FILOMAT (Niš) 9:3 (1995), 543-577
Algebra, Logic & Discrete Mathematics
Niš, April 14–16, 1995.

INVITED LECTURE

ITERATIVE METHODS FOR BOUNDING THE INVERSE OF A MATRIX (A SURVEY)

Miodrag S. Petković

ABSTRACT. The aim of this paper is to give a survey of iterative methods for bounding the inverse of a point or interval matrix. These methods are based on the generalized Schulz's method and developed in interval arithmetic. The interest in bounding roundoff errors in matrix computations has come from the impossibility of exact representation of elements of matrices in those cases when numbers are represented in the computer by strings of bits of finite length or elements were experimentally determined by measurement which leads to the uncertainty in initial data. A posed problem can be usefully solved by interval analysis, a new powerful tool of applied mathematics. A detailed study of the basic inclusion method and its modifications, including the convergence features, conditions for a safe convergence, the monotonicity property, the choice of initial inclusion matrices and a number of remarks concerning a practical realization, were presented. A special attention is devoted to the construction of efficient methods for the inclusion of the inverse of a matrix.

1. Introduction

The demands of the computer age at the beginning of the sixties years with its "finite" arithmetic dictate the need for a structure which has come to be called *interval analysis* or later *interval mathematics* - a new, growing, and fruitful branch of applied mathematics. *"Although interval analysis is in a sense just a new language for inequalities, it is very powerful language and is one that has direct applicability to the important problem of significance in large computations"* (R.D. Richtmeyer, Math. Comput. 22 (1968), p. 221). The starting point for the application of interval analysis, described for the first time by Moore [21], is the desire in numerical mathematics to be able to implement algorithms on digital computers capturing all the roundoff errors

This work is supported by the Science Fund of Serbia under Grant No. 0401

automatically and therefore to calculate strict errors automatically. Interval arithmetic is powerful tool for bounding a result of some computation or a solution of an equation so that interval methods are often called *self-validiting algorithms.*

Anyone using a computer, whether in engineering design, physical sciences, technical disciplines, or whatever has surely inquired about the effect of rounding error and propagated error due to uncertain initial data or uncertain values of parameters in mathematical models. A standard question should be *"what is the error in the obtained results?"*. Numerical algorithms using interval arithmetic supply techniques for keeping track of errors and provide the machine computation of rigorous error bounds on approximate solutions or results.

The application of interval mathematics to computing has several objectives: to provide computer algorithms for finding sets containing unknown solutions; to make these sets as small as possible; and to do all this as efficiently as possible. Towards these objectives, set-to-set mappings replace point-to-point mappings, and set inclusions replace approximate equalities.

The purpose of this paper is to present iterative methods for bounding the inverse of a matrix. The interest in bounding roundoff errors in matrix computations has come from the impossibility of exact representation of elements of matrices in some cases since numbers are represented in the computer by strings of bits of fixed, *finite* length. Besides, there are elements which are experimentally determined by measurement which leads to the *uncertainty* in initial data and it is only known that their values belong to some intervals. Finally, nearly all numerical computation is carried out with "fixed-precision", approximate arithmetic. In the commonly used approach, one assumes that the worst possible roundoff error occurs in each numerical step. One then determines (or bounds) how these errors can accumulate as the computation proceeds. This procedure is usually called *ordinary method for error bounding* and the abbreviation $\mathcal{OM}$ is used to refer to it. The second approach uses interval arithmetic (abbreviated as $\mathcal{IA}$) which has the advantage of an automatic control of rounding errors and, at the same time, an inclusion of the exact result of computation. For this reason, the main subject of this paper is concerned with iterative methods which use $\mathcal{IA}$ for bounding errors in matrix inversion.

In Section 2 we will give the basic matrix operations needed for the construction and analysis of iterative algorithms for the inclusion of real or interval matrices. A general approach to the problem of the inversion of matrices is described in Section 3. The two basic interval iterative methods, based on the generalized Schulz's method, are considered in Section 4. Conditions for the monotonicity of interval sequence of inclusion matrices

are the subject of Section 5. In Section 6 we study the problem of finding a suitable initial matrix which insures the convergence of the presented interval algorithms. Efficient iterative methods for bounding the inverse of a matrix, which combine the efficiency of floating-point arithmetic and the control of accuracy of results by interval arithmetic, are presented in Section 7. A special attention is devoted to the choice of parameters which define the most efficient inclusion algorithm. Finally, in Section 8, we describe an iterative method for the inclusion of an *interval* matrix. Throughout this paper several numerical examples are given to illustrate presented methods as well as difficulties which appear in solving the studied problem.

The presented study is a two-way bridge between linear algebra and computing. Its aim is to encourage mathematicians to look further to computing as a source of challenging new problems, and researchers in computing to turn more frequently to contemporary mathematics in their day-to-day use of the digital machine.

2. Interval matrix operations

A subset of the set of real numbers $\mathbb{R}$ of the form

$$A = [a_1, a_2] = \{x | \, a_1 \leq x \leq a_2, \; a_1, a_2 \in \mathbb{R}\}$$

is called a closed *real interval*. The set of all closed real intervals will be denoted by $I(\mathbb{R})$. If $a_2 = a_1$ then the interval $A = [a_1, a_2]$ degenerates to the real number a_1 and A is called a *point interval*. The basic operations and properties in the set $I(\mathbb{R})$ are described in the book [3, Ch. 1 and 2]. Real intervals will be denoted by capital letters.

A *real interval matrix* is a matrix whose elements are real intervals. Since we deal in this paper only with real intervals and real interval matrices, we will use the shorter terms *interval* and *interval matrix*. The set of $m \times n$ matrices over the real numbers is denoted by $M_{mn}(\mathbb{R})$ and the set of $m \times n$ matrices over the real intervals by $M_{mn}(I(\mathbb{R}))$. An interval matrix whose all components are point intervals is called a *point matrix*. Point matrices (elements from $M_{mn}(\mathbb{R})$) will be denoted by capital letters $A, B, C, \ldots$, while interval matrices (elements from $M_{mn}(I(\mathbb{R}))$) by capital letters $\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots$ in **bold**. Interval matrices are represented, as is customary for real or complex matrices, by their components in the form $\mathbf{A} = (A_{ij})$.

Definition 1. Two $m \times n$ interval matrices $\mathbf{A} = (A_{ij})$ and $\mathbf{B} = (B_{ij})$ are equal if and only if there is equality between all corresponding components of the matrices, that is, $\mathbf{A} = \mathbf{B} \Leftrightarrow A_{ij} = B_{ij}$ $(i = 1, \ldots, m; \; j = 1, \ldots, n)$.

A partial ordering on the set of interval matrices $M_{mn}(I(\mathbb{R}))$ is introduced by

Definition 2. Let $\mathbf{A} = (A_{ij})$ and $\mathbf{B} = (B_{ij})$ be two $m \times n$ interval matrices. Then

$$\mathbf{A} \subseteq \mathbf{B} \Leftrightarrow A_{ij} \subseteq B_{ij} \quad (i = 1, \dots, m,\ j = 1, \dots, n).$$

In particular, if $A = (a_{ij})$ is a point matrix, then we write $A \in \mathbf{B}$. Each interval matrix may be regarded as a set of point matrices.

In the following we give a short review of the basic operations between interval matrices which formally correspond to the operations on point matrices.

Definition 3. For two $m \times n$ matrices $\mathbf{A} = (A_{ij})$ and $\mathbf{B} = (B_{ij})$ interval matrix addition and subtraction are defined by

$$\mathbf{A} \pm \mathbf{B} := (A_{ij} \pm B_{ij}).$$

Definition 4. Let $\mathbf{A} \in M_{mr}(I(\mathbb{R}))$ and $\mathbf{B} \in M_{rn}(I(\mathbb{R}))$. An interval matrix computation is defined by

$$\mathbf{AB} := \left(\sum_{k=1}^{r} A_{ik} B_{kj} \right).$$

Definition 5. If $\mathbf{A} = (A_{ij})$ is an interval matrix and X an interval, then

$$X\mathbf{A} = \mathbf{A}X := (XA_{ij}).$$

It is easy to prove that

$$\mathbf{A} \pm \mathbf{B} = \{A \pm B |\ A \in \mathbf{A}, B \in \mathbf{B}\},$$

while

$$\mathbf{AB} \supseteq \{AB |\ A \in \mathbf{A}, B \in \mathbf{B}\}.$$

In the following theorem the basic properties of the introduced operations are given (see [3, Ch. 10]):

Theorem 1. *If* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *are interval matrices, then*

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \quad \textit{(commutativity)},$$
$$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C} \quad \textit{(associativity)},$$
$$\mathbf{A} + O = O + \mathbf{A} = \mathbf{A} \quad (O - \textit{zero matrix}),$$
$$\mathbf{A}I = I\mathbf{A} = \mathbf{A} \quad (I - \textit{unit matrix}),$$
$$\begin{aligned} &(\mathbf{A} + \mathbf{B})\mathbf{C} \subseteq \mathbf{AC} + \mathbf{BC} \\ &\mathbf{C}(\mathbf{A} + \mathbf{B}) \subseteq \mathbf{CA} + \mathbf{CB} \end{aligned} \quad \textit{(subdistributivity)},$$
$$(\mathbf{A} + \mathbf{B})C = \mathbf{A}C + \mathbf{B}C,$$
$$C(\mathbf{A} + \mathbf{B}) = C\mathbf{A} + C\mathbf{B},$$
$$\mathbf{A}(BC) \subseteq (\mathbf{A}B)C.$$

Let us note that the associative low is not, in general, valid for interval matrices. This low is not valid even if two of three matrices are point matrices (the last property in the above theorem).

The inclusion isotonicity property for the matrix operations is given in the following theorem ([3, Ch. 10]):

Theorem 2. *Let* $\mathbf{A}_k, \mathbf{B}_k$ $(k = 1, 2)$ *be interval matrices and* X *and* Y *real intervals. If* $* \in \{+, -, \cdot\}$ *is one of matrix operations then the conditions*

$$\mathbf{A}_k \subseteq \mathbf{B}_k \ (k = 1, 2) \quad \textit{and} \quad X \subseteq Y$$

imply $\mathbf{A}_1 * \mathbf{A}_2 \subseteq \mathbf{B}_1 * \mathbf{B}_2$ *and* $X\mathbf{A}_k \subseteq Y\mathbf{B}_k$.

In particular, from Theorem 2 we obtain

$$A \in \mathbf{A},\ B \in \mathbf{B} \ \Rightarrow\ A + B \in \mathbf{A} + \mathbf{B},$$
$$\lambda \in X,\ A \in \mathbf{A} \ \Rightarrow\ \lambda A \in X\mathbf{A} \quad (\lambda \in \mathbb{R}).$$

Definition 6. Matrix norm of an interval matrix $\mathbf{A}$ is defined by

$$\| \mathbf{A} \| := \max_{A \in \mathbf{A}} \| A \|,$$

where $\| \cdot \|$ is an arbitrary monotone norm.

Thus, the norm of an interval matrix is an extension of the norm of a point matrix and directly depends on the type of this norm. Most frequently, we use "maximum row-sum" norm $\| \cdot \|_\infty$,

$$\| \mathbf{A} \|_\infty := \max_{A \in \mathbf{A}} \| A \|_\infty = \max_i \sum_j |A_{ij}| \tag{2.1}$$

and "maximum column-sum" norm $\| \cdot \|_1$,

$$\| \mathbf{A} \|_1 := \max_{A \in \mathbf{A}} \| A \|_1 = \max_j \sum_i |A_{ij}|. \tag{2.2}$$

Both norms are monotonic and multiplicative, that is (omitting subscript indices),

$$\mathbf{B} \subseteq \mathbf{A} \ \Rightarrow \| \mathbf{B} \| \leq \| \mathbf{A} \| \quad \text{and} \quad \| \mathbf{AB} \| \leq \| \mathbf{A} \| \cdot \| \mathbf{B} \| .$$

In the sequel, we will omit the subscript indices (indicating the type of norm) and assume that the used matrix norm is monotonic and multiplicative. The application of some specific norm will be accented.

Before introducing the concept of width, absolute value and midpoint for interval matrices, we recall to the corresponding definitions for a given real interval $X = [a, b]$:

$$d(X) = b - a \quad (\textit{width});$$
$$|X| = \max(|a|, |b|) \quad (\textit{absolute value});$$
$$m(X) = \frac{a+b}{2} \quad (\textit{midpoint}).$$

Definition 7. For an interval matrix $\mathbf{A} = (A_{ij})$ the following point matrices are associated:

a) the width matrix $d(\mathbf{A}) := (d(A_{ij}))$;
b) the absolute value matrix $|\mathbf{A}| := (|A_{ij}|)$;
c) the midpoint matrix $m(\mathbf{A}) := (m(A_{ij}))$.

The matrix $d(\mathbf{A})$ and $|\mathbf{A}|$ have nonnegative components. The elements of the midpoint matrix $m(\mathbf{A})$ are real numbers which are equal to the midpoints of the corresponding (interval) components of the interval matrix $\mathbf{A}$, so that $m(\mathbf{A}) \in \mathbf{A}$.

Definition 8. A sequence of interval matrices $\{\mathbf{A}_k\}$ is monotonically nonincreasing if $\mathbf{A}_0 \supseteq \mathbf{A}_1 \supseteq \mathbf{A}_2 \supseteq \cdots$, and monotonically nondecreasing if $\mathbf{A}_0 \subseteq \mathbf{A}_1 \subseteq \mathbf{A}_2 \subseteq \cdots$.

Definition 9. The intersection of two interval matrices $\mathbf{A} = (A_{ij})$ and $\mathbf{B} = (B_{ij})$ of the same type is defined as

$$\mathbf{A} \cap \mathbf{B} := (A_{ij} \cap B_{ij}).$$

It is easy to see that the intersection of interval matrices has the property

$$\mathbf{A} \subseteq \mathbf{C},\ \mathbf{B} \subseteq \mathbf{D} \Rightarrow \mathbf{A} \cap \mathbf{B} \subseteq \mathbf{C} \cap \mathbf{D} \quad (\textit{inclusion isotonicity}).$$

Definition 10. Let $X = (x_{ij})$ and $Y = (y_{ij})$ be point matrices from $M_{mn}(\mathbb{R})$. Then

$$X \leq Y \Leftrightarrow x_{ij} \leq y_{ij}\ (i = 1, \ldots, m;\ j = 1, \ldots, n)$$

defines the relation of partial ordering "$\leq$" in $M_{mn}(\mathbb{R})$.

Using Definition 10 the following properties for real matrices, introduced in Definition 7, can be proved ([3, Ch. 10]):

Theorem 3. *If* $\mathbf{A} = (A_{ij})$ *and* $\mathbf{B} = (B_{ij})$ *are interval matrices of the same type, then*

$$
\begin{aligned}
&(1)\ \mathbf{A} \subseteq \mathbf{B} \Rightarrow d(\mathbf{A}) \leq d(\mathbf{B}),\\
&(2)\ \mathbf{A} \subseteq \mathbf{B} \Rightarrow |\mathbf{A}| \leq |\mathbf{B}|,\\
&(3)\ d(\mathbf{A} \pm \mathbf{B}) = d(\mathbf{A}) + d(\mathbf{B}),\\
&(4)\ |\mathbf{A} + \mathbf{B}| \leq |\mathbf{A}| + |\mathbf{B}|,\\
&(5)\ |\lambda\mathbf{A}| = |\mathbf{A}\lambda| = |\lambda||\mathbf{A}| \quad (\lambda \in \mathbb{R}),\\
&(6)\ |\mathbf{AB}| \leq |\mathbf{A}||\mathbf{B}|,\\
&(7)\ d(\mathbf{AB}) \leq d(\mathbf{A})|\mathbf{B}| + |\mathbf{A}|d(\mathbf{B}),\\
&(8)\ d(\mathbf{AB}) \geq |\mathbf{A}|d(\mathbf{B}), \quad d(\mathbf{AB}) \geq d(\mathbf{A})|\mathbf{B}|,\\
&(9)\ d(A\mathbf{B}) = |A|d(\mathbf{B}), \quad d(\mathbf{B}A) = d(\mathbf{B})|A|,\\
&(10)\ O \in \mathbf{A} \Rightarrow |\mathbf{A}| \leq d(\mathbf{A}) \leq 2|\mathbf{A}|,\\
&(11)\ m(\mathbf{A} \pm \mathbf{B}) = m(\mathbf{A}) \pm m(\mathbf{B}),\\
&(12)\ m(C\mathbf{A}) = Cm(\mathbf{A}), \quad m(\mathbf{A}C) = m(\mathbf{A})C.\\
&(13)\ m(C) = C.
\end{aligned}
$$

3. Problems of bounding the inverse of a matrix

In this section we will consider the problem of bounding the inverse of a matrix in the presence of rounding errors applying digital computers with the arithmetic of limited precision as well as uncertain data in elements of a given matrix.

