Supplementary Materials for
Fastened CROWN: Tightened Neural Network Robustness Certificates

Contents

In the appendix, we adopt the same notations as in the main text. When new notations arise, we will explain their meanings at the places where they first appear. In addition, to simply notation, we omit the brackets around the layer index in the superscripts when no ambiguity is caused in the remainder of the appendix. For example, we will use W^k instead of $W^{(k)}$ to denote the weight of the k-th layer in the neural network, s^{kU}, t^{kU} instead of $s^{(k)U}, t^{(k)U}$ to denote the parameters of the upper bounding lines in the k-th layer. However, brackets in variables like $s^{(k-1)U}, t^{(k-1)U}$ are remained to prevent ambiguity.

A.1 Comparison of related works

Table 1: Comparison of methods for providing certifiable adversarial robustness quantification in NNs.

Method	Multi-layer	Beyond ReLU/ MLP	Large Networks	Non-trivial Bound	Implementation
(Szegedy et al. 2014)	✓	✓/×	✓	×	Not Available
(Hein and Andriushchenko 2017)	×	differentiable/×	×	✓	TensorFlow
(Wong and Kolter 2018)	✓	✓/✓	✓	✓	Pytorch (GPU)
Fast-lin / Fast-lip (Weng et al. 2018b)	✓	×/ ×	✓	✓	NumPy
CROWN (Zhang et al. 2018)	✓	✓/ ×	✓	✓	NumPy
LP-ALL (Salman et al. 2019)	✓	✓/ ✓	×	✓	Gurobi
FROWN (This work)	✓	✓/ ✓	✓	✓	PyTorch (GPU)

"Large networks" refer to the computational applicability in handling deep networks

A.2 More about CROWN

Given the bounds of the previous $k-1, (k \geq 2)$ layers, $l^{[k-1]}$ and $u^{[k-1]}$, one could compute bounds of the k-th layer through the following steps. Repeat the following iteration:

$$\begin{aligned}(\tilde{w}_i^v)^\top &= [relu(\tilde{w}_i^{v+1}) \odot s^{vL} + neg(\tilde{w}_i^{v+1}) \odot s^{vU}]^\top W^v, \\ \tilde{b}_i^v &= relu(\tilde{w}_i{}^{v+1})^\top (s^{vL} \odot b^v + t^{vL}) + neg(\tilde{w}_i^{v+1})^\top (s^{vU} \odot b^v + t^{vU}) + \tilde{b}_i^{v+1}, \\ v &= k-1, k-2, \cdots, 1\end{aligned} \tag{1}$$

to obtain $\tilde{w}_i^1$ and $\tilde{b}_i^1$, where the starting variable is defined as

$$\begin{aligned}\tilde{w}_i^k &= W_{i,:}^k{}^\top, \\ \tilde{b}_i^k &= b_i^k.\end{aligned}$$

The parameters of the bounding lines $s^{[k-1]L}, s^{[k-1]U}, t^{[k-1]L}, t^{[k-1]U}$ are chosen to satisfy the following constraints:

$$\begin{aligned}s_i^{vL} z_i^v + t_i^{vL} \leq \sigma(z_i^v) \leq s_i^{vU} z_i^v + t_i^{vU}, \\ \forall\, z_i^v \in [l_i^v, u_i^v], i \in [n_v], v \in [k-1].\end{aligned}$$

See details of how the CROWN choose bounding lines in Section A.4. Then the lower bound $\gamma_i^{kL}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L})$ of z_i^k can be computed by the following formula:

$$\begin{aligned}z_i^k &= W_{i,:}^k a^{k-1} + b_i^k \geq (\tilde{w}_i^1)^\top x + \tilde{b}_i^1 \\ &\geq (\tilde{w}^1)^\top x_0 - \epsilon ||\tilde{w}^1||_q + \tilde{b}_i^1 \\ &= \gamma_i^{kL}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L}), \\ &\forall x \in \mathbb{B}_p(x_0, \epsilon).\end{aligned} \tag{2}$$

To compute the upper bound $\gamma_i^{kU}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L})$ of z_i^k, replace the iteration in Equation (1) with the following iteration:

$$\begin{aligned}(\tilde{w}_i^v)^\top &= [neg(\tilde{w}_i^{v+1}) \odot s^{vL} + relu(\tilde{w}_i^{v+1}) \odot s^{vU}]^\top W^v, \\ \tilde{b}_i^v &= neg(\tilde{w}_i{}^{v+1})^\top (s^{vL} \odot b^v + t^{vL}) + relu(\tilde{w}_i^{v+1})^\top (s^{vU} \odot b^v + t^{vU}) + \tilde{b}_i^{v+1},\end{aligned}$$

and then

$$\gamma_i^{kU}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L}) = (\tilde{w}^1)^\top x_0 + \epsilon ||\tilde{w}^1||_q + \tilde{b}_i^1$$

The bounds of the first layer can be computed directly througth the following formula:

$$\begin{aligned}l_i^1 &= \min_{a^0 \in B_p(x_0,\epsilon)} W_{i:}^1 a^0 + b_i^1 = W_{i:}^1 x_0 + b_i^1 - \epsilon ||W_{i:}^1||_q, \\ u_i^1 &= \max_{a^0 \in B_p(x_0,\epsilon)} W_{i:}^1 a^0 + b_i^1 = W_{i:}^1 x_0 + b_i^1 + \epsilon ||W_{i:}^1||_q.\end{aligned} \tag{3}$$

Then by repeating the process of computing bounds for the k-th layer from $k=2$ to $k=m$, we can obtain the bounds of the final output z_i^m, $\gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})$ and $\gamma_i^{mU}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})$.

A.3 Details of the Lp-based method

Given the bounds of the previous $m-1$ layers, $l^{[m-1]}$ and $u^{[m-1]}$, one could compute lower bound of the final output by solving the following LP problem:

$$\min_{a^{(0)}\in\mathbb{B}_p(x_0,\epsilon),a^{[m-1]},z^{[m-1]}} W^m_{i,:}a^{m-1}+b^m_i \tag{4}$$
$$\text{s.t.}\begin{cases} z^k = W^k a^{k-1}+b^k \\ h^{kL}(z^k)\preccurlyeq a^k \preccurlyeq h^{kU}(z^k)\text{ , for } k\in[m-1], \\ l^k\preccurlyeq z^k\preccurlyeq u^k \end{cases}$$

where

$$\begin{aligned} h^{kL}_i(z^k_i) &= s^{kL}_i z^k_i + t^{kL}_i, \\ h^{kU}_i(z^k_i) &= s^{kU}_i z^k_i + t^{kU}_i. \end{aligned} \tag{5}$$

The parameters of the bounding lines $s^{[k-1]L}, s^{[k-1]U}, t^{[k-1]L}, t^{[k-1]U}$ are chosen to satisfy the following constraints:

$$\begin{aligned} &s^{vL}_i z^v_i + t^{vL}_i \le \sigma(z^v_i) \le s^{vU}_i z^v_i + t^{vU}_i, \\ &\quad \forall\, z^v_i \in [l^v_i, u^v_i], i\in[n_v], v\in[k-1]. \end{aligned}$$

See details of how to choose these bounding lines in Section A.4. Note that $h^{kL}_i(z^k_i)$ and $h^{kU}_i(z^k_i)$ can be taken as the pointwise supremum and infimum of several linear functions, respectively, which is equivalent to using multiple linear constraints to replace the second constraint in Problem (4). The upper bound of the final output can be computed by simply replace the "min" operation in Problem (4) with "max".

The lower and upper bounds of the preactivation of k-th layer can be obtained by replacing the index m in Problem (4) with the corresponding layer index k, which leads us to the following two optimization problems:

$$l^k_i = \min_{a^{(0)}\in\mathbb{B}_p(x_0,\epsilon),a^{[k-1]},z^{[k-1]}} W^k_{i,:}a^{k-1}+b^k_i \tag{6}$$
$$\text{s.t.}\begin{cases} z^v = W^v a^{v-1}+b^v \\ h^{vL}(z^v)\preccurlyeq a^v \preccurlyeq h^{vU}(z^v)\text{ , for } v\in[k-1], \\ l^v\preccurlyeq z^v\preccurlyeq u^v \end{cases}$$

$$u^k_i = \max_{a^{(0)}\in\mathbb{B}_p(x_0,\epsilon),a^{[k-1]},z^{[k-1]}} W^k_{i,:}a^{k-1}+b^k_i \tag{7}$$
$$\text{s.t.}\begin{cases} z^v = W^v a^{v-1}+b^v \\ h^{vL}(z^v)\preccurlyeq a^v \preccurlyeq h^{vU}(z^v)\text{ , for } v\in[k-1]. \\ l^v\preccurlyeq z^v\preccurlyeq u^v \end{cases}$$

Note that Problems (6) and (7) are only applicable when $k\ge 2$. When $k=1$, the lower and upper bounds can be computed directly using the following formula:

$$l^1_i = \min_{a^0\in B_p(x_0,\epsilon)} W^1_{i:}a^0+b^1_i = W^1_{i:}x_0+b^1_i-\epsilon||W^1_{i:}||_q \tag{8}$$
$$u^1_i = \max_{a^0\in B_p(x_0,\epsilon)} W^1_{i:}a^0+b^1_i = W^1_{i:}x_0+b^1_i+\epsilon||W^1_{i:}||_q \tag{9}$$

Then one could recursively solve the LP problems (6)and (7) from the 2nd layer to the m-th layer to obtain the pre-activation bounds for all layers.

Similar to the self-consistency condition defined for CROWN in the main text, we could define the self-consistency condition for the LP process.

Condition 1 ***Self-consistency condition of the LP process.*** *Suppose* $\{\tilde{s}^{[v-1]U}, \tilde{s}^{[v-1]L}, \tilde{t}^{[v-1]U}, \tilde{t}^{[v-1]L}\}$ *are used to calculate* l_i^v *and* u_i^v*,* $\{\hat{s}^{[k-1]U}, \hat{s}^{[k-1]L}, \hat{t}^{[k-1]U}, \hat{t}^{[k-1]L}\}$ *are used to calculate* l_j^k *and* u_j^k*, then the following condition holds,*

$$\tilde{s}^{[v-1]U} = \hat{s}^{[v-1]U}, \tilde{s}^{[v-1]L} = \hat{s}^{[v-1]L},$$
$$\tilde{t}^{[v-1]U} = \hat{t}^{[v-1]U}, \tilde{t}^{[v-1]L} = \hat{t}^{[v-1]L},$$

for $\forall\, i \in [n_v], \forall j \in [n_k], 2 \leq v \leq k \leq m$*.*

A.4 The bounding lines in CROWN and the LP-based method

Refer the following tables to see how the CROWN and the LP based method choose bounding lines for ReLU, Sigmoid, Tanh activation. respectively.

As stated in the main text, the LP-based method gives the same bounds as the CROWN when the LP-based method only use one bounding line on each side. Therefore, we allow the LP-based method to use up to three bounding lines on each side in order to make improvement. In theory, the more bounding lines are used, the tighter bounds we can potentially obtain. However, as every bounding line corresponds to a linear constraint in Problem (4), it will inevitably takes more time to solve Problem (4) when more bounding lines are used.

Table 2: Illustration of how LP and CROWN choose bounding lines for $a = \text{ReLU}(z)$ activation. $s_0 = [\text{ReLU}(u) - \text{ReLU}(l)]/(u - l)$ and $t_0 = \text{ReLU}(l) - s_0 l$.

LP	$l < u \leq 0$	$l < 0 < u$	$0 \leq l < u$
Lower	$a = 0$	$a = 0$ and $a = z$	$a = z$
Upper	$a = 0$	$a = s_0 z + t_0$	$a = z$
CROWN	$l < u \leq 0$	$l < 0 < u$	$0 \leq l < u$
Lower	$a = 0$	$a = 0$, if $\lvert l\rvert > \lvert u\rvert$; $a = z$, else	$a = z$
Upper	$a = 0$	$a = s_0 z + t_0$	$a = z$

Table 3: Illustration of how LP and CROWN choose bounding lines for $a = \text{Sigmoid}(z)$ or $a = \text{Tanh}(z)$ activation. We use σ to denote Sigmoid and Tanh. $s_0 = [\sigma(u) - \sigma(l)]/(u - l)$ and the function $t(s, y)$ is defined as $t(s, y) = \sigma(y) - sy$. Case 1 refers to $\sigma'(u)l + t(\sigma'(u), u) \geq \sigma(l)$; and case 2, otherwise. Case 3 refers to $\sigma'(l)u + t(\sigma'(l), l) \leq \sigma(u)$; and case 4, otherwise. See definitions of u_d, l_d in Table 2 in the main text.

