On Singular Interval Systems

Götz Alefeld[1] and Günter Mayer[2]

[1] Universität Karlsruhe, 76128 Karlsruhe, Germany
goetz.alefeld@math.uni-karlsruhe.de
[2] Universität Rostock, 18051 Rostock, Germany
guenter.mayer@mathematik.uni-rostock.de

Abstract. We consider the interval iteration $[x]^{k+1} = [A][x]^k + [b]$ with $\rho(|[A]|) \le 1$ where $|[A]|$ denotes the absolute value of the given interval matrix $[A]$. If $|[A]|$ is irreducible we derive a necessary and sufficient criterion for the existence of the limit $[x]^* = [x]^*([x]^0)$ of each sequence $([x]^k)$ of interval iterates. In this way we generalize a well–known theorem of O. Mayer [6] on the above–mentioned iteration, and we are able to enclose solutions of certain singular systems $(I - A)x = b$ with $A \in [A]$ and degenerate interval vectors $[b] \equiv b$. Moreover, we give a connection between the convergence of $([x]^k)$ and the convergence of the powers of $[A]$.

1 Introduction

Consider Poisson's equation

$$\frac{\partial^2 u}{\partial s^2} + \frac{\partial^2 u}{\partial t^2} = -f(s,t) \tag{1}$$

on the unit square $Q = [0,1] \times [0,1]$ with a continuous function f defined on Q. If one looks for a solution $u(s,t)$ of (1) subject to the periodic boundary conditions

$$\left.\begin{array}{l} u(0,t) = u(1,t),\ 0 \le t \le 1 \\ u(s,0) = u(s,1), 0 \le s \le 1 \end{array}\right\} \tag{2}$$

and if one discretizes (1) using an equidistant grid of mesh size $h = \frac{1}{n}$, $n \in \mathbb{N}\backslash\{1,2\}$, a row–wise ordering and the well–known five point central difference approximation one ends up with a system

$$Cx = b \tag{3}$$

of linear equations in which $C \in \mathbb{R}^{n^2 \times n^2}$ is defined by

$$C = \frac{1}{4}\begin{pmatrix} D & -I & O & \dots & O & -I \\ -I & D & -I & O & \dots & O \\ O & -I & D & -I & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & O \\ O & \dots & O & -I & D & -I \\ -I & O & \dots & O & -I & D \end{pmatrix}, \quad D = \begin{pmatrix} 4 & -1 & 0 & \dots & 0 & -1 \\ -1 & 4 & -1 & 0 & \dots & 0 \\ 0 & -1 & 4 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 4 & -1 \\ -1 & 0 & \dots & 0 & -1 & 4 \end{pmatrix},$$

R. Alt et al. (Eds.): Num. Software with Result Verification, LNCS 2991, pp. 191–197, 2004.

$D \in \mathbb{R}^{n\times n}$, $I =$ identity matrix. The components b_i of $b \in \mathbb{R}^{n^2}$ are given by

$$b_i = \frac{h^2}{4} f(s_l, t_m), \quad i = 1, \dots, n^2,$$

with $s_l = t_l = lh$, $i = (m-1)\cdot n + l$, $l, m = 1, \dots, n$. When discretizing one assumes a periodic continuation of u across the boundary of Q. The unknowns x_i refer to the inner grid points and to the grid points of the right and upper boundary of Q. It is known (cf. [2], p. 196 ff) that C is a singular matrix of rank n^2-1. This follows from the fact that it is a singular irreducible M matrix with property c (cf. [2], Definition 6.4.10, Theorem 6.4.16 and p. 201). Richardson splitting applied to C yields to the iterative process

$$x^{k+1} = Ax^k + b, \quad k = 0, 1, \dots , \tag{4}$$

where $A = I - C$. Since every diagonal element of C is 1 the iteration (4) coincides here with the Jacobi method for (3). If n is odd the matrix A has the spectral radius $\rho(A) = 1$, and all eigenvalues λ of A with $|\lambda| = 1$ are one and have only linear elementary divisors, i.e., the corresponding Jordan blocks are 1×1. Such matrices – together with those of spectral radius less than one – are called semi–convergent ([2], p. 152). They represent just the matrices for which the limit $A^\infty = \lim_{k\to\infty} A^k$ exists.

