SQNL: A New Computationally Efficient Activation Function

Ms. Adedamola Wuraola, Dr. Nitish Patel

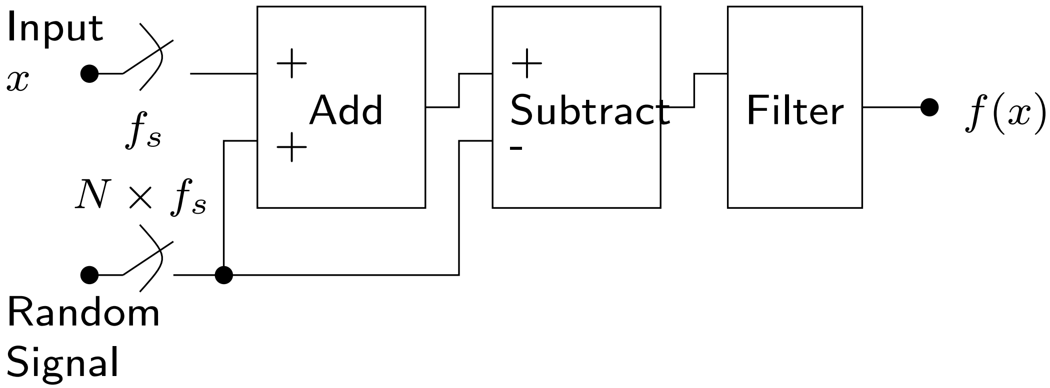
July 2018

Motivation

- Hardware ANN
  - Playback (forward pass)
  - Training (backpropagation)
- FPGA based, maybe ASIC based (neither Neuromorphic nor GPU/TPU styled)