Upper Bouding Line	$l < u \leq 0$	$l < 0 < u$		$0 \leq l < u$
		case1	case2	
LP	$a = s_0 z + t(s_0, l)$	tangents at l_d, u, $(l_d + u)/2$	$a = s_0 z + t(s_0, l)$	tangents at l, u, $(l + u)/2$
CROWN	$a = s_0 z + t(s_0, l)$	tangent at l_d,	$a = s_0 z + t(s_0, l)$	tangent at $(l + u)/2$
Lower Bounding Line	$l < u \leq 0$	$l < 0 < u$		$0 \leq l < u$
		case3	case4	
LP	tangents at l, u, $(l + u)/2$	tangents at l, u_d, $(l + u_d)/2$	$a = s_0 z + t(s_0, l)$	$a = s_0 z + t(s_0, l)$
CROWN	tangent at $(l + u)/2$	tangent at u_d	$a = s_0 z + t(s_0, l)$	$a = s_0 z + t(s_0, l)$

A.5 Proof of Theorem 1 in the main text - CROWN is optimal to the LP problem under certain conditions

We first prove Theorom 1 in the main text under a slightly different condition:

Condition 2

- *The LP process satisfies self-consistency condition stated in Condition* (1)*.*
- *CROWN shares the same bounding lines as in the LP process.*

We will prove that under Condition 2, the lower bound in Equation (2) given by CROWN is the optimal value of Problem (4).

Let's first prove that the lower bound in Equation (2) given by CROWN is the objective value of a feasible solution of the dual problem of Problem (4), which means that $\gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})$ is less or equal to the optimal value of Problem (4).

Derive the dual of Problem (4)**.** The lagrangian associate with Problem (4) is

$$
\begin{aligned}
& L(a^0, a^{[m-1]}, z^{[m-1]} \mid \nu^{[m-1]}, \mu^{[m-1]}, \tau^{[m-1]}, \lambda^{[m-1]}, \theta^{[m-1]}) \\
&= W_{i:}^m a^{m-1} + b_i^m + \sum_{k=1}^{m-1} {\nu^k}^\top (z^k - W^k a^{k-1} - b^k) + \sum_{k=1}^{m-1} {\mu^k}^\top (a^k - s^{kU} \odot z^k - t^{kU}) \\
&+ \sum_{k=1}^{m-1} {\tau^k}^\top (a^k - s^{kL} \odot z^k - t^{kL}) + \sum_{k=1}^{m-1} {\lambda^k}^\top (z^k - u^k) + \sum_{k=1}^{m-1} {\theta^k}^\top (z^k - l^k) \\
&= ({W_{i:}^m}^\top + \mu^{m-1} + \tau^{m-1})^\top a^{m-1} + {\nu^1}^\top W^1 a^0 + b_i^m + \sum_{k=1}^{m-2} (\mu^k + \tau^k - {W^{k+1}}^\top \nu^{k+1})^\top a^k \\
&+ \sum_{k=1}^{m-1} (\nu^k + \lambda^k + \theta^k - \mu^k \odot s^{kU} - \tau^k \odot s^{kL})^\top z^k \\
&- \sum_{k=1}^{m-1} ({u^k}^\top \lambda^k + {l^k}^\top \theta^k + {b^k}^\top \nu^k + {t^{kU}}^\top \mu^k + {t^{kL}}^\top \tau^k).
\end{aligned} \tag{10}
$$

Define the dual objective as

$$
\begin{aligned}
& \phi(\nu^{[k-1]}, \mu^{[k-1]}, \tau^{[k-1]}, \lambda^{[k-1]}, \theta^{[k-1]}) \\
&= \min_{a^0 \in B_p(x_0, \epsilon), a^{[m-1]}, z^{[m-1]}} L(a^0, a^{[m-1]}, z^{[m-1]} \mid \nu^{[k-1]}, \mu^{[k-1]}, \tau^{[k-1]}, \lambda^{[k-1]}, \theta^{[k-1]}) \\
&= \begin{cases} \min_{a^0 \in B_p(x_0,\epsilon)} {\nu^1}^\top W^1 a^0 + b_i^m - \sum_{k=1}^{m-1} ({u^k}^\top \lambda^k + {l^k}^\top \theta^k + {b^k}^\top \nu^k + {t^{kU}}^\top \mu^k + {t^{kL}}^\top \tau^k), & \text{if Condition 3 is met} \\ -\infty, & \text{else} \end{cases} \\
&= \begin{cases} {\nu^1}^\top W^1 x_0 + b_i^m - \epsilon ||{\nu^1}^\top W^1||_q - \sum_{k=1}^{m-1} ({u^k}^\top \lambda^k + {l^k}^\top \theta^k + {b^k}^\top \nu^k + {t^{kU}}^\top \mu^k + {t^{kL}}^\top \tau^k), & \text{if Condition 3 is met} \\ -\infty, & \text{else} \end{cases}
\end{aligned}
$$

Condition 3 is the following

Condition 3

$$
\begin{cases} {W_{i:}^m}^\top + \mu^{m-1} + \tau^{m-1} = 0 \\ \mu^k + \tau^k - {W^{k+1}}^\top \nu^{k+1} = 0\,,\ k = 1, 2, .., m-2 \\ \nu^k + \lambda^k + \theta^k - \mu^k \odot s^{kU} - \tau^k \odot s^{kL} = 0\,,\ k = 1, 2, .., m-1 \end{cases}
$$

Then the dual of Problem (4) is

$$\max_{\substack{\nu^k\in\mathbb{R}^{n_k},\mu^k\geq 0,\tau^k\leq 0,\\ \lambda^k\geq 0,\theta^k\leq 0,\, k=1,2,\ldots,m-1}} \phi(\nu^{[m-1]},\mu^{[m-1]},\tau^{[m-1]},\lambda^{[m-1]},\theta^{[m-1]}),$$

which is equivalent to

$$\begin{aligned}\max_{\substack{\lambda^k\geq 0,\theta^k\leq 0,\nu^k\in\mathbb{R}^{n_k},\\ \mu^k\geq 0,\tau^k\leq 0,\, k=1,2,\ldots,m-1}} \quad & {\nu^1}^\top W^1 x_0 + b_i^m - \epsilon||{\nu^1}^\top W^1||_q - \sum_{k=1}^{m-1}({u^k}^\top\lambda^k + {l^k}^\top\theta^k + {b^k}^\top\nu^k + {t^{kU}}^\top\mu^k + {t^{kL}}^\top\tau^k)\\ \text{s.t. } & \mu^k+\tau^k = {W^{k+1}}^\top\nu^{k+1},\ k=1,2,...,m-2\\ & \mu^{m-1}+\tau^{m-1} = -{W_{i,:}^m}^\top\\ & \nu^k+\lambda^k+\theta^k = s^{kU}\odot\mu^k + s^{kL}\odot\tau^k,\ k=1,2,...,m-1\end{aligned} \tag{11}$$

Prove CROWN is a feasible solution to the dual problem (11)**.** We can verify that the following solution is a feasible solution to Problem (11).

$$\begin{aligned}\mu^k &= relu({W^{k+1}}^\top\nu^{k+1}),\ k=1,2,...,m-2\\ \tau^k &= neg({W^{k+1}}^\top\nu^{k+1}),\ k=1,2,...,m-2\\ \mu^{m-1} &= relu(-{W_{i,:}^m}^\top) = -neg({W_{i,:}^m}^\top)\\ \tau^{m-1} &= neg(-{W_{i,:}^m}^\top) = -relu({W_{i,:}^m}^\top)\\ \nu^k &= s^{kU}\odot\mu^k + s^{kL}\odot\tau^k,\ k=1,2,...,m-1\\ \theta^k &= 0,\ k=1,2,...,m-1\\ \lambda^k &= 0,\ k=1,2,...,m-1\end{aligned} \tag{12}$$

Observe the iteration for ν, we have

$$\begin{aligned}{\nu^k}^\top &= (s^{kU}\odot\mu^k + s^{kL}\odot\tau^k)\\ &= [s^{kU}\odot relu({W^{k+1}}^\top\nu^{k+1}) + s^{kL}\odot neg({W^{k+1}}^\top\nu^{k+1})]\\ &= [s^{kU}\odot relu(-\tilde{w}^{k+1}) + s^{kL}\odot neg(-\tilde{w}^{k+1})]\\ &= -[s^{kU}\odot neg(\tilde{w}^{k+1}) + s^{kL}\odot relu(\tilde{w}^{k+1})]\\ k &= m-2, m-3, ..., 1\end{aligned}$$

where

$$\nu^{m-1} = [s^{(m-1)U}\odot neg({W_{i,:}^m}^\top) + s^{(m-1)L}\odot relu({W_{i,:}^m}^\top)]$$

and $\tilde{w}^k$ is defined as

$$(\tilde{w}^k)^\top = -{\nu^k}^\top W^k,\ k=1,2,...,m-1.$$

Then we can get the iteration of $\tilde{w}$

$$\begin{aligned}(\tilde{w}^k)^\top &= -{\nu^k}^\top W^k\\ &= -(s^{kU}\odot\mu^k + s^{kL}\odot\tau^k)W^k\\ &= -[s^{kU}\odot relu({W^{k+1}}^\top\nu^{k+1}) + s^{kL}\odot neg({W^{k+1}}^\top\nu^{k+1})]W^k\\ &= -[s^{kU}\odot relu(-\tilde{w}^{k+1}) + s^{kL}\odot neg(-\tilde{w}^{k+1})]W^k\\ &= [s^{kU}\odot neg(\tilde{w}^{k+1}) + s^{kL}\odot relu(\tilde{w}^{k+1})]W^k\\ k &= m-2, m-3, ..., 1\end{aligned}$$

where

$$\tilde{w}^{m-1} = [s^{(m-1)U} \odot neg(W^{m\top}_{i,:}) + s^{(m-1)L} \odot relu(W^{m\top}_{i,:})]W^{m-1\top}.$$

Compare the above iteration of $\tilde{w}$ with the iteration of $\tilde{w}_i$ in Equation (1) in CROWN. We can find that they are exactly the same iteration, which means at the end of both iterations, $\tilde{w}^1$ here equals to $\tilde{w}^1_i$ in the CROWN process in Equation (1).