We remark that the matrix A arising from the discretization of (1), (2) is symmetric. Therefore, all eigenvalues of A have only linear elementary divisors. In addition, A is non–negative and irreducible. Hence the Theorem of Perron and Frobenius guarantees that the eigenvalue $\lambda = 1$ is even algebraically simple which is not required in the definition of semi–convergence and which is not necessary for our subsequent considerations. Moreover, A is primitive if n is odd, and cyclic of index 2 if n is even. This can be seen by inspecting the lengths of the circuits in the directed graph associated with A ([2], § 2.2). Therefore, the theory of Perron and Frobenius on non–negative irreducible matrices shows that $\lambda = 1$ is the only eigenvalue of A with $|\lambda| = \rho(A) = 1$ in the case of n being odd while $\lambda = -1$ is another eigenvalue with this property in the case of even n.

In this short note we will consider the case where A is allowed to vary within a given interval matrix $[A]$ such that the absolute value $|[A]|$ of $[A]$ is irreducible and semi–convergent. We present – in a condensed form – results on the corresponding interval iteration

$$[x]^{k+1} = [A][x]^k + [b], \quad k = 0, 1, \dots \tag{5}$$

generalizing in this way a well–known theorem of O. Mayer [6]; cf. also [1], pp. 143 ff. By lack of space we must omit the very lengthy and by no means straightforward proofs. They will be published elsewhere.

We finally remark that singular linear systems also occur in other situations – cf. [2], § 7.6, in this respect.

2 Results

In order to recall some results for the iterative process (4) with (general) semi–convergent matrices A we first define the Drazin inverse A^D of an arbitrary $n \times n$ matrix $A = S \begin{pmatrix} \hat{J}_0 & O \\ O & \hat{J}_r \end{pmatrix} S^{-1}$ by $A^D = S \begin{pmatrix} O & O \\ O & (\hat{J}_r)^{-1} \end{pmatrix} S^{-1}$. Here, $J = \begin{pmatrix} \hat{J}_0 & O \\ O & \hat{J}_r \end{pmatrix}$ is the Jordan canonical form of A with square blocks $\hat{J}_0$, $\hat{J}_r$, whose diagonal blocks are just the singular Jordan blocks of J, and the non–singular ones, respectively; cf. for instance [2], § 5.4.

The following theorem which is contained in Lemma 7.6.13 in [2] answers completely the question on the convergence of (4).

Theorem 1. *Let (3) (with a matrix C not necessary equal to the one obtained by discretizing (1) and (2)) be solvable. Then each sequence (x^k) of iterates defined by (4) is convergent if and only if A is semi-convergent. The limit is independent of x^0 if and only if $\rho(A) < 1$. In any case this limit x^* is a solution of (3) and a fixed point of (4). By means of Drazin inverses it can be expressed as*

$$x^* = (I - A)^D b + \{I - (I - A)(I - A)^D\} x^0.$$

If $\rho(A) < 1$ then $(I - A)^{-1}$ exists. Hence (3) is uniquely solvable and by virtue of $(I - A)^{-1} = (I - A)^D$ Theorem 1 reduces to a basic result of numerical analysis in this case. Therefore, it is essentially the case $\rho(A) = 1$ which is of interest in our paper.

For the interval iteration (5) we will replace the assumption of solvability in Theorem 1 by the existence of a fixed point of (5). For interval matrices $[A]$ with $\rho(|[A]|) < 1$ the above–mentioned theorem of O. Mayer [6] guarantees that (5) has a unique fixed point. If $|[A]|$ is irreducible and satisfies $\rho(|[A]|) = 1$ we could prove in [5] an exhaustive result on the existence and the shape of such fixed points. In order to formulate our main result we need the following definition.

Definition 1. *([3], [4]) Let $[A]$ be an $n \times n$ interval matrix. Let*

$$[A]^0 = I, \qquad [A]^{k+1} = [A]^k \cdot [A], \quad k = 0, 1, \dots .$$

If $[A]^\infty = \lim_{k\to\infty} [A]^k$ exists then we call $[A]$ semi–convergent.

