Consequences of a properly implemented computer arithmetic for periodicities of iterative methods

R. Klatte and Ch. Ullrich

Abstract - In ordered sets it is possible to show under certain assumptions two basic theorems concerning the cycle length of sequences of iterates generated by monotone operators. These results are applied to different iterative methods, where the conclusions are valid for the sequences of iterates produced by the numerical computations only, if the used computer arithmetic is properly implemented.

Index terms - Rounding invariant structures, cycle, weakly cyclic vector function, floating-point arithmetic.

1. Introduction and Summary

Because of the technical conditions the set R of the numbers representable in a computer is finite. If we execute an algorithm of the form $x_{n+1} = Fx_n$, $n = o,1,2,\ldots$, beginning with any element $x_o \in R$, it is clear, that the sequence $\{x_n\}$ of the iterates has only a finite number of different elements. Moreover, from an element x_k the values of the m preceding iterates $x_{k-m}, x_{k-m+1}, \ldots, x_{k-1}$ return in the same order, i.e. the algorithm is ending in a cycle of the length m. After receiving a cycle it is useless to continue the iteration because we cannot obtain any further information. Therefore we can regard the end of the first cycle as the end of the numerical computation which easily can be recognized, if the cycle length is known.

For the studies of cycles it is usefull to know the mathematical structure of the arithmetic in R. Those algebraic and order properties, which should hold for any floating-point system, namely the usual rules for the neutral elements $0,1$ and the minus sign as well as compatibility properties between the algebraic and the order structure well-known from the real number field $\mathbb{R}$, have been summarized to the concept of the completely linearly ordered division ringoid $\{R,N,+,\cdot,/,\leq\}$ ([3]). In [5] it could be shown that R with a properly implemented floating-point arithmetic represents such a structure. In a similar way it is possible to describe computations with vectors $x \in V_nR$ and $n \times n$ matrices $A \in M_nR$ as well as the matrix vector multiplication $\cdot : M_nR \times V_nR \to V_nR$ by the structure of the ordered vectoid over an ordered division ringoid ([6]). The intent of this paper is to show the possibility of determining or estimating the length of cycles of iterative methods only by means of the algebraic and order properties of the above structures.

The following results are essentially based on two simple theorems for monotone operators in an ordered set $\{M, \leq\}$: an isotone operator always generates a sequence of iterates with the cycle length 1, if two elements x_{i_o}, x_{i_o+1} of the sequence are comparable with respect to $\leq$ and an antitone operator produces at most cycles of length 2, if three consecutive iterates are comparable.

It will be shown that for all non-negative resp. non-positive elements $A \in M_nR$ and for all $b \in V_nR$ the operator $T : V_nR \to V_nR$ with $Tx := Ax + b$ is an isotone resp. antitone operator. Therefore we can apply these theorems to T immediately. Without the assumption $A \geq O$ resp. $A \leq O$, it is nevertheless possible to get results for special matrices, which follow from general studies of weakly cyclic vector functions in V_nR ([2]). So, for example we are able to determine the cycle length of iteration sequences produced analogously to the Jacobi resp. Gauss-Seidel method , if A is a weakly cyclic matrix ([7]) or a generalization of it. All results are illustrated by numerical examples.

2. Rounding Invariant Structures

In recent investigations it was shown that the algebraic and order properties of the mathematical space given on a computer allow to be described by the structure of the completely linearly ordered division ringoid ([3]). We shortly repeat this definition:

Definition 1: A non empty ordered set $\{R, \leq\}$ with two binary operations $+,\cdot$ is called an "ordered ringoid" $\{R,+,\cdot,\leq\}$, if the properties (D1) to (D6) and (OD1) to (OD3) hold:

(D1) $\bigwedge_{a,b \in R} a + b = b + a$.

(D2) $\bigvee_{o \in R} \quad \bigwedge_{a \in R} a + o = a$.

(D3) $\bigvee_{e \in R} \quad \bigwedge_{a \in R} a \cdot e = e \cdot a = a$.

(D4) $\bigwedge_{a \in R} a \cdot o = o \cdot a = o$.

(D5) There exists a uniquely defined element $x \neq e$

of R such that with $-a := x \cdot a$ hold :

(a) $x \cdot x = e$,

(b) $\bigwedge_{a,b \in R} -(a \cdot b) = (-a) \cdot b = a \cdot (-b)$,

(c) $\bigwedge_{a,b \in R} -(a + b) = (-a) + (-b)$.

(OD1) $\bigwedge_{a,b,c \in R} (a \leqq b \Rightarrow a + c \leqq b + c)$.

(OD2) $\bigwedge_{a,b \in R} (a \leqq b \Rightarrow -b \leqq -a)$.

(OD3) $\bigwedge_{a,b,c \in R} (o \leqq a \leqq b \wedge c \geqq o \Rightarrow a \cdot c \leqq b \cdot c \wedge c \cdot a \leqq c \cdot b)$.

The elements $-e, o, e \in R$ are called the special elements of $\{R,+,\cdot,\leqq\}$. An ordered ringoid $\{R,+,\cdot,\leqq\}$ with a further inner operation $/ : R \times R \setminus N \Rightarrow R$ with $o \in N \subseteq R$ is called an "ordered division ringoid" $\{R,N,+,\cdot,/,\leqq\}$, if the properties (D6) to (D8) and (OD4) are fulfilled:

(D6) $\bigwedge_{a \in R} a / e = a$.

(D7) $\bigwedge_{a \in R \setminus N} o / a = o$.

(D8) $\bigwedge_{a \in R} \bigwedge_{b \in R \setminus N} -(a / b) = (-a)/b = a / (-b)$.

(OD4) $\bigwedge_{a,b,c \in R}$ $(o <$[1] $a \leqq b \wedge c > o \Rightarrow o \leqq a/c \leqq b/c \wedge o \leqq c/b \leqq c/a)$.

An ordered ringoid $\{R,+,\cdot,\leqq\}$ (resp. an ordered division ringoid $\{R,N,+,\cdot,/,\leqq\}$) is called a "completely ordered ringoid" (resp. a "completely ordered division ringoid"), if $\{R, \leqq\}$ is a complete lattice.

