ON THE OPTIMALITY OF INCLUSION ALGORITHMS

Henryk Kołacz

Institute of Mathematics

Technical University of Poznań

Poznań, Poland

Abstract. In this paper a general concept of inclusion algorithm is introduced. Any inclusion algorithm provides a set that includes the solution of a given problem. Inclusion algorithms are studied with respect to the information used by them.

Some examples illustrate the presented concepts and results.

1. Introduction

In computational practice we must take into consideration that the rounding and propagated errors can give a large and inestimable error of the final result.

In general, it is difficult to provide a priori estimates of this error, and even in case they are available they produce bounds so pessimistic that they are of little practical importance.

Therefore there is a need for automatic error control in numerical computations.

A very useful tool for it is the interval analysis introduced by Moore [2]. The basic idea of this analysis is the inclusion of the solution of a given problem by intervals.

In this paper we introduce the concept of inclusion algorithm. It is defined as an arbitrary operator φ such that it provides a set including the solution of a given problem. We shall assume that there exists an arithmetic such that the computed values of φ are outer approximations of the exact values of φ .

We present a model of optimality for inclusion algorithms. It is based on the methodology introduced by Traub and Woźniakowski in [7].

The optimality of inclusion algorithms is studied with respect to error and computational complexity. It is shown that the intersection algorithm is a strongly optimal inclusion algorithm with respect to error. There are some connections between our optimality model and the ideas of Ratschek [6].

To illustrate concepts and results we present two examples: integration and range approximation.

2. Basic definitions

Let E,F be two given sets. By $\mathbb{P}(E)$ we denote the power set of E, that is, the class of all subsets of E.

Let $R_E \subset \mathbb{P}(E)$ be a fixed class of subsets of E. The family R_E is called a class of set representations in E.

For example R_E is the class of all closed balls in a pseudometric space E or the class of all closed intervals in an ordered space E.

We assume that there exists an operator $H\colon \mathbb{P}(E) \rightarrow R_E$ such that:

(2.1) $H(X) = X$ for all $X \in R_E$,

(2.2) $X \subset H(X)$ for all $X \in \mathbb{P}(E)$,

(2.3) $X \subset Y$ implies $H(X) \subset H(Y)$ for all $X,Y \in \mathbb{P}(E)$.

The operator H satisfying the properties (2.1)-(2.3) is called a monotone upwardly directed rounding (see [2]).

In our model we assume that the distance between elements of the family R_E is measured by elements of a complete lattice K. Then every subset of K has an infimum and a supremum.
Moreover, let $\inf K = \Theta$, that is, $m \geqslant \Theta$ for all $m \in K$.

Definition 2.1. We shall say that $d: R_E \times R_E \rightarrow K$ is a distance operator in the class R_E if

(2.4) $X,Y \subset Z \Longrightarrow d(X,Y) \leqslant d(X,Z),\ d(Y,Z)$

for all $X,Y,Z \in R_E$.

Let ε be a given element of K, $\varepsilon \geqslant \Theta$.

Definition 2.2. We shall say that $X \in R_E$ is an ε-inclusion of an element $x \in E$ if

1^o $x \in X$,

2^o $d(H(x),X) \leqslant \varepsilon$.

We illustrate the above concepts by an example.

Example 2.1. Let E be a normed linear space over the real or complex field. Let R_E be an arbitrary class of set representations in E such that it includes the class of all singletons in E.
We define the distance operator d in R_E as

$$d(X,Y) = \|X - Y\| ,$$

where $\|X\| = \sup[\|x\|: x \in X]$. The set $U(x,\varepsilon)$ defined as

$$U(x,\varepsilon) = \{X \in R_E: x \in X, \|x-X\| \leqslant \varepsilon\},$$

is the family of all ε-inclusions of an element $x \in E$, where ε is a fixed nonnegative real number.

3. Information operators

Let $S: F \rightarrow E$ be an arbitrary operator. We want for any $f \in F$ to find an $\mathcal{E}$-inclusion of $S(f)$. To find it, we must know something about the element f. Let

$$N: F \rightarrow \mathcal{H} \tag{3.1}$$

be an arbitrary operator, where $\mathcal{H}$ is a given space.
The operator N is called the basic information operator for F and the element N(f) is called the basic information of f.

