Research Article

Convergence of Batch Split-Complex Backpropagation Algorithm for Complex-Valued Neural Networks

Huisheng Zhang,[1,2] Chao Zhang,[1] and Wei Wu[1]

[1] *Applied Mathematics Department, Dalian University of Technology, Dalian 116024, China*
[2] *Department of Mathematics, Dalian Maritime University, Dalian 116026, China*

Correspondence should be addressed to Wei Wu, wuweiw@dlut.edu.cn

The batch split-complex backpropagation (BSCBP) algorithm for training complex-valued neural networks is considered. For constant learning rate, it is proved that the error function of BSCBP algorithm is monotone during the training iteration process, and the gradient of the error function tends to zero. By adding a moderate condition, the weights sequence itself is also proved to be convergent. A numerical example is given to support the theoretical analysis.

1. Introduction

Neural networks are widely used in the fields of control, signal processing, and time series analysis [1]. Traditional neural networks' parameters are usually real numbers for dealing with real-valued signals. However, complex-valued signals also appear in practical applications. As a result, complex-valued neural network (CVNN), whose weights, threshold values, input and output signals are all complex numbers, is proposed [2, 3]. CVNN has been extensively used in processing complex-valued signals [4]. By encoding real-valued signals into complex numbers, CVNN also has shown more powerful capability than real-valued neural networks in processing real-valued signals. For example, two-layered CVNN [5] can successfully solve the XOR problem which cannot be solved by two-layered real-valued neural networks. CVNN can be trained by two types of complex backpropagation (BP) algorithms: fully complex BP algorithm and split-complex BP algorithm. Different from the fully complex BP algorithm [6], the operation of activation function in the split-complex BP algorithm is split into real part and imaginary part [2–4, 7], and this makes the split-complex BP algorithm avoid the occurrence of singular points in the adaptive training process.

Complex BP algorithms can be done using either a batch method or an online method. In online training, weights are updated after the presentation of each training example, while in batch training, weights are not updated until all of the examples are inputted into the networks. Compared with batch learning, online learning is hard to parallelize.

The convergence of neural networks learning algorithms is crucial for practical applications. The dynamical behaviors of many neural networks have been extensively analyzed [8, 9]. However, the existing convergence results of complex BP algorithm are mainly focusing on fully complex BP algorithm for two-layered CVNN (see, e.g., [10, 11]) and the convergence of split-complex BP algorithm is seldom investigated. Nitta [12] used CVNN as a complex adaptive pattern classifier and presented some heuristic convergence results. The purpose of this paper is to give some rigorous convergence results of batch split-complex BP (BSCBP) algorithm for three-layered CVNN. The monotonicity of the error function during the training iteration process is also guaranteed.

The remainder of this paper is organized as follows. The three-layered CVNN model and the BSCBP algorithm are described in the next section. Section 3 presents the main convergence theorem. A numerical example is given in Section 4 to verify our theoretical results. The details of the convergence proof are provided in Section 5. Some conclusions are drawn in Section 6.

2. Network Structure and Learning Method

Figure 1 shows the structure of the network we considered in this paper. It is a three-layered CVNN consisting of L input neurons, M hidden neurons, and 1 output neuron. For any positive integer d, the set of all d-dimensional complex vectors is denoted by $\mathbb{C}^d$ and the set of all d-dimensional real vectors is denoted by $\mathbb{R}^d$. Let us write $\mathbf{w}_m = \mathbf{w}_m^R + i\mathbf{w}_m^I = (w_{m1}, w_{m2}, \ldots, w_{mL})^T \in \mathbb{C}^L$ as the weight vector between the input neurons and mth hidden neuron, where $w_{ml} = w_{ml}^R + iw_{ml}^I$, w_{ml}^R and $w_{ml}^I \in \mathbb{R}^1$, $i = \sqrt{-1}$, $m = 1, \ldots, M$, and $l = 1, \ldots, L$. Similarly, write $\mathbf{v} = \mathbf{v}^R + i\mathbf{v}^I = (v_1, v_2, \ldots, v_M)^T \in \mathbb{C}^M$ as the weight vector between the hidden neurons and the output neuron, where $v_m = v_m^R + iv_m^I$, v_m^R and $v_m^I \in \mathbb{R}^1$, $m = 1, \ldots, M$. For simplicity, all the weight vectors are incorporated into a total weight vector

