Polynomial Optimization for Estimating the Lipschitz Constant of Neural Networks

Tong Chen, LAAS-CNRS

September 11, 2020

Background
Deep Neural Networks

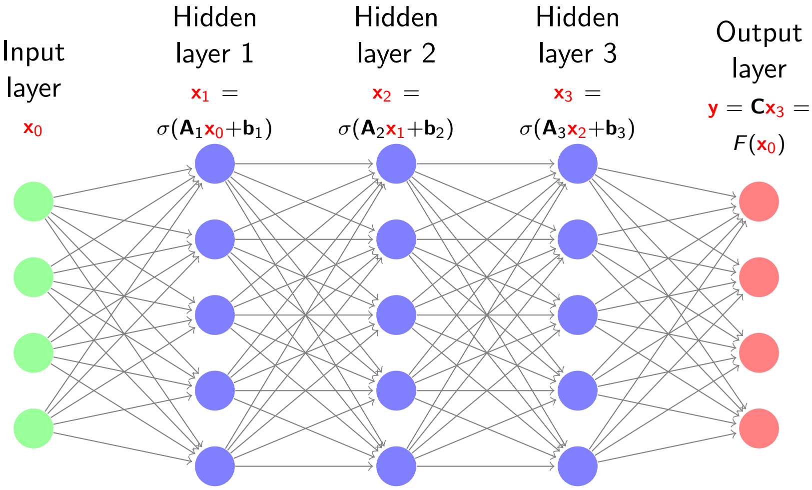
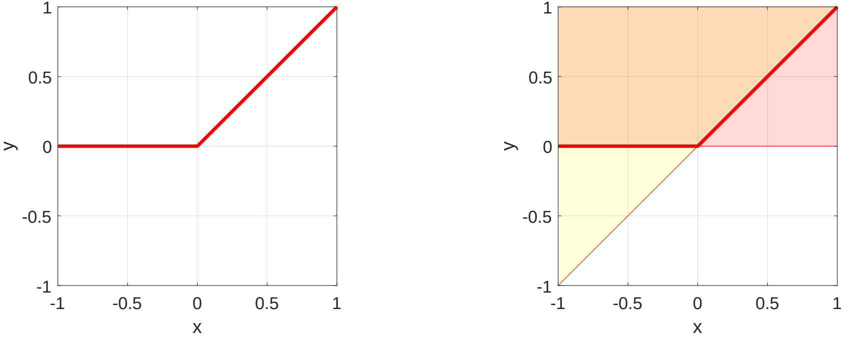


Figure: Fully-connected neural network

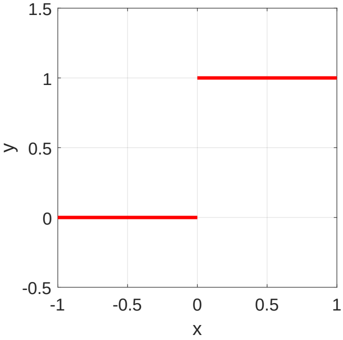
ReLU function (left) and its semialgebraicity (right)



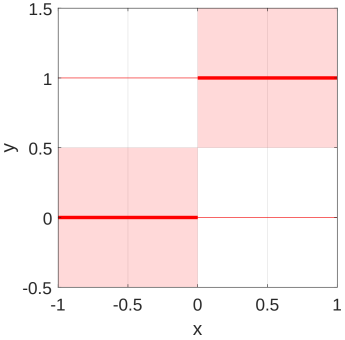
(a) $y = \max\{x, 0\}$

(b) $y(y - x) = 0, y \geq x, y \geq 0$

Derivative of ReLU function (left) and its semialgebraicity (right)



(a) $y = \mathbf{1}_{\{x \geq 0\}}$



(b) $y(y-1) = 0, (y - \frac{1}{2})x \geq 0$

Lipschitz Constant of Neural Networks

- Lipschitz constant (general):

$$L_f^{||\cdot||} := \inf\{L : \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}, |f(\mathbf{x}) - f(\mathbf{y})| \leq L||\mathbf{x} - \mathbf{y}||\}\,.$$

- Lipschitz constant (for neural network):

$$L_F^{||\cdot||_\infty} = \max_{\mathbf{t}, \mathbf{x}_i, \mathbf{u}_i} \mathbf{t}^T \left(\prod_{i=1}^{m} \mathbf{A}_i^T \mathrm{diag}(\mathbf{u}_i) \right) \mathbf{c}$$

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

Sparse Lasserre's Relaxation

- Simplest case: $m = 1$, $\mathbf{A}$ of size $p \times p$:

$$L_F^{||\cdot||_\infty} = \max_{\mathbf{t},\mathbf{x},\mathbf{u}} \mathbf{t}^T \mathbf{A}^T \mathrm{diag}(\mathbf{u}) \mathbf{c}$$

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

- Cliques:
  $I = \{x_1, \ldots, x_p, u_1, \ldots, u_p\}, J_i = \{u_1, \ldots, u_p, t_i\}, i = 1, \ldots, p.$