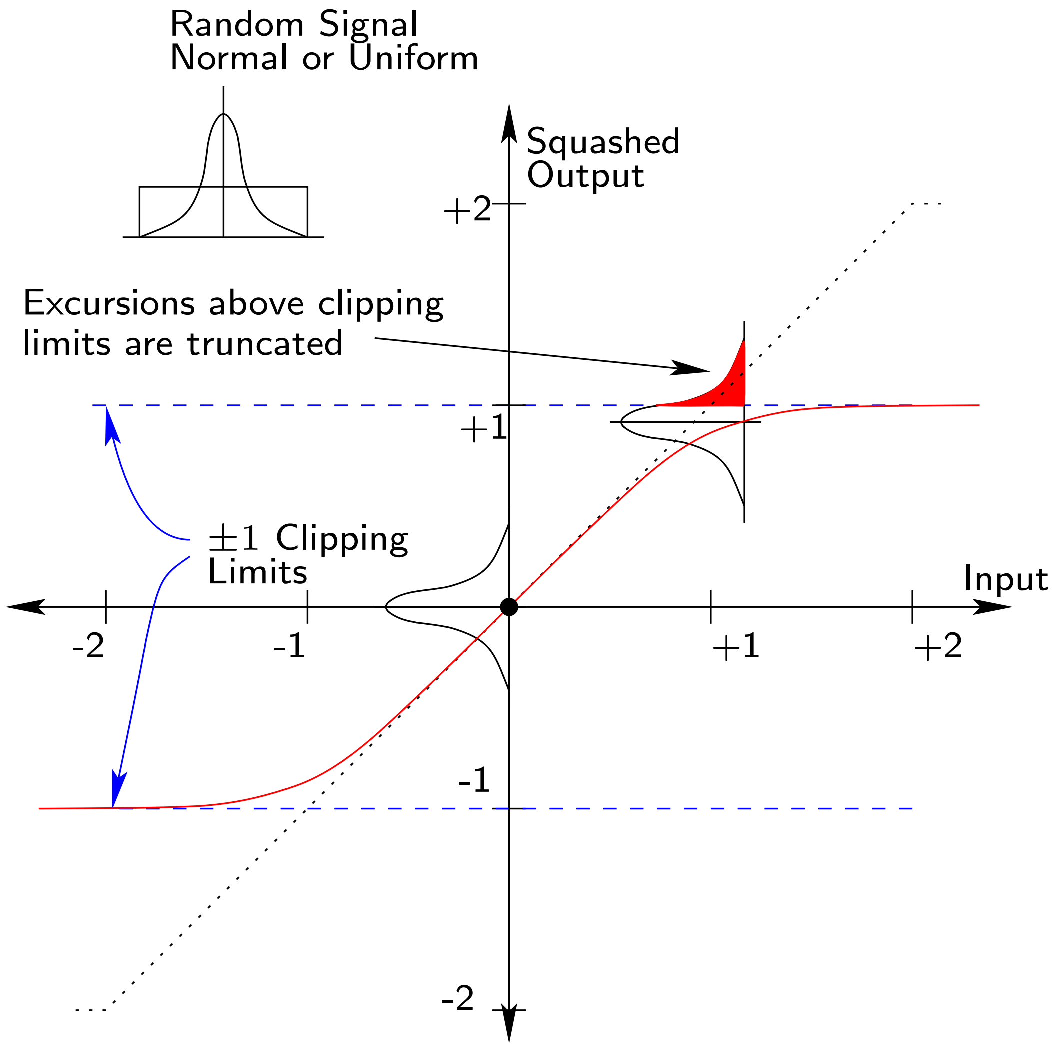
Building a non-linearity: Concept

Saturating Adder and Subtractor

