Ingeniería Industrial
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

Tightening Convex Relaxations of Trained Neural Networks

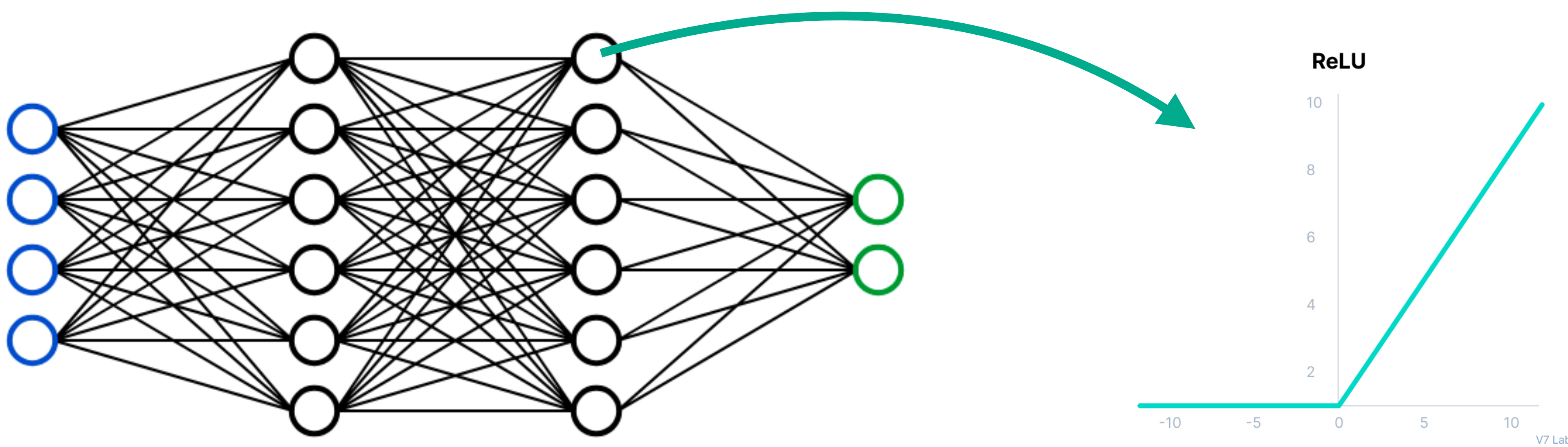
A unified approach for convex and S-shaped activations

Pablo Carrasco (Universidad de O’Higgins)
Gonzalo Muñoz (Universidad de Chile)

Mixed-Integer Programming Workshop 2025

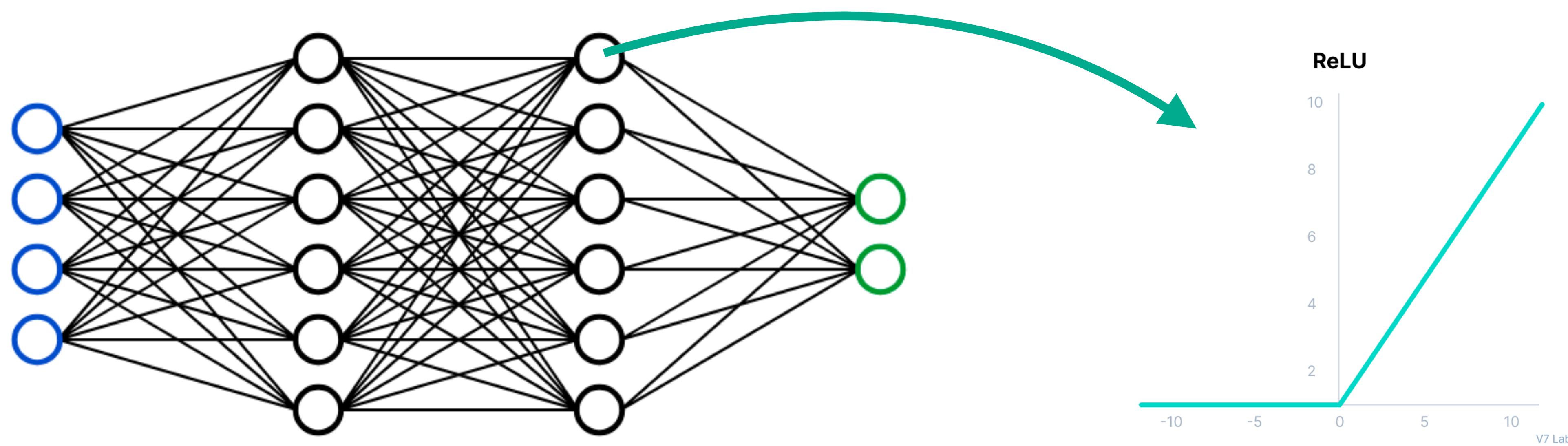
Optimization using trained neural networks

High level goal: to convexity $y = \mathrm{NN}(x)$



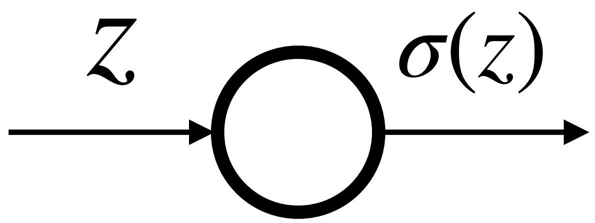
Optimization using trained neural networks

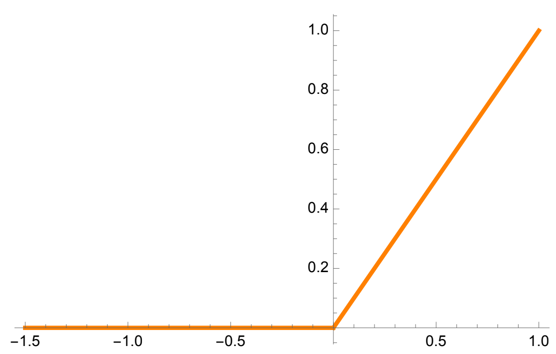
High level goal: to convexity $y = \mathrm{NN}(x)$

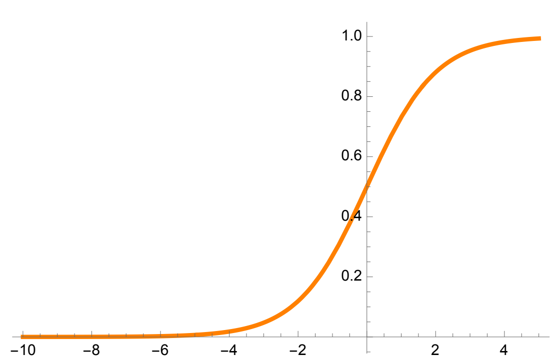


This can be useful in either using or evaluating a trained neural network via an optimization model