An ordered ringoid $\{R,+,\cdot,\leqq\}$ (resp. ordered division ringoid $\{R,\{o\},+,\cdot,/,\leqq\}$) is called linearly ordered, if $\{R, \leqq\}$ is a linearly ordered set (resp. and

(OD5) $\bigwedge_{a \in R \setminus \{o\}} a / a = e$

holds). □

Let M_nR be the set of $n \times n$ - matrices with elements of the ordered ringoid $\{R,+,\cdot,\leqq\}$. With the usual addition and multiplication for matrices $A = (a_{ij})$, $B = (b_{ij}) \in M_nR$

$$A + B := (a_{ij} + b_{ij})$$

$$A \cdot B := (\sum_{j=1}^{n} a_{ij} \cdot b_{jk})\ ^{2)}$$

and the order relation

$$A \leqq B :\Leftrightarrow (a_{ij} \leqq b_{ij} \quad \text{for} \quad i,j = 1(1)n) ,$$

$\{M_nR,+,\cdot,\leqq\}$ is an ordered ringoid with the special elements

$$-E = \begin{bmatrix} -e & & 0 \\ & -e & \\ 0 & & \ddots \; -e \end{bmatrix}, O = \begin{bmatrix} o & \ldots\ldots & o \\ o & \ldots\ldots & o \\ o & \ldots\ldots & o \end{bmatrix}, E = \begin{bmatrix} e & & 0 \\ & e & \\ 0 & & \ddots \; e \end{bmatrix} .$$

Let further V_nR be the set of n-tuples over an ordered ringoid $\{R,+,\cdot,\leqq\}$ with the usual inner operation $+$

$$a + b := (a_i + b_i),\ a = (a_i) \text{ and } b = (b_i), i=1(1)n$$

and the order relation

$$a \leqq b :\Leftrightarrow (a_i \leqq b_i \quad \text{for} \quad i = 1(1)n) .$$

Additionally we can define an outer operation $\cdot$: $M_nR \times V_nR \Rightarrow V_nR$ by

$$A \cdot b := (\sum_{j=1}^{n} a_{ij} \cdot b_j),\ A = (a_{ij}),\ b = (b_j) ,\quad i,j=1(1)n .$$

To be able to describe the properties of these operations we need the following

Definition 2: Let $\{R,+,\cdot,\leqq\}$ be an ordered ringoid with the special elements $\{-e,o,e\}$ and $\{V, \leqq\}$ an ordered set with a commutative binary operation $+$ and the neutral element o . V is called an "ordered vectoid over R " (or an "ordered R-vectoid"), if additionally an outer operation $\cdot : R \times V \Rightarrow V$ is defined so that the following properties hold:

(VD1) $\bigwedge_{a \in R} \bigwedge_{a \in V} (a \cdot o = o \wedge o \cdot a = o)$.

(VD2) $\bigwedge_{a \in V} e \cdot a = a$.

(VD3) $\bigwedge_{a \in R} \bigwedge_{a \in V} (-e) \cdot (a \cdot a) = ((-e) \cdot a) \cdot a = a \cdot ((-e) \cdot a)$.

(VD4) $\bigwedge_{a,b \in V} (-e) \cdot (a + b) = (-e) \cdot a + (-e) \cdot b$.

(OV1) $\bigwedge_{a,b,c \in V} (a \leqq b \Rightarrow a + c \leqq b + c)$.

[1] The relation $<$ is defined by $a < b :\Leftrightarrow (a \leqq b \wedge a \neq b)$.

[2] We define $\sum_{i=1}^{n} a_i := (\ldots((a_1 + a_2) + a_3)\ldots) + a_n$ and determine the multiplication to be carried out before the addition. Furthermore we shall use $a \cdot b$ and ab equivalently.

$$(OV2)\quad \bigwedge_{a,b\in V} (a \leqq b \Longrightarrow (-e)\cdot b \leqq (-e)\cdot a)\ .$$

$$(OV3)\quad \bigwedge_{a\in R}\ \bigwedge_{a,b\in V} (o \leqq a \wedge o \leqq a \leqq b \Rightarrow a\cdot a \leqq a\cdot b).$$

□

Theorem 1: With the above defined order relation $\leqq$ and the described operations, V_nR is an ordered M_nR-vectoid $\{V_nR, M_nR, \leqq\}$ and therefore also a R-vectoid.

Proof: Obviously V_nR fulfills the assumptions with the element $o = (o_i)$, $o_i = o$ for $i=1(1)n$. The properties (VD1) and (VD2) follow immediately from the definition of the outer operation and from (D3), (D4). Let now be $A = (a_{ij}) \in M_nR$, $a = (a_i)$, $b = (b_i)$ and $c = (c_i) \in V_nR$. Then we get:

$$(VD3): (-E)\cdot(A\cdot a) =$$

$$= \begin{pmatrix} -e & & 0 \\ & \ddots & \\ 0 & & -e \end{pmatrix} \cdot \begin{pmatrix} \sum_{j=1}^{n} a_{1j}\cdot a_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj}\cdot a_j \end{pmatrix} = \begin{pmatrix} (-e)\cdot \sum_{j=1}^{n} a_{1j}\cdot a_j \\ \vdots \\ (-e)\cdot \sum_{j=1}^{n} a_{nj}\cdot a_j \end{pmatrix} =$$

$$\underset{(D5c)}{=} \begin{pmatrix} \sum_{j=1}^{n} (-e)\cdot(a_{1j}\cdot a_j) \\ \vdots \\ \sum_{j=1}^{n} (-e)\cdot(a_{nj}\cdot a_j) \end{pmatrix} \underset{(D5b)}{=} \begin{pmatrix} \sum_{j=1}^{n} ((-e)\cdot a_{1j})\cdot a_j \\ \vdots \\ \sum_{j=1}^{n} ((-e)\cdot a_{nj})\cdot a_j \end{pmatrix} =$$

$$\underset{(D5b)}{=} \begin{pmatrix} \sum_{j=1}^{n} a_{1j}\cdot((-e)\cdot a_j) \\ \vdots \\ \sum_{j=1}^{n} a_{nj}\cdot((-e)\cdot a_j) \end{pmatrix} = ((-E)\cdot A)\cdot a = A\cdot((-E)\cdot a).$$

