Four Models of Hopfield-Type Octonion Neural Networks and Their Existing Conditions of Energy Functions

1st Yasuaki Kuroe
Faculty of Engeneering Science
Kansai University
Suita, Japan
kuroe@ieee.org

2nd Hitoshi Iima
Faculty of Information and Human Sciences
Kyoto Institute of Technology
Kyoto, Japan
iima@kit.ac.jp

3rd Yutaka Maeda
Faculty of Engeneering Science
Kansai University
Suita, Japan
maedayut@kansai-u.ac.jp

***Abstract*—Recently, models of neural networks in the real domain have been extended into the high dimensional domain such as the complex number and quaternion domain, and several high-dimensional models have been proposed. These extensions are generalized by introducing Clifford algebra (geometric algebra). In this paper we extend conventional real-valued Hopfield-type neural networks into the octonion domain and discuss their dynamics. The octonions represent a particular extension of the quaternions which also represent a particular extension of the complex numbers and have 7 imaginary parts. They are non-commutative and non-associative on multiplication and do not belong to Clifford algebra due to the latter fact. With this in mind we propose four models of octonion Hopfield-type neural networks. We derive existence conditions of an energy function and construct energy function for each model.**

***Index Terms*—Hopfield neural network, octonion neural network, energy function, existing condition**

I. Introduction

In recent years, there have been increasing research interests of artificial neural networks and many efforts have been made on applications of neural networks to various fields. As applications of the neural networks spread more widely, developing neural network models which can directly deal with complex numbers is desired in various fields. Several models of complex-valued neural networks have been proposed and their abilities of information processing have been investigated [1], [2]. Furthermore those studies are extended into the quaternion numbers domain, and models of quaternion neural networks are proposed and actively studied [2], [12]. These extensions are generalized [3] by introducing Clifford algebra (also called geometric algebra) [4]–[6].

It is well known that one of the pioneering works that triggered the research interests of neural networks in the last three decades is a proposal of models for neural networks by Hopfield [7]–[9]. He introduced the idea of an energy function to formulate a way of understanding the computation performed by fully connected recurrent neural networks and showed that a combinatorial optimization problem can be solved by them. The energy functions have been applied to various problems such as qualitative analysis of neural networks, synthesis of associative memories, optimization problems and so on ever since.

The extensions of the Hopfield-type neural networks to the complex domain and the quaternion domain have been studied. The existence condition of an energy function was derived for the complex-valued Hopfield-type neural networks [10], [11] and for quaternion Hopfield-type neural networks [12]. Those studies were also extended into the domains of the dual numbers and the hyperbolic numbers [13]. In all those studies the application of the energy functions to qualitative analysis of the Hopfield-type neural networks was also discussed. All the domains into which the real valued Hopfield-type neural networks were extended so far, that is, the complex, dual, hyperbolic numbers and the quaternions belong to Clifford algebra [4]–[6], [14].

In this paper we extend conventional real-valued neural networks into the octonion domain. The octonions represent a particular extension of the quaternions which also represent a particular extension of the complex numbers, and have 7 imaginary parts. They are non-commutative and non-associative on multiplication and do not belong to Clifford algebra due to the latter fact. There have been various attempts to find applications for the octonions mainly in geometry and physics [15]–[17] and they are expected to be applicable to high dimensional signal processing. Some studies on octonion neural networks, whose inputs, outputs, weights and biases are all octonions, also have been done [18]–[20]. C.-A. Popa presented the gradient descent algorithm for training octonion feedforward neural networks [19]. In [20] C.-A. Popa studied the stability analysis of neutral-type octonion neural networks with time varying delays.

This paper presents models of fully connected recurrent neural networks, which are extensions of the real-valued Hopfield-type neural networks to the octonions and discuss dynamics of those models from the point view of existence of an energy function. We have already proposed a model of Hopfield-type octonion neural networks and derived the existing condition of energy functions for it [18]. Due to the fact that the octonions are non-commutative and non-

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

associative on multiplication, a couple of different models can be considered. In this paper, with this in mind, we propose four models of Hopfield-type octonion neural networks. We also derive the existence conditions of energy functions for each of them and construct an energy function for each model. Similar to the real-valued ones, the energy functions enable us to analyze qualitative behaviors of the recurrent octonion neural networks and to apply to various problems such as synthesis of associative memories, optimization problems and so on.

II. Octonions

The octonions, which we denote by $\mathbb{O}$, are an 8-dimensional algebra with basis

$$\{1, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_5, \mathbf{e}_6, \mathbf{e}_7\}$$

and their multiplication is given in Table I, which describes the result of multiplying the element in the ith row by the element in the jth column [15]. An octonion number $x \in \mathbb{O}$ is described by

$$\begin{aligned} x =& x^{(0)} + \mathbf{e}_1 x^{(1)} + \mathbf{e}_2 x^{(2)} + \mathbf{e}_3 x^{(3)} \\ &+ \mathbf{e}_4 x^{(4)} + \mathbf{e}_5 x^{(5)} + \mathbf{e}_6 x^{(6)} + \mathbf{e}_7 x^{(7)} \end{aligned} \tag{1}$$

where $x^{(0)}, x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(5)}, x^{(6)}, x^{(7)}$ are real numbers. It is found from Table I that the octonions are non-

