On the Transfer Functions for Complex Neurons*

Marco Budinich

Physics Department, Trieste University and INFN
Via A.Valerio 2, 34127 Trieste, Italy
http://wwwusers.ts.infn.it/~mbh/MBHgeneral.html
marco.budinich@ts.infn.it

Abstract. Since some time the extension of neural networks to the complex field $\mathbb{C}$ has received considerable attention for possible applications to phenomena in which the variables contain also phase information. In this paper we focus the attention on the complex transfer function. In compliance to Liouville's theorem, that forbids entire bounded analytic functions, a frequent choice has been to separate real and imaginary part of the output of a neuron. We show that this choice corresponds to throwing the baby with the water since these kind of complex neurons are identical to 2 purely real ones thus losing any possible complex advantage. On the contrary we show that a neuron with an analytic transfer function provides an output that is not approximable by 2 real neurons, like in the previous case, thus indirectly proving the superior computing power of truly complex neurons.

1 Introduction

Since some years the idea of extending real neurons to the complex field $\mathbb{C}$ received considerable attention, see e.g. [1], with many papers appearing both in theory [2–5] and in applications [6–9]. In particular it appears particularly attractive the implicit possibility to treat properly data containing both amplitude and phase information. This could find applications in many fields like pattern recognition and classification, image processing, time series prediction etc.

The first part of this paper introduces the notation and the choices one should make when shifting to neurons with complex activations. Subsequently the paper concentrates on the transfer functions for a complex valued neuron.

2 Real neurons

Before entering the complex case we review the notation in the more familiar case of real neurons. One usually considers a model of neuron, descending from

* DRAFT, April 11, 2007
McCulloch and Pitts', with $n$ inputs $(x_1, x_2, \ldots, x_n) := \mathbf{x} \in [-1, 1]^n$ and the activation $z$ is defined as a weighted average of the inputs

$$z = \sum_{k=1}^{n} w_k x_k := \mathbf{w} \cdot \mathbf{x}$$

where the weights $w_k$ represent the synapses (we consider the threshold as one of the weights) and finally produces an output $y$ given by

$$y = f(z) = f(\mathbf{w} \cdot \mathbf{x})$$

where the function $f()$ is the transfer function of the neuron. A popular choice for the transfer function is

$$f(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

that offers several advantages: it is bounded in $(-1, 1)$, so that it can be fed directly to a successive neuron; it is continuous on the whole real axis together with its derivative, and its derivative is easy to calculate since

$$\tanh'(z) = 1 - \tanh^2(z)$$

Moreover one can introduce a scale parameter $\beta$ with which $f_\beta(z) = \tanh(\beta z)$ and one can fine tune its steepness since $f_\beta(0) = \beta$ and $f_\beta(z) \to \text{sgn}(z)$ when $\beta \to \infty$ so that one can change smoothly between continuous and discrete neurons.

### 2.1 Interpretation of $z$ and $y$

In the real case the subspace of $\mathbb{R}^n$ defined by $\mathbf{w} \cdot \mathbf{x} = 0$ is an hyperplane that represent the border separating two adjacent regions in which the output of the neuron takes respectively positive or negative values. So the activation $z = \mathbf{w} \cdot \mathbf{x}$ can be seen as a 'measure of belonging' of the input $\mathbf{x}$ to one of these regions. Mathematically this is obtained by the linear map defined by the scalar product $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ whose output has an unambiguous interpretation since $\mathbb{R}$ is a fully ordered set and, given two different activations, we can always say which is larger. The transfer function $f(z)$ maps successively the unbounded activation $z$ in a bounded interval $y \in (-1, 1)$ in a monotone way.

### 3 Extensions to $\mathbb{C}$

Many authors introduced complex neural networks as a direct extension of standard ones. One of the goals was that of exploiting complex number properties to better encode the cases in which the inputs contain also phase information. The standard approach has been to extend the traditional McCulloch and Pitts neuron model to one in which all quantities are complex and precisely: inputs,
weights and output. One obvious request is that as the imaginary part goes to 0 one returns to the standard case.

Let us suppose that the \( n \) inputs become complex so that for each input and its corresponding weight we will write

\[
x_k = x_{kR} + ix_{kI} \quad \text{and} \quad w_k = w_{kR} + iw_{kI}
\]

where the \( \text{R} \) and \( \text{I} \) subscripts indicate the real and imaginary parts respectively that render explicitly the isomorphism existing between \( \mathbb{C}^n \) and \( \mathbb{R}^{2n} \).

