Semialgebraic Optimization for Lipschitz Constants of ReLU Networks

Tong Chen, Jean-Bernard Lasserre, Victor Magron, Edouard Pauwels
tchen@laas.fr, lasserre@laas.fr, vmagron@laas.fr, edouard.pauwels@irit.fr

Problem Settings

Computing the upper bounds of Lipschitz constants (with respect to norm $||\cdot||$) of fully-connected ReLU networks. Notations:
$\mathbf{A}, \mathbf{b}, \mathbf{c}$: parameters of the network;
m: number of hidden layers;
$\mathbf{t}$: variables that dualize the norm $||\cdot||$;
$\mathbf{u}$: lifting variables of the derivative of ReLU function;
$\mathbf{x}$: variables in each layer;
$\mathbf{xy}$: product of two vectors is considered as coordinate-wise product.

Mathematical Formulation

$$\max_{\mathbf{x}_i,\mathbf{u}_i,\mathbf{t}} \mathbf{t}^T\Big(\prod_{i=1}^{m}\mathbf{A}_i^T\mathrm{diag}(\mathbf{u}_i)\Big)\mathbf{c} \qquad \text{(L)}$$

$$\text{s.t.} \begin{cases} (\mathbf{u}_i - \frac{1}{2})(\mathbf{A}_i\mathbf{x}_{i-1} + \mathbf{b}_i) \geq 0, \mathbf{u}_i(\mathbf{u}_i - 1) = 0; \\ \mathbf{x}_{i-1}(\mathbf{x}_{i-1} - \mathbf{A}_{i-1}\mathbf{x}_{i-2} - \mathbf{b}_{i-1}) = 0, \\ \mathbf{x}_{i-1} \geq 0, \mathbf{x}_{i-1} \geq \mathbf{A}_{i-1}\mathbf{x}_{i-2} + \mathbf{b}_{i-1}; \\ \mathbf{t}^2 \leq 1, (\mathbf{x}_0 - \bar{\mathbf{x}}_0 + \varepsilon)(\mathbf{x}_0 - \bar{\mathbf{x}}_0 - \varepsilon) \leq 0\,. \end{cases}$$

Methods

SHOR: Shor's relaxation applied to (L);
HR-1/2: 1st/2nd-order heuristic relaxation applied to (L);
LipOpt-3/4: LP-based method by Latorre et. al. with degree 3/4;
LBS: Lower bound computed by random sampling.

Experimental Settings

For **SHOR** and **HR-1/2**, use MATLAB with Mosek as a backend; for **LipOpt-3/4**, use Python with Gurobi as a backend.
OfM means running out of memory during building the model. Computational time is considered as the solver running time with unit second. All experiments are run on a personal laptop with a 4-core i5-6300HQ 2.3GHz CPU and 8GB of RAM.

References

[1] Fabian Latorre, Paul Rolland, Volkan Cevher: *Lipschitz constant estimation of Neural Networks via sparse polynomial optimization*, ICLR2020.
[2] Tong Chen, Jean-Bernard Lasserre, Victor Magron, Edouard Pauwels: *Semialgebraic Optimization for Lipschitz Constants of ReLU Networks*, NeurIPS 2020.

Acknowledgements

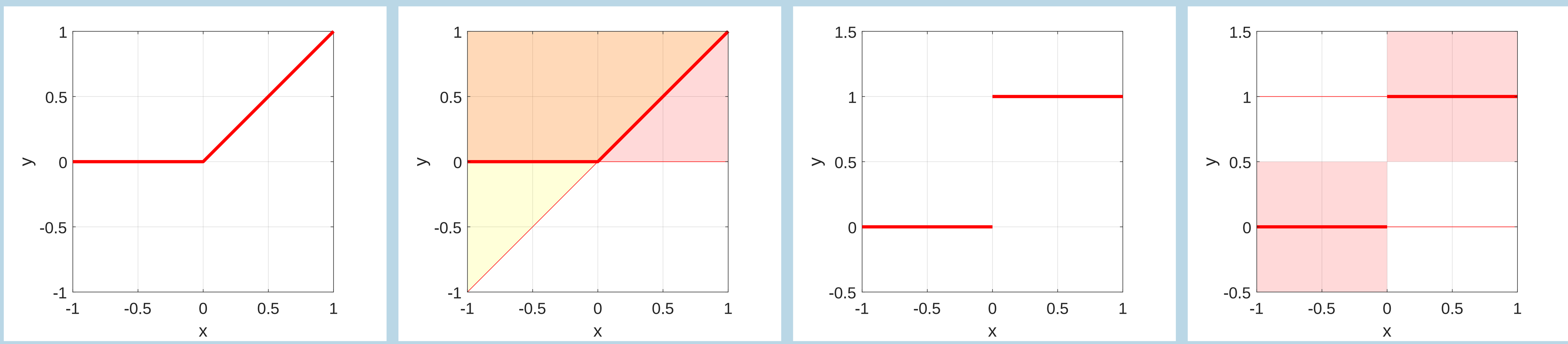
This work has benefited from the AI Interdisciplinary Institute ANITI funding, through the French "Investing for the Future - PIA3" program under the Grant agreement n°ANR-19-PI3A-0004. Edouard Pauwels acknowledges the support of Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant numbers FA9550-19-1-7026 and FA9550-18-1-0226, and ANR MasDol. Victor Magron was supported by the FMJH Program PGMO (EPICS project) and EDF, Thales, Orange et Criteo, the Tremplin ERC Stg Grant ANR-18-ERC2-0004-01 (T-COPS project) as well as the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Actions, grant agreement 813211 (POEMA).