TABLE I
Octonion multiplication table

	1	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_6$	$\mathbf{e}_7$
1	1	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_6$	$\mathbf{e}_7$
$\mathbf{e}_1$	$\mathbf{e}_1$	-1	$\mathbf{e}_4$	$\mathbf{e}_7$	$-\mathbf{e}_2$	$\mathbf{e}_6$	$-\mathbf{e}_5$	$-\mathbf{e}_3$
$\mathbf{e}_2$	$\mathbf{e}_2$	$-\mathbf{e}_4$	-1	$\mathbf{e}_5$	$\mathbf{e}_1$	$-\mathbf{e}_3$	$\mathbf{e}_7$	$-\mathbf{e}_6$
$\mathbf{e}_3$	$\mathbf{e}_3$	$-\mathbf{e}_7$	$-\mathbf{e}_5$	-1	$\mathbf{e}_6$	$\mathbf{e}_2$	$-\mathbf{e}_4$	$\mathbf{e}_1$
$\mathbf{e}_4$	$\mathbf{e}_4$	$\mathbf{e}_2$	$-\mathbf{e}_1$	$-\mathbf{e}_6$	-1	$\mathbf{e}_7$	$\mathbf{e}_3$	$-\mathbf{e}_5$
$\mathbf{e}_5$	$\mathbf{e}_5$	$-\mathbf{e}_6$	$\mathbf{e}_3$	$-\mathbf{e}_2$	$-\mathbf{e}_7$	-1	$\mathbf{e}_1$	$\mathbf{e}_4$
$\mathbf{e}_6$	$\mathbf{e}_6$	$\mathbf{e}_5$	$-\mathbf{e}_7$	$\mathbf{e}_4$	$-\mathbf{e}_3$	$-\mathbf{e}_1$	-1	$\mathbf{e}_2$
$\mathbf{e}_7$	$\mathbf{e}_7$	$\mathbf{e}_3$	$\mathbf{e}_6$	$-\mathbf{e}_1$	$\mathbf{e}_5$	$-\mathbf{e}_4$	$-\mathbf{e}_2$	-1

commutative on multiplication:

$$\mathbf{e}_i \mathbf{e}_j = -\mathbf{e}_j \mathbf{e}_i \neq \mathbf{e}_j \mathbf{e}_i$$

for $i \neq j$, non-associative on multiplication:

$$(\mathbf{e}_i \mathbf{e}_j)\mathbf{e}_k = -\mathbf{e}_i(\mathbf{e}_j \mathbf{e}_k) \neq \mathbf{e}_i(\mathbf{e}_j \mathbf{e}_k)$$

for $i \neq j \neq k$, and the octonion $\mathbb{O}$ does not belong to the Clifford algebra because of the latter fact.

$x^{(0)}$ of (1) is called real part, and is represented by $Re(x)$. An octonion whose real part is equal to zero is called pure octonion. The octonion conjugate x^* is defined by

$$\begin{aligned} x^* =& x^{(0)} - \mathbf{e}_1 x^{(1)} - \mathbf{e}_2 x^{(2)} - \mathbf{e}_3 x^{(3)} \\ &- \mathbf{e}_4 x^{(4)} - \mathbf{e}_5 x^{(5)} - \mathbf{e}_6 x^{(6)} - \mathbf{e}_7 x^{(7)}. \end{aligned} \tag{2}$$

The norm of an octonion number $|x|$ is defined by

$$|x|^2 = x^* x = \sum_{i=0}^{7} {x^{(i)}}^2. \tag{3}$$

III. Four Models of Hopfield-type Octonion Neural Networks

In this section we propose models of fully connected recurrent neural networks, which are extensions of real valued continuous-time Hopfield neural networks into the octonion domain. Since the octonions are non-commutative and non-associative on multiplication, a couple of different models can be considered. We present four models of Hopfield type octonion neural networks.

The first model is a direct extension of the Hopfield neural networks, described by differential equations of the form [18]:
Model 1:

$$\begin{cases} \tau_i \dfrac{du_i}{dt} &= -u_i + \displaystyle\sum_{j=1}^{n} w_{ij} v_j + b_i \\ \quad v_i &= f(u_i) \quad (i = 1, 2, ..., n) \end{cases} \tag{4}$$

where n is the number of neurons, τ_i is the time constant of the ith neuron, u_i and v_i are the state and the output of the ith neuron at time t, respectively, b_i is the threshold value, w_{ij} is the connection weight coefficient from the jth neuron to the ith one, and $f(\cdot)$ is the activation function of the neurons. In the model u_i, v_i, b_i and w_{ij} are all octonions: $u_i \in \mathbb{O}$, $v_i \in \mathbb{O}$, $b_i \in \mathbb{O}$ and $w_{ij} \in \mathbb{O}$. The time constant τ_i is a positive real number: $\tau_i \in \mathbb{R}, \tau_i > 0$. The product $w_{ij} v_j$ is performed according to the octonion multiplication table shown in Table I. The activation function $f(\cdot)$ is a nonlinear function which maps from an octonion to an octonion: $f : \mathbb{O} \to \mathbb{O}$, and

$$\frac{du_i}{dt} := \frac{d}{dt} u_i^{(0)} + \sum_{j=1}^{7} \mathbf{e}_j \frac{d}{dt} u_i^{(j)}.$$

Note that the neural network described by (4) is a direct octonion-domain extension of the real-valued continuous-time neural network of Hopfield type.

Since the octonions are non-commutative on multiplication, the model in which the product $w_{ij} v_j$ in the model (4) is replaced by $v_j w_{ij}$ is a different model. As the second model we consider the model which is described by differential equations of the form:
Model 2:

$$\begin{cases} \tau_i \dfrac{du_i}{dt} &= -u_i + \displaystyle\sum_{j=1}^{n} v_j w_{ij} + b_i \\ \quad v_i &= f(u_i) \quad (i = 1, 2, ..., n) \end{cases}. \tag{5}$$

Noting that in Model 1 (4) the signal v_i is weighted from the left hand side and in Model 2 (5) the signal v_i is weighted from the right side, the other models in which the signal v_i is weighted from both sides can be considered. Letting the weight by which v_i is multiplied from the left be w_{ij}^{ℓ} and from the right be w_{ij}^{r}, such models are obtained by replacing $w_{ij} v_j$ in (4) or $v_j w_{ij}$ in (5) by $w_{ij}^{\ell} v_j w_{ij}^{r}$. Note that, since the octonions are non-associative on multiplication, $w_{ij}^{\ell}(v_j w_{ij}^{r})$ and $(w_{ij}^{\ell} v_j) w_{ij}^{r}$ are different. We consider the additional following two models.