### 3.1 The activation \( z \)

When we accept that all quantities can be complex the first problem we are faced with is that of the activation \( z = \sum_{k=1}^{n} w_k x_k \) that, if on one hand immediately extends to the algebra of complex numbers, on the other hand contemporarily introduces some subtle problems.

We enumerate a list of problems together with some comments:

- the extension of the real scalar product to complex numbers

\[
z = \mathbf{w} \cdot \mathbf{x} \quad \text{a linear map} \quad \mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C}
\]

even if immediate, violates the usual axioms of an inner product, e.g. one can have \( \mathbf{z} \cdot \mathbf{z} = 0 \) with \( \mathbf{z} \neq 0 \) (for example for \( \mathbf{z} = (1, i) \in \mathbb{C}^2 \)).

- The loci of \( \mathbf{w} \cdot \mathbf{x} = 0 \) does not extend immediately the concept of border it had in \( \mathbb{R}^n \) to \( \mathbb{C}^n \) and in general can contain null (or isotropic) subspaces. Also this point can be partially circumvented, for example with the choice of a proper inner product, but is not clear that this is an advantage, for example the rich structure of null subspaces could be useful for discriminating the inputs. Some flavour of this has already been pointed out by some authors that recently introduced Clifford algebras in the field [3].

- \( \mathbb{C} \) is not an ordered set so that the activation \( z \) loses its characteristics of being a measure of belonging to a certain region. One could circumvent this problem considering, for example, the phase difference between the activation \( z \) and a given direction in \( \mathbb{C} \) so that the output is essentially a phase difference between \( z \) and a fixed direction (a similar approach is taken in [8]). In this way one reintroduces the concept of measure since the phase difference is a real number that can be interpreted in this way, but at this point the output is no more fully complex and part of the utility of complex neurons get lost (see below).

### 4 The transfer function \( f(\cdot) \)

We focus now to the second and major problem one encounters when extending neurons to \( \mathbb{C} \) is that of the transfer function. If the activation \( z \) would be real there would be no problems but for a complex \( z \) here is a possible list of requests to be fulfilled: a suitable transfer function should be:
– defined on the whole complex plane,
– non-constant,
– non-linear,
– holomorphic on its domain, i.e. entire,
– bounded within the whole domain.

Already these list of items cannot be contemporarily fulfilled because of Liouville’s theorem that states that every bounded entire function must be constant. So we can take two roads: give up to analyticity or give up to boundedness.

4.1 The case of bounded, non constant, functions \( f() \)

We first consider the case of giving up to analyticity: in this case several authors took for \( z = u + iv \) \([10, 11, 4]\)

\[
y = f(z) = \tanh(u) + i \tanh(v)
\]

(1)

we show now that with this choice any complex flavour disappears. In particular we show that a complex neuron with this transfer function is identical to a net formed by 2 standard real neurons. In our notation we may write the activation \( z \) as

\[
z = \sum_{k=1}^{n} w_k x_k = \sum_{k=1}^{n} (x_k \Re w_k - x_k \Im w_k) + i(\Re x_k w_k + \Im x_k w_k) := u + iv
\]

(2)

from which we get that the output of this neuron is given by

\[
y = f(z) = \tanh\left(\sum_{k=1}^{n} x_k \Re w_k - x_k \Im w_k \right) + i \tanh\left(\sum_{k=1}^{n} x_k \Im w_k + x_k \Re w_k \right)
\]

and this output is identical to that produced by 2 standard real neurons that read the \( 2n \) inputs \( x : = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^{2n} \) with respective weights \( w_1 : = (w_\Re, -w_\Im) \) and \( w_2 : = (w_\Im, w_\Re) \) and standard real transfer function \( f(z) = \tanh(z) \). The two outputs provided by these two real neurons can be considered as the real and imaginary part of the complex neuron (1). It is also clear, as pointed out in [4], that the separation hyperplanes of these two real neurons in \( \mathbb{R}^{2n} \) are necessarily orthogonal since \( w_1 \cdot w_2 = 0 \).

It should be also not difficult to prove that the weights obtained by standard gradient descent learning in the two equivalent networks are identical. We can conclude that when we consider complex neural networks with transfer functions given by (1), or similar formulas, there is nothing to gain with respect to standard, real, neural networks.