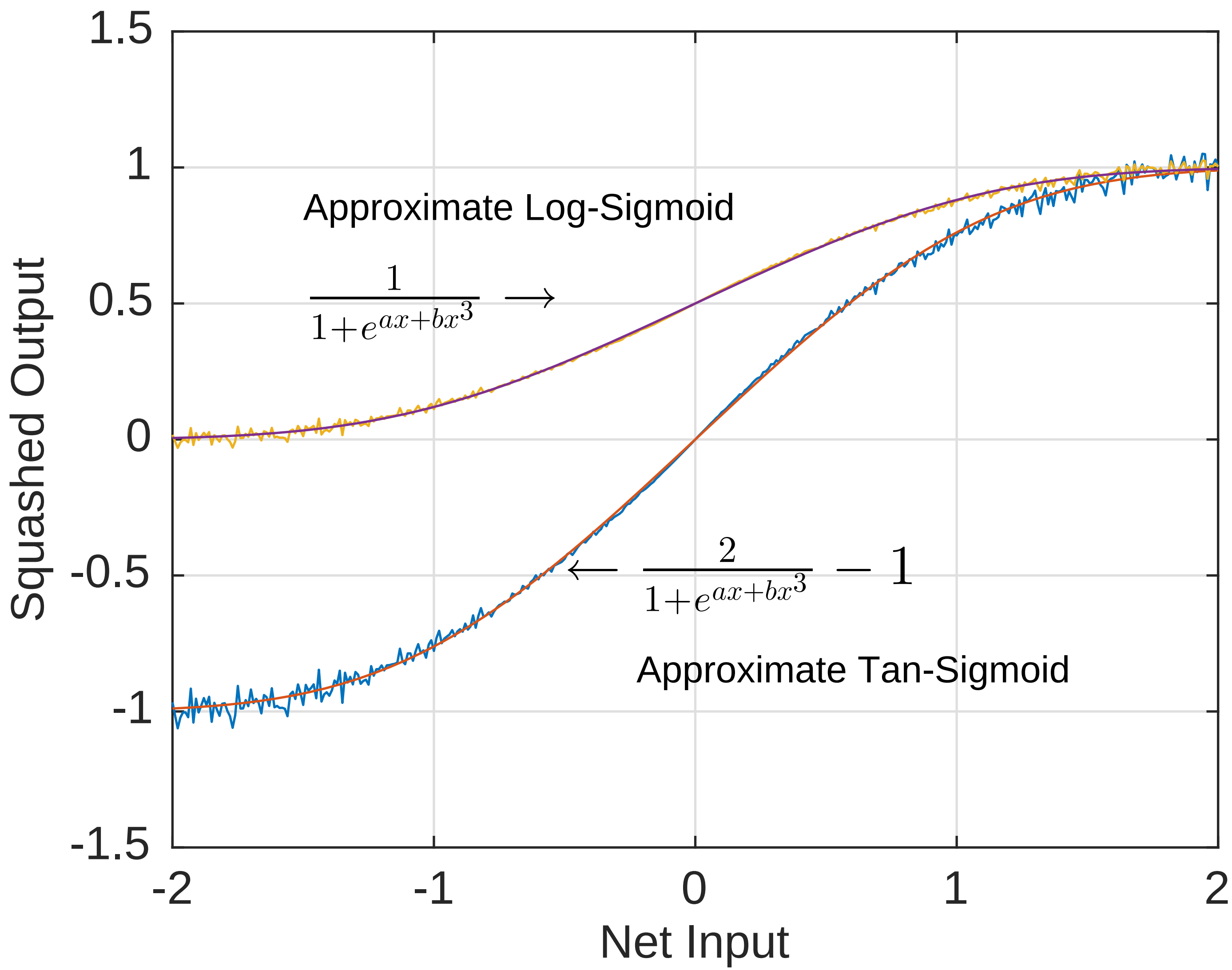


- Random source is uniformly distributed
- Adder and subtractors are saturating