$$\mathbf{W} = \left((\mathbf{w}_1)^T, (\mathbf{w}_2)^T, \ldots, (\mathbf{w}_M)^T, \mathbf{v}^T\right)^T \in \mathbb{C}^{M(L+1)}. \tag{2.1}$$

For input signals $\mathbf{z} = (z_1, z_2, \ldots, z_L)^T = \mathbf{x} + i\mathbf{y} \in \mathbb{C}^L$, where $\mathbf{x} = (x_1, x_2, \ldots, x_L)^T \in \mathbb{R}^L$, and $\mathbf{y} = (y_1, y_2, \ldots, y_L)^T \in \mathbb{R}^L$, the input of the mth hidden neuron is

$$\begin{aligned}
U_m &= U_m^R + iU_m^I \\
&= \sum_{l=1}^{L}(w_{ml}^R x_l - w_{ml}^I y_l) + i\sum_{l=1}^{L}(w_{ml}^I x_l + w_{ml}^R y_l) \\
&= \begin{pmatrix} \mathbf{w}_m^R \\ -\mathbf{w}_m^I \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} + i\begin{pmatrix} \mathbf{w}_m^I \\ \mathbf{w}_m^R \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}.
\end{aligned} \tag{2.2}$$

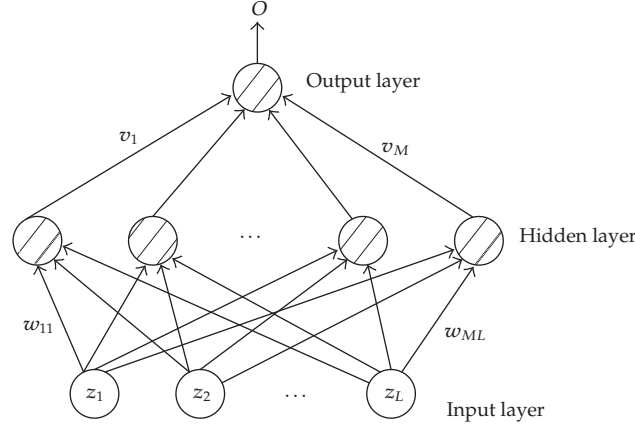Here "$\cdot$" denotes the inner product of two vectors.

Figure 1: CVVN with $L - M - 1$ structure.

For the sake of using BSCBP algorithm to train the network, we consider the following popular real-imaginary-type activation function [5]:

$$f_C(U) = f_R(U^R) + if_R(U^I) \tag{2.3}$$

for any $U = U^R + iU^I \in \mathbb{C}^1$, where f_R is a real function (e.g., sigmoid function). If we simply denote f_R as f, the output H_m for the hidden neuron m is given by

$$H_m = H_m^R + iH_m^I = f(U_m^R) + if(U_m^I). \tag{2.4}$$

Similarly, the input of the output neuron is

$$
\begin{aligned}
S &= S^R + iS^I \\
&= \sum_{m=1}^{M} (v_m^R H_m^R - v_m^I H_m^I) + i\sum_{m=1}^{M} (v_m^I H_m^R + v_m^R H_m^I) \\
&= \begin{pmatrix} \mathbf{v}^R \\ -\mathbf{v}^I \end{pmatrix} \cdot \begin{pmatrix} \mathbf{H}^R \\ \mathbf{H}^I \end{pmatrix} + i \begin{pmatrix} \mathbf{v}^I \\ \mathbf{v}^R \end{pmatrix} \cdot \begin{pmatrix} \mathbf{H}^R \\ \mathbf{H}^I \end{pmatrix}
\end{aligned}
\tag{2.5}
$$

and the output of the network is given by

$$O = O^R + iO^I = g(S^R) + ig(S^I), \tag{2.6}$$

where $\mathbf{H}^R = (H_1^R, H_2^R, \ldots, H_M^R)^T$, $\mathbf{H}^I = (H_1^I, H_2^I, \ldots, H_M^I)^T$, and g is a real function.