$$(VD4): (-E)\cdot(a + b) = ((-e)\cdot(a_i + b_i)) \underset{(D5c)}{=}$$

$$= ((-e)\cdot a_i + (-e)\cdot b_i) = (-E)\cdot a + (-E)\cdot b\ .$$

$$(OV1): a \leqq b \Rightarrow \bigwedge_{i=1(1)n} a_i \leqq b_i \underset{(OD1)}{\Longrightarrow}$$

$$\bigwedge_{i=1(1)n} a_i + c_i \leqq b_i + c_i \Rightarrow a + c \leqq b + c\ .$$

$$(OV2): a \leqq b \Rightarrow \bigwedge_{i=1(1)n} a_i \leqq b_i \underset{(OD2)}{\Longrightarrow}$$

$$\bigwedge_{i=1(1)n} (-e)\cdot b_i \leqq (-e)\cdot a_i \Rightarrow (-E)\cdot b \leqq (-E)\cdot a\ .$$

$$(OV3): (O \leqq A \wedge o \leqq a \leqq b) \Longrightarrow \bigwedge_{i,j=1(1)n} (o \leqq a_{ij} \wedge$$

$$\wedge\, o \leqq a_j \leqq b_j) \underset{(OD3)}{\Longrightarrow} \bigwedge_{i,j=1(1)n} a_{ij}\cdot a_j \leqq a_{ij}\cdot b_j \underset{(OD1)}{\Longrightarrow}$$

$$\Longrightarrow \bigwedge_{i=1(1)n} \sum_{j=1}^{n} a_{ij}\cdot a_j \leqq \sum_{j=1}^{n} a_{ij}\cdot b_j) \Longrightarrow$$

$$\Longrightarrow A\cdot a \leqq A\cdot b\ .\quad □$$

In an ordered set $\{M, \leqq\}$ an isotone (resp. antitone) operator $T : M \to M$ is defined by the property·

$$\bigwedge_{a,b\in M} (a \leqq b \Longrightarrow Ta \leqq Tb)$$

(resp. $\bigwedge_{a,b\in M} (a \leqq b \Longrightarrow Ta \geqq Tb)$. As an immediate consequence of Theorem 1 and the property

$$\bigwedge_{a,b,c\in R} (a \leqq b \wedge c \geqq o \Rightarrow a\cdot c \leqq b\cdot c \wedge c\cdot a \leqq c\cdot b),$$

which holds for linearly ordered ringoids (see [3]), we get the

Corollary: Let $\{V_nR, M_nR, \leqq\}$ be the ordered M_nR-vectoid over the linearly ordered ringoid $\{R,+,\cdot,\leqq\}$ and the operator $T: V_nR \to V_nR$ defined by

$$Tx := A\cdot x + b,\ A \in M_nR,\quad b \in V_nR\ ,$$

with $A \geqq O$ (resp. $A \leqq O$). Then T is an isotone (resp. antitone) operator. □

Furthermore, for isotone (resp. antitone) matrix operators we next show the following

Theorem 2: Let $\{R,+,\cdot,\leqq\}$ be a linearly ordered ringoid with the special elements $\{-e,o,e\}$. Then, an element $A \in M_nR$ with $A : V_nR \to V_nR$ is isotone (resp. antitone), iff $A \geqq O$ (resp. $A \leqq O$).

Proof: The first part of Theorem 2 is the statement of the above Corollary, if we choose $b = o$. Conversely, we assume, that $A = (a_{ij})$ is not non-negative. Then there exists at least one negative entry a_{kl} of A . Now we choose the vectors $a = o$ and $b = (b_i)$ with $b_l = e$ and $b_i = o$ for $i \neq l$ and obtain $A\cdot a = o$, $(A\cdot b)_k = a_{kl}\cdot e = a_{kl} < o$, i.e. $A\cdot a \nleqq A\cdot b$ with $a < b$. Therefore A ist not isotone, which completes the proof. □

3. Cycles in ordered sets

The numerical computation of an iterative method produces a sequence,in which beginning with a certain index a finite number of different elements repeat periodically. We precise this fact in the following

Definition 3: Let M be a non empty set. A sequence $\{x_n\}$ in M is called "cyclically ending", if the property (Z) holds:

$$\text{(Z)} \quad \bigvee_{\mu,n_o \in \mathbb{N}} \; \bigwedge_{n \in \mathbb{N}} (n \geqq n_o \Rightarrow x_{n+\mu} = x_n) \; .$$

The number $m := \min\{\mu \in \mathbb{N} \mid (Z)\}$ is called the "length of the cycle" and the set $\{x_{n_o}, x_{n_o+1}, \dots, x_{n_o+m-1}\}$ is called "cycle of length m" (see [1]).
An iterative method

$$\text{(IT)} \quad x_{n+1} = \phi(x_n) \; , \; x_o \in M \; , \quad n = o,1,2,\dots \; ,$$

generated by an operator $\phi : M \to M$ is called "cyclically ending in the cycle $Z(x_o)$", if the sequence $\{x_n\}$ is cyclically ending. □

In ordered sets we get for a sequence produced by an iterative method the following for the further investigations fundamental result:

Theorem 3: Let $\{M,\leqq\}$ be an ordered set and $\phi : M \to M$ an isotone operator. The sequence $\{x_n\}$ of iterates produced by ϕ be cyclically ending in a cycle of length m. Then,

$$\bigwedge_{i,j \in \mathbb{N}} (x_i \leqq x_j \Rightarrow m \,|\, |i-j|) \; .$$

Proof: Let $\{z_1, z_2, \dots, z_s\}$ be the cycle produced by (IT). First we show,that all cycle elements are incomparable:
Assume $z_r < z_{r+d}$ with $d > o$. Since ϕ is isotone, we have $z_{r+d} = \phi^d z_r < \phi^{2d} z_r < \dots < \phi^{sd} z_r = z_r$, i.e. $z_r < z_r$, a contradiction.

If the sequence $\{x_n\}$ ends cyclically,there exists a $r \in \mathbb{N}$ with $\phi^r(x_i), \phi^r(x_j) \in Z(x_o)$. The assumption $x_i \leqq x_j$ leads to $\phi^r x_i \leqq \phi^r x_j$ and therefore to $\phi^r x_i = \phi^r x_j$. Moreover, we get with $l := j - i$

$$z = \phi^r x_i = \phi^r(x_{i+l}) = \phi^{r+l}(x_i) = \phi^l(\phi^r x_i) = \phi^l z \; ,$$

which completes the proof. □

The assumption of Theorem 3 the sequence of iterates being cyclically ending means no restriction for our purposes, since the set of numbers representable on a computer is finite.