- Filter can be complex or simple.

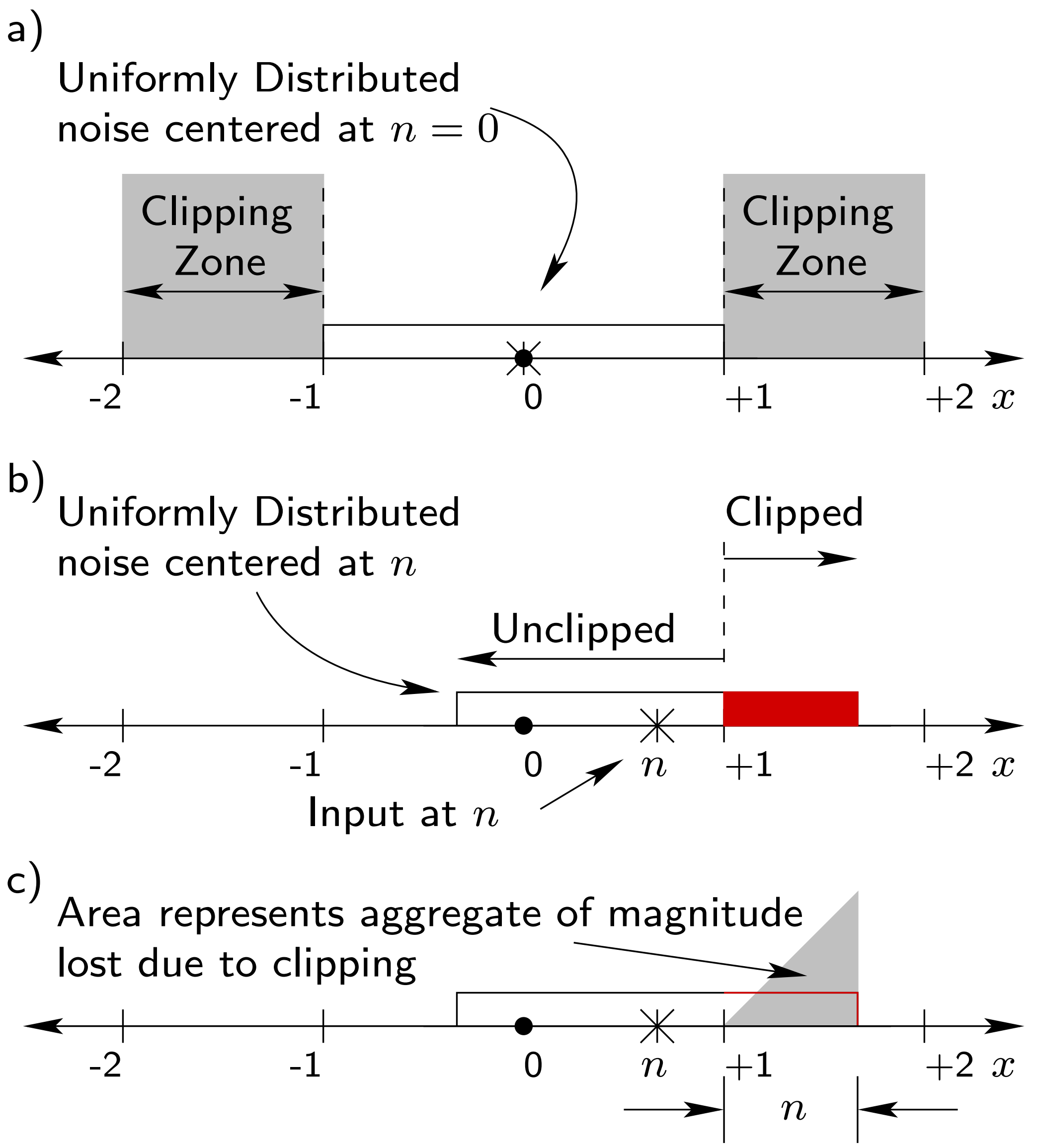
Concept (cont)