Lasserre's Hierarchy and Its Sparse Version

	Dense	Sparse
Original Problems	$\inf_{\mathbf{x}\in\mathbb{R}^n} \{f(\mathbf{x}) : g_i(\mathbf{x}) \geq 0, i \in [p]\}$	$\inf_{\mathbf{x}\in\mathbb{R}^n} \{f(\mathbf{x}) : g_i(\mathbf{x}_{I_{k(i)}}) \geq 0, i \in [p]\}$
Moment Problems	$\inf_{\mathbf{y}}\{ L_{\mathbf{y}}(f) : L_{\mathbf{y}}(1) = 1, \mathbf{M}_d(\mathbf{y}) \succeq 0, \mathbf{M}_{d-\omega_i}(g_i\mathbf{y}) \succeq 0, i \in [p]\}$	$\inf_{\mathbf{y}}\{ L_{\mathbf{y}}(f) : L_{\mathbf{y}}(1) = 1, \mathbf{M}_d(\mathbf{y}, I_k) \succeq 0, k \in [l]; \mathbf{M}_{d-\omega_i}(g_i\,\mathbf{y}, I_{k(i)}) \succeq 0, i \in [p]\}$
Number of SDPs	$1+p$	$l+p$
Size of SDPs	$\binom{n+2d}{2d}, \binom{n+2(d-\omega_i)}{2(d-\omega_i)}$	$\binom{\lvert I_k\rvert+2d}{2d}, \binom{\lvert I_k\rvert+2(d-\omega_i)}{2(d-\omega_i)}$

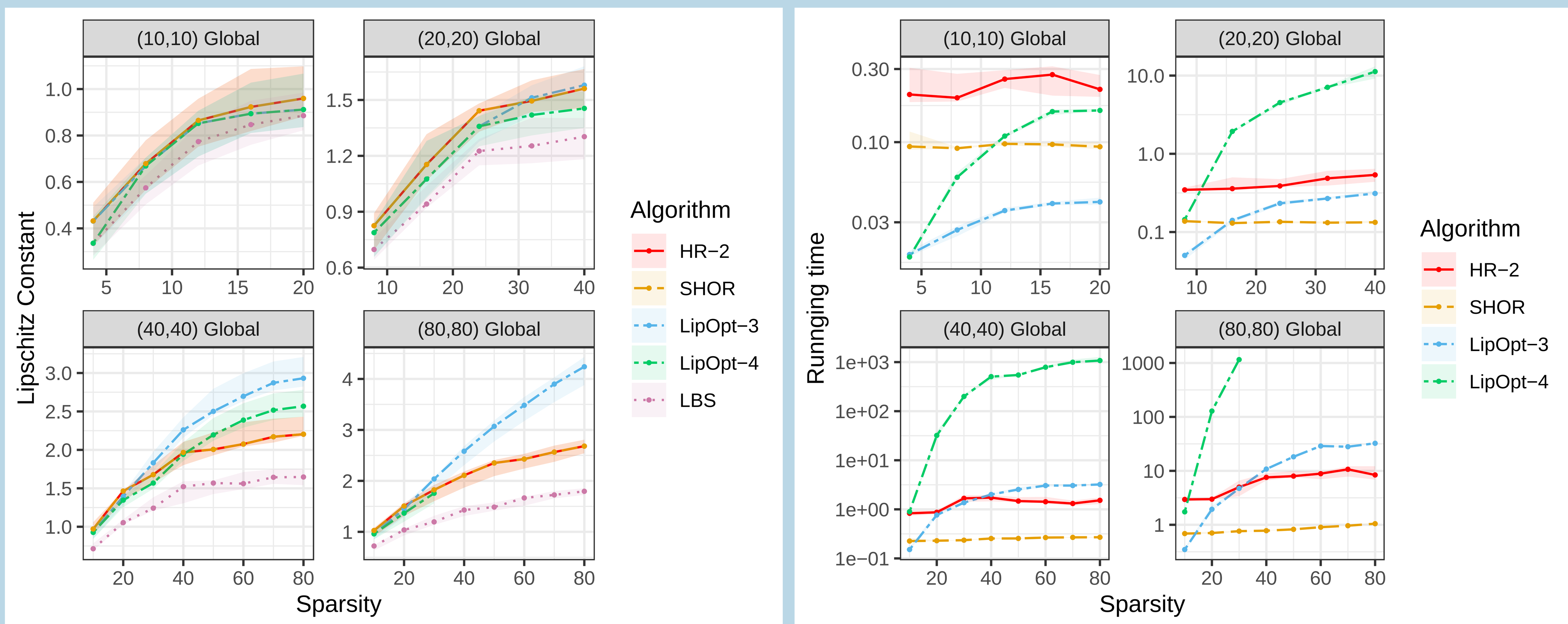
Semialgebraic Expression of ReLU Function and Its Derivative

