MAX-T LEARNING OF APPROXIMATE WEIGHT MATRICES FROM FUZZY DATA

Ismaïl Baaj

UNIVERSITÉ D'ARTOIS

For an inconsistent system of $\max - T$ fuzzy relational equations $A \Box_T^{\max} x = b$ (where A is a matrix, b a vector and T is a t-norm such as min):

Objectives

For an inconsistent system of $\max - T$ fuzzy relational equations $A \Box_T^{\max} x = b$ (where A is a matrix, b a vector and T is a t-norm such as min):

- Using the L_∞ norm, we compute by an explicit analytical formula the Chebyshev distance:

$$\Delta = \inf_{c \in \mathcal{C}} \|b - c\|$$

  where $\mathcal{C}$ is the set of second members of the consistent systems defined with the same matrix A

For an inconsistent system of $\max - T$ fuzzy relational equations $A \Box_T^{\max} x = b$ (where A is a matrix, b a vector and T is a t-norm such as min):

- Using the L_∞ norm, we compute by an explicit analytical formula the Chebyshev distance:

  $$\Delta = \inf_{c \in \mathcal{C}} \|b - c\|$$

  where $\mathcal{C}$ is the set of second members of the consistent systems defined with the same matrix A

- We give the structure of:
  - the set $\mathcal{C}_b$ of Chebyshev approximations of the second member b i.e., vectors $c \in \mathcal{C}$ such that $\|b - c\| = \Delta$
  - the approximate solutions set Λ_b of the system

  which are related in the following sense:

  an element of Λ_b is a solution vector x^* of a system $A \Box_T^{\max} x = c$ where $c \in \mathcal{C}_b$

Objectives

For an inconsistent system of $\max - T$ fuzzy relational equations $A \Box_T^{\max} x = b$ (where A is a matrix, b a vector and T is a t-norm such as min):

- Using the L_∞ norm, we compute by an explicit analytical formula the Chebyshev distance:
$$\Delta = \inf_{c \in \mathcal{C}} \|b - c\|$$
where $\mathcal{C}$ is the set of second members of the consistent systems defined with the same matrix A
- We give the structure of:
  - the set $\mathcal{C}_b$ of Chebyshev approximations of the second member b i.e., vectors $c \in \mathcal{C}$ such that $\|b - c\| = \Delta$
  - the approximate solutions set Λ_b of the system

  which are related in the following sense:

  an element of Λ_b is a solution vector x^* of a system $A \Box_T^{\max} x = c$ where $c \in \mathcal{C}_b$
- We introduce a paradigm for learning approximate weight matrices governed by max-T composition ; application in possibilistic learning

The work that is going to be presented is available with the proofs in the following two preprints:

- Baaj, Ismaïl. "Max-min Learning of Approximate Weight Matrices from Fuzzy Data." arXiv preprint arXiv:2301.06141 (2023) (submitted)
- Baaj, Ismaïl. "Chebyshev distances associated to the second members of Systems of Max-product/Lukasiewicz Fuzzy Relational Equations." arXiv preprint arXiv:2302.08554 (2023).

- Systems of $\max - T$ fuzzy relational equations play an important role in fuzzy modeling (numerous fuzzy models have been represented using it)

- Systems of $\max - T$ fuzzy relational equations play an important role in fuzzy modeling (numerous fuzzy models have been represented using it)
- The seminal work of (Sanchez 1977) led to the emergence of numerous AI applications:
  - He gave necessary and sufficient conditions for a system to be consistent
  - If the system is consistent, he showed there is a greater solution and a finite number of minimal solutions

- Systems of $\max - T$ fuzzy relational equations play an important role in fuzzy modeling (numerous fuzzy models have been represented using it)
- The seminal work of (Sanchez 1977) led to the emergence of numerous AI applications:
  - He gave necessary and sufficient conditions for a system to be consistent
  - If the system is consistent, he showed there is a greater solution and a finite number of minimal solutions
- However, addressing the inconsistency of these systems remains an open problem (raised as early as (Yager 1978) and still relevant today):
  - what are the "best" approximate solutions for an inconsistent system?
  - how to perturb as slightly as possible the second member of an inconsistent system to obtain a consistent system? (Pedrycz 1990)

Let:

- $A = [a_{ij}]_{1 \leq i \leq n, 1 \leq j \leq m} \in [0,1]^{n \times m}$ be a matrix
- $b = [b_i]_{1 \leq i \leq n} \in [0,1]^{n \times 1}$ be a vector of n components
- T is a t-norm among min, product or the one of Łukasiewicz
- $\mathcal{I}_T$ the residual implicator associated to the t-norm T (for min: *it is the Gödel implication* $\rightarrow_G$)

Let:

- $A = [a_{ij}]_{1 \leq i \leq n, 1 \leq j \leq m} \in [0,1]^{n \times m}$ be a matrix
- $b = [b_i]_{1 \leq i \leq n} \in [0,1]^{n \times 1}$ be a vector of n components
- T is a t-norm among min, product or the one of Łukasiewicz
- $\mathcal{I}_T$ the residual implicator associated to the t-norm T (for min: *it is the Gödel implication* $\rightarrow_G$)

The system of $\max - T$ fuzzy relational equations associated to (A, b) is:

$$(S) : A \Box_T^{\max} x = b$$

where:

- $x = [x_j]_{1 \leq j \leq m} \in [0,1]^{m \times 1}$ is an unknown vector of m components
- the operator $\Box_T^{\max}$ is the matrix product that uses the t-norm T as the product and max as the addition

Equivalently, the system can also be written as:

$$\max_{1 \leq j \leq m} T(a_{ij}, x_j) = b_i, \ \forall i \in \{1, 2, \ldots, n\}$$

- To check if the system (S) is consistent, we compute:

$$e = A^t \Box_{\mathcal{I}_T}^{\min} b,$$

where:

- A^t is the transpose of A
- the matrix product $\Box_{\mathcal{I}_T}^{\min}$ uses the residual implicator $\mathcal{I}_T$ as the product and min as the addition

- Thanks to Sanchez's seminal work for max − min composition and Pedrycz and Di Nola et al. for max − T composition, we have:

$$\text{the system } (S) \text{ is consistent} \Longleftrightarrow A \Box_T^{\max} e = b$$

The set of solutions of the system (S) is denoted by:

$$\mathcal{S} = \mathcal{S}(A, b) = \{v \in [0,1]^{m \times 1} \mid A \Box_T^{\max} v = b\}$$

If the system (S) is consistent, the vector e is its greatest solution and it has many minimal solutions

We have:

Lemma

The maps:

$$[0,1]^{m\times 1} \to [0,1]^{n\times 1} : x \mapsto A \Box_T^{\max} x,$$

$$[0,1]^{n\times 1} \to [0,1]^{m\times 1} : c \mapsto A^t \Box_{\mathcal{I}_T}^{\min} c$$

are increasing with respect to the usual order relation between vectors

and the following well-known result:

Lemma

Let $c, c' \in [0,1]^{n\times 1}$ *such that* $c \leq c'$ *then we have:*

$$\forall v \in [0,1]^{m\times 1}, A \Box_T^{\max} v = c \Longrightarrow v \leq A^t \Box_{\mathcal{I}_T}^{\min} c'$$

Let us construct a system of max − min fuzzy relational equations:

$$A = \begin{bmatrix} 0.06 & 0.87 & 0.95 \\ 0.75 & 0.13 & 0.88 \\ 0.82 & 0.06 & 0.19 \end{bmatrix} \text{ and } b = \begin{bmatrix} 0.4 \\ 0.7 \\ 0.7 \end{bmatrix}$$

$A^t = \begin{bmatrix} 0.06 & 0.75 & 0.82 \\ 0.87 & 0.13 & 0.06 \\ 0.95 & 0.88 & 0.19 \end{bmatrix}$. We compute the potential greatest solution:

$$e = A^t \Box_{\rightarrow_G}^{\min} b = \begin{bmatrix} \min(1.0, 0.7, 0.7) \\ \min(0.4, 1.0, 1.0) \\ \min(0.4, 0.7, 1.0) \end{bmatrix} = \begin{bmatrix} 0.7 \\ 0.4 \\ 0.4 \end{bmatrix}$$

where the Gödel implication $\rightarrow_G$ is the residual implicator associated to min defined by $x \rightarrow_G y = \begin{cases} 1 & \text{if } x \leq y \\ y & \text{if } x > y \end{cases}$

The system $A \Box_{\min}^{\max} x = b$ is consistent because:

$$A \Box_{\min}^{\max} e = \begin{bmatrix} 0.4 \\ 0.7 \\ 0.7 \end{bmatrix} = b$$

$\Delta = \inf_{c \in \mathcal{C}} \|b - c\|$ of a system $A \Box_T^{\max} x = b$ – *Application F*

For the system $(S) : A \Box_T^{\max} x = b$, we introduce the following application (Baaj 2023ab):

$$F : [0,1]^{n \times 1} \to [0,1]^{n \times 1} : c \mapsto F(c) = A \Box_T^{\max} (A^t \Box_{\mathcal{I}_T}^{\min} c) \tag{1}$$

$\Delta = \inf_{c \in \mathcal{C}} \|b - c\|$ of a system $A \Box_T^{\max} x = b$ – *Application F*

For the system $(S) : A \Box_T^{\max} x = b$, we introduce the following application (Baaj 2023ab):

$$F : [0,1]^{n \times 1} \to [0,1]^{n \times 1} : c \mapsto F(c) = A \Box_T^{\max} (A^t \Box_{\mathcal{I}_T}^{\min} c) \tag{1}$$

The application F allows us to check if a system is consistent (Reformulation of Sanchez's result):

Proposition

For any vector $c \in [0,1]^{n \times 1}$ the following conditions are equivalent:

1. $F(c) = c$,

2. the system $A \Box_T^{\max} x = c$ is consistent.