Concerning Theorem 3 we still give two Remarks: 1. If two consecutive iterates are comparable, we get a cycle of length 1. Especially, in linearly ordered sets we always obtain this result.

2. $\bigwedge_{y \in M} (x_i \leqq y \leqq x_j \Rightarrow Z(y) = Z(x_o))$.

For antitone operators we derive from Theorem 3 the following

Corollary: Let $\{M, \leqq\}$ be an ordered set and $\phi : M \to M$ an antitone operator. The sequence $\{x_n\}$ of iterates produced by ϕ be cyclically ending in a cycle of length m . Then we have

$$\bigwedge_{i,j \in \mathbb{N}} ((x_i \leqq x_j \wedge 2 \,|\, |i-j|) \Rightarrow m \,|\, |i-j|) \; .$$

The assumption of comparability of iterates with even-numbered difference of indices is fulfilled especially for generalized alternating sequences, i.e. for sequences with the property

$$\bigwedge_{i \in \mathbb{N}} \min(x_i, x_{i+d}) \leqq x_{i+2d} \leqq \max(x_i, x_{i+d})$$

resp.

$$\bigwedge_{i \in \mathbb{N}} \min(x_{i+d}, x_{i+2d}) \leqq x_i \leqq \max(x_{i+d}, x_{i+2d})$$

with $d \in \mathbb{N}$.

4. Applications to iterative methods

First we consider iterative methods, where Theorem 3 and the corresponding corollary lead immediately to results concerning cycles. So, let be $\{R,+,\cdot,\leqq\}$ a linearly ordered ringoid with the special elements $\{-e,o,e\}$. We express the element $A = (a_{ij}) \in M_n R$ as the matrix sum $A = L + D + R$ with

$$L = \begin{pmatrix} o & \cdots & & o \\ a_{21} & o & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & a_{n,n-1} & o \end{pmatrix} , \quad D = \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix} ,$$

$$R = \begin{pmatrix} o & a_{12} & \cdots & a_{1n} \\ \vdots & & \ddots & \vdots \\ \vdots & & & a_{n-1,n} \\ o & \cdots & \cdots & o \end{pmatrix} \in M_n R \; .$$

Now we can formulate the following well-known iterative methods in $V_n R$ with arbitrary elements $x^{(o)}, \; b \in V_n R$:

(a) $x^{(k+1)} = Ax^{(k)} + b$,   total-step method

(b) $x^{(k+1)} = \mathcal{L}x^{(k+1)} + \{(\mathcal{D} + \mathcal{R})x^{(k)} + b\}$, single-step method

(c) $x^{(k+1)} = (\omega\mathcal{L})x^{(k+1)} + \{[(e - \omega)\mathcal{E} + \omega(\mathcal{D} + \mathcal{R})]x^{(k)} + {}$
$+ \omega \cdot b)\}$, $\omega \in R$, $k=o,1,2,\dots$. successive relaxation method

By reason of monotony properties of the operators $A,\mathcal{L},\mathcal{D},\mathcal{R}$, we are able to verify the isotony resp. antitony of the operators defined by (a), (b) and (c):

Theorem 4: Let $\{R,+,\cdot,\leqq\}$ be a linearly ordered ringoid with the special elements $\{-e,o,e\}$. Then,

1. If $A : V_nR \Rightarrow V_nR$ is an isotone resp. antitone matrix, then the operator defined by (a) is also isotone resp. antitone.
2. If $\mathcal{L}$ is an isotone and $\mathcal{D} + \mathcal{R}$ an isotone (resp. antitone) matrix, then the operator defined by (b) is isotone (resp. antitone).
3. If $\mathcal{L}$ is an isotone and $\mathcal{D} + \mathcal{R}$ an isotone (resp. antitone) matrix, then the operator defined by (c) is isotone for $o < \omega \leqq e$ (resp. antitone for $\omega \geqq e$).

Proof: 1. This is the statement of the corollary of Theorem 2.

2. With $\mathcal{L}$ and $\mathcal{D} + \mathcal{R}$ isotone follows for the operator $Tx = \mathcal{L}(Tx) + (\mathcal{D} + \mathcal{R})x + b$:

(*) $x \leqq y \Longrightarrow (\mathcal{D} + \mathcal{R})x + b \leqq (\mathcal{D} + \mathcal{R})y + b$ and therefore $(Tx)_1 \leqq (Ty)_1$. Now suppose

(**) $\bigwedge_{i=1(1)k<n} (Tx)_i \leqq (Ty)_i$.

Then follows from (**) and $\mathcal{L}$ isotone :

$(\mathcal{L}(Tx))_{k+1} \leqq (\mathcal{L}(Ty))_{k+1}$

and with (*)

$(Tx)_{k+1} \leqq (Ty)_{k+1}$, i.e. T is isotone.

The proof of the second assertion turns out analogously.

3. In a linearly ordered ringoid we have the property $\bigwedge_{a \in R} a - a = o$. Hence, we get

$$o < \omega \leqq e \underset{(OD2)}{\Longrightarrow} -e \leqq -\omega \leqq o \underset{(OD1)}{\Longrightarrow} o = e - e \leqq e - \omega \leqq e.$$

Now it is clear, that $\omega\mathcal{L}$, $\omega(\mathcal{D} + \mathcal{R})$, $(e - \omega)\mathcal{E}$, $(e - \omega)\mathcal{E} + \omega(\mathcal{D} + \mathcal{R})$ are isotone operators, if $\mathcal{L}$ and $\mathcal{D} + \mathcal{R}$ are isotone, and therefore 3. is reduced to 2.