We remark that, in practice, there should be thresholds involved in the above formulas for the output and hidden neurons. Here we have omitted the bias so as to simplify the presentation and deduction.

Let the network be supplied with a given set of training examples $\{\mathbf{z}^q, d^q\}_{q=1}^Q \subset \mathbb{C}^L \times \mathbb{C}^1$. For each input $\mathbf{z}^q = \mathbf{x}^q + i\mathbf{y}^q$ $(1 \le q \le Q)$ from the training set, we write $U_m^q = U_m^{q,R} + iU_m^{q,I}$ $(1 \le m \le M)$ as the input for the hidden neuron m, $H_m^q = H_m^{q,R} + iH_m^{q,I}$ $(1 \le m \le M)$ as the output for the hidden neuron m, $S^q = S^{q,R} + iS^{q,I}$ as the input to the output neuron, and $O^q = O^{q,R} + iO^{q,I}$ as the actual output. The square error function of CVNN trained by BSCBP algorithm can be represented as follows:

$$
\begin{aligned}
E(\mathbf{W}) &= \frac{1}{2}\sum_{q=1}^Q (O^q - d^q)(O^q - d^q)^* \\
&= \frac{1}{2}\sum_{q=1}^Q \left[(O^{q,R} - d^{q,R})^2 + (O^{q,I} - d^{q,I})^2 \right] \\
&= \sum_{q=1}^Q \left[\mu_{qR}(S^{q,R}) + \mu_{qI}(S^{q,I}) \right],
\end{aligned}
\tag{2.7}
$$

where "$*$" signifies complex conjugate, and

$$
\mu_{qR}(t) = \frac{1}{2}(g(t) - d^{q,R})^2, \quad \mu_{qI}(t) = \frac{1}{2}(g(t) - d^{q,I})^2, \quad t \in \mathbb{R}^1,\ 1 \le q \le Q.
\tag{2.8}
$$

The purpose of the network training is to find $\mathbf{W}^\star$ which can minimize $E(\mathbf{W})$. The gradient method is often used to solve the minimization problem. Writing

$$
\mathbf{H}^q = \mathbf{H}^{q,R} + i\mathbf{H}^{q,I} = \left(H_1^{q,R}, H_2^{q,R}, \ldots, H_M^{q,R}\right)^T + i\left(H_1^{q,I}, H_2^{q,I}, \ldots, H_M^{q,I}\right)^T,
\tag{2.9}
$$

and differentiating $E(\mathbf{W})$ with respect to the real parts and imaginary parts of the weight vectors, respectively, give

$$
\frac{\partial E(\mathbf{W})}{\partial \mathbf{v}^R} = \sum_{q=1}^Q \left[\mu'_{qR}(S^{q,R})\mathbf{H}^{q,R} + \mu'_{qI}(S^{q,I})\mathbf{H}^{q,I} \right],
\tag{2.10}
$$

$$
\frac{\partial E(\mathbf{W})}{\partial \mathbf{v}^I} = \sum_{q=1}^Q \left[-\mu'_{qR}(S^{q,R})\mathbf{H}^{q,I} + \mu'_{qI}(S^{q,I})\mathbf{H}^{q,R} \right],
\tag{2.11}
$$

