Semialgebraic Representation of Monotone Deep Equilibrium Models (monDEQs) and Applications to Certification

Tong Chen, Jean-Bernard Lasserre, Victor Magron and Edouard Pauwels

LAAS-CNRS & IRIT & ANITI, Toulouse, France

Outline of the paper

We introduce the semialgebraic representations of ReLU function to describe the input-output relation of monDEQs, and propose three semidefinite programming (SDP) models for robustness certification.

- Robustness model: semialgebraicity of ReLU
- Lipschitz model: semialgebraicity of ∂ReLU
- Ellipsoid model: sum-of-square (SOS) decomposition

For simplicity, we only present the certification model. The detailed information can be referred to [3].

Structure of monDEQ

A fully-connected monDEQ [1] consists of one input layer $\mathbf{x}$, one implicit layer $\mathbf{z}$ and one output layer. The values of the implicit layer is the solution of an fixed-point equation of the input layer: $\mathbf{z} = \sigma(\mathbf{W}\mathbf{z} + \mathbf{U}\mathbf{x} + \mathbf{u})$, where $\mathbf{W}, \mathbf{U}, \mathbf{u}$ are parameters of the network, and we take $\sigma = \text{ReLU}$ as the activation function.

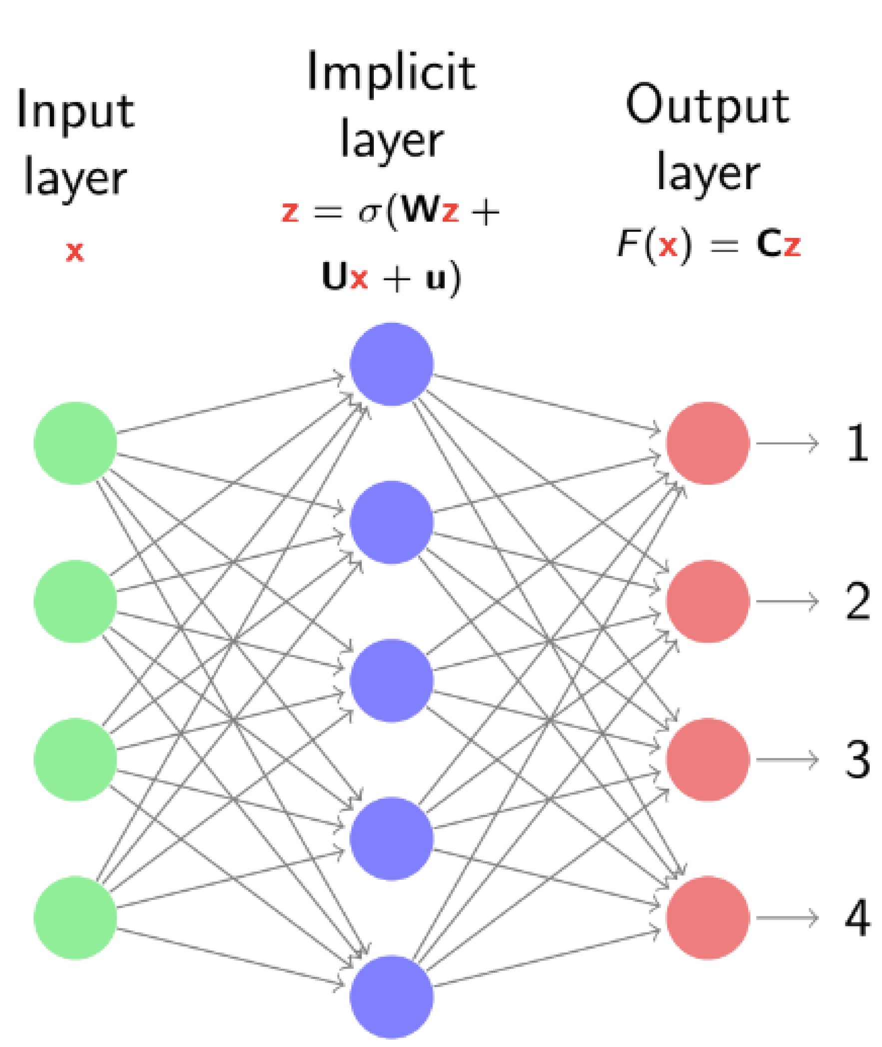
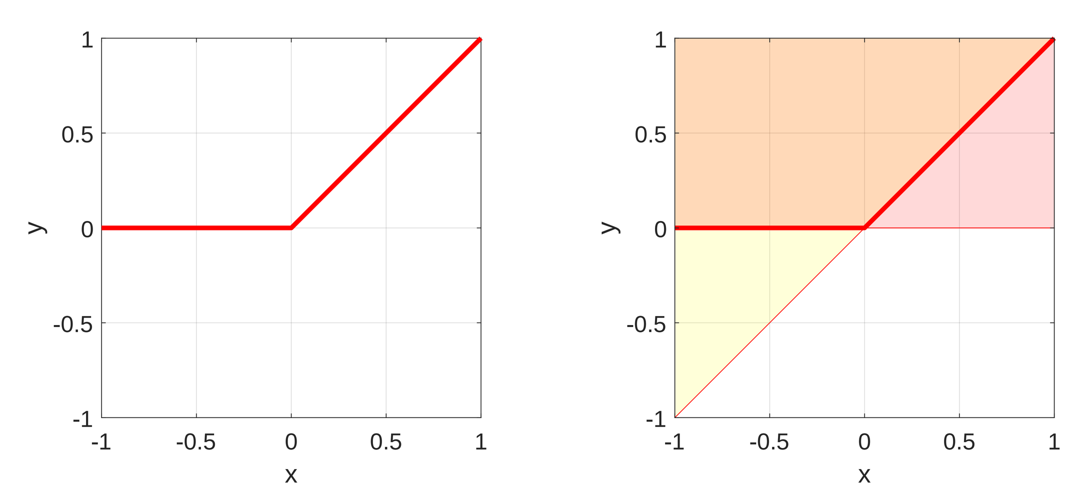