Concept (cont)



Concept (cont)



Mapping Function

The saturation results in a clipping $o(x)$ that is given by

$$o(x)\Big|_{x=n} = \frac{1}{2}\int_{1}^{1+n}(x-1)dx = \frac{n^2}{4}$$

The resultant effect is a non-linear mapping defined by

$$f_B(x) = \begin{cases} 1 & : x > 2.0 \\ x - \frac{x^2}{4} & : 0 \leq x \leq 2.0 \\ x + \frac{x^2}{4} & : -2.0 \leq x < 0 \\ -1 & : x < -2.0 \end{cases}$$

Properties of the SQNL

The function has been named the Square-Law Non-Linear (SQNL) function due to its inherent square operation.

Simple Non-Linearity The square law is, arguably, the simplist non-linearity

Symmetrical and Continuous It is symetrical around zero and it is continuous between $-\infty$ and $+\infty$

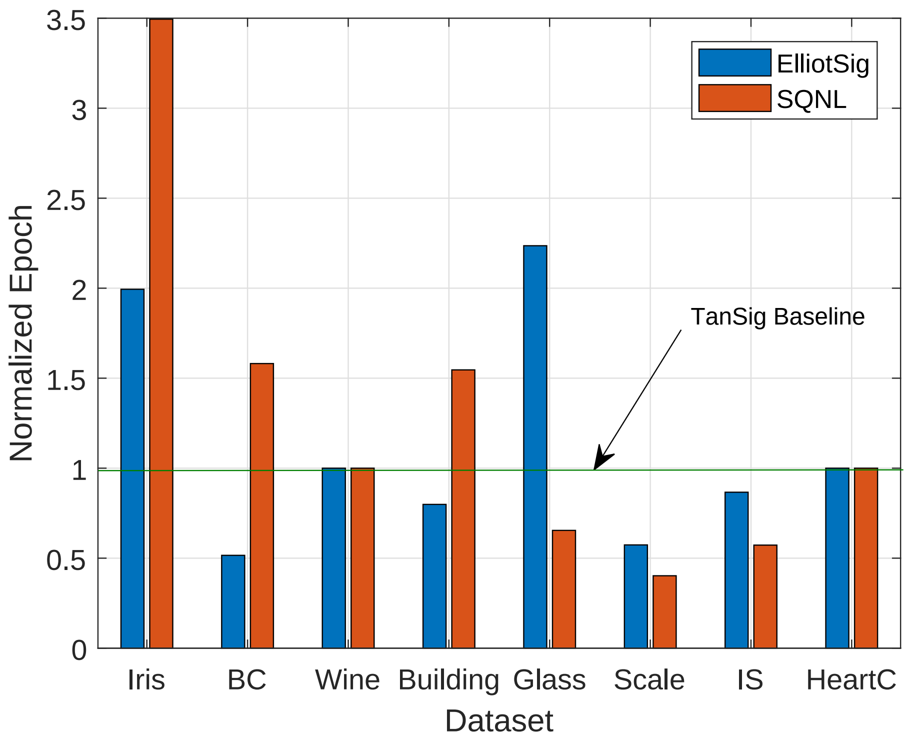
Linear Derivative The derivative of the SQNL is linear.

