9 APPENDIX

Define $[n_k^+] = \{i | w_i^k > 0\}$ and $[n_k^-] = \{i | w_i^k < 0\}$

Proof:neuron-wise tighest is equivalent to layer-wise tighest. We define linear bounds are layer-wise tightest is when $\iint_{x \in \times_{q \in [n_{k-1}]}[l_q^{k-1}, u_q^{k-1}]} (\sum_{i \in [n_k^+]} W_{ji}^{+,k+1} u_i^k(x) - \sum_{i \in [n_k^-]} W_{ji}^{-,k+1} l_i^k(x)) dx$ is minimal.

As $l_q^{k-1}, u_q^{k-1}, W_i^{+,k+1}, W_i^{-,k+1}$ are constant and $[n_k^+] \cap [n_k^-] = \emptyset$, thus we need to minimize $\iint_{x \in \times_{q \in [n_{k-1}]}[l_q^{k-1}, u_q^{k-1}]} u_i^k(x) dx, i \in [n_k^+]$ and $\iint_{x \in \times_{q \in [n_{k-1}]}[l_q^{k-1}, u_q^{k-1}]} (-l_i^k(x)) dx, i \in [n_k^-]$ respectively. □

Proof:Neuron-wise Tightest Linear Bounds. Because $u_i^k(x)$ is a linear combination of x. Without loss of generality, we assume $u_i^k(x) = u_i^k(x_1, \cdots, x_{n_{k-1}}) = \sum_{q \in [n_{k-1}]} a_{u,iq}^k x_q + b_{u,i}^k$. Then,

$$
\begin{aligned}
&\iint_{x \in \times_{q \in [n_{k-1}]}[l_q^{k-1}, u_q^{k-1}]} u_i^k(x) dx \\
=&\iint_{x \in \times_{q \in [n_{k-1}]}[l_q^{k-1}, u_q^{k-1}]} (\sum_{q \in [n_{k-1}]} a_{u,iq}^k x_q + b_{u,i}^k) dx \\
=&\sum_{q \in [n_{k-1}]} \frac{a_{u,iq}^k}{2} ((u_q^{k-1})^2 - (l_q^{k-1})^2) + b_{u,i}^k (u_q^{k-1} - l_q^{k-1}) \\
=&(u_q^{k-1} - l_q^{k-1}) (\sum_{q \in [n_{k-1}]} a_{u,iq}^k \frac{u_q^{k-1} + l_q^{k-1}}{2} + b_{u,i}^k) \\
=&(u_q^{k-1} - l_q^{k-1}) u_i^k(m)
\end{aligned}
$$

where $m = (\frac{u_1^{k-1} + l_1^{k-1}}{2}, \cdots, \frac{u_{n_{k-1}}^{k-1} + l_{n_{k-1}}^{k-1}}{2})$, because $u_q^{k-1}, l_q^{k-1}, q \in [n_{k-1}]$ are constant, and the minimize target has been transformed into minimizing $u_i^k(m), i \in [n_k^+]$.

Symmetric to the proof of minimizing $\iint_{x \in \times_{q \in [n_{k-1}]}[l_q^{k-1}, u_q^{k-1}]} u_i^k(x) dx$, minimizing $-\iint_{x \in \times_{q \in [n_{k-1}]}[l_q^{k-1}, u_q^{k-1}]} (l_i^k(x)) dx, i \in [n_k^-]$ is equivalent to minimize $-l_i^k(m), i \in [n_k^-]$.

And symmetric to the above proof, minimizing $-\iint_{x \in \times_{q \in [n_{k-1}]}[l_q^{k-1}, u_q^{k-1}]} (\sum_{i \in [n_k^+]} W_{ji}^{+,k+1} l_i^k(x) - \sum_{i \in [n_k^-]} W_{ji}^{-,k+1} u_i^k(x)) dx$, is equivalent to $-l_i^k(m), i \in [n_k^+]$,$u_i^k(m), i \in [n_k^-]$.

This completes the proof. □

Here, we provide a way to find neuron-wise tightest linear bounds of non-linear function $f(x)$.