Model 3:

$$\begin{cases} \tau_i \dfrac{du_i}{dt} & = -u_i + \displaystyle\sum_{j=1}^{n} w_{ij}^{\ell}(v_j w_{ij}^{r}) + b_i \\ \quad v_i & = f(u_i) \quad (i = 1, 2, ..., n) \end{cases} . \tag{6}$$

Model 4:

$$\begin{cases} \tau_i \dfrac{du_i}{dt} & = -u_i + \displaystyle\sum_{j=1}^{n} (w_{ij}^{\ell} v_j) w_{ij}^{r} + b_i \\ \quad v_i & = f(u_i) \quad (i = 1, 2, ..., n) \end{cases} . \tag{7}$$

IV. Definition of Energy Functions

We are now in the position to give the definition of energy functions for the octonion neural networks Model 1 (4), Model 2 (5), Model 3 (6) and Model 4 (7). If the neural network of Model 1 (4) is real valued, that is, u_i, v_i, b_i and w_{ij} are all real, $u_i \in \mathbb{R}$, $v_i \in \mathbb{R}$, $b_i \in \mathbb{R}$, $w_{ij} \in \mathbb{R}$ and the activation function is a real nonlinear function $f : \mathbb{R} \to \mathbb{R}$, the existence condition of an energy function which Hopfield et al. obtained is that the weight matrix $W = \{w_{ij}\}$ is a symmetric matrix ($w_{ij} = w_{ji}$) and the activation function is continuously differentiable, bounded and monotonically increasing. The following function $E : \mathbb{R}^n \to \mathbb{R}$ was proposed as an energy function for the network.

$$\begin{aligned} E(\boldsymbol{v}) \quad &= \quad -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} v_i v_j - \sum_{i=1}^{n} b_i v_i \\ &\quad + \sum_{i=1}^{n} \int_{0}^{v_i} f^{-1}(\rho) d\rho \end{aligned} \tag{8}$$

where $\boldsymbol{v} = [v_1, v_2, \cdots, v_n]^T \in \mathbb{R}^n$ and f^{-1} is the inverse function of f. Hopfield et al. showed that, if the existence conditions hold, the network (4) has the function $E(\boldsymbol{v})$ and it has the following property; the time derivative of E along the trajectories of (4), denoted by $\frac{dE}{dt}\big|_{(4)^R}$ is less or equal to 0, $\frac{dE}{dt}\big|_{(4)^R} \le 0$, and furthermore $\frac{dE}{dt}\big|_{(4)^R} = 0$ if and only if $\frac{dv_i}{dt} = 0$ ($i = 1, 2, \cdots, n$).

We define an energy function for the octonion neural networks of Model 1 (4), Model 2 (5), Model 3 (6) and Model 4 (7), by the analogy to that for Hopfield type real-valued neural networks as follows.

Definition 1: Consider the octonion neural network ($\mathcal{N}$) where $\mathcal{N}$ is the equation number, 4, 5, 6 or 7. E is an energy function of the octonion neural network ($\mathcal{N}$), if the following conditions are satisfied.

(i) $E(\cdot)$ is a mapping $E : \mathbb{O} \to \mathbb{R}$.

(ii) The derivative of E along the trajectories of the network ($\mathcal{N}$), denoted by $\frac{dE}{dt}\big|_{(\mathcal{N})}$, satisfies $\frac{dE}{dt}\big|_{(\mathcal{N})} \le 0$. Furthermore, $\frac{dE}{dt}\big|_{(\mathcal{N})} = 0$ if and only if $\frac{dv_i}{dt} = 0$ ($i = 1, 2, \cdots, n$).

V. Existence Conditions of Energy Functions

A. Octonion Activation Function

One of the important factors to characterize dynamics of recurrent neural networks is their activation functions which are nonlinear functions. It is therefore, important to discuss which type of nonlinear functions is chosen as activation functions for the octonion neural networks of Model 1 (4), Model 2 (5), Model 3 (6) and Model 4 (7). In the real-valued neural networks, the activation is usually chosen to be a smooth and bounded function such as a sigmoidal function. Recall that, in the complex domain, the Liouvill's theorem says that 'if $f(\cdot)$ is analytic at all points of the complex plane and bounded, then $f(\cdot)$ is constant'. Since a suitable $f(\cdot)$ should be bounded, it follows from the theorem that if we choose an analytic function for $f(\cdot)$, it is constant over the entire complex plane, which is clearly not suitable. In the complex-valued neural networks in [10], [11], in place of analytic function, a function whose real and imaginary parts are continuously differentiable with respect to the real and imaginary variables of its argument, respectively, is chosen for the activation function and the existence conditions of an energy function are derived.

In this paper, according to the discussion on the activation function of the complex-valued neural networks [10], [11], we choose a function which satisfies the following conditions as the activation function for the octonion neural networks of Model 1 (4), Model 2 (5), Model 3 (6) and Model 4 (7).

Let us express the nonlinear octonion function $f(u) : \mathbb{O} \to \mathbb{O}$ as:

$$f(u) \quad = \quad f^{(0)}(u^{(0)}, u^{(1)}, \cdots, u^{(7)}) \tag{9}$$

$$\quad + \quad \sum_{i=1}^{7} \mathbf{e}_i f^{(i)}(u^{(0)}, u^{(1)}, \cdots, u^{(7)}) \tag{10}$$

where

$$u = u^{(0)} + \sum_{i=1}^{7} \mathbf{e}_i u^{(i)} \tag{11}$$

and $f^{(i)}$, $i = 0, 1, 2, 3, 4, 5, 6, 7$ is a real function: $f^{(i)} : \mathbb{R}^8 \to \mathbb{R}$. We assume the following conditions on the activation function $f(u) : \mathbb{O} \to \mathbb{O}$ of the octonion neural networks of Model 1 (4), Model 2 (5), Model 3 (6) and Model 4 (7).

(i) $f^{(l)}(\cdot), (l = 0, 1, \cdots, 7)$ are continuously differentiable with respect to $u^{(m)}, (m = 0, 1, \cdots, 7)$.

