Exactness of SDP Relaxation for Robustness Verification of Neural Networks

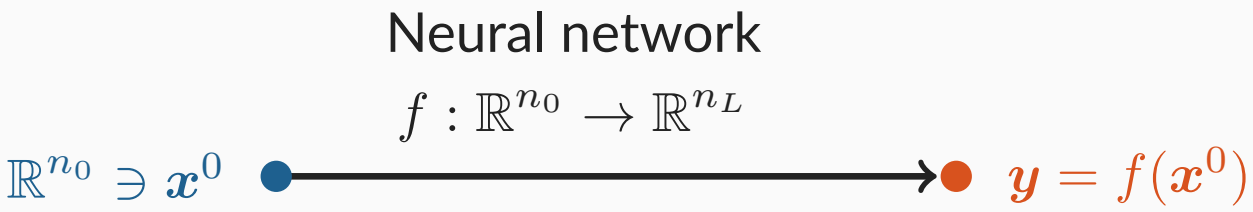
Godai Azuma (Aoyama Gakuin University)

Joint work with Sunyoung Kim (Ewha W. University)
Makoto Yamashita (Institute of Science Tokyo)

CONTO2025, Kyoto University (Dec 02, 2025)

This work was supported by JSPS KAKENHI JP24K20738, JP22KJ1307 and JP20H04145
NRF 2021-R1A2C1003810

Safety Verification of Neural Networks (NNs)



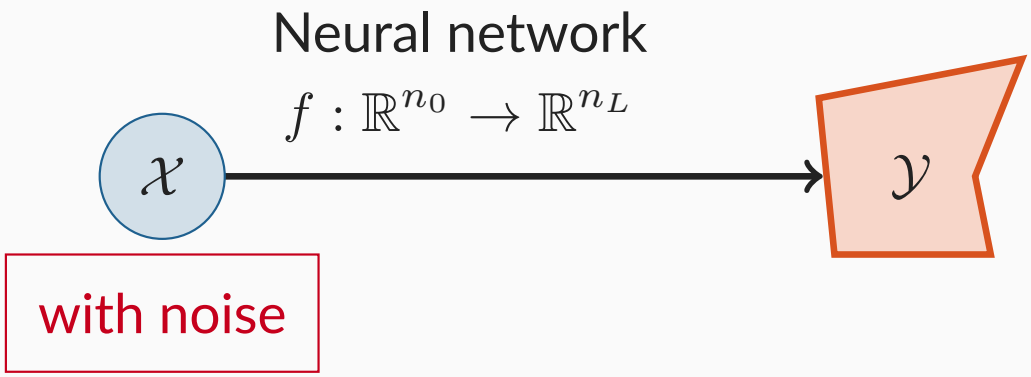
+ noise

- adversarial attacks
- uncertainty

Safety Verification

- Evaluation of the robustness of $\boldsymbol{x}^0$
- By verification of $\mathcal{Y} \subseteq S_y$ from given $\mathcal{X}, f, S_y$

Safety Verification of Neural Networks (NNs)

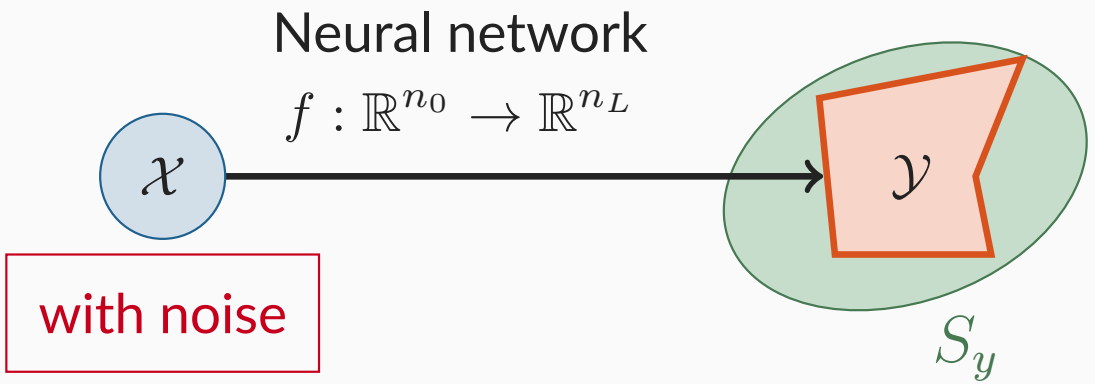


- adversarial attacks
- uncertainty

Safety Verification

- Evaluation of the robustness of $\boldsymbol{x}^0$
- By verification of $\mathcal{Y} \subseteq S_y$ from given $\mathcal{X}, f, S_y$

Safety Verification of Neural Networks (NNs)



with noise

- adversarial attacks
- uncertainty

same class with y

safety specification set

Safety Verification

- Evaluation of the robustness of x^0
- By verification of $\mathcal{Y} \subseteq S_y$ from given $\mathcal{X}, f, S_y$

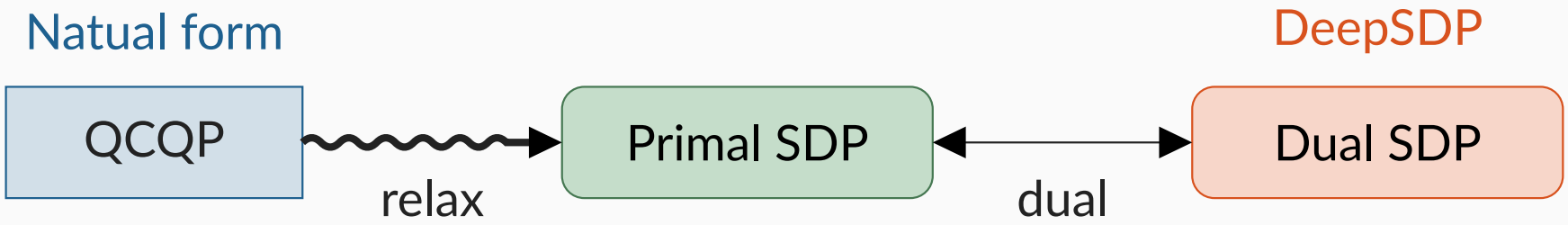
DeepSDP

Semidefinite programming(SDP)-based method can be considered

- Quadratic constraints can formulate NN with ReLU ϕ

$$y = \phi(x) := \max\{0, x\} \quad \Longleftrightarrow \quad \{y \geq 0, \quad y \geq x, \quad y(y - x) = 0\}$$

- DeepSDP is dual of SDP relaxation
  (solvable in polynomial time)



Exactness of Relaxation and Our Motivation

Two gaps exist

- Duality gap (Primal SDP and Dual SDP)
  - $= 0$ under strong duality

- Relaxation gap (QCQP and Primal SDP)
  - $= 0 \iff$ SDP relaxation is *exact*

Motivation of this talk

- Assume strong duality $\Rightarrow$ no duality gap
- Under what conditions is the primal SDP relaxation exact?