Figure 1:Fully-connected monDEQ

Semialgebraicity of ReLU

If $y = \text{ReLU}(x) = \max\{0, x\}$, it is equivalent to say that $y(y-x) = 0, y \geq x, y \geq 0$.



Hence the implicit layer of monDEQ can be written as:

$$\begin{aligned} &\mathbf{z}(\mathbf{z} - \mathbf{W}\mathbf{z} - \mathbf{U}\mathbf{x}_0 - \mathbf{u}) = 0, \\ &\mathbf{z} \geq \mathbf{W}\mathbf{z} + \mathbf{U}\mathbf{x}_0 + \mathbf{u}, \\ &\mathbf{z} \geq 0. \end{aligned}$$

Using this semialgebraic formulation, we are able to translate the certification problem into *polynomial optimization problem (POP)* which has the following general form:

$$\max_{\mathbf{x} \in \mathbb{R}^n} \{f(\mathbf{x}) : g_i(\mathbf{x}) \geq 0, i = 1, \ldots, p\}$$

where f and g_i are polynomials.

Robustness Model (POP)

Fix an input $\mathbf{x}_0 \in \mathbb{R}^{p_0}$. Let y_0 be the label of $\mathbf{x}_0$ and let $\mathbf{z} \in \mathbb{R}^p$ be the variables in the monDEQ implicit layer. Let $\mathbf{W}, \mathbf{U}, \mathbf{u}, \mathbf{C}$ be the parameters of monDEQ and denote by $\xi_i = (\mathbf{C}_{i,:} - \mathbf{C}_{y_0,:})^T$. The Robustness Model for monDEQ reads:

$$\delta_i := \max \ \xi_i^T \mathbf{z} \quad \text{s.t.} \begin{cases} \mathbf{z} = \text{ReLU}(\mathbf{W}\mathbf{z} + \mathbf{U}\mathbf{x} + \mathbf{u}), \\ \mathbf{x} \in \mathcal{E} \subseteq \mathbb{R}^{p_0}, \mathbf{z} \in \mathbb{R}^p. \end{cases} \tag{1}$$

where the input region $\mathcal{E}$ is the ball (w.r.t. norm $\|\cdot\|$) around $\mathbf{x}_0$ of a preset radius ε, i.e., $\mathcal{E} = \{\mathbf{x} \in \mathbb{R}^{p_0} : \|\mathbf{x} - \mathbf{x}_0\| \leq \varepsilon\}$. Using the semialgebraicity of ReLU function, it is easy to see that problem (1) is a POP.

Semidefinite Programming (SDP)

A real symmetric $n \times n$ matrix $\mathbf{M}$ is said to be *positive semidefinite (PSD)*, denoted by $\mathbf{M} \succeq 0$, if $\mathbf{z}^T\mathbf{M}\mathbf{z} \geq 0$ for all $\mathbf{z} \in \mathbb{R}^n$. A *semidefinite programming (SDP)* can be written in the form:

$$\min_{\mathbf{X} \in \mathbb{S}^n} \{\langle \mathbf{C}, \mathbf{X} \rangle_{\mathbb{S}^n} : \langle \mathbf{A}_k, \mathbf{X} \rangle_{\mathbb{S}^n} = b_k, k = 1, \ldots, m; \mathbf{X} \succeq 0\},$$

where $\mathbb{S}^n$ denotes the the space of all real symmetric $n \times n$ matrices, and $\langle \cdot, \cdot \rangle_{\mathbb{S}^n}$ denotes the Frobenius scalar product in $\mathbb{S}^n$.