$\Delta = \inf_{c \in \mathcal{C}} \|b - c\|$ of a system $A \Box_T^{\max} x = b$ – *Application F*

For the system $(S) : A \Box_T^{\max} x = b$, we introduce the following application (Baaj 2023ab):

$$F : [0,1]^{n \times 1} \to [0,1]^{n \times 1} : c \mapsto F(c) = A \Box_T^{\max} (A^t \Box_{\mathcal{I}_T}^{\min} c) \tag{1}$$

The application F allows us to check if a system is consistent (Reformulation of Sanchez's result):

Proposition

For any vector $c \in [0,1]^{n \times 1}$ the following conditions are equivalent:
1. $F(c) = c$,
2. the system $A \Box_T^{\max} x = c$ is consistent.

The properties of the application F justify its introduction:

Proposition

- $\forall c \in [0,1]^{n \times 1}$, $F(c) \leq c$
- *F is idempotent i.e., $\forall c \in [0,1]^{n \times 1}, F(F(c)) = F(c)$*
- *F is increasing and right-continuous*

The application F being right-continuous at a point $c \in [0,1]^{n \times \mathbf{1}}$ means: for any sequence $(c^{(k)})$ in $[0,1]^{n \times \mathbf{1}}$ such that $(c^{(k)})$ converges to c when $k \to \infty$ and verifying $\forall k, c^{(k)} \geq c$, we have: $F(c^{(k)}) \to F(c)$ when $k \to \infty$.

Notation

For $x, y, z, u, \delta \in [0, 1]$*, we use the following notations:*

$$x^+ = \max(x, 0),$$

$$\overline{z}(\delta) = \min(z + \delta, 1),$$

$$\underline{z}(\delta) = \max(z - \delta, 0) = (z - \delta)^+.$$

We remark the following equivalence in $[0, 1]$: $\mid x - y \mid \leq \delta \Longleftrightarrow \underline{x}(\delta) \leq y \leq \overline{x}(\delta)$ (1)

Notation

For $x, y, z, u, \delta \in [0, 1]$, *we use the following notations:*

$$x^+ = \max(x, 0),$$
$$\overline{z}(\delta) = \min(z + \delta, 1),$$
$$\underline{z}(\delta) = \max(z - \delta, 0) = (z - \delta)^+.$$

We remark the following equivalence in $[0, 1]$: $\mid x - y \mid \leq \delta \Longleftrightarrow \underline{x}(\delta) \leq y \leq \overline{x}(\delta)$ (1)

Notation

For our work, to the second member $b = [b_i]_{1 \leq i \leq n}$ *of the system* (S) *and a number* $\delta \in [0, 1]$, *we associate two vectors:*

$$\underline{b}(\delta) = [(b_i - \delta)^+]_{1 \leq i \leq n} \quad \textit{and} \quad \overline{b}(\delta) = [\min(b_i + \delta, 1)]_{1 \leq i \leq n}$$

$\underline{b}(\delta)$, $\overline{b}(\delta)$ *were already used in (Cuninghame-Green and Cechlárová 1995), (Li and Fang 2010).*

Then, from (1), we deduce for any $c = [c_i]_{1 \leq i \leq n} \in [0, 1]^{n \times 1}$:

$$\|b - c\| \leq \delta \Longleftrightarrow \underline{b}(\delta) \leq c \leq \overline{b}(\delta)$$

where $\|b - c\| = \max_{1 \leq i \leq n} \mid b_i - c_i \mid$

$\Delta = \inf_{c \in \mathcal{C}} \|b - c\|$ of a system $A \Box_T^{\max} x = b$ – *Definition*

To the matrix A and the vector b of the system (S), let us associate the set of vectors $c = [c_i] \in [0,1]^{n \times 1}$ such that the system $A \Box_T^{\max} x = c$ is consistent:

$$\mathcal{C} = \{c = [c_i] \in [0,1]^{n \times 1} \mid A \Box_T^{\max} x = c \text{ is consistent}\}$$

$\mathcal{C}$ is non-empty : $\mathbf{0}$ is one of its element and has solution (sic)

To the matrix A and the vector b of the system (S), let us associate the set of vectors $c = [c_i] \in [0, 1]^{n \times 1}$ such that the system $A \Box_T^{\max} x = c$ is consistent:

$$\mathcal{C} = \{c = [c_i] \in [0, 1]^{n \times 1} \mid A \Box_T^{\max} x = c \text{ is consistent}\}$$

$\mathcal{C}$ is non-empty : $\mathbf{0}$ is one of its element and has solution (sic)

This set allows us to define:

Definition

The Chebyshev distance associated to b of the system $(S) : A \Box_T^{\max} x = b$ is:

$$\Delta = \Delta(A, b) = \inf_{c \in \mathcal{C}} \|b - c\|$$

where:

$$\|b - c\| = \max_{1 \leq i \leq n} \mid b_i - c_i \mid$$

$\Delta = \inf_{c \in \mathcal{C}} \|b - c\|$ of a system $A \Box_T^{\max} x = b$ – *Definition*

For $\max - \min$ composition, we have the following fundamental result proven in (Cuninghame-Green and Cechlárová 1995) for computing the Chebyshev distance (rewritten with F in (Baaj 2023a)):

Theorem

$$\Delta = \min\{\delta \in [0,1] \mid \underline{b}(\delta) \leq F(\overline{b}(\delta))\}$$

This result was extended to $\max - T$ composition (Baaj 2023b)

For $\max - \min$ composition, we have the following fundamental result proven in (Cuninghame-Green and Cechlárová 1995) for computing the Chebyshev distance (rewritten with F in (Baaj 2023a)):

Theorem

$$\Delta = \min\{\delta \in [0, 1] \mid \underline{b}(\delta) \leq F(\overline{b}(\delta))\}$$

This result was extended to $\max - T$ composition (Baaj 2023b)

Algorithms were proposed for computing Δ for $\max - \min$ composition: (Cuninghame-Green and Cechlárová 1995), (Li and Fang 2010), (Cimler, Gavalec and Zimmerman 2018) : they output an approximate value of Δ (which can be far from Δ) and/or require too many calculations

$\Delta = \inf_{c \in \mathcal{C}} \|b - c\|$ of a system $A \Box_{\min}^{\max} x = b$

The Chebyshev distance Δ associated to b of the system $A \Box_{\min}^{\max} x = b$ is given by the following analytical formula (Baaj 2023a):

The Chebyshev distance Δ associated to b of the system $A \Box_{\min}^{\max} x = b$ is given by the following analytical formula (Baaj 2023a):

Theorem

$$\Delta = \max_{1 \leq i \leq n} \min_{1 \leq j \leq m} \max[(b_i - a_{ij})^+, \max_{1 \leq k \leq n} \sigma_G\,(b_i, a_{kj}, b_k)]$$

where

$$\sigma_G(x, y, z) = \min(\frac{(x - z)^+}{2}, (y - z)^+)$$

Hint: For any $\delta \in [0, 1]$, we have $\underline{x}(\delta) \leq y \rightarrow_G \overline{z}(\delta) \Longleftrightarrow \sigma_G(x, y, z) \leq \delta$

For a system of max-product fuzzy relational equations $A \Box_*^{\max} x = b$, the Chebyshev distance is given by the following formula (Baaj 2023b):

Theorem

$$\Delta_P = \max_{1 \le i \le n} \min_{1 \le j \le m} \max_{1 \le k \le n} \sigma_{GG}(a_{ij}, b_i, a_{kj}, b_k), \tag{2}$$

where

$$\sigma_{GG}(u, x, y, z) = \max[(x - u)^+, \min(\varphi(u, x, y, z), (y - z)^+)]$$

and

$$\varphi(u, x, y, z) = \begin{cases} \frac{(x \cdot y - u \cdot z)^+}{u + y} & \text{if } u > 0 \\ x & \text{if } u = 0 \end{cases}$$