Outline

- Introduction

- Quadratically constrained quadratic programming (QCQP)
- Exact semidefinite programming (SDP) relaxation

- Single-layer feed-forward neural network
- QCQP formulation for safety verification
- Exactness conditions for SDP relaxation

- Graphical explanation of exactness

- Summary

QCQP: Quadratically Constrained Quadratic Programming

Quadratic objective function and quadratic constraints

$$
\begin{aligned}
v^* \;:=\; \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \boldsymbol{x}^{\mathrm{T}} Q^0 \boldsymbol{x} + 2\big(\boldsymbol{q}^0\big)^{\mathrm{T}} \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{x}^{\mathrm{T}} Q^p \boldsymbol{x} + 2\big(\boldsymbol{q}^p\big)^{\mathrm{T}} \boldsymbol{x} \le b_p, \quad p \in [m] := \{1, \ldots, m\}.
\end{aligned}
\tag{$\mathcal{P}$}
$$

- Generally non-convex & NP-hard
- Approximately solvable via SDP relaxation

SDP Relaxation

QCQP

$$v^* = \min\left\{ \boldsymbol{x}^{\mathrm{T}} Q^0 \boldsymbol{x} + 2\left(\boldsymbol{q}^0\right)^{\mathrm{T}} \boldsymbol{x} \,\middle|\, \boldsymbol{x}^{\mathrm{T}} Q^p \boldsymbol{x} + 2\left(\boldsymbol{q}^p\right)^{\mathrm{T}} \boldsymbol{x} \le b_p,\ p \in [m] \right\} \qquad (\mathcal{P})$$

$$= \min\left\{ Q^0 \bullet X + 2\left(\boldsymbol{q}^0\right)^{\mathrm{T}} \boldsymbol{x} \,\middle|\, \begin{array}{l} X = \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} \\ Q^p \bullet X + 2\left(\boldsymbol{q}^0\right)^{\mathrm{T}} \boldsymbol{x} \le b_p,\ p \in [m] \end{array} \right\}$$

$$\ge \min\left\{ Q^0 \bullet X + 2\left(\boldsymbol{q}^0\right)^{\mathrm{T}} \boldsymbol{x} \,\middle|\, \begin{array}{l} X \succeq \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} \\ Q^p \bullet X + 2\left(\boldsymbol{q}^0\right)^{\mathrm{T}} \boldsymbol{x} \le b_p,\ p \in [m] \end{array} \right\} \qquad (\mathcal{P}_{\mathrm{R}})$$

Semidefinite Programming (SDP) Relaxation

$$=: v^*_{\mathrm{SDP}}$$

Notation

- $Q^p \bullet X := \sum_{i,j} Q^p_{ij} X_{ij}$: Frobenius inner product.
- $X \succeq \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} \iff X - \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}$ is positive semidefinite.

Exact SDP Relaxation

Def: Exactness

SDP relaxation ($\mathcal{P}_{\mathrm{R}}$) is exact (tight) if $v^* = v^*_{\mathrm{SDP}}$

- Exact $\iff$ ($\mathcal{P}_{\mathrm{R}}$) has a rank-1 solution X^*

$$X^* \succeq \boldsymbol{x}^*(\boldsymbol{x}^*)^{\mathrm{T}} \quad \iff \quad \begin{bmatrix} 1 & (\boldsymbol{x}^*)^{\mathrm{T}} \\ \boldsymbol{x}^* & X^* \end{bmatrix} \succeq O.$$

- Sufficient conditions for general QCQPs have been studied
  - Sparsity and graphs
  - Projection of epigraphs
  - Collinearity of Gram matrix representation

Gram Matrix Transformation

Replace (x, X) in SDP relaxation:

- Fix $e \in \mathbb{R}^s$ with $\|e\| = 1$
- Introduce $u^1, \ldots, u^n \in \mathbb{R}^s$ so that

$$\left[\begin{array}{c|c} 1 & x^{\mathrm{T}} \\ \hline x & X \end{array}\right] = \left[\begin{array}{c|ccc} e^{\mathrm{T}}e & e^{\mathrm{T}}u^1 & \cdots & e^{\mathrm{T}}u^n \\ \hline e^{\mathrm{T}}u^1 & (u^1)^{\mathrm{T}}u^1 & \cdots & (u^1)^{\mathrm{T}}u^n \\ \vdots & \vdots & \ddots & \vdots \\ e^{\mathrm{T}}u^n & (u^n)^{\mathrm{T}}u^1 & \cdots & (u^n)^{\mathrm{T}}u^n \end{array}\right]$$

$$\begin{aligned} Q^p \bullet X + 2(q^p)^{\mathrm{T}}x &= \sum_i \sum_j Q^p_{ij} X_{ij} + 2\sum_i q^p_i x_i \\ &= \sum_i \sum_j Q^p_{ij} (u^i)^{\mathrm{T}} u^j + 2\sum_i q^p_i e^{\mathrm{T}} u^i \end{aligned}$$

Exactness Condition for Gram Matrix Representation

Finally, we obtain

$$
\begin{aligned}
v^*_{\mathrm{SDP}} = \min \quad & \sum_i \sum_j Q^0_{ij} \left(\boldsymbol{u}^i\right)^{\mathrm{T}} \boldsymbol{u}^j + 2 \sum_i q^0_i \boldsymbol{e}^{\mathrm{T}} \boldsymbol{u}^i \\
\text{s.t.} \quad & \sum_i \sum_j Q^p_{ij} \left(\boldsymbol{u}^i\right)^{\mathrm{T}} \boldsymbol{u}^j + 2 \sum_i q^p_i \boldsymbol{e}^{\mathrm{T}} \boldsymbol{u}^i \leq b_p, \quad p \in [m], \\
& \boldsymbol{u}^1, \ldots, \boldsymbol{u}^n \in \mathbb{R}^n.
\end{aligned}
$$

Proposition [Z20]

SDP relaxation is exact if there exist an optimal solution $\left(\boldsymbol{u}^1\right)^*, \ldots, \left(\boldsymbol{u}^n\right)^*$ which are collinear to $\boldsymbol{e}$, i.e.,

$$
\left|\boldsymbol{e}^{\mathrm{T}} \boldsymbol{u}^i\right| = \left\|\boldsymbol{u}^i\right\| \quad \text{for all } i \in [n].
$$

[Z20] Zhang, On the tightness of semidefinite relaxations for certifying robustness to adversarial examples, NeurIPS, 2020.