Definition 3.1. Let $f \in F$ and $\mathcal{U}$ be a given set. We shall say that $L: f \rightarrow \mathcal{U}$ is an information operator for f (generated by N) if $N(f) \subset L(f)$.

We denote the family of all information operators for f, $f \in F$ by $\hat{I}_N(f)$. Obviously $\hat{I}_N(f)$ is nonempty for all $f \in F$ because $N \in \hat{I}_N(f)$. We illustrate the concept of information operator by the following example.

Example 3.1. Let M be a Banach space over the field of real numbers $\mathbb{R}$ and A be a nonempty subset of M.
Let $\mathcal{F}$ be a nonempty class of operators mapping A into M, which are n-times Frechet differentiable on A, where n denotes a fixed natural number. We take $F = \mathcal{F} \times R_A$ and $\mathcal{H} = R_M \times R_M \times \ldots \times R_M$ ((n+1)-times), where R_A and R_M denote fixed classes of set representations in A and M, respectively. Let $H: \mathbb{P}(M) \rightarrow R_M$ be a monotone upwardly directed rounding.

We define the basic information operator N in the following way:

$$N(g,X) = \left[H(\bar{g}(X)), H(\bar{g}'(X)), \ldots, H(\bar{g}^{(n)}(X)) \right],$$

where $\bar{g}^{(j)}(X)$ denotes the range of the jth Frechet derivative of $g \in \mathcal{F}$ over X. Then every information operator $L_g \in \hat{I}_N(g)$ has the following form:

$$L_g(X) = \left[G(X), G'(X), \ldots, G^{(n)}(X) \right],$$

where $\mathcal{H} = \mathcal{U}$ and $G^{(j)}$ is an extension of $g^{(j)}$ i.e. $\bar{g}^{(j)}(X) \subset$ $\subset G^{(j)}(X)$ for all $X \in R_A$ and $j = 0,1,2,\ldots,n$.

The inclusion between elements of the space $\mathcal{H}$ is meant componentwise.

It is often necessary to impose some restrictions on $L \in \hat{I}_N(f)$ in order to guarantee that the information $L(f)$ can be easily computed and enjoys some useful properties.
Let I_N be an operator defined on the set F such that $I_N(f)$ is a given family of information operators for $f \in F$, $I_N(f) \subset \hat{I}_N(f)$.
The operator I_N is called an information selection operator for F.
We denote

(3.2) $$I_N(F) = \{ N_f : \ N_f \in I_N(f), \quad f \in F . \}.$$

Example 3.2. Let $\mathcal{H}$ and $\mathcal{U}$ be given nonempty families of subsets of a space T. Let d be a distance operator in $\mathbb{P}(T)$ with values in $C = [0,+\infty)$ and be a fixed nonnegative real number. Then the operator I_N defined as

$$I_N(f) = \{ L \in \hat{I}_N(f) : \ d(N(f), L(f)) \leqslant \varepsilon \},$$

is an information selection operator for F.

For a given element $f \in F$ and an information operator $L \in I_N(f)$ we define the set $V(f,L)$ as follows:

(3.3) $$V(f,L) = \{ g \in F : \ \text{there exists } M \in I_N(g) \text{ such that } L(f)=M(g) \}.$$

Therefore $V(f,L)$ is the set of all elements $g \in F$ which have the same information as f under L. It is obvious that $V(f,L)$ is nonempty for every $f \in F$, $L \in I_N(f)$ because $f \in V(f,L)$.
Knowing $L(f)$, it is impossible to recognize which element $S(f)$ or $S(g)$ is being actually approximated for all $g \in V(f,L)$.
Analogously as in [7] we introduce the following definition.

Definition 3.2. We shall say din (I_N,f) is the local diameter of information if

(3.4) $$\mathrm{din}(I_N,f) = \sup_{L \in I_N(f)} \ \sup_{g_1,g_2 \in V(f,L)} d(H(S(g_1)),H(S(g_2))).$$

We shall say $\mathrm{din}(I_N)$ is the (global) diameter of information if

(3.5) $$\operatorname{din}(I_N) = \sup_{f\in F} \operatorname{din}(I_N,f).$$

4. Error of inclusion algorithms

To determine an ε-inclusion of $S(f)$ we use an inclusion algorithm which is an operator defined as follows.