First, we point out some more general problems in matrix inversion. Let B be an exact matrix whose elements can be exactly represented in arithmetic of finite (say, double) precision in a computer. Let $A = (a_{ij})$ be a matrix whose elements are subject to error. Suppose we know only that a_{ij} $(i, j = 1, 2, \ldots, n)$ lies in the real interval $[\underline{a}_{ij}, \overline{a}_{ij}]$, where $\underline{a}_{ij}$ and $\overline{a}_{ij}$ can be exactly represented in double precision.

Problem 1. Compute B^{-1} (approximately) and bound the errors resulting from roundoff.

Problem 2. For a given matrix $A = (a_{ij})$ with $a_{ij} \in [\underline{a}_{ij}, \overline{a}_{ij}]$ for $i, j = 1, 2, \ldots, n$, compute A^{-1} (approximately) and bound both the errors resulting from roundoff and the errors from possible errors in A itself.

Problem 3. Define the set

$$(A^I)^{-1} = \{A^{-1} \mid a_{ij} \in [\underline{a}_{ij}, \overline{a}_{ij}],\ A^{-1}A = I\}.$$

Compute $(A^I)^{-1}$ approximately and bound the errors due to roundoff.

Problem 4. Find $(A^I)^{-1}$ exactly.

Problem 1 can be easily solved by $\mathcal{OM}$ or by inverting B using $\mathcal{IA}$. Moreover, by use of arithmetic of sufficiently high precision, arbitrary accuracy with arbitrary sharp bounds can be obtained.

Using $\mathcal{IA}$, *Problem 2* can be solved as easily as *Problem 1*. Using $\mathcal{OM}$, only slightly more effort is required to solve *Problem 2* than *Problem 1*.

$\mathcal{OM}$ obviously cannot solve *Problem 4* and cannot solve *Problem 3* except in a very crude sense. It can be shown (see [6]) that *Problem 4* cannot be solved using $\mathcal{IA}$, even if infinite precision arithmetic is used. An approximate solution of arbitrary high accuracy can be obtained but the amount of work quickly becomes prohibitive.

Hence, we direct our attention to *Problem 3* which can be solved using $\mathcal{IA}$. Two approaches for solving this problem by $\mathcal{IA}$ have been developed in the literature: hyperpower method [4], [6] and Alefeld-Herzberger's modification of generalized Schulz' method [1].

The hyperpower method is defined by a matrix-valued fuction $\Phi(A, X)$ for real $n \times n$ matrix X in the range of the real $n \times n$ matrices, where A is a given matrix whose inverse A^{-1} has to be found. By means of the iteration

$$X^{(k+1)} = \Phi(A, X^{(k)}), \quad \text{given } X^{(0)},\ k \geq 0,$$

we get an iterative method which generates a sequence $\{X^{(k)}\}$ of matrices. Following Altman [4] we call this iterative method a *hyperpower method* for A^{-1} of order $p > 1$ if and only if the equation

$$I - AX^{(k+1)} = (I - AX^{(k)})^p, \quad k \geq 0$$

is filfilled. If the initial matrix $X^{(0)}$ is chosen so that $\rho(I - AX^{(0)}) < 1$ (ρ denotes the spectral radius), then the sequence of matrices $\{X^{(k)}\}$ converges to the inverse A^{-1} of the matrix A with the order of convergence p. Using suitable error-bounds for the hyperpower method it is possible to derive inclusion set for A^{-1}. Further improvements can be attained using interval Schulz-Herzberger's method in the final step, as it was proposed in [16] and [17]. Let us note that Herzberger presented in [10] a class of iterative methods for inverting a linear bounded operator in a Banach space, which can be considered as a kind of hyperpower method.

The second method uses an iterative procedure to bound the inverse not only for a point matrix but also for an interval matrix. As mentioned above, this method was introduced by Alefeld and Herzberger [1] and analysed later in their books [2] and [3]. It is based on the generalized Schulz's method for point matrices and realized in real interval arithmetic. Since a number of outstanding results concerning improvements and modifications of this methods, including detailed studies of many convergence properties and behaviours, and a practical realization, were given by Prof. J. Herzberger throughout about twenty papers, it is quite natural that the mentioned methods and their modifications are referred to as Schulz-Herzberger's methods, or the S-H methods, for brevity. A survey of these interval methods will be given in the following sections.

Before we present iterative methods of Schulz-Herzberger's type, we give an example to illustrate difficulties appearing in bounding inverse matrices.

Example 1. Let us consider the interval matrix

$$\mathbf{A} = \begin{bmatrix} 1 & [0.999995, 1.000005] \\ 2 & 1 \end{bmatrix}$$

and the point matrix

$$C(x) = \begin{bmatrix} 1 & x \\ 2 & 1 \end{bmatrix}, \quad x \in X = [0.999995, 1.000005].$$

Let $\mathbf{A}^{-1} = (A'_{ij})$ and $C(x)^{-1} = (c'_{ij}(x))$ be the inverse matrices of $\mathbf{A}$ and $C(x)$, respectively. Then $A'_{ij} = \{c'_{ij}(x) | \, x \in X\}$. Let us determine, for instance, the component A'_{12} of the inverse matrix $\mathbf{A}^{-1}$. First, we have $c'_{12} = x/(2x-1)$. For $x \in X = [0.999995, 1.000005]$ the component $c'_{12}(x)$ is a monotone function so that the endpoints of the interval X yield the extreme values (minimum and maximum) of $c'_{12}(x)$. According to this, using 10 significant digits, we obtain $A'_{12} = [0.9999950000, 1.0000050000]$.

On the other hand, using interval arithmetic of infinite precision and the rounding of results to 10 digits to find A'_{12}, we calculate

$$\frac{X}{2X-1} = [0.9999850001, 1.0000150001],$$

which differs from the exact result given above by A'_{12}.

4. Interval versions of Schulz's method

Let $p > 1$ be a fixed natural number and I the unit matrix. If A is a given nonsingular point matrix and $X^{(0)}$ is an initial matrix such that

$\| I - AX^{(0)} \| < 1$, then for finding the inverse of A the generalized iterative method of Schulz of the order p

$$X^{(k+1)} = X^{(k)} \sum_{r=0}^{p-1} (I - AX^{(k)})^r \quad (k = 0, 1, \dots) \tag{4.1}$$

can be applied (see [4], [26], [27], [32]). In particular, for $p = 2$, one obtains Schulz's method of the second order for calculating the inverse matrix [31]

$$X^{(k+1)} = X^{(k)}(2I - AX^{(k)}) \quad (k = 0, 1, \dots) . \tag{4.2}$$

Let $\mathbf{X}$ be an interval matrix containing the inverse matrix A^{-1} of a given nonsingular matrix A, and let $X \in \mathbf{X}$ (for example, $X = m(\mathbf{X})$). For $B = I - AX$ we have the identity

$$I^{p-1} - B^{p-1} = (I - B)(I + B + B^2 + \cdots + B^{p-2}) = AX \sum_{r=0}^{p-2} B^r,$$

that is, after multiplying by A^{-1},

$$A^{-1} - A^{-1}(I - AX)^{p-1} = X \sum_{r=0}^{p-2} (I - AX)^r.$$

Hence, since $A^{-1} \in \mathbf{X}$,

$$\begin{aligned} A^{-1} &= X \sum_{r=0}^{p-2} (I - AX)^r + A^{-1}(I - AX)^{p-1} \\ &\in X \sum_{r=0}^{p-2} (I - AX)^r + \mathbf{X}(I - AX)^{p-1}. \end{aligned} \tag{4.3}$$

The last relation suggests the following iterative interval version of (4.1) for the inclusion of the matrix A :

$$\mathbf{X}^{(k+1)} = m(\mathbf{X}^{(k)}) \sum_{r=0}^{p-2} \left(I - Am(\mathbf{X}^{(k)})\right)^r + \mathbf{X}^{(k)} \left(I - Am(\mathbf{X}^{(k)})\right)^{p-1}, \tag{4.4}$$

$(k = 0, 1, \dots)$, assuming that the initial matrix $\mathbf{X}^{(0)}$ contains A^{-1}.

The properties of the inclusion iterative method (4.4) are given in the following theorem ([3, Ch. 18]):

Theorem 4. *Let A be a nonsingular $n \times n$ matrix and $\mathbf{X}^{(0)}$ an $n \times n$ interval matrix such that $A^{-1} \in \mathbf{X}^{(0)}$. A sequence $\{\mathbf{X}^{(k)}\}$ of interval matrices is calculated according to* (4.4). *Then*

(4a) *each matrix $\mathbf{X}^{(k)}$ $(k \geq 0)$ contains A^{-1};*

(4b) *the sequence $\{\mathbf{X}^{(k)}\}$ converges to A^{-1} if and only if the spectral radius $\rho(I - Am(X^{(0)}))$ is smaller than 1;*

(4c) *using a matrix norm $\| \cdot \|$ the sequence $\{d(\mathbf{X}^{(k)})\}$ satisfies*

$$\| d(\mathbf{X}^{(k+1)}) \| \leq \gamma \| d(\mathbf{X}^{(k)}) \|^p, \quad \gamma \geq 0,$$

that is, the order of convergence of the method (4.4) *is at least p.*

Proof. Of (4a): Setting $\mathbf{X}^{(k)} = \mathbf{X}$ and $m(\mathbf{X}^{(k)}) = X$ in (4.3) and taking into account the iterative formula (4.4), we obtain

$$A^{-1} \in m(\mathbf{X}^{(k)}) \sum_{r=0}^{p-2} \left(I - Am(\mathbf{X}^{(k)})\right)^r + \mathbf{X}^{(k)} \left(I - Am(\mathbf{X}^{(k)})\right)^{p-1} = \mathbf{X}^{(k+1)}.$$

Since, in addition, $A^{-1} \in \mathbf{X}^{(0)}$, the proof of (4a) follows by complete induction.

Of (4b): Using the rules from Theorem 3 for the midpoint matrices, the midpoint mapping in the iterative procedure (4.4) gives the following iterative formula for the sequence $\{m(\mathbf{X}^{(k)})\}$:

$$m(\mathbf{X}^{(k+1)}) = m(\mathbf{X}^{(k)}) \sum_{r=0}^{p-1} \left(I - Am(\mathbf{X}^{(k)})\right)^r.$$

This is a generalization of Schulz's iterative procedure given also by (4.1). Multiplying both sides of this equation by A one obtains

$$\begin{aligned} Am(\mathbf{X}^{(k+1)}) &= \left(I - \left(I - Am(\mathbf{X}^{(k)})\right)\right) \sum_{r=0}^{p-1} \left(I - Am(\mathbf{X}^{(k)})\right)^r \\ &= I - \left(I - Am(\mathbf{X}^{(k)})\right)^p, \end{aligned}$$

or

$$I - Am(\mathbf{X}^{(k+1)}) = \left(I - Am(\mathbf{X}^{(k)})\right)^p = \left(I - Am(\mathbf{X}^{(0)})\right)^{p^{k+1}}.$$

Hence, there follows

$$\lim_{k \to \infty} m(\mathbf{X}^{(k)}) = A^{-1} \Leftrightarrow \lim_{k \to \infty} \left(I - Am(\mathbf{X}^{(0)})\right)^p = O \Leftrightarrow \rho\left(I - Am(\mathbf{X}^{(0)})\right) < 1.$$

Let us show that the sequence $\{\mathbf{X}^{(k)}\}$ converges to A^{-1} if and only if the sequence of midpoint matrices $\{m(\mathbf{X}^{(k)})\}$ converges to A^{-1}. This follows from the consideration of the sequence $\{d(\mathbf{X}^{(k)})\}$ of the width matrices which satisfy

$$d(\mathbf{X}^{(k+1)}) = d(\mathbf{X}^{(k)})|(I - Am(\mathbf{X}^{(k)}))^{p-1}|$$

(see the properties (3) and (9) of Theorem 3). If $\lim_{k\to\infty} m(\mathbf{X}^{(k)}) = A^{-1}$, then the last relation implies that $\lim_{k\to\infty} d(\mathbf{X}^{(k)}) = O$. Conversely, using the continuity of m and (13) of Theorem 3 it follows trivially that $\lim_{k\to\infty} \mathbf{X}^{(k)} = A^{-1}$ implies $\lim_{k\to\infty} m(\mathbf{X}^{(k)}) = A^{-1}$. Since it was shown above that the condition $\rho(I - Am(\mathbf{X}^{(0)})) < 1$ was necessary and sufficient for the convergence of $\{m(\mathbf{X}^{(k)})\}$, it follows that (4b) is valid.

Of (4c): First we estimate

$$\begin{aligned} d(\mathbf{X}^{(k+1)}) &= d(\mathbf{X}^{(k)})|(I - Am(\mathbf{X}^{(k)}))^{p-1}| \\ &= d(\mathbf{X}^{(k)})|(AA^{-1} - Am(\mathbf{X}^{(k)}))^{p-1}| \\ &\leq d(\mathbf{X}^{(k)})(|A||A^{-1} - m(\mathbf{X}^{(k)})|)^{p-1}| \\ &\leq d(\mathbf{X}^{(k)})2^{-(p-1)}(|A|d(\mathbf{X}^{(k)}))^{p-1}. \end{aligned}$$

Using a monotonic and multiplicative matrix norm $\|\cdot\|$ and the last relation, we get

$$\| d(\mathbf{X}^{(k+1)}) \| \leq 2^{-(p-1)} \| A \|^{p-1} \| d(\mathbf{X}^{(k)}) \|^p$$

Since the inequality

$$\| B \| \gamma_1 \leq \| B \| \leq \gamma_2 \| B \|, \quad \gamma_1 > 0, \ \gamma_2 > 0,$$

is valid for every matrix norm $\|\cdot\|$, from this inequality we get

$$\| d(\mathbf{X}^{(k+1)}) \| \gamma_1 \leq 2^{-(p-1)} \gamma_2^{(p-1)} \| A \|^{p-1} \gamma_2^p \| d(\mathbf{X}^{(k)}) \|^p,$$

which proves (4c). □

Remark 1. From the proof given above we see that the assertion of the theorem is also valid even if $\mathbf{X}^{(0)}$ is an arbitrary interval matrix not necessarily containing A^{-1}. In that case we will not have the inclusion $A^{-1} \in \mathbf{X}^{(k)}$ in general. We observe that the criterion (4b) depended only on the midpoint matrix $m(\mathbf{X}^{(0)})$ of the given inclusion matrix $\mathbf{X}^{(0)}$, while the width $d(\mathbf{X}^{(0)})$ can be arbitrary. For this reason, taking $m(\mathbf{X}^{(0)})$ to be an approximation to A^{-1} (but so that the condition (4b) holds) and choosing the elements of the matrix $\mathbf{X}^{(0)}$ to be large enough so that the enclosure of A^{-1} by $\mathbf{X}^{(0)}$ be ensured, we can provide not only the convergence of the method (4.4) but also the inclusion $A^{-1} \in \mathbf{X}^{(k)}$ $(k = 1, 2, \ldots)$.