Outline

- Introduction
- Quadratically constrained quadratic programming (QCQP)
- Exact semidefinite programming (SDP) relaxation
- Single-layer feed-forward neural network
- QCQP formulation for safety verification
- Exactness conditions for SDP relaxation
- Graphical explanation of exactness
- Summary

Single-layer Neural Networks

W^0, W^1: weight matrices, $\boldsymbol{b}^0, \boldsymbol{b}^1$: bias vectors

Neural network

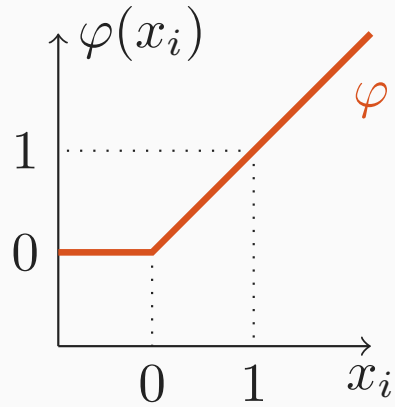
$$\boldsymbol{x}^1 := \phi(W^0 \boldsymbol{x}^0 + \boldsymbol{b}^0),$$
$$f(\boldsymbol{x}^0) := W^1 \boldsymbol{x}^1 + \boldsymbol{b}^1.$$

Note we consider the case that

- $W^1 = I$, and $\boldsymbol{b}^1 = \boldsymbol{0}$.
- ϕ is an element-wise ReLU function, i.e.,

$$\phi(\boldsymbol{x}) := \begin{bmatrix} \varphi(x_1) & \cdots & \varphi(x_n) \end{bmatrix}^{\mathrm{T}},$$

where $\varphi(x_i) := \max\{0, x_i\}$.



Input Set $\mathcal{X} \subseteq \boldsymbol{R}^{n_0}$

Set $\mathcal{X}$ contains the uncertainty and attacks.

- Each input $\boldsymbol{x}^0$ is chosen from $\mathcal{X}$.
- The safety of $\boldsymbol{x}^0$ is evaluated by $\boldsymbol{S}_y$.

Note

$\mathcal{X}$ is not the domain of NN f.

Various shapes are possible.

- hyper-ellipsoid $\quad \mathcal{X} = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \leq \rho\}$.
- hyper-rectangle $\quad \mathcal{X} = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_\infty \leq \rho\}$.

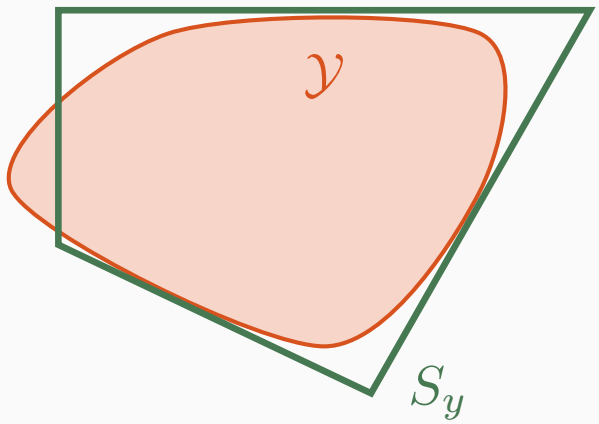
Setting II

This talk covers the case where $\mathcal{X}$ is a hyper-ellipsoid.

Polytope Safety Specification Set

Consider polytope safety specification set S_y

- Let S_y be a quadrilateral below.
- $\mathcal{Y} \subseteq S_y$ can be verified via four half-spaces.



Setting I

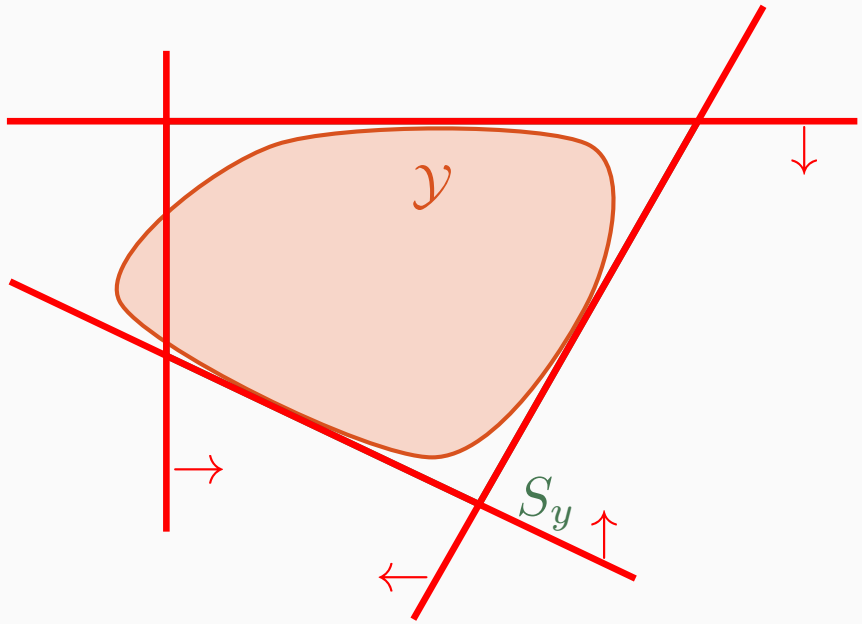
Assume that safety specification set S_y is a half-space

$$H \coloneqq \left\{ \boldsymbol{y} \in \mathbb{R}^{n_2} \,\middle|\, \boldsymbol{c}^{\mathrm{T}} \boldsymbol{y} - d \geq 0 \right\}.$$

Polytope Safety Specification Set

Consider polytope safety specification set S_y

- Let S_y be a quadrilateral below.
- $\mathcal{Y} \subseteq S_y$ can be verified via four half-spaces.



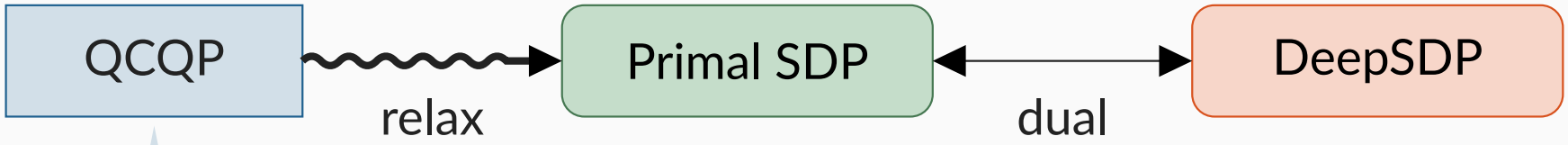
Setting I

Assume that safety specification set S_y is a half-space

$$H \coloneqq \left\{ \boldsymbol{y} \in \mathbb{R}^{n_2} \,\middle|\, \boldsymbol{c}^{\mathrm{T}} \boldsymbol{y} - d \geq 0 \right\}.$$

QCQP Formulation

Natual form



$$
\begin{aligned}
\min \quad & 2\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^1 \\
\text{s.t.} \quad & \boldsymbol{x}^1 \geq \boldsymbol{0}, \quad \boldsymbol{x}^1 \geq W^0\boldsymbol{x}^0 + \boldsymbol{b}^0, \quad \boldsymbol{x}^1 \odot \left(W^0\boldsymbol{x}^0 + \boldsymbol{b}^0 - \boldsymbol{x}^1\right) \leq \boldsymbol{0}, \\
& \text{(+ some valid cuts)}, \\
& \boldsymbol{x}^0 \in \mathcal{X} = \left\{\boldsymbol{x} \,\middle|\, \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \leq \rho\right\}, \quad \boldsymbol{x}^1 \in \mathbb{R}^{n_1}.
\end{aligned}
$$