Robustness Model (SDP)

Applying Shor's relaxation to POP (1), we obtain an SDP:

$$\max \ \xi_i^T \mathbf{P}[\mathbf{z}] \quad \text{s.t.} \begin{cases} \text{diag}(\mathbf{P}[\mathbf{z}\mathbf{z}^T] - \mathbf{W}\mathbf{P}[\mathbf{z}\mathbf{z}^T] - \mathbf{U}\mathbf{P}[\mathbf{x}\mathbf{z}^T] \\ \quad -\mathbf{u}\mathbf{P}[\mathbf{z}^T]) = 0, \mathbf{P} \succeq 0, \mathbf{P}[1] = 1, \\ \mathbf{P}[\mathbf{z}] \geq \mathbf{W}\mathbf{P}[\mathbf{z}] + \mathbf{U}\mathbf{P}[\mathbf{x}] + \mathbf{u}, \mathbf{P}[\mathbf{z}] \geq 0, \\ \mathbf{1}^T\text{diag}(\mathbf{P}[\mathbf{x}\mathbf{x}^T]) - 2\mathbf{x}_0^T\mathbf{P}[\mathbf{x}] + \mathbf{x}_0^T\mathbf{x}_0 \geq 0, \ (L_2) \\ \text{diag}(\mathbf{P}[\mathbf{x}\mathbf{x}^T] - 2\mathbf{x}_0\mathbf{P}[\mathbf{x}^T] + \mathbf{x}_0\mathbf{x}_0^T). \ (L_\infty) \end{cases} \tag{2}$$

where the symmetric matrix $\mathbf{P}$ is defined by

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}[1] & \mathbf{P}[\mathbf{x}^T] & \mathbf{P}[\mathbf{z}^T] \\ \mathbf{P}[\mathbf{x}] & \mathbf{P}[\mathbf{x}\mathbf{x}^T] & \mathbf{P}[\mathbf{x}\mathbf{z}^T] \\ \mathbf{P}[\mathbf{z}] & \mathbf{P}[\mathbf{z}\mathbf{x}^T] & \mathbf{P}[\mathbf{z}\mathbf{z}^T] \end{bmatrix}.$$

The optimal solution $\tilde{\delta}_i$ of (2) is an upper bound of δ_i, i.e., $\delta_i \leq \tilde{\delta}_i$. One can certify robustness of monDEQs based on the values of $\tilde{\delta}_i$: if $\tilde{\delta}_i < 0$ for all $i \neq y_0$, then the network F is robust at $\mathbf{x}_0$.

References

[1] Ezra Winston and J. Zico Kolter.
Monotone operator equilibrium networks.

[2] Chirag Pabbaraju, Ezra Winston, and J. Zico Kolter.
Estimating lipschitz constants of monotone deep equilibrium models.

[3] Tong Chen, Jean B Lasserre, Victor Magron, and Edouard Pauwels.
Semialgebraic representation of monotone deep equilibrium models and applications to certification.

Numerical Results

Based on the first 100 test examples of MNIST dataset, we compute the ratio of certified examples for robustness model. We compare our SDP-based method with the state-of-the-art in [2]. We consider L_2 norm with $\varepsilon = 0.1$ and L_∞ norm with $\varepsilon = 0.1, 0.05, 0.01$.

Norm	ε	Our method	Pabbaraju et. al.
L_2	0.1	99%	91%
L_∞	0.1	0%	0%
	0.05	24%	0%
	0.01	99%	24%

Table 1:Ratio of certified test examples

From Table 1, we see that our method outperforms the method in [2] for all the cases. Another interesting phenomenon is that, for $\varepsilon = 0.1$, we can certify 99% of the examples for L_2 norm while 0% for L_∞ norm. Compared to the traditional deep neural networks, this means monDEQs are less robust with respect to L_∞ norm.

Contact Information

- Web: http://www.github.com/TongCHEN779
- Email: tchen@laas.fr

NEURAL INFORMATION PROCESSING SYSTEMS

ANITI ARTIFICIAL & NATURAL INTELLIGENCE TOULOUSE INSTITUTE

LAAS CNRS

UNIVERSITÉ TOULOUSE III PAUL SABATIER