(ii) $f(\cdot)$ is a bounded function, that is, there exists some $M > 0$ such that $|f(\cdot)| \le M$.

From this assumption, we can define the Jacobian matrix of the activation function f at a point u, denoted by $\boldsymbol{J}_f(u) = \{\alpha_{lm}(u)\} \in \mathbb{R}^{8\times 8}$ where

$$\alpha_{lm}(u) = \left.\frac{\partial f^{(l)}}{\partial u^{(m)}}\right|_{u} . \tag{12}$$

B. Derivation of Existence Conditions

We now discuss existence conditions of the energy functions for the octonion neural networks of Model 1 (4), Model 2 (5), Model 3 (6) and Model 4 (7). We need the following assumptions on the activation function.

Assumption 1: The activation function f satisfies

(i) f is an injective function,

(ii) $\boldsymbol{J}_f(u)$ is a symmetric matrix for all $u \in \mathbb{O}$,
(iii) $\boldsymbol{J}_f(u)$ is positive definite for all $u \in \mathbb{O}$.

Because of the condition (i) of Assumption 1 and boundedness of f, there exists the inverse function of f, denoted by $g = f^{-1}$. We express g as $u = g(v)$:

$$g(v) = g^{(0)}(v^{(0)}, v^{(1)}, \cdots, v^{(7)}) + \sum_{i=1}^{7} \mathbf{e}_i g^{(i)}(v^{(0)}, v^{(1)}, \cdots, v^{(7)}) \qquad (13)$$

where $g^{(l)} : \mathbb{R}^8 \to \mathbb{R}$ $(l = 0, 1, \cdots, 7)$. Then, the following lemma holds.

Lemma 1: If f satisfies Assumption 1, there exists a scalar function $G(v) : \mathbb{O} \to \mathbb{R}$ such that

$$\frac{\partial G}{\partial v^{(l)}} = g^{(l)}(v^{(0)}, v^{(1)}, \cdots, v^{(7)}) \qquad (l = 0, 1, \cdots, 7). \quad (14)$$

Proof: We define the Jacobian matrix of g at v by $\boldsymbol{J}_g(v) = \{\beta_{lm}\} \in \mathbb{R}^{8\times 8}$ where $\beta_{lm} = \partial g^{(l)}/\partial v^{(m)}$ By partially differentiating both sides of the equations $u^{(l)} = g^{(l)}(v^{(0)}, v^{(1)}, \cdots, v^{(7)})$ with respect to $u^{(m)}$ $(l, m = 0, 1, \cdots, 7)$, the relation $\boldsymbol{I} = \boldsymbol{J}_g(v)\boldsymbol{J}_f(u)$ is obtained for all u, where $\boldsymbol{I} \in \mathbb{R}^{8\times 8}$ is identity matrix. From this relation and the conditions (ii) and (iii) of Assumption 1, the fact $\boldsymbol{J}_g(v) = \{\boldsymbol{J}_f(u)^{-1}\}^t = \boldsymbol{J}_g(v)^t$ holds. Therefore we have

$$\left.\frac{\partial g^{(l)}}{\partial v^{(m)}}\right|_v = \left.\frac{\partial g^{(m)}}{\partial v^{(l)}}\right|_v \qquad (l, m = 0, 1, \cdots, 7) \qquad (15)$$

for all $v \in \mathbb{O}$. Let us define a function G by

$$\begin{aligned} G(v) = &\int_0^{v^{(0)}} g^{(0)}(\rho, 0, 0, 0, 0, 0, 0, 0)d\rho \\ &+ \int_0^{v^{(1)}} g^{(1)}(v^{(0)}, \rho, 0, 0, 0, 0, 0, 0)d\rho \\ &+ \int_0^{v^{(2)}} g^{(2)}(v^{(0)}, v^{(1)}, \rho, 0, 0, 0, 0, 0)d\rho \\ &+ \int_0^{v^{(3)}} g^{(3)}(v^{(0)}, v^{(1)}, v^{(2)}, \rho, 0, 0, 0, 0)d\rho \\ &+ \int_0^{v^{(4)}} g^{(4)}(v^{(0)}, v^{(1)}, v^{(2)}, v^{(3)}, \rho, 0, 0, 0)d\rho \\ &+ \int_0^{v^{(5)}} g^{(5)}(v^{(0)}, v^{(1)}, v^{(2)}, v^{(3)}, v^{(4)}, \rho, 0, 0)d\rho \\ &+ \int_0^{v^{(6)}} g^{(6)}(v^{(0)}, v^{(1)}, v^{(2)}, v^{(3)}, v^{(4)}, v^{(5)}, \rho, 0)d\rho \\ &+ \int_0^{v^{(7)}} g^{(7)}(v^{(0)}, v^{(1)}, v^{(2)}, v^{(3)}, v^{(4)}, v^{(5)}, v^{(6)}, \rho)d\rho. \end{aligned} \qquad (16)$$

It can be easily shown by using (15) that the function G satisfies the equations (14). ■

In the followings we derive the existing conditions of energy functions of the octonion neural networks and construct energy functions by using the function G.

The following theorem holds for the octonion neural networks of Model 1 (4) and Model 2 (5).

Theorem 1: Consider the octonion neural networks of Model 1 (4) and Model 2 (5). If the weight coefficients w_{ij} satisfy

$$w_{ji} = w_{ij}^* \quad (i, j = 1, 2, \cdots, n), \qquad (17)$$

and the activation function f satisfies Assumption 1, then there exist the energy functions for them. The energy functions are constructed as:

$$E(\boldsymbol{v}) = -\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\frac{1}{2}Re(v_i^* w_{ij} v_j + 2b_i^* v_i) - G(v_i)\right\} \quad (18)$$

for the network of Model 1 (4) and

$$E(\boldsymbol{v}) = -\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\frac{1}{2}Re(v_i^* v_j w_{ij} + 2b_i^* v_i) - G(v_i)\right\} \quad (19)$$

for the network of Model 2 (5).