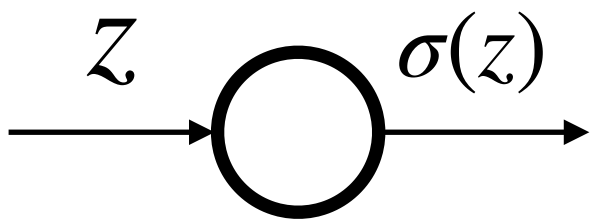
Single neuron convexifications



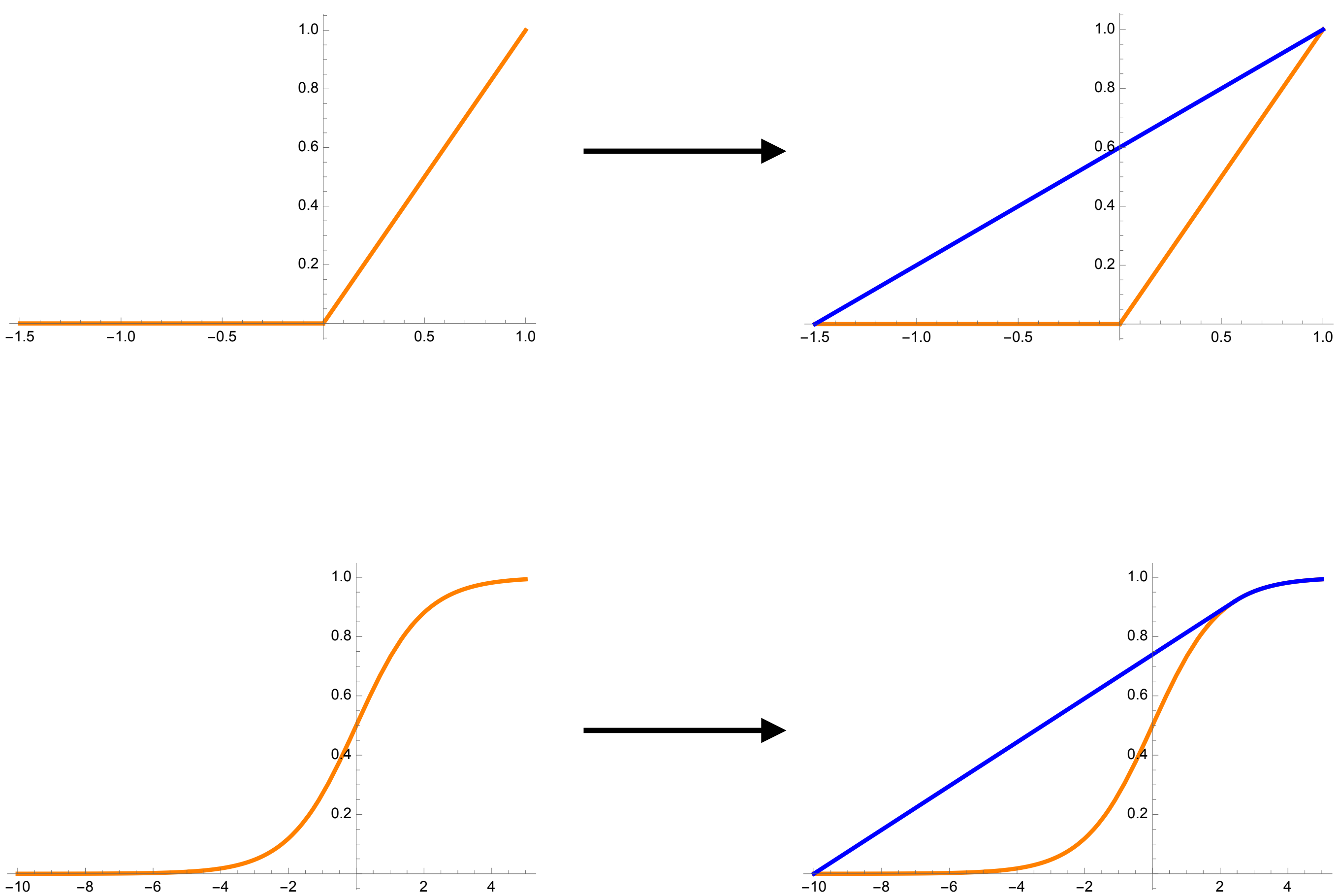




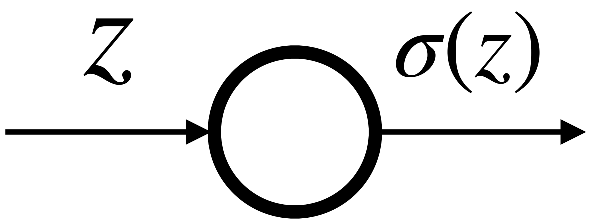
Single neuron convexifications



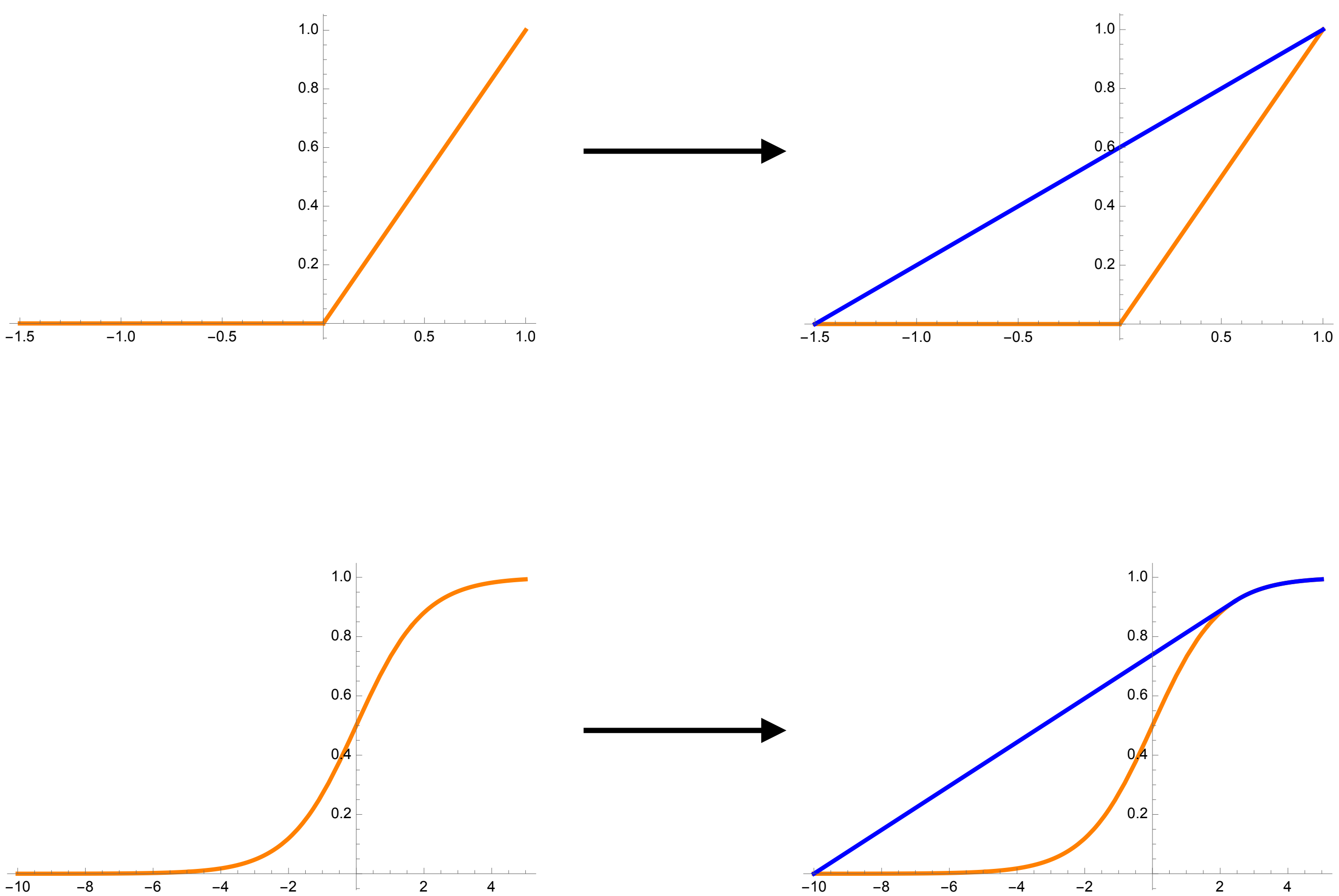
Function envelopes



Single neuron convexifications

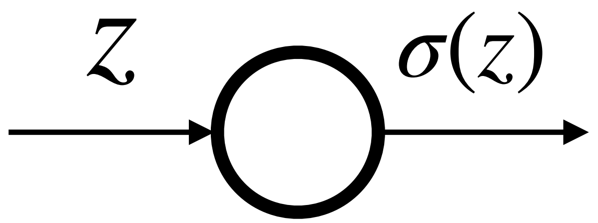


Function envelopes

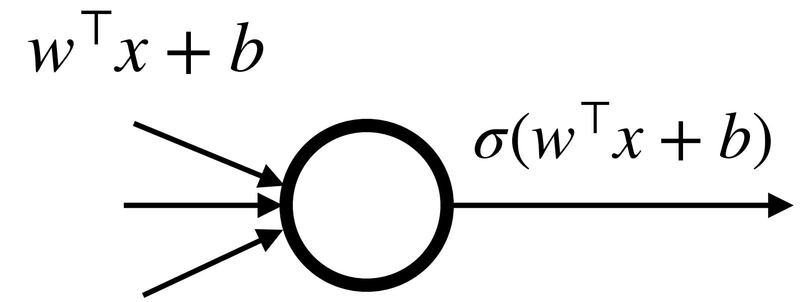


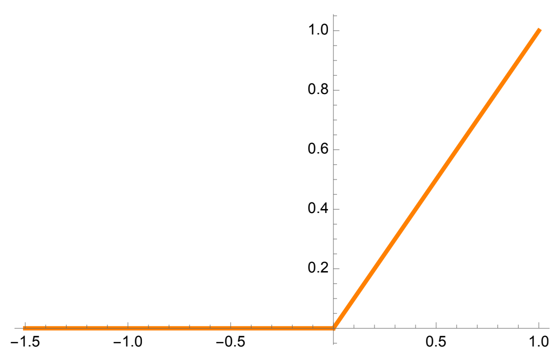
(Locally)

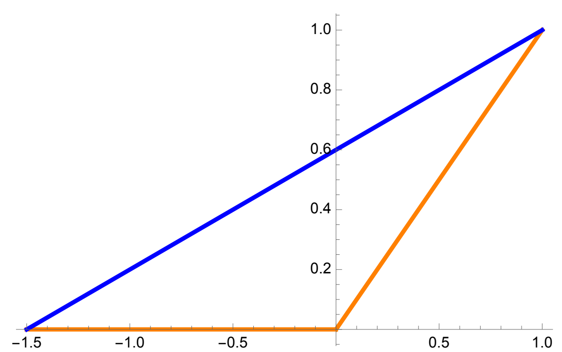
Single neuron convexifications

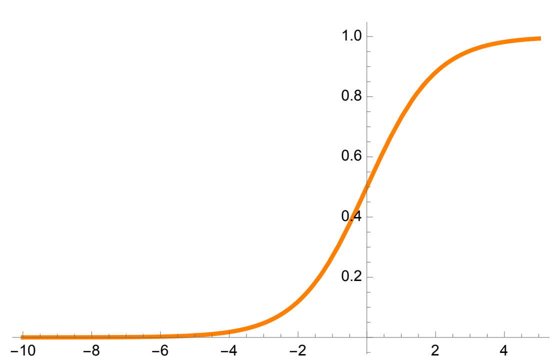


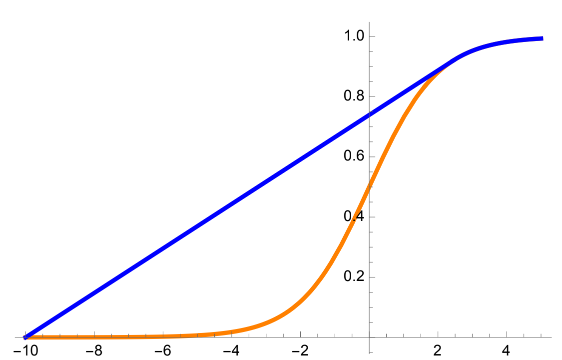
Function envelopes





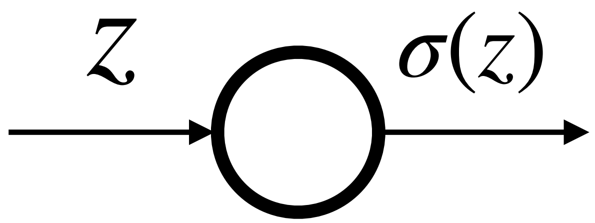




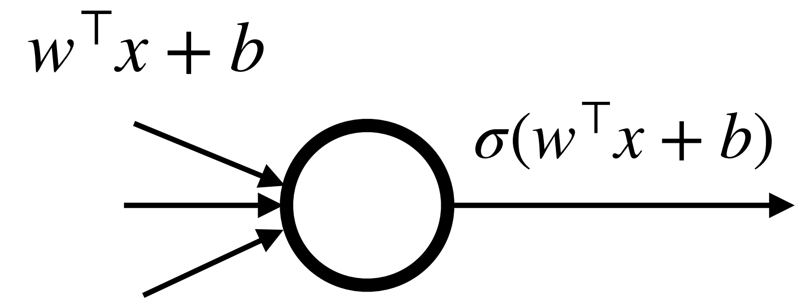


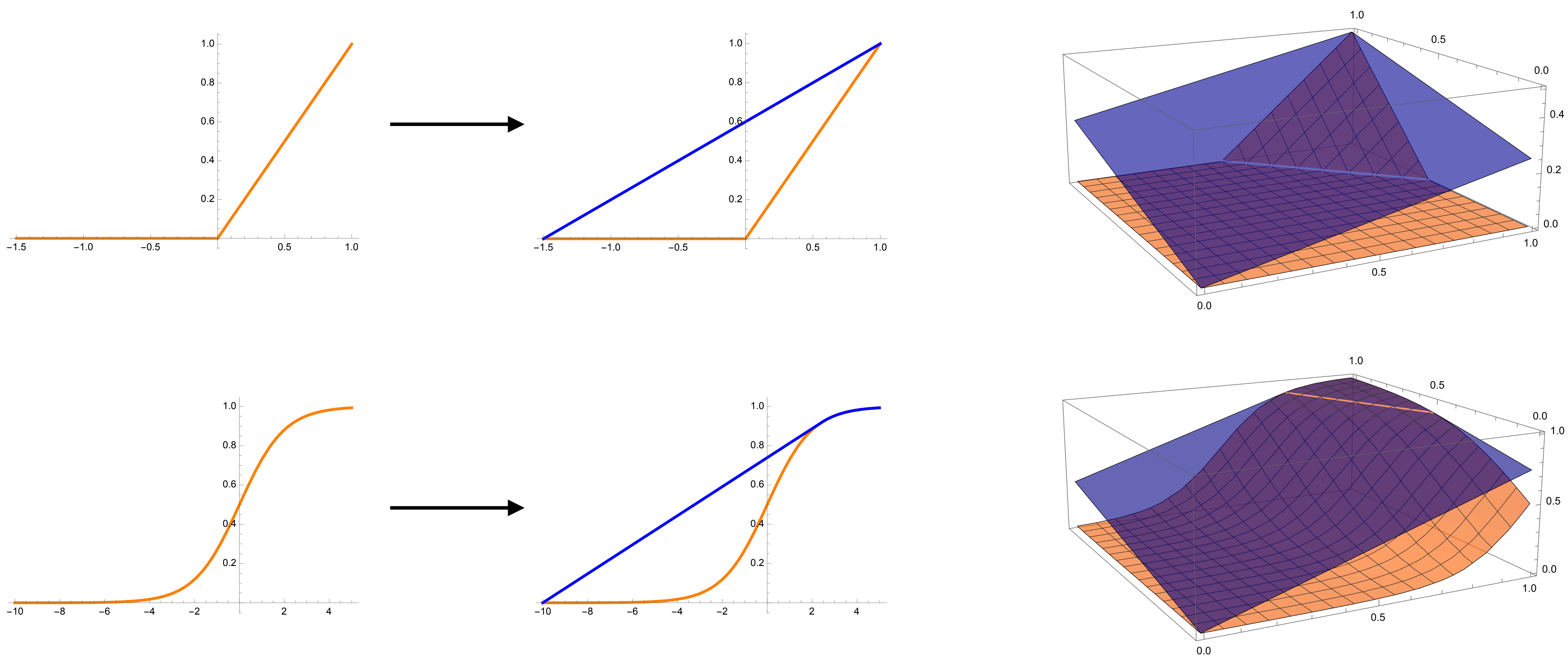
(Locally)