Theorem 2. *Let $[A]$ be a non–degenerate $n \times n$ interval matrix with irreducible absolute value $|[A]|$. Let the iteration (5) have a fixed point $[z]^*$ (which implies $[b] \equiv b \in \mathbb{R}^n$ in the case $\rho(|[A]|) = 1$ according to Theorem 8 in [5]). Then the following three statements are equivalent.*

a) Each sequence $([x]^k)$ of (5) is convergent.
b) The interval matrix $[A]$ is semi–convergent.
c) The absolute value $|[A]|$ is semi–convergent. Moreover, if $\rho(|[A]|) = 1$ and if $[A]$ contains only one matrix $\dot{A}$ with $|\dot{A}| = |[A]|$ then $\dot{A} \neq -D|[A]|D$ for all matrices D with $|D| = I$.

In case of convergence of (5) the limit $[x]^* = [x]^*([x]^0)$ of $([x]^k)$ is a fixed point of the iteration (5). It contains the set $S([x]^0)$ of all solutions of (3) which one obtains as limits of sequences (x^k) of iterates defined by (4) with $x^0 \in [x]^0$, i.e.,

$$S([x]^0) = \Big\{x^* \mid x^* = (I-A)^D b + \{I-(I-A)(I-A)^D\}x^0, \\ A \in [A],\ x^0 \in [x]^0,\ b \in [b]\Big\} \subseteq [x]^*([x]^0)\ .$$

In case of convergence of (5) the limit $[x]^*$ of $([x]^k)$ does not depend on the starting vector $[x]^0$ if and only if one of the following equivalent properties holds:

(i) $\rho(|[A]|) < 1$.
(ii) $\lim_{k\to\infty} |[A]|^k = O$.
(iii) $\lim_{k\to\infty} [A]^k = O$.

Note that the equivalence 'a) ⇔ c)' remains true even if $[A]$ is degenerate (and $|[A]|$ is irreducible) while 'b) ⇒ c)' becomes false as the example

$$[A] \equiv A = \begin{pmatrix} 2/3 & 2/3 \\ 2/3 & -2/3 \end{pmatrix} \tag{6}$$

shows. (Cf. [3], [4] for further details.) Since the statements b), c) do not depend on $[b]$ and since (5) has always the fixed point $[z]^* \equiv 0$ for $[b] \equiv 0$ the existence of $[z]^*$ does not need to be assumed in Theorem 2 for the equivalence of b) and c). If $|[A]|$ is reducible the equivalence of (ii) and (iii) becomes false as can be seen, e.g., by the 2×2 block diagonal matrix $[A] = \mathrm{diag}([0,1/2], B)$ where B is the matrix denoted by A in (6). We refer to [3] or [7] in this case.

We conclude our contribution with a numerical example which illustrates the theory.

Example 1. Define the $n \times n$ interval matrix $[D]$ by

$$[D] = [D]_{[\alpha],[\beta]} = \begin{pmatrix} [\alpha] & [\beta] & 0 & \dots & 0 & [\beta] \\ [\beta] & [\alpha] & [\beta] & 0 & \dots & 0 \\ 0 & [\beta] & [\alpha] & [\beta] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & [\beta] & [\alpha] & [\beta] \\ [\beta] & 0 & \dots & 0 & [\beta] & [\alpha] \end{pmatrix},$$

and the $n^2 \times n^2$ interval matrix $[A] = [\underline{A}, \overline{A}]$ in block form by

$$[A] = \begin{pmatrix} [D] & [\gamma]I & O & \dots & O & [\gamma]I \\ [\gamma]I & [D] & [\gamma]I & O & \dots & O \\ O & [\gamma]I & [D] & [\gamma]I & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & O \\ O & \dots & O & [\gamma]I & [D] & [\gamma]I \\ [\gamma]I & O & \dots & O & [\gamma]I & [D] \end{pmatrix},$$

where $[\alpha]$, $[\beta]$, $[\gamma]$ are intervals which are still to be chosen. By means of the Kronecker product $\otimes$ the matrix $[A]$ can be written as

$$[A] = I \otimes [D]_{[\alpha],[\beta]} + [D]_{0,[\gamma]} \otimes I, \qquad I \in \mathbb{R}^{n \times n}.$$

In this way it can easily be constructed in software packages like INTLAB [8] whose version 4.1.1 we used for our interval computations. We choose $[\alpha]$, $[\beta] \neq 0$, $[\gamma] \neq 0$ such that $|[\alpha]| + 2|[\beta]| + 2|[\gamma]| = 1$ holds. Then $|[A]|$ is irreducible and $\rho(|[A]|) = 1$ is guaranteed. Moreover, $[b] \equiv b = (b_i) \in \mathbb{R}^{n^2}$ is necessary for the existence of a fixed point of (5) which is required as assumption in Theorem 2.