Theorem 9.1 (Finding Neuron-wise Tightest Linear Bounds). *Given $f(x)$ is a continuous function, then*
(1) if $f'(m)(x - m) + f(m)$ is an upper(lower) bound, then $u(m)(-l(m))$ reach its minimum.
(2)considering two points $(d_1, f(d_1)), (d_2, f(d_2))$ where $d_1, d_2 \in \times_{i=1}^n [l_i, u_i]$. When $m = \lambda d_1 + (1 - \lambda) d_2, \lambda \in [0, 1]$, if the upper(lower) bounding line/plane passes through $(d_1, f(d_1)), (d_2, f(d_2))$, then $u(m)(-l(m))$ has reached its minimum.

In essence, the theorem "Finding neuron-wise Tightest Linear Bounds" shows that if the tangent line/plane at m is the upper/lower bound of $f(x)$, then the tangent line/plane at $x = m$ is obviously the neuron-wise tightest linear upper/lower bound of $f(x)$. Otherwise, we try to find two points that satisfy $m = \lambda d_1 + (1 - \lambda) d_2, \lambda \in [0, 1]$, so that if a line/plane passing through $(d_1, f(d_1)), (d_2, f(d_2))$ is the linear bound of $f(x)$, then this line/plane is the neuron-wise tightest bounding constraints.

Proof:Finding neuron-wise Tightest Linear Bounds. Obviously, (1) is right.

And we prove (2) by contradiction. Without loss of generality, we suppose the line/plane to be proved is an upper bound $u(x)$ and $u(d_1) = f(d_1), u(d_2) = f(d_2), m = \lambda d_1 + (1 - \lambda) d_2, \lambda \in [0, 1]$. We assume there is another bounding line/plane $u'(x)$ and $u'(m) < u(m)$. Because $u(x), u'(x)$ are linear combinations of x and $m = \lambda d_1 + (1 - \lambda) d_2, \lambda \in [0, 1]$,then $u(m) = \lambda u(d_1) + (1 - \lambda) u(d_2), u'(m) = \lambda u'(d_1) + (1 - \lambda) u'(d_2), \lambda \in [0, 1]$. Therefore, $\lambda u'(d_1) + (1 - \lambda) u'(d_2) < \lambda u(d_1) + (1 - \lambda) u(d_2) = \lambda f(d_1) + (1 - \lambda) f(d_2)$ which contradicts $u'(d_1) \geq f(d_1), u'(d_2) \geq f(d_2)$.

This completes the proof. □

Proof:Sigmoid/Tanh/Arctan Linear Bounds. $m = F^k(x_0) \in [l, u] \in \mathbb{R}$, and if $d_1 \leq m \leq d_2$, then there exists $\lambda = \frac{d_2 - m}{d_2 - d_1}$, s.t. $\lambda d_1 + (1 - \lambda) d_2 = m$

case 1: $m \geq 0$

upper bound:

(1) when $k_{ml} > f'(m)$:

As $m \geq 0$, we have $u \geq 0$, and when $x \geq 0$, $f(x)$ is concave. Therefore, $f(x) \leq f'(m)(x-m) + f(m), x \in [m,u]$.

Considering $x < 0$, because $k_{ml} > f'(m)$, we have $f'(m)(l-m) + f(m) > k_{ml}(l-m) + f(m) = f(l)$ and $f'(m)(l-m) + f(m) \geq f(0)$. Thus, as $f(x), x < 0$ is convex, then $f'(m)(l-m) + f(m) \geq f(x)$

Thus, $f'(m)(x-m) + f(m)$ is the upper bounding line and $u(m)$ is minimum obviously, by theorem 9.1(1).

(2) when $k_{ml} < f'(m)$ and $k > f'(u)$:

$k > f'(u)$, means the tangent point x^* exists. Because $f(x)$ is continuous, $f'(0) > \frac{f(0)-f(l)}{0-l}$ and $f'(u) < \frac{f(u)-f(l)}{u-l}$, thus there exists a solution $x^* \in [0,u]$ for equation $f'(x^*) = \frac{f(x^*)-f(l)}{x^*-l}$.