Single neuron convexifications



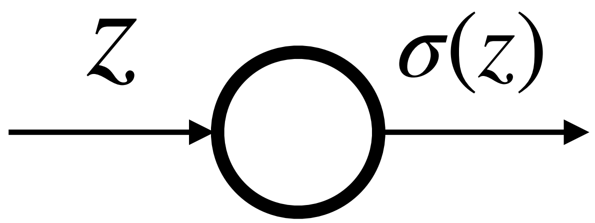
Function envelopes



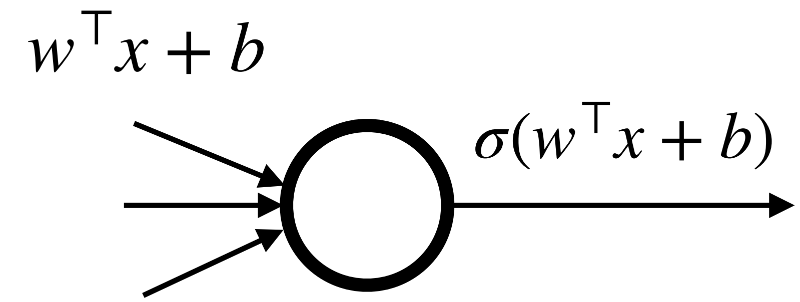


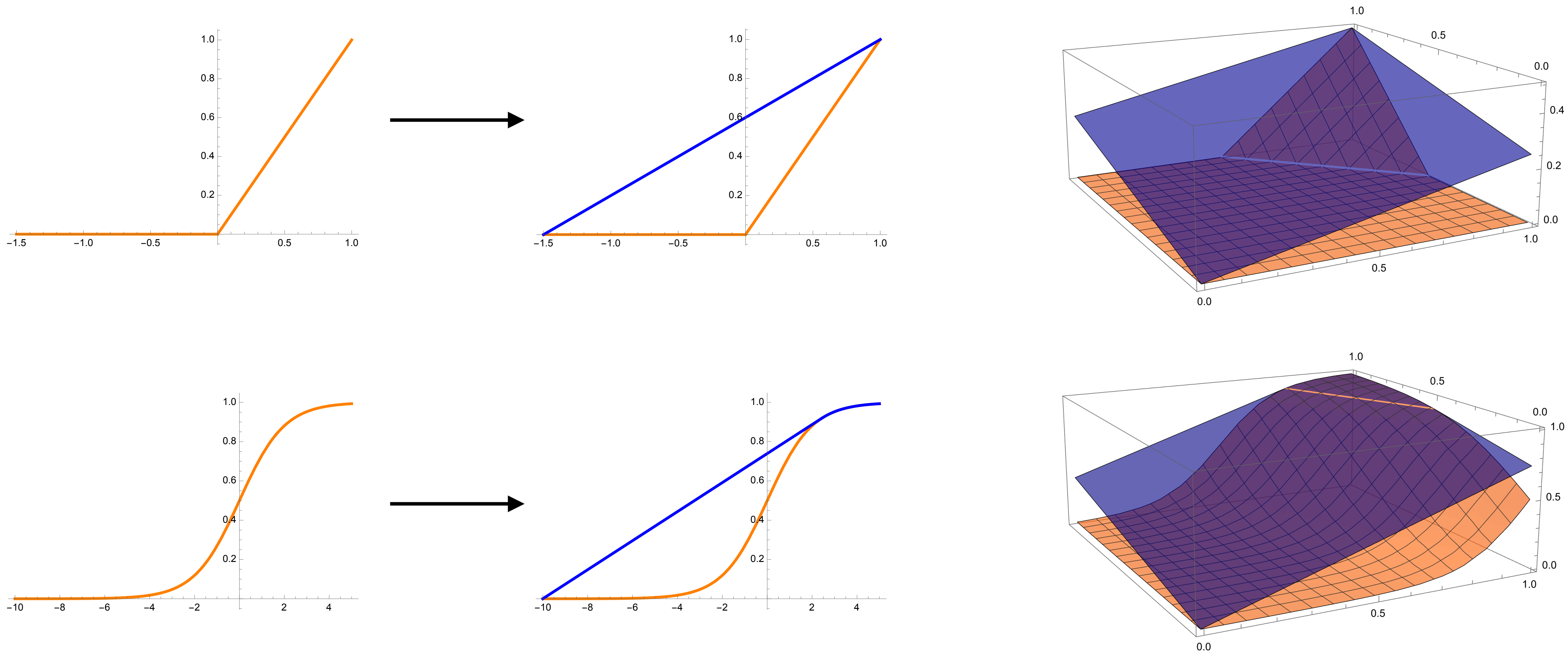
(Locally)

Single neuron convexifications



Function envelopes



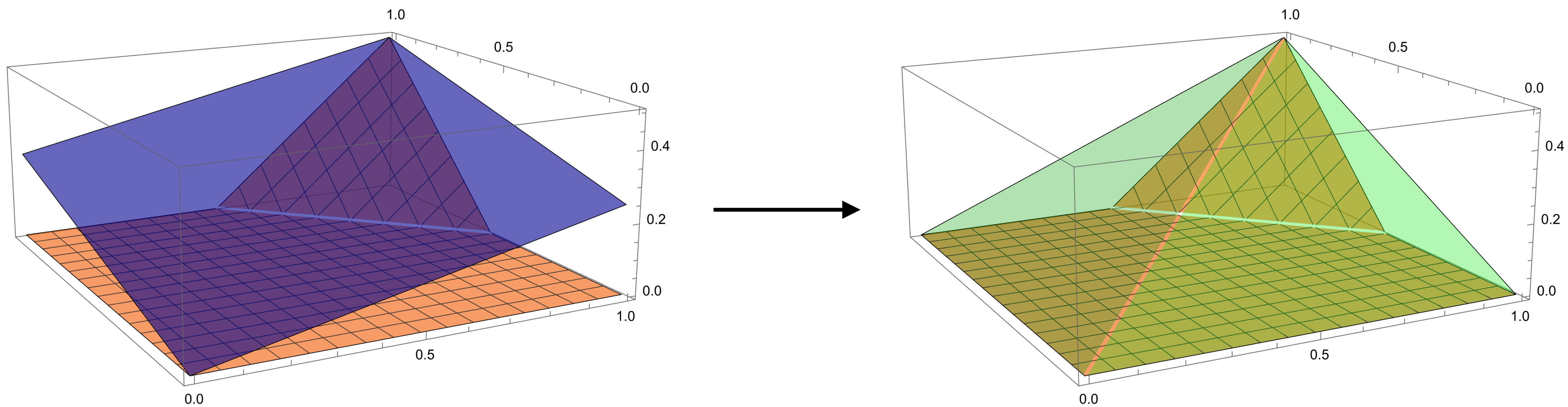


(Locally)

A path forward for the ReLU

Anderson et al. (2020) and Tjandraatmadja et al. (2020) showed how to tightly convexify

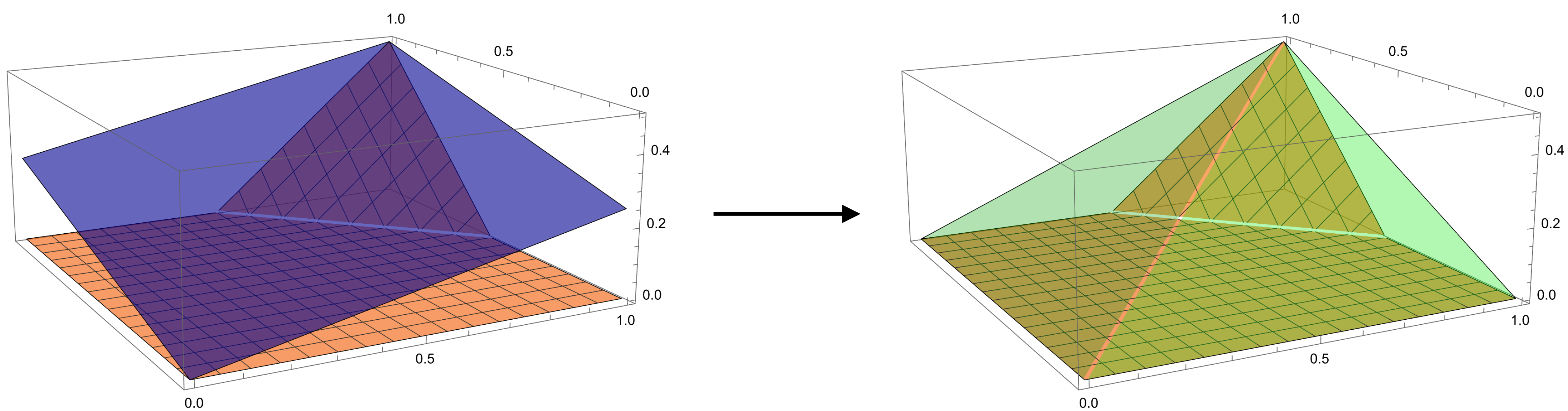
$$y = \sigma(w^\top x + b)$$



A path forward for the ReLU

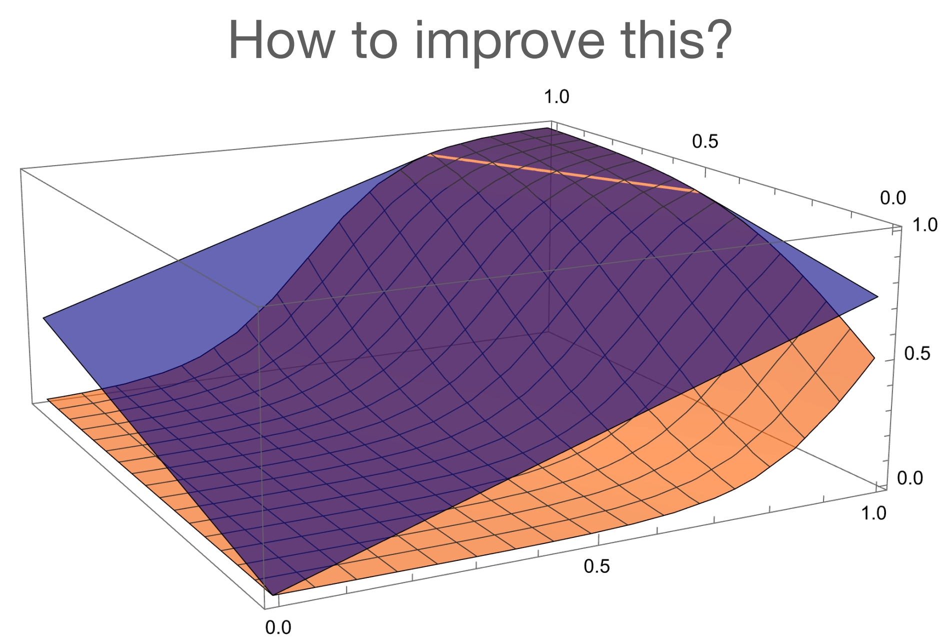
Anderson et al. (2020) and Tjandraatmadja et al. (2020) showed how to tightly convexify

$$y = \sigma(w^\top x + b)$$



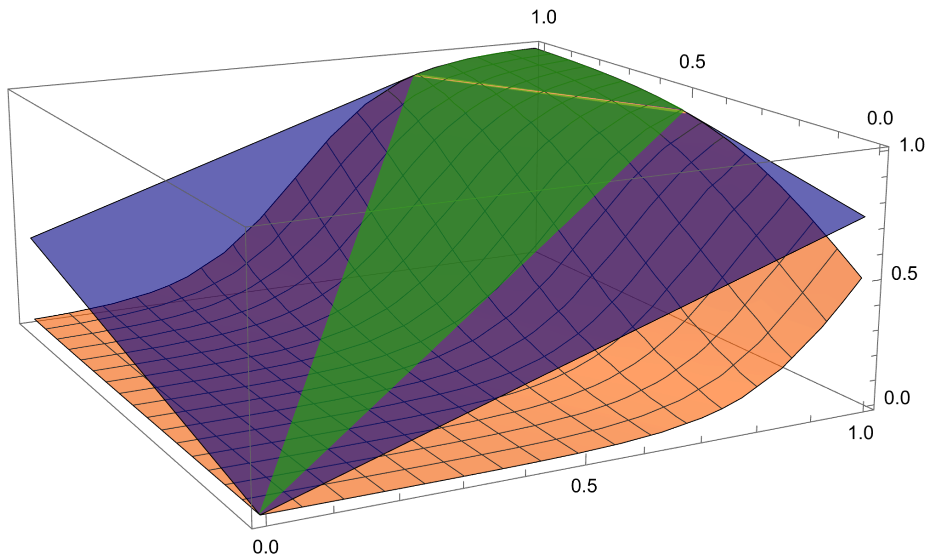
Their approach can handle polyhedral domains, piecewise linear convex activations, and they develop a polynomial time separation