Definition 4.1. We shall say that $\varphi : I_N(F) \rightarrow R_E$ is an inclusion algorithm for the problem S if

(4.1) $$S(f) \in \varphi(N_f)$$

for all $f \in F$ and $N_f \in I_N(f)$.

We denote the class of all inclusion algorithms using the information generated by the information selection operator I_N by $\hat{A}(I_N)$.

Let us observe that $\hat{A}(I_N)$ is an ordered set with the order relation $\leqslant$ defined as follows:

(4.2) $$\varphi_1 \leqslant \varphi_2 \Longleftrightarrow \varphi_1(N_f) \subset \varphi_2(N_f)$$

for all $f \in F$ and $N_f \in I_N(f)$, where $\varphi_1, \varphi_2 \in \hat{A}(I_N)$.

Definition 4.2. We shall say $e(\varphi,f)$ is the local error of $\varphi \in \hat{A}(I_N)$ if

(4.3) $$e(\varphi,f) = \sup_{N_f \in I_N(f)} \ \sup_{g \in V(f,N_f)} d(H(S(g)), \varphi(N_f)).$$

We shall say $e(\varphi)$ is the (global) error of φ if

(4.4) $$e(\varphi) = \sup_{f\in F} e(\varphi,f).$$

It is obvious that if $\varphi_1 \leqslant \varphi_2$ then $e(\varphi_1) \leqslant e(\varphi_2)$ for all inclusion algorithms $\varphi_1, \varphi_2 \in \hat{A}(I_N)$.

From the inclusion (4.1) it follows that the local diameter of information is a lower bound on the local error of any inclusion algorithm. A formal proof is provided by

<u>Theorem</u> 4.1. For any inclusion algorithm $\Phi \in \hat{A}(I_N)$,

$$e(\Phi, f) \geqslant \operatorname{din}(I_N, f) \tag{4.5}$$

for all $f \in F$. Moreover,

$$e(\Phi) \geqslant \operatorname{din}(I_N). \tag{4.6}$$

<u>Proof</u>. Let $f \in F$ and $N_f \in I_N(f)$. It is obvious that $S(g) \in \Phi(N_f)$ for all $g \in V(f, N_f)$. From this by the formula (2.4) we obtain the inequality (4.5).
The inequality (4.6) is a simple consequence of (4.5). The proof is complete.

<u>Example</u> 4.1. For $f \in F$, $N_f \in I_N(f)$ we define

$$U^*(N_f) = H(\{S(g): \ g \in V(f, N_f)\}). \tag{4.7}$$

It is obvious that U^* is an inclusion algorithm, $U^* \in \hat{A}(I_N)$.
From the inclusion (4.1) it follows that

$$U^*(N_f) \subset \Phi(N_f), \tag{4.8}$$

for all $f \in F$, $N_f \in I_N(f)$ and any inclusion algorithm Φ.
Moreover, taking $R_E := \mathbb{P}(E)$ and $d(X, Y) = \|X - Y\|$ we obtain

$$e(U^*, f) = \operatorname{din}(I_N, f) \tag{4.9}$$

for all $f \in F$. This means that the inequalities (4.5), (4.6) cannot be improved in general.

Let $A(I_N)$ be a nonempty class of inclusion algorithms using the information generated by I_N.

<u>Definition</u> 4.3. We shall say that $P \in A(I_N)$ is a strongly optimal error inclusion algorithm in the class $A(I_N)$ if

$$\inf \left[e(\Phi, f): \Phi \in A(I_N) \right] = e(P, f) \tag{4.10}$$

for all $f \in F$.
We shall say that $P \in A(I_N)$ is an optimal error inclusion algorithm in the class $A(I_N)$ if

$$\inf \left[e(\Phi): \Phi \in A(I_N) \right] = e(P). \tag{4.11}$$

Theorem 4.2. Let $A(I_N)$ be a nonempty family of inclusion algorithms such that $A(I_N) = \hat{A}(I_N) \cap W$, where W is a class of set operators. We define the operator φ^* as

$$\varphi^*(N_f) = H\Big(\bigcap_{\varphi \in A(I_N)} \varphi(N_f)\Big) . \tag{4.12}$$

Suppose $\varphi^* \in W$. Then φ^* is a strongly optimal error inclusion algorithm in $A(I_N)$.