$k_{ml} < f'(m)$, means $m \in [l, x^*]$. Because $k_{ml} < f'(m)$ and $f'(u) < \frac{f(u)-f(l)}{u-l}$, thus $m < x^* < u$. Because $f(x), x \geq 0$ is concave, $f(x) \leq f'(x^*)(x-x^*) + f(x^*), x \geq 0$. Because $f'(x^*)(0-x^*) + f(x^*) \geq f(0)$, $f'(x^*)(l-x^*) + f(x^*) = f(l)$ and $f(x), x < 0$ is convex, thus $f'(x^*)(x-x^*) + f(x^*) \geq f(x), x < 0$

Therefore, $f'(x^*)(x-x^*) + f(x^*)$ is the upper bound, and as $l \leq m < x^*$, by theorem 9.1(2), $u(m)$ reaches minimum.

(3) when $k_{ml} < f'(m)$ and $k \leq f'(u)$:

$k \leq f'(u)$, which means $k(x-l) + f(l) \geq f(x)$. Because when $x > 0$, we have $k(x-l) + f(l) = k(x-u) + f(u) > f'(u)(x-u) + f(u)$ and $f(x), x > 0$ is concave, thus $k(x-l) + f(l) \geq f(x), x > 0$. When $x \leq 0$, we have $k \leq f'(u) < \frac{f(0)-f(u)}{0-u}$, thus $k(0-u) + f(u) > \frac{f(0)-f(u)}{0-u}(0-u) + f(u) = f(0)$. As $f(x)$ is convex when $x \leq 0$. Thus $k(x-l) + f(l)$ is the upper bound and as $m \in [l,u]$, $u(m)$ reaches minimum by theorem 9.1(2).

lower bound:

(4) when $k \geq f'(l)$,

As proved above, $k \geq f'(l)$ means the tangent point x^{**} exists and the $f'(x^{**})(x-u) + f(u)$ is the lower bound. As $u \geq m > 0 > x^**$, $l(m)$ reaches its minimum by theorem 9.1(2).

(5) when $k < f'(l)$,

As proved above, $k(x-l) + f(l)$ is the lower bound and as $l \leq m \leq u$, $l(m)$ achieves it minimum by theorem 9.1(2).

case 2: $m < 0$ The proof of case 2 is symmetric to the proof of case 1, for the reason that $f(x)$ is centrosymmetry about the origin.

□

Similar to proof " Sigmoid/Tanh/Arctan Linear Bounds" , when $m = \frac{u+l}{2}$, the linear bounds of theorem 3.1 is neuron-wise tightest according to theorem 9.1 and theorem 3.2. Although the neuron-wise tightest technique[9] can compute a larger robustness bounds than other state-of-the-art works and stands for the highest precision among related work, experimental results have shown that Ti-Lin is better than the neuron-wise tightest linear bounds of Sigmoid/Tanh/Arctan function.

Proof:Maxpool Linear Bounds. $m = (m_1, \cdots, m_n) \in \mathbb{R}^n$.

upper bound:

case 1: When $u_i \geq l_i \geq \cdots, f(x_1, \cdots, x_n) = x_i, \forall (x_1, \cdots, x_n)$, then we set $u(x_1, \cdots, x_n) = x_i$, then $u(m) = f(m)$ which is the neuron-wise tightest upper bounding plane by theorem 9.1(1).

case 2:When $u_i \geq u_j \geq l_j \geq \cdots, f(x_1, \cdots, x_n) = max(x_i, x_j), \forall (x_1, \cdots, x_n)$.

If $f(x_1, \cdots, x_n) = x_i$,

$$
\begin{aligned}
u(x_1, \cdots, x_n) - x_i &= \frac{u_i - l_j}{u_i - l_i}(x_i - l_i) + (x_j - l_j) + l_j - x_i \\
&= \frac{u_i - l_j}{u_i - l_i}(x_i - l_i) + (x_j - l_j) + l_j - l_i + l_i - x_i \\
&= (x_i - l_i)(\frac{u_i - l_j}{u_i - l_i} - 1) + (x_j - l_j) + l_j - l_i \\
&= (x_i - l_i)\frac{l_i - l_j}{u_i - l_i} + (x_j - l_j) + l_j - l_i \\
&= (l_j - l_i)(1 - \frac{x_i - l_i}{u_i - l_i}) + (x_j - l_j) \\
&= (l_j - l_i)\frac{u_i - x_i}{u_i - l_i} + (x_j - l_j) \\
&\geq 0
\end{aligned}
$$