$\odot$ = Hadamard product

Determine a constant term d for fixed $\boldsymbol{c}$ in

$$\left\{\boldsymbol{x}^1 \in \mathbb{R}^{n_2} \,\middle|\, \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^1 - d \geq 0\right\}.$$

minimize in primal side

maximize in dual side

Primal SDP Relaxation

Natual form

$$
\begin{aligned}
\min \quad & 2\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^1 \\
\text{s.t.} \quad & I \bullet X^{00} - 2\hat{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x}^0 \le \rho^2 - \hat{\boldsymbol{x}}^{\mathrm{T}}\hat{\boldsymbol{x}}, \\
& \left(\boldsymbol{e}^i\right)^{\mathrm{T}}\boldsymbol{x}^1 \ge 0, \quad \left(\boldsymbol{e}^i\right)^{\mathrm{T}}\boldsymbol{x}^1 \ge \left(\boldsymbol{e}^i\right)^{\mathrm{T}}\left(W^0\boldsymbol{x}^0 + \boldsymbol{b}^0\right), \quad i = 1, \ldots, n_1, \\
& X^{11}_{ii} - b^0_i x^1_i - \textstyle\sum_j W^0_{ij} X^{10}_{ij} \le 0, \qquad\qquad i = 1, \ldots, n_1, \\
& \begin{bmatrix} 1 & \left(\boldsymbol{x}^0\right)^{\mathrm{T}} & \left(\boldsymbol{x}^1\right)^{\mathrm{T}} \\ \boldsymbol{x}^0 & X^{00} & \left(X^{10}\right)^{\mathrm{T}} \\ \boldsymbol{x}^1 & X^{10} & X^{11} \end{bmatrix} \succeq O
\end{aligned}
$$

$\boldsymbol{e}^i$: the ith unit vector

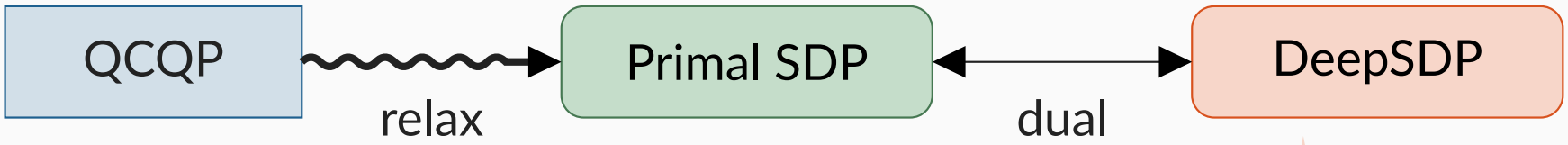
Objective of this work

Verify the exactness of (primal) SDP relaxation

- checking "*collinearity*" of another form

Dual SDP relaxation

Natual form



$$
\begin{aligned}
\max_{\gamma,\boldsymbol{\lambda},\boldsymbol{\nu},\boldsymbol{\eta},d} \quad & 2d \\
\text{s.t.} \quad & \gamma \begin{bmatrix} \hat{\boldsymbol{x}}^{\mathrm{T}}\hat{\boldsymbol{x}} - \rho^2 & -\hat{\boldsymbol{x}}^{\mathrm{T}} & \mathbf{0} \\ -\hat{\boldsymbol{x}} & I & O \\ \mathbf{0} & O & O \end{bmatrix} + \begin{bmatrix} -2d & \mathbf{0}^{\mathrm{T}} & \boldsymbol{c}^{\mathrm{T}} \\ \mathbf{0} & O & O \\ \boldsymbol{c} & O & O \end{bmatrix} \\
& \quad + \begin{bmatrix} 0 & \boldsymbol{\nu}^{\mathrm{T}}W^0 & -\boldsymbol{\nu}^{\mathrm{T}} - \boldsymbol{\eta}^{\mathrm{T}} \\ (W^0)^{\mathrm{T}}\boldsymbol{\nu} & O & -(W^0)^{\mathrm{T}}\operatorname{diag}(\boldsymbol{\lambda}) \\ -\boldsymbol{\nu} - \boldsymbol{\eta} & -\operatorname{diag}(\boldsymbol{\lambda})W^0 & 2\operatorname{diag}(\boldsymbol{\lambda}) \end{bmatrix} \succeq O, \\
& \gamma \in \mathbb{R}_+, \quad \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\eta} \in \mathbb{R}^n_+, \quad d \in \mathbb{R}.
\end{aligned}
$$

Our Theoretical Results

For previous QCQP,

Theorem

The primal SDP relaxation is exact if

- $\mathcal{X} = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \leq \rho\}$; or
- $\mathcal{X} = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_\infty \leq \rho\}$, $W^0 = I$, and $\hat{\boldsymbol{x}} \geq -\boldsymbol{b}^0$.

In addition, DeepSDP is also exact under strong duality.

- No extra assumption in hyper-ellipsoid case.
- Proof depending on
  Gram matrix representation with $\boldsymbol{e} = \boldsymbol{e}^1$.

Gram Matrix Representation of Safety Verification

- Assign $\boldsymbol{u}^j \in \mathbb{R}^s$ for $\boldsymbol{x}^0$, and $\boldsymbol{v}^i \in \mathbb{R}^s$ for $\boldsymbol{x}^1$:

$$\boldsymbol{x}^0 = \left[(\boldsymbol{e}^1)^{\mathrm{T}}\boldsymbol{u}^1, \ldots, (\boldsymbol{e}^1)^{\mathrm{T}}\boldsymbol{u}^{n_0}\right]^{\mathrm{T}}, \ \boldsymbol{x}^1 = \left[(\boldsymbol{e}^1)^{\mathrm{T}}\boldsymbol{v}^1, \ldots, (\boldsymbol{e}^1)^{\mathrm{T}}\boldsymbol{v}^{n_1}\right]^{\mathrm{T}}$$

- Similarly,

$$[X^{00}]_{ij} = (\boldsymbol{u}^i)^{\mathrm{T}}\boldsymbol{u}^j, \quad [X^{10}]_{ij} = (\boldsymbol{u}^i)^{\mathrm{T}}\boldsymbol{v}^j, \quad [X^{11}]_{ij} = (\boldsymbol{v}^i)^{\mathrm{T}}\boldsymbol{v}^j$$

Primal SDP is equivalently

$$\begin{aligned}
\min_{\boldsymbol{u}^j, \boldsymbol{v}^i} \ & 2\textstyle\sum_{i=1}^{n_1} c_i \, (\boldsymbol{e}^1)^{\mathrm{T}}\boldsymbol{v}^i \\
\text{s.t.} \ & (\boldsymbol{e}^1)^{\mathrm{T}}\boldsymbol{v}^i \geq 0, && i = 1, \ldots, n_1, \\
& (\boldsymbol{e}^1)^{\mathrm{T}}\boldsymbol{v}^i \geq (\boldsymbol{e}^1)^{\mathrm{T}} \left(\textstyle\sum_{j=1}^{n_0} W_{ij}\boldsymbol{u}^j + b_i^0 \boldsymbol{e}^1\right), && i = 1, \ldots, n_1, \\
& \left\|\boldsymbol{v}^i\right\|_2^2 \leq \left(\textstyle\sum_{j=1}^{n_0} W_{ij}\boldsymbol{u}^j + b_i^0 \boldsymbol{e}^1\right)^{\mathrm{T}} \boldsymbol{v}^i, && i = 1, \ldots, n_1, \\
& \sum_{j=1}^{n_0} \|\boldsymbol{u}^j - \hat{x}_j \boldsymbol{e}^1\|_2^2 \leq \rho^2.
\end{aligned}$$

