3 Basic notation and ideas

- Basic notation
- Basic interval notation, arithmetics, operations and relations
- Interval structures, expressions and functions
- Intervals and rounding
- Comparison of interval structures
- Testing of interval algorithms

This is a preliminary chapter containing the elementary building blocks for this work. We start with the basic notation for real mathematical objects. Then we move towards interval related material. We briefly introduce interval arithmetics and other operations on intervals. Later, we explain how to work with more complex interval structures – vectors, matrices, expressions and functions. Relation of intervals and computer rounded arithmetics is discussed. Because in almost every subsequent chapter we compare various algorithms with interval outputs, we explain how to compare quality of interval results here. We also state what software and computational power we use for such testing..

3.1 Notation

For the sake of clarity we provide a list of notation that we are going to use for real structures:

notation	**explanation**
A	a real matrix
x, b	a real column vector
I or I_n	identity matrix of the corresponding size
e_i	ith column of I
E	all-ones matrix of the corresponding size

notation	explanation
A_{ij}	the coefficient in ith row and jth column of a matrix A
a_{ij} or $a_{(i,j)}$	the coefficient in ith row and jth column of a matrix A
A_{i*}	ith row of a matrix A
A_{*j}	ith column of a matrix A
$A_{1,(1:3)}$	a vector (a_{11}, a_{12}, a_{13}) (notation borrowed from Matlab)
A^T	A transposed
$\vert \cdot \vert$	absolute value (for vectors and matrices works element-wise)
$\Vert \cdot \Vert_p$	vector or matrix p-norm
A^+	the Moore-Penrose pseudoinverse of A
A^{-1}	inverse matrix
A^{-T}	inverse of A^T
$\varrho(A)$	spectral radius of A
S^n	a set of all vectors of length n with coefficients from the set S
Y_n	the set $\{\pm 1\}^n$

For every vector $x \in \mathbb{R}^n$ we define its sign vector $\operatorname{sign}(x) \in \{\pm 1\}^n$ as

$$\operatorname{sign}(x)_i = \begin{cases} 1, & \text{if } x_i \geq 0, \\ 0, & \text{if } x_i = 0. \\ -1, & \text{if } x_i < 0. \end{cases}$$

Functions $\max, \min$ applied on a vector are understand in a similar way as in Matlab, they choose maximum/minimum of the vector coefficients. For a given vector $x \in \mathbb{R}^n$ we denote

$$D_x = \operatorname{diag}(x_1, \dots, x_n) = \begin{pmatrix} x_1 & 0 & \dots & 0 \\ 0 & x_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x_n \end{pmatrix}.$$

By writing $|x - y| < \epsilon$ for two vectors x, y of the same length we mean $|x_i - y_i| < \epsilon$ for each i. Hence, when relation operators such as $>, <, \leq, \geq, =$ are applied to vectors or matrices, then, unless not stated otherwise, they are understood component-wise.

3.2 Interval

The key notion of this work is an *interval*. Even though, there are various types of intervals, here we understand it as a synonym for a real closed interval.

Definition 3.1 (Interval). For $\underline{a}, \overline{a} \in \mathbb{R}$ a real closed interval $\boldsymbol{a}$ is defined as

$$\boldsymbol{a} = [\underline{a}, \overline{a}] = \{a \in \mathbb{R} \mid \underline{a} \leq a \leq \overline{a}\},$$

$\underline{a}, \overline{a}$ are called the lower and upper bound respectively.

If it holds that $\underline{a} = \overline{a}$, then we call the interval *degenerate*. If $\underline{a} = -\overline{a}$, then we call the interval *symmetric*. We denote the set of all real closed intervals by $\mathbb{IR}$. Open intervals will be only rarely needed and their use will be explicitly announced. They will be typeset with parentheses (i.e., (a, b)).

An interval can be also defined using a center and a distance from this center.

Definition 3.2 (Interval 2). For $a_c \in \mathbb{R}$ and positive $a_\Delta \in \mathbb{R}$ a real closed interval $\boldsymbol{a}$ can be also defined as

$$\boldsymbol{a} = [a_c - a_\Delta, a_c + a_\Delta],$$

a_c and a_Δ are called the *midpoint* and *radius* respectively.

Sometimes it simplifies the notation to move the subscripts to the top, i.e., a^c, a^Δ, especially when other subscripts are used. We use this notation interchangeably. To be concise, when speaking about an interval $\boldsymbol{a}$ we implicitly assume that a_c, a_Δ are respectively its midpoint and radius.

Even though, the two definitions are obviously equivalent, using a proper definition may save excessive notation. Intervals and derived interval structures are denoted in boldface (i.e., $\boldsymbol{x}, \boldsymbol{A}, \boldsymbol{b}, \boldsymbol{f}$). Real numbers, vectors, matrices, functions, etc., are typeset in normal font (i.e., x, A, b, f).

3.3 Set operations

Intervals can be viewed as sets and therefore the typical set operations can be defined for them.

Definition 3.3 (Set operations). Let us have two intervals $\boldsymbol{a} = [\underline{a}, \overline{a}]$ and $\boldsymbol{b} = [\underline{b}, \overline{b}]$. Then $\boldsymbol{a} \cap \boldsymbol{b} = \emptyset$ if $\overline{a} < \underline{b}$ or $\overline{b} < \underline{a}$. Otherwise

$$\begin{aligned} \boldsymbol{a} \cap \boldsymbol{b} &= [\max(\underline{a}, \underline{b}),\ \min(\overline{a}, \overline{b})], \\ \boldsymbol{a} \cup \boldsymbol{b} &= \{x \in \boldsymbol{a} \vee x \in \boldsymbol{b}\}. \end{aligned}$$

Since the result of the operation $\cup$ is not always a single interval we define the hull as

$$\square(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a} \sqcup \boldsymbol{b} = [\min(\underline{a}, \underline{b}),\ \max(\overline{a}, \overline{b})].$$

Note that for the hull we use two different notations, that can be interchanged. Generally, the hull is understood as the interval of the minimal width containing the sets $\boldsymbol{a}$ and $\boldsymbol{b}$. The set operations can be easily extended to take more intervals as arguments.