Example 2. The S-H method (4.4) for $p = 2$ was applied for the inclusion of the inverse of the point matrix

$$A = \begin{bmatrix} \frac{4}{5} & \frac{1}{5} \\ \frac{3}{10} & \frac{9}{10} \end{bmatrix}.$$

The initial inclusion matrix was constructed according to the procedure (6.2) given in Section 6. Thus, with $a = 1/(1- \| I - A \|) = \frac{5}{3}$, for the initial matrix $\mathbf{X}^{(0)}$ we choose

$$\mathbf{X}^{(0)} = \begin{bmatrix} [-a, 2+a] & [-a, a] \\ [-a, a] & [-a, 2+a] \end{bmatrix} = \begin{bmatrix} [-\frac{5}{3}, \frac{11}{3}] & [-\frac{5}{3}, \frac{5}{3}] \\ [-\frac{5}{3}, \frac{5}{3}] & [-\frac{5}{3}, \frac{11}{3}] \end{bmatrix}.$$

In this way we ensure that $A^{-1} \in \mathbf{X}^{(0)}$ holds. Besides, we have $\rho(I - Am(\mathbf{X}^{(0)})) = \rho(I - A) = 0.8 < 1$, which provides the convergence of the iterative procedure (Theorem 4). The first four iterations give the following inclusion interval matrices (using arithmetic with 7 significant digits):

$$\mathbf{X}^{(1)} = \begin{bmatrix} [0.1666666, 2.2333316] & [-0.8999999, 0.4999999] \\ [-1.4333324, 0.8333329] & [0.5000000, 1.6999988] \end{bmatrix},$$

$$\mathbf{X}^{(2)} = \begin{bmatrix} [1.1716651, 1.5043325] & [-0.3969995, -0.1749999] \\ [-0.5963338, -0.2616657] & [1.0849990, 1.3049983] \end{bmatrix},$$

$$\mathbf{X}^{(3)} = \begin{bmatrix} [1.3587207, 1.3672409] & -0.3054331, -0.2997532] \\ [-0.4581502, -0.4496299] & [1.2088432, 1.2145233] \end{bmatrix},$$

$$\mathbf{X}^{(4)} = \begin{bmatrix} [1.3636322, 1.3636379] & [-0.3030319, -0.3030281] \\ [-0.4545477, -0.4545422] & [1.2121181, 1.2121219] \end{bmatrix}.$$

The applied iterative methods converges quadratically starting from the third iteration. Besides, in each iteration step we have

$$A^{-1} = \begin{bmatrix} \frac{15}{11} & -\frac{10}{33} \\ -\frac{15}{11} & \frac{40}{33} \end{bmatrix} = \begin{bmatrix} 1.36363636\ldots & -0.30303030\ldots \\ -0.45454545\ldots & 1.21212121\ldots \end{bmatrix} \in \mathbf{X}^{(k)}.$$

The sequence of the matrices produced by (4.4) always contains A^{-1} according to (4a) and, thus, it seems natural to form the intersection of the

new inclusion matrix $\mathbf{X}^{(k+1)}$ and the former matrix $\mathbf{X}^{(k)}$ in order to decrease the resulting matrix, which leads to the iterative method

(4.5)
$$\begin{cases} \mathbf{Y}^{(k+1)} = m(\mathbf{X}^{(k)}) \sum_{r=0}^{p-2} (I - Am(\mathbf{X}^{(k)}))^r + \mathbf{X}^{(k)} (I - Am(\mathbf{X}^{(k)}))^{p-1}, \\ \mathbf{X}^{(k+1)} = \mathbf{Y}^{(k+1)} \cap \mathbf{X}^{(k)}, \quad (k = 0, 1, \dots). \end{cases}$$

Using this iteration procedure one obtains a monotonic sequence

$$\mathbf{X}^{(0)} \supseteq \mathbf{X}^{(1)} \supseteq \mathbf{X}^{(2)} \supseteq \cdots$$

of inclusions for A^{-1}. The following numerical example does show, however, that the convergence criterion (4b) is not sufficient for convergence in general.

Example 3. ([3, Ch. 18]) We choose $p = 2$ and the matrices

$$A = \begin{bmatrix} 0.4 & 0.6 \\ -0.6 & 0.4 \end{bmatrix}, \quad \mathbf{X}^{(0)} = \begin{bmatrix} [-2, 4] & [-3, 3] \\ [-3, 3] & [-2, 4] \end{bmatrix},$$

which implies that $m(\mathbf{X}^{(0)}) = I$. We obtain

$$I - Am(\mathbf{X}^{(0)}) = \begin{bmatrix} 0.6 & -0.6 \\ 0.6 & 0.6 \end{bmatrix}$$

and calculate $\rho(I - Am(\mathbf{X}^{(0)})) = 0.6\sqrt{2} \approx 0.85 < 1$. Therefore, the procedure (4.4) converges to A^{-1} using this interval matrix. Applying (4.5) we find

$$\mathbf{Y}^{(1)} = m(\mathbf{X}^{(0)}) + \mathbf{X}^{(0)}(I - Am(\mathbf{X}^{(0)})) = \begin{bmatrix} [-2, 5.2] & [-4.2, 3] \\ [-3, 4.2] & [-2, 5.2] \end{bmatrix},$$

which implies that $\mathbf{X}^{(1)} = \mathbf{X}^{(0)}$. The sequence of matrices generated by (4.5) therefore does not converge to A^{-1} in contrast to the sequence computed by (4.4).

A convergence statement for the iteration (4.5) is contained in the following theorem.

Theorem 5. *Let A be a nonsingular $n \times n$ matrix and $\mathbf{X}^{(0)}$ an $n \times n$ interval matrix for which $A^{-1} \in \mathbf{X}^{(0)}$. If the sequence of matrices $\{\mathbf{X}^{(k)}\}$ is produced by* (4.5), *then*

(5a) *each matrix $\mathbf{X}^{(k)}$, $k \geq 0$, contains A^{-1};*

(5b) *if the inequality* $\rho(|I - AX|) < 1$ *is satisfied for all* $X \in \mathbf{X}^{(0)}$, *then the sequence* $\{\mathbf{X}^{(k)}\}$ *converges toward* A^{-1};

(5c) *the sequence* $\{d(\mathbf{X}^{(k)})\}$ *is bounded as follows:*

$$\| d(\mathbf{X}^{(k+1)}) \| \leq \gamma' \| d(\mathbf{X}^{(k)}) \|^p, \quad \gamma' \geq 0,$$

that is, the order of convergence of the iterative process (4.5) *is at least* p.

Proof. Of (5a): As in the proof of (4a) of Theorem 4 we first show that $A^{-1} \in \mathbf{Y}^{(k+1)}$, from which follows immediately that $A^{-1} \in \mathbf{X}^{(k+1)}$ since $A^{-1} \in \mathbf{X}^{(k)}$.

Of (5b): We shall use the fact that every sequence $\{\mathbf{X}^{(k)}\}$, for which $\mathbf{X}^{(0)} \supseteq \mathbf{X}^{(1)} \supseteq \mathbf{X}^{(2)} \supseteq \cdots$ holds, converges to an interval matrix $\mathbf{X} = (X_{ij})$, where

$$X_{ij} = \bigcap_{k=0}^{+\infty} X_{ij}^{(k)} \quad (i = 1, \ldots, m;\ j = 1, \ldots, n)$$

(see [3, Corollary 8 in Ch. 10]). Therefore, the sequence $\{\mathbf{X}^{(k)}\}$ obtained by (4.5) always converges to an interval matrix $\mathbf{X}$. We now show that under the assumptions of the theorem we must necessarily have $d(\mathbf{X}) = O$. We define

$$\mathbf{Y} = m(\mathbf{X}) \sum_{r=0}^{p-2} \big(I - Am(\mathbf{X})\big)^r + \mathbf{X}\big(I - Am(\mathbf{X})\big)^{p-1}$$

and obtain $\mathbf{X} = (X_{ij} \cap Y_{ij}) \subseteq \mathbf{Y}$ from (4.5). By (1) of Theorem 3 we get $d(\mathbf{X}) \leq d(\mathbf{Y})$. For $d(\mathbf{X})$ we obtain from (4.5)

$$d(\mathbf{X})|I - Am(\mathbf{X})|^{p-1} \geq d(\mathbf{X})|(I - Am(\mathbf{X}))^{p-1}| = d(\mathbf{Y}) \geq d(\mathbf{X}),$$

which implies that

$$d(\mathbf{X})\big(I - |I - Am(\mathbf{X})|^{p-1}\big) \leq O.$$

The assumption $\rho(|I - Am(\mathbf{X})|) < 1$ implies the existence of $\big(I - |I - Am(\mathbf{X})|^{p-1}\big)^{-1}$. It can be shown that this inverse is also nonnegative. From this it follows that $d(\mathbf{X}) \leq O$, and hence $d(\mathbf{X}) = O$. Taking into account (5a) we obtain $\mathbf{X} = A^{-1}$.

Of (5c): As in the proof of (4c) one first derives the inequality

$$\| d(\mathbf{Y}^{(k+1)}) \| \leq \gamma \| d(\mathbf{X}^{(k)}) \|^p$$

for a monotonic and multiplicative matrix norm $\|\cdot\|$. From this it follows that the inequality

$$\| d(\mathbf{X}^{(k+1)}) \| \leq \| d(\mathbf{Y}^{(k+1)}) \| \leq \gamma \| d(\mathbf{X}^{(k)}) \|^p$$

is valid since $\mathbf{X}^{(k+1)} \subseteq \mathbf{Y}^{(k+1)}$ as well as using (1) of Theorem 3 and the monotonicity of the norm $\|\cdot\|$. Analogous to the proof of (4c) we use the norm equivalence theorem to prove the final statement. □

5. Monotonicity of Schulz-Herzberger's method

J.W. Schmidt has proved in [28] that the inclusion $\mathbf{X}^{(1)} \subseteq \mathbf{X}^{(0)}$ is a necessary and sufficient condition for the monotonicity of the interval Schulz's method

$$\mathbf{X}^{(k+1)} = m(\mathbf{X}^{(k)}) + \mathbf{X}^{(k)}\Big(I - Am(\mathbf{X}^{(k)})\Big). \tag{5.1}$$

Starting from the above inclusion J. Herzberger has derived in [7] the necessary and sufficient condition which is of practical importance. Furthermore, using Schmidt's remark (given without a proof) that the inclusion $\mathbf{X}^{(1)} \subseteq \mathbf{X}^{(0)}$ is also necessary and sufficient for the monotonicity of the higher-order method (4.4) (see [28]), J. Herzberger has considered in [9] the monotonicity of (4.4).

The aim of this section is to give a useful sufficient condition for the monotonicity of the S-H method (4.4). Our consideration reduces to Herzberger's results [7] concerning the iterative method (5.1), which can be generalized for the method (4.4).

Lemma 1. *Let* $\mathbf{X}^{(0)}, \mathbf{X}^{(1)}, \ldots$ *be the sequence of interval matrices produced by the iterative formula* (4.4) *and let* $\rho(|I - Am(\mathbf{X}^{(0)})|) < 1$. *If the inequality*

$$2\big|m(\mathbf{X}^{(k)})\Big(I - Am(\mathbf{X}^{(k)})\Big)\big| \leq d(\mathbf{X}^{(k)})(I - |I - Am(\mathbf{X}^{(k)})|) \tag{5.2}$$

is valid for $k = 0$, *then it holds for each* $k = 0, 1, 2, \ldots$.

Proof. For brevity, let us introduce the notations

$$\mathbf{C}_k = I - Am(\mathbf{X}^{(k)}), \quad \mathbf{B}_k = |\mathbf{C}_k|.$$

From (4.4) we find the midpoint matrix $m(\mathbf{X}^{(k+1)})$ and the width matrix $d(\mathbf{X}^{(k+1)})$,

$$m(\mathbf{X}^{(k+1)}) = m(\mathbf{X}^{(k)}) \sum_{r=0}^{p-1} \mathbf{C}_k^r, \tag{5.3}$$

$$d(\mathbf{X}^{(k+1)}) = d(\mathbf{X}^{(k)}) |\mathbf{C}_k^{p-1}|. \tag{5.4}$$

Using inequalities

$$|\mathbf{XY}| \leq |\mathbf{X}||\mathbf{Y}|, \quad |\mathbf{X}+\mathbf{Y}| \leq |\mathbf{X}|+|\mathbf{Y}|$$

for the absolute value matrices, in the special case of the point matrices we obtain

$$|\mathbf{C}_k^r| \leq |\mathbf{C}_k|^r = \mathbf{B}_k^r, \tag{5.5}$$

$$\left|\sum_{r=0}^{p-1} \mathbf{C}_k^r\right| \leq \sum_{r=0}^{p-1} |\mathbf{C}_k|^r = \sum_{r=0}^{p-1} \mathbf{B}_k^r. \tag{5.6}$$

Starting from (5.3), we find

$$Am(\mathbf{X}^{(k+1)}) = \left(I - \left(I - Am(\mathbf{X}^{(k)})\right)\right) \sum_{r=0}^{p-1} \mathbf{C}_k^r = I - \mathbf{C}_k^p,$$

wherefrom

$$\mathbf{C}_{k+1} = I - Am(\mathbf{X}^{(k+1)}) = \mathbf{C}_k^p = \mathbf{C}_0^{p^{k+1}}. \tag{5.7}$$

Since $\rho(|I - Am(\mathbf{X}^{(0)})|) = \rho(\mathbf{B}_0) < 1$ implies $\rho(\mathbf{B}_0^\nu) < 1$ $(\nu > 1)$, we have

$$\rho(|\mathbf{C}_k|) = \rho(|\mathbf{C}_0^{p^k}|) \leq \rho(\mathbf{B}_0^{p^k}) < 1,$$

that is

$$\rho(\mathbf{B}_0) < 1 \quad \text{implies} \quad \rho(\mathbf{B}_k) < 1, \quad k = 0, 1, \ldots . \tag{5.8}$$

Furthermore, because of $\rho(\mathbf{B}_k) < 1$ there exists the inverse matrix $(I - \mathbf{B}_k)^{-1} \geq O$ and the following identity is valid

$$\sum_{r=0}^{p-1} \mathbf{B}_k^r = (I - \mathbf{B}_k)^{-1}(I - \mathbf{B}_k^p).$$

We shall now prove that the inequality (5.2), where $\mathbf{X}^{(k)}$ is given by (4.4), is valid for each $k = 1, 2, \ldots$ if

$$2|m(\mathbf{X}^{(0)})\mathbf{C}_0| \leq d(\mathbf{X}^{(0)})(I - \mathbf{B}_0) \tag{5.9}$$

(the inequality (5.2) for $k = 0$) holds.

Let us rewrite (5.2) in a (shorter) form

$$2|m(\mathbf{X}^{(k)})\mathbf{C}_k| \le d(\mathbf{X}^{(k)})(I - \mathbf{B}_k) \tag{5.10}$$

and assume that this inequality holds for some index $k \ge 0$. Multiplying both sides of (5.10) by $(I - \mathbf{B}_k)^{-1}(I - \mathbf{B}_k^p)|\mathbf{C}_k^{p-1}|$, one obtains

$$2|m(\mathbf{X}^{(k)})\mathbf{C}_k|(I - \mathbf{B}_k)^{-1}(I - \mathbf{B}_k^p)|\mathbf{C}_k^{p-1}| \le d(\mathbf{X}^{(k)})(I - \mathbf{B}_k^p)|\mathbf{C}_k^{p-1}|$$

or

$$2|m(\mathbf{X}^{(k)})\mathbf{C}_k|\sum_{r=0}^{p-1}\mathbf{B}_k^r|\mathbf{C}_k^{p-1}| \le d(\mathbf{X}^{(k)})(I - \mathbf{B}_k^k)|\mathbf{C}_k^{p-1}|. \tag{5.11}$$

Using inequalities

$$\mathbf{B}_k^p = |\mathbf{C}_k|^p \ge |\mathbf{C}_k^{p-1}||\mathbf{C}_k| \ge |\mathbf{C}_k^p|,$$

we find

$$\begin{aligned}(I - \mathbf{B}_k^p)|\mathbf{C}_k^{p-1}| &\le (I - |\mathbf{C}_k^{p-1}||\mathbf{C}_k|)|\mathbf{C}_k^{p-1}| \\ &\le |\mathbf{C}_k^{p-1}| - |\mathbf{C}_k^{p-1}||\mathbf{C}_k^p| = |\mathbf{C}_k^{p-1}|(I - |\mathbf{C}_k^p|) \\ &= |\mathbf{C}_k^{p-1}|(I - \mathbf{B}_{k+1}).\end{aligned}$$

According to (5.6) and the last inequality, from (5.11) we obtain

$$2\Big|m(\mathbf{X}^{(k)})(\sum_{r=0}^{p-1}\mathbf{C}_k^r)\mathbf{C}_k^p\Big| \le d(\mathbf{X}^{(k)})|\mathbf{C}_k^{p-1}|(I - \mathbf{B}_{k+1}). \tag{5.12}$$

Taking into account formulas (5.3), (5.4) and (5.7), the inequality (5.12) becomes

$$2|m(\mathbf{X}^{(k+1)})\mathbf{C}_{k+1}| \le d(\mathbf{X}^{(k+1)})(I - \mathbf{B}_{k+1}).$$

This proves (5.10) (that is, (5.2)) by complete induction since (5.9) holds as the assumption of Lemma 1. □

Theorem 6. *Let* $A^{-1} \in \mathbf{X}^{(0)}$ *and* $\rho(|I - Am(\mathbf{X}^{(0)})|) < 1$. *Then the generalized interval method* (4.4) *converges to* A^{-1}, *where* $A^{-1} \in \mathbf{X}^{(k)}$ $(k = 0, 1, \dots)$, *and if*

$$2|m(\mathbf{X}^{(0)})(I - Am(\mathbf{X}^{(0)}))| \le d(\mathbf{X}^{(0)})(I - |I - Am(\mathbf{X}^{(0)})|) \tag{5.13}$$

holds, then the method (4.4) *is monotone.*

Proof. First, we observe that under the given assumption, there follows that (4.4) converges because

$$\rho\big(|I - Am(\mathbf{X}^{(0)})|\big) < 1 \quad \text{implies} \quad \rho\big(I - Am(\mathbf{X}^{(0)})\big) < 1.$$

The inclusion $A^{-1} \in \mathbf{X}^{(k)}$ for each $k \geq 0$ has been proved in Theorem 4.