Decomposition according to u^j and v^i

$$\min_{\boldsymbol{u}^j, \boldsymbol{v}^i} \; 2\sum_{i=1}^{n_1} c_i \left(\boldsymbol{e}^1\right)^{\mathrm{T}} \boldsymbol{v}^i \tag{1}$$

$$\text{s.t.} \; \left(\boldsymbol{e}^1\right)^{\mathrm{T}} \boldsymbol{v}^i \geq 0, \qquad i = 1, \ldots, n_1, \tag{2}$$

$$\left(\boldsymbol{e}^1\right)^{\mathrm{T}} \boldsymbol{v}^i \geq \left(\boldsymbol{e}^1\right)^{\mathrm{T}} \left(\sum_{j=1}^{n_0} W_{ij} \boldsymbol{u}^j + b_i^0 \boldsymbol{e}^1\right), \quad i = 1, \ldots, n_1, \tag{3}$$

$$\left\|\boldsymbol{v}^i\right\|_2^2 \leq \left(\sum_{j=1}^{n_0} W_{ij} \boldsymbol{u}^j + b_i^0 \boldsymbol{e}^1\right)^{\mathrm{T}} \boldsymbol{v}^i, \qquad i = 1, \ldots, n_1, \tag{4}$$

$$\sum_{j=1}^{n_0} \left\|\boldsymbol{u}^j - \hat{x}_j \boldsymbol{e}^1\right\|_2^2 \leq \rho^2. \tag{5}$$

Inner problem

$$\Psi(\boldsymbol{v}^1, \ldots, \boldsymbol{v}^{n_1}) :=$$

$$\min_{\boldsymbol{u}^1, \ldots, \boldsymbol{u}^{n_0}} \; \sum_{j=1}^{n_0} \left\|\boldsymbol{u}^j - \hat{x}_j \boldsymbol{e}\right\|_2^2$$

$$\text{s.t.} \quad (3),\ (4).$$

Outer problem

$$\min_{\boldsymbol{v}^1, \ldots, \boldsymbol{v}^{n_1}} \quad (1)$$

$$\text{s.t.} \quad (2)$$

$$\Psi(\boldsymbol{v}^1, \ldots, \boldsymbol{v}^{n_1}) \leq \rho^2.$$

Relationship Between Their Solutions

A part of KKT condition of Inner problem :

$$\begin{bmatrix} \boldsymbol{u}^1 \\ \vdots \\ \boldsymbol{u}^{n_0} \end{bmatrix} = \begin{bmatrix} \hat{x}_1 \boldsymbol{e}^1 \\ \vdots \\ \hat{x}_{n_0} \boldsymbol{e}^1 \end{bmatrix} - \sum_{i=1}^{n_1} \frac{\nu_i}{2} \begin{bmatrix} W_{i1} \boldsymbol{e}^1 \\ \vdots \\ W_{in} \boldsymbol{e}^1 \end{bmatrix} + \sum_{i=1}^{n_1} \frac{\lambda_i}{2} \begin{bmatrix} W_{i1} \boldsymbol{v}^i \\ \vdots \\ W_{in} \boldsymbol{v}^i \end{bmatrix}$$

Lemma: Linear Combination

For any optimal solution $(\boldsymbol{u}^1)^*, \ldots, (\boldsymbol{u}^{n_0})^*$ of Inner problem , there exist $\boldsymbol{m} \in \mathbb{R}^{n_0}$ and $M \in \mathbb{R}^{n_1 \times n_0}$ such that

$$(\boldsymbol{u}^j)^* = m_j \boldsymbol{e}^1 + \sum_{i=1}^{n_1} M_{ij} \boldsymbol{v}^i \quad \text{for each } j \in \{1, \ldots, n_0\}.$$

Collinearity in Outer-problem

Using $\boldsymbol{m}$ and M changes Outer problem into

$$\left.\begin{aligned} \min_{\boldsymbol{v}^1,\ldots,\boldsymbol{v}^{n_1}} \quad & 2\sum_{i=1}^{n_1} c_i\, \boldsymbol{e}^{\mathrm{T}}\boldsymbol{v}^i && (1)\\ \text{s.t.} \quad & \boldsymbol{e}^{\mathrm{T}}\boldsymbol{v}^i \geq 0,\ i=1,\ldots,n_1, && (2)\\ & \sum_{j=1}^{n_0}\left\|(m_j-\hat{x}_j)\,\boldsymbol{e}^1+\sum_{i=1}^{n_1} M_{ij}\boldsymbol{v}^i\right\|_2^2 \leq \rho^2. \end{aligned}\right\}$$

Lemma: Collinearity of $(\boldsymbol{v}^i)^*$

Outer problem has an optimal solution $(\boldsymbol{v}^1)^*,\ldots,(\boldsymbol{v}^{n_1})^*$ collinear to $\boldsymbol{e}^1$.

Therefore, the SDP relaxation is exact due to the collinearity.

Collinearity in Outer-problem

Using $\boldsymbol{m}$ and M changes Outer problem into

$$\left.\begin{aligned} \min_{\boldsymbol{v}^1,\ldots,\boldsymbol{v}^{n_1}} \quad & 2\sum_{i=1}^{n_1} c_i\, \boldsymbol{e}^{\mathrm{T}}\boldsymbol{v}^i && (1)\\ \text{s.t.} \quad & \boldsymbol{e}^{\mathrm{T}}\boldsymbol{v}^i \geq 0,\ i=1,\ldots,n_1, && (2)\\ & \sum_{j=1}^{n_0}\left\|(m_j-\hat{x}_j)\,\boldsymbol{e}^1+\sum_{i=1}^{n_1} M_{ij}\boldsymbol{v}^i\right\|_2^2 \leq \rho^2. \end{aligned}\right\}$$

Essence of Proof.

- Let $(\bar{\boldsymbol{v}}^1,\ldots,\bar{\boldsymbol{v}}^{n_1})$ be an optimal solution.
- Assume at least one of $\bar{\boldsymbol{v}}^1,\ldots,\bar{\boldsymbol{v}}^{n_1}$ is not collinear to $\boldsymbol{e}^1$.
- Define
$$\hat{\boldsymbol{v}}^i := \left[\bar{v}_1^i,\ 0,\ \cdots,\ 0\right]^{\mathrm{T}}.$$
- Then, $(\hat{\boldsymbol{v}}^1,\ldots,\hat{\boldsymbol{v}}^{n_1})$ is another optimal solution.