3.4 Interval arithmetics

An arithmetics can be defined on intervals. We are going to use a standard definition mentioned in, e.g, [133]. What we need from an arithmetical operation $\circ$ on two intervals $\boldsymbol{a}, \boldsymbol{b}$ is

$$\boldsymbol{a} \circ \boldsymbol{b} = \square\{a \circ b \mid a \in \boldsymbol{a}, b \in \boldsymbol{b}\}.$$

The following definition of the basic operations satisfies such a demand.

Definition 3.4 (Interval arithmetics)**.** Let us have two intervals $\boldsymbol{a} = [\underline{a}, \overline{a}]$ and $\boldsymbol{b} = [\underline{b}, \overline{b}]$. Arithmetical operations $+, -, \cdot, /$ are defined as

$$\begin{aligned}
\boldsymbol{a} + \boldsymbol{b} &= [\underline{a} + \underline{b}, \overline{a} + \overline{b}],\\
\boldsymbol{a} - \boldsymbol{b} &= [\underline{a} - \overline{b}, \overline{a} - \underline{b}],\\
\boldsymbol{a} \cdot \boldsymbol{b} &= [\min(M), \max(M)], \quad \text{where } M = \{\underline{a}\underline{b},\, \underline{a}\overline{b},\, \overline{a}\underline{b},\, \overline{a}\overline{b}\},\\
\boldsymbol{a}/\boldsymbol{b} &= \boldsymbol{a} \cdot (1/\boldsymbol{b}), \quad \text{where } 1/\boldsymbol{b} = [1/\overline{b}, 1/\underline{b}],\ 0 \notin \boldsymbol{b}.
\end{aligned}$$

In the definition of division we presume $\boldsymbol{b}$ does not contain 0. When we need to divide with intervals containing zero, an extended version of interval arithmetics can be used [114, 155].

The set $\mathbb{IR}$ with the defined interval arithmetics does not form a field. Only some properties of a field hold. There exist distinct zero element $\mathbf{0} = [0, 0]$ and unit element $\mathbf{1} = [1, 1]$ (we will denote them just 0 and 1 respectively). Moreover, for all $\boldsymbol{a} \in \mathbb{IR}$ it holds that

$$\begin{aligned}
0 + \boldsymbol{a} &= \boldsymbol{a},\\
1 \cdot \boldsymbol{a} &= \boldsymbol{a},\\
0 \cdot \boldsymbol{a} &= 0.
\end{aligned}$$

By definition, the addition and multiplication are commutative and associative.

$$\begin{aligned}
\boldsymbol{x} + \boldsymbol{y} &= \boldsymbol{y} + \boldsymbol{x}, & \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) &= (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z},\\
\boldsymbol{x}\boldsymbol{y} &= \boldsymbol{y}\boldsymbol{x}, & \boldsymbol{x}(\boldsymbol{y}\boldsymbol{z}) &= (\boldsymbol{x}\boldsymbol{y})\boldsymbol{z}.
\end{aligned}$$

Unfortunately, there is no inverse element with respect to addition and multiplication.

Proposition 3.5. *For a nondegenerate interval $\boldsymbol{a} = [\underline{a}, \overline{a}]$ there does not exist an inverse element with respect to addition.*

Proof. Let $\boldsymbol{a}$ be a nondegenerate interval and let $\boldsymbol{b}$ be its inverse element with respect to addition. According to the definition of the zero interval we get

$$\mathbf{0} = [0, 0] = \boldsymbol{a} + \boldsymbol{b} = [\underline{a}, \overline{a}] + [\underline{b}, \overline{b}] = [\underline{a} + \underline{b}, \overline{a} + \overline{b}].$$

Thus $\underline{a} + \underline{b} = 0$ and $\overline{a} + \overline{b} = 0$. It follows that

$$\underline{b} = -\underline{a},\ \overline{b} = -\overline{a}.$$

Hence

$$\boldsymbol{b} = [-\underline{a}, -\overline{a}].$$

For the bounds of the interval $\boldsymbol{a}$ it holds that $\underline{a} \leq \overline{a}$. However, the bounds of $\boldsymbol{b}$ are contradiction to the definition of interval since $-\underline{a} \geq -\overline{a}$. □

According to the definition of an interval the inverse element with respect to addition exists only for a degenerate interval. The proof for nonexistence of inverse element with respect to multiplication can be provided similarly, but requires more tedious elaboration.

Moreover, the distributivity does not hold either. Generally,

$$\boldsymbol{a}(\boldsymbol{b} + \boldsymbol{c}) \neq \boldsymbol{ab} + \boldsymbol{ac}.$$

Example 3.6. For intervals $\boldsymbol{a} = [1, 2], \boldsymbol{b} = [1, 1]$ and $\boldsymbol{c} = [-1, -1]$ we obtain the following results.

$$\begin{aligned} \boldsymbol{a}(\boldsymbol{b} + \boldsymbol{c}) &= [0, 0], \\ \boldsymbol{ab} + \boldsymbol{ac} &= [-1, 1]. \end{aligned}$$

However, the *subdistributivity* always holds

$$\boldsymbol{a}(\boldsymbol{b} + \boldsymbol{c}) \subseteq \boldsymbol{ab} + \boldsymbol{ac}.$$

Such an overestimation caused by the second formula is a result of a so-called *dependency problem*. Whenever real number is chosen from $\boldsymbol{a}$ the same value should be fixed for the second occurrence of the second $\boldsymbol{a}$. However, the interval arithmetics does not see both $\boldsymbol{a}$'s as one and the same variable, but rather as two different variables. We will touch dependency in Section 3.8 in more detail.

3.5 Relations

Basic relations of intervals can be defined in the following way.

Definition 3.7 (Relations)**.** For two intervals $\boldsymbol{a} = [\underline{a}, \overline{a}]$ and $\boldsymbol{b} = [\underline{b}, \overline{b}]$.
The relation $\boldsymbol{a} = \boldsymbol{b}$ holds if

$$\underline{a} = \underline{b} \quad \text{and} \quad \overline{a} = \overline{b}.$$

The relation $\boldsymbol{a} \leq \boldsymbol{b}$ holds if

$$\overline{a} \leq \underline{b}.$$

The relation $\boldsymbol{a} < \boldsymbol{b}$ holds if

$$\overline{a} < \underline{b}.$$

Similarly for the relations $\geq, <, >$.