$$
\begin{aligned}
\frac{\partial E(\mathbf{W})}{\partial \mathbf{w}_m^R} = \sum_{q=1}^Q \Big[&\mu'_{qR}(S^{q,R})\left(v_m^R f'(U_m^{q,R})\mathbf{x}^q - v_m^I f'(U_m^{q,I})\mathbf{y}^q\right) \\
&+ \mu'_{qI}(S^{q,I})\left(v_m^I f'(U_m^{q,R})\mathbf{x}^q + v_m^R f'(U_m^{q,I})\mathbf{y}^q\right) \Big], \quad 1 \le m \le M,
\end{aligned}
\tag{2.12}
$$

$$
\begin{aligned}
\frac{\partial E(\mathbf{W})}{\partial \mathbf{w}_m^I} = \sum_{q=1}^Q \Big[&\mu'_{qR}(S^{q,R})\left(-v_m^R f'(U_m^{q,R})\mathbf{y}^q - v_m^I f'(U_m^{q,I})\mathbf{x}^q\right) \\
&+ \mu'_{qI}(S^{q,I})\left(-v_m^I f'(U_m^{q,R})\mathbf{y}^q + v_m^R f'(U_m^{q,I})\mathbf{x}^q\right) \Big], \quad 1 \le m \le M.
\end{aligned}
\tag{2.13}
$$

Starting from an arbitrary initial value $\mathbf{W}^0$ at time 0, BSCBP algorithm updates the weight vector $\mathbf{W}$ iteratively by

$$\mathbf{W}^{n+1} = \mathbf{W}^n + \Delta\mathbf{W}^n, \quad n = 0, 1, \ldots, \tag{2.14}$$

where $\Delta\mathbf{W}^n = ((\Delta\mathbf{w}_1^n)^T, \ldots, (\Delta\mathbf{w}_M^n)^T, \Delta\mathbf{v}^n)^T)^T$, with

$$\Delta\mathbf{w}_m^n = -\eta\left(\frac{\partial E(\mathbf{W}^n)}{\partial\mathbf{w}_m^R} + i\frac{\partial E(\mathbf{W}^n)}{\partial\mathbf{w}_m^I}\right), \quad m = 1, \ldots, M,$$

$$\Delta\mathbf{v}^n = -\eta\left(\frac{\partial E(\mathbf{W}^n)}{\partial\mathbf{v}^R} + i\frac{\partial E(\mathbf{W}^n)}{\partial\mathbf{v}^I}\right). \tag{2.15}$$

Here $\eta > 0$ stands for the learning rate. Obviously, we can rewrite (2.14) and (2.15) by dealing with the real parts and the imaginary parts of the weights separately

$$\Delta\mathbf{w}_m^{n,R} = \mathbf{w}_m^{n+1,R} - \mathbf{w}_m^{n,R} = -\eta\frac{\partial E(\mathbf{W}^n)}{\partial\mathbf{w}_m^R},$$

$$\Delta\mathbf{w}_m^{n,I} = \mathbf{w}_m^{n+1,I} - \mathbf{w}_m^{n,I} = -\eta\frac{\partial E(\mathbf{W}^n)}{\partial\mathbf{w}_m^I},$$

$$\Delta\mathbf{v}^{n,R} = \mathbf{v}^{n+1,R} - \mathbf{v}^{n,R} = -\eta\frac{\partial E(\mathbf{W}^n)}{\partial\mathbf{v}^R},$$

$$\Delta\mathbf{v}^{n,I} = \mathbf{v}^{n+1,I} - \mathbf{v}^{n,I} = -\eta\frac{\partial E(\mathbf{W}^n)}{\partial\mathbf{v}^I}, \tag{2.16}$$

where $m = 1, \ldots, M$.