Now let's write out the objective in Problem (11) at the feasible point given by Equation (12).

$$\begin{aligned}
&\nu^{1\top} W^1 x_0 + b^m_i - \epsilon||\nu^{1\top} W^1||_q - \sum_{k=1}^{m-1}(u^{k\top}\lambda^k + l^{k\top}\theta^k + b^{k\top}\nu^k + t^{kU\top}\mu^k + t^{kL\top}\tau^k) \\
=&(\tilde{w}^1)^\top x_0 + b^m_i - \epsilon||\tilde{w}^1||_q - \sum_{k=1}^{m-1}(b^{k\top}\nu^k + t^{kU\top}\mu^k + t^{kL\top}\tau^k) \\
=&(\tilde{w}^1)^\top x_0 + b^m_i - \epsilon||\tilde{w}^1||_q - \sum_{k=1}^{m-1}(-b^{k\top}[s^{kU} \odot neg(\tilde{w}^{k+1}) + s^{kL} \odot relu(\tilde{w}^{k+1})] + t^{kU\top} relu(-\tilde{w}^{k+1}) + t^{kL\top} neg(-\tilde{w}^{k+1}) \\
=&(\tilde{w}^1)^\top x_0 + b^m_i - \epsilon||\tilde{w}^1||_q + \sum_{k=1}^{m-1}\{b^{k\top}[s^{kU} \odot neg(\tilde{w}^{k+1}) + s^{kL} \odot relu(\tilde{w}^{k+1})] + t^{kU\top} neg(\tilde{w}^{k+1}) + t^{kL\top} relu(\tilde{w}^{k+1})\}
\end{aligned} \tag{13}$$

Compare the last line of Equation (13) with the lower bound given by CROWN in Equation (2). Using the proved conclusion $\tilde{w}^1 = \tilde{w}^1_i$, if we can further prove

$$b^m_i + \sum_{k=1}^{m-1}\{b^{k\top}[s^{kU} \odot neg(\tilde{w}^{k+1}) + s^{kL} \odot relu(\tilde{w}^{k+1})] + t^{kU\top} neg(\tilde{w}^{k+1}) + t^{kL\top} relu(\tilde{w}^{k+1}) = \tilde{b}^1_i,$$

then we can conclude that Equation (13) equals to the lower bound in Equation (2). Consider the definition of $\tilde{b}_i$ in Equation (1):

$$\tilde{b}^v_i = relu(\tilde{w}_i{}^{v+1})^\top(s^{vL} \odot b^v + t^{vL}) + neg(\tilde{w}^{v+1}_i)^\top(s^{vU} \odot b^v + t^{vU}) + \tilde{b}^{v+1}_i,\ v = m-1, m-2, \cdots, 1.$$

Repeat the above iteration then we can obtain:

$$\tilde{b}^1_i = \sum_{v=1}^{m-1}\{[relu(\tilde{w}_i{}^{v+1}) \odot s^{vL} + neg(\tilde{w}_i{}^{v+1}) \odot s^{vU}]^\top b^v + relu(\tilde{w}_i{}^{v+1})^\top t^{vL} + neg(\tilde{w}_i{}^{v+1})^\top t^{vU}\} + b^m_i,$$

which is exactly what we want to prove. Now we conclude that the lower bound given by CROWN in Equation (2) is the objective value of the dual problem (11) at the feasible point (12).

Prove CROWN is the optimal solution to Problem (11)**.** Next we prove that when the self-consistency condition in the LP process is met, the feasible point given by Equation (12) is actually the optimal solution to Problem (11), which means the lower bound given by CROWN in Equation (2) is the optimal value to Problem (11).

Consider the prime solution

$$\begin{aligned}
a^0 &= \underset{x \in B_p(x_0,\epsilon)}{\operatorname{argmin}}\ \nu^{1\top} W^1 x \\
z^k &= W^k a^{k-1} + b^k,\ k = 1, 2, ..., m-1 \\
a^k &= sgn(W^{k+1\top}\nu^{k+1}) \odot [s^{kU} \odot z^k + t^{kU}] + [1 - sgn(W^{k+1\top}\nu^{k+1})] \odot [s^{kL} \odot z^k + t^{kL}],\ k = 1, 2, ..., m-1
\end{aligned} \tag{14}$$

where $sgn(x) = 1$, if $x > 0; sgn(x) = 0$, else. Here we are able to choose a consistent set of bounding line parameters $\{s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L}\}$ in the above definition is because of the self-consistency condition of the LP process. We will prove that the prime-dual solution pair given by Equations (14) and (11) are the optimal solutions to the prime problem (4) and the dual problem (11), respectively.

First, it's easy to verify that the prime-dual solution pair satisfies complementarity:

$$\begin{aligned}
{\nu^k}^\top(z^k - W^k a^{k-1} - b^k) &= 0 \\
{\mu^k}^\top(a^k - s^{kU} \odot z^k - t^{kU}) &= 0 \\
{\tau^k}^\top(a^k - s^{kL} \odot z^k - t^{kL}) &= 0 \\
{\lambda^k}^\top(z^k - u^k) &= 0 \\
{\theta^k}^\top(z^k - l^k) &= 0 \\
k = 1, 2, ..., m-1.&
\end{aligned}$$

Now let's show they satisfy Lagrangian Optimality. It's obvious that the dual solution given by Equation (12) satisfies

$$\mu^k \geq 0, \tau^k \leq 0, \lambda^k \geq 0, \theta^k \leq 0,\ k = 1, 2, ..., m-1$$

and Condition (3). We need to prove that at the dual solution given by Equation (12), the prime solution in Equation (14) is the optimal solution to the following problem:

$$\underset{a^0 \in B_p(x_0,\epsilon), a^{[m-1]}, z^{[m-1]}}{\operatorname{argmin}} L(a^0, a^{[m-1]}, z^{[m-1]} \mid \nu^{[m-1]}, \mu^{[m-1]}, \tau^{[m-1]}, \lambda^{[m-1]}, \theta^{[m-1]}). \tag{15}$$

At the dual solution in Equation (12),

$$\begin{aligned}
&L(a^0, a^{[m-1]}, z^{[m-1]} \mid \nu^{[m-1]}, \mu^{[m-1]}, \tau^{[m-1]}, \lambda^{[m-1]}, \theta^{[m-1]}) \\
&= (W_{i:}^{m\top} + \mu^{m-1} + \tau^{m-1})^\top a^{m-1} + {\nu^1}^\top W^1 a^0 + b_i^m + \sum_{k=1}^{m-2} (\mu^k + \tau^k - {W^{k+1}}^\top \nu^{k+1})^\top a^k \\
&+ \sum_{k=1}^{m-1} (\nu^k + \lambda^k + \theta^k - \mu^k \odot s^{kU} - \tau^k \odot s^{kL})^\top z^k \\
&- \sum_{k=1}^{m-1} ({u^k}^\top \lambda^k + {l^k}^\top \theta^k + {b^k}^\top \nu^k + {t^{kU}}^\top \mu^k + {t^{kL}}^\top \tau^k) \\
&= {\nu^1}^\top W^1 a^0 + b_i^m - \sum_{k=1}^{m-1} ({u^k}^\top \lambda^k + {l^k}^\top \theta^k + {b^k}^\top \nu^k + {t^{kU}}^\top \mu^k + {t^{kL}}^\top \tau^k)
\end{aligned}$$

Observe that the last 2 terms in the above formula is independent of the prime variables $\{a^0, a^{[m-1]}, z^{[m-1]}\}$, then it's obvious that the dual solution in Equation (14) is the optimal solution to Problem (15) at the dual solution in Equation (12). Therefore the prime-dual solution pair satisfies Lagrangian Optimality.

Now let's prove the primal feasibility, namely, the prime solution in Equation (14) is a feasible solution to Problem (4). First, its obvious $a^0 \in B_p(x_0, \epsilon)$ since $a^0 = \underset{x \in B_p(x_0,\epsilon)}{\operatorname{argmin}} {\nu^1}^\top W^1 x$. The constraints in the primal problem (4) are

$$z^k = W^k a^{k-1} + b^k,\ k \in [m-1] \tag{16}$$

$$s^{kL} \odot z^k + t^{kL} \leq a^k \leq s^{kU} \odot z^k + t^{kU},\ k \in [m-1] \tag{17}$$

$$l^k \leq z^k \leq u^k,\ k \in [m-1] \tag{18}$$

It's obvious that the primal solution given by Equation (14) satisfies constraints (16) and (17). We need to prove that z^k defined in Equation (14) satisfies constraint (18). For $k = 1$, its obvious that

$$l^1 \leq z^1 = W^1 a^0 + b^1 \leq u^1$$

is true according to the definition of l^1 and u^1 in Problems (6) and (7).

Assume constraint (18) holds true for $1, 2, ..., k-1$. Now let's prove it's true for k. Since constraint (18) holds true for $1, 2, ..., k-1$, we know $a^0, a^1, ..., a^{k-1}, z^1, z^2, ..., z^{k-1}$ given by Equation (14) is a feasible solution to

Problems (6) and (7) due to the self-consistency condition of the LP process: The bounding line parameters s and t in Equation (14)) are the same of those in Problems (6) and (7). That is to say z^k given by Equation (14) is the objective function value at a feasible solution of Problems (6) and (7). According to the definition of l^k and u^k, where l^k is the minimum value of Problems (6) and u^k is the maximum value of Problems (7), we must have $l^k \leq z^k \leq u^k$. The constraint (18) holds true for k. Therefore we conclude that constraint (18) holds true for any $k \in [m-1]$.

So far we have proven the prima-dual solution pair given by Equation (14) and (12) satisfies complementarity, Lagrangian optimality, and primal feasibility, namely, the solution pair is a saddle point of the Lagrangian function (10). Therefore we conclude that the solution pair are the optimal solutions to the prime problem and dual problem respectively, and the duality gap is 0. Therefore, the solution given by CROWN is the optimal value of Problem (4).

Prove the theorem under the original condition. Although we are proving the above conclusion for the last layer of the neural network, namely, the bounds of the final output, which can be also seen as the preactivation of the m-th layer, given by CROWN and LP are the same under Condition 2, the same argument goes for an arbitrary intermediate layer (except for the first layer) by simply replace m with the index of the corresponding intermediate layer. In other words, the bounds of the preactivation of an arbitrary intermediate layer (except for the first layer) given by CROWN and LP are the same under Condition 2. While for the first layer, both LP and CROWN don't need bounding lines of the previous layer (There is no previous layers in this case anyway.) to compute bounds. Both of them can give the bounds of the first layer directly and we can verify that they are the same according to their definitions.

Combining all the above argument, we conclude that if we run LP and CROWN independently, with the constraint that both LP and CROWN satisfy self-consistency and use the same set of bounding lines, they will give exactly the same bounds of all the preactivations and the final output.

A.6 Tighter bounding lines lead to tighter bounds under weaker conditions

We briefly repeat the CROWN process here. Given the bounds of the previous $m-1$ layers, $l^{[m-1]}$ and $u^{[m-1]}$, one could compute bounds of the m-th layer through the following steps. Repeat the following iteration

$$\begin{aligned}(\tilde{w}_i^v)^\top &= [relu(\tilde{w}_i^{v+1}) \odot s^{vL} + neg(\tilde{w}_i^{v+1}) \odot s^{vU}]^\top W^v, \\ \tilde{b}_i^v &= relu(\tilde{w}_i{}^{v+1})^\top (s^{vL} \odot b^v + t^{vL}) + neg(\tilde{w}_i^{v+1})^\top (s^{vU} \odot b^v + t^{vU}) + \tilde{b}_i^{v+1}, \\ v &= m-1, m-2, \cdots, 1,\end{aligned} \tag{19}$$

to obtain $\tilde{w}_i^1$ and $\tilde{b}_i^1$, where the starting variable is defined as

$$\begin{aligned}\tilde{w}_i^m &= W_{i,:}^{m\top}, \\ \tilde{b}_i^m &= b_i^m,\end{aligned}$$

and the parameters of the bounding lines $s^{[m-1]L}, s^{[m-1]U}, t^{[m-1]L}, t^{[m-1]U}$ are chosen to satisfy the following constraints:

$$\begin{aligned}&s_i^{vL} z_i^v + t_i^{vL} \leq \sigma(z_i^v) \leq s_i^{vU} z_i^v + t_i^{vU}, \\ &\quad \forall\, z_i^v \in [l_i^v, u_i^v], i \in [n_v], v \in [m-1].\end{aligned}$$

Then the lower bound $\gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})$ of z_i^m can be computed by the following formula:

$$\begin{aligned}z_i^m &= W_{i,:}^m a^{m-1} + b_i^m \geq (\tilde{w}_i^1)^\top x + \tilde{b}_i^1 \\ &\geq (\tilde{w}^1)^\top x_0 - \epsilon ||\tilde{w}^1||_q + \tilde{b}_i^1 \\ &= \gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L}), \\ &\forall x \in \mathbb{B}_p(x_0, \epsilon).\end{aligned} \tag{20}$$

We emphasize that the bounds of the preactivations, $l^{[m-1]}$ and $u^{[m-1]}$, need not to be computed by the CROWN. They only need to be valid bounds of $z^{[m-1]}$. They could be computed by the CROWN, the LP-based method or other possible methods. Then we have the following theorem:

Theorem 1 *Suppose the bounds of the preactivations in the previous $m-1$ layers, $l^{[m-1]}$ and $u^{[m-1]}$, are known, and the robustness of a neural network is evaluated by CROWN on two trials with two different sets of bounding lines characterized by $\{\tilde{s}^{[m-1]U}, \tilde{s}^{[m-1]L}, \tilde{t}^{[m-1]U}, \tilde{t}^{[m-1]L}\}$ and $\{\hat{s}^{[m-1]U}, \hat{s}^{[m-1]L}, \hat{t}^{[m-1]U}, \hat{t}^{[m-1]L}\}$. If $\{\hat{s}^{[m-1]U}, \hat{s}^{[m-1]L}, \hat{t}^{[m-1]U}, \hat{t}^{[m-1]L}\}$, $l^{[m-1]}$ and $u^{[m-1]}$ satisfy Condition 4, the closed-form bounds obtained via CROWN satisfy*

$$\begin{aligned}&\gamma_i^{(m)L}(\tilde{s}^{[m-1]U}, \tilde{s}^{[m-1]L}, \tilde{t}^{[m-1]U}, \tilde{t}^{[m-1]L}) \geq \\ &\gamma_i^{(m)L}(\hat{s}^{[m-1]U}, \hat{s}^{[m-1]L}, \hat{t}^{[m-1]U}, \hat{t}^{[m-1]L}), \\ &\gamma_i^{(m)U}(\tilde{s}^{[m-1]U}, \tilde{s}^{[m-1]L}, \tilde{t}^{[m-1]U}, \tilde{t}^{[m-1]L}) \leq \\ &\gamma_i^{(m)U}(\hat{s}^{[m-1]U}, \hat{s}^{[m-1]L}, \hat{t}^{[m-1]U}, \hat{t}^{[m-1]L}),\end{aligned}$$

for $\forall\, i \in [n_m]$, when bounding lines determined by $\{\tilde{s}^{[m-1]U}, \tilde{s}^{[m-1]L}, \tilde{t}^{[m-1]U}, \tilde{t}^{[m-1]L}\}$ are the same as or tighter than those determined by $\{\hat{s}^{[m-1]U}, \hat{s}^{[m-1]L}, \hat{t}^{[m-1]U}, \hat{t}^{[m-1]L}\}$.

Condition 4 *The bounding line parameters $\{s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L}\}$ and bounds of the preactivations $l^{[m-1]}, u^{[m-1]}$ satisfy the following constraints:*

$$\gamma_i^{kL}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L}) \geq l_i^k, \gamma_i^{kU}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L}) \leq u_i^k, \forall i \in [n_k], \forall k \in [m-1].$$

Note that Condition 4 is weaker than the self-consistency condition defined in Condition 1 in the main text. That is because when the preactivations are obtained by the CROWN and the self-consistency condition is met in this process. We will have

$$l_i^k = \gamma_i^{kL}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L}), u_i^k = \gamma_i^{kU}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L}), \forall i \in [n_k], \forall k \in [m-1],$$

which makes Condition 4 hold true.

Below we begin our proof of Theorem 1:

Overview. Introduce two new variables r_j^{1kU} and r_j^{2kU}:

$$\begin{aligned} s_j^{kU} l_j^k + t_j^{kU} &= r_j^{1kU}, \\ s_j^{kU} u_j^k + t_j^{kU} &= r_j^{2kU}. \end{aligned} \tag{21}$$

They are the function values of the the function $s_j^{kU} z_j^k + t_j^{kU}$ at the two endpoints of the interval $[l_j^k, u_j^k]$. It's obvious that if we decrease r_j^{1kU} or r_j^{2kU}, the upper bounding line charaterized by r_j^{1kU} and r_j^{2kU} will become tighter in the interval $[l_j^k, u_j^k]$. Similarly, we can define r_j^{1kL} and r_j^{2kL} as

$$\begin{aligned} s_j^{kL} l_j^k + t_j^{kL} &= r_j^{1kL}, \\ s_j^{kL} u_j^k + t_j^{kL} &= r_j^{2kL}. \end{aligned}$$

Define the linear function $f_i^{mL}(x)$ as the following:

$$z_i^m = W_{i,:}^m a^{m-1} + b_i^m \geq (\tilde{w}_i^1)^\top a^0 + \tilde{b}_i^1 = (\tilde{w}_i^1)^\top x + \tilde{b}_i^1 = f_i^{mL}(x),\ \forall x \in \mathbb{B}_p(x_0, \epsilon). \tag{22}$$

Note that $f_i^{mL}(x)$ is linear in term of x. However, its explicit form depends on the bounding line parameters $\{s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L}\}$. We charaterize its dependency on these parameters as

$$f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L}).$$

The relationship between lower bound $\gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})$ given by CROWN and $f_i^{mL}(x)$ is

$$\gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L}) = \min_{x \in \mathbb{B}_p(x_0,\epsilon)} f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L}) \tag{23}$$

If we can prove

$$\frac{\partial f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{1kU}} \leq 0, \frac{\partial f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{2kU}} \leq 0,$$
$$\forall x \in \mathbb{B}_p(x_0, \epsilon), \forall j \in [n_k], \forall k \in [m-1],$$

we will have

$$\frac{\partial \gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{1kU}} \leq 0, \frac{\partial \gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{2kU}} \leq 0,$$
$$\forall j \in [n_k], \forall k \in [m-1].$$

In other words, tighter upper bounding lines will lead to tighter lower bound of z_i^m. The same argument goes for the lower bounding line as well.

Now let's prove the above inequality.

Define auxiliary variable y

$$y^0 = x$$

and

$$\begin{aligned} y^v &= sgn(\tilde{w}_i^{v+1}) \odot [s^{vL} \odot (W^v y^{v-1} + b^v) + t^{vL}] + [1 - sgn(\tilde{w}_i^{v+1})] \odot [s^{vU} \odot (W^v y^{v-1} + b^v) + t^{vU}] \\ v &= 1, 2, ..., m-1, \end{aligned} \tag{24}$$

where $sgn(x)$ is defined as $sgn(x) = 1$, if $x \geq 0$; 0, else. Note that y^v is dependent of x. Using the definition of $\tilde{w}_i^v$ and $\tilde{b}_i^v$ in Equation (19), we have

$$\begin{aligned} &(\tilde{w}_i^v)^\top y^{v-1} + \tilde{b}_i^v \\ &= relu(\tilde{w}_i{}^{v+1})^\top [s^{vL} \odot (W^v y^{v-1} + b^v) + t^{vL}] + neg(\tilde{w}_i^{v+1})^\top [s^{vU} \odot (W^v y^{v-1} + b^v) + t^{vU}] + \tilde{b}_i^{v+1} \\ &= \tilde{w}_i{}^{v+1} \{sgn(\tilde{w}_i{}^{v+1}) \odot [s^{vL} \odot (W^v y^{v-1} + b^v) + t^{vL}] + [1 - sgn(\tilde{w}_i^{v+1})] \odot [s^{vU} \odot (W^v y^{v-1} + b^v) + t^{vU}]\} + \tilde{b}_i^{v+1} \\ &= (\tilde{w}_i^{v+1})^\top y^v + \tilde{b}_i^{v+1}. \end{aligned}$$

That is to say

$$(\tilde{w}_i^v)^\top y^{v-1} + \tilde{b}_i^v = (\tilde{w}_i^{v+1})^\top y^v + \tilde{b}_i^{v+1}, \forall v \in [m-1] \tag{25}$$

Recall $f_i^{mL}(x)$ defined in Equation (35). We have

$$f_i^{mL}(x) = (\tilde{w}_i^1)^\top x + \tilde{b}_i^1 = (\tilde{w}_i^2)^\top y^1 + \tilde{b}_i^2 = \cdots = (w_i^m)^\top y^{m-1} + b_i^m. \tag{26}$$

Compute $\frac{\partial f_i^{mL}(x)}{\partial s_j^{kU}}$. Using the definition of $\tilde{w}_i^v$ and $\tilde{b}_i^v$ in Equation (19), we have

$$\frac{\partial \tilde{w}_i^v}{\partial s_j^{kU}} = \begin{cases} (W^v)^\top [s^{vL} \odot \frac{\partial relu(\tilde{w}_i^{v+1})}{\partial s_j^{kU}} + s^{vU} \odot \frac{\partial neg(\tilde{w}_i^{v+1})}{\partial s_j^{kU}}], & \text{if } v < k, \\ neg(\tilde{w}_{ij}^{k+1}) W_{j,:}^k, & \text{if } v = k, \\ 0, & \text{if } v > k. \end{cases} \tag{27}$$

$$\frac{\partial \tilde{b}_i^v}{\partial s_j^{kU}} = \begin{cases} \frac{\partial relu(\tilde{w}_i^{v+1})}{\partial s_j^{kU}}^\top (s^{vL} \odot b^v + t^{vL}) + \frac{\partial neg(\tilde{w}_i^{v+1})}{\partial s_j^{kU}}^\top (s^{vU} \odot b^v + t^{vU}) + \frac{\partial \tilde{b}_i^{v+1}}{\partial s_j^{kU}}, & \text{if } v < k, \\ neg(\tilde{w}_{ij}^{k+1}) b_j^k, & \text{if } v = k, \\ 0, & \text{if } v > k. \end{cases} \tag{28}$$

Note that both sides of Equation (27) is a vector. Combining the above conclusion with Equation (26), we have

$$\begin{aligned} \frac{\partial f_i^{mL}(x)}{\partial s_j^{kU}} = \frac{\partial ((\tilde{w}_i^1)^\top y^0 + \tilde{b}_i^k)}{\partial s_j^{kU}} = \cdots = \frac{\partial ((\tilde{w}_i^k)^\top y^{k-1} + \tilde{b}_k^1)}{\partial s_j^{kU}} &= (\frac{\partial \tilde{w}_i^k}{\partial s_j^{kU}})^\top y^{k-1} + \frac{\partial \tilde{b}_i^k}{\partial s_j^{kU}} \\ &= neg(\tilde{w}_{ij}^{k+1})(W_{j,:}^k)^\top y^{k-1} + \tilde{w}_{ij}^{k+1} b_j^k \\ &= neg(\tilde{w}_{ij}^{k+1})[(W_{j,:}^k)^\top y^{k-1} + b_j^k] \end{aligned} \tag{29}$$

In the same manner, we can prove

$$\frac{\partial f_i^{mL}(x)}{\partial t_j^{kU}} = neg(\tilde{w}_{ij}^{k+1}) \tag{30}$$

Compute $\frac{\partial f_i^{mL}(x)}{\partial r_j^{1kU}}$ and $\frac{\partial f_i^{mL}(x)}{\partial r_j^{2kU}}$. According to the definition of r_j^{1kU} and r_j^{2kU} in Equation (21), we have

$$\frac{\partial s_j^{kU}}{\partial r_j^{1kU}} = -\frac{1}{u_j^k - l_j^k}, \ \frac{\partial t_j^{kU}}{\partial r_j^{1kU}} = \frac{u_j^k}{u_j^k - l_j^k}$$

and

$$\frac{\partial s_j^{kU}}{\partial r_j^{2kU}} = \frac{1}{u_j^k - l_j^k}, \ \frac{\partial t_j^{kU}}{\partial r_j^{2kU}} = -\frac{l_j^k}{u_j^k - l_j^k}.$$

Combining ths result with Equations (29) and (30), we have

$$\begin{aligned} \frac{\partial f_i^{mL}(x)}{\partial r_j^{1kU}} &= \frac{\partial f_i^{mL}(x)}{\partial s_j^{kU}} \frac{\partial s_j^{kU}}{\partial r_j^{1kU}} + \frac{\partial f_i^{mL}(x)}{\partial t_j^{kU}} \frac{\partial t_j^{kU}}{\partial r_j^{1kU}} \\ &= -\frac{1}{u_j^k - l_j^k} \frac{\partial f_i^{mL}(x)}{\partial s_j^{kU}} + \frac{u^k}{u^k - l^k} \frac{\partial f_i^{mL}(x)}{\partial t_j^{kU}} \\ &= \frac{1}{u_j^k - l_j^k} neg(\tilde{w}_{ij}^{k+1})[u_j^k - (W_{j,:}^k{}^\top y^k + b_j^k)] \end{aligned}$$

and

$$\begin{aligned}\frac{\partial f_i^{mL}(x)}{\partial r_j^{2kU}} &= \frac{\partial f_i^{mL}(x)}{\partial s_j^{kU}}\frac{\partial s_j^{kU}}{\partial r_j^{2kU}} + \frac{\partial f_i^{mL}(x)}{\partial t_j^{kU}}\frac{\partial t_j^{kU}}{\partial r_j^{2kU}} \\ &= \frac{1}{u_j^k - l_j^k}\frac{\partial f_i^{mL}(x)}{\partial s_j^{kU}} - \frac{l_j^k}{u_j^k - l_j^k}\frac{\partial f_i^{mL}(x)}{\partial t_j^{kU}} \\ &= \frac{1}{u_j^k - l_j^k} neg(\tilde{w}_{ij}^{k+1})({W_{j,:}^k}^\top y^{k-1} + b_j^k - l_j^k).\end{aligned}$$

We know that $u_j^k - l_j^k \geq 0$ and $neg(\tilde{w}_{ij}^{k+1}) \leq 0$. Therefore, if we can prove $u_j^k \geq {W_{j,:}^k}^\top y^{k-1} + b_j^k \geq l_j^k$, $\forall x \in \mathbb{B}_p(x_0, \epsilon)$, we will have $\frac{\partial f_i^{mL}(x)}{\partial r_j^{1kU}} \leq 0$ and $\frac{\partial f_i^{mL}(x)}{\partial r_j^{2kU}} \leq 0$, $\forall x \in \mathbb{B}_p(x_0, \epsilon)$.