Note, that some intervals are incomparable, e.g.,

$$[1, 3] \not\leq [2, 4] \quad \text{and} \quad [1, 3] \not\geq [2, 4].$$

3.6 More interval notation

Regarding intervals we need to define more notation:

notion	formula	explanation
$\mathrm{wid}(\boldsymbol{a})$	$\overline{a} - \underline{a}$	width of an interval
$\mathrm{mid}(\boldsymbol{a})$	$a_c = (\underline{a} + \overline{a})/2$	midpoint of an interval
$\mathrm{rad}(\boldsymbol{a})$	$\mathrm{wid}(\boldsymbol{a})/2$	radius of an interval
$\mathrm{mig}(\boldsymbol{a})$	$\min(\lvert\underline{a}\rvert, \lvert\overline{a}\rvert)$ or 0 when $0 \in \boldsymbol{a}$	mignitude of an interval
$\mathrm{mag}(\boldsymbol{a})$	$\max(\lvert\underline{a}\rvert, \lvert\overline{a}\rvert)$	magnitude of an interval
$\lvert\boldsymbol{a}\rvert$	$\{\lvert a\rvert,\ a \in \boldsymbol{a}\}$	absolute values of an interval

Note the difference between the absolute value and magnitude. Sometimes these two notions are used interchangably. Nevertheless, here in our work, we are going to strictly distinguish between them. The magnitude of an interval is a number while the absolute value of an interval is an interval:

$$\begin{aligned} \text{If } \boldsymbol{a} > 0 \quad & |\boldsymbol{a}| = [\underline{a}, \overline{a}]\,, \\ \text{If } \boldsymbol{a} < 0 \quad & |\boldsymbol{a}| = [|\overline{a}|, |\underline{a}|]\,, \\ \text{If } 0 \in \boldsymbol{a} \quad & |\boldsymbol{a}| = [0, \max\{|\underline{a}|, |\overline{a}|\}]\,. \end{aligned}$$

For many important properties of the introduced functions and operations see [139].

3.7 Vectors and matrices

Intervals can be used as building blocks for more complex structures. In this section we address interval vectors and matrices. An interval matrix (or an interval vector as its special case) can be defined as a matrix having intervals as its coefficients.

Definition 3.8 (Interval vector and matrix)**.** Let $\boldsymbol{b}_i, \boldsymbol{a}_{ij}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$ be intervals then an m-dimensional interval vector $\boldsymbol{b}$ and an $m \times n$ interval matrix $\boldsymbol{A}$ are defined as

$$\boldsymbol{b} = \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_m \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_{11} & \ldots & \boldsymbol{a}_{1n} \\ \vdots & & \vdots \\ \boldsymbol{a}_{m1} & \ldots & \boldsymbol{a}_{mn} \end{pmatrix},$$

When we talk about *square* matrices, we always assume the size of $\boldsymbol{A}$ is $n \times n$. Otherwise, the size $m \times n$ is assumed. Note that an n-dimensional interval vector actually represents an n-dimensional box aligned with axes. That is why we use the phrases “interval vector” and “interval box” interchangeably.

The relations $=, \leq, \geq, <, >, \subseteq, \in$ are understood component-wise. So are the set operations $\cup, \cap, \square, \sqcup$. Hence an $m \times n$ interval matrix can be also defined using two real $m \times n$ matrices $\underline{A}, \overline{A}$ as

$$\boldsymbol{A} = \{A \mid \underline{A} \leq \overline{A}\}.$$

Formally, it is slightly different from Definition 3.8, however it is simple to transit between the two points of view. We can also define an interval matrix using its midpoint A_c and radius A_Δ matrix as

$$\boldsymbol{A} = [A_c - A_\Delta, A_c + A_\Delta].$$

For the sake of concise notation, when speaking about $\boldsymbol{A}$ we always implicitly assume that $\underline{A}, \overline{A}$ are its lower and upper bound respectively and that A_c, A_Δ are its midpoint and radius respectively.

For two interval matrices $\boldsymbol{A}, \boldsymbol{B}$ of the same size the interval arithmetics operations $+$ and $-$ are performed component-wise as

$$(\boldsymbol{A} + \boldsymbol{B})_{ij} = \boldsymbol{a}_{ij} + \boldsymbol{b}_{ij},$$
$$(\boldsymbol{A} - \boldsymbol{B})_{ij} = \boldsymbol{a}_{ij} - \boldsymbol{b}_{ij}.$$

For an $m \times n$ matrix $\boldsymbol{A}$ and an $n \times p$ matrix $\boldsymbol{B}$ the matrix multiplication $\boldsymbol{AB}$, can be carried out similarly as in the case of real matrices.

$$(\boldsymbol{AB})_{ij} = \sum_{k=1}^{n} \boldsymbol{a}_{ik}\boldsymbol{b}_{kj}.$$

Even though, the result gives sharp bounds on the matrix product, it can contain matrices that cannot be obtained by any product of $A \in \boldsymbol{A}, B \in \boldsymbol{B}$. Here is an example from [133].

Example 3.9. For two matrices

$$\boldsymbol{A} = \begin{pmatrix} [1,2] & [3,4] \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} [5,6] & [7,8] \\ [9,10] & [11,12] \end{pmatrix}$$

the product is $\boldsymbol{AB} = \begin{pmatrix} [32,52] & [40,64] \end{pmatrix}$. Let us take the matrix $\begin{pmatrix} 32 & 64 \end{pmatrix}$; the element 32 is obtained by multiplying $\underline{A}$ by lower bound of the right column of $\boldsymbol{B}$ and the element 64 is obtained by multiplying $\overline{A}$ by upper bound of the right column of $\boldsymbol{B}$.