Similarly, we verify the second assertion by applying

$$e \leqq \omega \underset{(OD2)}{\Longrightarrow} -\omega \leqq -e \underset{(OD1)}{\Longrightarrow} e - \omega \leqq e - e = o . \quad \square$$

The operators considered in Theorem 4 fulfill the assumptions of monotony in Theorem 3 and the corresponding corollary, i.e. we are able to determine the cycle length of the iteration sequences generated by the methods (a), (b) and (c), if comparable iterates occur. Consequently, the problem consists in starting the iteration with an element $x^{(o)}$, such that comparable iterates are produced with indices as small as possible. Because of this difficulty we will consider the question, whether the assumption of the comparability can be avoided for certain operators. To this, we generalize the concept of the weakly cyclic matrices of index k to vector functions by

Definition 4: Let $\{R,+,\cdot,\leqq\}$ be a linearly ordered ringoid. A vector function $F = (f_i): B \subseteq V_nR \Rightarrow V_nR$ is called "weakly cyclic of index k" , if the set $N := \{1,2,\dots,n\}$ is partitioned into k disjoint subsets $N_1,\dots,N_k$ with $\bigcup_{i=1}^{k} N_i = N$ and the following property is valid:

$$\bigwedge_{i=1(1)k} \; \bigwedge_{l \in N_i} f_l = f_l(x_{N_{P(i)}}) \text{ with}$$

$x_{N_i} := \{x_j | j \in N_i\}$, where P is a cyclic permutation of $\{1,2,\dots,k\}$. If P is a permutation of $\{1,2,\dots,k\}$ partitioned into r cyclic permutations of k_i, $i = 1(1)r$, elements, the vector function F is called "$(k_1,k_2,\dots,k_r)$-weakly cyclic of index k".

Theorem 5: Let $\{R, \leqq\}$ be an ordered set, $F = (f_i)$: $V_nR \Rightarrow V_nR$ a weakly cyclic vector function of index k . Further IS and AN be disjoint subsets of $\{1,\dots,k\}$ with $IS \cup AN = \{1,\dots,k\}$ and $\bigwedge_{i \in IS} \bigwedge_{l \in N_i} f_l$ isotone and $\bigwedge_{i \in AN} \bigwedge_{l \in N_i} f_l$ antitone.

Then, $F^k: V_nR \Rightarrow V_nR$ is an isotone operator, if #AN is even, and an antitone operator, if #AN is odd.

Proof: By means of the definition of the weakly cyclic vector function we get by k applications of F to an element $a \in V_nR$

$$F^k(a) = \Big(\prod_{l=o}^{k-1} (f_{N_{P^l(i)}})(a)\Big) .$$

Therefore, F^k is isotone, if #AN is even, and antitone, if #AN is odd. $\square$

In the next theorem we extend the results of Theorem 5 to $(k_1,\dots,k_r)$-weakly cyclic vector functions of index k .

Theorem 6: Let $\{R, \leqq\}$ be an ordered set, $F = (f_i)$:

$V_nR \to V_nR$ a $(k_1,\dots,k_r)$-weakly cyclic vector function of index k and $p := 2 \cdot \text{s.c.m.}\,(k_1,\dots,k_r)$. Let further IS and AN be disjoint subsets of $\{1,2,\dots,k\}$ with $IS \cup AN = \{1,2,\dots,k\}$ and $\bigwedge_{i\in IS} \bigwedge_{l\in N_i} f_l$ isotone and $\bigwedge_{i\in AN} \bigwedge_{l\in N_i} f_l$ antitone.

Then, F^p is an isotone operator.

Proof: We consider the partitioning of F into the r weakly cyclic vector functions F_i of index k_i, $i=1(1)r$. Applying Theorem 5,the $2k_i$-fold application of F_i is isotone in any case. Therefore the s.c.m. $(2k_1,\dots,2k_r)$-fold application of F is isotone, which yields immediately the statement. □

To give an idea of the magnitude of the number 1 in Theorem 6, we give an estimation in the following

Corollary: Let be given the assumptions of Theorem 6. Then F^p is isotone with

$$p \leqq 2\cdot g(k) \text{ and } g(k)= \begin{cases} 3^{\frac{k}{3}} & k \equiv o \bmod 3 \\ 4\cdot 3^{[\frac{k}{3}-1]} & k \equiv 1 \bmod 3 \\ 2\cdot 3^{[\frac{k}{3}]} & k \equiv 2 \bmod 3 \end{cases} .$$

Proof: The first part of the statement follows from Theorem 5 applied to the r weakly cyclic vector functions F_i of index k_i, $i = 1(1)r$. It is

$$m_{Max} \leqq \max_{\sum_{i=1}^{r} k_i = k,\ r \leqq k} 2 \cdot \text{s.c.m.}(k_1,k_2,\dots,k_r)$$

and hence,

$$m_{Max} \leqq 2 \cdot \max_{\sum_{i=1}^{r} k_i = k} \prod_{i=1}^{r} k_i \ , \ r \leqq k \ .$$

For the determination of that additive partitioning $k_1,\dots,k_r$ of k, for which the product $\prod_{i=1}^{r} k_i$ becomes maximal, the following considerations are possible.

Let be $k = k_1 + k_2$, $k,k_1,k_2 \in \mathbb{N}$. Then,

$k_1 + k_2 < k_1 \cdot k_2$, if $k_1 \geqq 3$ and $k_2 \geqq 2$.

Hence, we get the estimation

$$5 \leqq k = k_1 + k_2 \leqq k_1 \cdot k_2 = (k_{11} + k_{12})\cdot k_2 \leqq$$
$$\leqq k_{11} \cdot k_{12} \cdot k_2 \leqq \dots .$$

The product on the right side of this inequality increases, if k is further partitioned in terms notless 2. So we get a sum of the numbers 2 and 3. Since $2 + 2 + 2 = 6 = 3 + 3$ and $2^3 < 3^2$, we get the maximal product by a partition into the numbers 2 and 3, where the number 2 does not occur more than twice. Now expression $g(k)$ follows immediately. □

For $k = 5(1)12$ the function $g(k)$ assumes the following values:

k	5	6	7	8	9	10	11	12
g(k)	12	18	24	36	54	72	108	162

The results of Theorem 5 and Theorem 6 resp. of the corresponding Corollary are now applied to matrices

Theorem 7: Let $\{R,+,\cdot,\leq\}$ be a linearly ordered rin-goid and $A=(a_{ij})\in M_nR$ a weakly cyclic matrix of index n. With $\#AN = \#\{a_{ij} \mid a_{ij} < o\}$ hold

1. If the iterative method defined by (a) is cyclically ending, then the cycle length m divides n resp. $2n$ for #AN even resp. #AN odd.
2. If A is given in normal form and the iterative method (b) is cyclically ending, then the cycle has length 1 resp. 2 for #AN even resp. #AN odd.