Hint: for all $\delta \in [0, 1]$, we have: $\underline{x}(\delta) \le u \cdot (y \underset{GG}{\longrightarrow} \overline{z}(\delta)) \Longleftrightarrow \sigma_{GG}(u, x, y, z) \le \delta$ where $\underset{GG}{\longrightarrow}$ is the Goguen product associated to T_P

For a system of max-Łukasiewicz fuzzy relational equations $A \Box_{T_L}^{\max} x = b$, the Chebyshev distance is given by the following formula (Baaj 2023b):

Theorem

$$\Delta_L = \max_{1 \le i \le n} \min_{1 \le j \le m} \max_{1 \le k \le n} \sigma_L(1 - a_{ij}, b_i, a_{kj}, b_k),$$

where

$$\sigma_L(u, x, y, z) = \min(x, \max(v^+, \frac{(v + y - z)^+}{2})) \textit{ with } v = x + u - 1$$

Hint: for all $\delta \in [0, 1]$, we have: $\underline{x}(\delta) \le \max(0, y \underset{L}{\longrightarrow} \overline{z}(\delta) - u) \iff \sigma_L(u, x, y, z) \le \delta$ where $\underset{L}{\longrightarrow}$ is the Łukasiewicz implication associated to T_L

Let an inconsistent system of max – min fuzzy relational equations $A \Box_{\min}^{\max} x = b$ where:

$$A = \begin{bmatrix} 0.03 & 0.38 & 0.26 \\ 0.98 & 0.10 & 0.03 \\ 0.77 & 0.15 & 0.85 \end{bmatrix} \text{ and } b = \begin{bmatrix} 0.54 \\ 0.13 \\ 0.87 \end{bmatrix} \tag{3}$$

We compute $\Delta = \max_{1 \leq i \leq 3} \delta_i$ where $\delta_i = \min_{1 \leq j \leq 3} \max[(b_i - a_{ij})^+, \max_{1 \leq k \leq 3} \sigma_G(b_i, a_{kj}, b_k)]$

For computing δ_1, we have:

$$[(b_{\mathbf{1}} - a_{\mathbf{1}j})^+]_{\mathbf{1} \leq j \leq \mathbf{3}} = \begin{bmatrix} 0.54 - 0.03 \\ 0.54 - 0.38 \\ 0.54 - 0.26 \end{bmatrix} = \begin{bmatrix} 0.51 \\ 0.16 \\ 0.28 \end{bmatrix}, [\sigma_G(b_{\mathbf{1}}, a_{kj}, b_k)]_{\mathbf{1} \leq k \leq \mathbf{3}, \mathbf{1} \leq j \leq \mathbf{3}} = \begin{bmatrix} 0.0 & 0.0 & 0.0 \\ 0.205 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \end{bmatrix}$$

$(\sigma_G(b_{\mathbf{1}}, a_{\mathbf{21}}, b_{\mathbf{2}}) = \sigma_G(0.54, 0.98, 0.13) = 0.205)$

Therefore: $\delta_1 = \min(\max(0.51, 0.205), \max(0.16, 0), \max(0.28, 0)) = 0.16$ and similarly we compute $\delta_2 = 0, \delta_3 = 0.02$

The Chebyshev distance Δ associated to b is $\Delta = \max(\delta_1, \delta_2, \delta_3) = 0.16$

Chebyshev approximations of b of $A \Box_T^{\max} x = b$

For an inconsistent system $A \Box_T^{\max} x = b$ we define:

Definition

The set of Chebyshev approximations of the second member b is defined using the set $\mathcal{C}$ and the Chebyshev distance Δ associated to the second member b:

$$\mathcal{C}_b = \{c \in \mathcal{C} \mid \|b - c\| = \Delta(A, b)\}$$

Reminder: $\mathcal{C}$ **is the set of second members of the consistent systems defined with the same matrix** A

Chebyshev approximations of b of $A\Box_T^{\max}x = b$

For an inconsistent system $A\Box_T^{\max}x = b$ we define:

Definition

The set of Chebyshev approximations of the second member b is defined using the set $\mathcal{C}$ and the Chebyshev distance Δ associated to the second member b:

$$\mathcal{C}_b = \{c \in \mathcal{C} \mid \|b - c\| = \Delta(A, b)\}$$

Reminder: $\mathcal{C}$ is the set of second members of the consistent systems defined with the same matrix A

We can prove that the set $\mathcal{C}_b$ is non-empty as it has a greater element:

Chebyshev approximations of b of $A \Box_T^{\max} x = b$

For an inconsistent system $A \Box_T^{\max} x = b$ we define:

Definition

The set of Chebyshev approximations of the second member b is defined using the set $\mathcal{C}$ and the Chebyshev distance Δ associated to the second member b:

$$\mathcal{C}_b = \{c \in \mathcal{C} \mid \|b - c\| = \Delta(A, b)\}$$

Reminder: $\mathcal{C}$ is the set of second members of the consistent systems defined with the same matrix A

We can prove that the set $\mathcal{C}_b$ is non-empty as it has a greater element:

Proposition

(1) $F(\overline{b}(\Delta)) \in \mathcal{C}_b$,
(2) $\forall c \in \mathcal{C}_b,\ c \leq F(\overline{b}(\Delta))$,
So, $F(\overline{b}(\Delta))$ *is the greatest Chebyshev approximation of* b

Sketch of the proof of (1): We have $\underline{b}(\Delta) \leq F(\overline{b}(\Delta)) \leq \overline{b}(\Delta)$, $\|F(\overline{b}(\Delta)) - b\| \leq \Delta$, but $\Delta = \inf_{c \in \mathcal{C}} \|b - c\|$, so $\|F(\overline{b}(\Delta)) - b\| = \Delta$ then we get the property. The proof of (2) is given by: for $c \in \mathcal{C}_b$, we have $c \leq \overline{b}(\Delta)$, as F is increasing, $F(c) = c \leq F(\overline{b}(\Delta))$

As a consequence we have:

Corollary

$$\Delta = \min_{c \in \mathcal{C}} \|b - c\|$$

$$\Delta = 0 \iff \textit{the system } (S) \textit{ is consistent}$$

As a consequence we have:

Corollary

$$\Delta = \min_{c \in \mathcal{C}} \|b - c\|$$

$$\Delta = 0 \Longleftrightarrow \textit{ the system } (S) \textit{ is consistent}$$

Therefore, $\Delta = 0$ is a necessary and sufficient condition for the system (S) to be consistent

As a consequence we have:

Corollary

$$\Delta = \min_{c \in \mathcal{C}} \|b - c\|$$

$$\Delta = 0 \iff \textit{the system } (S) \textit{ is consistent}$$

Therefore, $\Delta = 0$ is a necessary and sufficient condition for the system (S) to be consistent

It is much more difficult to obtain *minimal Chebyshev approximations of b*. In the next section, we will show that the set:

$$\mathcal{C}_{b,\min} = \{c \in \mathcal{C}_b \mid c \text{ minimal in } \mathcal{C}_b\}$$

is non-empty (we will construct elements of this set) and finite

We continue with the inconsistent system of $\max - \min$ fuzzy relational equations $A \Box_{\min}^{\max} x = b$ where:

$$A = \begin{bmatrix} 0.03 & 0.38 & 0.26 \\ 0.98 & 0.10 & 0.03 \\ 0.77 & 0.15 & 0.85 \end{bmatrix} \text{ and } b = \begin{bmatrix} 0.54 \\ 0.13 \\ 0.87 \end{bmatrix}. \tag{4}$$

whose Chebyshev distance associated to b was computed $\Delta = 0.16$.

From $b = \begin{bmatrix} 0.54 \\ 0.13 \\ 0.87 \end{bmatrix}$, we compute $\overline{b}(\Delta) = \begin{bmatrix} 0.70 \\ 0.29 \\ 1.00 \end{bmatrix}$. Then, the greatest Chebyshev approximation of b is:

$$F(\overline{b}(\Delta)) = A \Box_{\min}^{\max} (A^t \Box_{\rightarrow_G}^{\min} \overline{b}(\Delta)) = \begin{bmatrix} 0.38 \\ 0.29 \\ 0.85 \end{bmatrix}.$$

We check:

- the Chebyshev distance between $F(\overline{b}(\Delta))$ and b is equal to $\Delta = 0.16$
- the system $A \Box_{\min}^{\max} x = F(\overline{b}(\Delta))$ is consistent

In this section, for an inconsistent system $A \Box_T^{\max} x = b$, we study:

Λ_b : the approximate solutions set of the system

$\mathcal{C}_b$: the set of Chebyshev approximations of the second member b

Relating approx solutions set Λ_b to Chebyshev approx. set $\mathcal{C}_b$

In this section, for an inconsistent system $A \Box_T^{\max} x = b$, we study:

Λ_b : the approximate solutions set of the system

$\mathcal{C}_b$: the set of Chebyshev approximations of the second member b

which are related as follows:

The approximate solutions are any vector $x^* \in [0,1]^{m \times 1}$ such that the vector $c = A \Box_T^{\max} x^*$ is a Chebyshev approximation of b i.e., $c \in \mathcal{C}_b$

In this section, for an inconsistent system $A \Box_T^{\max} x = b$, we study:

Λ_b : the approximate solutions set of the system

$\mathcal{C}_b$: the set of Chebyshev approximations of the second member b

which are related as follows:

The approximate solutions are any vector $x^* \in [0,1]^{m\times 1}$ such that the vector $c = A \Box_T^{\max} x^*$ is a Chebyshev approximation of b i.e., $c \in \mathcal{C}_b$

Then, we have:

for all $c \in \mathcal{C}_b$, the solutions of the system $A \Box_T^{\max} x = c$ belong to Λ_b i.e., they are approximate solutions

Relating Λ_b to $\mathcal{C}_b$ – *Preliminaries*

Formally, we introduce the following surjective and increasing map:

$$\theta : [0,1]^{m\times 1} \to \mathcal{C} : x \mapsto A \Box_T^{\max} x$$

Reminder: $\mathcal{C}$ is the set of second members of the consistent systems defined with the same matrix A

Formally, we introduce the following surjective and increasing map:

$$\theta : [0,1]^{m\times 1} \to \mathcal{C} : x \mapsto A \Box_T^{\max} x$$

Reminder: $\mathcal{C}$ **is the set of second members of the consistent systems defined with the same matrix** A

Definition

The approximate solutions set is defined by:

$$\Lambda_b = \theta^{-1}(\mathcal{C}_b) = \{x \in [0,1]^{m\times 1} \mid \theta(x) \in \mathcal{C}_b\}$$

Relating Λ_b to $\mathcal{C}_b$ – *Preliminaries*

Formally, we introduce the following surjective and increasing map:

$$\theta : [0,1]^{m\times 1} \to \mathcal{C} : x \mapsto A\Box_T^{\max} x$$

Reminder: $\mathcal{C}$ is the set of second members of the consistent systems defined with the same matrix A

Definition

The approximate solutions set is defined by:

$$\Lambda_b = \theta^{-1}(\mathcal{C}_b) = \{x \in [0,1]^{m\times 1} \mid \theta(x) \in \mathcal{C}_b\}$$

- As θ is surjective we have:

$$\mathcal{C}_b = \{\theta(x) \mid x \in \Lambda_b\} = \theta(\Lambda_b)$$

Formally, we introduce the following surjective and increasing map:

$$\theta : [0,1]^{m\times 1} \to \mathcal{C} : x \mapsto A \Box_T^{\max} x$$

Reminder: $\mathcal{C}$ is the set of second members of the consistent systems defined with the same matrix A

Definition

The approximate solutions set is defined by:

$$\Lambda_b = \theta^{-1}(\mathcal{C}_b) = \{x \in [0,1]^{m\times 1} \mid \theta(x) \in \mathcal{C}_b\}$$

- As θ is surjective we have:

$$\mathcal{C}_b = \{\theta(x) \mid x \in \Lambda_b\} = \theta(\Lambda_b)$$

- We have: $$x \in \Lambda_b \iff \|\theta(x) - b\| = \Delta(A, b)$$

  Reminder: $\Delta(A,b)$ is the Chebyshev distance associated to $A \Box_T^{\max} x = b$

- The approximate solutions set Λ_b is non-empty as $\eta = A^t \Box_{\mathcal{I}_T}^{\min} F(\overline{b}(\Delta))$ is one of its element and we prove that it is the greatest one

Relating approx solutions set Λ_b to Chebyshev approx. set $\mathcal{C}_b$

In what follows, we shall look for a finite non-empty set denoted $\Lambda_{b,\min}$ of minimal approximate solutions, which satisfies:

$$\Lambda_{b,\min} \subseteq \Lambda_b \text{ and } \mathcal{C}_{b,\min} = \{\theta(x) \mid x \in \Lambda_{b,\min}\} = \theta(\Lambda_{b,\min})$$

The existence of $\Lambda_{b,\min}$ implies that $\mathcal{C}_{b,\min}$ is also non-empty and finite

In what follows, we shall look for a finite non-empty set denoted $\Lambda_{b,\min}$ of minimal approximate solutions, which satisfies:

$$\Lambda_{b,\min} \subseteq \Lambda_b \text{ and } \mathcal{C}_{b,\min} = \{\theta(x) \mid x \in \Lambda_{b,\min}\} = \theta(\Lambda_{b,\min})$$

The existence of $\Lambda_{b,\min}$ implies that $\mathcal{C}_{b,\min}$ is also non-empty and finite

It is deduced from this characterization of the set Λ_b (Baaj 2023a):

Proposition

For any $x \in [0,1]^{m\times 1}$, we have:

x is an approximate solution i.e., $x \in \Lambda_b \Longleftrightarrow \underline{b}(\Delta) \leq A \Box_T^{\max} x$ and $x \leq \eta$

Relating approx solutions set Λ_b to Chebyshev approx. set $\mathcal{C}_b$

In what follows, we shall look for a finite non-empty set denoted $\Lambda_{b,\min}$ of minimal approximate solutions, which satisfies:

$$\Lambda_{b,\min} \subseteq \Lambda_b \text{ and } \mathcal{C}_{b,\min} = \{\theta(x) \mid x \in \Lambda_{b,\min}\} = \theta(\Lambda_{b,\min})$$

The existence of $\Lambda_{b,\min}$ implies that $\mathcal{C}_{b,\min}$ is also non-empty and finite

It is deduced from this characterization of the set Λ_b (Baaj 2023a):

Proposition

For any $x \in [0,1]^{m\times 1}$, we have:

x is an approximate solution i.e., $x \in \Lambda_b \iff \underline{b}(\Delta) \leq A \Box_T^{\max} x$ and $x \leq \eta$

and to get minimal approximate solutions we use (Matusiewicz and Drewniak 2013)'s result on the solving of a system of max $-*$ inequalities ($*$: increasing and continuous function): *a system of max-$*$ inequalities has a finite non-empty set of solutions*

We use (Matusiewicz and Drewniak 2013)'s result (they also give an algorithm for it):

Notation

$\{v^{(1)}, v^{(2)}, \dots, v^{(h)}\}$ *is the set of minimal solutions of the system of inequalities* $\underline{b}(\Delta) \leq A \Box_T^{\max} x$ *such that* $\forall i \in \{1, 2, \dots, h\}, v^{(i)} \leq \eta$

Relating approx solutions set Λ_b to Chebyshev approx. set $\mathcal{C}_b$

We use (Matusiewicz and Drewniak 2013)'s result (they also give an algorithm for it):

Notation

$\{v^{(1)}, v^{(2)}, \dots, v^{(h)}\}$ *is the set of minimal solutions of the system of inequalities* $\underline{b}(\Delta) \leq A \Box_T^{\max} x$ *such that* $\forall i \in \{1, 2, \dots, h\}, v^{(i)} \leq \eta$

Then we get the minimal Chebyshev approximations of b:

Proposition

We put:

$\widetilde{\mathcal{C}} = \{\theta(v^{(1)}), \theta(v^{(2)}), \dots, \theta(v^{(h)})\}$ *and* $(\widetilde{\mathcal{C}})_{\min} = \{c \in \widetilde{\mathcal{C}} \mid c$ *is minimal in* $\widetilde{\mathcal{C}}\}$

Then, we have: $\widetilde{\mathcal{C}} \subseteq \mathcal{C}_b$ *and* $\mathcal{C}_{b,\min} = (\widetilde{\mathcal{C}})_{\min}$

We use (Matusiewicz and Drewniak 2013)'s result (they also give an algorithm for it):

Notation

$\{v^{(1)}, v^{(2)}, \dots, v^{(h)}\}$ is the set of minimal solutions of the system of inequalities $\underline{b}(\Delta) \leq A \Box_T^{\max} x$ such that $\forall i \in \{1, 2, \dots, h\}, v^{(i)} \leq \eta$

Then we get the minimal Chebyshev approximations of b:

Proposition

We put:
$\widetilde{\mathcal{C}} = \{\theta(v^{(1)}), \theta(v^{(2)}), \dots, \theta(v^{(h)})\}$ *and* $(\widetilde{\mathcal{C}})_{\min} = \{c \in \widetilde{\mathcal{C}} \mid c \text{ is minimal in } \widetilde{\mathcal{C}}\}$

Then, we have: $\widetilde{\mathcal{C}} \subseteq \mathcal{C}_b$ *and* $\mathcal{C}_{b,\min} = (\widetilde{\mathcal{C}})_{\min}$

So the set $\mathcal{C}_{b,\min}$ is non-empty and finite, and we are able to define a non-empty and finite set of minimal approximate solutions:
$\Lambda_{b,\min} = \{x \in \{v^{(1)}, v^{(2)}, \dots, v^{(h)}\} \mid \theta(x) \in \mathcal{C}_{b,\min}\}$