3. Main Results

Throughout the paper $\|\cdot\|$ denotes the usual Euclidean norm. We need the following assumptions:

(A1) there exists a constant $c_1 > 0$ such that

$$\max_{t\in\mathbb{R}^1}\{|f(t)|, |g(t)|, |f'(t)|, |g'(t)|, |f''(t)|, |g''(t)|\} \leq c_1; \tag{3.1}$$

(A2) there exists a constant $c_2 > 0$ such that $\|\mathbf{v}^{n,R}\| \leq c_2$ and $\|\mathbf{v}^{n,I}\| \leq c_2$ for all $n = 0, 1, 2, \ldots$;

(A3) the set $\Phi_0 = \{\mathbf{W} \mid (\partial E(\mathbf{W})/\partial\mathbf{w}_m^R) = 0, (\partial E(\mathbf{W})/\partial\mathbf{w}_m^I) = 0, (\partial E(\mathbf{W})/\partial\mathbf{v}^R) = 0, (\partial E(\mathbf{W})/\partial\mathbf{v}^I) = 0, m = 1, \ldots, M\}$ contains only finite points.

Theorem 3.1. *Suppose that Assumptions (A1) and (A2) are valid and that $\{\mathbf{W}^n\}$ are the weight vector sequence generated by (2.14)–(2.16) with arbitrary initial values $\mathbf{W}^0$. If $\eta \leq c_8$, where c_8 is a constant defined in (5.21) below, then one has*

(i) $E(\mathbf{W}^{n+1}) \le E(\mathbf{W}^n)$, $n = 0, 1, 2, \ldots$;

(ii) $\lim_{n \to \infty} \|(\partial E(\mathbf{W}^n)/\partial \mathbf{w}_m^R)\| = 0$, $\lim_{n \to \infty} \|(\partial E(\mathbf{W}^n)/\partial \mathbf{w}_m^I)\| = 0$, $\lim_{n \to \infty} \|(\partial E(\mathbf{W}^n)/\partial \mathbf{v}^R)\| = 0$, and $\lim_{n \to \infty} \|(\partial E(\mathbf{W}^n)/\partial \mathbf{v}^I)\| = 0$, $0 \le m \le M$.

Furthermore, if Assumption (A3) also holds, then there exists a point $\mathbf{W}^\star \in \Phi_0$ such that

(iii) $\lim_{n \to \infty} \mathbf{W}^n = \mathbf{W}^\star$.

The monotonicity of the error function $E(\mathbf{W})$ during the learning process is shown in the statement (i). The statement (ii) indicates the convergence of the gradients for the error function with respect to the real parts and the imaginary parts of the weights. The statement (iii) points out that if the number of the stationary points is finite, the sequence $\{\mathbf{W}^n\}$ will converge to a local minimum of the error function.

4. Numerical Example

In this section, we illustrate the convergence behavior of BSCBP algorithm by using a simple numerical example. The well-known XOR problem is a benchmark in literature of neural networks. As in [5], the training samples of the encoded XOR problem for CVNN are presented as follows:

$$\{\mathbf{z}^1 = -1 - i, \ d^1 = 1\}, \qquad \{\mathbf{z}^2 = -1 + i, \ d^2 = 0\},$$
$$\{\mathbf{z}^3 = 1 - i, \ d^3 = 1 + i\}, \qquad \{\mathbf{z}^4 = 1 + i, \ d^4 = i\}. \tag{4.1}$$

This example uses a network with one input neuron, three hidden neurons, and one output neuron. The transfer function is $tansig(\cdot)$ in MATLAB, which is a commonly used sigmoid function. The learning rate η is set to be 0.1. We carry out the test with the initial components of the weights stochastically chosen in $[-0.5, 0.5]$. Figure 2 shows that the gradient tends to zero and the square error decreases monotonically as the number of iteration increases and at last tends to a constant. This supports our theoretical findings.

5. Proofs

In this section, we first present two lemmas; then, we use them to prove the main theorem.