The operation $+$ is for interval matrices commutative and associative. There are cases when associativity of multiplication multiplication fails [139]. As in the case of intervals, for matrices we again get subdistributivity [139].

$$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) \subseteq \boldsymbol{AB} + \boldsymbol{AC},$$
$$(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} \subseteq \boldsymbol{AB} + \boldsymbol{AC}.$$

The already mentioned functions and operations $\text{wid}(\cdot)$, $\text{mid}(\cdot)$, $\text{rad}(\cdot)$, $\text{mig}(\cdot)$, $\text{mag}(\cdot)$ and $|\cdot|$ are for interval vectors and matrices understood component-wise. They posses several useful properties:

$$\operatorname{mag}(\boldsymbol{A}) = |A_c| + A_\Delta, \tag{3.1}$$
$$\operatorname{mid}(\boldsymbol{A} \pm \boldsymbol{B}) = A_c \pm B_c, \tag{3.2}$$
$$\operatorname{mid}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{mid}(\boldsymbol{A})\operatorname{mid}(\boldsymbol{B}), \quad \text{if } \boldsymbol{A} \text{ or } \boldsymbol{B} \text{ is thin}, \tag{3.3}$$
$$\operatorname{rad}(\boldsymbol{A}\boldsymbol{B}) = |A|\operatorname{rad}(\boldsymbol{B}) \quad \text{if } \boldsymbol{A} \text{ is thin or } B_c = 0, \tag{3.4}$$
$$\operatorname{rad}(\boldsymbol{A} + \boldsymbol{B}) = \operatorname{rad}(\boldsymbol{A}) + \operatorname{rad}(\boldsymbol{B}). \tag{3.5}$$

Interval version of other notation such as $\varrho(\cdot)$ or $(\cdot)^{-1}$ will be introduced in the corresponding chapters later when needed. Next, we introduce the useful concept of interval matrix norms.

Definition 3.10 (Interval matrix norm)**.** For interval matrices a matrix norm $\|\cdot\|$ can be defined as

$$\|\boldsymbol{A}\| = \max\{\|A\|,\ A \in \boldsymbol{A}\}.$$

Regarding the computation of matrix norms, there are easily computable matrix norms:

$$\|\boldsymbol{A}\|_1 = \max_j \sum_i \operatorname{mag}(\boldsymbol{A}_{ij}),$$
$$\|\boldsymbol{A}\|_\infty = \max_i \sum_j \operatorname{mag}(\boldsymbol{A}_{ij}).$$

Furthermore, we can use a so-called *scaled maximum norm* as a generalization of the maximum norm $\|\cdot\|_\infty$. For any vector $\boldsymbol{x} \in \mathbb{IR}^n$ and a vector $0 < u \in \mathbb{R}^n$ we define

$$\|\boldsymbol{x}\|_u := \max\{\operatorname{mag}(\boldsymbol{x}_i)/u_i \mid i = 1, \ldots, n\},$$

and

$$\|\boldsymbol{A}\|_u := \|\operatorname{mag}(\boldsymbol{A})u\|_u.$$

Note that for $u = (1, \ldots, 1)^T$ we get the maximum norm. The following holds for such a norm [139]

$$\|\boldsymbol{A}\|_u < \alpha \iff \operatorname{mag}(\boldsymbol{A})u < \alpha u, \tag{3.6}$$
$$\|\boldsymbol{A}\|_u \le \alpha \iff \operatorname{mag}(\boldsymbol{A})u \le \alpha u. \tag{3.7}$$

In the further text, many of our results will be in terms of matrix norms. We will use only *consistent* matrix norms, i.e, those that satisfy

$$\|\boldsymbol{A} \cdot \boldsymbol{x}\| \le \|\boldsymbol{A}\| \cdot \|\boldsymbol{x}\|.$$

All the mentioned norms satisfy this property [126, 139]. Note all norms were defined for interval matrices. To define $\|\cdot\|_1, \|\cdot\|_\infty, \|\cdot\|_u$ for real matrices it is enough to replace $\operatorname{mag}(\cdot)$ with $|\cdot|$.

3.8 Interval expressions and functions

One of the important tasks is to enclose the range of a real-valued function. This section is loosely inspired by [139] and [215]. Let us consider a function $f : D \mapsto \mathbb{R}$, where $D \subseteq \mathbb{R}^n$, the range is then

$$f(D) = \{f(x) \mid x \in D\}.$$

For a monotone (or piece-wise monotone) function the range can be expressed exactly. Elementary functions such as $\cos(x), \sin(x), |x|, a^x, \log(x)$ satisfy this property. We can use these functions as building blocks for more complex functions.

Generally, we want to extend a real-valued function f to an interval function $\boldsymbol{f}$

$$\boldsymbol{f} : \mathbb{IR}^n \mapsto \mathbb{IR}.$$

Such a generalization should poses some favorable characteristics. First, it would be useful if

$$\boldsymbol{f}(x) = f(x), \ \forall x \in D.$$

Here, x can be seen as a degenerate interval vector. Or an interval function should at least satisfy

$$f(x) \in \boldsymbol{f}(\boldsymbol{x}), \ \text{for } x \in \boldsymbol{x} \subseteq \square D.$$

Such a function is called *interval extension.* Another favorable property is *inclusion monotonicity*, i.e.,

$$\boldsymbol{x} \subseteq \boldsymbol{y} \Rightarrow \boldsymbol{f}(\boldsymbol{x}) \subseteq \boldsymbol{f}(\boldsymbol{y}).$$

For a function a *natural interval extension* can be obtained by viewing its variables as intervals and its operators/subfunctions as interval operators/subfunctions. In [131] we can find the following theorem by Moore.

Theorem 3.11. *The natural interval extension $\boldsymbol{f}$ associated with a real function f that is a combination of only constants, variables, arithmetical operations and elementary functions (*$\sin(x), \cos(x), |x|, a^x, \log(x) \ldots$*) is an inclusion monotone interval extension such that*

$$\{f(x) \mid x \in \boldsymbol{x}\} \subseteq \boldsymbol{f}(\boldsymbol{x}),$$

for any $\boldsymbol{x}$, where $\boldsymbol{f}(\boldsymbol{x})$ is defined.

Note that there is a difference between $f(\boldsymbol{x})$ and $\boldsymbol{f}(\boldsymbol{x})$. The first denotes computation of range of a function over $\boldsymbol{x}$ and the second is an interval function. The following example demonstrates that not every interval extension is narrow.

Example 3.12. For $x \in \boldsymbol{x} = [-1, 2]$ compare the following:

$$\begin{aligned} \boldsymbol{x} - \boldsymbol{x} &= [-3, 1] \quad \text{vs.} \quad 0, \\ \boldsymbol{x} \cdot \boldsymbol{x} &= [-2, 4] \quad \text{vs.} \quad \boldsymbol{x}^2 = [0, 4]. \end{aligned}$$

The previous examples suffered from a so-called *dependency problem* – the interval arithmetic as is defined does not see the double occurrence of $\boldsymbol{x}$ and treats both occurrences as separate variables. We have actually met this phenomenon when talking about subdistributivity of interval arithmetic operations or interval matrix multiplication. Not surprisingly, we have the following theorem by Moore [131].