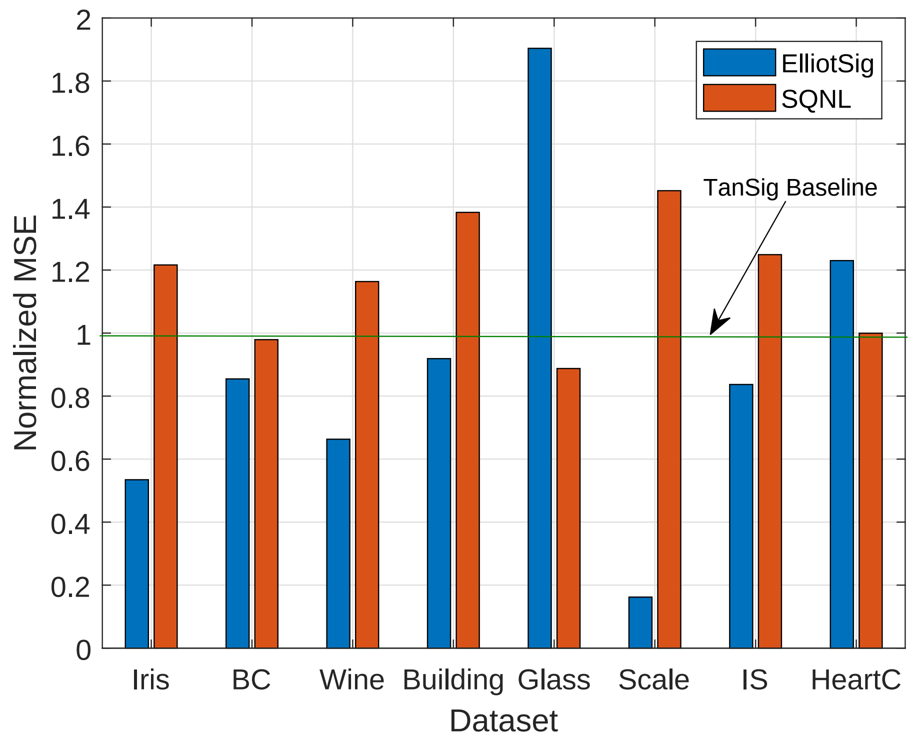
Is this mapping comparable to other similar mappings?

Method: By attempting to quantify both convergerce speed and the ability to generalise

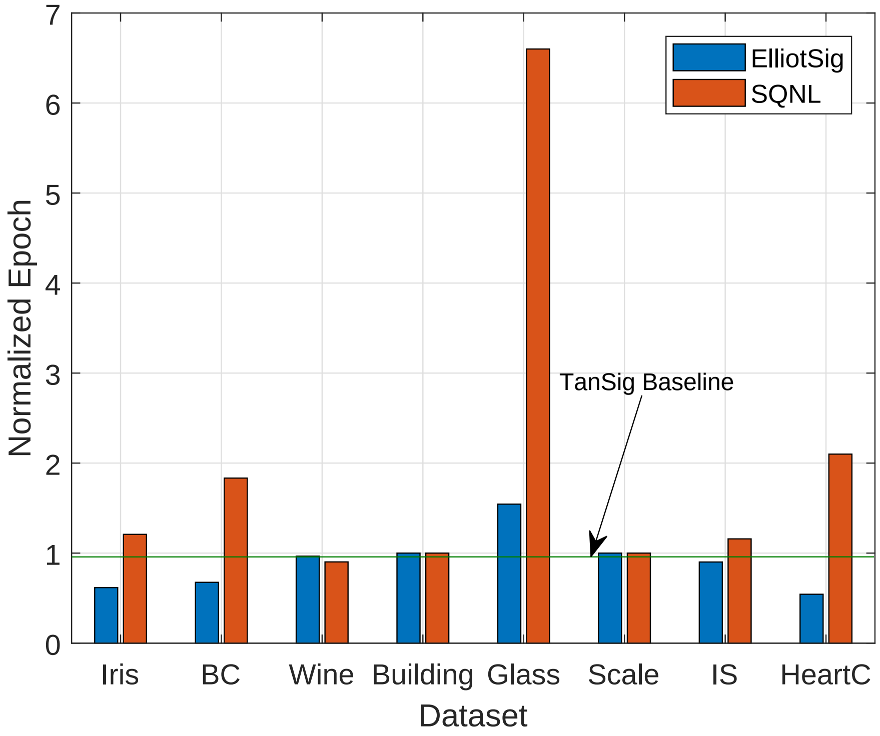
- 100 networks trained: 100 weight-sets stored and reused with every experiment
- Only mapping function changed
- Both epochs and mean square error logged