If $f(x_1, \cdots, x_n) = x_j$,

$$\begin{aligned} u(x_1, \cdots, x_n) - x_j &= \frac{u_i - l_j}{u_i - l_i}(x_i - l_i) + (x_j - l_j) + l_j - x_j \\ &= \frac{u_i - l_j}{u_i - l_i}(x_i - l_i) \\ &\geq 0 \end{aligned}$$

Because

$$\begin{aligned} u(u_1, \cdots, u_{i-1}, u_i, u_{i+1}, \cdots, u_{j-1}, l_j, u_{j+1}, \cdots, u_n) &= \frac{u_i - l_j}{u_i - l_i}(u_i - l_i) + (l_j - l_j) + l_j \\ &= u_i - l_j + l_j \\ &= f(u_1, \cdots, u_{i-1}, u_i, u_{i+1}, \cdots, u_{j-1}, l_j, u_{j+1}, \cdots, u_n) \end{aligned}$$

and

$$\begin{aligned} u(l_1, \cdots, l_{i-1}, l_i, l_{i+1}, \cdots, l_{j-1}, u_j, l_{j+1}, \cdots, l_n) &= \frac{u_i - l_j}{u_i - l_i}(l_i - l_i) + (u_j - l_j) + l_j \\ &= u_j - l_j + l_j \\ &= f(l_1, \cdots, l_{i-1}, l_i, l_{i+1}, \cdots, l_{j-1}, u_j, l_{j+1}, \cdots, l_n) \end{aligned}$$

we notice that $(u_1, \cdots, u_{i-1}, u_i, u_{i+1}, \cdots, u_{j-1}, l_j, u_{j+1}, \cdots, u_n)$ and $(l_1, \cdots, l_{i-1}, l_i, l_{i+1}, \cdots, l_{j-1}, u_j, l_{j+1}, \cdots, l_n)$ are the space diagonal of $\times_{i=1}^{n}[l_i, u_i]$, and $m = \frac{1}{2}(u_1, \cdots, u_{i-1}, u_i, u_{i+1}, \cdots, u_{j-1}, l_j, u_{j+1}, \cdots, u_n) + \frac{1}{2}(l_1, \cdots, l_{i-1}, l_i, l_{i+1}, \cdots, l_{j-1}, u_j, l_{j+1}, \cdots, l_n)$, thus the plane is the neuron-wise tightest linear upper plane by theorem 9.1(2).

case 3:When $u_i \geq u_j \geq u_k \geq \cdots$, $f(x_1, \cdots, x_n) = max(x_1, \cdots, x_n), \forall (x_1, \cdots, x_n)$. If $f(x_1, \cdots, x_n) = x_i$,

$$\begin{aligned} u(x_1, \cdots, x_n) - x_i &= \frac{u_i - u_k}{u_i - l_i}(x_i - l_i) + \frac{u_j - u_k}{u_j - l_j}(x_j - l_j) + u_k - x_i \\ &= \frac{u_i - u_k}{u_i - u_k}(x_i - l_i) + \frac{u_j - u_k}{u_j - l_j}(x_j - l_j) + u_k - l_i + l_i - x_i \\ &= (x_i - l_i)(\frac{u_i - u_k}{u_i - l_i} - 1) + \frac{u_j - u_k}{u_j - l_j}(x_j - l_j) + u_k - l_i \\ &= (x_i - l_i)\frac{l_i - u_k}{u_i - l_i} + \frac{u_j - u_k}{u_j - l_j}(x_j - l_j) + u_k - l_i \\ &= (u_k - l_i)(1 - \frac{x_i - l_i}{u_i - l_i}) + \frac{u_j - u_k}{u_j - l_j}(x_j - l_j) \\ &= (u_k - l_i)\frac{u_i - x_i}{u_i - l_i} + \frac{u_j - u_k}{u_j - l_j}(x_j - l_j) \\ &\geq 0 \end{aligned}$$

If $f(x_1, \cdots, x_n) = x_j$, the proof is the same as above.