We first use $n = 5$, $[\alpha] = 0$, $[\beta] = [\gamma] = 1/4$. This leads to the particular situation of Section 1 in which we showed that $[A] \equiv A = |[A]| \in \mathbb{R}^{n^2 \times n^2}$ is semi–convergent with $\rho(|[A]|) = 1$. If $b = (I - A)\check{z}$ for some $\check{z} = (\check{z}_i) \in \mathbb{R}^{n^2}$ then Theorem 8 in [5] guarantees that (5) has the fixed points

$$[z]^* = \check{z} + se + t[-1, 1]e, \tag{7}$$

where s, t are any real numbers with $t \geq 0$ and where $e = (1, 1, \ldots, 1)^T \in \mathbb{R}^{n^2}$ is an eigenvector of $A \geq O$ associated with the eigenvalue $\lambda = \rho(A) = 1$. Therefore, by virtue of Theorem 2 a), c), extended by the first remark following this theorem, the limits $[x]^* = [x]^*([x]^0)$ exist for any starting vectors $[x]^0$ and are precisely the vectors $[z]^*$ in (7). We choose $b_i = b_{n^2+1-i} = 0.5$ for $i \in \{1, 3, 4, 7\}$, $b_2 = b_{n^2-1} = -2$, $b_i = 0$ otherwise. Then $\check{z} = (1, -1, 1, 1, \ldots, 1, 1, -1, 1)^T \in \mathbb{R}^{n^2}$ satisfies $b = (I - A)\check{z}$ as required above. We iterated according to (5) with different starting vectors $[x]^0$. We stopped the iteration either when the criterion

$$[\tilde{x}]^k = [\tilde{x}]^{k-1} \tag{8}$$

was fulfilled for some $k = k_0$ or when k reached a given upper bound $k_{\max}$, where here and in the sequel the tilde denotes computed, i.e., rounded quantities. By the outward rounding of the machine interval arithmetic (cf., e.g., [1]) we always have $[x]^k \subseteq [\tilde{x}]^k = ([\tilde{x}]_i^k)$, $k = 0, 1, \ldots$. Moreover, in the case (8) we can guarantee $[x]^k \subseteq [\tilde{x}]^{k_0}$, $k = k_0, k_0 + 1, \ldots$, whence $[x]^* \subseteq [\tilde{x}]^{k_0}$.

If (8) cannot be obtained, i.e., in the case where k reaches $k_{\max}$ one can compute the midpoints $\tilde{m}_i = \operatorname{mid}\left([\tilde{x}]_i^{k_{\max}} - \check{z}_i\right)$, $i = 1, \ldots, n^2$, and the radii $\tilde{r}_i = \operatorname{rad}\left([\tilde{x}]_i^{k_{\max}} - \check{z}_i\right)$, $i = 1, \ldots, n^2$. Here, we assume that the *computed* values $\tilde{m}_i, \tilde{r}_i$ satisfy $[\tilde{x}]_i^{k_{\max}} - \check{z}_i \subseteq \tilde{m}_i + \tilde{r}_i[-1, 1]$. Define

$$\tilde{s} = \left(\max_i \tilde{m}_i + \min_i \tilde{m}_i\right) / 2 \in \mathbb{R} \tag{9}$$

and

$$\tilde{t} = \max_i \left(\tilde{r}_i + |\tilde{s} - \tilde{m}_i|\right) \in \mathbb{R} \tag{10}$$

using upward rounding in the latter case. According to (7) the vector

$$[\hat{z}]^* = \check{z} + \tilde{s}e + \tilde{t}[-1, 1]e$$

(not to be confused with $[z]^*$ in Theorem 2) is a fixed point of (5) provided that $[\hat{z}]^*$ is computed with *exact* arithmetic. By construction, $[\hat{z}]^*$ contains $[\tilde{x}]^{k_{\max}}$. From $[x]^{k_{\max}} \subseteq [\tilde{x}]^{k_{\max}} \subseteq [\hat{z}]^*$ we get

$$[x]^k \subseteq [\hat{z}]^*, \quad k = k_{\max}, k_{\max}+1, \dots , \quad \text{whence } [x]^* \subseteq [\hat{z}]^*.$$

This holds also if $k_{\max}$ is replaced by k_0 in the case (8). In our tables we list $[\hat{z}]^*$ in both cases.