Proof. First let us observe that φ^* is an inclusion algorithm. Therefore $\varphi^* \in A(I_N)$. Since $\varphi^* \leq \varphi$ for any inclusion algorithm

$\varphi \in A(I_N)$, $e(\varphi^*,f) \leqslant e(\varphi, f)$ for all $f \in F$.

From this we obtain that φ^* is a strongly optimal error inclusion algorithm in $A(I_N)$. The proof is complete.

Corollary 4.1. The algorithm U^* defined by the formula (4.7) is a strongly optimal error inclusion algorithm in the class $\hat{A}(I_N)$.

Proof. It is a simple consequence of the inclusion (4.8).

Remark 4.1. A strongly optimal error inclusion algorithm is also an optimal error inclusion algorithm but the converse is, in general, not true. Obviously U^* is an optimal error inclusion algorithm in $\hat{A}(I_N)$.

5. Complexity of inclusion algorithms

In this section we present a model of computation which consists of a set of primitive operations, permissible information operators, and permissible inclusion algorithms. This model is based on the general setting given in [7].

(i) Let t be a primitive operation in a given class of set representations R_E in E. Examples of primitive operations in $I(E)$ are interval operations (the addition of two intervals, the multiplication of an interval by a real number etc.). Usually primitive operations in R_E are defined by some corresponding operations in the space E (see [2]).

Let T be a given set of primitive operations in R_E. We denote the complexity (the total cost) of t by comp(t). We assume that comp(t) is finite.

(ii) Let $f \in F$ and $L \in I_N(f)$. We say that L is a permissible information operator for f with respect to T if there exists a program using a finite number of primitive operations from T which computes $L(f)$. We assume that if $L(f)$ requires the evaluation of operations $t_1, t_2, \ldots, t_k \in T$, then $\mathrm{comp}(L(f)) = \sum_{i=1}^{k} \mathrm{comp}(t_i)$.

(iii) Let $I_N(f)$ be a nonempty class of permissible information operators for f, $f \in F$. Let $\Phi \in \hat{A}(I_N)$. We say that Φ is a permissible inclusion algorithm with respect to T if for every $f \in F$ and $L \in I_N(f)$ there exists a program using a finite number of primitive operations from T which computes $Z \in R_E$ such that $Z \supset \Phi(Y)$, where $Y = L(f)$.

Let $\mathrm{comp}(\Phi(Y))$ be the complexity of computing $\Phi(Y)$. We assume that if $\Phi(Y)$ requires the evaluation of $s_1, s_2, \ldots, s_m \in T$, then $\mathrm{comp}(\Phi(Y)) = \sum_{i=1}^{m} \mathrm{comp}(s_i)$. We denote the class of all permissible inclusion algorithms with respect to T in $\hat{A}(I_N)$ by $\hat{A}_T(I_N)$.

We define the complexity of $\Phi \in \hat{A}_T(I_N)$ as

$$(5.1) \qquad \mathrm{comp}(\Phi) = \sup_{f \in F} \sup_{L \in I_N(f)} \left[\mathrm{comp}(L(f)) + \mathrm{comp}(\Phi(L(f)))\right].$$

Let $\varepsilon \geqslant \theta$ be a fixed element of a complete lattice K.

Let $A_T(I_N, \varepsilon)$ be a nonempty subset of $\hat{A}_T(I_N)$ such that $e(\Phi) \leq \varepsilon$ for all $\Phi \in A_T(I_N, \varepsilon)$.

<u>Definition</u> 5.1. We shall say that $P \in A_T(I_N, \varepsilon)$ is an ε-complexity optimal inclusion algorithm in the class $A_T(I_N, \varepsilon)$ if

$$(5.2) \qquad \inf\left[\mathrm{comp}(\Phi) : \Phi \in A_T(I_N, \varepsilon)\right] = \mathrm{comp}(P).$$

The analysis needed to characterize and construct an ε-complexity optimal algorithm for a particular problem can be a difficult mathematical problem.

6. Applications

In this section we show some examples of how the above analysis can be applied to some concrete problems.
We present two examples: integration and range approsimation.