Beyond the ReLU



Beyond the ReLU

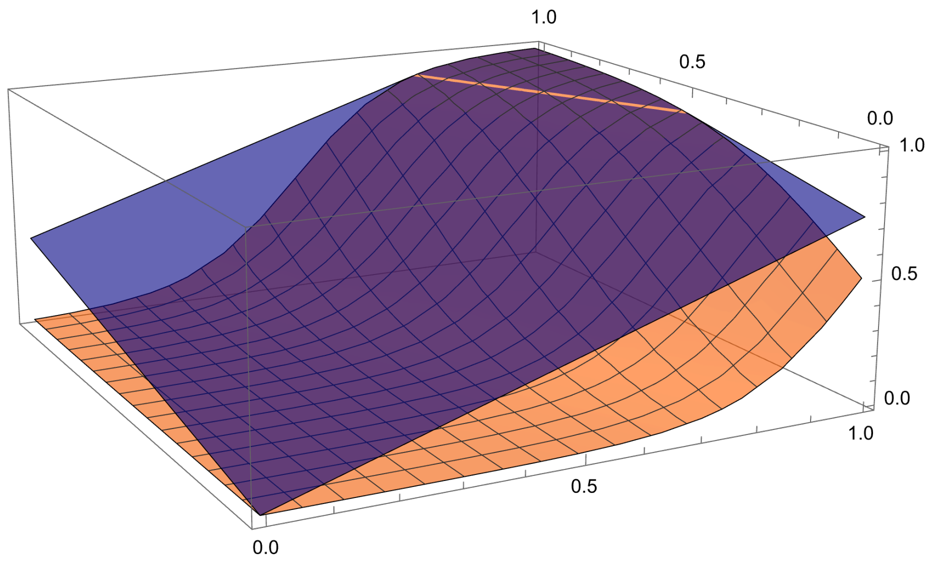
How to improve this?



The green part is OK

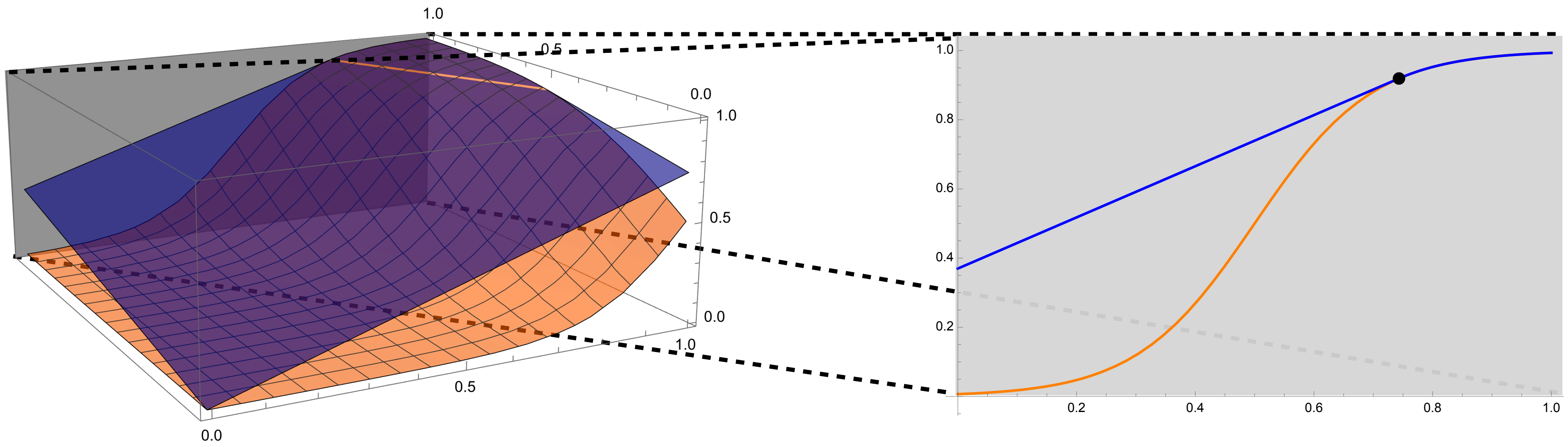
Beyond the ReLU

How to improve this?



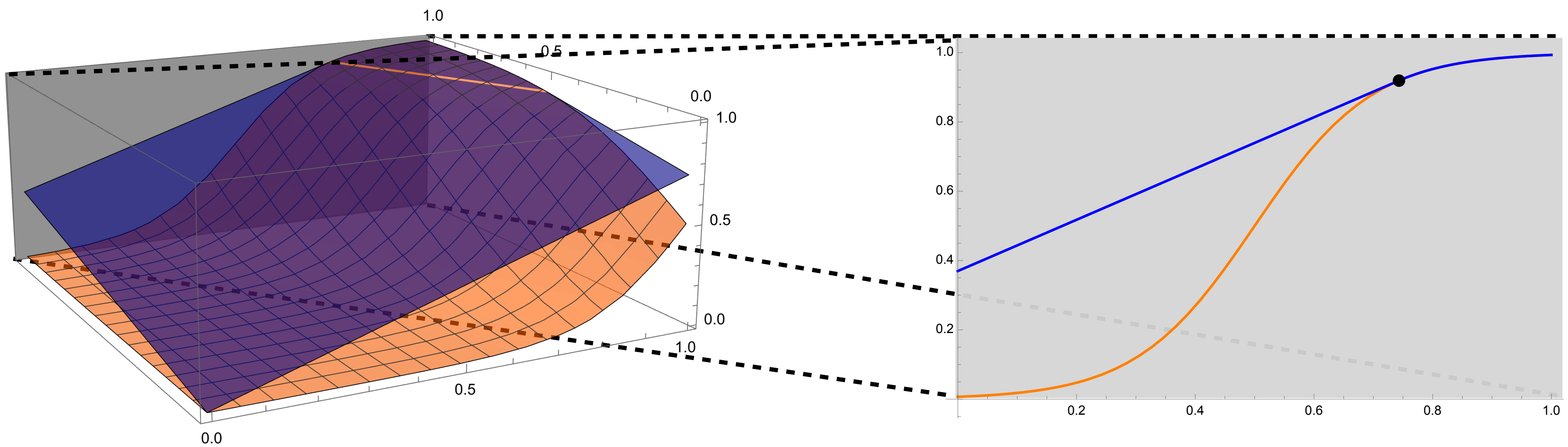
Beyond the ReLU

How to improve this?

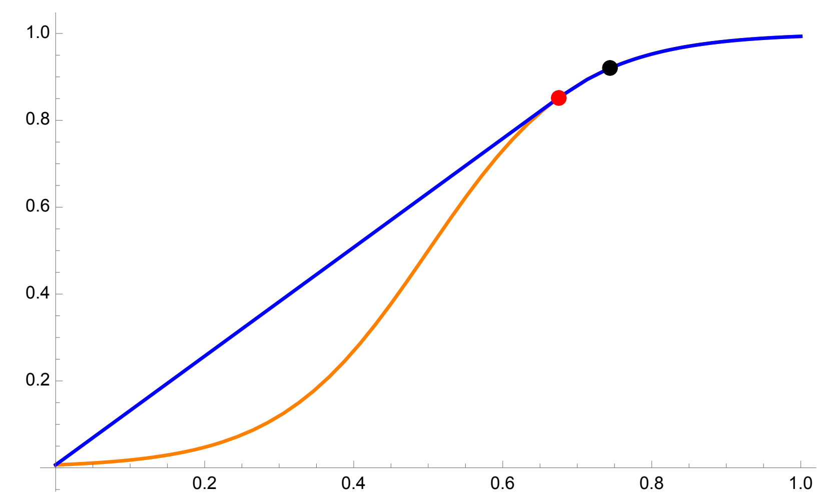


Beyond the ReLU

How to improve this?

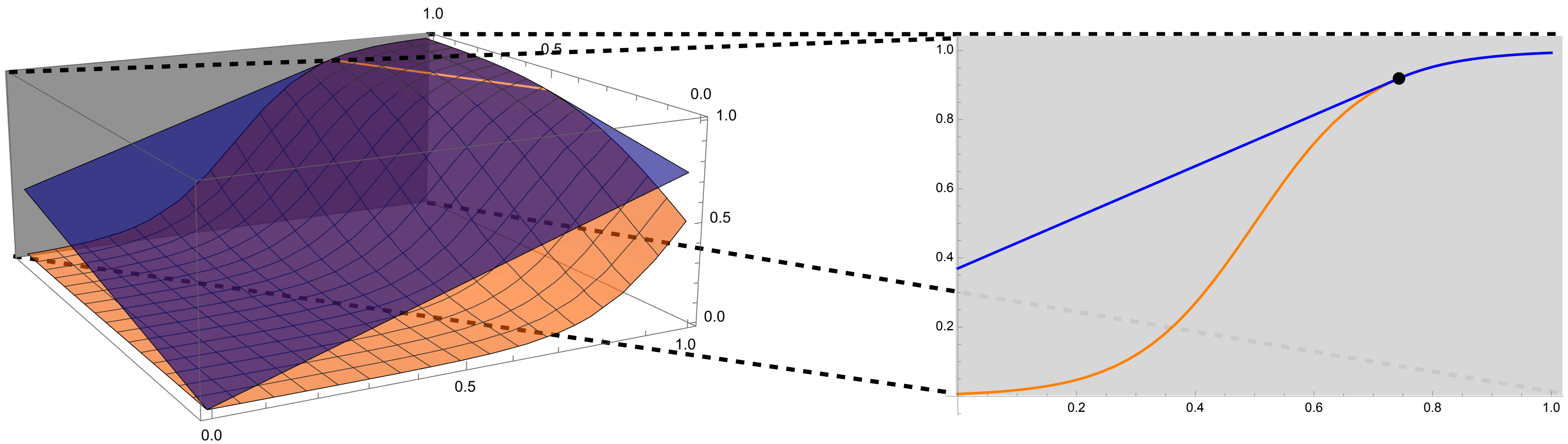


Tawarmalani and Sahinidis (2002) showed that, on the boundary, the envelope should look like this

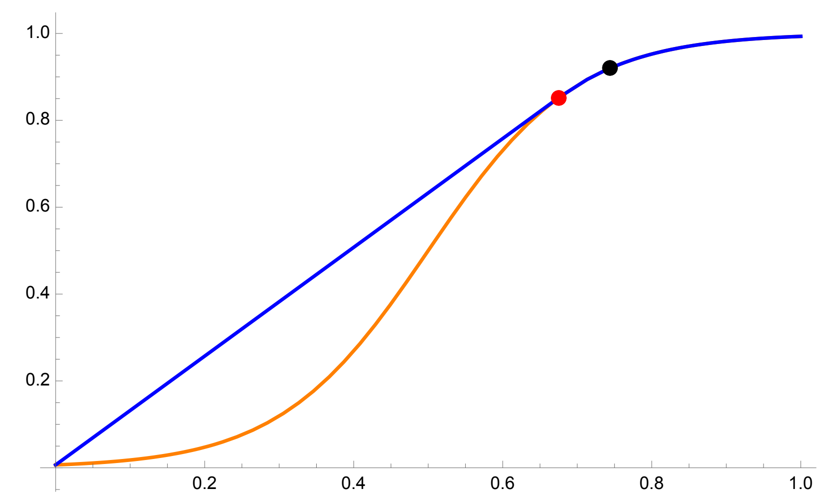


Beyond the ReLU

How to improve this?