Table 1. Starting vector vs. enclosure $[\hat{z}]^* = \check{z} + \tilde{s}e + \tilde{t}[-1,1]e$

$[x]^0$	$\tilde{s}$	$\tilde{t}$	k_0	$k_{\max}$
0	-0.84	$1.021405182655144 \cdot 10^{-14}$	192	
e	$0.16 + 10^{-14}$	$3.996802888650564 \cdot 10^{-14}$	–	200
$[-1,1]e$	$-0.84 + 2 \cdot 10^{-14}$	1.00000000000006	–	200
$[-2,1]e$	$-1.34 + 10^{-14}$	1.50000000000006	172	
$\left((-1)^i[-1,2]\right)_{i=1}^{n^2}$	$-0.86 + 10^{-14}$	1.50000000000006	–	200

Without further knowledge on a relation between $[x]^*$ and $[x]^0$ we cannot, of course, assess the quality which the enclosure $[\tilde{x}]^{k_0}$ or $[\hat{z}]^*$ of the true limit $[x]^*$ has with respect to $[x]^*$. For degenerate starting vectors $[x]^0 \equiv x^0$, however, the radius of $[\tilde{x}]^{k_0}$, and $[\hat{z}]^*$, respectively, may indicate this quality. In theory this radius is zero for such starting vectors, in practice it is not by virtue of rounding errors during the iteration. Table 1 contains the parameters $\tilde{s}$, $\tilde{t}$ from (9), (10) for different starting vectors.

Table 2. Starting vector vs. enclosure $[\hat{z}]^* = \check{z} + \tilde{s}e + \tilde{t}[-1,1]e$

$[x]^0$	$\tilde{s}$	$\tilde{t}$	k_0	$k_{\max}$
0	0	1	814	
e	0	1	796	
$[-1,1]e$	$-0.42 - 10^{-14}$	$1.42 + 10^{-14}$	796	
$[-2,1]e$	$-0.92 - 2 \cdot 10^{-14}$	$1.92 + 2 \cdot 10^{-14}$	796	
$\left((-1)^i[-1,2]\right)_{i=1}^{n^2}$	$-0.68 - 2 \cdot 10^{-14}$	$1.68 + 2 \cdot 10^{-14}$	777	

We next choose $n = 5$, $[\alpha] = [0, 1/4]$, $[\beta] = [0, 1/8]$, $[\gamma] = [1/8, 1/4]$ and $b = (I - \overline{A})\check{z}$ with $\check{z}$ as above. Then nearly all earlier remarks hold analogously, and we obtain the results of Table 2. By virtue of Theorem 8 in [5] we get the restriction $t \geq |\pm 1 + s| = 1 + |s|$ for s, t from (7). A short glance at Table 2 reveals that this inequality also holds for our computed values $\tilde{s}, \tilde{t}$ instead of s, t.

Acknowledgement. The authors are grateful to two anonymous referees for a series of comments and remarks which improved the paper.

References

1. Alefeld, G., Herzberger, J.: Introduction to Interval Computations. Academic Press, New York, 1983
2. Berman, A., Plemmons, R.J.: Nonnegative Matrices in the Mathematical Sciences. Academic Press, New York, 1979
3. Mayer, G.: On the convergence of powers of interval matrices. Linear Algebra Appl. **58** (1984) 201 – 216
4. Mayer, G.: On the convergence of powers of interval matrices (2). Numer. Math. **46** (1985) 69 – 83
5. Mayer, G., Warnke, I.: On the fixed points of the interval function $[f]([x]) = [A][x] + [b]$. Linear Algebra Appl. **363** (2003) 201 – 216
6. Mayer, O.: Über die in der Intervallrechnung auftretenden Räume und einige Anwendungen, Ph.D. Thesis, Universität Karlsruhe, Karlsruhe, 1968
7. Pang, C.-T., Lur, Y.-Y., Guu, S.-M.: A new proof of Mayer's theorem, Linear Algebra Appl. **350** (2002) 273 – 278
8. Rump, S.M.: INTLAB – INTerval LABoratory. In: Csendes T. (ed.), Developements in Reliable Computing, Kluwer, Dordrecht, 1999, 77 – 104