Theorem 3.13. *Let $f(x_1, \ldots, x_n, y_1, \ldots, y_m)$ be an real function from the previous theorem with $n + m$ variables and let $\boldsymbol{f}$ be its natural extension. Suppose that the variables $y_1, \ldots, y_m$ occur only once in f. Then*

$$\square\{f(x, y) \mid x \in \boldsymbol{x},\, y \in \boldsymbol{y}\} = \bigcup_{x \in \boldsymbol{x}} \boldsymbol{f}(x, \boldsymbol{y}),$$

for $(\boldsymbol{x}, \boldsymbol{y})$ where $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})$ is defined.

Especially, if each variable in arithmetical expression occurs only once, then the following holds.

$$\boldsymbol{f}(\boldsymbol{x}) = \square\{f(x) \mid x \in \boldsymbol{x}\}.$$

There are many other methods for enclosing the range of $f(x)$ for $x \in \boldsymbol{x}$. One way is to use the mean value form

$$f(x) = f(x_c) + f'(\psi)(x - x_c),$$

where ψ lies on a line segment between x and x_c. Since $\psi \in \boldsymbol{x}$ we have the interval extension

$$\boldsymbol{f}(\boldsymbol{x}) \subseteq f(x_c) + f'(\boldsymbol{x})(\boldsymbol{x} - x_c).$$

The function f' is a gradient of f. Its range can be estimated in various ways e.g., using slopes [139]. We are going to use such methods in Chapter 10. For more on range of real-valued functions and polynomials see, e.g., [46, 134, 194, 196].

3.9 Rounded interval arithmetic

Now, we briefly touch the topic that we will address only rarely in the text. It is a well-known fact that computers cannot represent all numbers from $\mathbb{R}$. Let us denote the set of machine representable numbers by $\mathbb{R}^{pc}$. When a number cannot be represented it is necessary to round it to some representable number. Speaking of rounding procedure for a real number a, we are interested in the two main types: rounding to $+\infty$ ($\uparrow a$) and rounding to $-\infty$ ($\downarrow a$).

These roundings preserve the property of the relation $\leq$ on $\mathbb{R}$. Thus, for $a, b \in \mathbb{R}$

$$a \leq b \Rightarrow \uparrow a \leq \uparrow b.$$

Also for every $a \in \mathbb{R}$ it holds that

$$\uparrow\uparrow a = \uparrow a.$$

We have already defined the set $\mathbb{IR}$ as a set of real closed intervals. We can also define the set $\mathbb{IR}^{pc}$, the set of real closed intervals with machine representable endpoints. We can switch from $\mathbb{IR}$ to $\mathbb{IR}^{pc}$ with use of directed rounding:

$$[a,b] \in \mathbb{IR} \quad \mapsto \quad [\downarrow a, \uparrow b] \in \mathbb{IR}^{pc}.$$

Such an implementation needs switching of rounding mode. If a being machine representable implies $-a$ is also machine representable, then only one directed rounding is enough:

$$[\downarrow a, \uparrow b] = [\downarrow a, -\downarrow -b] = [-\uparrow -a, \uparrow b].$$

The interval arithmetics can be defined also on $\mathbb{IR}^{pc}$. For example addition of two intervals $\boldsymbol{a} = [\underline{a}, \overline{a}], \boldsymbol{b} = [\underline{b}, \overline{b}] \in \mathbb{IR}^{pc}$ can be defined as

$$\boldsymbol{a} +^{pc} \boldsymbol{b} = [\downarrow(\underline{a} + \underline{b}), \uparrow(\overline{a} + \overline{b})].$$

In the following text, we will keep working with $\mathbb{IR}$, however, we keep in mind that to obtain verified results algorithms must be implemented via $\mathbb{IR}^{pc}$. That is, when we talk about the hull or enclosure we implicitly assume that its end-points are machine representable.

There are many packages that handle computing with $\mathbb{IR}^{pc}$, e.g., Intlab for Matlab and Octave [188], Octave interval package [62], `libieee1788` for C++ [135] and many others [113]. However, not all of them conform to the interval arithmetics standard IEEE 1788-2015 [162]. More on rounding and interval arithmetics can be found in, e.g. [132, 133, 139].

3.10 Comparing quality of interval results

In this work we need to compare intervals or interval vectors (boxes) returned by various methods. If two methods `a` and `b` return single intervals $\boldsymbol{a}$ and $\boldsymbol{b}$ respectively (e.g., methods for computing the determinant of an interval matrix), the returned solutions can be compared as

$$rat(\boldsymbol{a}, \boldsymbol{b}) = \frac{\mathrm{wid}(\boldsymbol{a})}{\mathrm{wid}(\boldsymbol{b})}. \tag{3.8}$$

If the two methods return n-dimensional interval vectors $\boldsymbol{a} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ and $\boldsymbol{b} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ respectively, their quality is compared as the average ratio of widths of the corresponding elements

$$\frac{\sum_{i=1}^{n} rat(\boldsymbol{a}_i, \boldsymbol{b}_i)}{n}. \tag{3.9}$$

Only rarely will it be compared as

$$\frac{\sum_{i=1}^{n} \mathrm{wid}(\boldsymbol{a}_i)}{\sum_{i=1}^{n} \mathrm{wid}(\boldsymbol{b}_i)}. \tag{3.10}$$

In each comparison we use a reference method, i.e., a method to which other methods are compared. In the text, the previous formulas are used in the following way. The method `b` (the second one) is always the reference method and `a` is the method compared to it. Hence, if the ratio is > 1, then the method `a` is worse than `b`, if the ratio is < 1, then the intervals returned by `a` are tighter than the ones by `b`.

3.11 How we test

Most of the chapters need to compare more methods for solving a certain interval problem. Features of each method can be demonstrated by special cases (particular interval matrix or system, etc.). Nevertheless, to compare methods more thoroughly, we test them on larger sets of random problems. Of course, the problems in real applications are not exactly random, however, in some cases the testing on random systems gives us a hint about the natural behavior of the methods.