Under the condition (5.13) of Theorem 6 (and Lemma 1, too) the inequality

$$2|m(\mathbf{X}^{(k)})\mathbf{C}_k| \leq d(\mathbf{X}^{(k)})(I - \mathbf{B}_k)$$

holds for each $k \geq 0$. Multiplying both sides of the last inequality by

$$\sum_{r=0}^{p-2} \mathbf{B}_k^r = (I - \mathbf{B}_k)^{-1}(I - \mathbf{B}_k^{p-1}) \geq O,$$

we obtain

$$2|m(\mathbf{X}^{(k)})\mathbf{C}_k| \sum_{r=0}^{p-2} \mathbf{B}_k^r \leq d(\mathbf{X}^{(k)})(I - \mathbf{B}_k^{p-1}). \tag{5.14}$$

Since

$$\Big|m(\mathbf{X}^{(k)}) \sum_{r=1}^{p-1} \mathbf{C}_k^r\Big| \leq |m(\mathbf{X}^{(k)})\mathbf{C}_k| \sum_{r=0}^{p-2} \mathbf{B}_k^r$$

and

$$I - \mathbf{B}_k^{p-1} \leq I - |\mathbf{C}_k^{p-1}|,$$

from (5.14) we obtain

$$2\Big|m(\mathbf{X}^{(k)}) \sum_{r=1}^{p-1} \mathbf{C}_k^r\Big| \leq d(\mathbf{X}^{(k)})(I - |\mathbf{C}_k^{p-1}|)$$

or

$$\Big|m(\mathbf{X}^{(k)}) \sum_{r=0}^{p-1} \mathbf{C}_k^r - m(\mathbf{X}^{(k)})\Big| \leq \frac{1}{2}\big(d(\mathbf{X}^{(k)}) - d(\mathbf{X}^{(k)})|\mathbf{C}_k^{p-1}|\big).$$

Finally, according to the formulas (5.3) and (5.4) for the matrices $m(\mathbf{X}^{(k+1)})$ and $d(\mathbf{X}^{(k+1)})$, the last inequality becomes

$$|m(\mathbf{X}^{(k+1)}) - m(\mathbf{X}^{(k)})| \leq \frac{1}{2}\big(d(\mathbf{X}^{(k)}) - d(\mathbf{X}^{(k+1)})\big). \tag{5.15}$$

But, the inequality (5.15) is necessary and sufficient for the inclusion

$$\mathbf{X}^{(k+1)} \subseteq \mathbf{X}^{(k)}. \tag{5.16}$$

Therefore, if the condition (5.13) is satisfied, then the inclusion (5.16) holds for each $k \geq 0$, which means that the generalized iterative method (4.4) is monotone. This completes the proof of the theorem. □

Remark 2. The condition (5.13) can be rewritten in the form

$$2|m(\mathbf{X}^{(0)})(I - Am(\mathbf{X}^{(0)}))|(I - |I - Am(\mathbf{X}^{(0)})|)^{-1} \leq d(\mathbf{X}^{(0)}). \tag{5.17}$$

Since this condition depends only on the given matrix A and the initial approximation $m(\mathbf{X}^{(0)})$ for A^{-1}, the matrix $d(\mathbf{X}^{(0)}) \geq O$ can always be chosen so that (5.17) is satisfied. Since the convergence condition $\rho(|I - Am(\mathbf{X}^{(0)})|) < 1$ does not depend on the width matrix $d(\mathbf{X}^{(0)})$, this matrix can be taken so that

(i) an initial interval matrix $\mathbf{X}^{(0)}$ safely includes A^{-1} and
(ii) the monotonicity of the iterative method (4.4) is provided.

We observe that (5.13) coincides with the corresponding condition obtained for the interval Schulz's method (5.1). Since the construction of the proof of the assertion which gives a sufficient condition for the monotonicity of (5.1) is directly based on the relation (5.13) (see [7, Theorem 2]), for the higher-order interval method (4.4) ($p > 1$) we immediately have the following theorem:

Theorem 7. *Let* $\|I - Am(\mathbf{X}^{(0)})\| < 1$ ($\|\cdot\|$ *the column-sum norm), then the method* (4.4) *converges to* A^{-1}. *In addition, this method is monotone if the following is valid*

$$d(X_{ij}^{(0)}) = h \geq \frac{2 \cdot \max\limits_{i,j} |m(X_{ij}^{(0)})|}{1 - \|I - Am(\mathbf{X}^{(0)})\|} \quad \text{for } i \neq j, \quad d(X_{ii}^{(0)}) \geq h. \tag{5.18}$$

Theorem 7 gives a sufficient condition for the monotonicity of the generalized interval method (4.4). Under the given assumptions of this theorem it is always possible to choose the width matrix $d(\mathbf{X}^{(0)})$ in such a way that the method (4.4) is monotone. A detailed description of the construction of the initial including matrix $\mathbf{X}^{(0)}$ which guarantees for $A^{-1} \in \mathbf{X}^{(0)}$ is given in the next section.

6. Construction of the initial inclusion matrix

The convergence criterion (5b) in Theorem 5 depends on the width of the inclusion matrix $\mathbf{X}^{(0)}$ for A^{-1}, which is not a case with the criterion (4b) in Theorem 4. Nevertheless, it is not difficult to find a relation between these criteria. For instance, if an interval matrix $\mathbf{X}^{(0)}$ satisfies the inequality $\| I - Am(\mathbf{X}^{(0)}) \| < 1$, for a monotonic and multiplicative norm $\| \cdot \|$, then we have that

$$\| d(\mathbf{X}^{(0)}) \| < \alpha = 2(1 - \| I - Am(\mathbf{X}^{(0)}) \|) / \| A \| \tag{6.1}$$

is a sufficient criterion for the statement that $\| I - AX \| < 1$ for all $X \in \mathbf{X}^{(0)}$. To construct a suitable interval matrix $\mathbf{X}^{(0)}$ let us assume that A may be represented as $A = I - B$ with $\| B \| < 1$. The choice $m(\mathbf{X}^{(0)}) := I$ gives

$$\| I - Am(\mathbf{X}^{(0)}) \| = \| B \| < 1,$$

and, according to the criterion (4b), the inclusion method (4.4) is convergent for every interval matrix $\mathbf{X}^{(0)}$ for which $m(\mathbf{X}^{(0)}) = I$. In order to insure the inclusion $A^{-1} \in \mathbf{X}^{(0)}$ we consider the equation $AX = (I - B)X = I$ or $X = BX + I$. In regard to this there follows (using a multiplicative matrix norm) that

$$\| X \| \leq a := \frac{1}{1 - \| B \|},$$

wherefrom (using the row-sum or the column-sum norm)

$$-a \leq x_{ij} \leq a \quad (1 \leq i, j \leq n)$$

for all the elements of $X = (x_{ij})$. For the matrix $\mathbf{X}^{(0)} = (X_{ij})$ defined by

$$X_{ij}^{(0)} = \begin{cases} [-a, a] & \text{for } i \neq j \\ [-a, 2 + a] & \text{for } i = j, \end{cases} \tag{6.2}$$

we have $A^{-1} \in \mathbf{X}^{(0)}$ and also $m(\mathbf{X}^{(0)}) = I$. By virtue of Theorem 4 the iterative method converges to A^{-1}.

From the above consideration, we see that the iterative method (4.4) requires weaker convergence conditions compared to (4.5). For this reason, it is convenient to start with the method (4.4) as soon as the sufficient condition (6.1) is fulfilled provided $\| I - Am(\mathbf{X}^{(0)}) \| < 1$ and then to continue with the method (4.5). Such a combined process has been described in details by Alefeld and Herzberger [1].

The sufficient condition (6.1) can be weakened, which is the subject of the following assertion:

Theorem 8. *If* $\mathbf{X}^{(k)}$ *is an inclusion matrix for* A^{-1}, *then*

$$\| d(\mathbf{X}^{(k)}) \| < \beta = \frac{2}{\| A \|} \tag{6.3}$$

is a sufficient condition for the convergence of (4.5) *to* A^{-1}.

Proof. Applying the width operator d to the iterative formula (4.5), we obtain

$$\begin{aligned} d(\mathbf{X}^{(k+1)}) &\leq d\Big(m(\mathbf{X}^{(k)}) \sum_{r=0}^{p-2} \Big(I - Am(\mathbf{X}^{(k)})\Big) + \mathbf{X}^{(k)} \Big(I - Am(\mathbf{X}^{(k)})\Big)^{p-1}\Big) \\ &\leq d(\mathbf{X}^{(k)}) 2^{-p+1} \Big(|A| d(\mathbf{X}^{(k)})\Big)^{p-1}. \end{aligned}$$

Using a monotonic and multiplicative matrix norm, we get

$$\| d(\mathbf{X}^{(k+1)}) \| \leq \left(\frac{\| A \|}{2}\right)^{p-1} \| d(\mathbf{X}^{(k)}) \|^p,$$

which proves that (6.3) is sufficient for $\| d(\mathbf{X}^{(k+1)}) \| \to 0$, and whence, $\mathbf{X}^{(k+1)} \to A^{-1}$. □

Remark 3. Comparing the numbers α and β appearing in (6.1) and (6.3) we infer that $\alpha < \beta$, which means that the condition (6.3) is weaker than (6.1). Furthermore, β is considerably simpler to calculate and has the same value for all the matrices $\mathbf{X}^{(k)}$. Finally, the criterion (6.3) from Theorem 6 is even considerably less restrictive than that of Theorem 5, as it was shown in [8].

The result given in the following theorem provides a better inclusion for A^{-1} compared with (6.2).

Theorem 9. *For the initial inclusion matrix* $\widehat{\mathbf{X}}^{(0)}$ *defined by*

$$\widehat{\mathbf{X}}^{(0)} = I + ([-c, c]) \quad \textit{with} \quad c = \frac{\| B \|}{1 - \| B \|}$$

we have $A^{-1} \in \widehat{\mathbf{X}}^{(0)}$ *and the iterative process* (4.4) *converges to* A^{-1} ($\| \cdot \|$ *row-sum or colomn-sum norm*).

Proof. Starting from the obvious equalities

$$A^{-1} - I = (I - B)^{-1} - I = (I - B)^{-1} B$$

and using a multiplicative matrix norm $\| \cdot \|$ and the well-known inequality

$$\| (I - B)^{-1} \| \leq \frac{1}{1 - \| B \|},$$

we obtain

$$\| A^{-1} - I \| \leq \| (I - B)^{-1} \| \, \| B \| \leq \frac{\| B \|}{1 - \| B \|}.$$

In this way the inclusion $A^{-1} \in \hat{\mathbf{X}}^{(0)}$ is proved. Further, since

$$\| I - Am(\hat{\mathbf{X}}^{(0)}) \| = \| B \| < 1,$$

the iterative method (4.4) converges to A^{-1} (see Theorem 4). □

Remark 4. The computation of $\mathbf{X}^{(0)}$ and $\hat{\mathbf{X}}^{(0)}$ requires the same amount of work but we have $\hat{\mathbf{X}}^{(0)} \subset \mathbf{X}^{(0)}$.

For nonsingular matrices A which do not have the same property as in the previous, some other approach which uses Theorem 7 has to be applied for constructing a starting matrix for (4.4) with $A^{-1} \in \mathbf{X}^{(0)}$. Namely, according to Remark 1, the iterative method (4.4) converges to A^{-1} even if $\mathbf{X}^{(0)}$ does not include A^{-1}. But the construction (5.18) guarantees the monotonicity of the interval matrices produced by (4.4),

$$\mathbf{X}^{(0)} \supseteq \mathbf{X}^{(1)} \supseteq \mathbf{X}^{(2)} \supseteq \dots,$$

and thus we necessarily have $A^{-1} \in \mathbf{X}^{(0)}$.

7. Combined Schulz-type methods

In this section we describe a general approach to the construction of new methods of *Schulz's type* for improving bounds for the inverse A^{-1} of a given $n \times n$ nonsingular matrix A. These methods, proposed by J. Herzberger and Lj. Petković [18], [19], possess a great computational efficiency.

It is well known that interval evaluations are more costly than ordinary floating-point computations. For this reason, it would be advisable to apply the necessary interval computations only in a part of the algorithm. The aim of this section is to present an approach for solving this problem, which combines iterative methods in floating-point arithmetic as well as in interval arithmetic. In this way, we take advantage of comparatively small computational costs of floating-point arithmetic and the very important inclusion property of interval arithmetic (the enclosure of the exact result).

Definition 11. The mapping Φ from the set of $n \times n$-matrices onto itself is called a Schulz-type method of order $p \geq 2$ for A^{-1} if and only if for $Y = \Phi(X, A)$ the equation

$$I - AY = (I - AX)^p \tag{7.1}$$

holds true.

Remark 5. For practical computations Φ should only consist of matrix multiplications and additions.

Two the most frequently used examples of the mapping Φ are given below:

Example 4. Let $p \geq 2$, then

$$\Phi_p(X, A) = X \sum_{i=0}^{p-1} (I - AX)^i \tag{7.2}$$

defines a Schulz-type method for A^{-1} of order p.

Example 5. We can use Ostrowski's identity (see [22])

$$\begin{aligned} \Phi_5(X, A) =& X \cdot \left(I + \frac{\sqrt{5}+1}{2}(I - AX) + (I - AX)^2 \right) \\ & \times \left(I - \frac{\sqrt{5}+1}{2}(I - AX) + (I - AX)^2 \right) \\ =& X \cdot \sum_{\nu=0}^{4} (I - AX)^\nu, \end{aligned} \tag{7.3}$$

which also gives a Schulz's type method of order 5.

By means of a Schulz-type method for A^{-1} we can construct an iteration method in ordinary floating-point arithmetic as follows:

$$X^{(k+1)} = \Phi(X^{(k)}, A), \quad X^{(0)} \text{ given}, \quad k \geq 0. \tag{7.4}$$

The following assertion has been proved in [18]:

Theorem 10. *Let Φ be a Schulz-type method for A^{-1} of order p. Then the sequences of matrices $\{X^{(k)}\}$ produced by (7.4) have the following properties:*

(a) $X^{(k)} \to A^{-1} \Leftrightarrow \rho(I - AX^{(0)}) < 1$,

(b) *if the method (7.4) is convergent, then its order of convergence is at least p.*

where for $x \in A$, $\mathcal{A}(x)$ denotes the principal element of $\mathcal{A}$ generated by x.

For a semigroup S, $\mathcal{Q}(S)$ will denote the lattice of quasi-orders on S, $\mathcal{E}(S)$ will denote the lattice of equivalence relations on S and $\mathrm{Con}\,(S)$ will denote the lattice of congruence relations on S. It is well-known that $\mathrm{Con}\,(S)$ is a complete sublattice of $\mathcal{E}(S)$ and $\mathcal{E}(S)$ is a complete sublattice of $\mathcal{Q}(S)$. By $\mathcal{E}^*(S)$ we denote the lattice of 0-restricted equivalence relation on a semigroup $S = S^0$, which is the principal ideal of $\mathcal{E}(S)$ generated by the equivalence relation χ determined by the partition $\{S^*, 0\}$.

An ideal A of a semigroup S is a *prime ideal* if for $x, y \in S$, $xSy \subseteq A$ implies that either $x \in A$ or $y \in A$, or, equivalently, if for all ideals M and N of S, $MN \subseteq A$ implies that either $M \subseteq A$ or $N \subseteq A$. A completely 0-simple semigroup with the property that the structure group of its Rees-matrix representation is the one-element group, is called a *rectangular 0-band*. Equivalently, a rectangular 0-band can be defined as a semigroup $S = S^0$ in which 0 is a prime ideal and for all $a, b \in S$, either $aba = a$ or $aba = 0$.

For undefined notions and notations we refer to the following books: G. Birkhoff [2], S. Bogdanović [4], S. Bogdanović and M. Ćirić [7], S. Burris and H. P. Sankappanavar [17], A. H. Clifford and G. B. Preston [35], [36], G. Grätzer [45], J. M. Howie [48], E. S. Lyapin [62], M. Petrich [72], [73], L. N. Shevrin [98], L. N. Shevrin and A. Ya. Ovsyanikov [102], [103], O. Steinfeld [105] and G. Szász [109].

1.2. A classification of decompositions

In this section we classify decompositions of semigroups into few classes and we single out the most important types of decompositions.