Remark: 1: Since the octonions are non-associative on multiplication, the order of the multiplications of $v_i^* w_{ij} v_j$ in (18) ($v_i^* v_j w_{ij}$ in (19)) should be specified: $v_i^*(w_{ij}v_j)$ or $(v_i^* w_{ij})v_j$ ($v_i^*(v_j w_{ij})$ or $(v_i^* v_j)w_{ij}$). However it can be shown that the equality $Re(v_i^*(w_{ij}v_j)) = Re((v_i^* w_{ij})v_j)$ ($Re(v_i^*(v_j w_{ij})) = Re((v_i^* v_j)w_{ij})$) holds, and hence the order of the multiplications is not specified in (18) (in (19)).

Proof: First we prove the existence of the energy function for the network of Model 1 (4) by showing that the function given by (18) satisfies the definition of the energy function (Definition 1). Let us define the gradient operator in the octonion domain as:

$$\nabla_{v_i} = \frac{d}{dv_i^{(0)}} + \sum_{j=1}^{7} \mathbf{e}_j \frac{d}{dv_i^{(j)}}. \qquad (20)$$

Let $\hat{\mathbf{v}}_i$ be $\hat{\mathbf{v}}_i = [v_i^{(0)}, v_i^{(1)}, \cdots, v_i^{(7)}]^t \in \mathbb{R}^8$. Under the assumption (17) on the weight coefficients w_{ij}, the gradient of the energy function E given by (18) with respect to $\hat{\mathbf{v}}_i$, denoted by $\nabla_{\hat{\mathbf{v}}_i}E(v)$, can be calculated as follows:

$$\begin{aligned} \nabla_{\hat{\mathbf{v}}_i}E(\boldsymbol{v}) &= -(-\hat{\mathbf{u}}_i + \sum_{j=1}^{n} \hat{\mathbf{W}}_{ij}^1 \hat{\mathbf{v}}_j + \hat{\mathbf{b}}_i) \\ &= -\tau_i \frac{d\hat{\mathbf{u}}_i}{dt} \end{aligned} \qquad (21)$$

where $\hat{\mathbf{u}}_i = [u_i^{(0)}, u_i^{(1)}, \cdots, u_i^{(7)}]^t \in \mathbb{R}^8$, $\hat{\mathbf{b}}_i =$

$[b_i^{(0)}, b_i^{(1)}, \cdots, b_i^{(7)}]^t \in \mathbb{R}^8$ and

$$\hat{\mathbf{W}}_{ij}^1 = \begin{bmatrix} w_{ij}^{(0)} & -w_{ij}^{(1)} & -w_{ij}^{(2)} & -w_{ij}^{(3)} & -w_{ij}^{(4)} & -w_{ij}^{(5)} & -w_{ij}^{(6)} & -w_{ij}^{(7)} \\ w_{ij}^{(1)} & w_{ij}^{(0)} & -w_{ij}^{(4)} & -w_{ij}^{(7)} & w_{ij}^{(2)} & -w_{ij}^{(6)} & w_{ij}^{(5)} & w_{ij}^{(3)} \\ w_{ij}^{(2)} & w_{ij}^{(4)} & w_{ij}^{(0)} & -w_{ij}^{(5)} & -w_{ij}^{(1)} & w_{ij}^{(3)} & -w_{ij}^{(7)} & w_{ij}^{(6)} \\ w_{ij}^{(3)} & w_{ij}^{(7)} & w_{ij}^{(5)} & w_{ij}^{(0)} & -w_{ij}^{(6)} & -w_{ij}^{(2)} & w_{ij}^{(4)} & -w_{ij}^{(1)} \\ w_{ij}^{(4)} & -w_{ij}^{(2)} & w_{ij}^{(1)} & w_{ij}^{(6)} & w_{ij}^{(0)} & -w_{ij}^{(7)} & -w_{ij}^{(3)} & w_{ij}^{(5)} \\ w_{ij}^{(5)} & w_{ij}^{(6)} & -w_{ij}^{(3)} & w_{ij}^{(2)} & w_{ij}^{(7)} & w_{ij}^{(0)} & -w_{ij}^{(1)} & -w_{ij}^{(4)} \\ w_{ij}^{(6)} & -w_{ij}^{(5)} & w_{ij}^{(7)} & -w_{ij}^{(4)} & w_{ij}^{(3)} & w_{ij}^{(1)} & w_{ij}^{(0)} & -w_{ij}^{(2)} \\ w_{ij}^{(7)} & -w_{ij}^{(3)} & -w_{ij}^{(6)} & w_{ij}^{(1)} & -w_{ij}^{(5)} & w_{ij}^{(4)} & w_{ij}^{(2)} & w_{ij}^{(0)} \end{bmatrix}.$$

Thus the gradient of the function E given by (18), $\nabla_{v_i} E(\boldsymbol{v})$, is obtained as follows.

$$\begin{aligned} \nabla_{v_i} E(\boldsymbol{v}) &= -(-u_i + \sum_{j=1}^{n} w_{ij} v_j + b) \\ &= -\tau_i \frac{du_i}{dt}. \end{aligned} \tag{22}$$

By using (22) the derivative of the energy function E given by (18) along the trajectories of the network of Model 1 (4) is calculated as follows.

$$\begin{aligned} \left.\frac{dE(\boldsymbol{v})}{dt}\right|_{(4)} &= \sum_{i=1}^{n} \sum_{l=0}^{7} \frac{dE}{dv_i^{(l)}} \frac{dv_i^{(l)}}{dt} \\ &= Re\{\sum_{i=1}^{n} \nabla_{v_i} E(\boldsymbol{v})^* \frac{dv_i}{dt}\}. \end{aligned} \tag{23}$$