The structure of the set $\mathcal{C}_b$ of Chebyshev approximations of b is described by the following result:

Theorem

For all $c \in [0,1]^{n \times 1}$, we have:

$$c \in \mathcal{C}_b \iff F(c) = c \text{ and } \exists\, c' \in \mathcal{C}_{b,\min} \text{ s.t. } c' \leq c \leq F(\overline{b}(\Delta)).$$

In (Baaj 2023a), for max − min composition, a structure theorem for the approximate solutions set Λ_b is established

We continue with the same system $A\Box_{\min}^{\max} x = b$ whose Chebyshev distance is $\Delta = 0.16$.
We compute:

$$\underline{b}(\Delta) = \begin{bmatrix} 0.38 \\ 0.00 \\ 0.71 \end{bmatrix}, \overline{b}(\Delta) = \begin{bmatrix} 0.70 \\ 0.29 \\ 1.00 \end{bmatrix} \text{ and } \eta = A^t \Box_{\to_G}^{\min} F(\overline{b}(\Delta)) = \begin{bmatrix} 0.29 \\ 1 \\ 1 \end{bmatrix}$$

The vector $\theta(\eta) = \begin{bmatrix} 0.38 \\ 0.29 \\ 0.85 \end{bmatrix}$ is the greatest Chebyshev approximation of b

The system of inequalities $\underline{b}(\Delta) \leq A\Box_{\min}^{\max} x$ is: $\begin{bmatrix} 0.38 \\ 0.00 \\ 0.71 \end{bmatrix} \leq \begin{bmatrix} 0.03 & 0.38 & 0.26 \\ 0.98 & 0.10 & 0.03 \\ 0.77 & 0.15 & 0.85 \end{bmatrix} \Box_{\min}^{\max} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$.

Using the approach of (Matusiewicz and Drewniak 2013), we obtain two minimal solutions:
$v = \begin{bmatrix} 0.00 \\ 0.38 \\ 0.71 \end{bmatrix}$ and $v' = \begin{bmatrix} 0.71 \\ 0.38 \\ 0.00 \end{bmatrix}$ Among these minimal solutions, only v is lower than η

The set $\widetilde{\mathcal{C}}$ contains one element, which is $A\Box_{\min}^{\max} v = \begin{bmatrix} 0.38 \\ 0.10 \\ 0.71 \end{bmatrix}$ and we have $\widetilde{\mathcal{C}} = (\widetilde{\mathcal{C}})_{\min}$

Therefore, the unique minimal Chebyshev approximation of b is $\check{b} = \begin{bmatrix} 0.38 \\ 0.10 \\ 0.71 \end{bmatrix}$

Some approximate solutions of the system (S) are the solutions of the consistent system $\theta(\eta) = A\Box_{\min}^{\max} x$ and the solutions of the consistent system $\check{b} = A\Box_{\min}^{\max} x$

We introduce a paradigm to approximately learn a weight matrix W relating input and output data by $\max - T$ composition from training data:

$$(x^{(i)})_{1 \leq i \leq N}, x^{(i)} \in [0,1]^{m \times 1} \quad ; \quad (y^{(i)})_{1 \leq i \leq N}, y^{(i)} \in [0,1]^{n \times 1}$$

The learning error is expressed by L_∞ norm: $E(W) = \max_{1 \leq i \leq N} \|y^{(i)} - W \Box_T^{\max} x^{(i)}\|$
(*where the norm of a vector z of n components is* $\|z\| = \max_{1 \leq k \leq n} \mid z_k \mid$)

Learning Approximate Weight Matrices

We introduce a paradigm to approximately learn a weight matrix W relating input and output data by $\max - T$ composition from training data:

$$(x^{(i)})_{1 \leq i \leq N}, x^{(i)} \in [0,1]^{m \times 1} \quad ; \quad (y^{(i)})_{1 \leq i \leq N}, y^{(i)} \in [0,1]^{n \times 1}$$

The learning error is expressed by L_∞ norm: $E(W) = \max_{1 \leq i \leq N} \|y^{(i)} - W \Box_T^{\max} x^{(i)}\|$

(*where the norm of a vector z of n components is* $\|z\| = \max_{1 \leq k \leq n} \mid z_k \mid$)

Main results:

- we can compute by an analytical formula a positive constant μ, *which depends only on the training data*, such that the following equality holds:

$$\mu = \min_{W \in [0,1]^{n \times m}} E(W)$$

- Whatever if $\mu = 0$ or $\mu > 0$ our method give a matrix W^* s.t $E(W^*) = \mu$

We introduce a paradigm to approximately learn a weight matrix W relating input and output data by $\max - T$ composition from training data:

$$(x^{(i)})_{1 \leq i \leq N}, x^{(i)} \in [0,1]^{m \times 1} \quad ; \quad (y^{(i)})_{1 \leq i \leq N}, y^{(i)} \in [0,1]^{n \times 1}$$

The learning error is expressed by L_∞ norm: $E(W) = \max\limits_{1 \leq i \leq N} \|y^{(i)} - W \Box_T^{\max} x^{(i)}\|$
(*where the norm of a vector z of n components is* $\|z\| = \max\limits_{1 \leq k \leq n} \mid z_k \mid$)
Main results:

- we can compute by an analytical formula a positive constant μ, *which depends only on the training data*, such that the following equality holds:

$$\mu = \min_{W \in [0,1]^{n \times m}} E(W)$$

- Whatever if $\mu = 0$ or $\mu > 0$ our method give a matrix W^* s.t $E(W^*) = \mu$

existing approaches: numerous $\max - \min$ gradient descent e.g. (Pedrycz 1991), associative memory (Sussner and Valle 2006)

We begin by studying the learning of approximate weight matrices when the training data are composed of N data whose outputs are *scalars*:

$$(x^{(i)})_{1 \leq i \leq N}, x^{(i)} \in [0,1]^{m \times 1} \quad ; \quad (y^{(i)})_{1 \leq i \leq N}, y^{(i)} \in [0,1]$$

(For $i = 1, 2, \dots, N$, each pair $(x^{(i)}, y^{(i)})$ is a training datum, where $x^{(i)}$ is an input data vector and $y^{(i)}$ is the targeted output data value in $[0,1]$)

We want to learn a weight matrix $V \in [0,1]^{1 \times m}$ such that:

$$\forall i \in \{1, 2, \dots, N\}, \; V \Box_T^{\max} x^{(i)} = y^{(i)}$$

Learning Approximate Weight Matrices

To tackle this problem, we introduce the following system:

$$(S) : L \Box_T^{\max} u = b,$$

where:

$$L = \left[x_j^{(i)}\right]_{1 \leq i \leq N, 1 \leq j \leq m} = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_m^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_m^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_m^{(N)} \end{bmatrix} \quad \text{and} \quad b = [y^{(i)}]_{1 \leq i \leq N} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix}.$$

(S) is canonically associated to the training data: the rows of L are the transpose of the $x^{(1)}, x^{(2)}, \dots, x^{(N)}$ and the components of b are the $y^{(1)}, y^{(2)}, \dots, y^{(N)}$

To tackle this problem, we introduce the following system:

$$(S) : L\Box_T^{\max} u = b,$$

where:

$$L = \left[x_j^{(i)}\right]_{1\leq i\leq N, 1\leq j\leq m} = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_m^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_m^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_m^{(N)} \end{bmatrix} \quad \text{and} \quad b = [y^{(i)}]_{1\leq i\leq N} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{bmatrix}.$$

(S) is canonically associated to the training data: the rows of L are the transpose of the $x^{(1)}, x^{(2)}, \dots, x^{(N)}$ and the components of b are the $y^{(1)}, y^{(2)}, \dots, y^{(N)}$

To relate the learning problem formulated previously to the system (S), we will use:

Lemma

Let $v = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \in [0,1]^{m\times 1}$ *be a column-vector and* $V = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \in [0,1]^{1\times m}$ *is the row matrix which is the transpose of* v. *We put* $b' = \left[b'_i\right]_{1\leq i\leq N} = L\Box_T^{\max} v$. *Then, we have:*

1. $\forall i \in \{1,2,\dots,N\}$, $b'_i = V\Box_T^{\max} x^{(i)} \in [0,1]$,
2. $\|b - b'\| = E(V) = \max_{1\leq i\leq N} \mid y^{(i)} - V\Box_T^{\max} x^{(i)} \mid$

The second statement implies that V is a weight matrix of the training data $((x^{(i)})_{1\leq i\leq N}, (y^{(i)})_{1\leq i\leq N})$ if and only if v is a solution of the system (S)

Therefore, the learning problem is related to the system (S) by:

Proposition

Let $v = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \in [0,1]^{m\times 1}$ *be a column-vector and* $V = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \in [0,1]^{1\times m}$ *is the row matrix which is the transpose of* v*. We have:*

$$v \text{ is a solution of the system } (S) \iff \forall i \in \{1,2,\dots,N\}\; V \Box_T^{\max} x^{(i)} = y^{(i)}$$