Let us say again that by a decompositions of a semigroup S we mean a family $\mathcal{D} = \{S_\alpha\}_{\alpha \in Y}$ of subsets of S satisfying the condition

$$S = \bigcup_{\alpha \in Y} S_\alpha, \qquad \text{where } S_\alpha \cap S_\beta = \varnothing, \text{ for } \alpha, \beta \in Y, \alpha \neq \beta.$$

Various special kinds of decompositions we obtain in two general ways: imposing some requirements on the structure of the components S_α, and imposing some requirements on products of elements from different classes.

The first general type of decompositions that we single out are *decompositions S onto subsemigroups*, determined by the property that any S_α is a subsemigroup of S. Clearly, decompositions onto subsemigroups correspond to equivalence relations satisfying the *cm*-property, so the following theorem can be easily proved:

Theorem 1.1. *The poset of decompositions of a semigroup S onto subsemigroups is a complete lattice which is dually isomorphic to the lattice of equivalence relations on S satisfying the cm-property.*

If to a decomposition of a semigroup S onto subsemigroups we impose an additional condition

$$ab \in \langle a \rangle \cup \langle b \rangle,$$

for all elements $a, b \in S$ belonging to the different components, then we obtain so called $\cup$-*decompositions*. Decompositions of this type will be considered in Section 5.1.

The second general class of decompositions that we single out form decompositions whose related equivalence relations are congruences. Decompositions of this type are called *decompositions by congruences*. When the decomposition $\mathcal{D}$ is a decomposition by a congruence relation, then the index set Y is a factor semigroup of S and many properties of S are determined by structure of the semigroup Y. Special types of decompositions by congruences we obtain imposing some requirements on the structure of the related factor semigroup. If a class $\mathfrak{C}$ of semigroups and a semigroup S are given, then a congruence relation θ on S is called a $\mathfrak{C}$-*congruence* on S if the related factor S/θ is in $\mathfrak{C}$, the related decomposition is given a $\mathfrak{C}$-*decomposition*, and the related factor semigroup is called a $\mathfrak{C}$-*homomorphic image* of S. When there exists the greatest $\mathfrak{C}$-decomposition of S, i.e. the smallest $\mathfrak{C}$-congruence on S, then we say that the factor semigroup corresponding to this congruence is the *greatest* $\mathfrak{C}$-*homomorphic image* of S. A semigroup S is called $\mathfrak{C}$-*indecomposable* if the universal relation is the unique $\mathfrak{C}$-congruence on S. Of course, when the class $\mathfrak{C}$ contains the trivial (one-element) semigroup, then the $\mathfrak{C}$-decompositions determine a decomposition type.

If the decomposition $\mathcal{D}$ is both a decomposition by a congruence relation and a decomposition onto subsemigroups, then it is called a *band decomposition* of S and the related congruence relation is called a *band congruence* on S. Equivalently, the type of band decompositions is defined as the type of $\mathfrak{C}$-decompositions, where $\mathfrak{C}$ equals the variety $[x^2 = x]$ of bands. Moreover, by some subvarieties of the variety of bands we define the following very important special types of band decompositions and band congruences:

- *semilattice decompositions* and *congruences*, determined by the variety $[x^2 = x, xy = yx]$ of *semilattices*;
- *matrix decompositions* and *congruences*, determined by the variety $[x^2 = x, xyx = x] = [x^2 = x, xyz = xz]$ of *rectangular bands*;
- *left (right) zero band decompositions* and *congruences*, determined by the variety $[x^2 = x, xy = x]$ ($[x^2 = x, xy = y]$) of *left (right) zero bands*;

- *normal band decompositions* and *congruences*, determined by the variety $[x^2 = x, xyzx = xzyx] = [x^2 = x, xyzu = xzyu]$ of *normal bands*;
- *left (right) normal band decompositions* and *congruences*, determined by the variety $[x^2 = x, xyz = xzy]$ ($[x^2 = x, xyz = yxz]$) of *left (right) normal bands*.

Also, *chain decompositions* and *congruences* are determined by the class of *chains* (linearly ordered semilattices). The decomposition $\mathcal{D}$ is called an *ordinal decomposition* if it is a chain decomposition, i.e. Y is a chain, and for all $a, b \in S$,

$$a \in S_\alpha,\ b \in S_\beta,\ \alpha < \beta \quad \Rightarrow \quad ab = ba = a.$$

These decompositions will be considered in Section 5.2. In the last chapter of this paper we will also consider I-matrix decompositions and semilattice-matrix decompositions, which will be precisely defined in Sections 5.3 and 5.4, respectively.

Semigroups with zero have a specific structure and in studying of such semigroups it is often convenient to represent a semigroup $S = S^0$ in the form:

$$S = \bigcup_{\alpha \in Y} S_\alpha, \qquad \text{where } S_\alpha \cap S_\beta = 0, \text{ for } \alpha, \beta \in Y, \alpha \neq \beta.$$

In this case, the partition $\mathcal{D}$ of S, whose components are 0 and $S_\alpha^\bullet$, $\alpha \in Y$, is called a *0-decomposition* of S. If, moreover, any S_α is a subsemigroup of S, we say that $\mathcal{D}$ is a *0-decomposition* of S *onto subsemigroups* and that S is a *0-sum* of semigroups S_α, $\alpha \in Y$, and the semigroups S_α will be called the *summands* of this decomposition. Equivalence relations corresponding to such decompositions are exactly the ones which satisfy the 0-*cm*-property, so the following theorem follows:

Theorem 1.2. *The poset of 0-decompositions of a semigroup $S = S^0$ onto subsemigroups is a complete lattice which is dually isomorphic to the lattice of equivalence relations on S satisfying the 0-cm-property.*

Special decompositions of this type may be determined by some properties of the index set Y. Namely, it is often convenient to assume that Y is a partial groupoid whose all elements are idempotents, and to require that the multiplication on S is carried by Y, by the following condition:

$$\begin{cases} S_\alpha S_\beta \subseteq S_{\alpha\beta} & \text{if } \alpha\beta \text{ is defined in } Y \\ S_\alpha S_\beta = 0 & \text{otherwise} \end{cases}, \qquad \text{for all } \alpha, \beta \in Y.$$

For example, if Y is a semigroup, i.e. a band, we obtain so called *0-band decompositions*. If the product $\alpha\beta$ is undefined, whenever $\alpha \neq \beta$, then $S_\alpha S_\beta = 0$, whenever $\alpha \neq \beta$, and such decompositions are called *orthogonal decompositions*. If Y is a left (right) zero band, then the corresponding decomposition is called a *decomposition into a left (right) sum* of semigroups. If Y is a nonempty subset of $I \times \Lambda$, where I and Λ are nonempty sets, and the partial multiplication on Y is defined by: for $(i,\lambda),(j,\mu) \in Y$, the product $(i,\lambda)(j,\mu)$ equals (i,μ), if $(i,\mu) \in Y$, and it is undefined, otherwise, then the decomposition $\mathcal{D}$ carried by Y is called a *decomposition into a matrix sum* of semigroups S_α, $\alpha \in Y$. Finally, if Y is an arbitrary poset and for $\alpha, \beta \in Y$, the product $\alpha\beta$ is defined as the meet of α and β, if it exists, then the related decomposition is called a *quasi-semilattice decomposition* of S.

1.3. Decompositions by congruences

Given a nonempty class $\mathfrak{C}$ of semigroups. Let $\mathrm{Con}_{\mathfrak{C}}(S)$ denotes the set of all $\mathfrak{C}$-congruences on S. Of course, $\mathrm{Con}_{\mathfrak{C}}(S)$ is a subset of $\mathrm{Con}(S)$ and it can be treated as a poset with respect to the usual ordering of congruences. Properties of posets of $\mathfrak{C}$-congruences inside the lattice $\mathrm{Con}(S)$ have been first investigated by T. Tamura and N. Kimura in [123], 1955, where they proved the following theorem:

Theorem 1.3. (T. Tamura and N. Kimura [123]) *If $\mathfrak{C}$ is a variety of semigroups, then $\mathrm{Con}_{\mathfrak{C}}(S)$ is a complete lattice, for any semigroup S.*

For the variety of semilattices, the previous theorem has been proved also by T. Tamura and N. Kimura [122], 1954 (see Theorem 2.1).

The problem of existence of the greatest decomposition of a given type has been solved in a special case, for so-called μ-decompositions, by T. Tamura [110], 1956. The solution of this problem in the general case has been given by N. Kimura [54], 1958, by the next theorem. Note that by an *algebraic class of* of semigroups we mean a class of semigroups closed under isomorphisms.

Theorem 1.4. (N. Kimura [54]) *Let $\mathfrak{C}$ be a nonempty algebraic class of semigroups. Then $\mathfrak{C}$ is closed under subdirect products if and only if $\mathrm{Con}_{\mathfrak{C}}(S)$ has the smallest element, for any semigroup S for which $\mathrm{Con}_{\mathfrak{C}}(S) \neq \varnothing$.*

As N. Kimura [54] noted, this theorem has been also found by E. J. Tully. Note that if $\mathrm{Con}_{\mathfrak{C}}(S)$ has the smallest elements, then it is a complete meet-subsemilattice of $\mathrm{Con}(S)$.

The converse of Theorem 1.3 has been proved in a recent paper of M. Ćirić and S. Bogdanović [24]. Namely, they proved the following theorem:

Theorem 1.5. (M. Ćirić and S. Bogdanović [24]) *Let $\mathfrak{C}$ be a nonempty algebraic class of semigroups. Then $\mathfrak{C}$ is a variety if and only if $\mathrm{Con}_{\mathfrak{C}}(S)$ is a complete sublattice of $\mathrm{Con}(S)$, for any semigroup S.*

By the proof of the previous theorem, given in [24], the next theorem also follows:

Theorem 1.6. (T. Tamura and N. Kimura [123]) *If $\mathfrak{C}$ is a variety of semigroups, then $\mathrm{Con}_{\mathfrak{C}}(S)$ is a principal dual ideal of $\mathrm{Con}(S)$, for any semigroup S.*

Note that Theorems 1.4, 1.5 and 1.6 holds also for any algebra.

The following theorem, proved by M. Petrich in [72], 1973, has been very useful in his investigations of some greatest decompositions of semigroups.

Theorem 1.7. (M. Petrich [72]) *Let $\mathfrak{C}$ be a variety of semigroups, $\mathfrak{D}$ the class of subdirectly irreducible semigroups from $\mathfrak{C}$ and S any semigroup. Then a congruence θ on a semigroup S, different from the universal congruence, is a $\mathfrak{C}$-congruence if and only if it is the intersection of some family of $\mathfrak{D}$-congruences.*

If we define the trivial semigroup to be subdirectly irreducible, then Theorem 1.7 says that $\mathrm{Con}_{\mathfrak{D}}(S)$ is meet-dense in $\mathrm{Con}_{\mathfrak{C}}(S)$.

2. Semilattice decompositions

Semilattice decompositions of semigroups have been first defined and studied by A. H. Clifford [29], 1941. After that they have been investigated by many authors and they have been systematically studied in several monographs: by E. S. Lyapin [62], 1960, A. H. Clifford and G. B. Preston [35], 1961, M. Petrich [72], 1973, and [73], 1977, S. Bogdanović [4], 1985, S. Bogdanović and M. Ćirić [7], 1993, and other.

First general results concerning semilattice decompositions of semigroups have been the results of T. Tamura and N. Kimura from [122], 1954. There they proved a theorem, given below as Theorem 2.1, by which it follows the existence of the greatest semilattice decomposition on any semigroup. This theorem initiated intensive studying of the greatest semilattice decompositions of semigroups and Section 2.1 is devoted to the results from this area. We present various characterizations of the greatest semilattice decomposition of a semigroup, the smallest semilattice congruence on a semigroup and the greatest semilattice homomorphic image of a semigroup, given by M. Yamada [132], 1955, T. Tamura [110], 1956, [112], 1964, and [117], 1972, M.

Petrich [69], 1964, and [72], 1973, M. S. Putcha [79], 1973, and [80], 1975, and M. Ćirić and S. Bogdanović [21]. We also quote the famous theorem of T. Tamura [110], 1956, on atomicity of semilattice decompositions, which is probably the most important result of the theory of semilattice decompositions of semigroups, and we give several characterizations of semilattice indecomposable semigroups given by M. Petrich [69], 1964, and [72], 1973, and T. Tamura [117], 1972. For the related results concerning decompositions of groupoids we refer to G. Thierrin [127], 1956.

Section 2.2 is devoted to lattices of semilattice decompositions of a semigroup, i.e. to lattices of semilattice congruence on a semigroup. We present characterizations of these lattices of T. Tamura [120], 1975, M. Ćirić and S. Bogdanović [23], and S. Bogdanović and M. Ćirić [12].

2.1. The greatest semilattice decomposition

As we noted above, the first general result concerning semilattice decompositions of semigroups is the one of T. Tamura and N. Kimura [122], 1954, which is given by the following theorem:

Theorem 2.1. (T. Tamura and N. Kimura [122]) *The poset of semilattice decompositions of any semigroup is a complete lattice.*

By the previous theorem it follows that any semigroup has a greatest semilattice decomposition. The first characterization of the greatest semilattice decomposition has been given by M. Yamada [132], 1955, in terms of P-subsemigroups. A subsemigroup T of a semigroup S is called a *P-semigroup* of S if for all $a_1, \ldots, a_n \in S$,

$$a_1 \cdots a_n \in T \quad \Rightarrow \quad C(a_1, \ldots, a_n) \subseteq T,$$

where $C(a_1, \ldots, a_n)$ denotes the subsemigroup of S consisting of all products of elements $a_1, \ldots, a_n \in S$ with each a_i appearing at least once [132]. Recall that $P(\mathcal{A})$ denotes the intersection of all principal congruences defined by elements of a nonempty family $\mathcal{A}$ of subsets of a semigroup.

Theorem 2.2. (M. Yamada [132]) *A relation θ on a semigroup S is a semilattice congruence if and only if $\theta = P(\mathcal{A})$, for some nonempty family $\mathcal{A}$ of P-subsemigroups of S.*

As a consequence of the previous theorem it can be deduced the following theorem:

Theorem 2.3. (M. Yamada [132]) *The smallest semilattice congruence on a semigroup S equals the congruence $P(\mathcal{X})$, where $\mathcal{X}$ denotes the set of all P-subsemigroups of S.*

Another approach to the greatest decompositions of semigroups, through completely prime ideals and filters, has been developed by M. Petrich [69], 1964. He proved the following four theorems:

Theorem 2.4. (M. Petrich [69]) *A relation θ on a semigroup S is a semilattice congruence if and only if $\theta = \Theta(\mathcal{A})$, for some nonempty family $\mathcal{A}$ of completely prime ideals of S.*

Theorem 2.5. (M. Petrich [69]) *The smallest semilattice congruence on a semigroup S equals the congruence $\Theta(\mathcal{X})$, where $\mathcal{X}$ denotes the set of all completely prime ideals of S.*

Theorem 2.6. (M. Petrich [69]) *A relation θ on a semigroup S is a semilattice congruence if and only if $\theta = \Theta(\mathcal{A})$, for some nonempty family $\mathcal{A}$ of filters of S.*

Theorem 2.7. (M. Petrich [69]) *The smallest semilattice congruence on a semigroup S equals the congruence $\Theta(\mathcal{X})$, where $\mathcal{X}$ denotes the set of all filters of S.*

Another proofs of the previous two theorems have been given by the authors in [21].

The role of completely prime ideals and filters in semilattice decompositions of semigroups can be explained by Theorem 1.7. Namely, the two-element chain Y_2 is, up to an isomorphism, the unique subdirectly irreducible semilattice, and any homomorphism of a semigroup S onto Y_2 determines a partition of S whose one component is a completely prime ideal and other is a filter of S. This approach has been used by M. Petrich in [72], 1973.

M. Petrich [69], 1964, also gave a method to construct the principal filters of a semigroup:

Theorem 2.8. (M. Petrich [69]) *The principal filter of a semigroup S generated by an element $a \in S$ can be computed using the following formulas:*

$$N_1(a) = \langle a \rangle\,, \quad N_{n+1}(a) = \langle \{x \in S \mid N_n(a) \cap J(x) \neq \varnothing\} \rangle\,, \ n \in \mathbb{Z}^+$$

$$N(a) = \bigcup_{n \in \mathbb{Z}^+} N_n(a).$$

The third approach to the greatest decompositions of semigroups is the one of T. Tamura from [117], 1972. Using the *division relation* $|$ on a semigroup S defined by:

$$a \mid b \quad \Leftrightarrow \quad b \in S^1 a S^1,$$

T. Tamura defined the relation $\longrightarrow$ on S by:

$$a \longrightarrow b \quad \Leftrightarrow \quad (\exists n \in \mathbf{Z}^+)\, a \mid b^n,$$

and he gave an efficient characterization of the smallest semilattice congruence on a semigroup:

Theorem 2.9. (T. Tamura [117]) *The smallest semilattice congruence on a semigroup S equals the natural equivalence of the relation $\longrightarrow^{\infty}$.*

Another proof of this theorem has been given by T. Tamura [118], 1973.