If not stated otherwise, the tests are computed using the two settings:

1. DESKTOP – computationally demanding tests run on a desktop machine with 8-CPU machine Intel(R) Core(TM) i7-4790K, 4.00GHz, 15937 MB RAM, Octave 4.0.3., Octave Interval package 3.0.0.
2. LAPTOP – computationally not so exhaustive tests are executed on laptop with Intel Core i5-7200U – 2.5GHz, TB 3.1GHz, HyperThreading; 8GB DDR4 memory. Octave 4.2.2, Octave Interval package 3.2.0.

Most of the methods tested here are implemented in interval toolbox LIME (see more in Chapter 12), which is built on Oliver Heimlich's [62] interval package for Octave.

4 Interval matrix ZOO

- M-matrices and inverse nonnegative matrices
- Strictly diagonally dominant matrices
- H-matices
- Regular and strongly regular matrices
- Relations between matrix classes
- Other types of matrices

We defined general interval matrices in the previous chapter. However, there is a large variety of special types of interval matrices. Many of them emerged as generalization of notions from the classical linear algebra. Nevertheless, in this chapter we focus only on interval matrices. For more insight into real matrices see, for example, the works [19, 41, 91, 150].

Some of the classes of interval matrices have favorable properties (easily computable inverse, regularity etc.) and algorithms usually work well for them. We feel the necessity of characterizing the distinct classes of interval matrices, their features and links between them, since we believe it would increase the understanding of the rest of the work. This chapter is loosely based on Chapter 3 and 4 from Neumaier's book [139]. We focus on the most common types of matrices that are usually used in connection with quality of solving interval linear systems and related problems. In this short chapter we re-structure Neumaier's material and add some new examples and comments to make the relations between the classes of matrices more visible and clear. The final Figure 4.1 illustrates the relationships between the classes of interval matrices.

4.1 Regular matrices

Regular matrices are of special importance.

Definition 4.1 (Regular matrix)**.** A square interval matrix $\boldsymbol{A}$ is *regular* if every $A \in \boldsymbol{A}$ is nonsingular.

Note that there is a slight terminology confusion in addressing the similar quality of real and interval matrices.

real matrices	$\rightarrow$	nonsingular
interval matrices	$\rightarrow$	regular

The interval matrices that are not regular are called *singular* as in the case of real matrices.

In Chapter 11 we show that checking regularity is generally a coNP-complete problem. There exist a lot of sufficient and necessary conditions for regularity of interval matrices [179]. All of them are of exponential nature.

Fortunately, there are some polynomially computable sufficient conditions and not explicitly exponential algorithms for checking regularity [38, 96, 163, 164]. The following useful condition is from [164].

Theorem 4.2. *A square interval matrix $\boldsymbol{A}$ is regular if for some real matrix R the following condition holds*

$$\varrho(|I - RA_c| + |R|A_\Delta) < 1.$$

Particularly, if A_c is regular, then for $R = A_c^{-1}$ the condition reads $\varrho(|A_c^{-1}|A_\Delta) < 1$.

It can be shown that if the first condition holds for some R then

$$\varrho(|A_c^{-1}|A_\Delta) \leq \varrho(|I - RA_c| + |R|A_\Delta),$$

which makes the midpoint inverse a kind of optimal choice. Later we use the following simple consequence of Theorem 4.2.

Corollary 4.3. *A square interval matrix $\boldsymbol{A}$ with $A_c = I$ is regular if*

$$\varrho(A_\Delta) < 1.$$

4.2 M-matrices

In many applications matrices of a special form, called Z-matrices, appear [7, 18, 19, 39].

Definition 4.4 (Z-matrix)**.** A square real matrix A is called a *Z-matrix* if $a_{ij} \leq 0$ for every $i \neq j$. A square interval matrix $\boldsymbol{A}$ is called a *Z-matrix*, if every $A \in \boldsymbol{A}$ is a Z-matrix.

By adding more restriction to Z-matrices we obtain an important subclass of interval matrices.

Definition 4.5 (M-matrix)**.** An interval matrix $\boldsymbol{A}$ is an *M-matrix* if it is a Z-matrix and there exists $0 < u \in \mathbb{R}^n$ such that $\boldsymbol{A}u > 0$ (understood component-wise).

According to [150] the term "M-matrix" was first used by Ostrowski in [146] where he studied such matrices extensively. They are often connected to various problems in mathematics, biology, physics, etc. For more applications and properties of real M-matrices one can see [19, 41, 150]. Another feature of M-matrices is computational, since many algorithms behave well when working with an M-matrix.

Before stating the equivalent characterization of M-matrices, we need to specify what do we mean by an inverse interval matrix, a principal minor and a P-matrix.

Definition 4.6 (Inverse interval matrix)**.** Let us have a regular interval matrix $\boldsymbol{A}$. We define its interval inverse matrix $\boldsymbol{A}^{-1}$ as

$$\boldsymbol{A}^{-1} = [\underline{B}, \overline{B}],$$

$$\underline{B}_{ij} = \min\{(A^{-1})_{ij} \mid A \in \boldsymbol{A}\},$$
$$\overline{B}_{ij} = \max\{(A^{-1})_{ij} \mid A \in \boldsymbol{A}\},$$

for $i, j = 1, \ldots, n$.

Definition 4.7 (Principal minor)**.** For a square matrix a *principal matrix* occurs when deleting some rows of the matrix and also the corresponding columns with the same indices. A determinant of a principal matrix is called a *principal minor*.

Definition 4.8 (P-matrix)**.** A square real matrix is a *P-matrix* if its every principal minor is positive. A square interval matrix $\boldsymbol{A}$ is a *P-matrix* if every $A \in \boldsymbol{A}$ is a P-matrix.