Substituting (22) into the right side of the above equation, we have

$$\begin{aligned} \left.\frac{dE(\boldsymbol{v})}{dt}\right|_{(4)} &= -\sum_{i=1}^{n} (\frac{du_i^{(0)}}{dt} \tau_i \frac{dv_i^{(0)}}{dt} + \frac{du_i^{(1)}}{dt} \tau_i \frac{dv_i^{(1)}}{dt} + \cdots \\ &\quad + \frac{du_i^{(6)}}{dt} \tau_i \frac{dv_i^{(6)}}{dt} + \frac{du_i^{(7)}}{dt} \tau_i \frac{dv_i^{(7)}}{dt}) \\ &= -\sum_{i=1}^{n} ((\sum_{m=0}^{7} \frac{dg^{(0)}(v_i)}{dv_i^{(m)}} \frac{dv_i^{(m)}}{dt}) \tau_i \frac{dv_i^{(0)}}{dt} \\ &\quad + (\sum_{m=0}^{7} \frac{dg^{(1)}(v_i)}{dv_i^{(m)}} \frac{dv_i^{(m)}}{dt}) \tau_i \frac{dv_i^{(1)}}{dt} \\ &\quad + \ldots \\ &\quad + (\sum_{m=0}^{6} \frac{dg^{(6)}(v_i)}{dv_i^{(m)}} \frac{dv_i^{(m)}}{dt}) \tau_i \frac{dv_i^{(6)}}{dt} \\ &\quad + (\sum_{m=0}^{7} \frac{dg^{(7)}(v_i)}{dv_i^{(m)}} \frac{dv_i^{(m)}}{dt}) \tau_i \frac{dv_i^{(7)}}{dt}) \\ &= -\sum_{i=1}^{n} (\frac{d\hat{\mathbf{v}}_i}{dt})^t \tau_i \boldsymbol{J}_g(v_i)^t (\frac{d\hat{\mathbf{v}}_i}{dt}) \end{aligned} \tag{24}$$

where $\boldsymbol{J}_g(v_i)$ is the Jacobian matrix of the function $g(\cdot)$ with respect to v_i.

Since $\tau_i > 0$ for all i and $\boldsymbol{J}_g(v_i)$ is positive definite for any v_i $(i = 1, 2, \cdots, n)$, the condition $\frac{dE}{dt}\big|_{(4)} \leq 0$ holds, and furthermore $\frac{dE}{dt}\big|_{(4)} = 0$ if and only if $\frac{d\hat{\mathbf{v}}}{dt} = \mathbf{0} \Leftrightarrow \frac{dv_i}{dt} = 0$. Hence the function E satisfies the definition of energy functions (Definition 1).

The existence of the energy function for the network of Model 2 (5) can be proved as follows. Under the assumption (17) on the weight coefficients w_{ij}, the gradient of the function E given by (19) with respect to $\hat{\mathbf{v}}_i$ can be calculated as follows:

$$\begin{aligned} \nabla_{\hat{\mathbf{v}}_i} E(\boldsymbol{v}) &= -(-\hat{\mathbf{u}}_i + \sum_{j=1}^{n} \hat{\mathbf{W}}_{ij}^2 \hat{\mathbf{v}}_j + \hat{\mathbf{b}}_i) \\ &= -\tau_i \frac{d\hat{\mathbf{u}}_i}{dt} \end{aligned} \tag{25}$$

where

$$\hat{\mathbf{W}}_{ij}^2 = \begin{bmatrix} w_{ij}^{(0)} & -w_{ij}^{(1)} & -w_{ij}^{(2)} & -w_{ij}^{(3)} & -w_{ij}^{(4)} & -w_{ij}^{(5)} & -w_{ij}^{(6)} & -w_{ij}^{(7)} \\ w_{ij}^{(1)} & w_{ij}^{(0)} & w_{ij}^{(4)} & w_{ij}^{(7)} & -w_{ij}^{(2)} & w_{ij}^{(6)} & -w_{ij}^{(5)} & -w_{ij}^{(3)} \\ w_{ij}^{(2)} & -w_{ij}^{(4)} & w_{ij}^{(0)} & w_{ij}^{(5)} & w_{ij}^{(1)} & -w_{ij}^{(3)} & w_{ij}^{(7)} & -w_{ij}^{(6)} \\ w_{ij}^{(3)} & -w_{ij}^{(7)} & -w_{ij}^{(5)} & w_{ij}^{(0)} & w_{ij}^{(6)} & w_{ij}^{(2)} & -w_{ij}^{(4)} & w_{ij}^{(1)} \\ w_{ij}^{(4)} & w_{ij}^{(2)} & -w_{ij}^{(1)} & -w_{ij}^{(6)} & w_{ij}^{(0)} & w_{ij}^{(7)} & w_{ij}^{(3)} & -w_{ij}^{(5)} \\ w_{ij}^{(5)} & -w_{ij}^{(6)} & w_{ij}^{(3)} & -w_{ij}^{(2)} & -w_{ij}^{(7)} & w_{ij}^{(0)} & w_{ij}^{(1)} & w_{ij}^{(4)} \\ w_{ij}^{(6)} & w_{ij}^{(5)} & -w_{ij}^{(7)} & w_{ij}^{(4)} & -w_{ij}^{(3)} & -w_{ij}^{(1)} & w_{ij}^{(0)} & w_{ij}^{(2)} \\ w_{ij}^{(7)} & w_{ij}^{(3)} & w_{ij}^{(6)} & -w_{ij}^{(1)} & w_{ij}^{(5)} & -w_{ij}^{(4)} & -w_{ij}^{(2)} & w_{ij}^{(0)} \end{bmatrix}.$$

Note that the equation (25) becomes equal to (21) by replacing $\hat{\mathbf{W}}_{ij}^2$ in (25) by $\hat{\mathbf{W}}_{ij}^1$ in (21). By using the fact the existence condition of the energy functions for Model 2 can be proved in the similar way to that for Model 1. ■

The following theorem holds for the octonion neural networks of Model 3 (6) and Model 4 (7).