Three different characterizations of the smallest semilattice congruence on a semigroup have been also obtained by M. S. Putcha in [79], 1973, and [80], 1975.

Theorem 2.10. (M. S. Putcha [80]) *The smallest semilattice congruence on a semigroup S equals the equivalence on S generated by the relation $xy \equiv xyx \equiv yx$, for all $x, y \in S^1$.*

Another proof of this theorem has been given by T. Tamura [119], 1973.

Theorem 2.11. (M. S. Putcha [80]) *The smallest semilattice congruence on a semigroup S equals the relation $—^{\infty}$, where $— = \longrightarrow \cap \longrightarrow^{-1}$.*

Theorem 2.12. (M. S. Putcha [79]) *The smallest semilattice congruence on a semigroup S equals the relation θ on S defined by: $a\,\theta\,b$ if and only if for all $x, y \in S^1$ there exists a semilattice indecomposable subsemigroup T of S such that $xay, xby \in T$.*

An approach to semilattice decompositions of semigroups, different to the one of M. Petrich and T. Tamura, has been developed by M. Ćirić and S. Bogdanović in [21]. As we will see later, the results obtained there explained the connections between the above presented results of M. Petrich and T. Tamura. M. Ćirić and S. Bogdanović [21] started from the completely semiprime ideals and they first gave the following representations of the principal radicals of a semigroup:

Theorem 2.13. (M. Ćirić and S. Bogdanović [21]) *The principal radical of a semigroup S generated by an element $a \in S$ has the following representation:*

$$\Sigma(a) = \{x \in S \mid a \longrightarrow^{\infty} x\}.$$

Theorem 2.14. (M. Ćirić and S. Bogdanović [21]) *The principal radical of a semigroup S generated by an element $a \in S$ can be computed using the following formulas:*

$$\Sigma_1(a) = \sqrt{SaS},\ \Sigma_n(a) = \sqrt{S\Sigma_n(a)S},\ n \in \mathbb{Z}^+,\ \Sigma(a) = \bigcup_{n\in\mathbb{Z}^+} \Sigma_n(a).$$

Recall that $\mathcal{I}d^{\mathrm{cs}}(S)$ denotes the lattice of all completely semiprime ideals of a semigroup S. By means of Theorems 2.13 and 2.9, the authors in [21] obtained the following characterization of the smallest semilattice congruence on a semigroup:

Theorem 2.15. (M. Ćirić and S. Bogdanović [21]) *The smallest semilattice congruence on a semigroup S equals the equivalence $\Theta(\mathcal{I}d^{\mathrm{cs}}(S))$.*

A characterization of the greatest semilattice homomorphic image of a semigroup has been given by M. Ćirić and S. Bogdanović [21], through principal radicals of a semigroup:

Theorem 2.16. (M. Ćirić and S. Bogdanović [21]) *If a, b is any pair of elements of a semigroup S, then*

$$\Sigma(a) \cap \Sigma(b) = \Sigma(ab),$$

i.e. the set Σ_S of all principal radicals of S, partially ordered by inclusion, is a semilattice and it is the greatest semilattice homomorphic image of S.

As a consequence of the previous theorem, the authors in [21] proved the next theorem without use of the Zorn's lemma arguments:

The next theorem, which gives a connection between Theorems 2.15 and 2.5, has been proved by M. Petrich [72], 1973. Another proof of this theorem, without use of the Zorn's lemma arguments, has been given by the authors in [21], as a consequence of Theorem 2.16.

Theorem 2.17. (M. Petrich [72]) *Any completely semiprime ideal of a semigroup S is the intersection of some family of completely prime ideals of S.*

In other words, Theorem 2.17 says that the set of completely prime ideals of a semigroup S is meet-dense in $\mathcal{I}d^{\mathrm{cs}}(S)$.

Another consequence of Theorem 2.16 is the next theorem which gives a representation of the principal filters better than the one from Theorem 2.8.

Theorem 2.18. (M. Ćirić and S. Bogdanović [21]) *The principal filter of a semigroup S generated by an element a has the following representation:*

$$N(a) = \{x \in S \mid x \longrightarrow^{\infty} a\}.$$

The components of the greatest semilattice decomposition of a semigroup are characterized by the next theorem, which is clearly a consequence of Theorems 2.13, 2.18 and 2.9.

Theorem 2.19. (M. Petrich [72]) *The component of the greatest semilattice decomposition of a semigroup S containing an element a of S is precisely the subsemigroup $\Sigma(a) \cap N(a)$.*

The most significant theorem of the theory of semilattice decompositions of semigroup is probably the theorem of T. Tamura [110], 1956, on atomicity of semilattice decompositions of semigroups, given here as Theorem 2.20. Note that another proofs of this theorem have been given by T. Tamura in [112], 1964, by means of the concept of "contents", in [117], 1972, using the relation $\longrightarrow^{\infty}$, in [118], 1973, and [120], 1975, by M. Petrich [69], 1964, by means of completely prime ideals, and by M. S. Putcha [79], 1973, using the relation defined in Theorem 2.12 and the subsemigroups of the form $C(a_1, \ldots, a_n)$.

Theorem 2.20. (T. Tamura [110]) *Any component of the greatest semilattice decomposition of a semigroup is a semilattice indecomposable semigroup.*

Semilattice indecomposable semigroups have been described by T. Tamura [117] and M. Petrich [69], [72], by the following

Theorem 2.21. *The following conditions on a semigroup S are equivalent:*

(i) *S is semilattice indecomposable;*
(ii) $(\forall a, b \in S)\ a \longrightarrow^{\infty} b$;
(iii) *S has no proper completely semiprime ideals;*
(iv) *S has no proper completely prime ideals.*

The equivalence of conditions (i) and (ii) has been established by T. Tamura [117], 1972, (i) $\Leftrightarrow$ (iii) has been proved by M. Petrich [69], 1964, and (i) $\Leftrightarrow$ (iv) by M. Petrich [72], 1973.

Note that in the class of semilattice indecomposable semigroup the mostly investigated were Archimedean semigroups, defined by: $a \longrightarrow b$, for all elements a and b. Semilattices of such semigroups have been studied by many authors. The most important results from this area have been obtained by M. S. Putcha [79], 1973, T. Tamura [116], 1972, M. Ćirić and S. Bogdanović

[19], 1993, and [21], S. Bogdanović and M. Ćirić [6], 1992, and [14], and L. N. Shevrin [99] and [100], 1994. For more informations about semilattices of Archimedean semigroups the reader is also referred to the survey paper of S. Bogdanović and M. Ćirić [8], 1993, or their book [7], 1993.

2.2. The lattice of semilattice decompositions

T. Tamura [120] got an idea of studying semilattice decompositions of a semigroup through quasi-orders on this semigroup having some suitable properties. We say that a quasi-order ξ on a semigroup S is *positive* if $a\,\xi\,ab$ and $b\,\xi\,ab$, for all $a, b \in S$. These quasi-orders have been introduced by B. M. Schein [88], 1965, and they were since studied from different points of view by T. Tamura, M. S. Putcha, S. Bogdanović, M. Ćirić and other. By a *half-congruence* T. Tamura in [120], 1975, called a compatible quasi-order on a semigroup, and by a *lower-potent* quasi-order he called a quasi-order ξ on a semigroup satisfying the condition: $a^2\,\xi\,a$, for all elements a. Using these notions, T. Tamura proved the following theorem:

Theorem 2.22. (T. Tamura [120]) *The lattice of semilattice congruences on a semigroup S is isomorphic to the lattice of positive lower-potent half-congruences on S.*

As the authors noted in [23], the notion "lower-potent half-congruence" in Theorem 2.22 can be replaced by "quasi-order satisfying the *cm*-property". Recall from Section 1.1 that a relation ξ on a semigroup S satisfies the *common multiple property*, briefly the *cm-property*, if for all $a, b, c \in S$, $a\,\xi\,c$ and $b\,\xi\,c$ implies $ab\,\xi\,c$. Using this notion, introduced by T. Tamura in [116], 1972, Theorem 2.22 can be written as follows:

Theorem 2.23. *The lattice of semilattice congruences on a semigroup S is isomorphic to the lattice of positive quasi-orders on S satisfying the cm-property.*

Using the Tamura's approach, the authors in [23] connected semilattice decompositions of a semigroup with some sublattices of the lattice $\mathcal{I}d^{\mathrm{cs}}(S)$ of completely simple ideals of a semigroup. Recall from Section 1.1 that a subset K of a lattice L is *meet-dense* in L if any element of L can be written as the meet of some family of elements of K. We will say that a sublattice L of $\mathcal{I}d^{\mathrm{cs}}(S)$ satisfies the *completely prime ideal property*, shortly the *cpi-property*, if the set of completely prime ideals from L is meet-dense in L, i.e. if any element of L can be written as the intersection of some family of completely prime ideals from L. As we seen before, this property was proved for $\mathcal{I}d^{\mathrm{cs}}(S)$ by Theorem 2.17. M. Ćirić and S. Bogdanović [23] showed

that the *cpi*-property plays a crucial role in semilattice decompositions of semigroups:

Theorem 2.24. (M. Ćirić and S. Bogdanović [23]) *The lattice of semilattice decompositions of a semigroup S is isomorphic to the lattice of complete 1-sublattices of $\mathcal{I}d^{cs}(S)$ satisfying the cpi-property.*

Another connection of semilattice decompositions of a semigroup, with some sublattices of the lattice of subsets of a semigroup, has been established by S. Bogdanović and M. Ćirić in [12]. There they proved the following theorem:

Theorem 2.25. (S. Bogdanović and M. Ćirić [12]) *The lattice of semilattice decompositions of a semigroup S is isomorphic to the lattice of complete 1-sublattices of $\mathcal{P}(S)$ whose principal elements are filters of S.*

For more informations about the role of quasi-orders in semilattice decompositions of semigroups we refer to another survey paper of S. Bogdanović and M. Ćirić [16].

3. Band decompositions

Although the existence of the greatest band decomposition has been established by T. Tamura and N. Kimura in [123], 1955, by the theorem which is given here as Theorem 1.3, there are no sufficiently efficient characterizations of the greatest band decomposition of a semigroup in the general case. But, there are very nice descriptions of greatest decompositions for some special types of band decompositions, as semilattice decompositions, treated in the previous chapter, matrix decompositions, where left zero band and right zero band decompositions are included, and normal band decompositions, where left normal band and right normal band decompositions are included. This chapter is devoted to the results concerning the greatest matrix decomposition of a semigroup, which will be presented in Section 3.1, and to the results concerning the greatest normal band decomposition of a semigroup, which will be presented in Section 3.2.

Matrix decompositions, as well as left zero band and right zero band decompositions, have appeared first in studying of completely simple semigroups. Namely, by the famous Rees-Sushkevich theorem on matrix representations of completely simple semigroups, any completely simple semigroup can be decomposed into a matrix of groups, and also into a left zero band of right groups and into a right zero band of left groups. First general results concerning these decompositions have been obtained by P. Dubreil

[41], 1951, who constructed the smallest left zero band congruence on a semigroup, and by G. Thierrin [128], 1956, who characterized the components of the greatest left zero band decomposition of a semigroup. The general theory of matrix decompositions of semigroups has been founded by M. Petrich in [70], 1996. These results will be a topic of Section 3.1.

Normal bands have been introduced by M. Yamada and N. Kimura [133], 1958, whereas left normal bands have been first defined and studied by V. V. Vagner [129], 1962, and B. M. Schein [86], 1963, and [87], 1965. The general results concerning left normal band, right normal band and normal band decompositions of a semigroup, presented in Section 3.2, have been obtained by M. Petrich in [71], 1966.

For additional informations about matrix and normal band decompositions the reader is referred to the book of M. Petrich [73], 1977.

3.1. Matrix decompositions

As we noted before, the first general result concerning left zero band decompositions of a semigroup is the one of P. Dubreil [41], 1951. Define the relations $\overset{l}{\approx}$ and $\overset{r}{\approx}$ on a semigroup S by:

$$a \overset{l}{\approx} b \Leftrightarrow L(a) \cap L(b) \neq \varnothing, \qquad a \overset{r}{\approx} b \Leftrightarrow R(a) \cap R(b) \neq \varnothing, \quad (a, b \in S).$$

The relation $\overset{r}{\approx}$ has been introduced in above mentioned paper of P. Dubreil, where he proved the following theorem:

Theorem 3.1. (P. Dubreil [41]) *The smallest left zero band congruence on a semigroup S equals the relation $\overset{r}{\approx}{}^{\infty}$.*

The components of the greatest left zero band decomposition of a semigroup have been first described by G. Thierrin [128], 1956, by the following theorem:

Theorem 3.2. (G. Thierrin [128]) *The components of the greatest left zero band decomposition of a semigroup S are the minimal left consistent right ideals.*

Other characterizations of the greatest left zero band decomposition of a semigroup have been obtained by M. Petrich in [70], 1966. In this paper he proved the following two theorems:

Theorem 3.3. (M. Petrich [70]) *A relation θ on a semigroup S is a left zero band congruence on S if and only if $\theta = \Theta(\mathcal{A})$, for some nonempty family $\mathcal{A}$ of left consistent right ideals of S.*

Theorem 3.4. *The smallest left zero band congruence on a semigroup S equals the relation $\Theta(\mathcal{RI}d^{\mathrm{lc}}(S))$.*

The key theorem in theory of matrix decompositions of semigroups is the next theorem, proved by M. Petrich in [70], 1966, which gives a connection between left zero band, right zero band and matrix congruences on a semigroup:

Theorem 3.5. (M. Petrich [70]) *The intersection of a left zero band congruence and a right zero band congruence on a semigroup S is a matrix congruence on S.*

Conversely, any matrix congruence on S can be written uniquely as the intersection of a left zero band congruence and a right zero band congruence on S.

Combining Theorems 3.1 and 3.5, the following characterization of the smallest matrix congruence on a semigroup follows:

Theorem 3.6. (M. Petrich [70]) *The smallest matrix congruence on a semigroup S equals the relation $\overset{l}{\approx}{}^{\infty} \cap \overset{r}{\approx}{}^{\infty}$.*

Combining Theorem 3.3 and its dual, M. Petrich [70] obtained the following two theorems:

Theorem 3.7. (M. Petrich [70]) *A relation θ on a semigroup S is a matrix congruence on S if and only if $\theta = \Theta(\mathcal{A})$, for some nonempty subset $\mathcal{A}$ of $\mathcal{X}$, where $\mathcal{X} = \mathcal{LI}d^{\mathrm{rc}}(S) \cup \mathcal{RI}d^{\mathrm{lc}}(S)$.*

Theorem 3.8. (M. Petrich [70]) *The smallest matrix congruence on a semigroup S equals the relation $\Theta(\mathcal{X})$, where $\mathcal{X} = \mathcal{LI}d^{\mathrm{rc}}(S) \cup \mathcal{RI}d^{\mathrm{lc}}(S)$.*

M. Petrich in [70] also gave an alternative approach to matrix decompositions of semigroups, through so-called quasi-consistent subsemigroups. Namely, by a *quasi-consistent* subset of a semigroup S he defined a completely semiprime subset A of S satisfying the condition: for all $x, y, z \in S$, $xyz \in A$ if and only if $xy \in A$. Quasi-consistent subsemigroups of a semigroup M. Petrich connected with left consistent right ideals and right consistent left ideals by the following theorem:

Theorem 3.9. (M. Petrich [70]) *The intersection of a left consistent right ideal and a right consistent left ideal of a semigroup S is a quasi-consistent subsemigroup of S.*

Conversely, any quasi-consistent subsemigroup of S can be written uniquely as the intersection of a left consistent right ideal and a right consistent left ideal.