(i) Integration

Let F be the class of all continuous real functions defined on the interval $[a,b] \subset \mathbb{R}$. We take $E = \mathbb{R}$ and $R_E = I(\mathbb{R})$, where $I(\mathbb{R})$ denotes the class of all closed intervals over $\mathbb{R}$.
We define the distance operator d in $I(\mathbb{R})$ as

$$(6.1) \qquad d(X,Y) = \sup\left[\,|x-y| : x \in X,\ y \in Y\,\right].$$

We define the operator $S: F \longrightarrow \mathbb{R}$ as

$$(6.2) \qquad S(g) = \int_a^b g(t)\,dt,$$

for $g \in F$.
Let M be a positive integer and subdivide $[a,b]$ into M subintervals $X_1, X_2, \ldots, X_M$, so that

$$(6.3) \qquad a = \underline{X}_1 < \overline{X}_1 = \underline{X}_2 < \overline{X}_2 < \ldots < \overline{X}_M = b,$$

where $X_i = [\underline{X}_i, \overline{X}_i]$ for $i = 1,2,\ldots,M$.

We define the basic information operator N as

$$(6.4) \qquad N(g) = \left[\bar{g}(X_1), \bar{g}(X_2), \ldots, \bar{g}(X_M)\right],$$

where $g \in F$. Then any information operator for g has the form:

$$(6.5) \qquad L(g) = \left[G(X_1), G(X_2), \ldots, G(X_M)\right],$$

where G is an interval extension of g. The inclusion between elements of $I^M(\mathbb{R})$ (the Cartesian product of $I(\mathbb{R})$, M-times) is meant componentwise. For $g \in F$ and $L_g \in I_N(g)$ we define the interval operator as follows (see [3]):

$$(6.6) \qquad \Phi(L_g) = \sum_{i=1}^{M} G(X_i)\,w(X_i),$$

where $w(X)$ denotes the width of an interval $X \in I(\mathbb{R})$.
Obviously by the mean value theorem Φ is an inclusion algorithm.
Let $Ex(g)$ be a nonempty family of interval extensions of $g \in F$.

Let $I_N(g)$ be the family of all information operators for g of the form (6.5) with $G \in Ex(g)$.
Then it is not difficult to verify that

$$(6.7) \qquad din(I_N,g) = \sup_{G \in Ex(g)} \sum_{i=1}^{M} w(G(X_i))w(X_i).$$

Moreover, let us observe that

$$(6.8) \qquad e(\varphi,g) \leqslant \sup_{G \in Ex(g)} \sum_{i=1}^{M} w(G(X_i))w(X_i).$$

From this by Theorem 4.1 we obtain that φ is a strongly optimal error inclusion algorithm.

(ii) Range approximation

Let U be the family of all real functions defined on an interval $D \subset \mathbb{R}$ and differentiable n-times on $D \in I(\mathbb{R})$.
We take $F = U \times I(D)$, $E = I(\mathbb{R})$ and $R_E = I(\mathbb{R})$.
We define the distance operator d in $I(\mathbb{R})$ by the formula (6.1).
For $X \in I(D)$ we define the power X^n of X by $X^n = \{x^n: \ x \in X\}$, where $n \geqslant 0$. We denote the absolute value of $X \in I(D)$ by $|X|$.
We define the operator $S: U \times I(D) \to I(\mathbb{R})$ as

$$(6.9) \qquad S(g,X) = \bar{g}(X),$$

where $\bar{g}(X)$ denotes the range of g over X.
Let N be the basic information operator defined as

$$(6.10) \qquad N(g,X) = \left[g(c),g'(c),\dots,g^{(n-1)}(c),\bar{g}^{(n)}(X)\right],$$

where $c = m(X)$ is the midpoint of X, $n \in \mathbb{N}$ and $g^{(j)}$ denotes the jth derivative of g. We define an information operator N_g for g as

$$(6.11) \qquad N_g(X) = \left[g(c),g'(c),\dots,g^{(n-1)}(c),G^{(n)}(X)\right],$$

where $G^{(n)}$ is an interval extension of $g^{(n)}$.
Let $Ex(g^{(n)})$ be a nonempty class of interval extensions of $g^{(n)}$.
Let $I_N(g)$ be the family of information operators for g of the form (6.11) with $G^{(n)} \in Ex(g^{(n)})$. We denote $w := w(D)$ and

$$(6.12) \qquad |E_n| := \sup_{G^{(n)} \in Ex(g^{(n)})} \left|G^{(n)}(D)\right|.$$