Therefore, the learning problem is related to the system (S) by:

Proposition

Let $v = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \in [0,1]^{m \times 1}$ *be a column-vector and* $V = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \in [0,1]^{1 \times m}$ *is the row matrix which is the transpose of* v*. We have:*

$$v \text{ is a solution of the system } (S) \iff \forall i \in \{1, 2, \dots, N\}\ V \Box_T^{\max} x^{(i)} = y^{(i)}$$

From the Lemma (second statement) it follows that for any $V \in [0,1]^{1 \times m}$, we have:

$$E(V) = \max_{1 \leq i \leq N} \mid y^{(i)} - V \Box_T^{\max} x^{(i)} \mid = \|b - L \Box_T^{\max} v\| \geq \Delta(L, b) \quad \text{(Chebyshev distance of } (S))$$

Learning Approximate Weight Matrices

Therefore, the learning problem is related to the system (S) by:

Proposition

Let $v = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \in [0,1]^{m\times 1}$ *be a column-vector and* $V = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \in [0,1]^{1\times m}$ *is the row matrix which is the transpose of* v*. We have:*

$$v \text{ is a solution of the system } (S) \iff \forall i \in \{1,2,\dots,N\}\ V \Box_T^{\max} x^{(i)} = y^{(i)}$$

From the Lemma (second statement) it follows that for any $V \in [0,1]^{1\times m}$, we have:

$$E(V) = \max_{1\leq i\leq N} \mid y^{(i)} - V\Box_T^{\max} x^{(i)} \mid = \|b - L\Box_T^{\max} v\| \geq \Delta(L,b) \quad \text{(Chebyshev distance of } (S))$$

This leads to the definition of the positive constant μ:

Definition

The positive constant μ minimizing the learning error $E(V) = \max_{1\leq i\leq N} \mid y^{(i)} - V\Box_T^{\max} x^{(i)} \mid$ according to training data is the Chebyshev distance associated to b of (S): $\mu = \Delta(L,b)$

Therefore, the learning problem is related to the system (S) by:

Proposition

Let $v = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \in [0,1]^{m\times 1}$ *be a column-vector and* $V = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \in [0,1]^{1\times m}$ *is the row matrix which is the transpose of* v*. We have:*

$$v \text{ is a solution of the system } (S) \Longleftrightarrow \forall i \in \{1,2,\dots,N\}\ V\Box_T^{\max} x^{(i)} = y^{(i)}$$

From the Lemma (second statement) it follows that for any $V \in [0,1]^{1\times m}$, we have:

$$E(V) = \max_{1\leq i\leq N} \mid y^{(i)} - V\Box_T^{\max} x^{(i)} \mid = \|b - L\Box_T^{\max} v\| \geq \Delta(L,b) \quad \text{(Chebyshev distance of } (S))$$

This leads to the definition of the positive constant μ:

Definition

The positive constant μ minimizing the learning error $E(V) = \max_{1\leq i\leq N} \mid y^{(i)} - V\Box_T^{\max} x^{(i)} \mid$ according to training data is the Chebyshev distance associated to b of (S): $\mu = \Delta(L,b)$

As:
v is an approx. solution of $(S) \Longleftrightarrow E(V) = \max_{1\leq i\leq N} \mid y^{(i)} - V\Box_T^{\max} x^{(i)} \mid = \Delta(L,b) = \mu$,
we deduce that the equality $\mu = \min_{V\in[0,1]^{1\times m}} E(V)$ holds

Let us consider the following training data:

$$\begin{array}{l|l} x^{(1)} = (0.7, 0.4, 0.4)^t & y^{(1)} = 0.7 \\ x^{(2)} = (1.0, 0.2, 0.5)^t & y^{(2)} = 1.0 \\ x^{(3)} = (0.2, 0.3, 0.8)^t & y^{(3)} = 0.3 \end{array}$$

We construct the system $(S) : L \Box_{\min}^{\max} u = b$ where $L = \begin{bmatrix} 0.7 & 0.4 & 0.4 \\ 1.0 & 0.2 & 0.5 \\ 0.2 & 0.3 & 0.8 \end{bmatrix}$ and $b = \begin{bmatrix} 0.7 \\ 1.0 \\ 0.3 \end{bmatrix}$.
The system is consistent because the Chebyshev distance associated to b is equal to zero: $\Delta(L, b) = 0$, so $\mu = \Delta(L, b) = 0$. The greatest solution of (S) is $\begin{bmatrix} 1.0 \\ 1.0 \\ 0.3 \end{bmatrix}$ and there are two minimal solutions $\begin{bmatrix} 1.0 \\ 0.3 \\ 0.0 \end{bmatrix}$ and $\begin{bmatrix} 1.0 \\ 0.0 \\ 0.3 \end{bmatrix}$. Let us use the solution $v = \begin{bmatrix} 1.0 \\ 0.7 \\ 0.3 \end{bmatrix}$ of the system (S) and we put $V = v^t = \begin{bmatrix} 1.0 & 0.7 & 0.3 \end{bmatrix}$. The weight matrix V relates input and output data of the training data:

$$V \Box_{\min}^{\max} x^{(1)} = y^{(1)},$$
$$V \Box_{\min}^{\max} x^{(2)} = y^{(2)},$$
$$V \Box_{\min}^{\max} x^{(3)} = y^{(3)}.$$

Learning Approximate Weight Matrices

We continue by learning approximate weight matrices W when the training data are composed of N data whose outputs are *vectors*:

$$(x^{(i)})_{1 \leq i \leq N}, x^{(i)} \in [0,1]^{m \times 1} \quad ; \quad (y^{(i)})_{1 \leq i \leq N}, y^{(i)} \in [0,1]^{n \times 1}$$

For $i = 1, 2, \ldots, N$, each pair $(x^{(i)}, y^{(i)})$ is a training datum, where $x^{(i)}$ is the input data vector and $y^{(i)}$ is the targeted output data vector.

We study the following problems:
1. Is there a weight matrix W of size (n, m) such that:

$$\forall i \in \{1, 2, \cdots, N\}, \; W \Box_T^{\max} x^{(i)} = y^{(i)}$$

2. If this not the case, how to get a suitable approximate weight matrix W ?

We associate to the training data n systems of $\max - T$ fuzzy relational equations denoted by $(S_1), (S_2), \dots, (S_n)$

We associate to the training data n systems of $\max - T$ fuzzy relational equations denoted by $(S_1), (S_2), \dots, (S_n)$

For $1 \leq k \leq n$, the system (S_k) is of the form:

$$(S_k) : L \Box_T^{\max} u = b^{(k)}, \tag{5}$$

where:

- we reuse the matrix $L = [l_{ij}]_{1 \leq i \leq N, 1 \leq j \leq m} = [x_j^{(i)}]_{1 \leq i \leq N, 1 \leq j \leq m}$ of size (N, m), which is defined by the transpose of the input data column vectors $x^{(1)}, x^{(2)}, \dots, x^{(N)}$,
- the unknown part is a column vector $u \in [0, 1]^{m \times 1}$
- for $k = 1, 2, \dots, n$, the components of the column vector $b^{(k)} = [b_i^{(k)}]_{1 \leq i \leq N}$ are defined by: $b_i^{(k)} = y_k^{(i)}$; $1 \leq i \leq N$

We associate to the training data n systems of $\max - T$ fuzzy relational equations denoted by $(S_1), (S_2), \dots, (S_n)$

For $1 \leq k \leq n$, the system (S_k) is of the form:

$$(S_k) : L \Box_T^{\max} u = b^{(k)}, \tag{5}$$

where:

- we reuse the matrix $L = [l_{ij}]_{1 \leq i \leq N, 1 \leq j \leq m} = [x_j^{(i)}]_{1 \leq i \leq N, 1 \leq j \leq m}$ of size (N, m), which is defined by the transpose of the input data column vectors $x^{(1)}, x^{(2)}, \dots, x^{(N)}$,
- the unknown part is a column vector $u \in [0,1]^{m \times 1}$
- for $k = 1, 2, \dots, n$, the components of the column vector $b^{(k)} = [b_i^{(k)}]_{1 \leq i \leq N}$ are defined by: $b_i^{(k)} = y_k^{(i)}$; $1 \leq i \leq N$

We remark that for $k = 1, 2, \dots, n$ and $i = 1, 2, \dots, N$, each component $b_i^{(k)}$ of the second member $b^{(k)}$ of the system (S_k) is equal to the component $y_k^{(i)}$ of the targeted output data vector $y^{(i)}$:

$$b^{(k)} = \begin{bmatrix} y_k^{(1)} \\ y_k^{(2)} \\ \vdots \\ y_k^{(N)} \end{bmatrix} \tag{6}$$

In (Baaj 2023a), we prove the following results:

• There is a positive constant denoted μ which can be computed by an analytical formula according to the training data and which satisfies:

$$\forall W \in [0,1]^{n\times m}, E(W) = \max_{1\leq i\leq N} \|y^{(i)} - W \Box_T^{\max} x^{(i)}\| \geq \mu \tag{7}$$

This positive constant minimizes the learning error $E(W)$ and is expressed in terms of Chebyshev distances associated to the second member of the systems:

$$\mu := \max_{1\leq k\leq n} \Delta(L, b^{(k)}) \tag{8}$$

Learning Approximate Weight Matrices

In (Baaj 2023a), we prove the following results:

• There is a positive constant denoted μ which can be computed by an analytical formula according to the training data and which satisfies:

$$\forall W \in [0,1]^{n\times m}, E(W) = \max_{1\leq i\leq N} \|y^{(i)} - W \Box_T^{\max} x^{(i)}\| \geq \mu \tag{7}$$

This positive constant minimizes the learning error $E(W)$ and is expressed in terms of Chebyshev distances associated to the second member of the systems:

$$\mu := \max_{1\leq k\leq n} \Delta(L, b^{(k)}) \tag{8}$$

• We show the following equivalence: having a weight matrix that perfectly relates the input data to the output data is equivalent to having $\mu = 0$ i.e.,

$$\exists W \in [0,1]^{n\times m},\ \text{s.t. } \forall i \in \{1,2,\cdots,N\}, W \Box_T^{\max} x^{(i)} = y^{(i)} \iff \mu = 0 \tag{9}$$

Learning Approximate Weight Matrices

In (Baaj 2023a), we prove the following results:

• There is a positive constant denoted μ which can be computed by an analytical formula according to the training data and which satisfies:

$$\forall W \in [0,1]^{n\times m}, E(W) = \max_{1\leq i\leq N} \|y^{(i)} - W \Box_T^{\max} x^{(i)}\| \geq \mu \tag{7}$$

This positive constant minimizes the learning error $E(W)$ and is expressed in terms of Chebyshev distances associated to the second member of the systems:

$$\mu := \max_{1\leq k\leq n} \Delta(L, b^{(k)}) \tag{8}$$

• We show the following equivalence: having a weight matrix that perfectly relates the input data to the output data is equivalent to having $\mu = 0$ i.e.,

$$\exists W \in [0,1]^{n\times m},\ \text{s.t. } \forall i \in \{1,2,\cdots,N\}, W \Box_T^{\max} x^{(i)} = y^{(i)} \iff \mu = 0 \tag{9}$$

• We show that the set of approximate weight matrices:

$$\mathcal{A} = \left\{ W \in [0,1]^{n\times m} \;\middle|\; E(W) = \max_{1\leq i\leq N} \|y^{(i)} - W \Box_T^{\max} x^{(i)}\| = \mu \right\} \tag{10}$$

is non-empty. This implies that $\mu = \min_{W\in[0,1]^{n\times m}} E(W)$

In (Baaj 2023a), we give a method for learning approximate weight matrices W:

Method

Let W be a matrix defined row by row, which satisfies the following conditions:

- *If the system (S_k) is consistent, we define the k-th row of W as the transpose of a solution $u^{(k)}$ of the system (S_k). For instance, its greatest solution $L^t \Box_{\mathcal{I}_T}^{\min} b^{(k)}$. With this choice, we have:*

$$\| b^{(k)} - L \Box_T^{\max} u^{(k)} \| = 0 = \Delta(L, b^{(k)})$$

- *If the system (S_k) is inconsistent, we take a Chebyshev approximation $b^{(k),*}$ of $b^{(k)}$ (an element of the non-empty set $\mathcal{C}_{b^{(k)}}$). With this choice, we define the k-th row of W as the transpose of a solution $u^{(k)}$ of the system $L \Box_T^{\max} u = b^{(k),*}$, for instance the greatest solution $L^t \Box_{\mathcal{I}_T}^{\min} b^{(k),*}$. With this choice, we have:*

$$\| b^{(k)} - L \Box_T^{\max} u^{(k)} \| = \Delta(L, b^{(k)})$$

Therefore $W \in \mathcal{A}$

Let us consider the following training data:

$$\begin{array}{l|l} x^{(1)} = (0.7, 0.4, 0.4)^t & y^{(1)} = (0.7, 0.1, 0.3)^t \\ x^{(2)} = (1.0, 0.2, 0.5)^t & y^{(2)} = (1.0, 0.7, 0.0)^t \end{array}$$

Table 1: Training data of the example. We have $N = 2, m = 3$ and $n = 3$.

We have $L = \begin{bmatrix} 0.7 & 0.4 & 0.4 \\ 1.0 & 0.2 & 0.5 \end{bmatrix}$, $b^{(1)} = \begin{bmatrix} 0.7 \\ 1.0 \end{bmatrix}$, $b^{(2)} = \begin{bmatrix} 0.1 \\ 0.7 \end{bmatrix}$ and $b^{(3)} = \begin{bmatrix} 0.3 \\ 0.0 \end{bmatrix}$. We form three systems $(S_1), (S_2)$ and (S_3):

$$(S_1) : L \Box_{\min}^{\max} u_1 = b^{(1)},$$
$$(S_2) : L \Box_{\min}^{\max} u_2 = b^{(2)},$$
$$(S_3) : L \Box_{\min}^{\max} u_3 = b^{(3)}.$$

- The system (S_1) is consistent because $\Delta(L, b^{(1)}) = 0$. It has $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ as greatest solution, and it has a unique minimal solution $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

- The system (S_2) is inconsistent because $\Delta(L, b^{(2)}) = 0.3$. We get $\eta = \begin{bmatrix} 0.4 \\ 1 \\ 1 \end{bmatrix}$ and the greatest Chebyshev approximation of $b^{(2)}$ is: $\begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix}$ because $L \Box_{\min}^{\max} \eta = \begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix}$. The vectors $\begin{bmatrix} 0.4 \\ 0.0 \\ 0.0 \end{bmatrix}$ and $\begin{bmatrix} 0.0 \\ 0.0 \\ 0.4 \end{bmatrix}$ are solutions of the system of inequalities $\underline{b^{(2)}}(\Delta(L, b^{(2)})) \leq L \Box_{\min}^{\max} x$ and lower than η. We have $L \Box_{\min}^{\max} \begin{bmatrix} 0.4 \\ 0.0 \\ 0.0 \end{bmatrix} = L \Box_{\min}^{\max} \begin{bmatrix} 0.0 \\ 0.0 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.4 \end{bmatrix}$, therefore, we have a unique minimal Chebyshev approximation of $b^{(2)}$ which is $\begin{bmatrix} 0.4 \\ 0.4 \end{bmatrix}$. We use the greatest Chebyshev approximation. The system $(S_2') : \begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.7 & 0.4 & 0.4 \\ 1.0 & 0.2 & 0.5 \end{bmatrix} \Box_{\min}^{\max} u_2'$ is consistent and it has $\begin{bmatrix} 0.4 \\ 1 \\ 1 \end{bmatrix}$ as greatest solution and one unique minimal solution $\begin{bmatrix} 0 \\ 0 \\ 0.5 \end{bmatrix}$.

- The system (S_3) is inconsistent because $\Delta(L, b^{(3)}) = 0.15$. We use the greatest Chebyshev approximation of $b^{(3)}$: $\begin{bmatrix} 0.15 \\ 0.15 \end{bmatrix}$. The system $(S_3') : \begin{bmatrix} 0.15 \\ 0.15 \end{bmatrix} = \begin{bmatrix} 0.7 & 0.4 & 0.4 \\ 1.0 & 0.2 & 0.5 \end{bmatrix} \Box_{\min}^{\max} u_3'$ is consistent and it has $\begin{bmatrix} 0.15 \\ 0.15 \\ 0.15 \end{bmatrix}$ as greatest solution and three minimal solutions $\begin{bmatrix} 0.15 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0.15 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \\ 0.15 \end{bmatrix}$.

As, we have $\Delta(L, b^{(1)}) = 0$, $\Delta(L, b^{(2)}) = 0.3$ and $\Delta(L, b^{(3)}) = 0.15$, we have $\mu = 0.3$.