Using the previous theorem, matrix congruences on a semigroup can be characterized through quasi-consistent subsemigroups of a semigroup as follows:

Theorem 3.10. (M. Petrich [70]) *A relation θ on a semigroup S is a matrix congruence on S if and only if $\theta = \Theta(\mathcal{A})$, for some nonempty family $\mathcal{A}$ of the set of quasi-consistent subsemigroups of S.*

Theorem 3.11. (M. Petrich [70]) *The smallest matrix congruence on a semigroup S equals the relation $\Theta(\mathcal{X})$, where $\mathcal{X}$ denotes the set of all quasi-consistent subsemigroups of S.*

Using Theorem 3.5 and the fact that the join of any left zero band congruence and any right zero band congruence on a semigroup equals the universal congruence on this semigroup, the lattice of matrix congruences on a semigroup can be characterized in the following way:

Theorem 3.12. *The lattice of matrix congruences on a semigroup S is isomorphic to the direct product of the lattice of left zero band congruences and the lattice of right zero band congruences on S.*

A characterization of the lattice of right zero band decompositions of a semigroup can be obtained through left consistent right ideals of a semigroup, modifying the results of S. Bogdanović and M. Ćirić [13] to semigroups without zero. For related results concerning semigroups with zero we refer to Section 4.2.

Until the end of this section we will consider only semigroups without zero, because the definition of the lattice $\mathcal{RI}d(S)$ is different for semigroups with and without zero, and the set of right consistent left ideals of a semigroup with zero is one-element.

Theorem 3.13. *The poset $\mathcal{RI}d^{\mathrm{lc}}(S)$ of left consistent right ideals of a semigroup $S \neq S^0$ without zero is a complete atomic Boolean algebra and $\mathcal{RI}d^{\mathrm{lc}}(S) = \mathfrak{B}(\mathcal{RI}d(S))$.*

Theorem 3.14. *The lattice of left zero band decompositions of a semigroup $S \neq S^0$ is isomorphic to the lattice of complete Boolean subalgebras of $\mathcal{RI}d^{\mathrm{lc}}(S)$.*

The role of left zero band decompositions of a semigroup in direct decompositions of the lattice of right ideals of this semigroup is demonstrated by the following two theorems:

Theorem 3.15. *The lattice $\mathcal{RI}d(S)$ of right ideals of a semigroup $S \neq S^0$ is a direct product of lattices L_α, $\alpha \in Y$, if and only if S is a left zero band of semigroups S_α, $\alpha \in Y$, and $L_\alpha \cong \mathcal{RI}d(S_\alpha)$, for any $\alpha \in Y$.*

Theorem 3.16. *If S_α, $\alpha \in Y$, are components of the greatest left zero band decomposition of a semigroup $S \neq S^0$, then the lattice $\mathcal{RI}d(S)$ can be decomposed into a direct product of its intervals $[0, S_\alpha]$, $\alpha \in Y$, which are directly indecomposable.*

3.2. Normal band decompositions

In the introduction of Chapter 3 we said that the general theory of normal band decompositions of semigroups, including here left normal band and right normal band decompositions, has been founded by M. Petrich in [71], 1966. The methods used in this paper has been obtained by combination of the methods which M. Petrich used in [69], in studying of semilattice decompositions, and the ones used in [70], in studying of matrix decompositions.

M. Petrich in [71] defined a *left (right) normal complex* of a semigroup S as a nonempty subset A of S which is a left (right) consistent right (left) ideal of the smallest filter $N(A)$ of S containing A, and he defined a *normal complex* of S as a subset A of S which is a quasi-consistent subsemigroup of $N(A)$. He also introduced the following relations on a semigroup S: for a nonempty subset A of S, Φ_A is the equivalence relation on S whose classes are nonempty sets among the sets A, $N(A) - A$ and $S - N(A)$, and for a nonempty family $\mathcal{A}$ of subsets of S, $\Phi(\mathcal{A})$ is the equivalence relation on S defined by:

$$\Phi(\mathcal{A}) = \bigcap_{A \in \mathcal{A}} \Phi_A.$$

Theorem 3.17. (M. Petrich [71]) *A relation θ on a semigroup S is a left normal band congruence on S if and only if $\theta = \Phi(\mathcal{A})$, for some nonempty family $\mathcal{A}$ of left normal complexes of S.*

Theorem 3.18. (M. Petrich [71]) *The smallest left normal band congruence on a semigroup S equals the relation $\theta = \Phi(\mathcal{X})$, where $\mathcal{X}$ denotes the set of all left normal complexes of S.*

In order to study normal band congruences on a semigroup through left normal band congruences and right normal band congruences, M. Petrich proved the following theorem, similar to Theorem 3.5 concerning matrix congruences:

Theorem 3.19. (M. Petrich [71]) *The intersection of a left normal band congruence and a right normal band congruence on a semigroup S is a normal band congruence on S.*

Conversely, any normal band congruence on S can be written as the intersection of the smallest left normal band congruence and the smallest right normal band congruence on S containing it.

Let $\mathbf{X}^{(0,0)} \ni A^{-1}$ be an initial inclusion for A^{-1} and $\Phi(X, A)$ define a Schulz-type method of order p for A^{-1}. Then for fixed integers $k \geq 0$ and $p \geq 1$ we define:

$$
\begin{aligned}
(7.5) \qquad & X^{(n,0)} = m(\mathbf{X}^{(n,0)}), \\
& X^{(n,i)} = \Phi(X^{(n,i-1)}, A), \quad 1 \leq i \leq k,
\end{aligned}
$$

(empty statement in case $k = 0$)

$$
\mathbf{X}^{(n+1,0)} = X^{(n,k)} \sum_{i=0}^{r-1} (I - AX^{(n,k)})^i + \mathbf{X}^{(n,0)} (I - AX^{(n,k)})^r,
$$

(*Horner-scheme evaluation in interval arithmetic*), $n \geq 0$,

and

$$
\begin{aligned}
(7.6) \qquad & X^{(n,0)} = m(\mathbf{X}^{(n,0)}), \\
& X^{(n,i)} = \Phi(X^{(n,i-1)}, A), \quad 1 \leq i \leq k,
\end{aligned}
$$

(empty statement in case k=0)

$$
\mathbf{X}^{(n+1,0)} = \left\{ X^{(n,k)} \sum_{i=0}^{r-1} (I - AX^{(n,k)})^i + \mathbf{X}^{(n,0)} (I - AX^{(n,k)})^r \right\} \cap \mathbf{X}^{(n,0)},
$$

(*Horner-scheme evaluation in interval arithmetic*), $n \geq 0$,

where $m(\mathbf{X}) = (m(X_{ij}))$ is the midpoint matrix.

Remark 6. For $k = 0$ we get as special cases the methods (4.4) and (4.5) discussed in Section 4.

In particular, for the fixed $n = 0$ in (7.5) and (7.6), we obtain the combined methods

$$
(7.7) \qquad \begin{cases} X^{(i)} = \Phi(X^{(i-1)}, A), \quad 1 \leq i \leq k, \\ \mathbf{X}^{(1,k)} = X^{(k)} \sum_{i=0}^{r-1} (I - AX^{(k)})^i + \mathbf{X}^{(0)} (I - AX^{(k)})^r. \end{cases}
$$

and the monotonic version

$$(7.8)\quad \begin{cases} X^{(i)} & = \Phi(X^{(i-1)}, A), \quad 1 \le i \le k, \\ \mathbf{X}^{(1,k)} & = \left\{ X^{(k)} \sum_{i=0}^{r-1} (I - AX^{(k)})^i + \mathbf{X}^{(0)} (I - AX^{(k)})^r \right\} \bigcap \mathbf{X}^{(0)}. \end{cases}$$

The combined methods (7.7) and (7.8) are, therefore, performed applying k iterations in floating-point arithmetic in order to obtain sufficiently good approximation (point matrix) $X^{(k)}$ to the inverse matrix A^{-1} and then, in the final step, the inclusion method of the order r to provide the guaranteed error bounds to A^{-1}. Such a combination is of a great interest in practice and, for this reason, it was studied extensively in the papers [19] and [25].

For the combined methods (7.5) and (7.6) the following theorem has been proved in [19].

Theorem 11. *For the methods* (7.5) *and* (7.6) *the inclusion* $A^{-1} \in \mathbf{X}^{(n,0)}$ $(n \ge 0)$ *holds.*

Theorem 12. *The sequence* $\{\mathbf{X}^{(n,0)}\}$ *obtained by the method* (7.5) *converges to* A^{-1} *if and only if* $\rho(I - AX^{(0,0)}) < 1$.

As presented in Sections 4 and 6, the convergence criterion for monotonic methods like (7.6) for which

$$\mathbf{X}^{(0,0)} \supseteq \mathbf{X}^{(1,0)} \supseteq \cdots \supseteq \mathbf{X}^{(n,0)} \ni A^{-1}$$

obviously holds, differ from those of the non-monotonic methods like (7.5). This is contained in the convergence theorem, which is quite similar to Theorem 8.

Theorem 13. *The sequence* $\{\mathbf{X}^{(n,0)}\}$ *generated by the method* (7.6) *converges to* A^{-1} *if the inequality*

$$\| d(\mathbf{X}^{(0,0)}) \| < 2/\| A \|,$$

with a monotonic matrix norm $\| \cdot \|$, *is fulfilled.*

According to Traub [33, Appendix C] the *efficiency index* of an iterative method of order q can be defined by $q^{1/\Theta}$, where Θ is the total amount of work for one iteration step. In methods like (7.5) and (7.6) one usually measures Θ in terms of matrix multiplications and all other computational costs are considered to be negligible compared with these. If we count the computational efforts by Traub's formula we get the following results, assuming

that one interval matrix multiplication costs at least about two times as much as a point matrix multiplication:

ks multiplications for the application of the Schulz-type method where **s** is the number of multiplications for the evaluation of Φ;

r + 1 interval matrix multiplications for the Horner-scheme interval evaluation or approximately **2(r + 1)** point matrix multiplications.

This makes a total cost of **ks + 2(r + 1)** multiplications for one step of methods (7.5) or (7.6), reduced to point matrix multiplications. A Schulz-type method of the form (7.4) requires only ordinary floating point operations whereas the Horner-scheme interval evaluation has to be done completely by rounded interval operations to ensure the inclusion property of Theorem 11.

From Theorem 11 and Theorem 12 we get lower bounds for the order of convergence of our methods (7.5) and (7.6) as $q = rp^k + 1$ so that lower bounds for the efficiency index are given by

$$E(p, r, k) = (rp^k + 1)^{1/(ks+2(r+1))}.$$

Before determining parameters p and r in order to establish the optimal combined method concerning the computational efficiency expressed by the efficiency index $E(p, r, k)$, we recall that the most efficient method of Schulz's type in ordinary floating-point arithmetic reads

$$X^{(k+1)} = X^{(k)} \cdot \left(I + \frac{\sqrt{5}+1}{2}(I - AX^{(k)}) + (I - AX^{(k)})^2 \right) \times \left(I - \frac{\sqrt{5}+1}{2}(I - AX^{(k)}) + (I - AX^{(k)})^2 \right),$$

which is constructed using the mapping Φ_5 given in Example 5. Namely, the number of multiplication is $\mathbf{s} = 4$ for the evaluation of $\Phi_5(X, A)$ given by (7.3) and $\mathbf{s} = p$ for $\Phi_p(X, A)$ $(p \neq 5)$ given by (7.2) when Horner-scheme evaluation is applied.

In the sequel, speaking about the function Φ_5 (the case $p = 5$), we will assume the function defined by Ostrowski's identity (7.3), while in the remaining cases Φ_p $(p \neq 5)$ will denote the mapping (7.2). According to this, we define the total amount of work (expressed by point matrix multiplications) by

$$\Theta = \begin{cases} 4k + 2(r+1), & p = 5, \\ pk + 2(r+1), & p \neq 5 \end{cases}$$

(see [19]). Therefore, the lower bound of the efficiency index is given by

$$E(p,r,k) = \begin{cases} (r5^k+1)^{1/(4k+2r+2)}, & p=5, \\ (rp^k+1)^{1/(pk+2r+2)}, & p\neq 5. \end{cases} \tag{7.9}$$

The detailed procedure for finding optimal values of p and r (with respect to definition (7.9)) has been done by M. Petković and J. Herzberger in the paper [25]. This problem is of a great practical importance in applying the combined methods (7.5) and (7.6), and also (7.7) and (7.8). It leads to an optimization problem in the field of integers. First, the following theorem has been proved:

Theorem 14. *Let $r \in \{1,\dots,7\}$ and let $k \geq 1$ and p ($p \geq 2$ and $p \neq 5$) be arbitrary integers. Then*

$$E(5,r,k) > E(p,r,k). \tag{7.10}$$

As explained in [25], the restriction for r to be less than 8 is made for practical reasons. Namely, for a sufficiently great r (at least $r = 16$ but usually considerably greater, even more than 100) it is possible to find $p \geq 6$ and k such that the inequality (7.10) becomes converse. But, such values of p (at least 6 iterations in floating-point arithmetic) and r (at least 16 iterations in interval arithmetic) are meaningless in practice, especially in the situation when it is easy to provide initial matrices which insures the safe convergence.

The optimal choice of the number of point iterations r has been considered in the following theorem, assuming that $p = 5$.

Theorem 15. *The function $q(r) := (r5^k+1)^{1/(4k+2r+2)}$ attains its maximum on the interval $(1,2)$ for arbitrary $k \geq 1$.*

Using the result of Theorem 15 and the fact that r is an integer, we conclude that the optimal r in the combined methods can be either $r = 1$ or $r = 2$, depending on the number of iterative steps. A short analysis has shown that

$$E(5,2,k) > E(5,1,k) \quad \text{for } k = 1(1)6$$

and

$$E(5,2,k) < E(5,1,k) \quad \text{for } k \geq 7.$$

Thus, if the number of point iterative steps k is less than 7 then $r = 2$ is the optimal value, while for $k \geq 7$ the optimal value is $r = 1$. However, the second case ($k \geq 7$) is only of theoretical importance due to the very fast

convergence of the applied point method (of the order 5). For example, if $\| d(\mathbf{X}^{(0)}) \| = 0.8$, using the estimation

$$\| d(\mathbf{X}^{(1,k)}) \| \sim \| d(\mathbf{X}^{(0)}) \|^{5k+1}$$

for $k = 4$ we obtain even $\| d(\mathbf{X}^{(1,k)} \| \sim 10^{-61}$, which is an indicative illustration that the use of a (relatively) great number of iterative steps (say, $k > 3$) is not only meaningless but also not profitable (because of the limited precision of digital computers).

Finally, according to the previous results and discussion, in a practical realization of the combined method it should be chosen $p = 5$ and $r = 2$ (optimal for $k \leq 6$). Thus, the most efficient combined method of Schulz-type is of the form

$$(7.11)\quad X^{(n,i+1)} = X^{(n,i)} \cdot \left(I + \frac{\sqrt{5}+1}{2}(I - AX^{(n,i)}) + (I - AX^{(n,i)})^2 \right) \times$$
$$\times \left(I - \frac{\sqrt{5}+1}{2}(I - AX^{(n,i)}) + (I - AX^{(n,i)})^2 \right),$$
$$i = 0, 1, \ldots, k-1 \ (\textit{in floating point arithmetic})$$

$$(7.12)\quad \mathbf{X}^{(n+1,0)} = \left(\mathbf{X}^{(0)}(I - AX^{(n,k)}) + X^{(n,k)}\right) \cdot (I - AX^{(n,k)}) + X^{(n,k)}$$
$$(\textit{in interval arithmetic}),$$

where $X^{(n,0)} = m(\mathbf{X}^{(n,0)})$ and $n \geq 0$ and the starting matrix $\mathbf{X}^{(0,0)}$ includes A^{-1}.

The combined methods (7.11) and (7.12) have been considered in details in [19].

Example 6. To illustrate numerically the combined method (7.11) - (7.12), we present the example taken from the paper [19], where a 9×9 nonsingular matrix A with $A = I - B$, $\| B \| < 1$, was considered. Here $\| \cdot \|$ denotes the column-sum norm and the matrix $B = (b_{ij})$ is defined by

$$b_{ij} = \begin{cases} 0.1 & i \neq j \\ 0 & i = j \end{cases}, \quad (1 \leq i, j \leq 9).$$

A starting inclusion matrix $\mathbf{X}^{(0,0)}$ is constructed according to Theorem 9, that is, $\mathbf{X}^{(0,0)} = I + ([-c, c])$, where $c = \frac{\|B\|}{1-\|B\|}$. Evidently $m(\mathbf{X}^{(0,0)}) = I$ and the inclusion $A^{-1} \in \mathbf{X}^{(0,0)}$ holds.

For the method (7.11) - (7.12), referred to as the method (a), it was taken $k = 2$. The result of this combined method was compared to the classical optimal method, referred to as method (b), which can be defined as

$$\mathbf{Y}^{(0)} = \mathbf{X}^{(0,0)} \quad \text{and}$$
$$\mathbf{Y}^{(n+1)} = m(\mathbf{Y}^{(n)}) + (m(\mathbf{Y}^{(n)}) + \mathbf{Y}^{(n)}(I - Am(\mathbf{Y}^{(n)}))(I - Am(\mathbf{Y}^{(n)}))$$

for $n = 0, 1, \ldots$.