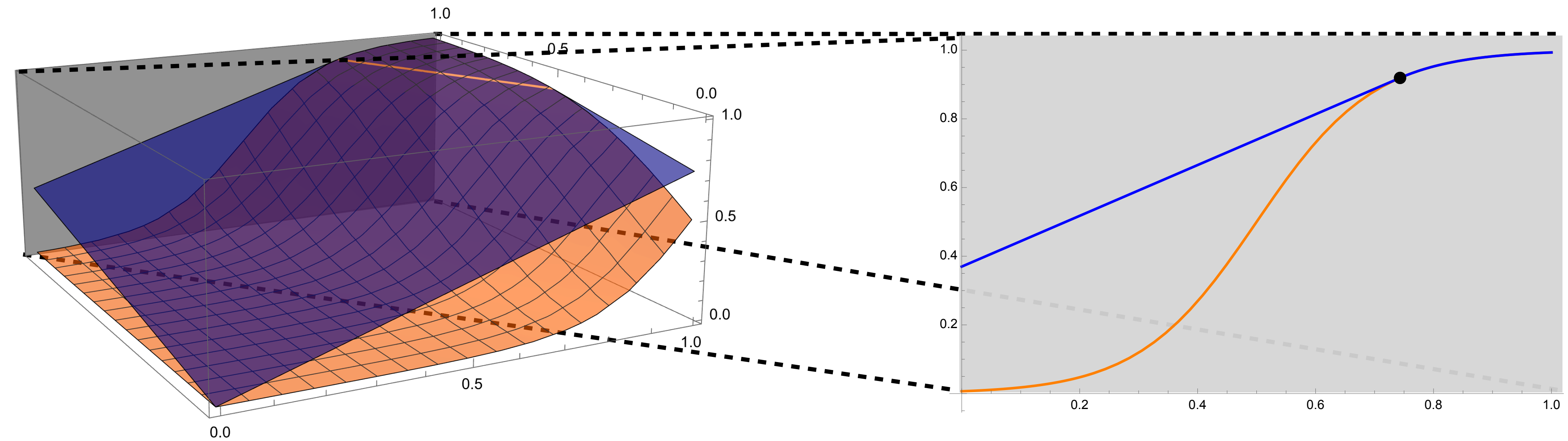
Tawarmalani and Sahinidis (2002) showed that, on the boundary, the envelope should look like this



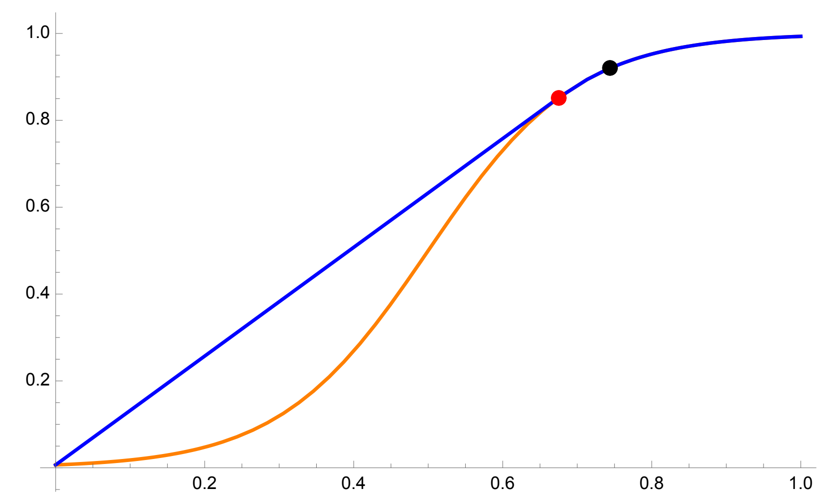
How to build from this?

Beyond the ReLU

How to improve this?



Tawarmalani and Sahinidis (2002) showed that, on the boundary, the envelope should look like this

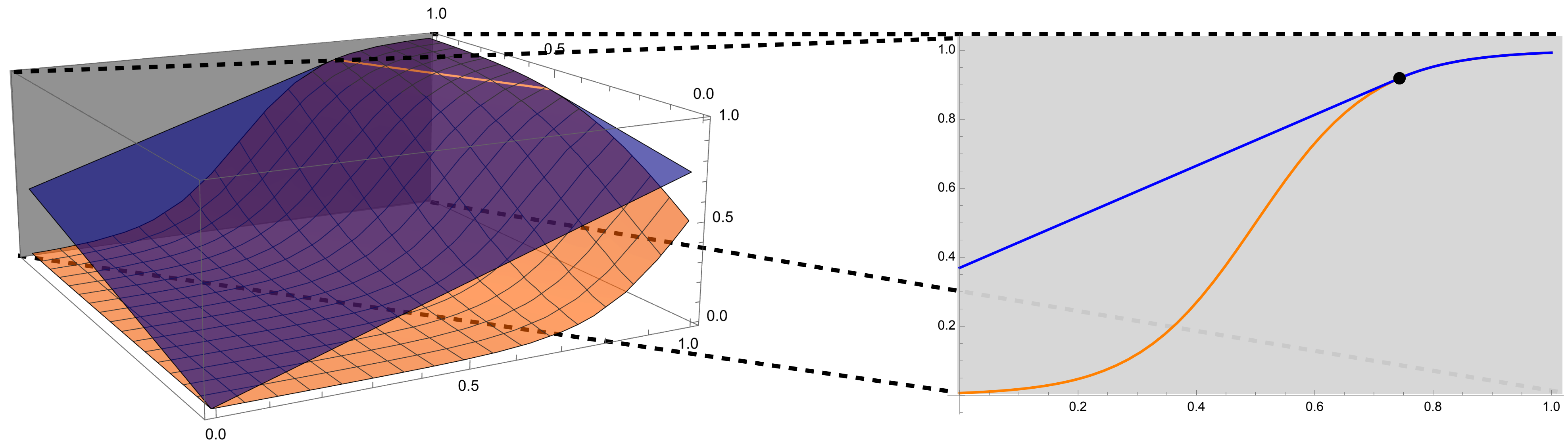


How to build from this?

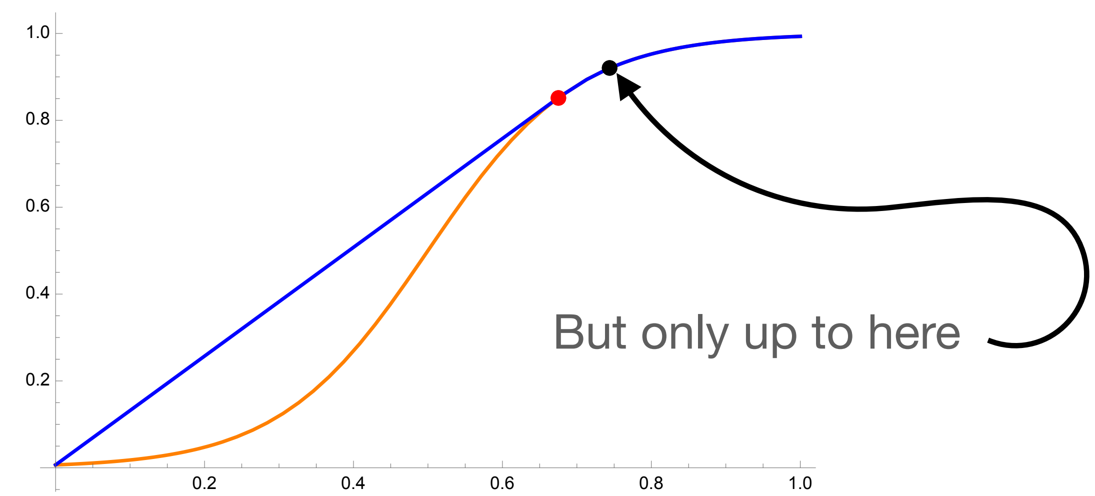
Taking perspective from the boundary!

Beyond the ReLU

How to improve this?



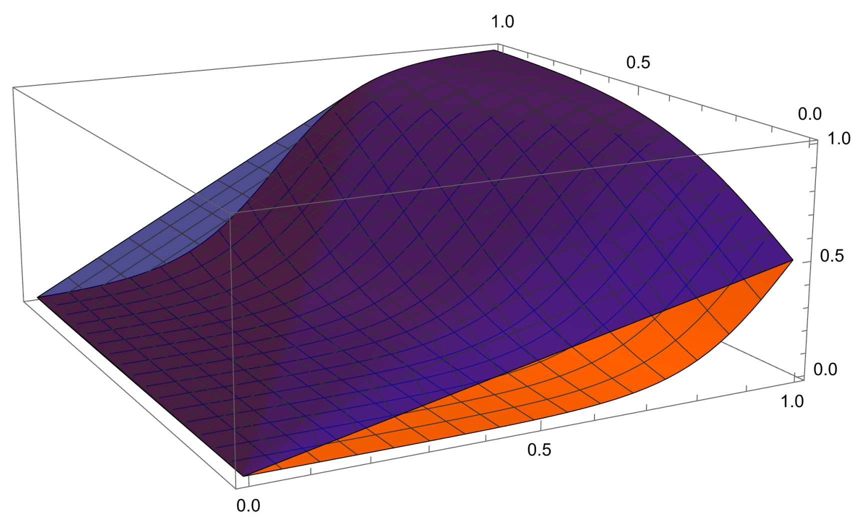
Tawarmalani and Sahinidis (2002) showed that, on the boundary, the envelope should look like this



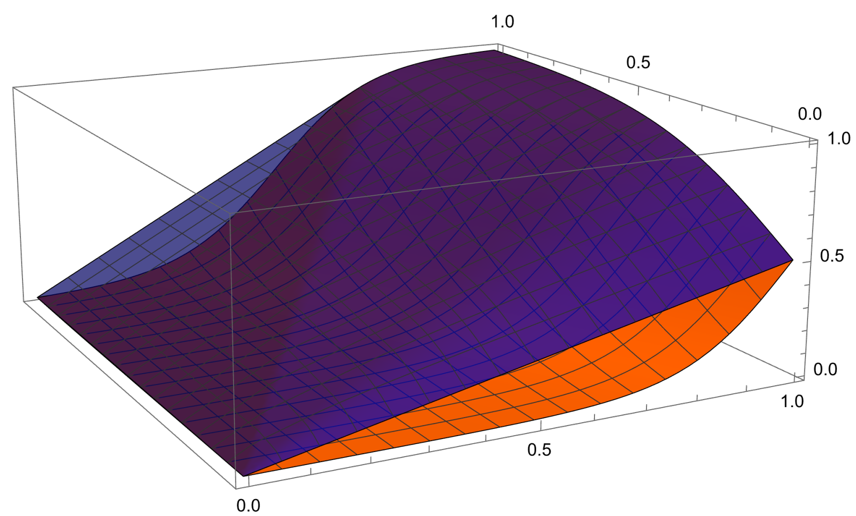
How to build from this?

Taking perspective from the boundary!