Theorem 2: Consider the octonion neural networks of Model 3 (6) and Model 4 (7). Assume that the weight coefficients w_{ij}^{ℓ} and w_{ij}^{r} satisfy ${w_{ij}^{\ell}}^* = w_{ij}^{r}$ and we rewrite them as

$$w_{ij} := {w_{ij}^{\ell}}^* = w_{ij}^{r}.$$

If the weight coefficients w_{ij} satisfy the condition (17) and the activation function f satisfies Assumption 1, then there exist the energy functions for them. The energy functions are constructed as:

$$E(\boldsymbol{v}) = -\sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ \frac{1}{2} Re(v_i^* (w_{ij}^* (v_j w_{ij})) + 2 b_i^* v_i) - G(v_i) \right\} \tag{26}$$

for the network of Model 3 (6) and

$$E(\boldsymbol{v}) = -\sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ \frac{1}{2} Re(v_i^* ((w_{ij}^* v_j) w_{ij}) + 2 b_i^* v_i) - G(v_i) \right\} \tag{27}$$

for the network of Model 4 (7).

Proof: The existence of the energy function for the network of Model 3 (6) can be proved as follows. Under the assumption (17) on the weight coefficients $w_{ij} := {w^{\ell}_{ij}}^* = w^r_{ij}$, the gradient of the function E given by (26) with respect to $\hat{\mathbf{v}}_i$ can be calculated as follows:

$$\begin{aligned} \nabla_{\hat{\mathbf{v}}_i} E(\boldsymbol{v}) &= -(-\hat{\mathbf{u}}_i + \sum_{j=1}^{n} \hat{\mathbf{W}}^3_{ij} \hat{\mathbf{W}}^2_{ij} \hat{\mathbf{v}}_j + \hat{\mathbf{b}}_i) \\ &= -\tau_i \frac{d\hat{\mathbf{u}}_i}{dt} \end{aligned} \tag{28}$$

where

$$\hat{\mathbf{W}}^3_{ij} = \begin{bmatrix} w^{(0)}_{ij} & w^{(1)}_{ij} & w^{(2)}_{ij} & w^{(3)}_{ij} & w^{(4)}_{ij} & w^{(5)}_{ij} & w^{(6)}_{ij} & w^{(7)}_{ij} \\ -w^{(1)}_{ij} & w^{(0)}_{ij} & w^{(4)}_{ij} & w^{(7)}_{ij} & -w^{(2)}_{ij} & w^{(6)}_{ij} & -w^{(5)}_{ij} & -w^{(3)}_{ij} \\ -w^{(2)}_{ij} & -w^{(4)}_{ij} & w^{(0)}_{ij} & w^{(5)}_{ij} & w^{(1)}_{ij} & -w^{(3)}_{ij} & w^{(7)}_{ij} & -w^{(6)}_{ij} \\ -w^{(3)}_{ij} & -w^{(7)}_{ij} & -w^{(5)}_{ij} & w^{(0)}_{ij} & w^{(6)}_{ij} & w^{(2)}_{ij} & -w^{(4)}_{ij} & w^{(1)}_{ij} \\ -w^{(4)}_{ij} & w^{(2)}_{ij} & -w^{(1)}_{ij} & -w^{(6)}_{ij} & w^{(0)}_{ij} & w^{(7)}_{ij} & w^{(3)}_{ij} & -w^{(5)}_{ij} \\ -w^{(5)}_{ij} & -w^{(6)}_{ij} & w^{(3)}_{ij} & -w^{(2)}_{ij} & -w^{(7)}_{ij} & w^{(0)}_{ij} & w^{(1)}_{ij} & w^{(4)}_{ij} \\ -w^{(6)}_{ij} & w^{(5)}_{ij} & -w^{(7)}_{ij} & w^{(4)}_{ij} & -w^{(3)}_{ij} & -w^{(1)}_{ij} & w^{(0)}_{ij} & w^{(2)}_{ij} \\ -w^{(7)}_{ij} & w^{(3)}_{ij} & w^{(6)}_{ij} & -w^{(1)}_{ij} & w^{(5)}_{ij} & -w^{(4)}_{ij} & -w^{(2)}_{ij} & w^{(0)}_{ij} \end{bmatrix}.$$

Note that the equation (28) becomes equal to (21) by replacing $\hat{\mathbf{W}}^3_{ij}\hat{\mathbf{W}}^2_{ij}$ by $\hat{\mathbf{W}}^1_{ij}$ in (21). It can be shown that the product of the matrices $\hat{\mathbf{W}}^3_{ij}$ and $\hat{\mathbf{W}}^2_{ij}$ is commutative, that is, $\hat{\mathbf{W}}^3_{ij}\hat{\mathbf{W}}^2_{ij} = \hat{\mathbf{W}}^2_{ij}\hat{\mathbf{W}}^3_{ij}$. By using these facts the existence condition of the energy functions for Model 3 can be proved in the similar way to that for Model 1.

The existence of the energy function for the network of Model 4 (7) can be proved as follows. Under the assumption (17) on the weight coefficients $w_{ij} := {w^{\ell}_{ij}}^* = w^r_{ij}$, the gradient of the function E given by (27) with respect to $\hat{\mathbf{v}}_i$ can be calculated as follows:

$$\begin{aligned} \nabla_{\hat{\mathbf{v}}_i} E(\boldsymbol{v}) &= -(-\hat{\mathbf{u}}_i + \sum_{j=1}^{n} \hat{\mathbf{W}}^2_{ij} \hat{\mathbf{W}}^3_{ij} \hat{\mathbf{v}}_j + \hat{\mathbf{b}}_i) \\ &= -\tau_i \frac{d\hat{\mathbf{u}}_i}{dt}. \end{aligned} \tag{29}$$

Note that the equation (29) becomes equal to (21) by replacing $\hat{\mathbf{W}}^2_{ij}\hat{\mathbf{W}}^3_{ij}$ by $\hat{\mathbf{W}}^1_{ij}$ in (21). By using the fact the existence condition of the energy functions for Model 4 can be proved in the similar way to that for Model 1. ■

Remark: 2: It can be shown that the relation $(b^*a)b = b^*(ab)$ holds for any octonion $a \in \mathbb{O}$ and $b \in \mathbb{O}$. Therefore the term $w^*_{ij}(v_j w_{ij})$ in the energy function (26) of Model 3 and the term $(w^*_{ij} v_j) w_{ij}$ in the energy function (27) of Model 4 are identical and they are same energy functions.