From the solutions of (S_1), (S_2') and (S_3'), we can construct an approximate weight matrix W row by row. For instance, $W = \begin{bmatrix} 1 & 0 & 0.2 \\ 0.2 & 1 & 0.5 \\ 0.15 & 0.15 & 0.0 \end{bmatrix}$ where $\begin{bmatrix} 1 \\ 0 \\ 0.2 \end{bmatrix}$ is a solution of (S_1), $\begin{bmatrix} 0.2 \\ 1.0 \\ 0.5 \end{bmatrix}$ is a solution of (S_2') and $\begin{bmatrix} 0.15 \\ 0.15 \\ 0 \end{bmatrix}$ is a solution of (S_3'). From the training data, we observe that:

$$W \Box_{\min}^{\max} x^{(1)} = \begin{bmatrix} 0.7 \\ 0.4 \\ 0.15 \end{bmatrix} \text{ and } \| \begin{bmatrix} 0.7 \\ 0.4 \\ 0.15 \end{bmatrix} - y^{(1)} \| = 0.3 = \mu,$$

$$W \Box_{\min}^{\max} x^{(2)} = \begin{bmatrix} 1 \\ 0.5 \\ 0.15 \end{bmatrix} \text{ and } \| \begin{bmatrix} 1 \\ 0.5 \\ 0.15 \end{bmatrix} - y^{(2)} \| = 0.2 < \mu.$$

In what follows, we study an application of our results: how to approximately learn the rule parameters of a possibilistic rule-based system

In what follows, we study an application of our results: how to approximately learn the rule parameters of a possibilistic rule-based system

Recently, (Dubois and Prade 2020) have emphasized the development of possibilistic learning methods that would be consistent with if-then rule-based reasoning.
For this purpose, (Baaj 2022) introduced a system of min − max fuzzy relational equations for learning the rule parameters of a possibilistic rule-based system according to a training datum:

$$(\Sigma) : Y = \Gamma \Box_{\max}^{\min} X,$$

where:

- $\Box_{\max}^{\min}$ is the matrix product which takes max as the product and min as the addition.
- the second member Y describes an output possibility distribution,
- the matrix Γ contains the possibility degrees of the rule premises and X is an unknown vector containing the rule parameters

If the system (Σ) is inconsistent, e.g., due to poor training data, an approximate solution is desirable.

The general method that we introduced for obtaining approximate solutions of a system of $\max - \min$ fuzzy relational equations can be applied to the case of a $\min - \max$ system such as (Σ).

In (Baaj 2023a):

- we show how to switch from a system of $\min - \max$ fuzzy relational equations such as (Σ) to a system of $\max - \min$ fuzzy relational equations and vice versa

The general method that we introduced for obtaining approximate solutions of a system of $\max - \min$ fuzzy relational equations can be applied to the case of a $\min - \max$ system such as (Σ).

In (Baaj 2023a):

- we show how to switch from a system of $\min - \max$ fuzzy relational equations such as (Σ) to a system of $\max - \min$ fuzzy relational equations and vice versa
- We introduce analogous tools for a system of $\min - \max$ fuzzy relational equations to those already introduced for a system of $\max - \min$ fuzzy relational equations, and we show their correspondences
  (In particular, the analytical formula for the Chebyshev distance ∇ of a $\min - \max$ system)

The general method that we introduced for obtaining approximate solutions of a system of $\max - \min$ fuzzy relational equations can be applied to the case of a $\min - \max$ system such as (Σ).

In (Baaj 2023a):

- we show how to switch from a system of $\min - \max$ fuzzy relational equations such as (Σ) to a system of $\max - \min$ fuzzy relational equations and vice versa
- We introduce analogous tools for a system of $\min - \max$ fuzzy relational equations to those already introduced for a system of $\max - \min$ fuzzy relational equations, and we show their correspondences
(In particular, the analytical formula for the Chebyshev distance ∇ of a $\min - \max$ system)
- From these results, we propose a method for finding approximate solutions of the rule parameters of possibilistic rule-based system when we have multiple training data

Notation : to any matrix $A = [a_{ij}]$, we associate $A^\circ = [1 - a_{ij}]$ and we have $(A^\circ)^\circ = A$.

	System: $A \Box_{\min}^{\max} x = b$	System: $G \Box_{\max}^{\min} x = d$	Relation iff $G = A^\circ$ and $d = b^\circ$
Set of solutions	$\mathcal{S}(A, b)$	$\mathcal{S}(G, d)$	$\mathcal{S}(G, d) = \mathcal{S}(A, b)^\circ$
Potential greatest/lowest solution	$e = A^t \Box_{\rightarrow_G}^{\min} b$ (greatest solution)	$r = G^t \Box_{\epsilon}^{\max} d$ (lowest solution)	$r = e^\circ$
Application computing the matrix product of the system matrix and a given vector in $[0,1]^{m\times 1}$	$\theta : [0,1]^{m\times 1} \to [0,1]^{n\times 1}$ $: x \mapsto A \Box_{\min}^{\max} x$	$\psi : [0,1]^{m\times 1} \to [0,1]^{n\times 1}$ $: x \mapsto G \Box_{\max}^{\min} x$	$\psi(x) = \theta(x^\circ)^\circ$
Set of second members of the consistent systems defined with the matrix	$\mathcal{C} = \{\theta(x) \mid x \in [0,1]^{m\times 1}\}$	$\mathcal{T} = \{\psi(x) \mid x \in [0,1]^{m\times 1}\}$	$\mathcal{T} = \mathcal{C}^\circ$
Application for checking if a system defined with the matrix and a given vector in $[0,1]^{n\times 1}$ as second member is a consistent system	$F : [0,1]^{n\times 1} \to [0,1]^{n\times 1}$ $c \mapsto A \Box_{\min}^{\max} (A^t \Box_{\rightarrow_G}^{\min} c)$	$U : [0,1]^{n\times 1} \to [0,1]^{n\times 1}$ $c \mapsto G \Box_{\max}^{\min} (G^t \Box_{\epsilon}^{\max} c)$	$U(c) = F(c^\circ)^\circ$
Chebyshev distance associated to the second member	$\Delta = \Delta(A, b)$	$\nabla = \nabla(G, d)$	$\nabla(G, d) = \Delta(A, b)$
Set of Chebyshev approximations of the second member	$\mathcal{C}_b$	$\mathcal{T}_d$	$\mathcal{T}_d = \mathcal{C}_b^\circ$
Extremal Chebyshev approximations of the second member	greatest: $F(\bar{b}(\Delta))$ minimal approx. set: $\mathcal{C}_{b,\min}$	lowest: $U(\underline{d}(\nabla))$ maximal approx. set: $\mathcal{T}_{d,\max}$	$U(\underline{d}(\nabla)) = F(\bar{b}(\Delta))^\circ$ $\mathcal{T}_{d,\max} = \mathcal{C}_{b,\min}^\circ$
Approximate solutions set	Λ_b	Υ_d	$\Upsilon_d = \Lambda_b^\circ$
Extremal approximate solutions	greatest: $\eta = A^t \Box_{\rightarrow_G}^{\min} F(\bar{b}(\Delta))$ a min. approx. sol. set: $\Lambda_{b,\min}$	lowest: $\nu = G^t \Box_{\epsilon}^{\max} U(\underline{d}(\nabla))$ a max. approx. sol. set: $\Upsilon_{d,\max}$	$\nu = \eta^\circ$ $\Upsilon_{d,\max} = \Lambda_{b,\min}^\circ$

Table 4: Tools of the systems $A \Box_{\min}^{\max} x = b$ and $G \Box_{\max}^{\min} x = d$ and their relations iff $G = A^\circ$ and $d = b^\circ$.

For an inconsistent system of $\max - T$ fuzzy relational equations $A \Box_T^{\max} x = b$ (where A is a matrix, b a vector and T is a t-norm such as min):

For an inconsistent system of $\max - T$ fuzzy relational equations $A \Box_T^{\max} x = b$ (where A is a matrix, b a vector and T is a t-norm such as min):

- Using the L_∞ norm, we gave an explicit formula for computing the Chebyshev distance associated to b of (S)
- We gave the structure of the approximate solutions set of (S) and that of the Chebyshev approximations set $\mathcal{C}_b$
- We showed how to learn approximate weight matrices whose learning error according to training data is minimal
- We proposed an application of our results: the learning of the rule parameters of possibilistic rule-based system

Currently, we are studying:

- how to get consistent subsystems of fuzzy relational equations (considering a subset of equations)
- how to obtain the "best" approximate solutions of Λ_b (criteria?)
- how to obtain an approximate inverse of a fuzzy matrix governed by $\max - T$ composition
- improve the learning paradigm of approximate weight matrices
- other applications based on systems of $\max - T$ fuzzy relational equations

References I

Ronald R Yager. On the lack of inverses in fuzzy arithmetic. Fuzzy Sets and Systems, 4(1):73–82, 1980.

Elie Sanchez. Resolution of composite fuzzy relation equations. Information and control, 30(1):38–48, 1976

RA Cuninghame-Green and Katarína Cechlárová. Residuation in fuzzy algebra and some applications. Fuzzy Sets and Systems, 71(2):227–239, 1995.

Zofia Matusiewicz and Józef Drewniak. Increasing continuous operations in fuzzy max-* equations and inequalities. Fuzzy Sets and Systems, 232:120–133, 2013

Didier Dubois and Henri Prade. From possibilistic rule-based systems to machine learning-a discussion paper. In International Conference on Scalable Uncertainty Management, pages 35–51. Springer, 2020

Ismaïl Baaj. Learning rule parameters of possibilistic rule-based system. In 2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pages 1–8. IEEE, 2022