Prove $u_j^k \geq {W_{j,:}^k}^\top y^{k-1} + b_j^k \geq l_j^k$, $\forall x \in \mathbb{B}_p(x_0, \epsilon)$**.** For $k = 1$, $u_j^1 \geq {W_{j,:}^1}^\top y^0 + b_j^1 = {W_{j,:}^1}^\top x + b_j^1 \geq l_j^1$, $\forall x \in \mathbb{B}_p(x_0, \epsilon)$. Obviously, the proposition is true according to the definition of u_j^1 and l_j^1 in Equations (3). That is to say

$$\frac{\partial f_i^{mL}(x)}{\partial r_j^{1(1)U}} \leq 0, \frac{\partial f_i^{mL}(x)}{\partial r_j^{2(1)U}} \leq 0, \forall x \in \mathbb{B}_p(x_0, \epsilon)$$

always holds true. Therefore, tighter upper bounding lines in the first layer always leads to tighter bound of z_i^m.

Now lets prove $u_j^k \geq {W_{j,:}^k}^\top y^{k-1} + b_j^k \geq l_j^k$, $\forall x \in \mathbb{B}_p(x_0, \epsilon)$ holds true in general. Assume the proposition is true for $1, 2, \cdots, k-1$. Now let's prove it's true for k. Namely we want to prove

$$u_j^k \geq {W_{j,:}^k}^\top y^{k-1} + b_j^k \geq l_j^k.$$

Recall the definition of $s^{(k-1)L}, t^{(k-1)L}, s^{(k-1)U}, t^{(k-1)U}$. They satisfy

$$s_j^{(k-1)L} z_j^{k-1} + t_j^{(k-1)L} \leq \sigma(z_j^{k-1}) \leq s_j^{(k-1)U} z_j^{k-1} + t_j^{(k-1)U}, \forall z_j^{k-1} \in [l_j^{k-1}, u_j^{k-1}], \forall j \in [n_{k-1}].$$

Therefore we have

$$u_j^{k-1} \geq {W_{j,:}^{k-1}}^\top y^{k-2} + b_j^{k-1} \geq l_j^{k-1} \Rightarrow [s^{(k-1)L} \odot (W^{k-1}y^{k-2} + b^{k-1}) + t^{(k-1)L}] \leq [s^{(k-1)U} \odot (W^{k-1}y^{k-1-1} + b^{k-1}) + t^{(k-1)U}].$$

Combing this with the definition of y^{k-1}:

$$y^{k-1} = sgn(\tilde{w}_i^k) \odot [s^{(k-1)L} \odot (W^{k-1}y^{k-2} + b^{k-1}) + t^{(k-1)L}] + [1 - sgn(\tilde{w}_i^k)] \odot [s^{(k-1)U} \odot (W^{k-1}y^{k-2} + b^{k-1}) + t^{(k-1)U}]$$

We can obtain that

$$[s^{(k-1)L} \odot (W^{k-1}y^{k-2} + b^{k-1}) + t^{(k-1)L}] \leq y^{k-1} \leq [s^{(k-1)U} \odot (W^{k-1}y^{k-2} + b^{k-1}) + t^{(k-1)U}].$$

Plugging in the bounds of y^{k-1}, we can derive a lower bound of ${W_{j,:}^k}^\top y^{k-1} + b_j^k$:

$$\begin{aligned}&{W_{j,:}^k}^\top y^{k-1} + b_j^k \\ &\geq relu(W_{j,:}^k)^\top [s^{(k-1)L} \odot (W^{k-1}y^{k-2} + b^{k-1}) + t^{(k-1)L}] + neg(W_{j,:}^k)^\top [s^{(k-1)U} \odot (W^{k-1}y^{k-2} + b^{k-1}) + t^{(k-1)U}] + b_j^k \\ &= [relu(W_{j,:}^k) \odot s^{(k-1)L} + neg(W_{j,:}^k) \odot s^{(k-1)U}]^\top W^{k-1}y^{k-2} \\ &+ relu(W_{j,:}^k)^\top (s^{(k-1)L} \odot b^{k-1} + t^{(k-1)L}) + neg(W_{j,:}^k)^\top (s^{(k-1)U} \odot b^{k-1} + t^{(k-1)U}) + b_j^k \\ &= (\tilde{w'}_j^{k-1})^\top y^{k-2} + \tilde{b'}_j^{k-1}.\end{aligned}$$

In the last line of the above equation , we have plugged in the definition of $\tilde{w'}_j^{k-1}$ and $\tilde{b'}_j^{k-1}$ in Equation (1) for $v = k-1$ by replacing i with j and adding superscript $'$ to $\tilde{w}$ and $\tilde{b}$. The superscript $'$ here on $\tilde{w'}_j^{k-1}$ and $\tilde{b'}_j^{k-1}$ is to hightlight their difference from $\tilde{w}_j^{k-1}$ and $\tilde{b}_j^{k-1}$ in the iteration defined in Equation (19).

Now we have proved that ${W_{j,:}^k}^\top y^{k-1} + b_j^k \geq (\tilde{w'}_j^{k-1})^\top y^{k-2} + \tilde{b'}_j^{k-1}$. Use the definition of y^{k-2} and the condition $u_j^{k-2} \geq {W_{j,:}^{k-2}}^\top y^{k-3} + b_j^{k-2} \geq l_j^{k-2}$ we can further prove

$$(W_{j,:}^k)^\top y^{k-1} + b_j^k \geq (\tilde{w'}_j^{k-1})^\top y^{k-2} + \tilde{b'}_j^{k-1} \geq (\tilde{w'}_j^{k-2})^\top y^{k-3} + \tilde{b'}_j^{k-2}$$

where $\tilde{w'}_j^{k-2}$ and $\tilde{b'}_j^{k-2}$ are obtained by the same iteration as in Equation (1) for $v = k-2$ by replacing i with j and adding superscript $'$ to $\tilde{w}$ and $\tilde{b}$. Repeat the above the process, eventually we get

$$(W_{j,:}^k)^\top y^{k-1} + b_j^k \geq (\tilde{w'}_j^{k-1})^\top y^{k-2} + \tilde{b'}_j^{k-1} \geq (\tilde{w'}_j^{k-2})^\top y^{k-3} + \tilde{b'}_j^{k-2} \geq \cdots \geq (\tilde{w'}_j^{1})^\top x + \tilde{b'}_j^{1},$$

where the variables $\tilde{w'}_j^{[k-1]}$ and $\tilde{b'}_j^{[k-1]}$ are obtained through the iteration in Equation (1) by replacing i with j and adding superscript $'$ to $\tilde{w}$ and $\tilde{b}$. Remind that the iteration in Equation (1) is used to compute the lower bound of z_i^k: $\gamma_i^{kL}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L})$. Therefore we can see the iteration of the variables $\tilde{w'}_j^{[k-1]}$ and $\tilde{b'}_j^{[k-1]}$ as the CROWN process to compute the lower bound of z_j^k: $\gamma_j^{kL}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L})$. That is to say

$$\begin{aligned} z_j^k &= W_{j,:}^k a^{k-1} + b_j^k \geq (\tilde{w'}_j^{1})^\top x + \tilde{b'}_j^{1} \\ &\geq (\tilde{w'}^{1})^\top x_0 - \epsilon ||\tilde{w'}^{1}||_q + \tilde{b'}_i^{1} \\ &= \gamma_j^{kL}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L}), \\ &\forall x \in \mathbb{B}_p(x_0, \epsilon). \end{aligned}$$

Therefore we have

$${W_{j,:}^k}^\top y^{k-1} + b_j^k \geq (\tilde{w'}_j^{1})^\top x + \tilde{b'}_j^{1} \geq \gamma_j^{kL}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L})$$

Using the condition

$$\gamma_j^{kL}(s^{[k-1]U}, s^{[k-1]L}, t^{[k-1]U}, t^{[k-1]L}) \geq l_j^k,$$

we conclude that ${W_{j,:}^k}^\top y^{k-1} + b_j^k \geq l_j^k$. Similarly, we can prove ${W_{j,:}^k}^\top y^{k-1} + b_j^k \leq u_j^k$. Therefore we conclude

$$u_j^k \geq {W_{j,:}^k}^\top y^{k-1} + b_j^k \geq l_j^k, \ \forall x \in \mathbb{B}_p(x_0, \epsilon), \forall k \in [m-1].$$

Conclusion. Now we have proven that

$$\frac{\partial f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{1(1)U}} \leq 0, \frac{\partial f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{2(1)U}} \leq 0,$$
$$\forall x \in \mathbb{B}_p(x_0, \epsilon), \forall j \in [n],$$

and

$$\frac{\partial f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{1kU}} \leq 0, \frac{\partial f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{2kU}} \leq 0,$$
$$\forall x \in \mathbb{B}_p(x_0, \epsilon), \forall j \in [n_k], 2 \leq k \leq m-1,$$
$$\text{if } \gamma_c^{vL}(s^{[v-1]U}, s^{[v-1]L}, t^{[v-1]U}, t^{[v-1]L}) \geq l_c^v, \gamma_c^{vU}(s^{[v-1]U}, s^{[v-1]L}, t^{[v-1]U}, t^{[v-1]L}) \leq u_c^v, \forall c \in [n_v], \forall v \in [k-1].$$

Using the relationship between $f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})$ and $\gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})$ in Equation (23), we conclude that

$$\frac{\partial \gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{1(1)U}} \leq 0, \frac{\partial \gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{2(1)U}} \leq 0, \forall j \in [n],$$

and

$$\frac{\partial \gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{1kU}} \leq 0, \frac{\partial \gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{2kU}} \leq 0,$$

$\forall j \in [n_k], 2 \leq k \leq m-1,$

if $\gamma_c^{vL}(s^{[v-1]U}, s^{[v-1]L}, t^{[v-1]U}, t^{[v-1]L}) \geq l_c^v, \gamma_c^{vU}(s^{[v-1]U}, s^{[v-1]L}, t^{[v-1]U}, t^{[v-1]L}) \leq u_c^v, \forall c \in [n_v], \forall v \in [k-1].$

The conclusion for the lower bounding line parameters can be similarly derived:

$$\frac{\partial \gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{1(1)L}} \geq 0, \frac{\partial \gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{2(1)L}} \geq 0, \forall j \in [n],$$

and

$$\frac{\partial \gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{1kL}} \geq 0, \frac{\partial \gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial r_j^{2kL}} \geq 0,$$

$\forall j \in [n_k], 2 \leq k \leq m-1,$

if $\gamma_c^{vL}(s^{[v-1]U}, s^{[v-1]L}, t^{[v-1]U}, t^{[v-1]L}) \geq l_c^v, \gamma_c^{vU}(s^{[v-1]U}, s^{[v-1]L}, t^{[v-1]U}, t^{[v-1]L}) \leq u_c^v, \forall c \in [n_v], \forall v \in [k-1].$