Proof: 1. From Theorem 5 follows, that T^n with $Tx := Ax + b$ is an isotone operator for # AN even resp. an antitone operator for #AN odd, i.e. the component functions of T^n are isotone resp. antitone. Therefore, the component sequences of the iterates $x^{(o)},x^{(n)},x^{(2n)},\dots$ resp. $x^{(o)},x^{(2n)},x^{(4n)},\dots$ are monotone. The application of Theorem 3 gives the desired result.

2. It is

$$L = \begin{bmatrix} o & & & \\ a_1 & \ddots & & 0 \\ & a_2 & & \\ 0 & & \ddots & \\ & & a_{n-1} & o \end{bmatrix}, \mathcal{D} + \mathcal{R} = \begin{bmatrix} o & \dots & o & a_n \\ & & & o \\ & & & \vdots \\ 0 & & \ddots & \vdots \\ & & & o \end{bmatrix}$$

and therefore,

$$Tx = \begin{bmatrix} a_n x_n + b_1 \\ a_1(a_n x_n + b_1) + b_2 \\ \vdots \\ a_{n-1}(a_{n-2}(\dots a_1(a_n x_n + b_1) + b_2) + \dots) + b_n \end{bmatrix} =:$$

$$=: (t_i(x_n)) \ .$$

Analogously, we get
$T^2x = (t_i(t_n(x_n)))$, $T^3x = (t_i(t_n^2(x_n))),\dots,$ $T^{k+1}x = (t_i(t_n^k(x_n))),\dots$. Since $\{R,+,\cdot,\leqq\}$ is a linearly ordered ringoid, we have $x_n \leqq t_nx_n$ resp. $x_n \geqq t_nx_n$. Hence, t_n isotone yields
$t_n^{k-1}x_n \leqq t_n^kx_n$ resp. $t_n^{k-1}x_n \geqq t_n^kx_n$. If t_n is antitone,we obtain from $x_n \leqq t_n^2x_n$ resp. $x_n \geqq t_n^2x_n$ the inequality $t_n^{2(k-1)}x_n \leqq t_n^{2k}x_n$ resp. $t_n^{2(k-1)}x_n \geqq t_n^{2k}x_n$. By Theorem 3 the sequence in each component i, $i=1(1)n$, has the cycle length 1, if the operator t_n is isotone, and the cycle length 1 or 2, if t_n is antitone. By the definition of t_n it is clear, that for #AN even resp. odd t_n is isotone resp. antitone. □

Remark: Theorem 6 gives us the hint that the determination of cycles is not only possible for weakly cyclic matrices. We consider a matrix A which has only one entry not equal o in each row and in each column. Then, by similar considerations the iterative method (a) generates a cycle of length m with $m|2\cdot\text{s.c.m.}\,(n_1,\dots n_r)$, where n_i, $i = 1(1)r$ are the numbers of elements in the r cyclic submatrices of A . In the following table we give for some values of n the greatest attainable cycle length m_{Max}, the corresponding partition of n and the estimation m_E computed by $g(n)$ (see Corollary of Theorem 6).

n	m_{Max}	partition of n	m_E
2	4	2	4
3	6	3	6
4	8	4	8
5	12	2,3	12
6	12	6	18
7	24	3,4	24
8	30	3,5	36
9	40	4,5	54
10	60	2,3,5	72
11	60	1,2,3,5,	108
12	120	3,4,5	162

After we have proved results concerning cycles produced by iterative methods for the solution of systems of equations, we finally want to include a theorem concerning iterative methods for the determination of zeros of real functions.

Theorem 8: Let $\{R,N,+,\cdot,/,\leqq\}$ be a linearly ordered division ringoid with the special elements $\{-e,o,e\}$ and $f,g\colon R \Rightarrow R$ mappings, where f is isotone (resp. antitone) and g is antitone (resp. isotone) in $J \subseteq R$ and $g > o$ (resp. $g < o$) in J . Furthermore,assume that the operator defined by $Tx := x - f(x)/g(x)$, $g(x) \notin N$, generates a sequence $\{x_n\} \subseteq J$, starting with $x_o \in J$. Then, if the sequence $\{x_n\}$ is cyclically ending, the cycle length may be 1 or 2 .

Proof: Let $\{z_1,\dots,z_s\}$ be the cycle of the sequence $\{x_n\}$ with $s > 2$. Since $\{R,\leqq\}$ is linearly ordered, with a permutation P we can write

$$z_{P(1)} < z_{P(2)} < \dots < z_{P(s)} .$$

From f isotone and g antitone and $g > o$ we obtain

$$f(z_{P(1)}) < f(z_{P(2)}) < \dots < f(z_{P(s)})$$

and $g(z_{P(1)}) > g(z_{P(2)}) > \dots > g(z_{P(s)}) > o$.

With the property

$$\bigwedge_{a,b,c\in R} (a \leqq b \wedge c \geqq o \Longrightarrow a/c \leqq b/c) ,$$

which is an immediate consequence of (OD4), we get

$$f(z_{P(1)})/g(z_{P(1)}) < \dots < f(z_{P(s)})/g(z_{P(s)}).$$

With (OD1) and (OD2) we have

$$z_{P(1)}-f(z_{P(1)})/g(z_{P(1)})>\dots>z_{P(s)})-f(z_{P(s)})/g(z_{P(s)})$$

and therefore ,

$$z_{P(s)} = T\,z_{P(1)}$$

$$z_{P(1)} = T\,z_{P(1)} ,$$

i.e. $\{z_{P(1)},z_{P(s)}\}$ forms a cycle of length 2, in contradiction to tne assumption. The other case is proved analogously. □

