *Interval analysis and applications to linear algebra*, D.Phil. Thesis, Oxford.

[17] J. Rohn (1989), *Systems of linear interval equations*, Linear Algebra Appl. **126**, pp 39 – 78.

[18] S. Skelboe (1974), *Computation of rational interval functions*, BIT **14**, pp 87 – 95.

[19] S. Skelboe (1977), *True worst-case analysis of linear electrical circuits by interval arithmetic*, Report I T 11, Institute of Circuit Theory and Telecommunications, Technical University of Denmark, Lyngby.

[20] O. Toft (1992), *Sequential and parallel solution of linear interval equations*, Master's thesis, Institute for Numerical Analysis, Technical University of Denmark, pp 1 – 98.

[21] P. Wonqwises (1975), *Experimentelle Untersuchungen zur numerischen Auflösung von linearen Gleichungssystemen mit Fehlererfassung*. pp 316 – 325 in K. Nickel (ed), "Interval Mathematics", Lecture Notes in Computer Science 29, Springer Verlag.

NOTATION

$\underline{a}, \overline{a}$	Lower and upper limit of interval A	p. 6
$\lvert A \rvert$	Upper bound of interval A	(2.1), (2.2)
$\lVert A \rVert$	Norm of interval A	(2.3), (2.4)
$d(A, B)$	Distance between intervals A and B	(5.1), (5.10)
$\overline{f}(X)$	Interval hull of f on X	(4.1)
$f_M(X)$	Mean value form	(6.3)
$J(x, y)$	Jacobian	p. 52
$K(x, X)$	Krawczyk operator	(9.10)
$m(A)$	Midpoint of interval A	(2.1), (2.2)
M_{ij}	(i,j)th element in the interval matrix M	p. 7
$N(x, X)$	Newton operator	(9.1)
$r(A)$	Radius of interval A	(2.1), (2.2)
$w(A)$	Width of interval A	(2.1), (2.3), (2.4), p. 22
X_i	ith element in the interval vector X	p. 6
$X_{(s)}$	sth element in a sequence of intervals	pp. 33, 47, 73
$\mathcal{X}$	Set of solutions to $Ax = b$	(10.5)

INDEX