Although we are proving the above conclusion for the lower bound of the m-th layer, the same argument actually goes for the upper and lower bounds of an arbitrary intermediate layer, which leads us to the following conclusion:

$$\begin{aligned}
&\frac{\partial \gamma_i^{kL}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{1(1)U}} \leq 0, \frac{\partial \gamma_i^{kL}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{2(1)U}} \leq 0, \\
&\frac{\partial \gamma_i^{kU}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{1(1)U}} \geq 0, \frac{\partial \gamma_i^{kU}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{2(1)U}} \geq 0, \\
&\frac{\partial \gamma_i^{kL}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{1(1)L}} \geq 0, \frac{\partial \gamma_i^{kL}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{2(1)L}} \geq 0, \\
&\frac{\partial \gamma_i^{kU}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{1(1)L}} \leq 0, \frac{\partial \gamma_i^{kU}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{2(1)L}} \leq 0, \\
&\forall j \in [n], 2 \leq g \leq m,
\end{aligned} \tag{31}$$

and

$$\frac{\partial \gamma_i^{gL}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{1kU}} \leq 0, \frac{\partial \gamma_i^{gL}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{2kU}} \leq 0,$$

$$\frac{\partial \gamma_i^{gU}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{1kU}} \geq 0, \frac{\partial \gamma_i^{gU}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{2kU}} \geq 0,$$

$$\frac{\partial \gamma_i^{gL}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{1kL}} \geq 0, \frac{\partial \gamma_i^{gL}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{2kL}} \geq 0,$$

$$\frac{\partial \gamma_i^{gU}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{1kL}} \leq 0, \frac{\partial \gamma_i^{gU}(s^{[g-1]U}, s^{[g-1]L}, t^{[g-1]U}, t^{[g-1]L})}{\partial r_j^{2kL}} \leq 0,$$

$\forall j \in [n_k], 2 \leq k \leq g-1, 3 \leq g \leq m$

if $\gamma_c^{vL}(s^{[v-1]U}, s^{[v-1]L}, t^{[v-1]U}, t^{[v-1]L}) \geq l_c^v, \gamma_c^{vU}(s^{[v-1]U}, s^{[v-1]L}, t^{[v-1]U}, t^{[v-1]L}) \leq u_c^v, \forall c \in [n_v], \forall v \in [k-1].$ (32)

From Equations (31) and (32), we conclude that Theorem 1 is true under Condition 4.

A.7 Problems (9) and (10) in the main text are non-convex in general

We repeat Problems (9) and (10) in the main text here:

$$\begin{aligned} \max_{s^{[k-1]L}, s^{[k-1]U}, t^{[k-1]L}, t^{[k-1]U}} \quad & \gamma_i^{(k)L} \\ \text{s.t.} \quad & s_i^{vL} z_i^v + t_i^{vL} \le \sigma(z_i^v) \le s_i^{vU} z_i^v + t_i^{vU}, \\ & \forall\, z_i^v \in [\mathbf{l}_i^v, \mathbf{u}_i^v], i \in [n_v], v \in [k-1], \end{aligned} \tag{33}$$

and

$$\begin{aligned} \min_{s^{[k-1]L}, s^{[k-1]U}, t^{[k-1]L}, t^{[k-1]U}} \quad & \gamma_i^{(k)U} \\ \text{s.t.} \quad & s_i^{vL} z_i^v + t_i^{vL} \le \sigma(z_i^v) \le s_i^{vU} z_i^v + t_i^{vU}, \\ & \forall\, z_i^v \in [\mathbf{l}_i^v, \mathbf{u}_i^v], i \in [n_v], v \in [k-1]. \end{aligned} \tag{34}$$

Now we prove Problems (33) and (34) are convex when there are two layers, but can be non-convex when there are three or more layers.

We only prove the conclusion for problem (33) and a similar analysis can be made to problem (34). First, the feasible region is convex. It can be derived from the definition of convexity:

$$\left.\begin{array}{l} s_i^{vL} z_i^v + t_i^{vL} \le \sigma(z_i^v) \\ s_i'^{vL} z_i^v + t_i'^{vL} \le \sigma(z_i^v) \end{array}\right\} \Rightarrow \left(\lambda s_i^{vL} + \lambda' s_i'^{vL}\right) z_i^v + \left(\lambda t_i^{vL} + \lambda' t_i'^{vL}\right) \le \sigma(z_i^v)$$

where $\lambda, \lambda' \in (0,1)$ with $\lambda + \lambda' = 1$.

Now, let's take a look at the objective function

$$\gamma_i^{kL} = \tilde{W}_{i,:}^1 x_0 - \epsilon \|\tilde{W}_{i,:}^1\|_q + \tilde{b}_i^1$$

When there are only two layers, we can expand the $\tilde{W}_{i,j}^1$ and $\tilde{b}^1$ as follows:

$$\tilde{W}_{i,j}^1 = \sum_{v=1}^{n_1} W_{v,j}^1 \left(relu(W_{i,v}^2) s_v^{1L} + neg(W_{i,v}^2) s_v^{1U}\right)$$

$$\tilde{b}_i^1 = [relu(W_{i,:}^2) \odot s^{1L} + neg(W_{i,:}^2) \odot s^{1U}] b^1 + relu(W_{i,:}^2) t^{1L} + neg(W_{i,:}^2) t^{1U} + b_i^2$$

Therefore, $\tilde{W}_{i,j}^1$ is linear in $s_v^{(1)L/U}$, and $\tilde{b}_i^1$ is linear in $s_v^{(1)L/U}$ and $t_v^{(1)L/U}$. Since $\|\cdot\|_q$ is convex, problem (33) is convex in this case.

When there are three layers, we show that $\tilde{W}_{i,j}^1$ is not convex, and $\tilde{W}_{i,:}^1 x_0 + \tilde{b}_i^1$ is not concave. First,

$$\tilde{W}_{i,j}^1 = \sum_{v=1}^{n_1} W_{v,j}^1 \left(relu(\tilde{W}_{i,v}^2) s_v^{1L} + neg(\tilde{W}_{i,v}^2) s_v^{1U}\right)$$

where

$$\tilde{W}_{i,j}^2 = \sum_{v=1}^{n_2} W_{v,j}^2 \left(relu(W_{i,v}^3) s_v^{2L} + neg(W_{i,v}^3) s_v^{2U}\right)$$

Since $\tilde{W}_{i,j}^2$ is a function in $s_v^{(2)L/U}$, $\tilde{W}_{i,j}^1$ is not convex in that a function times a new variable is not convex anymore. To explain, we assume $x \in \mathbb{R}^d, y \in \mathbb{R}$, and $\mathfrak{g} : \mathbb{R}^d \to \mathbb{R}$ is a (possibly convex) function with Hessian $H_{\mathfrak{g}}$ and gradient $\nabla_{\mathfrak{g}}$. Then, the Hessian of the function $\mathfrak{g}(x)y$ is

$$\left(\begin{array}{c:c} H_{\mathfrak{g}}(x) & \nabla_{\mathfrak{g}}(x) \\ \hdashline \nabla_{\mathfrak{g}}(x)^\top & 0 \end{array}\right)$$

Then, its quadratic form at (x, y) is given by

$$x^\top H_{\mathfrak{g}}(x)x + 2\nabla_{\mathfrak{g}}(x)^\top xy$$

This can always be less than 0 by choosing an appropriate y. Since $W^k(k = 1, 2, 3)$ can be arbitrary, $\|\tilde{W}^1_{i,:}\|_q$ can be non-convex. Similarly, we have $-\tilde{W}^1_{i,j}$ is not convex, so $\tilde{W}^1_{i,j}$ is not concave. Then, $\tilde{W}^1_{i,:}x_0 + \tilde{b}^1_i$ is not concave. This is because the terms involving s^{1L}_v can be extended as

$$\left(\sum_{j=1}^{n_1} relu(\tilde{W}^2_{i,v})(W^1_{v,j}(x_0)_j + b^1_j)\right) s^{1L}_v$$

By the same reason why $\tilde{W}^1_{i,j}$ is neither convex nor concave, this sum is not concave, and similar for s^{1U}_v and $t^{(1)L/U}_v$. Therefore, the objective function can be non-convex, and this analysis can be trivially extended to > 3 layers.

Combining these observations, we conclude that the problem (33) is convex when there are two layers, but can be non-convex when there are three or more layers.

A.8 Prove Theorem 3 in the main text and use it to narrow search space of bounding lines

Recall the CROWN process in Equations (1) and (2). Define the linear function $f_i^{mL}(x)$ as the following:

$$z_i^m = W_{i,:}^m a^{m-1} + b_i^m \geq (\tilde{w}_i^1)^\top a^0 + \tilde{b}_i^1 = (\tilde{w}_i^1)^\top x + \tilde{b}_i^1 = f_i^{mL}(x),\ \forall x \in \mathbb{B}_p(x_0, \epsilon). \tag{35}$$

Note that $f_i^{mL}(x)$ is linear in term of x. However, its explicit form depends on the bounding line parameters $\{s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L}\}$. We charaterize its dependency on these parameters as

$$f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L}).$$

The relationship between lower bound $\gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})$ given by CROWN and $f_i^{mL}(x)$ is

$$\gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L}) = \min_{x \in \mathbb{B}_p(x_0,\epsilon)} f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L}) \tag{36}$$

We first prove that $f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})$ increases if $t^{[m-1]U}$ decreases and $t^{[m-1]L}$ increases. Consider the iteration of $\tilde{w}_i$ in Equation (1):

$$(\tilde{w}_i^v)^\top = [relu(\tilde{w}_i^{v+1}) \odot s^{vL} + neg(\tilde{w}_i^{v+1}) \odot s^{vU}]^\top W^v, v = m-1, m-2, \cdots, 1.$$

We can see that $\tilde{w}_i^v$ is independent of the itercepts t, namely,

$$\frac{\partial \tilde{w}_i^1}{\partial t_j^{vU}} = \frac{\partial \tilde{w}_i^1}{\partial t_j^{vL}} = 0, \forall v \in [m-1], \forall j \in [n_v]. \tag{37}$$

Consider the iteration of $\tilde{b}_i$ in Equation (1):

$$\tilde{b}_i^v = relu(\tilde{w}_i{}^{v+1})^\top (s^{vL} \odot b^v + t^{vL}) + neg(\tilde{w}_i^{v+1})^\top (s^{vU} \odot b^v + t^{vU}) + \tilde{b}_i^{v+1}, v = m-1, m-2, \cdots, 1.$$

We can see that

$$\frac{\partial \tilde{b}_i^k}{\partial t_j^{vU}} = \begin{cases} \frac{\partial \tilde{b}_i^{k+1}}{\partial t_j^{vU}}, & \text{if } k < v, \\ relu(\tilde{w}_{ij}^{v+1}), & \text{if } k = v, \\ 0, & \text{if } k > v \end{cases} \quad \text{and} \quad \frac{\partial \tilde{b}_i^k}{\partial t_j^{vL}} = \begin{cases} \frac{\partial \tilde{b}_i^{k+1}}{\partial t_j^{vL}}, & \text{if } k < v, \\ neg(\tilde{w}_{ij}^{v+1}), & \text{if } k = v, \\ 0, & \text{if } k > v. \end{cases} \tag{38}$$

Then we can obtain

$$\begin{aligned} &\frac{\partial \tilde{b}_i^1}{\partial t_j^{vL}} = \frac{\partial \tilde{b}_i^2}{\partial t_j^{vL}} = \cdots = \frac{\partial \tilde{b}_i^v}{\partial t_j^{vL}} = relu(\tilde{w}_{ij}^{v+1}) \geq 0, \\ &\frac{\partial \tilde{b}_i^1}{\partial t_j^{vU}} = \frac{\partial \tilde{b}_i^2}{\partial t_j^{vU}} = \cdots = \frac{\partial \tilde{b}_i^v}{\partial t_j^{vU}} = neg(\tilde{w}_{ij}^{v+1}) \leq 0, \\ &\forall v \in [m-1], \forall j \in [n_v]. \end{aligned} \tag{39}$$