4.2 The case of holomorphic functions \( f() \)

In this case the output of the neuron is given by

\[
y = f(z) = f(u + iv) := \alpha(u, v) + i\beta(u, v)
\]
and the Cauchy-Riemann analyticity conditions require
\[
\begin{align*}
\frac{\partial u}{\partial v} &= \frac{\partial \beta}{\partial \alpha} \\
\frac{\partial \beta}{\partial u} &= -\frac{\partial u}{\partial v}
\end{align*}
\]

We want to investigate what there is to gain by a complex neural network in this case and we will prove that the output of this neuron cannot be calculated with a network made with 2 standard neurons as in the previous paragraph. Let us suppose the contrary i.e. that there exist two real neurons such that
\[
\begin{align*}
y_1 &= \tanh(x \cdot w_1) = \alpha(u, v) \\
y_2 &= \tanh(x \cdot w_2) = \beta(u, v)
\end{align*}
\]
where \( x = (x_R, x_I) \in \mathbb{R}^{2n} \) and the different output can be provided only by the different weight vectors \( w_1, w_2 \in \mathbb{R}^{2n} \). Should this be true the Cauchy-Riemann analyticity conditions would apply to these hypothetical neurons
\[
\begin{align*}
\frac{\partial y_1}{\partial u} &= \frac{\partial y_2}{\partial v} \\
\frac{\partial y_2}{\partial u} &= -\frac{\partial y_1}{\partial v}
\end{align*}
\]
and we can expand the partial derivative by the chain rule to get, e.g. from the first relation,
\[
\frac{\partial y_1}{\partial u} = \tanh'(x \cdot w_1) \sum_{k=1}^{2n} \frac{\partial (x \cdot w_1)}{\partial x_k} \frac{\partial x_k}{\partial u} = \frac{\partial y_2}{\partial v} = \tanh'(x \cdot w_2) \sum_{k=1}^{2n} \frac{\partial (x \cdot w_2)}{\partial x_k} \frac{\partial x_k}{\partial v}
\]
in this equality the sums evaluate to some constant value depending on the various weights\(^1\). Thus this equation cannot be satisfied for all values of \( u \) and \( v \) because of the non-linear part introduced by \( \tanh'(x \cdot w_1) = \tanh'(x \cdot w_2) \) implies \( w_1 = w_2 \) that, in turn, implies the trivial solution \( y_1 = y_2 = 0 \). In other words the functions \( \alpha(u, v) \) and \( \beta(u, v) \) are harmonic conjugate that cannot be represented by normal neurons.

To obtain an arbitrarily precise approximation of the output of a complex neuron by means of a traditional neural network one can invoke the universal approximation capability of neural networks. One can see the output of a complex neuron \( \alpha(u, v) \) and \( \beta(u, v) \) as continuous real functions of \( \mathbb{R}^{2n} \to \mathbb{R} \), defined on \( x \in [-1, 1]^{2n} \) that can thus be approximated with arbitrary precision by a sufficiently large network of standard neurons. This is a good argument in favor of

\(^1\) For example (with obvious notation)
\[
\frac{\partial (x \cdot w_1)}{\partial x_k} = w_{1k}
\]
and, from (2)
\[
\frac{\partial u}{\partial x_k} = \begin{cases} 
\frac{\partial u}{\partial x_k} = w_{kR} & \text{for } k \leq n \\
-\frac{\partial u}{\partial x_k} = w_{kI} & \text{for } k > n
\end{cases}
\]
the use of complex neurons since it indicates that the output of a single complex neuron can easily convey the information of a bunch of traditional neurons. The capabilities of neuron with holomorphic transfer function surely deserve further studies to assess its possibilities and to understand its shortcomings, like for example singularities, a matter of life for holomorphic functions.

5 Conclusions

In summary we have shown that:

- complex neurons with transfer function (1) are identical to 2 standard real neurons in parallel,
- complex neurons with holomorphic transfer function are much more promising, as already pointed out in [12, 5],
- there is considerable room for a better understanding of the definition of the activation $z$ and of its intimate relation with the transfer function; e.g.;
- there could be possible interesting surprises in introducing the use of null subspaces (and tightly connected Clifford algebras) in the activation $z$.

References