Theorem 4.9. *The following statements are equivalent*

1. $\boldsymbol{A}$ *is an M-matrix,*
2. *every* $A \in \boldsymbol{A}$ *is an M-matrix,*
3. $\underline{A}, \overline{A}$ *are M-matrices,*
4. $\boldsymbol{A}$ *is a regular Z-matrix and* $\boldsymbol{A}^{-1} = [\overline{A}^{-1}, \underline{A}^{-1}] \geq 0$*,*
5. $\boldsymbol{A}$ *is a Z-matrix and P-matrix.*

The statements *1.–4.* come from [139]. The statement *2.* implies that if $\boldsymbol{A}$ is an M-matrix and $\boldsymbol{B} \subseteq \boldsymbol{A}$, then $\boldsymbol{B}$ is also an M-matrix. From *4.* we can see that M-matrices are inverse nonnegative matrices (see the next section). The statement *5.* is a simple generalization of the similar claim for real matrices [150]. To check that a matrix is an M-matrix, the statement *4.* gives a hint how to find a positive vector u proving that $\boldsymbol{A}$ is an M-matrix. First, solve the system $\underline{A}u = e$. Because $\boldsymbol{A}^{-1}$ should be nonnegative, for the solution u it should hold that $u = \underline{A}^{-1}e > 0$. Second, check whether $\boldsymbol{A}u > 0$. The check needs to be performed in a verified way. It is also possible to exploit the statement *3.* If both $\underline{A}, \overline{A}$ are Z-matrices and their verified inverse is nonnegative, then $\boldsymbol{A}$ is an M-matrix. From *4.* it can be seen that M-matrices are regular. For a more detailed proof see e.g., [139] Regarding *5.*, computation of determinants of interval matrices can be used. For example, a tight enclosure of a determinant of a 2×2 interval matrix can be expressed as

$$\det \begin{pmatrix} \boldsymbol{a}_{11} & \boldsymbol{a}_{12} \\ \boldsymbol{a}_{21} & \boldsymbol{a}_{22} \end{pmatrix} = \boldsymbol{a}_{11} \cdot \boldsymbol{a}_{22} - \boldsymbol{a}_{12} \cdot \boldsymbol{a}_{21}.$$

This topic is further elaborated in Chapter 8.

Example 4.10. Let us have the matrix

$$\boldsymbol{A} = \begin{pmatrix} 2 & -1 \\ [-2, 0] & 2 \end{pmatrix}.$$

Clearly $\boldsymbol{A}$ is a Z-matrix. Furthermore, $\boldsymbol{A}$ is an M-matrix since all principal minors are positive $\det(\boldsymbol{A}_1) = 2, \det(\boldsymbol{A}_2) = 2, \det(\boldsymbol{A}_{12}) = [2, 4]$.

Example 4.11. Let us show another way to prove that the Z-matrix $\boldsymbol{A}$ from the previous example is an M-matrix. For

$$u = \begin{pmatrix} 1 \\ 1.5 \end{pmatrix} \text{ we see that } \boldsymbol{A}u = \begin{pmatrix} 0.5 \\ [1, 3] \end{pmatrix} > 0.$$

4.3 Inverse nonnegative matrices

From previous section it is already known that every M-matrix has a nonnegative inverse. M-matrices are part of a larger class of interval matrices.

Definition 4.12 (Inverse nonnegative)**.** A square interval matrix $\boldsymbol{A}$ is called *inverse nonnegative* if $\boldsymbol{A}$ is regular and $\boldsymbol{A}^{-1} \geq 0$.

For such a class of matrices there is a theorem by Kuttler in [117] which gives us explicit bounds on a matrix inverse.

Theorem 4.13 (Kuttler)**.** *Let $\boldsymbol{A}$ be an interval matrix. If its lower and upper bounds $\underline{A}, \overline{A}$ are regular and $\underline{A}^{-1}, \overline{A}^{-1} \geq 0$ then $\boldsymbol{A}$ is regular and*

$$\boldsymbol{A}^{-1} = [\overline{A}^{-1}, \underline{A}^{-1}] \geq 0.$$

Example 4.14. If we take the already known matrix

$$\boldsymbol{A} = \begin{pmatrix} 2 & -1 \\ [-2, 0] & 2 \end{pmatrix},$$

using the algebraic formula for inverse of a real 2×2 matrix we can inspect both inverses of $\underline{A}$ and $\overline{A}$

$$\underline{A}^{-1} = \begin{pmatrix} 1 & 0.5 \\ 1 & 1 \end{pmatrix}, \quad \overline{A}^{-1} = \begin{pmatrix} 0.5 & 0.25 \\ 0 & 0.5 \end{pmatrix},$$

and according to Kuttler's theorem we get

$$\boldsymbol{A}^{-1} = [\overline{A}^{-1}, \underline{A}^{-1}] = \begin{pmatrix} [0.5, 1] & [0.25, 0.5] \\ [0, 1] & [0.5, 1] \end{pmatrix},$$

which confirms that $\boldsymbol{A}$ is an inverse nonnegative matrix. Notice that $\boldsymbol{A}^{-1}$ is not regular since it contains, for example, the singular matrix

$$\begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}.$$

Example 4.15. According to Kuttler's theorem, the matrix

$$\boldsymbol{B} = \begin{pmatrix} -2 & 1 \\ [5, 6] & -2 \end{pmatrix}$$

has the inverse

$$\boldsymbol{B}^{-1} = \begin{pmatrix} [1, 2] & [0.5, 1] \\ [2.5, 5] & [1, 2] \end{pmatrix}.$$

which proves that $\boldsymbol{B}$ is inverse nonnegative, although it is not a Z-matrix (also not an M-matrix). Hence, not every inverse nonnegative matrix must be an M-matrix.

4.4 H-matrices

H-matrices are a generalization of M-matrices by lifting the condition on signs of matrix off-diagonal elements. The class of H-matrices inherits some favorable properties from M-matrices; regularity, for example (see [139]). We define an H-matrix using a comparison matrix.

Definition 4.16 (Comparison matrix)**.** For a square real matrix A its *comparison matrix* $\langle A \rangle$ is defined as

$$\langle A \rangle_{ii} = A_{ii},$$
$$\langle A \rangle_{ij} = -|A_{ij}| \quad \text{for } i \neq j.$$

For a square interval matrix $\boldsymbol{A}$ its *comparison matrix* $\langle \boldsymbol{A} \rangle$ is defined as

$$\langle \boldsymbol{A} \rangle_{ii} = \mathrm{mig}(\boldsymbol{A}_{ii}),$$
$$\langle \boldsymbol{A} \rangle_{ij} = -\,\mathrm{mag}(\boldsymbol{A}_{ij}) \quad \text{for } i \neq j.$$

Note that $\langle \boldsymbol{A} \rangle$ is forced to be a Z-matrix.