Combining Equations (37) and (39), we can obtain

$$\begin{aligned} &\frac{\partial f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial t_j^{vL}} \\ &= \frac{\partial [(\tilde{w}_i^1)^\top x + \tilde{b}_i^1]}{\partial t_j^{vL}} \\ &= x^\top \frac{\partial \tilde{w}_i^1}{\partial t_j^{vL}} + \frac{\partial \tilde{b}_i^1}{\partial t_j^{vL}} \\ &= relu(\tilde{w}_{ij}^{v+1}) \geq 0, \forall x \in \mathbb{B}_p(x_0, \epsilon) \end{aligned} \tag{40}$$

and

$$\begin{aligned}
&\frac{\partial f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})}{\partial t_j^{vU}} \\
&= \frac{\partial[(\tilde{w}_i^1)^\top x + \tilde{b}_i^1]}{\partial t_j^{vU}} \\
&= x^\top \frac{\partial \tilde{w}_i^1}{\partial t_j^{vU}} + \frac{\partial \tilde{b}_i^1}{\partial t_j^{vU}} \\
&= neg(\tilde{w}_{ij}^{v+1}) \leq 0, \forall x \in \mathbb{B}_p(x_0, \epsilon).
\end{aligned}$$

This means $f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, t^{[m-1]U}, t^{[m-1]L})$ increases if $t^{[m-1]U}$ decreases and $t^{[m-1]L}$ increases. Therefore

$$\begin{aligned}
f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, \tilde{t}^{[m-1]U}, \tilde{t}^{[m-1]L}) \geq f_i^{mL}(x \mid s^{[m-1]U}, s^{[m-1]L}, \hat{t}^{[m-1]U}, \hat{t}^{[m-1]L}), \forall x \in \mathbb{B}_p(x_0, \epsilon) \quad (41) \\
\text{if } \tilde{t}^{vU} \leq \hat{t}^{vU} \text{ and } \tilde{t}^{vL} \geq \hat{t}^{vL}, \forall v \in [m-1].
\end{aligned}$$

Minimize both sides of Inequality (41) over $x \in \mathbb{B}_p(x_0, \epsilon)$ and consider Equation (36), then we can get

$$\begin{aligned}
\gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, \tilde{t}^{[m-1]U}, \tilde{t}^{[m-1]L}) \geq \gamma_i^{mL}(s^{[m-1]U}, s^{[m-1]L}, \hat{t}^{[m-1]U}, \hat{t}^{[m-1]L}), \\
\text{if } \tilde{t}^{vU} \leq \hat{t}^{vU} \text{ and } \tilde{t}^{vL} \geq \hat{t}^{vL}, \forall v \in [m-1].
\end{aligned}$$

In the same manner, we can prove the same conclusion for the upper bound $\gamma_i^{mU}(s^{[m-1]U}, s^{[m-1]L}, \tilde{t}^{[m-1]U}, \tilde{t}^{[m-1]L})$, namely,

$$\begin{aligned}
\gamma_i^{mU}(s^{[m-1]U}, s^{[m-1]L}, \tilde{t}^{[m-1]U}, \tilde{t}^{[m-1]L}) \leq \gamma_i^{mU}(s^{[m-1]U}, s^{[m-1]L}, \hat{t}^{[m-1]U}, \hat{t}^{[m-1]L}), \\
\text{if } \tilde{t}^{vU} \leq \hat{t}^{vU} \text{ and } \tilde{t}^{vL} \geq \hat{t}^{vL}, \forall v \in [m-1].
\end{aligned}$$

Although we are proving the above conclusion for the bounds of the final output, which can be also seen as the preactivation of the m-th layer, the same argument goes for an arbitrary intermediate layer (except for the first layer) by simply replace m with the index of the corresponding intermediate layer. Therefore the follwowing conclusion can be made:

$$\begin{aligned}
&\gamma_i^{(k)L}(s^{[k-1]U}, s^{[k-1]L}, \tilde{t}^{[k-1]U}, \tilde{t}^{[k-1]L}) \geq \\
&\gamma_i^{(k)L}(s^{[k-1]U}, s^{[k-1]L}, \hat{t}^{[k-1]U}, \hat{t}^{[k-1]L}), \\
&\gamma_i^{(k)U}(s^{[k-1]U}, s^{[k-1]L}, \tilde{t}^{[k-1]U}, \tilde{t}^{[k-1]L}) \leq \\
&\gamma_i^{(k)U}(s^{[k-1]U}, s^{[k-1]L}, \hat{t}^{[k-1]U}, \hat{t}^{[k-1]L}), \\
&\text{if } \tilde{t}^{vU} \leq \hat{t}^{vU} \text{ and } \tilde{t}^{vL} \geq \hat{t}^{vL}, \forall v \in [k-1],
\end{aligned}$$

where $k = 2, 3, \cdots, m$. This is exactly what we want to prove.

Use Theorem 3 to narrow down search space of bounding lines. Theorem 3 essentially states that we should choose bounding lines with tighter intercepts in order to obtain tighter bounds when slopes are fixed. Figure 1 illustrates how we can use Theorem 3 to choose bounding lines for ReLU activaiton when the input interval $[l, u]$ crosses the origin: Bounding lines UB 2, UB 3, LB 2 and LB 3 can yield tighter bounds than bounding lines UB 2.1, UB 3.1, LB 2.1 and LB 3.1, respectively, because they have tighter intercepts. However, we can't determine which one of bounding lines UB1, UB 2 and UB 3 is better, or which one of bounding lines LB1, LB 2 and LB 3 is better. Using Theorem 3, the reduced search space of upper boundling line can be characterized by a single variable d_1:

$$\begin{cases}
\text{Slope} = k_u + d_1, \text{Intercept} = -(k_u + d_1)l, & \text{if } d_1 \geq 0, \\
\text{Slope} = k_u + d_1, \text{Intercept} = u - (k_u + d_1)u, & \text{if } d_1 < 0.
\end{cases} \quad (42)$$

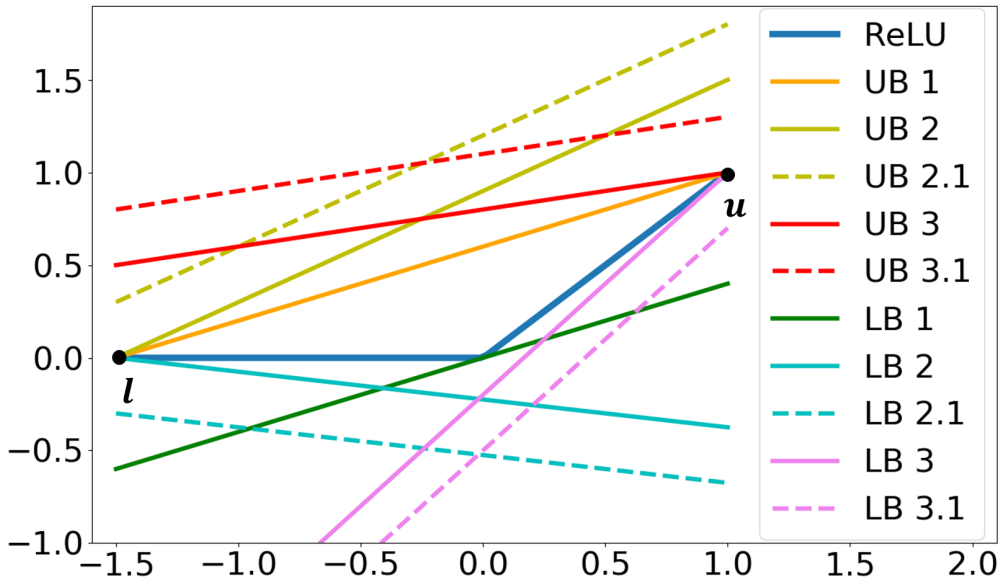

Figure 1: Use Theorem 3 to instruct how to choose bounding lines. Bounding lines UB 2.1, UB 3.1, LB 2.1 and LB 3.1 are paralell to bounding lines UB 2, UB 3, LB 2 and LB 3, respectively. Bounding lines UB 2, UB 3, LB 2 and LB 3 can yield tighter bounds than bounding lines UB 2.1, UB 3.1, LB 2.1 and LB 3.1, respectively

where $[l, u]$ is the input interval of ReLU activation, $l < 0$, $u > 0$, $k_u = u/(u-l)$, $b_u = -k_u l$. Note that this search space of the upper bounding line is characterized by a single variable and the upper bounding line changes continuously with the variable d_1.

Similarly, search space of lower bounding line can be characterized by a single variable d_2:

$$\begin{cases} \text{Slope} = d_2, \text{Intercept} = -d_2 l, & \text{if } d_2 < 0, \\ \text{Slope} = d_2, \text{Intercept} = 0, & \text{if } 0 \le d_2 \le 1, \\ \text{Slope} = d_2, \text{Intercept} = u - d_2 u, & \text{if } d_2 > 1. \end{cases} \tag{43}$$

Note that Theorem 3 can be used to narrow down search space of bounding lines of other activation functions (e.g., Sigmoid, and Tanh) and the corresponding search space can be characterized by a single variable continuously as well. Therefore, we can design gradient based method to search over candidate bounding lines to obtain tighter bounds. However, we further narrow down the search space by requiring the bounding lines to be the tightest bounding lines in order to simplify implementation of FROWN. We emphasize that this requirement is not necessary. FROWN can be readily generalized to the case in which the search space is the one defined in (42) or (43), and the obtained bounds should be even tighter as the search space is larger.

A.9 Experimental set-ups and complete experiment result

In the projected gradient descent (PGD) iterations of FROWN, we adopt early stop to improve efficiency: we stop the iteration if the relative improvement of the objective function is less than 1% for 5 consecutive steps. The certified targeted-attack bounds are average over multiple samples. All the samples and target labels are selected randomly. See Table 4 for the complete experiment result of the Sensorless Drive Diagnosis dataset. In this dataset, the input has been normalized by substracting the mean of all samples and then divided by the standard deviation of all the samples. We only compute l_∞ bounds for classifiers on the this dataset, because different elements of the input have different physical meanings in this dataset. It's meaningless to entangle them together like in the l_2 or l_1 norm.

See Tables 5 and 6 for the complete experiment result of the MNIST dataset. See Table 7 for the complete experiment result of the CIFAR 10 dataset. The input of these 2 datasets have been scaled to the range $[-1, 1]$.

Discussions. In some cases, the LP-based method gives even worse result than the CROWN, which should not happen in theory. We think the reason is twofold. First, the Gurobi solver doesn't always find the optimal solution. Sometimes it gives suboptimal solution. Since we specify to use the "dual simplex" method to solve the LP problem in Equations (4), (6) and (7), we can always guarantee that the solutions given by Gurobi are always valid bounds of the preactivations. Second, the bounds are averaged over randomly selected images. It may be caused by statistical fluctuations.

Table 4: (Experiment I) Averaged certified l_∞ bounds of different classifiers on Sensorless Drive Diagnosis dataset. Compared with CROWN, the improvements on certified bounds of FROWN and LP are computed respectively. The running time of FROWN and LP are measured by seconds. The bounds are averaged over 1000 samples.

Network	Certified Bounds			Improvement		Avg. Time per Image		Speed up
	CROWN	FROWN	LP	FROWN	LP	FROWN	LP	FROWN vs LP
$4 \times [20]$ ReLU	0.2019	0.2247	0.2269	11.27%	12.38%	0.35	0.96	2.8 X
$8 \times [20]$ ReLU	0.2094	0.2365	0.2526	12.95%	20.66%	1.81	4.19	2.3 X
$12 \times [20]$ ReLU	0.1996	0.2496	0.2740	25.05%	37.28%	4.39	9.46	2.2 X
$4 \times [20]$ Sigmoid	0.1019	0.1418	0.1388	39.08%	36.21%	0.74	2.11	2.8 X
$8 \times [20]$ Sigmoid	0.0858	0.1618	0.1626	88.62%	89.59%	4.15	32.15	7.7 X
$12 \times [20]$ Sigmoid	0.0782	0.1510	0.1081	93.06%	38.22%	8.71	152.91	17.6 X
$4 \times [20]$ Tanh	0.0975	0.1306	0.1361	34.04%	39.67%	0.55	2.05	3.7 X
$8 \times [20]$ Tanh	0.0754	0.1551	0.1343	105.77%	78.22%	3.35	30.33	9.1 X

Table 5: (Experiment I) Averaged certified l_p bounds of small size classifiers on MNIST dataset. "N/A" indicates no results can be obtained in the given runtime. The up arrow "↑" means "more than". The bounds of FROWN and CROWN are averaged over 1000 images, while those of LP are averaged over 100 images due to the long runnig time of LP.