Sparse Lasserre's Relaxation

- 2nd-order sparse relaxation:

$$\rho_2 = \max_{\mathbf{y}} \; L_{\mathbf{y}}(\mathbf{t}^T \mathbf{A}^T \mathrm{diag}(\mathbf{u})\mathbf{c})$$

$$\text{s.t.} \begin{cases} \mathbf{M}_2(\mathbf{y}, I) \succeq 0, \mathbf{M}_2(\mathbf{y}, J_i) \succeq 0, L_{\mathbf{y}}(1) = 1; \\ \mathbf{M}_1(u_i(u_i - 1)\mathbf{y}, J_i) = 0, \\ \mathbf{M}_1((u_i - 1/2)(\mathbf{A}_{i,:}\mathbf{x} + b_i)\mathbf{y}, I) \succeq 0\,; \\ \mathbf{M}_1((1 - t_i^2)\mathbf{y}, J_i) \succeq 0, \\ \mathbf{M}_1(-(x_i - \bar{x}_i + \varepsilon)(x_i - \bar{x}_i - \varepsilon)\mathbf{y}, I) \succeq 0\,. \end{cases}$$

- $|I| = 2p$, $\mathbf{M}_2(\mathbf{y}, I)$ of size $\binom{2p+2}{2} = (p+1)(2p+1) = O(p^2)$.
- $|J_i| = p + 1$, $\mathbf{M}_2(\mathbf{y}, J_i)$ of size $\binom{p+3}{2} = (p+3)(p+2)/2 = O(p^2)$.

Heuristic Relaxation

- Reduce the size of the cliques:

$$I = \{x_1, \ldots, x_p, u_1, \ldots, u_p\} \longrightarrow I_i = \{x_i\}$$
$$J_i = \{u_1, \ldots, u_p, t_i\} \longrightarrow J_i = \{u_i, t_i\}$$

  These cliques **no longer** satisfies the RIP condition.

- Reduce the order of the relaxation w.r.t. dense constraints:

$$\begin{aligned} &\mathbf{M}_1((u_i - 1/2)(\mathbf{A}_{i,:}\mathbf{x} + b_i)\mathbf{y}, I) \\ \longrightarrow &\mathbf{M}_0((u_i - 1/2)(\mathbf{A}_{i,:}\mathbf{x} + b_i)\mathbf{y}, I) = L_{\mathbf{y}}((u_i - 1/2)(\mathbf{A}_{i,:}\mathbf{x} + b_i)) \end{aligned}$$

- Add a full 1st-order moment matrix $\mathbf{M}_1(\mathbf{y})$ to make the relaxation feasible.

Heuristic Relaxation

- Recall: 2nd-order sparse relaxation:

$$\rho_2 = \max_{\mathbf{y}} \; L_{\mathbf{y}}(\mathbf{t}^T \mathbf{A}^T \mathrm{diag}(\mathbf{u})\mathbf{c})$$

$$\text{s.t.} \begin{cases} \mathbf{M}_2(\mathbf{y}, I) \succeq 0, \mathbf{M}_2(\mathbf{y}, J_i) \succeq 0, L_{\mathbf{y}}(1) = 1; \\ \mathbf{M}_1(u_i(u_i - 1)\mathbf{y}, J_i) = 0, \\ \mathbf{M}_1((u_i - 1/2)(\mathbf{A}_{i,:}\mathbf{x} + b_i)\mathbf{y}, I) \succeq 0\,; \\ \mathbf{M}_1((1 - t_i^2)\mathbf{y}, J_i) \succeq 0, \\ \mathbf{M}_1(-(x_i - \bar{x}_i + \varepsilon)(x_i - \bar{x}_i - \varepsilon)\mathbf{y}, I) \succeq 0\,. \end{cases}$$

Heuristic Relaxation

- 2nd-order heuristic relaxation:

$$h_2 = \max_{\mathbf{y}} \; L_{\mathbf{y}}(\mathbf{t}^T \mathbf{A}^T \mathrm{diag}(\mathbf{u})\mathbf{c})$$

$$\text{s.t.} \begin{cases} \boxed{\mathbf{M}_1(\mathbf{y}) \succeq 0}, \mathbf{M}_2(\mathbf{y}, \boxed{\{x_i\}}) \succeq 0, \mathbf{M}_2(\mathbf{y}, \boxed{\{u_i, t_i\}}) \succeq 0, L_{\mathbf{y}}(1) = 1; \\ \mathbf{M}_1(u_i(u_i - 1)\mathbf{y}, \boxed{\{u_i, t_i\}}) = 0, \\ \boxed{L_{\mathbf{y}}((u_i - 1/2)(\mathbf{A}_{i,:}\mathbf{x} + b_i)) \geq 0}; \\ \mathbf{M}_1((1 - t_i^2)\mathbf{y}, \boxed{\{u_i, t_i\}}) \succeq 0, \\ \mathbf{M}_1(-(x_i - \bar{x}_i + \varepsilon)(x_i - \bar{x}_i - \varepsilon)\mathbf{y}, \boxed{\{x_i\}}) \succeq 0\,. \end{cases}$$

- $\rho_1 \leq h_2 \leq \rho_2$.

Main Results (HR-2 v.s. SHOR)

Trained $(784, 500)$ network (**SDP-NN**)

		HR-2	**SHOR**	**LBS**
Glob.	Bound	14.56	17.85	9.69
	Time	12246	2869	-
Loc.	Bound	12.70	16.07	8.20
	Time	20596	4217	-

- LBS: lower bound given by random sampling.
- More information: `https://arxiv.org/abs/2002.03657`.