The main construction



The main construction

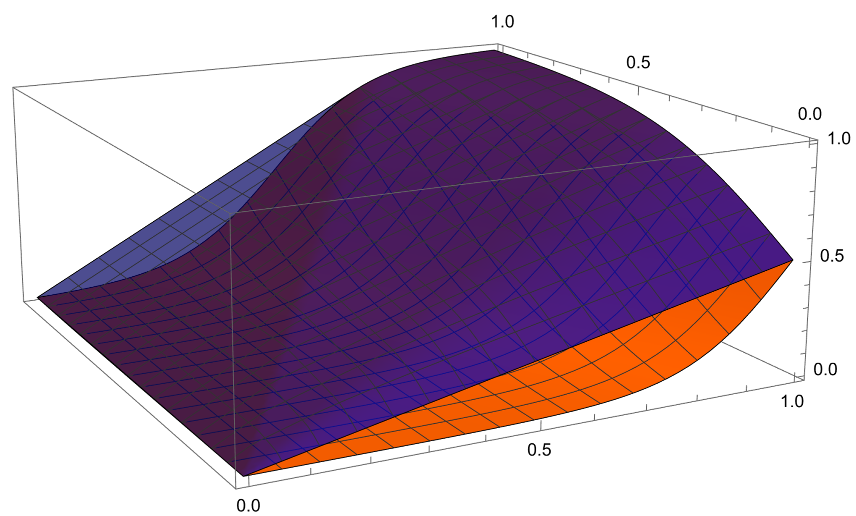


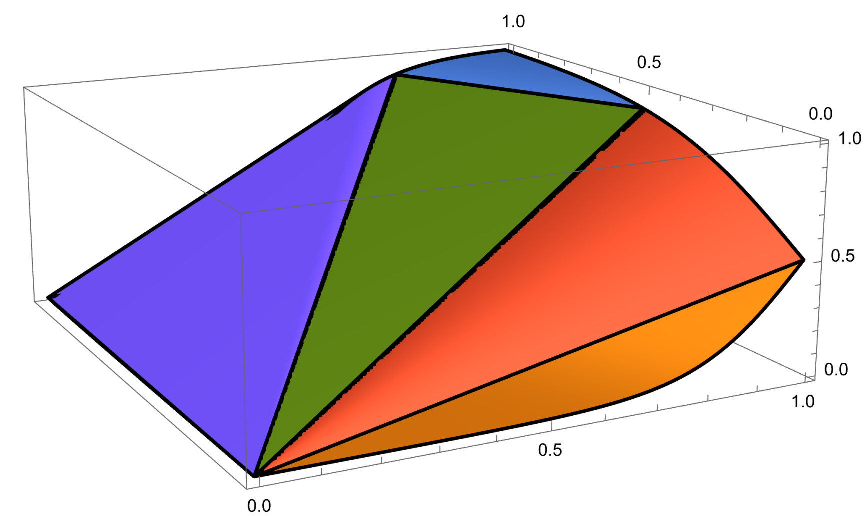
Theorem 1 *Consider* $w \in \mathbb{R}^n_+$, $b \in \mathbb{R}$ *and let* $f : [0,1]^n \to \mathbb{R}$ *be a function of the form* $f(x) = \sigma(w^\top x + b)$ *where* σ *satisfies the STFE property. Then,*

$$conc(f,[0,1]^n)(x) = \begin{cases} f(x) & \text{if } x \in R_f \\ \sigma(b) + \frac{\sigma(\hat{z})-\sigma(b)}{\hat{z}-b}(w^\top x) & \text{if } x \in R_l \\ \sigma(b) + x_i\, conc(f_{-i} - \sigma(b), [0,1]^{n-1})\left(\frac{x_{-i}}{x_i}\right) & \text{if } x \in R_i \end{cases}$$

where R_f, R_l *and* R_i *are defined using Definition 3 with* $\hat{z}$ *as the tie point of* σ *in* $[b, w^\top \mathbf{1} + b]$.

The main construction



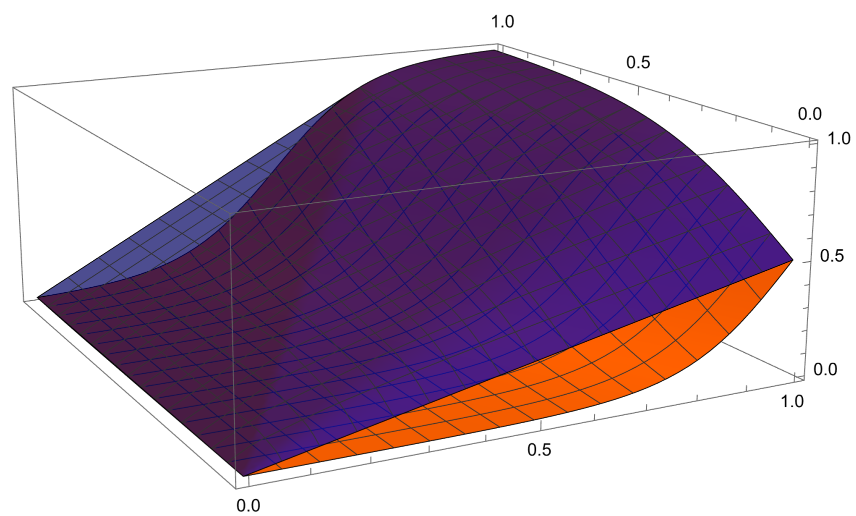


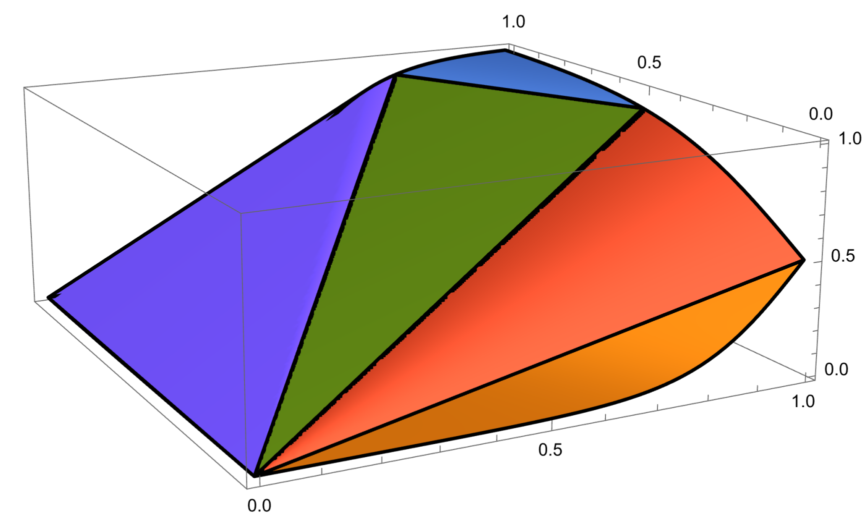
Theorem 1 *Consider* $w \in \mathbb{R}^n_+$, $b \in \mathbb{R}$ *and let* $f : [0,1]^n \to \mathbb{R}$ *be a function of the form* $f(x) = \sigma(w^\top x + b)$ *where* σ *satisfies the STFE property. Then,*

$$conc(f,[0,1]^n)(x) = \begin{cases} f(x) & \text{if } x \in R_f \\ \sigma(b) + \frac{\sigma(\hat{z})-\sigma(b)}{\hat{z}-b}(w^\top x) & \text{if } x \in R_l \\ \sigma(b) + x_i\, conc(f_{-i} - \sigma(b), [0,1]^{n-1}) \left(\frac{x_{-i}}{x_i}\right) & \text{if } x \in R_i \end{cases}$$

where R_f, R_l *and* R_i *are defined using Definition 3 with* $\hat{z}$ *as the tie point of* σ *in* $[b, w^\top \mathbf{1} + b]$.

The main construction





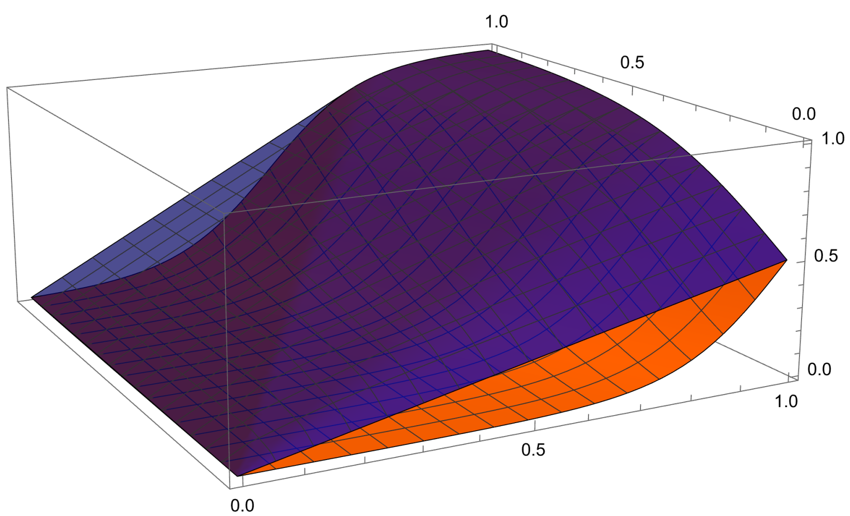
The envelope evaluation is either

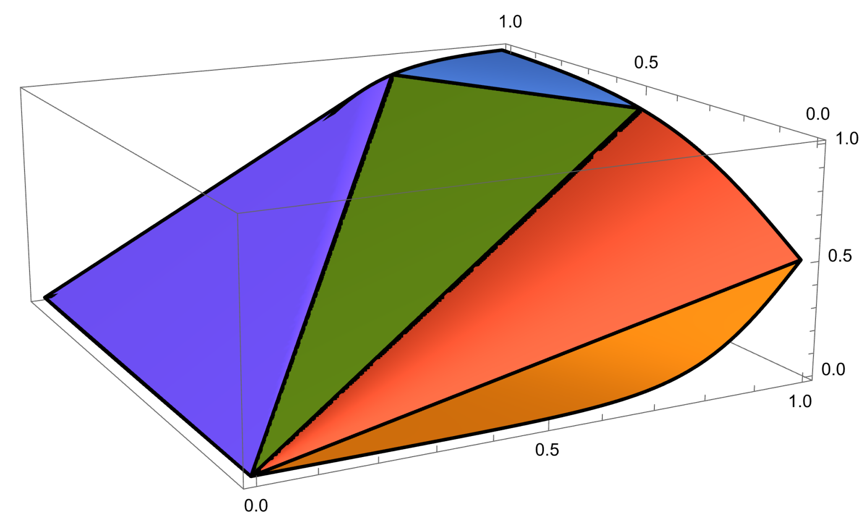
Theorem 1. *Consider $w \in \mathbb{R}^n$, $b \in \mathbb{R}$ and let $f : [0,1]^n \to \mathbb{R}$ be a function of the form $f(x) = \sigma(w^\top x + b)$ where σ satisfies the STFE property. Then,*

$$conc(f, [0,1]^n)(x) = \begin{cases} f(x) & \text{if } x \in R_f \\ \sigma(b) + \frac{\sigma(\hat{z}) - \sigma(b)}{\hat{z} - b}(w^\top x) & \text{if } x \in R_l \\ \sigma(b) + x_i \, conc(f_{-i} - \sigma(b), [0,1]^{n-1}) \left(\frac{x_{-i}}{x_i} \right) & \text{if } x \in R_i \end{cases}$$

where R_f, R_l and R_i are defined using Definition 3 with $\hat{z}$ as the tie point of σ in $[b, w^\top \mathbf{1} + b]$.