Outline

- Introduction
- Quadratically constrained quadratic programming (QCQP)
- Exact semidefinite programming (SDP) relaxation
- Single-layer feed-forward neural network
- QCQP formulation for safety verification
- Exactness conditions for SDP relaxation
- Graphical explanation of exactness
- Summary

Meaning of Exact Relaxation

Current situation

- Instead of checking whether $\mathcal{Y} \subseteq S_y$,

$$\mathcal{Y} \subseteq \quad H \coloneqq \bigcap_k \left\{ \boldsymbol{y} \in \mathbb{R}^{n_2} \,\middle|\, \left(\boldsymbol{c}^k\right)^{\mathrm{T}} \boldsymbol{y} - \boxed{d^k} \geq 0 \right\} \quad \subseteq S_y.$$

Optimal Solutions

Where does exactness appear?

- Half-space H supports $\mathcal{Y}$ at a face.
- Margin between H and S_y increases.
  - It is the robustness of the input set $\mathcal{X}$.
  - $\iff$ $\mathcal{X}$ can be made larger.

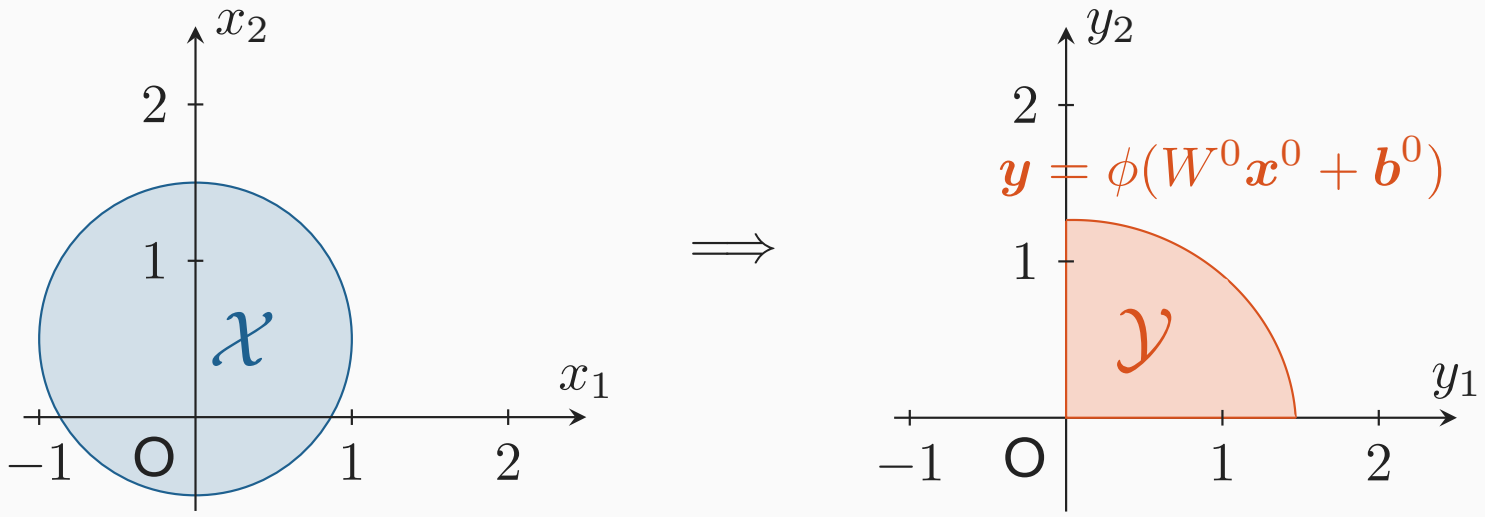
Instance with Exact Relaxation

Example: a single-layer NN

$$W^0 = \begin{bmatrix} 1/2 & 5/4 \\ -6/5 & 2/5 \end{bmatrix}, \quad W^1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{b}^0 = \begin{bmatrix} -0.5 \\ -0.2 \end{bmatrix}, \quad \boldsymbol{b}^1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

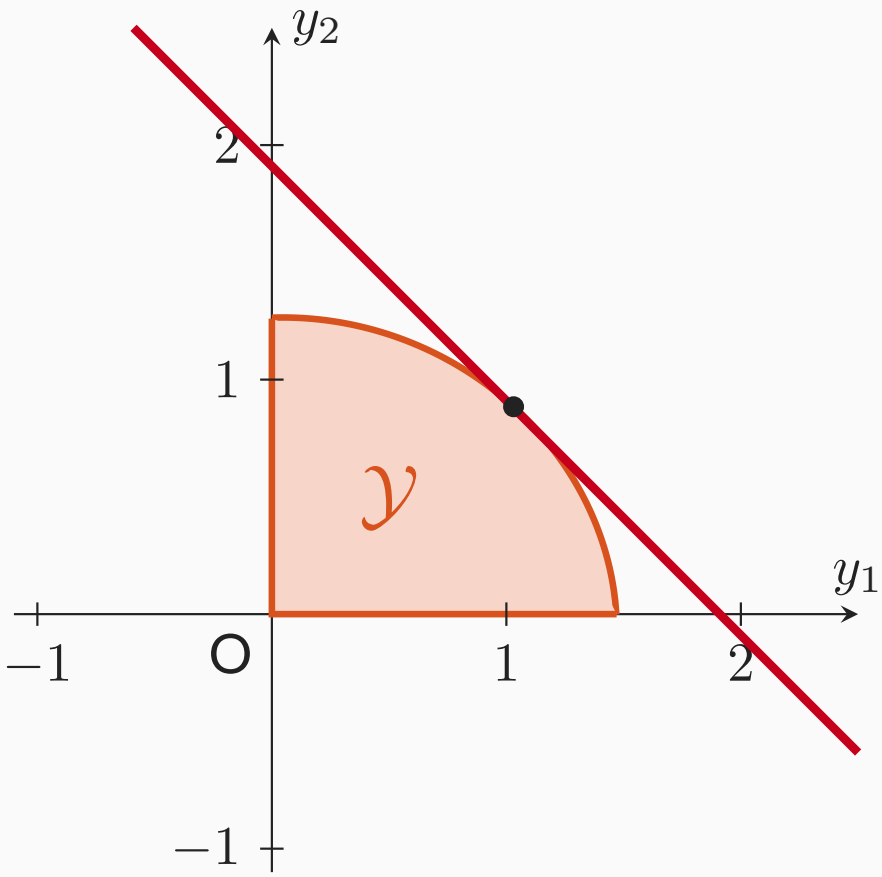
with the input set

$$\mathcal{X} := \left\{ \boldsymbol{x} \in \mathbb{R}^2 \,\middle|\, \left\| \boldsymbol{x} - \begin{bmatrix} 0 \\ 0.5 \end{bmatrix} \right\|_2 \leq 1 \right\}.$$



Experiment Result in Mosek and Julia

Slope c^*		Solution
c_1	c_2	d^*
1	0	-8.42×10^{-10}
0	1	-3.45×10^{-10}
-1	1	-1.47
1	-1	-1.26
$-1/4$	-1	-1.32
$-1/2$	-1	-1.46
-1	-1	-1.91
-2	-1	-3.16
-4	-1	-5.96