$$\begin{aligned} u(x_1, \cdots, x_n) - x_j &= \frac{u_i - u_k}{u_i - l_i}(x_i - l_i) + \frac{u_j - u_k}{u_j - l_j}(x_j - l_j) + u_k - x_j \\ &= \frac{u_i - u_k}{u_i - u_k}(x_i - l_i) + \frac{u_j - u_k}{u_j - l_j}(x_j - l_j) + u_k - l_j + l_j - x_j \\ &= \frac{u_i - u_k}{u_i - l_i}(x_i - l_i) + (\frac{u_j - u_k}{u_j - l_j} - 1)(x_j - l_j) + u_k - l_j \\ &= \frac{u_i - u_k}{u_i - l_i}(x_i - l_i) + \frac{l_j - u_k}{u_j - l_j}(x_j - l_j) + u_k - l_i \\ &= \frac{u_i - u_k}{u_i - l_i}(x_i - l_i) + (1 - \frac{x_j - l_j}{u_j - l_j})(u_k - l_j) \\ &= \frac{u_i - u_k}{u_i - l_i}(x_i - l_i) + \frac{u_j - x_j}{u_j - l_j}(u_k - l_j) \\ &\geq 0 \end{aligned}$$

If $f(x_1, \cdots, x_n) = x_k$,

$$\begin{aligned} u(x_1, \cdots, x_n) - x_j &= \frac{u_i - u_k}{u_i - l_i}(x_i - l_i) + \frac{u_j - u_k}{u_j - l_j}(x_j - l_j) + (u_k - x_k) \\ &\geq 0 \end{aligned}$$

If $f(x_1, \cdots, x_n) = x_l, l \neq i, j, k$,

$$\begin{aligned} u(x_1, \cdots, x_n) - x_l &= \frac{u_i - u_k}{u_i - l_i}(x_i - l_i) + \frac{u_j - u_k}{u_j - l_j}(x_j - l_j) + u_k - x_l \\ &= \frac{u_i - u_k}{u_i - l_i}(x_i - l_i) + \frac{u_j - u_k}{u_j - l_j}(x_j - l_j) + (u_k - u_l) + (u_l - x_l) \\ &\geq 0 \end{aligned}$$

Because

$$\begin{aligned} u(u_1, \cdots, u_{i-1}, u_i, u_{i+1}, \cdots, u_{j-1}, l_j, u_{j+1}, \cdots, u_n) &= \frac{u_i - u_k}{u_i - l_i}(u_i - l_i) + \frac{u_j - u_k}{u_j - l_j}(l_j - l_j) + u_k \\ &= u_i - u_k + u_k \\ &= f(u_1, \cdots, u_{i-1}, u_i, u_{i+1}, \cdots, u_{j-1}, l_j, u_{j+1}, \cdots, u_n) \end{aligned}$$

and

$$\begin{aligned} u(l_1, \cdots, l_{i-1}, l_i, l_{i+1}, \cdots, l_{j-1}, u_j, l_{j+1}, \cdots, l_n) &= \frac{u_i - u_k}{u_i - l_i}(l_i - l_i) + \frac{u_j - u_k}{u_j - l_j}(u_j - l_j) + u_k \\ &= u_j - u_k + u_k \\ &= f(l_1, \cdots, l_{i-1}, l_i, l_{i+1}, \cdots, l_{j-1}, u_j, l_{j+1}, \cdots, l_n) \end{aligned}$$

$m = \frac{1}{2}(u_1, \cdots, u_{i-1}, u_i, u_{i+1}, \cdots, u_{j-1}, l_j, u_{j+1}, \cdots, u_n) + \frac{1}{2}(l_1, \cdots, l_{i-1}, l_i, l_{i+1}, \cdots, l_{j-1}, u_j, l_{j+1}, \cdots, l_n)$, by theorem 9.1(2), the upper bound is the neuron-wise tightest bounding plane.

lower bound:

$l(x_1, \cdots, x_n) = x_j = argmax_i m_i$, and $\forall (x_1, \cdots, x_n) \in \times_{i=1}^{n} [l_i, u_i]$,

$$\begin{aligned} f(x_1, \cdots, x_n) &= max(x_1, \cdots, x_n) \\ &\geq x_j \\ &= l(x_1, \cdots, x_n) \end{aligned}$$

Furthermore, $l(m) = f(m_1, \cdots, m_n)$, hence, $l(x_1, \cdots, x_n)$ is the neuron-wise tightest lower bounding plane by theorem 9.1(1).

This completes the proof. □