Semialgebraic expression of ReLU function: $y = \max\{x, 0\} \Leftrightarrow y(y-x) = 0, y \geq x, y \geq 0$;
Semialgebraic expression of the derivative of ReLU function: $y = \mathbf{1}_{\{x\geq 0\}} \Leftrightarrow y(y-1) = 0, (y-\frac{1}{2})x \geq 0$.

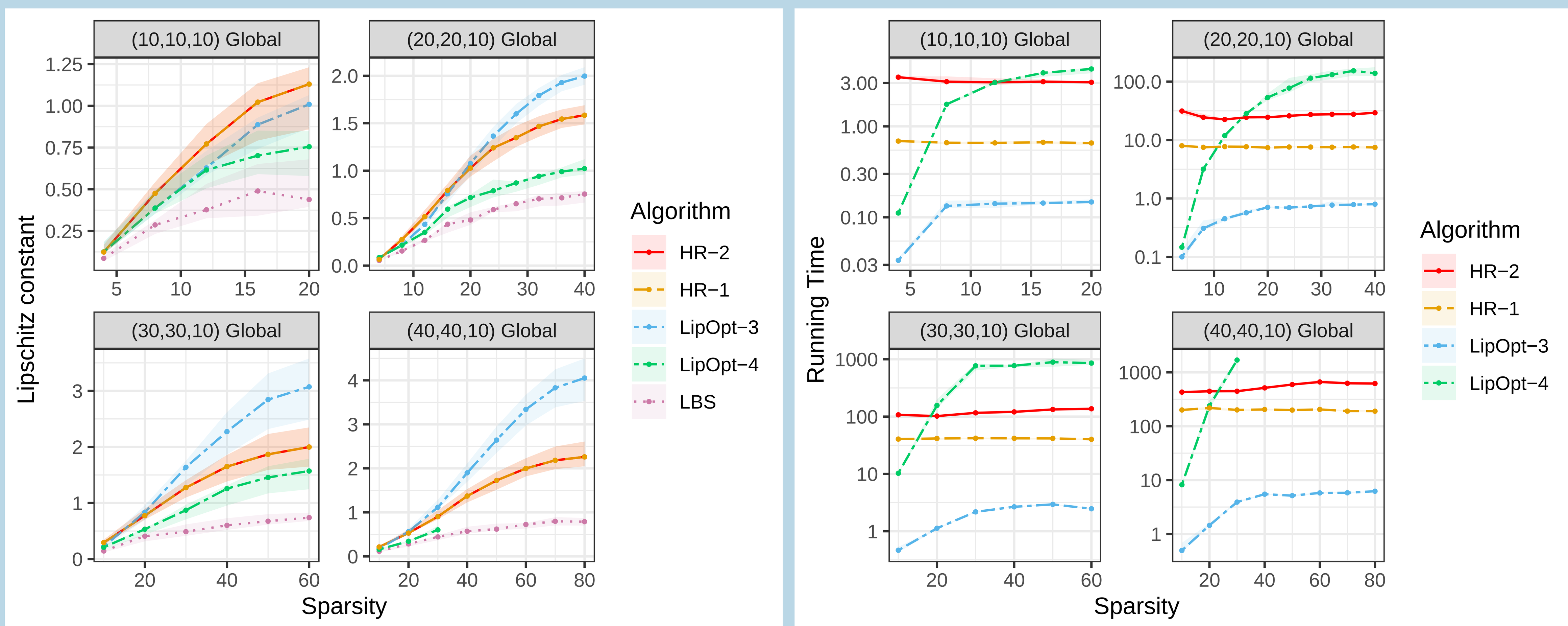


Experiments on Random Networks

Upper bounds of global Lipschitz constant and running time for 1-hidden layer networks.



Upper bounds of global Lipschitz constant and running time for 2-hidden layer networks.



Experiments on Trained Network (SDP-NN)

Upper bounds of Lipschitz constant and running time of various methods for SDP-NN network.

	Global				Local			
	HR-2	**SHOR**	**LipOpt-3**	**LBS**	**HR-2**	**SHOR**	**LipOpt-3**	**LBS**
Bound	14.56	17.85	OfM	9.69	12.70	16.07	OfM	8.20
Time	12246	2869	OfM	-	20596	4217	OfM	-