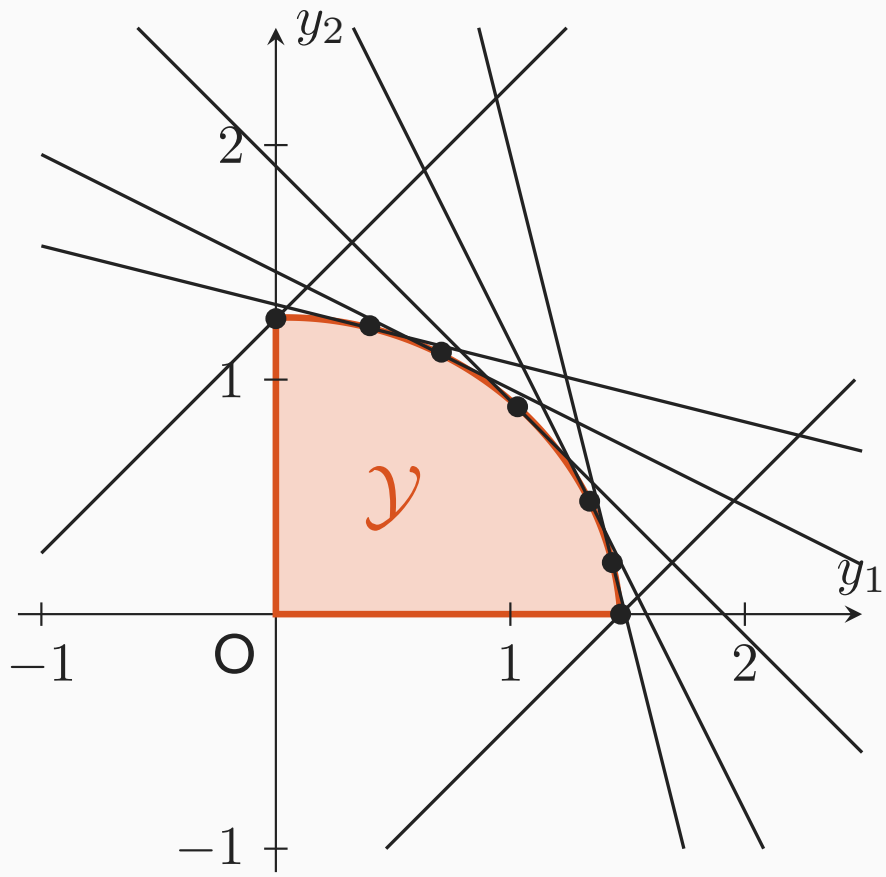


Observation

All boundaries intersect $\mathcal{Y}$ at a point

Experiment Result in Mosek and Julia

Slope c^*		Solution
c_1	c_2	d^*
1	0	-8.42×10^{-10}
0	1	-3.45×10^{-10}
-1	1	-1.47
1	-1	-1.26
$-1/4$	-1	-1.32
$-1/2$	-1	-1.46
-1	-1	-1.91
-2	-1	-3.16
-4	-1	-5.96



Observation

All boundaries intersect $\mathcal{Y}$ at a point

Conclusion

Summary

- Safety verification of single-layer neural networks
- Exactness conditions of SDP relaxation
- Graphical insight for exactness

Future works

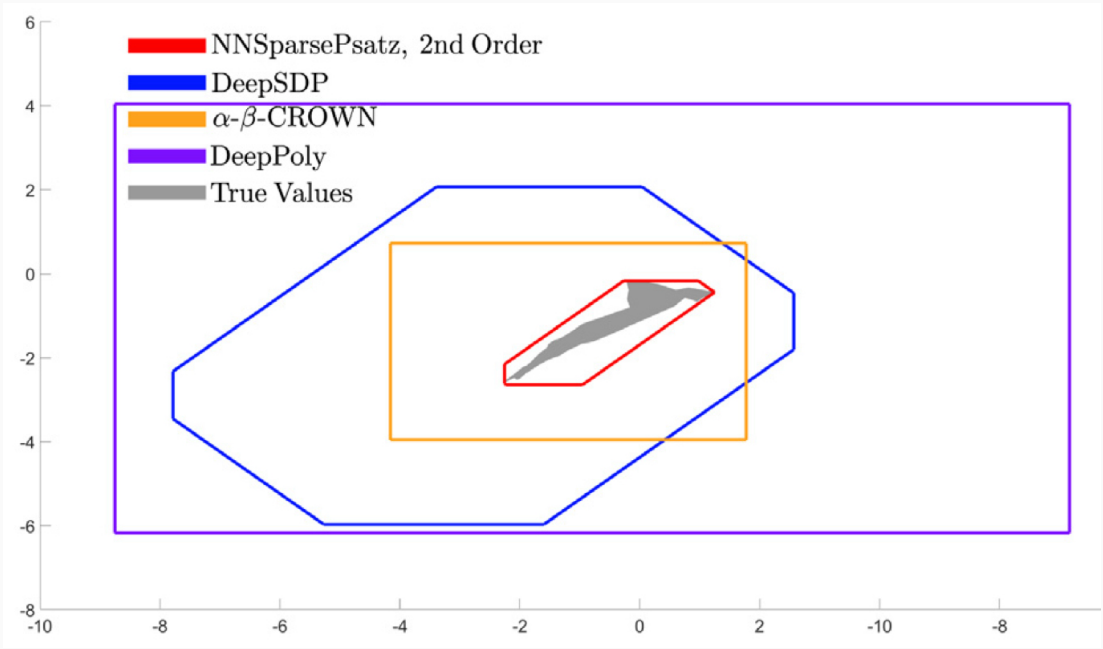
- Analyze the non-polyhedral case of S_y
- Extend results to other networks

Thank you for your attention!

For more details, see arXiv:2504.09934

Exact Semidefinite Relaxations for Verifying Robustness of Neural Networks

- Less accurate than other methods in general [NP21].
- Accurate when the relaxation is exact.



[NP21] Newton and Papachristodoulou, Neural network verification using polynomial optimisation, IEEE CDC, 2021.

Constraints for Hidden Layers

$$\begin{bmatrix} 1 \\ \boldsymbol{w}^0 \\ \phi(\boldsymbol{w}^0) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 & \mathbf{0}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & O & -\boldsymbol{e}^i \left(\boldsymbol{e}^i\right)^{\mathrm{T}} \\ \mathbf{0} & -\boldsymbol{e}^i \left(\boldsymbol{e}^i\right)^{\mathrm{T}} & 2\boldsymbol{e}^i \left(\boldsymbol{e}^i\right)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{w}^0 \\ \phi(\boldsymbol{w}^0) \end{bmatrix} = 0, \quad i = 1, \ldots, N, \tag{6a}$$

$$\begin{bmatrix} 1 \\ \boldsymbol{w}^0 \\ \phi(\boldsymbol{w}^0) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 & \left(\boldsymbol{e}^i\right)^{\mathrm{T}} & -\left(\boldsymbol{e}^i\right)^{\mathrm{T}} \\ \boldsymbol{e}^i & O & O \\ -\boldsymbol{e}^i & O & O \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{w}^0 \\ \phi(\boldsymbol{w}^0) \end{bmatrix} \leq 0, \quad i = 1, \ldots, N, \tag{6b}$$

$$\begin{bmatrix} 1 \\ \boldsymbol{w}^0 \\ \phi(\boldsymbol{w}^0) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 & \mathbf{0}^{\mathrm{T}} & -\left(\boldsymbol{e}^i\right)^{\mathrm{T}} \\ \mathbf{0} & O & O \\ -\boldsymbol{e}^i & O & O \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{w}^0 \\ \phi(\boldsymbol{w}^0) \end{bmatrix} \leq 0, \quad i = 1, \ldots, N. \tag{6c}$$

Valid Cuts

Let $\boldsymbol{w}^0 = W^0\boldsymbol{x}^0 + \boldsymbol{b}^0$.