Definition 4.17 (H-matrix)**.** A square real matrix A is an *H-matrix* if $\langle A \rangle$ is an M-matrix. A square interval matrix $\boldsymbol{A}$ is an *H-matrix* if $\langle \boldsymbol{A} \rangle$ is an M-matrix.

Hence checking of H-matrix property can be transformed to checking M-matrix property. The following equivalent conditions can be found in Neumaier [139].

Theorem 4.18. *The following statements are equivalent:*

1. $\boldsymbol{A}$ *is an H-matrix,*
2. *every* $A \in \boldsymbol{A}$ *is an H-matrix,*
3. $\langle \boldsymbol{A} \rangle$ *is regular and* $\langle \boldsymbol{A} \rangle^{-1} e > 0$.

Example 4.19. If $\boldsymbol{A}$ is an M-matrix, then $\langle A \rangle = \underline{A}$, which is according to Theorem 4.9 also an M-matrix. Therefore, every M-matrix is also an H-matrix.

Example 4.20. The slightly changed matrix from Example 4.10

$$\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ [0, 2] & 2 \end{pmatrix}$$

is not an M-matrix, however $\langle \boldsymbol{A} \rangle = \begin{pmatrix} 2 & -1 \\ -2 & 2 \end{pmatrix}$ which is an M-matrix, hence $\boldsymbol{A}$ is an H-matrix because its inverse is

$$\langle \boldsymbol{A} \rangle^{-1} = \begin{pmatrix} 1 & 0.5 \\ 1 & 1 \end{pmatrix} \geq 0.$$

Example 4.21. Every regular lower or upper triangular matrix is an H-matrix [139].

Example 4.22. Every matrix that is sufficiently close to the identity matrix is also an H-matrix, i.e., every matrix that satisfies

$$\| I - \boldsymbol{A} \| < 1,$$

for some consistent matrix norm is an H-matrix [139].

Nevertheless, there exist inverse nonnegative matrices that are not H-matrices.

Example 4.23. The inverse nonnegative matrix from Example 4.15 is not even an H-matrix, because its comparison matrix is not an M-matrix (its determinant is -2).

4.5 Strictly diagonally dominant matrices

The condition for H-matrices $\langle \boldsymbol{A} \rangle u > 0$ can be rewritten for $u = (1, \dots, 1)^T$ as

$$\operatorname{mig}(\boldsymbol{a}_{ii}) > \sum_{k \neq i} \operatorname{mag}(\boldsymbol{a}_{ik}), \quad \text{for } i = 1, \dots, n. \tag{4.1}$$

Definition 4.24 (Strictly diagonally dominant matrix)**.** A square interval matrix $\boldsymbol{A}$ satisfying the condition 4.1 is called *strictly diagonally dominant*.

Clearly, according to its definition, every strictly diagonally dominant matrix is an H-matrix. Therefore it is also regular. Whenever a (preconditioned) matrix is close to the identity matrix then it is strictly diagonally dominant (and also an H-matrix).

Example 4.25. If $\|I - \boldsymbol{A}\|_\infty < 1$ then $\boldsymbol{A}$ is strictly diagonally dominant.

Example 4.26. The matrix

$$\boldsymbol{A} = \begin{pmatrix} 2 & [-1, 0] \\ [-1, 0] & 2 \end{pmatrix}$$

is strictly diagonally dominant and also an M-matrix (hence also inverse nonnegative).

Example 4.27. Not every strictly diagonally dominant matrix is an M-matrix. The strictly diagonally dominant matrix

$$\boldsymbol{A} = \begin{pmatrix} -2 & [0, 1] \\ [0, 1] & -2 \end{pmatrix}$$

is not an M-matrix because it is not a Z-matrix. Moreover, $\boldsymbol{A}$ is not inverse nonnegative since

$$\underline{A}^{-1} = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}^{-1} = \begin{pmatrix} -0.5 & 0 \\ 0 & -0.5 \end{pmatrix}.$$

Example 4.28. There exists an H-matrix that is not an M-matrix, not strictly diagonally dominant and not inverse nonnegative. The matrix

$$\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ [0, 2] & 2 \end{pmatrix}$$

is not an M-matrix (it is not a Z-matrix), but it is an H-matrix. It is clearly not strictly diagonally dominant. And since

$$\underline{A}^{-1} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 0.5 & -0.25 \\ 0 & 0.5 \end{pmatrix},$$

it is not inverse nonnegative.

Example 4.29. The slightly adapted matrix from Example 4.26

$$\boldsymbol{A} = \begin{pmatrix} 2 & [-1,0] \\ [-1,0] & 1 \end{pmatrix},$$

is an M-matrix, however it is not strictly diagonally dominant.

Example 4.30. There exists an H-matrix that is neither an M-matrix nor strictly diagonally dominant, but it is inverse nonnegative. The matrix

$$\boldsymbol{A} = \begin{pmatrix} [1, 1+\varepsilon] & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{for } \varepsilon > 0,$$

is not an M-matrix (because it is not a Z-matrix), but it is an H-matrix since

$$\langle \boldsymbol{A} \rangle^{-1} = \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \geq 0.$$

$\boldsymbol{A}$ is clearly not strictly diagonally dominant. It is inverse nonnegative since

$$\underline{A}^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \geq 0$$

and, according to the Sherman–Morison formula,

$$\overline{A}^{-1} = \begin{pmatrix} \frac{1}{1+\varepsilon} & \frac{1}{1+\varepsilon} & \frac{1}{1+\varepsilon} \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \geq 0.$$

Example 4.31. There exists an H-matrix, that is not an M-matrix and is both strictly diagonally dominant and inverse nonnegative. The matrix

$$\boldsymbol{A} = \begin{pmatrix} [11/30, 11/30+\varepsilon] & -0.1 & 1/30 \\ -0.1 & 0.3 & -0.1 \\ 1/30 & -0.1 & 11/30 \end{pmatrix}, \quad \text{for some } \varepsilon > 0,$$

is not an M-matrix (because it is not a Z-matrix). The matrix is clearly strictly diagonally dominant. It is an H-matrix since for its comparison matrix

$$\langle \boldsymbol{A} \rangle = \begin{pmatrix} 11/30 & -0.1 & -1/30 \\ -0.1 & 0.3 & -0.1 \\ -1/30 & -0.1 & 11/30 \end{pmatrix},$$