Small Networks	p	Certified Bounds			Improvement		Avg. Time per Image		Speed up
		CROWN	FROWN	LP	FROWN	LP	FROWN	LP	FROWN vs LP
$5 \times [20]$ ReLU	1	3.8018	4.0835	4.2215	7.41%	11.04%	1.69	195.07	115.5 X
	2	0.5346	0.5710	0.5587	6.81%	4.52%	1.12	41.63	37.1 X
	∞	0.0261	0.0278	0.0287	6.52%	9.89%	1.26	11.93	9.5 X
$10 \times [20]$ ReLU	1	2.6866	3.1702	3.1071	18.00%	15.65%	5.15	571.21	110.9 X
	2	0.4036	0.4575	0.4615	13.35%	14.33%	6.59	100.80	15.3 X
	∞	0.0193	0.0223	0.0234	15.56%	21.47%	7.30	45.56	6.2 X
$15 \times [20]$ ReLU	1	2.1809	2.7292	2.8531	25.14%	30.82%	18.42	1438.25	78.1 X
	2	0.3491	0.4275	0.4220	22.46%	20.88%	15.35	235.32	15.3 X
	∞	0.0169	0.0210	0.0217	24.26%	28.27%	17.81	112.61	6.3 X
$20 \times [20]$ ReLU	1	2.3853	3.1062	3.1735	30.22%	33.04%	35.53	1301.11	36.6 X
	2	0.3656	0.4925	0.5074	34.71%	38.78%	31.10	229.87	7.4 X
	∞	0.0183	0.0240	0.0245	30.84%	33.75%	43.31	199.40	4.6 X
$5 \times [20]$ Sigmoid	1	1.8009	2.1340	2.1126	18.49%	17.31%	1.75	310.27	177.2 X
	2	0.3100	0.3581	0.3417	15.51%	10.20%	1.23	44.00	35.9 X
	∞	0.0153	0.0174	0.0170	13.94%	11.11%	1.39	17.19	12.4 X
$10 \times [20]$ Sigmoid	1	1.4980	1.9381	1.8209	29.38%	21.55%	5.09	649.97	127.8 X
	2	0.2652	0.3247	0.3062	22.47%	15.46%	6.94	239.19	34.5 X
	∞	0.0134	0.0164	0.0148	22.68%	10.39%	7.22	158.32	21.9 X
$15 \times [20]$ Sigmoid	1	1.5634	1.9365	1.9616	23.87%	25.47%	18.40	3188.63	173.3 X
	2	0.2685	0.3230	0.3279	20.31%	22.14%	16.58	401.33	24.2 X
	∞	0.0135	0.0162	0.0155	20.32%	15.28%	17.49	727.29	41.6 X
$20 \times [20]$ Sigmoid	1	1.5348	1.9779	1.9730	28.87%	28.55%	31.80	4904.99	154.3 X
	2	0.2524	0.3261	0.2657	29.21%	5.28%	31.77	1354.54	42.6 X
	∞	0.0131	0.0166	0.0153	27.01%	17.12%	44.86	1859.68	41.5 X
$5 \times [20]$ Tanh	1	2.3118	2.7823	2.5703	20.35%	11.18%	1.48	405.26	274.2 X
	2	0.3722	0.4360	0.4314	17.15%	15.91%	1.10	41.01	37.2 X
	∞	0.0181	0.0211	0.0207	16.68%	14.43%	1.22	16.78	13.8 X
$10 \times [20]$ Tanh	1	1.9444	2.5133	2.4822	29.26%	27.66%	5.00	1356.38	271.2 X
	2	0.3092	0.3833	0.3555	23.98%	14.98%	7.04	118.53	16.8 X
	∞	0.0153	0.0190	0.0178	23.63%	15.81%	7.15	131.57	18.4 X
$15 \times [20]$ Tanh	1	1.9776	2.5399	2.3660	28.43%	19.64%	20.76	2938.48	141.6 X
	2	0.3038	0.3923	0.3613	29.15%	18.96%	17.93	358.07	20.0 X
	∞	0.0150	0.0192	0.0185	28.14%	23.53%	20.00	808.17	40.4 X
$20 \times [20]$ Tanh	1	1.4078	1.8961	1.7902	34.68%	27.16%	38.56	4185.14	108.5 X
	2	0.2265	0.2996	0.2819	32.27%	24.47%	33.28	391.69	11.8 X
	∞	0.0113	0.0145	0.0149	28.46%	32.36%	44.07	1414.24	32.1 X

Table 6: (Experiment I) Averaged certified l_p bounds of large size classifiers on MNIST dataset. "N/A" indicates no results can be obtained in the given runtime. The up arrow "↑" means "more than". The bounds of FROWN and CROWN are averaged over 1000 images, while those of LP are averaged over 100 images due to the long runnig time of LP.

Large Networks	p	Certified Bounds			Improvement		Avg. Time per Image		Speed up
		CROWN	FROWN	LP	FROWN	LP	FROWN	LP	FROWN vs LP
$3 \times [100]$ ReLU	1	4.8414	5.2071	5.2775	7.55%	9.01%	2.99	532.82	178.1 X
	2	0.6966	0.7207	0.6901	3.46%	−0.94%	2.05	210.06	102.5 X
	∞	0.0335	0.0337	0.0333	0.37%	−0.78%	1.95	110.17	56.4 X
$5 \times [100]$ ReLU	1	3.8928	4.2343	N/A	8.77%	N/A	13.11	6000.00↑	457.8 X↑
	2	0.5499	0.5789	0.5934	5.28%	7.91%	10.90	826.79	75.9 X
	∞	0.0261	0.0276	0.0278	5.85%	6.57%	10.42	446.07	42.8 X
$7 \times [100]$ ReLU	1	3.5509	3.9907	N/A	12.38%	N/A	55.07	10000.00↑	181.6 X↑
	2	0.4997	0.5347	N/A	7.01%	N/A	49.49	10000.00↑	202.1 X↑
	∞	0.0241	0.0258	N/A	7.09%	N/A	44.94	10000.00↑	222.5 X↑
$3 \times [100]$ Sigmoid	1	2.7370	3.1822	3.1705	16.27%	15.84%	3.57	2364.50	662.2 X
	2	0.4320	0.4821	0.4465	11.59%	3.36%	2.56	469.55	183.6 X
	∞	0.0213	0.0234	0.0231	9.90%	8.21%	2.53	172.65	68.3 X
$5 \times [100]$ Sigmoid	1	2.0998	2.5184	N/A	19.93%	N/A	13.64	10000.00↑	733.0 X↑
	2	0.3233	0.3774	0.2494	16.74%	−22.84%	12.38	9368.58	757.1 X
	∞	0.0156	0.0186	0.0177	18.76%	13.10%	13.09	1319.25	100.8 X
$7 \times [100]$ Sigmoid	1	1.7294	2.2380	N/A	29.40%	N/A	48.96	10000.00↑	204.3 X↑
	2	0.2833	0.3427	N/A	20.96%	N/A	42.04	10000.00↑	237.9 X↑
	∞	0.0140	0.0168	N/A	19.96%	N/A	42.00	10000.00↑	238.1 X↑
$3 \times [100]$ Tanh	1	2.9399	3.5049	3.5369	19.22%	20.31%	3.25	2640.61	812.9 X
	2	0.4445	0.5131	0.5443	15.44%	22.44%	2.41	354.07	146.6 X
	∞	0.0215	0.0246	0.0253	14.79%	17.86%	2.40	169.86	70.7 X
$5 \times [100]$ Tanh	1	2.0709	2.5909	N/A	25.11%	N/A	13.29	15000.00↑	1128.5 X↑
	2	0.3059	0.3742	0.3711	22.31%	21.32%	12.31	4045.06	328.7 X
	∞	0.0149	0.0179	0.0176	20.05%	18.00%	12.92	1270.61	98.3 X
$7 \times [100]$ Tanh	1	1.8139	2.2823	N/A	25.82%	N/A	57.97	30000.00↑	517.5 X
	2	0.2667	0.3389	N/A	27.04%	N/A	46.19	10000.00↑	216.5 X
	∞	0.0132	0.0162	N/A	22.68%	N/A	46.91	10000.00↑	213.2 X

Table 7: (Experiment II) Averaged certified l_p bounds of different classifiers on CIFAR10 Networks. The bounds of FROWN and CROWN are averaged over 500 images. Compared with CROWN, the improvements on certified bounds of FROWN are computed. To speed up FROWN, instead of performing optimization for every neuron in a layer, we optimize the neurons group by group. For 4-layer and 6-layer networks, every 8 neurons in a layer are grouped together. For 8-layer, every 64 neurons in a layer are grouped together.

Network	p	Certified Bounds		Improvement	Network	p	Certified Bounds		Improvement
		CROWN	FROWN				CROWN	FROWN	
$4 \times [2048]$ ReLU	1	7.5460	7.5989	0.70%	$4 \times [2048]$ Sigmoid	1	3.6558	4.4726	22.34%
	2	0.6175	0.6303	2.09%		2	0.2847	0.3353	17.77%
	∞	0.0151	0.0157	4.14%		∞	0.0069	0.0082	18.33%
$6 \times [2048]$ ReLU	1	4.5990	4.7241	2.72%	$6 \times [2048]$ Sigmoid	1	1.5093	1.8430	22.11%
	2	0.3745	0.3723	−0.60%		2	0.1180	0.1448	22.71%
	∞	0.0092	0.0093	0.95%		∞	0.0029	0.0035	21.03%
$8 \times [2048]$ ReLU	1	4.2349	4.9416	16.69%	$8 \times [2048]$ Sigmoid	1	1.2875	1.6862	30.97%
	2	0.3476	0.3809	9.59%		2	0.1006	0.1295	28.65%
	∞	0.0087	0.0094	8.83%		∞	0.0024	0.0033	37.65%

A.10 Improve efficiency of FROWN and balance trade-off between tightness of bounds and time cost

FROWN can be speeded up by optimizing neurons in a layer group by group, instead of one by one: Considering the objective functions in Problems (9) and (10) in the main text, every neuron (including neurons in the intermediate layers) has such an objective function and need to be optimized one by one. To speed up this process, we can maximize (or minimize) the mean of the objective functions of a group of neurons in a layer at a time. The more neurons that we optimize at a time, the more speed up we will gain, and as expected, the looser bound we will obtain. Our code also enables GPU usage to further speed up FROWN. We conduct experiments on a $4 \times [1024]$ ReLU network on MNIST dataset to demonstrate the trade-off between tightness of bounds and efficiency. The result is shown in Table 8. We can see GPU usage alone provides about 4X speed up. Performing optimization in groups can decrease the time down to 1.41s, and this trick nearly doesn't hurt tightness of the bounds. In conclusion, FROWN has a parameter (group size) to balance the trade-off between tightness of bounds and efficiency. Researchers can determine this parameter based on their computational power and available time when applying FROWN in a specific task.

Table 8: Averaged certified l_∞ bounds of a $4 \times [1024]$ ReLU network on MNIST dataset. CPU refers to a single Intel Xeon E5-2640 v3 (2.60GHz) CPU and GPU refers to an NVIDIA GeForce GTX TITAN X GPU.

FROWN			
Device	Group size	Certified Bounds ($\times 10^{-2}$)	Avg. Time per Image (s)
CPU	1	2.942	89.90
GPU	1	2.937	23.44
GPU	8	2.937	3.53
GPU	16	2.898	1.96
GPU	32	2.897	1.41