Developing functions with the same information as g in Taylor series around c we obtain

$$(6.13)\qquad \mathrm{din}(I_N,g) \leqslant \sum_{k=1}^{n-1} 2^{\lambda_k - k} \frac{1}{k!} g^{(k)}(c)w^k + 2^{\lambda_n - n} \frac{1}{n!} |E_n| w^n ,$$

where

$$(6.14)\qquad \lambda_k = \begin{cases} 0 & \text{if } k \text{ is even,} \\ 1 & \text{if } k \text{ is odd.} \end{cases}$$

For $g \in U$ and $X \in I(D)$ the Taylor form of g of order n, is defined by (see [4]):

$$(6.15)\qquad \Phi(N_g,X) = \sum_{k=0}^{n-1} \frac{1}{k!} g^{(k)}(c)(X-c)^k + \frac{1}{n!} G^{(n)}(X)(X-c)^n .$$

It is obvious that Φ is an inclusion algorithm for the problem S. It is not difficult to verify that

$$(6.16)\qquad e(\Phi,g) \leqslant \sum_{k=1}^{n-1} 2^{\lambda_k - k} \frac{1}{k!} g^{(k)}(c)w^k + 2^{\lambda_n - n} \frac{1}{n!} |E_n| w^n .$$

Now let U be the class of all polynomials of degree at most $n-1$ defined on the interval D. We take $Ex(g^{(n)}) = \{G^{(n)}\}$, where $G^{(n)}(X) = [0,0]$ for all $X \in I(D)$.
Suppose $g^{(k)}(c) \geqslant 0$ or $g^{(k)}(c) \leqslant 0$ for $k = 1,2,\ldots,n-1$.
We shall present our consideration for the first assumption. The considerations for the second assumption are analogous.

1° Let $g^{(k)}(c) = 0$ for $k=2,4,6,\ldots$. Then it is easily verified that

$$(6.17)\qquad \mathrm{din}(I_N,g) = \sum_{k=1,\,\mathrm{odd}}^{n-1} 2^{1-k} \frac{1}{k!} g^{(k)}(c)w^k .$$

In this case the Taylor form has the following form:

$$(6.18)\qquad \Phi(N_g,X) = g(c) + \sum_{k=1,\,\mathrm{odd}}^{n-1} \frac{1}{k!} g^{(k)}(c)[-z^k,z^k],$$

where $z = w(X)/2$. It is easy to show that

$$(6.19)\qquad e(\Phi,g) \leqslant \mathrm{din}(I_N,g).$$

Therefore by Theorem 4.1, Φ is a strongly optimal error inclusion algorithm.

2° Let $g^{(k)}(c) = 0$ for $k = 1,3,5,\ldots$. From this it follows that

$$\mathrm{din}(I_N, g) = \sum_{k=2,\,\mathrm{even}}^{n-1} 2^{-k} \frac{1}{k!} g^{(k)}(c) w^k . \tag{6.20}$$

We have in this case

$$\Phi(N_g, X) = g(c) + \sum_{k=2,\,\mathrm{even}}^{n-1} \frac{1}{k!} g^{(k)}(c) \left[0, z^k\right] . \tag{6.21}$$

It is easy to verify that in this case the inequality (6.19) holds, too. Therefore Φ is a strongly optimal error inclusion algorithm.

The problems connected with the range approximation by Taylor forms were considered in [3],[4],[5] (see also bibliography in [4]).

References

[1] Kulisch, U.W. and Miranker, W.L.: Computer arithmetic in theory and practice, Academic Press, New York, 1981.

[2] Moore, R.E.: Interval analysis, Printice-Hall, Englewood Cliffs, New York, 1966.

[3] Moore, R.E.: Methods and applications of interval analysis, SIAM, Philadelphia, 1979.

[4] Ratschek, H. and Rokne, J.: Computer methods for the range of functions, Ellis Horwood Limited, 1984.

[5] Ratschek, H.: Optimality of the centered form for polynomials, Journal of Approximation Theory, 32, pp. 151-159, 1981.

[6] Ratschek, H.: Optimal approximations in interval analysis, in: Interval Mathematics, ed. K. Nickel, Academic Press, pp. 181-202, 1980.

[7] Traub, J.F. and Woźniakowski, H.: A general theory of optimal algorithms, Academic Press, New York, 1980.