The main construction





Th… …*of*

The envelope evaluation is either

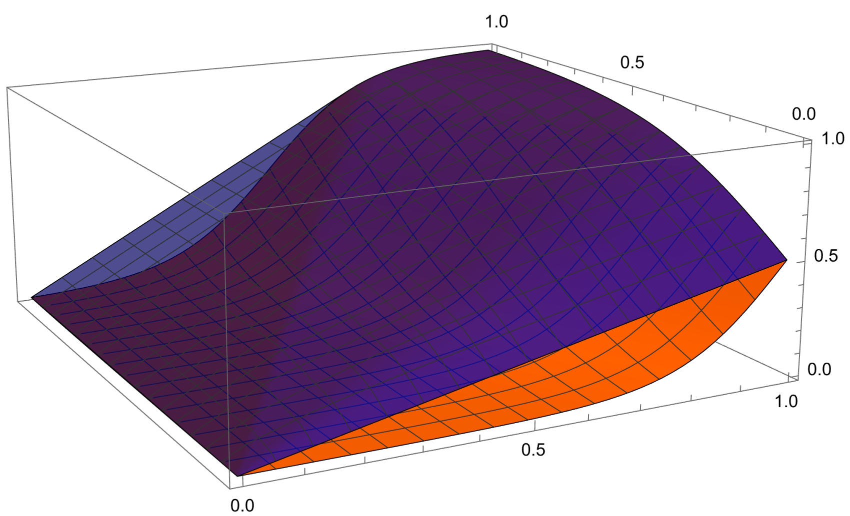
the … *property. Then,*

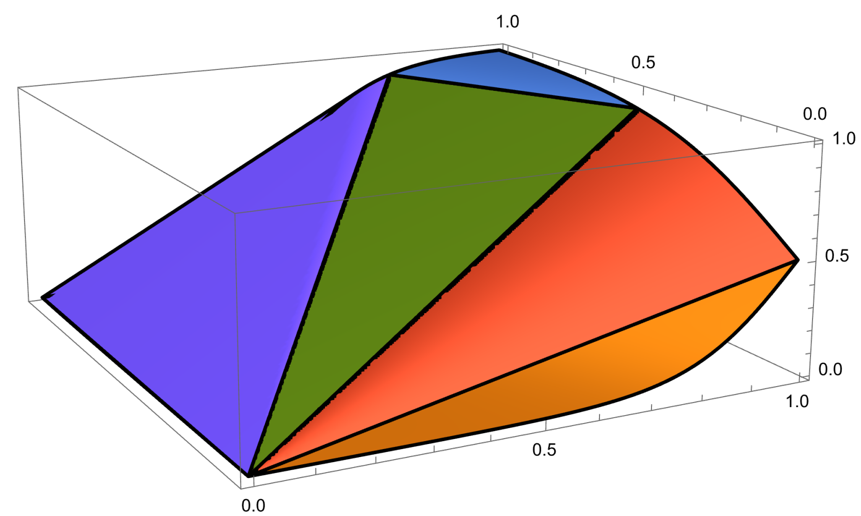
The function itself

$$conc(f,[0,1]^n)(x) = \begin{cases} \ldots & f \\ \sigma(b) + \frac{\sigma(\hat{z})-\sigma(b)}{\hat{z}-b}(w^\top x) & \text{if } x \in R_l \\ \sigma(b) + x_i\, conc(f_{-i} - \sigma(b), [0,1]^{n-1})\left(\frac{x_{-i}}{x_i}\right) & \text{if } x \in R_i \end{cases}$$

where R_f, R_l and R_i are defined using Definition 3 with $\hat{z}$ as the tie point of σ in $[b, w^\top \mathbf{1} + b]$.

The main construction





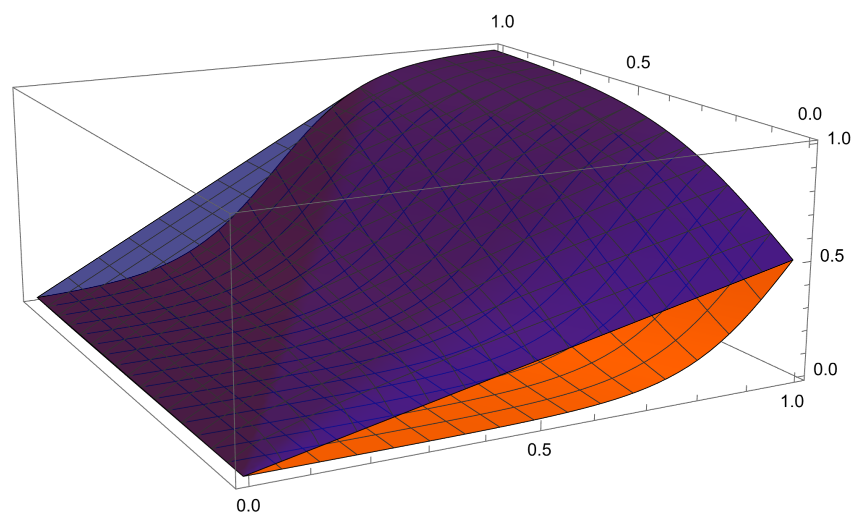
Th... *of the* ...

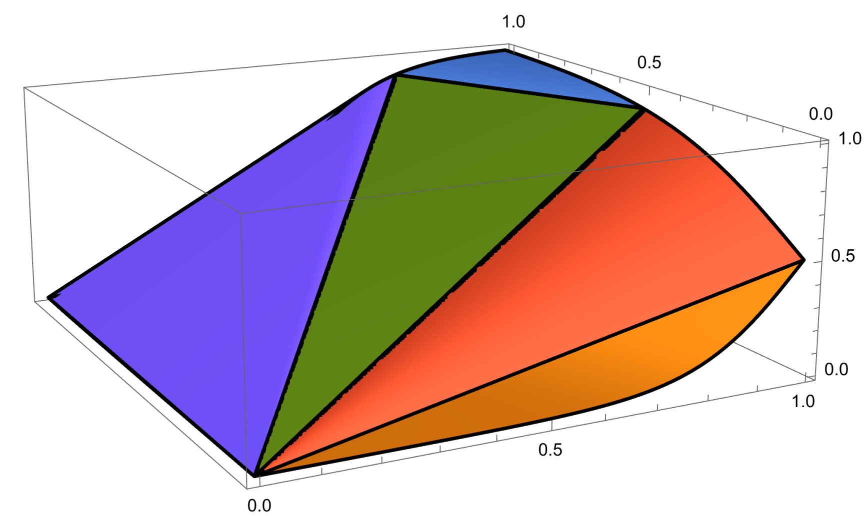
The envelope evaluation is either

$$conc(f, [0,1]^n)(x) = \begin{cases} \text{The function itself} & f \\ \text{A linear function} & l \\ \sigma(b) + x_i\, conc(f_{-i} - \sigma(b), [0,1]^{n-1}) \left(\frac{x_{-i}}{x_i}\right) & \text{if } x \in R_i \end{cases}$$

where R_f, R_l and R_i are defined using Definition 3 with $\hat{z}$ as the tie point of σ in $[b, w^\top \mathbf{1} + b]$.

The main construction



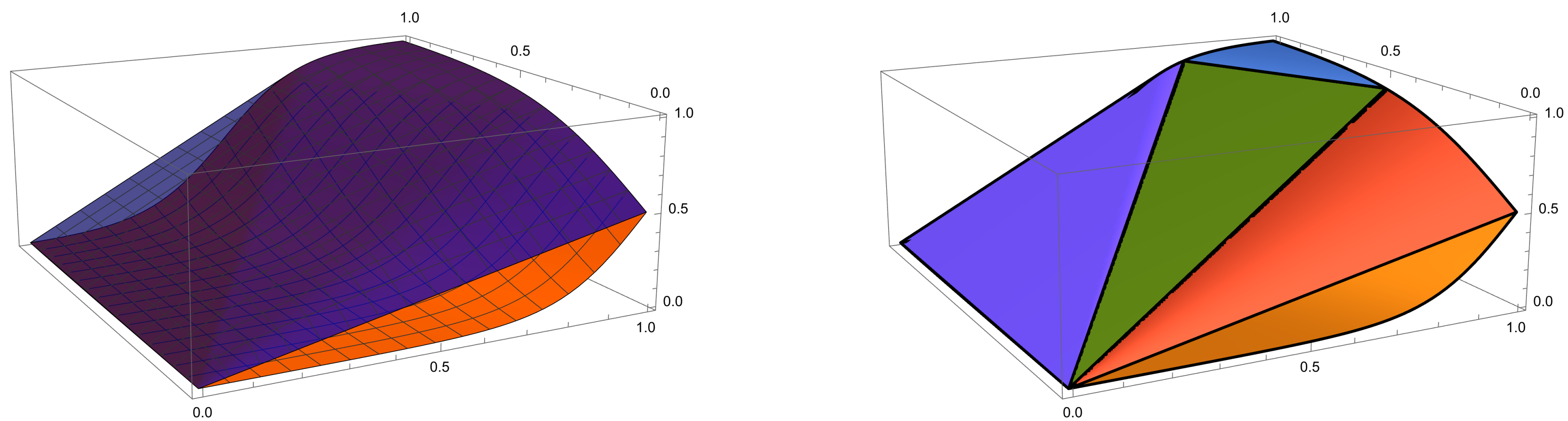


The envelope evaluation is either

$$conc(f, [0,1]^n)(x) = \begin{cases} \text{The function itself} \\ \text{A linear function} \\ \text{The perspective of an envelope of 1 dimension less} \end{cases}$$

where R_f, R_l and R_i are defined using Definition 3 with $\hat{z}$ as the tie point of σ in $[b, w^\top \mathbf{1} + b]$.

The main construction

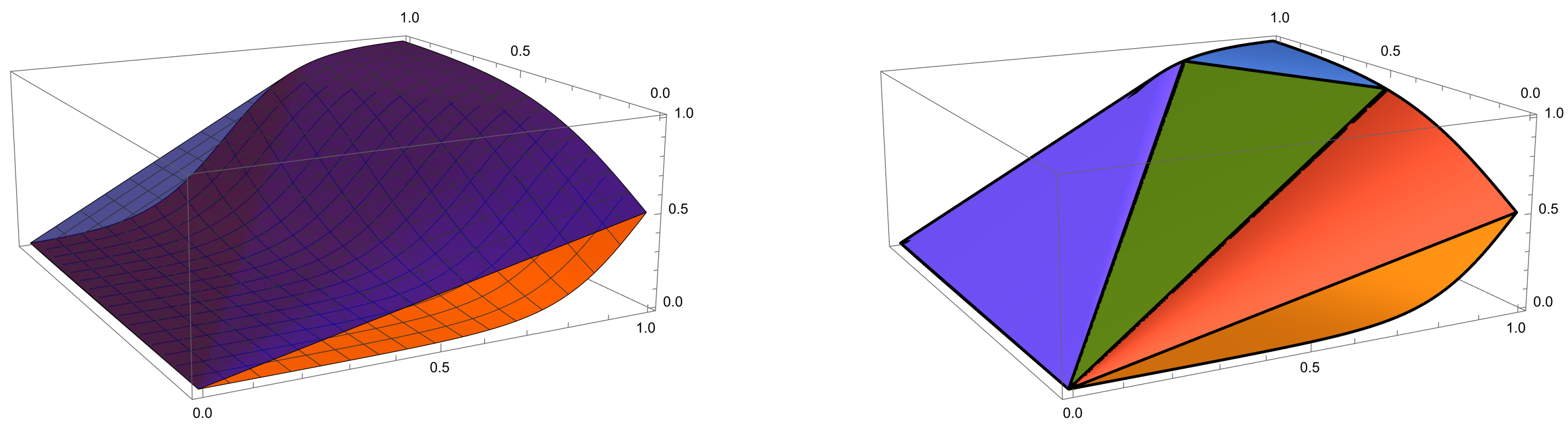


The envelope evaluation is either

$$conc(f,[0,1]^n)(x) = \begin{cases} \text{The function itself} \\ \text{A linear function} \\ \text{The perspective of an envelope of 1 dimension less} \end{cases}$$