Let the inequality $\| d(\mathbf{X}) \| < \varepsilon = 5 \times 10^{-10}$ define the stopping criterion. The results obtained by the methods (a) and (b) are given in Table 1.

The total computational amount of work in terms of point matrix multiplications under the same assumption for interval matrix multiplications as above is as follows:

for the method (a): $2 \times (4 + 6) = 20$
for the method (b): $5 \times 6 = 30$.

It is clear that method (a) converges faster with the smaller computational efforts. Moreover, the computational efficiency of the method (a) is greater the greater is k.

n	$\| d(\mathbf{X}^{(n,0)}) \|$	$\| d(\mathbf{Y}^{(n)}) \|$
0	$7.20000000000 \times 10^{1}$	$7.20000000000 \times 10^{1}$
1	$7.73094113280 \times 10^{0}$	$4.60800000000 \times 10^{1}$
2	$1.96000000000 \times 10^{-10}$	$1.20795955200 \times 10^{1}$
3		$2.17606647543 \times 10^{-1}$
4		$1.27215720000 \times 10^{-6}$
5		$2.56000000000 \times 10^{-10}$

Table 1

8. Bounding the inverse of an interval matrix

Let $\mathbf{A} = (A_{ij})$ be an $n \times n$ interval matrix for which A^{-1} exists for every real matrix $A \in \mathbf{A}$ and denote $\mathbf{A}^i = \{A^{-1} | \ A \in \mathbf{A}\}$. In this section the problem of computing an interval matrix $\mathbf{X}$ with $\mathbf{A}^i \subseteq \mathbf{X}$ is considered. In many cases one can find an initial inclusion $\mathbf{X}^{(0)} \supseteq \mathbf{A}^i$, for example, by means of norm inequalities. But, in that case, the question arises how to improve $\mathbf{X}^{(0)}$ in such a way that its width $d(\mathbf{X}^{(0)}) = d((X_{ij}^{(0)})) = (d(X_{ij}^{(0)}))$ will be reduced. Theoretically, it is possible to find the interval hull of $\mathbf{A}^i$ in the form $\widehat{\mathbf{X}} = \cap\{\mathbf{X} | \ \mathbf{X} \supseteq \mathbf{A}^i\}$, but this, in general, cannot be done without

an unreasonable amount of work. For this reason, we are not dealing with this kind of problem and we are looking for an improvement $\mathbf{X}^*$ for $\mathbf{X}^{(0)}$ with $\mathbf{A}^i \subseteq \mathbf{X}^* \subseteq \mathbf{X}^{(0)}$ and $d(\mathbf{X}^*) \leq d(\mathbf{X}^{(0)})$ such that at least for a monotone matrix norm $\| \cdot \|$ the strict inequality

$$\| d(\mathbf{X}^*) \| < \| d(\mathbf{X}^{(0)}) \|$$

is valid. Schmidt found in [28] and [30] a monotone algorithm for the iterative improvement of $\mathbf{X}^{(0)}$. Alefeld and Herzberger suggested in [3, Ch. 18] (see, also, Section 4 of this paper) a somewhat different approach by means of interval analysis. The proposed method is closely related to the monotone version of the interval Schulz method for the iterative improvement of bounds for the inverse of a real nonsingular matrix A and it can be read in the form

$$\mathbf{X}^{(k+1)} = \{m(\mathbf{X}^{(k)}) + \mathbf{X}^{(k)}(I - Am(\mathbf{X}^{(k)}))\} \cap \mathbf{X}^{(k)}, \tag{8.1}$$

where $m(\mathbf{X})$ is the midpoint matrix of $\mathbf{X}$. A similar generalization with $\mathbf{A}_k$ and $\lim_{k\to\infty} \mathbf{A}_k = A$ instead of $\mathbf{A}$ was already used in Chapter 20 in [3] in connection with the Newton-method. In the case $\mathbf{A} = A$ one obtains the well-known interval Schulz-method. For iteration (8.1) we get immediately the following lemma:

Lemma 2. *For* $\mathbf{A}^i \subseteq \mathbf{X}^{(0)}$ *the sequence of matrices* $\{\mathbf{X}^{(k)}\}$ *produced by* (8.1) *has the property*

$$\mathbf{A}^i \subseteq \mathbf{X}^{(k)}, \quad (k = 0, 1, \dots) .$$

Proof. Since $\mathbf{A}^i \subseteq \mathbf{X}^{(0)}$, we choose $A^{-1} \in \mathbf{X}^{(0)}$ and by the use of the inclusion property of the interval operations we find

$$A^{-1} = m(\mathbf{X}^{(0)}) + A^{-1}(I - Am(\mathbf{X}^{(0)})) \in m(\mathbf{X}^{(0)}) + \mathbf{X}^{(0)}(I - Am(\mathbf{X}^{(0)})) \subseteq \mathbf{X}^{(1)}.$$

For $k > 1$ the proof can be done analogously. □

From (8.1) there follows

$$\mathbf{X}^{(0)} \supseteq \mathbf{X}^{(1)} \supseteq \mathbf{X}^{(2)} \supseteq \mathbf{X}^{(3)} \supseteq \dots$$

and thus

$$\lim_{k\to\infty} \mathbf{X}^{(k)} = \mathbf{X}^*$$

is valid. But the iteration process (8.1) could already fail with $\mathbf{X}^* = \mathbf{X}^{(0)}$ especially if $d(\mathbf{A})$ is of considerable size. In that case, instead of improving $X^{(0)}$, the process starts reproducing the same disk. Therefore, a convergence analysis for (8.1) which gives sufficient conditions for

$$\| d(\mathbf{X}^*) \| < \| d(\mathbf{X}^{(0)}) \|$$

has to be done in such a way that the method (8.1) yields an improved inclusion $\mathbf{X}^*$. For a given matrix $\mathbf{A}$ these sufficient conditions will impose some restrictions for $\| |\mathbf{X}^{(0)}| \|$ as well as for $\| d(\mathbf{X}^{(0)}) \|$ and so determine a class of matrices $\mathbf{X}^{(0)}$ with $\mathbf{A}^i \subseteq \mathbf{X}^{(0)}$ for which method (8.1) improves $\mathbf{X}^{(0)}$. The main result is the following theorem whose proof was given in [14].

Theorem 16. *Let $\mathbf{A}$ be given, then the iteration process* (8.1) *converges to $\mathbf{X}^*$ with $\| d(\mathbf{X}^*) \| < \| d(\mathbf{X}^{(0)}) \|$ if the matrix $\mathbf{X}^{(0)}$ with $\mathbf{A}^i \subseteq \mathbf{X}^{(0)}$ fulfills the inequalities*

$$\| d(\mathbf{A}) \| < \frac{3}{\| |\mathbf{X}^{(0)}| \| \, (8 \cdot \| m(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \| + \frac{3}{2}} \tag{8.2}$$

and

$$\frac{4 \cdot \| d(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \|^2}{2 - \| d(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \|} < \| d(\mathbf{X}^{(0)}) \| < \frac{16}{19} \cdot \frac{1}{\| m(\mathbf{A}) \|}. \tag{8.3}$$

In addition to this, for $\mathbf{X}^*$ the inequality

$$\| d(\mathbf{X}^*) \| \leq \frac{2 \cdot \| d(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \|^2}{1 - \frac{1}{2} \cdot \| d(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \|} \tag{8.4}$$

holds.

Remark 7. A sufficient condition for $\| |\mathbf{X}^{(0)}| \|$ in terms of $\| m(\mathbf{A}) \|$ and $\| d(\mathbf{A}) \| > 0$ such that (8.2) is fulfilled can easily be derived as

$$\| |\mathbf{X}^{(0)}| \| < \frac{2}{\frac{\| d(\mathbf{A}) \|}{2} + \sqrt{\frac{\| d(\mathbf{A}) \|^2}{4} + \frac{32}{3} \| m(\mathbf{A}) \| \cdot \| d(\mathbf{A}) \|}}.$$

Remark 8. From (8.4) it follows that $\| d(\mathbf{X}^*) \| \to 0$ as $\| d(\mathbf{A}) \| \to 0$. Thus, the estimation (8.4) claims that for $\mathbf{A} = A$ the interval Schulz-method converges to A^{-1}. This is the reason why (8.1) can be regarded as a generalization of the Schulz-method (5.1) in the case of an interval matrix $\mathbf{A}$.

Remark 9. The condition (8.3) is more restrictive than the corresponding result for the interval Schulz-method in the case $\mathbf{A} = A$ (see [8]) where the sufficient condition for the convergence

$$\| d(\mathbf{X}^{(0)}) \| < \frac{2}{\| A \|}$$

is proved. Here, $\mathbf{X}^{(0)}$ can contain singular matrices as examples show.

Remark 10. Condition (8.3) implies that every $X \in \mathbf{X}^{(0)}$ is nonsingular. This can be seen taking $X \in \mathbf{X}^{(0)}$. Then $(m(\mathbf{A}))^{-1} \in \mathbf{X}^{(0)}$ and we have

$$\| X - (m(\mathbf{A}))^{-1} \| = \| |X - (m(\mathbf{A}))^{-1}| \| \leq \| d(\mathbf{X}^{(0)}) \| < \frac{16}{19} \cdot \frac{1}{\| m(\mathbf{A}) \|}$$

$$< \frac{1}{\| m(\mathbf{A}) \|}.$$

According to [4, Theorem 4 in Section 4] it follows that

$$(m(\mathbf{A}))^{-1} + (X - (m(\mathbf{A}))^{-1}) = X$$

is nonsingular.

As it was shown in [14], the assumption on $\| |\mathbf{X}^{(0)}| \|$ can be weakened. But this requires more complicated form of the upper bound for $\| d(\mathbf{X}^{(0)}) \|$. Both is given in

Corollary of Theorem 16. *Let* $\mathbf{A}$ *be given. Then the iteration process* (8.1) *converges to* $\mathbf{X}^*$ *with* $\| d(\mathbf{X}^*) \| < \| d(\mathbf{X}^{(0)}) \|$ *if the matrix* $\mathbf{X}^{(0)}$ *with* $\mathbf{A}^i \subseteq \mathbf{X}^{(0)}$ *fulfills the inequalities*

$$\| d(\mathbf{A}) \| < \frac{1}{\| |\mathbf{X}^{(0)}| \| \cdot (2 \| m(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \| + 1}$$

and

$$\frac{4 \cdot \| d(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \|^2}{2 - \| d(\mathbf{A}) \| \cdot \| |\mathbf{X}^{(0)}| \|} < \| d(\mathbf{X}^{(0)}) \|$$

$$< \frac{1}{\| m(\mathbf{A}) \|} \left(1 - \frac{\| d(\mathbf{A}) \| \| |\mathbf{X}^{(0)}| \|}{2} \right).$$

In addition to this, the inequality (8.4) *holds.*

In practical computations the quantity $\| d(\mathbf{A}) \|$ is of small size. The interval matrix $\mathbf{A}$ appears, for instance, because of inaccurate initial data for a real matrix A or from conversion errors which are usually not too large. Therefore, the necessary initial inclusion $\mathbf{X}^{(0)}$ for $\mathbf{A}$ can often be calculated by an application of an interval Gaussian elimination or even by norm inequalities (see [14]).

References

[1] G. Alefeld, J. Herzberger, *Matrizeninvertierung mit Fehlereffassung Elektron*, Datenverarbeit **12** (1970), 410–416.

[2] G. Alefeld, J. Herzberger, *Einführung in die Intervallrechnung*, Bibliographisches Institut AG, Mannheim, 1974.

[3] G. Alefeld, J. Herzberger, *Introduction to interval computation*, Academic Press, New York, 1983.

[4] M. Altman, *An optimum cubically convergent iterative method of inverting a linear bounded operator in Hilbert space*, Pacific J. Math. **10** (1960), 1107–1113.

[5] P. Foster, *Bemerkungen zum Iterationsverfahren von Schulz zur Bestimmung der Inversen einer Matrix*, Numer. Math. **12** (1968), 211–214.

[6] E. Hansen, *Interval arithmetic in matrix computations. Part I*, J. SIAM Numer. Anal. Ser. B **2** (1965), 308–320.

[7] J. Herzberger, *On the monotonicity of the interval version of Schulz's method*, Computing **38** (1987), 71–74.

[8] J. Herzberger, *Remarks on the interval version of Schulz's method*, Computing **39** (1987), 183–186.

[9] J. Herzberger, *Zur Monotonie der intervallmäßigen Schulz-Verfahren höherer Ordnung*, Z. Angew. Math. Mech. **67** (1987), 137–138.

[10] J. Herzberger, *A class of optimal iterative methods of inverting a linear bounded operator*, Numer. Funct. Anal. and Optimiz. **9** (1987), 521–533.

[11] J. Herzberger, *Ein effizienter Algorithmus zur iterativen Einschließung der inversen Matrix*, Aplikace Matematiky **32** (1987), 271–275.

[12] J. Herzberger, *Monotone Einschließungsalgorithmen für die inverse Matrix mit Hilfe von PASCAL-SC*, Angew. Informatik **30** (1988), 207–212.

[13] J. Herzberger, *Iterationsverfahren höherer Ordnung zur Einschließung der Inversen einer Matrix*, Z. Angew. Math. Mech. **69** (1989), 115–120.

[14] J. Herzberger, *On the convergence of an iterative method for bounding the inverse of an interval matrix*, Computing **41** (1989), 153–162.

[15] J. Herzberger, *Uber die Wirksamkeit eines iterationsverfahrens zur Einschließ ung der inversen einer intervallmatrix*, Z. Angew. Math. Mech. **70** (1990), 470–472.

[16] J. Herzberger, *Using error-bounds for hyperpower methods to calculate inclusions for the inverse of a matrix.*, BIT **30** (1990), 508–515.

[17] J. Herzberger, D. Bethke, *On two algorithms for bounding the inverses of an interval matrix*, Interval Computations **1** (1991), 44–53.

[18] J. Herzberger, Lj. Petković, *On the construction of efficient interval Schulz's methods for bounding the inverse matrix*, Z. Angew. Math. Mech. **71** (1991), 411–412.

[19] J. Herzberger, Lj. Petković, *Efficient iterative algorithms for bounding the inverse of a matrix*, Computing **44** (1990), 237–244.

[20] L. V. Kantorovich, G. P. Akilov, *Functional Analysis*, Pergamon Press, Oxford, 1982.

[21] R. E. Moore, *Interval analysis*, Prentice Hall, New Jersey, 1966.

[22] A. M. Ostrowski, *Sur quelques transformations de la série de Liouville-Neumann*, C.R. Acad. Sci. Paris **206** (1938), 1345–1347.

[23] Lj. D. Petković, M. S. Petković, *On the monotonicity of the higher-order Schulz's method*, Z. Angew. Math. Mech. **68** (1988), 455–456.

[24] M. S. Petković, *Introduction to interval mathematics*, Naučna knjiga, Beograd, 1989. (In Serbian)

[25] M. S. Petković, J. Herzberger, *On the efficiency of a class of combined Schulz's methods for bounding the inverse matrix*, Z. Angew. Math. Mech. **71** (1991), 181–187.

[26] W. V. Petryshyn, *On the inversion of matrices and linear operators*, Proc. Amer. Math. Soc. **16** (1965), 893–901.

[27] W. V. Petryshyn, *On generalized inverses and on the converses of* $(I - \beta K)^n$ *with application to iterative methods*, J. Math. Anal. Appl. **18** (1967), 417–439.

[28] J. W. Schmidt, *Einschließung inverser Elemente durch Fixpunktverfahren*, Numer. Math. **31** (1978), 313–320.

[29] J. W. Schmidt, *Monotone Eingrezung von inversen Elementen durch ein quadratisch konvergentes Verfahren ohne Durchschnittsbildung*, Z. Angew. Math. Mech. **60** (1980), 202–204.

[30] J. W. Schmidt, *Two-sided approximations of inverses, square-roots and Cholesky factors*, Computational Mathematics. Banach Center Publications **13** (1984), 483–497.

[31] G. Schulz, *Iterative Berechnung der reziproken Matrix*, Z. Angew. Math. Mech. **13** (1933), 57–59.

[32] E. Stickel, *On a class of high order methods for inverting matrices*, Z. Angew. Math. Mech. **67** (1987), 334–336.

[33] J. F. Traub, *Iterative methods for the solutions of equations*, Prentice-Hall, New Jersey, 1964.

Faculty of Electronic Engineering, Department of Mathematics, P.O. Box 73, 18000 Niš, Yugoslavia