The existence conditions of energy functions thus obtained are ones on the connection weight coefficients w_{ij} and the activation function $f(\cdot)$. As examples of the functions which satisfy Assumption 1,

$$f(u) = \frac{u}{1+|u|} \tag{30}$$

$$f(u) = \tanh(u^{(0)}) + \sum_{i=1}^{7} \mathbf{e}_i \tanh(u^{(i)}) \tag{31}$$

can be considered. Equation (30) has the same form as that of the complex-valued function which is often used in the complex-valued neural networks [10], [11]. The function (31) is a split activation function, that is, each component of its argument is transformed separately.

It is expected that the energy functions (18), (19), (26) and (27) can be applied to various problems. In the real valued neural networks energy functions have been applied to various problems such as qualitative analysis of neural networks, synthesis of associative memories and optimization problems. In [10] and [12], qualitative analysis of the complex valued and quaternion valued networks is performed by utilizing energy functions and some results are obtained. The similar results can be obtained for the octonion neural networks (4), (5), (6) and (7) by utilizing the energy functions (18), (19), (26) and (27).

VI. Conclusion

Recently models of neural networks in the real domain have been extended into the high dimensional domain such as the complex number and quaternion domain. In this paper we extended conventional real-valued models of recurrent neural networks into the octonion domain and discussed their dynamics. Since the octonions are non-commutative and non-associative on multiplication, a couple of different models of octonion neural networks can be considered. We proposed four models of fully connected recurrent octonion neural network, which are extensions of the real-valued Hopfield type neural networks to the octonion domain. We also studied dynamics of the proposed models from the point view of existence conditions of energy functions. We derive the existence conditions of energy functions for each of them and construct an energy function for each model. It is expected that those energy functions are applied to various problems such as qualitative analysis of neural networks, synthesis of associative memories, optimization problems and so on. Note also that, although we treat continuous-time models of Hopfield type neural networks in this paper, discrete-time ones can be considered and are of interest and useful especially from implementation viewpoint. It can be considered the similar discussions can be done on discrete-time models of Hopfield type octonion neural networks.

References

[1] A. Hirose (ed.): Complex-Valued Neural Networks Theories and Applications, World Scientific, 2003.

[2] T. Nitta (ed.): Complex-Valued Neural Networks Utilizing High-Dimensional Parameters, IGI Global, 2009.

[3] S. Buchholz: A Theory of Neural Computation with Clifford Algebra, Ph.D. Thesis, University. of Kiel, 2005.

[4] P. Lounesto: Clifford Algebras and Spinors 2nd Edition, Cambridge Univ. Press, 2001.

[5] L. Dorst, D. Fontijne and S. Mann: Geometric Algebra for Computer Science An object-oriented Approach to Geometry, Morgan Kaufmann Publisher, 2007.

[6] Christian Perwass: Geometric Algebra with Applications in Engineering, Springer-Verlag 2009.

[7] J. J. Hopfield: Neurons with graded response have collective computational properties like those of two-state neurons; *Proc. Natl. Acad. Sci. USA*, Vol.81, pp.3088–3092, 1984.

[8] J. J. Hopfield and D. W. Tank: Neural' computation of decisions in optimization problems; *Biol. Cybern.*, Vol.52, pp.141–152, 1985.

[9] D. W. Tank and J. J. Hopfield, "Simple 'neural' optimization networks: an A/D converter, signal decision circuit, and a linear programming circuit," *IEEE Trans. Circuits Syst.*, **33**, 533–541, 1986.

[10] N. Hashimoto, Y. Kuroe and T. Mori: On Energy Function for Complex-Valued Neural Networks; *Trans. of the Institute of System, Control and Information Engineers*, Vol.15, No.10, pp.559-565, 2002 (in Japanese).

[11] Y. Kuroe, M. Yoshida and T. Mori: On Activation Functions for Complex-Valued Neural Networks - Existence of Energy Functions -; *Artificial Neural Networks and Neural Information Processing - ICANN/ICONIP 2003, Okyay Kaynak et al.(Eds.), Lecture Notes in Computer Science*, 2714, pp.985-992, Springer, 2003.

[12] M. Yoshida, Y. Kuroe and T. Mori: Models of Hopfield-Type Quaternion Neural Networks and Their Energy Functions; *International Journal of Neural Systems*, Vol.15, Nos.1 & 2, pp.129-135, 2005.

[13] Y. Kuroe, S. Tanigawa and H. Iima: Models of Hopfield-type Clifford Neural Networks and Their Energy Functions - Hyperbolic and Dual Valued Networks -; Proceedings of International Conference on Neural Information Processing, Lecture Notes in Computer Science, 7062, pp. 560-569, Springer-Verlag, 2011.

[14] Y. Kuroe: Models of Clifford Recurrent Neural Networks and Their Dynamics, Proceedings of 2011 International Joint Conference on Neural Networks, pp.1035-1041, 2011.

[15] John C. Baez: The Octonions, American Mathematical Society, Vol.39, No.2, pp.145-205, 2001.

[16] John H. Conway and Derek A. Smith: On Quaternions and Octonions: Their Geometry, Arithmetic, and Symmetry, A K Peters Wellesley, Massachusetts, 2003.

[17] T. Dray and C. Manogue: The Geometry of the Octonions, World Scientific, 2015.

[18] Y. Kuroe and H. Iima: A model of Hopfield-type Octonion Neural Networks and Existing Conditions of Energy Functions, Proc. of 2016 International Joint Conference on Neural Networks, pp. 4426-4430, 2016.

[19] C.-A. Popa: Octonion-Valued Neural Networks; Proc. of International Conference on Artificial Neural Networks (ICANN2016), pp 435-443, 2016.

[20] C.-A. Popa: Global exponential stability of neutral-type octonion-valued neural networks with time-varying delays; Neurocomputing, Vol.309, No.2, pp.ages 117-133, 2018.