5. Numerical ex mples

Let $\hat{\mathbb{R}} := \{x \in \mathbb{R} \mid |x| \leqq s\}$ be a bounded subset of the real number field. Then, a finite symmetric subset $\hat{\mathbb{R}}_M$ of $\hat{\mathbb{R}}$ with the property $-s,s \in \hat{\mathbb{R}}_M$ is a symmetric screen of $\{\hat{\mathbb{R}}, \leqq\}$ (see [4]). Furthermore, $\{\hat{\mathbb{R}}, \leqq\}$ forms a completely, linearly ordered ringoid by using the real arithmetic. With a monotone, antisymmetric rounding

$$\square : \hat{\mathbb{R}} \to \hat{\mathbb{R}}_M \ ,$$

i.e. an isotone,idempotent mapping from $\hat{\mathbb{R}}$ to $\hat{\mathbb{R}}_M$ with the property

$$\bigwedge_{a \in \hat{\mathbb{R}}} \square(-a) = -\square a \ ,$$

and the inner operations $\boxed{*}$ defined by

$$\bigwedge_{a,b \in \hat{\mathbb{R}}_M} a \boxed{*} b := \square(a * b), \ * \in \{+,\cdot\} \ ,$$

$\hat{\mathbb{R}}_M$ forms a completely,linearly ordered ringoid $\{\hat{\mathbb{R}}_M, \boxplus, \boxdot, \leqq\}$ (see [3]). By analogous considerations we obtain a completely,linearly ordered division ringoid $\{\hat{\mathbb{R}}_M, \{o\}, \boxplus, \boxdot, \boxed{/}, \leqq\}$

Thus, $\{\hat{\mathbb{R}}_M, \boxplus, \boxdot, \leqq\}$ resp. $\{\hat{\mathbb{R}}_M, \{o\}, \boxplus, \boxdot, \boxed{/}, \leqq\}$ and $\{V_n\hat{\mathbb{R}}_M, M_n\hat{\mathbb{R}}_M, \leqq\}$ fulfill the assumptions of the theorems proved in the precedent sections.

Now we replace $\hat{\mathbb{R}}_M$ by the set $\hat{\mathbb{R}}_{m,b}$ of the normalized floating-point numbers $\hat{\mathbb{R}}_{m,b} :=$ $\{x \in \mathbb{R} \mid x = * o.d_1 d_2 \dots d_m \cdot b^e, \ * \in \{+,-\}$, $d_i \in \{o,1,\dots,b-1\}, \ i = 1(1)m, \ d_1 \neq o,$ $e_{min} \leqq e \leqq e_{max}, \ e \in \mathbb{Z}\} \cup \{o\}$ with $m,b \in \mathbb{N}, \ b \geqq 2$. Then, the set $\hat{\mathbb{R}}$ is given by $\hat{\mathbb{R}} :=$ $\{x \in \mathbb{R} \mid |x| \leqq o.d_1 d_2 \dots d_m \cdot b^{e_{max}}, d_i = b-1, \ i=1(1)m\}$. Furthermore, we apply the frequently used monotone, antisymmetric rounding

$$\square a := \begin{cases} \nabla a \text{ if } a < (\nabla a + \Delta a)/2 \text{ or } a = (\nabla a + \Delta a)/2 \wedge a \leqq o \\ \Delta a \text{ if } a > (\nabla a + \Delta a)/2 \text{ or } a = (\nabla a + \Delta a)/2 \wedge a > o \end{cases}$$

(rounding to the nearest number of $\hat{\mathbb{R}}_{m,b}$), where ∇ and Δ mean the uniquely defined monotone, directed roundings,([5]).

The examples 2. and 4. has been performed on the digital computer EL X8 of the Computer Center of the University of Karlsruhe. Note, that the decimal numbers of $\hat{\mathbb{R}}_{12,1o}$ used for the representation of the operators and cycles have been generated by conversion of binary numbers of $\hat{\mathbb{R}}_{4o,2}$. We cannot expect, that the described operators produce the same cycles by computation in $\hat{\mathbb{R}}_{12,1o}$.

Examples:

1. Let be n = 2, b = 1o, m = 2. The operator

$$Tx = \begin{bmatrix} o & -o.78 \\ o.93 & o \end{bmatrix} \boxdot x \boxplus \begin{bmatrix} 1.2 \\ o.75 \end{bmatrix} \ ,$$

generates a sequence $\{x^{(n)}\}$ by $x^{(n+1)} :=$ $Tx^{(n)}$, $x^{(o)} \in V_2\hat{\mathbb{R}}_{2,1o}$. For different starting vectors $x^{(o)}$ the sequence is ending in different cycles:

$$Z_1 = \left\{ \begin{bmatrix} o.26 \\ 1.2 \end{bmatrix}, \begin{bmatrix} o.26 \\ o.99 \end{bmatrix}, \begin{bmatrix} o.43 \\ o.99 \end{bmatrix}, \begin{bmatrix} o.43 \\ 1.2 \end{bmatrix} \right\}$$

$$Z_2 = \left\{ \begin{bmatrix} o.34 \\ 1.2 \end{bmatrix}, \begin{bmatrix} o.26 \\ 1.1 \end{bmatrix}, \begin{bmatrix} o.34 \\ o.99 \end{bmatrix}, \begin{bmatrix} o.43 \\ 1.1 \end{bmatrix} \right\}$$

$$Z_3 = \left\{ \begin{bmatrix} o.34 \\ 1.1 \end{bmatrix} \right\} \ .$$

According to Theorem 7(1) we get a maximal cycle length 4 and by Theorem 7 (2) the single-step method is ending in a cycle of length 1 or 2:

$$Z_3 \text{ and } Z_4 = \left\{ \begin{bmatrix} o.26 \\ o.99 \end{bmatrix}, \begin{bmatrix} o.43 \\ 1.2 \end{bmatrix} \right\} \ .$$

Applying the monotone rounding towards zero we get beginning with $x^{(o)} = (o.26,1.2)^T$ the cycle