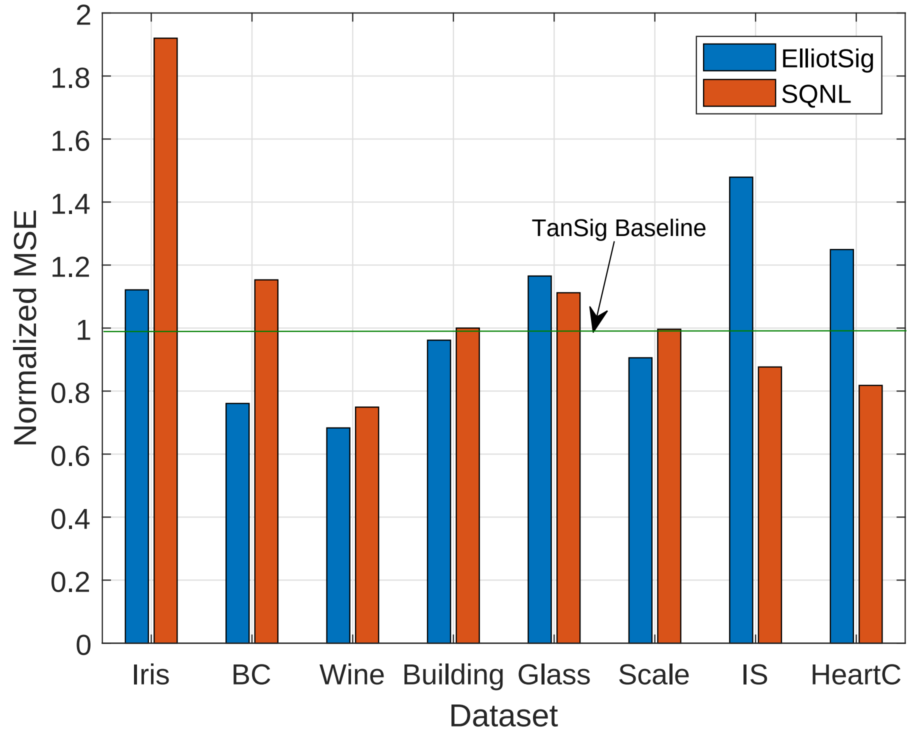
Backpropagation



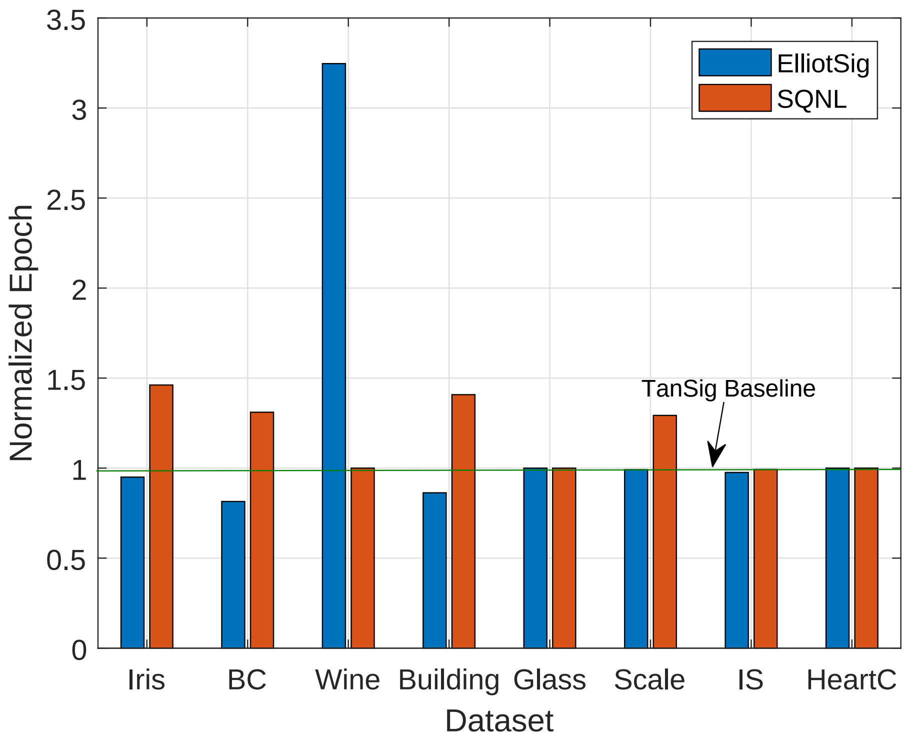


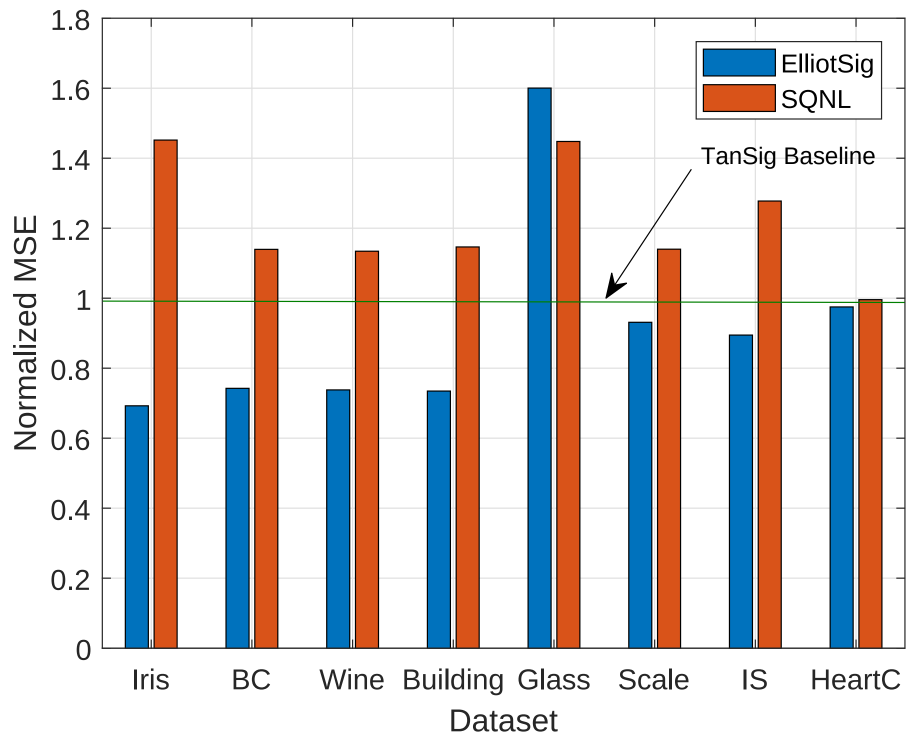
Levenberg-Marquandt





Resilient Backpropagation





Convergence Speed On MNIST Dataset

Function	Epoch
ElliotSig	17.74
TanSig	8.45
SQNL	8.11

Conclusions

- The SQNL is a simple non-linearity
- Mathematically speaking, both forward and derivative functions are simple
- The SQNL seems to perform better.
- However, the variation in performance suggests a strong data set dependence
- Importantly, the SQNL is not inferior to the well established TanSig.
- Digital circuit implementations are possible