Valid Cuts for ReLU

The following inequation always holds

$$\left[\phi(w_j^0) - \phi(w_i^0)\right]\left[\phi(w_j^0) - \phi(w_i^0) - (w_j - w_i)\right] \leq 0 \quad \forall (i,j) \in \{1,\ldots,n\}^2$$

$$\begin{bmatrix} \boldsymbol{w}^0 \\ \phi(\boldsymbol{w}^0) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} O & -(\boldsymbol{e}^i - \boldsymbol{e}_j)(\boldsymbol{e}^i - \boldsymbol{e}_j)^{\mathrm{T}} \\ -(\boldsymbol{e}^i - \boldsymbol{e}_j)(\boldsymbol{e}^i - \boldsymbol{e}_j)^{\mathrm{T}} & 2(\boldsymbol{e}^i - \boldsymbol{e}_j)(\boldsymbol{e}^i - \boldsymbol{e}_j)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}^0 \\ \phi(\boldsymbol{w}^0) \end{bmatrix} \leq 0, \quad (7)$$

Input Constraint

Since $\boldsymbol{x}^0 \in \mathcal{X} = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \leq \rho\}$.

In QCQP

$$\left\|\boldsymbol{x}^0 - \hat{\boldsymbol{x}}\right\|_2^2 \leq \rho^2$$

In SDP relaxation

$$\begin{bmatrix} \hat{\boldsymbol{x}}^{\mathrm{T}}\hat{\boldsymbol{x}} - \rho^2 & -\hat{\boldsymbol{x}}^{\mathrm{T}} & \mathbf{0} \\ -\hat{\boldsymbol{x}} & I & O \\ \mathbf{0} & O & O \end{bmatrix} \bullet \underbrace{\begin{bmatrix} 1 & \left(\boldsymbol{x}^0\right)^{\mathrm{T}} & \left(\boldsymbol{x}^1\right)^{\mathrm{T}} \\ \boldsymbol{x}^0 & X_{00} & X_{10}{}^{\mathrm{T}} \\ \boldsymbol{x}^1 & X_{10} & X_{11} \end{bmatrix}}_{=: G} \leq 0$$

In DeepSDP By introducing a dual variable γ,

$$\gamma \begin{bmatrix} \hat{\boldsymbol{x}}^{\mathrm{T}}\hat{\boldsymbol{x}} - \rho^2 & -\hat{\boldsymbol{x}}^{\mathrm{T}} & \mathbf{0} \\ -\hat{\boldsymbol{x}} & I & O \\ \mathbf{0} & O & O \end{bmatrix}$$

Safety Specification Set

Consider a half-space $H := \left\{ \boldsymbol{y} \in \mathbb{R}^{n_2} \,\middle|\, \boldsymbol{c}^{\mathrm{T}} \boldsymbol{y} - d \geq 0 \right\}$.

- The slope $\boldsymbol{c}$ according to each half-space is given.
- The largest d makes H smaller.

In SDP relaxation

$$\boldsymbol{x}^1 \in H \iff \begin{bmatrix} -2d & \mathbf{0}^{\mathrm{T}} & \boldsymbol{c}^{\mathrm{T}} \\ \mathbf{0} & O & O \\ \boldsymbol{c} & O & O \end{bmatrix} \bullet \begin{bmatrix} 1 & \left(\boldsymbol{x}^0\right)^{\mathrm{T}} & \left(\boldsymbol{x}^1\right)^{\mathrm{T}} \\ \boldsymbol{x}^0 & X_{00} & {X_{10}}^{\mathrm{T}} \\ \boldsymbol{x}^1 & X_{10} & X_{11} \end{bmatrix} \leq 0$$

In DeepSDP Let d behave as a dual variable.

Quadratic Formulation for ReLU Function

Review: ϕ applies element-wisely ReLU function φ

Let $\boldsymbol{w}^0 := W^0\boldsymbol{x}^0 + \boldsymbol{b}^0$. For any $i \in \{1, \ldots, n_1\}$,

$$\varphi(w_i^0) = \max\{0, w_i^0\} \quad \Longleftrightarrow \quad \begin{cases} \varphi(w_i^0)\left(\varphi(w_i^0) - w_i^0\right) \leq 0, \\ \varphi(w_i^0) \geq w_i^0, \quad \varphi(w_i^0) \geq 0. \end{cases}$$

In QCQP The first inequality is

$$\begin{bmatrix} 1 \\ \boldsymbol{w}^0 \\ \phi(\boldsymbol{w}^0) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 & \mathbf{0}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & O & -\boldsymbol{e}^i\left(\boldsymbol{e}^i\right)^{\mathrm{T}} \\ \mathbf{0} & -\boldsymbol{e}^i\left(\boldsymbol{e}^i\right)^{\mathrm{T}} & 2\boldsymbol{e}^i\left(\boldsymbol{e}^i\right)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{w}^0 \\ \phi(\boldsymbol{w}^0) \end{bmatrix} \leq 0, \quad i = 1, \ldots, n_1.$$

Transformation of $[1, w^0, \phi(w^0)]^{\mathrm{T}}$

Equivalently, for $i \in \{1, \ldots, n_1\}$,

$$\begin{bmatrix} 1 \\ x^0 \\ x^1 \end{bmatrix}^{\mathrm{T}} \underbrace{\begin{bmatrix} 1 & \mathbf{0}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} \\ b^0 & W^0 & O \\ \mathbf{0} & O & I \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 0 & \mathbf{0}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & O & -e^i (e^i)^{\mathrm{T}} \\ \mathbf{0} & -e^i (e^i)^{\mathrm{T}} & 2e^i (e^i)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} \\ b^0 & W^0 & O \\ \mathbf{0} & O & I \end{bmatrix}}_{=: L_i} \begin{bmatrix} 1 \\ x^0 \\ x^1 \end{bmatrix} \leq$$

In SDP relaxation $\quad L_i \bullet G \leq 0, \quad i \in \{1, \ldots, n_1\}$

In DeepSDP Introducing a dual variable $\boldsymbol{\lambda} \in \mathbb{R}^{n_1}_+$, $\sum_{i=1}^{n_1} \lambda_i L_i$

Constraint	$\varphi(w_i^0)\left(\varphi(w_i^0) - w_i^0\right) \leq 0$	$\varphi(w_i^0) \geq w_i^0$	$\varphi(w_i^0) \geq 0$
Dual variable	λ_i	ν_i	η_i