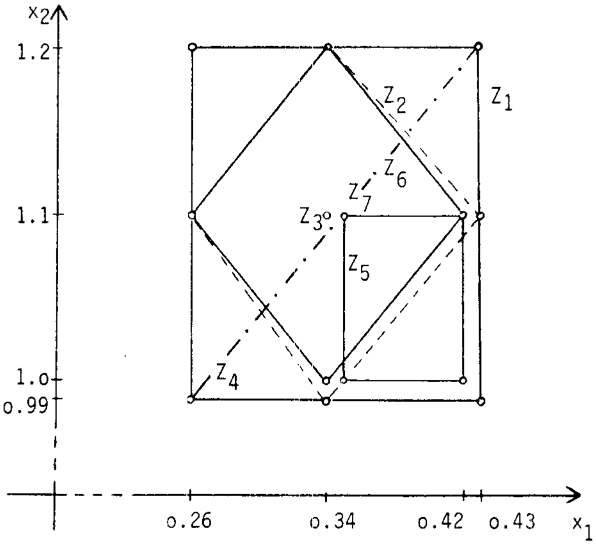
$$Z_5 = \left\{ \begin{bmatrix} o.42 \\ 1.1 \end{bmatrix}, \begin{bmatrix} o.35 \\ 1.1 \end{bmatrix}, \begin{bmatrix} o.35 \\ 1.o \end{bmatrix}, \begin{bmatrix} o.42 \\ 1.o \end{bmatrix} \right\}$$

and applying the monotone rounding away from zero we get beginning with $x^{(o)} = (o.27,1.2)^T$ the cycle

$$Z_6 = \left\{ \begin{bmatrix} o.26 \\ 1.1 \end{bmatrix}, \begin{bmatrix} o.34 \\ 1.o \end{bmatrix}, \begin{bmatrix} o.42 \\ 1.1 \end{bmatrix}, \begin{bmatrix} o.34 \\ 1.2 \end{bmatrix} \right\}$$

and the cycle $Z_7 = \left\{ \begin{bmatrix} o.34 \\ 1.1 \end{bmatrix} \right\}$.

We illustrate the cycles Z_1 to Z_7 graphically:



2. Another example of Theorem 7 (1) is given by the operator

$$Tx := \begin{pmatrix} o & o & -o.6638554373967 \\ -o.6686745454153 & o & o \\ o & -o.6795139955248 & o \end{pmatrix} \boxdot x \boxplus$$

$$\boxplus \begin{pmatrix} 4.23575o97976o \\ -2.5776794259o7 \\ o.7134761463822 \end{pmatrix}$$

in $V_3\hat{\mathbb{R}}_{13,1o}$, which produces e.g. the following cycle

$$\begin{pmatrix} 1.996957891o17 \\ -3.912994335897 \\ 3.37241o562o35 \end{pmatrix}, \begin{pmatrix} 1.996957891o17 \\ -3.912994335897 \\ 3.37241o562o31 \end{pmatrix}, \begin{pmatrix} 1.996957891o2o \\ -3.912994335897 \\ 3.37241o562o31 \end{pmatrix},$$

$$\begin{pmatrix} 1.996957891o2o \\ -3.9129943359oo \\ 3.37241o562o31 \end{pmatrix} \begin{pmatrix} 1.996957891o2o \\ -3.9129943359oo \\ 3.37241o562o35 \end{pmatrix} \begin{pmatrix} 1.996957891o17 \\ -3.9129943359oo \\ 3.37241o562o35 \end{pmatrix}$$

$$\begin{pmatrix} 1.996957891o17 \\ -3.912994335897 \\ 3.37241o562o35 \end{pmatrix}, \ldots .$$

3. Finally, we give an operator in $V_5\hat{\mathbb{R}}_{4o,2}$, which generates according to the remark of Theorem 7 a cycle of maximum length 12:

$$Tx := \begin{pmatrix} o & o & a_1 & o & o \\ a_2 & o & o & o & o \\ o & a_3 & o & o & o \\ o & o & o & o & a_4 \\ o & o & o & a_5 & o \end{pmatrix} \boxdot x \boxplus \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix}$$

with

a_1 =	-o.7551928744688	b_1 =	o.63o577475600l
a_2 =	o.1o32844o94188	b_2 =	o.57554821o81o7
a_3 =	o.6376119369179	b_3 =	o.265o77o197379
a_4 =	o.47474o285o753	b_4 =	o.119o3o2626437
a_5 =	o.8622977826562	b_5 =	-o.o32834795o5717

With the elements

x_1 =	o.1459944151620	y_1 =	o.1459944151626
x_2 =	o.59o6271577596	y_2 =	o.59o6271577587
x_3 =	o.6416679457934	y_3 =	o.6416679457925
x_4 =	o.o7339623131165	y_4 =	o.o7339623131184
x_5 =	-o.o961242o257263	y_5 =	-o.o961242o257272

we can describe the generated cycle in the following way:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}, \begin{pmatrix} x_1 \\ y_2 \\ x_3 \\ y_4 \\ x_5 \end{pmatrix}, \begin{pmatrix} x_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ x_4 \\ y_5 \end{pmatrix}, \begin{pmatrix} y_1 \\ x_2 \\ y_3 \\ x_4 \\ x_5 \end{pmatrix}, \begin{pmatrix} y_1 \\ x_2 \\ x_3 \\ y_4 \\ x_5 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ y_4 \\ y_5 \end{pmatrix}, \begin{pmatrix} x_1 \\ y_2 \\ x_3 \\ x_4 \\ y_5 \end{pmatrix}, \begin{pmatrix} x_1 \\ y_2 \\ y_3 \\ x_4 \\ x_5 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ x_5 \end{pmatrix},$$

$$\begin{pmatrix} y_1 \\ x_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix}, \begin{pmatrix} y_1 \\ x_2 \\ x_3 \\ x_4 \\ y_5 \end{pmatrix} .$$

References:

[1] COLLATZ, L.: Funktionalanalysis und numerische Mathematik, Springer Verlag, Berlin 1964

[2] KLATTE, R.: Zyklisches Enden bei Iterationsverfahren. Dr.-Dissertation, Universität Karlsruhe, 1975

[3] KULISCH, U.: Rounding Invariant Structures. Mathematics Research Center, University of Wisconsin, Technical Summary Report # 1103, September 1970, 1-47

[4] KULISCH, U.: An Axiomatic Approach to Rounded Computations. Numerische Mathematik 18, 1-17 (1971)

[5] KULISCH, U.: Implementation and Formalization of floating-point arithmetic. To appear in Computing

[6] ULLRICH, Chr.: Rundungsinvariante Strukturen mit äußeren Verknüpfungen. Dr.-Dissertation , Universität Karlsruhe , 1972

[7] VARGA, R.: Matrix Iterative Analysis, Prentice Hall, Inc., Englewood Cliffs, New Jersey 1962

Address:

Dr. R. Klatte, Dr. Chr. Ullrich
Institut für Angewandte Mathematik
Universität Karlsruhe

D-75 Karlsruhe
Kaiserstr. 12
West Germany