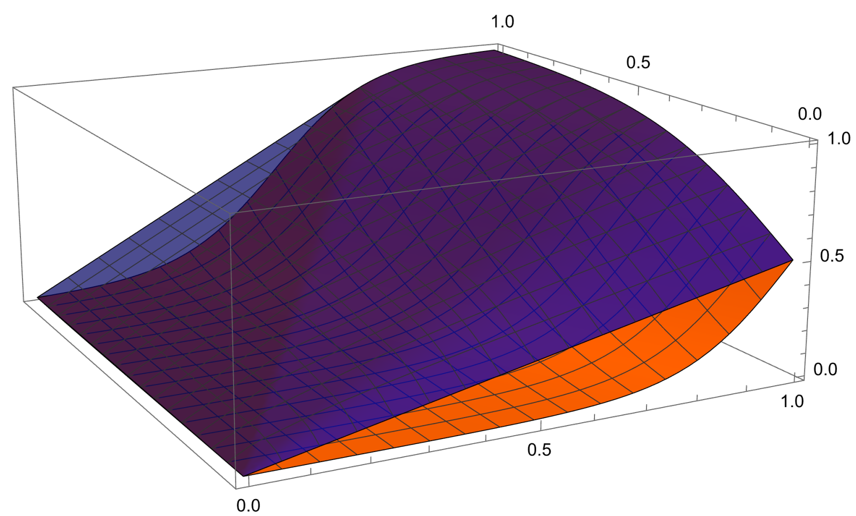
This works for any convex or S-shaped activation function on a box,
can construct supporting hyperplanes of the convex hull of the graph,
and requires at most n 1-dimensional convexifications

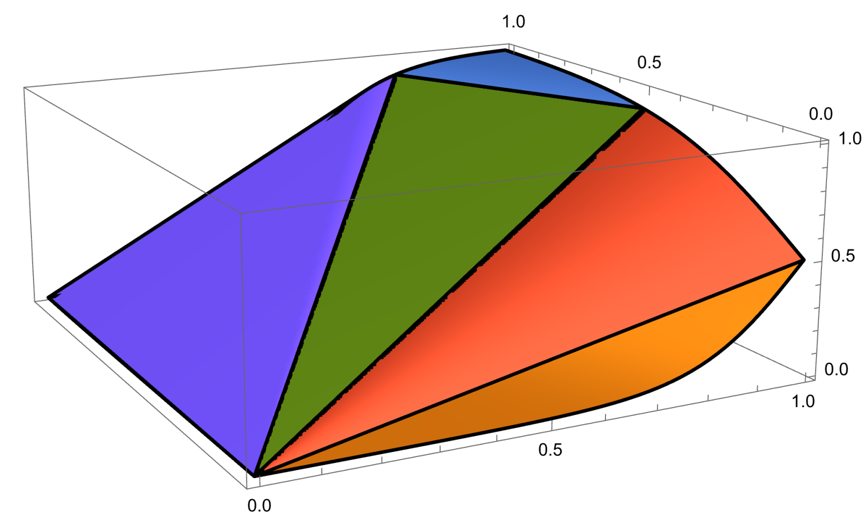
The key ingredients



An envelope construction like this does not work in general, here we heavily rely on

The key ingredients

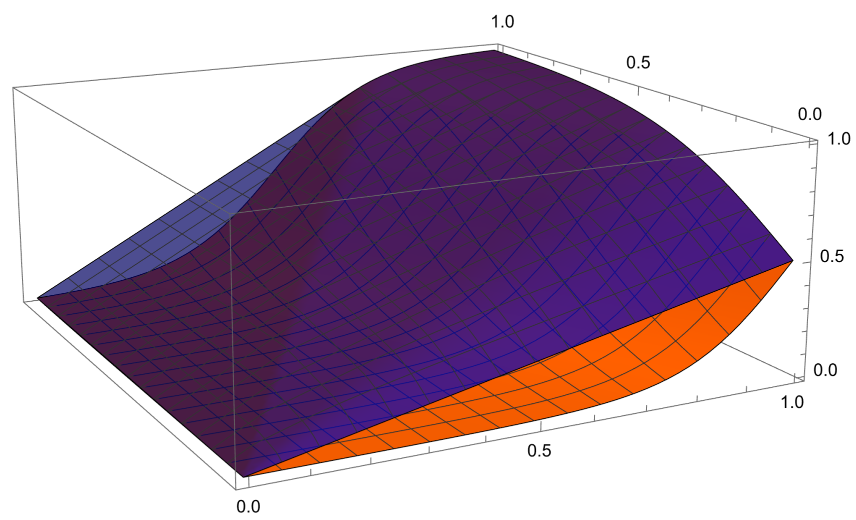


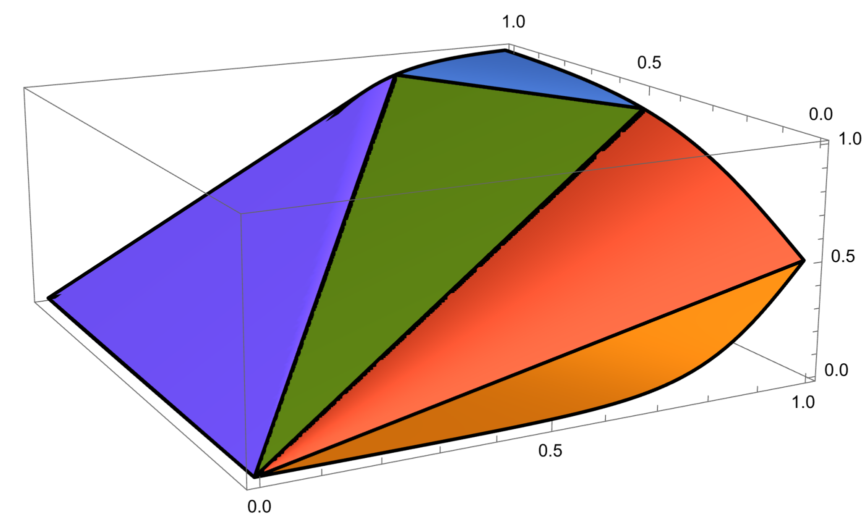


An envelope construction like this does not work in general, here we heavily rely on

The one-dimensional envelope is always a secant followed by the function

The key ingredients



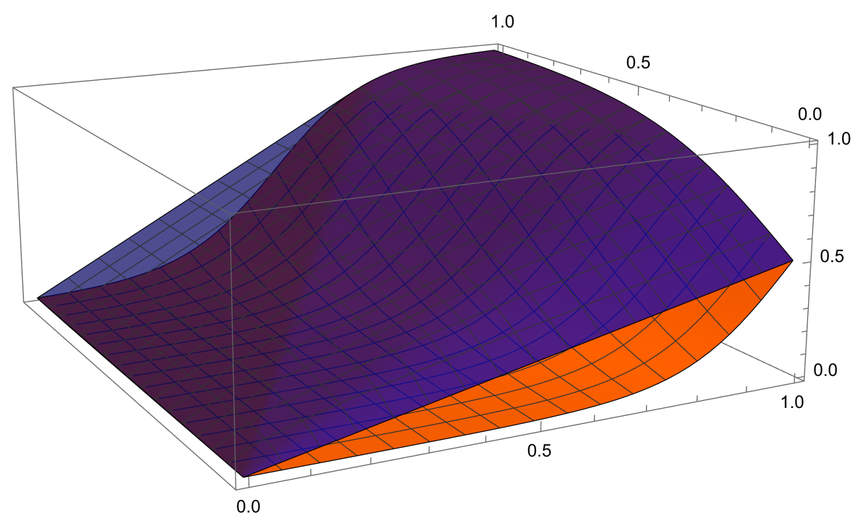


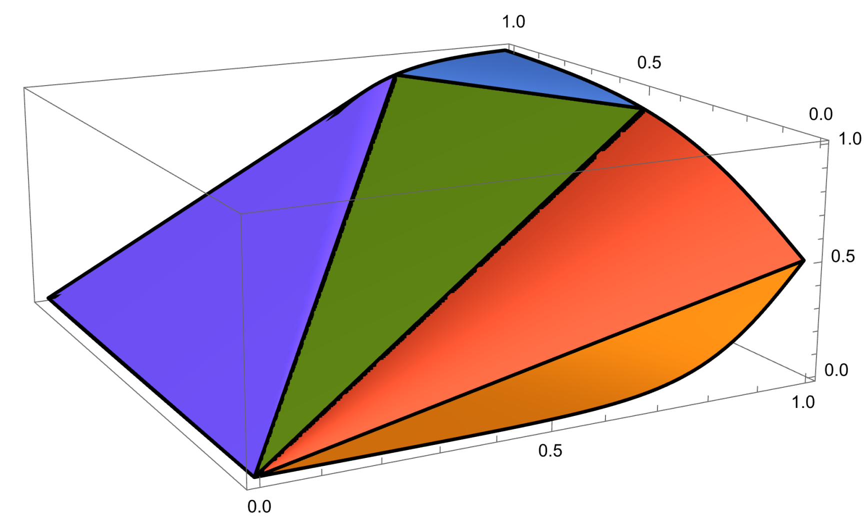
An envelope construction like this does not work in general, here we heavily rely on

The one-dimensional envelope is always a secant followed by the function

On the faces, the global structure repeats itself

The key ingredients





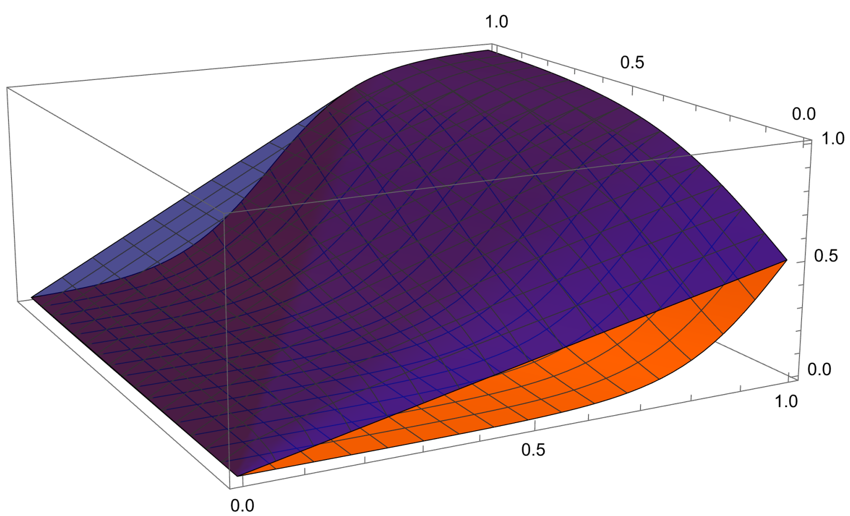
An envelope construction like this does not work in general, here we heavily rely on

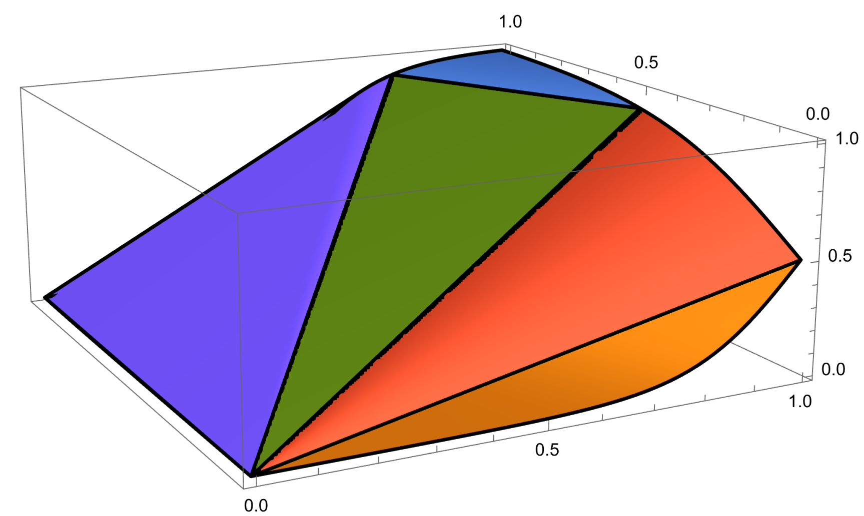
The one-dimensional envelope is always a secant followed by the function

On the faces, the global structure repeats itself

The function $\sigma(w^\top x + b)$ is constant in $w^\perp$

The key ingredients





An envelope construction like this does not work in general, here we heavily rely on

The one-dimensional envelope is always a secant followed by the function

On the faces, the global structure repeats itself

The function $\sigma(w^\top x + b)$ is constant in $w^\perp$

Thank you!