Lemma 5.1. *Suppose that the function $\mathcal{E} : \mathbb{R}^{2M(L+1)} \to \mathbb{R}^1$ is continuous and differentiable on a compact set $\Phi \subset \mathbb{R}^{2M(L+1)}$ and that $\Phi_1 = \{\boldsymbol{\theta} \mid (\partial \mathcal{E}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}) = 0\}$ contains only finite points. If a sequence $\{\boldsymbol{\theta}^n\}_{n=1}^\infty \subset \Phi$ satisfies*

$$\lim_{n \to \infty} \|\boldsymbol{\theta}^{n+1} - \boldsymbol{\theta}^n\| = 0, \qquad \lim_{n \to \infty} \left\| \frac{\partial \mathcal{E}(\boldsymbol{\theta}^n)}{\partial \boldsymbol{\theta}} \right\| = 0, \tag{5.1}$$

then there exists a point $\boldsymbol{\theta}^\star \in \Phi_1$ such that $\lim_{n \to \infty} \boldsymbol{\theta}^n = \boldsymbol{\theta}^\star$.

Proof. This result is almost the same as [13, Theorem 14.1.5], and the detail of the proof is omitted. □
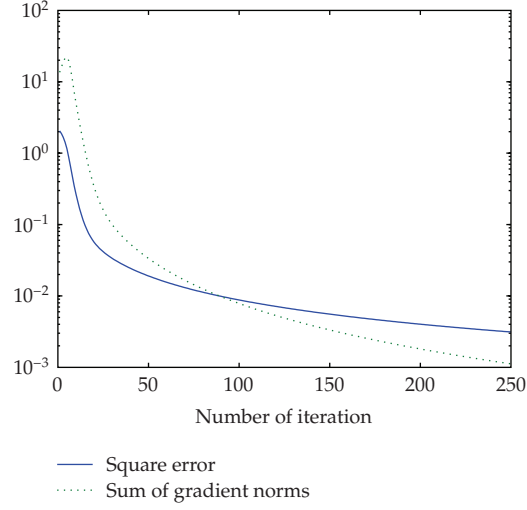
Figure 2: Convergence behavior of BSCBP algorithm for solving XOR problem. (sum of gradient norms = $\sum_{m=1}^{M}(\|\Delta\mathbf{w}_m^{n,R}\|^2 + \|\Delta\mathbf{w}_m^{n,I}\|^2) + \|\Delta\mathbf{v}^{n,R}\|^2 + \|\Delta\mathbf{v}^{n,I}\|^2$.)

For any $1 \le q < Q, 1 \le m \le M$ and $n = 0,1,2,\ldots$, write

$$U_m^{n,q} = U_m^{n,q,R} + iU_m^{n,q,I} = \begin{pmatrix} \mathbf{w}_m^{n,R} \\ -\mathbf{w}_m^{n,I} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x}^q \\ \mathbf{y}^q \end{pmatrix} + i\begin{pmatrix} \mathbf{w}_m^{n,I} \\ \mathbf{w}_m^{n,R} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x}^q \\ \mathbf{y}^q \end{pmatrix},$$

$$H_m^{n,q} = H_m^{n,q,R} + iH_m^{n,q,I} = f(U_m^{n,q,R}) + if(U_m^{n,q,I}),$$

$$\mathbf{H}^{n,q,R} = (H_1^{n,q,R},\ldots,H_M^{n,q,R})^T, \qquad \mathbf{H}^{n,q,I} = (H_1^{n,q,I},\ldots,H_M^{n,q,I})^T, \tag{5.2}$$

$$S^{n,q} = S^{n,q,R} + iS^{n,q,I} = \begin{pmatrix} \mathbf{v}^{n,R} \\ -\mathbf{v}^{n,I} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{H}^{n,q,R} \\ \mathbf{H}^{n,q,I} \end{pmatrix} + i\begin{pmatrix} \mathbf{v}^{n,I} \\ \mathbf{v}^{n,R} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{H}^{n,q,R} \\ \mathbf{H}^{n,q,I} \end{pmatrix},$$

$$\psi^{n,q,R} = \mathbf{H}^{n+1,q,R} - \mathbf{H}^{n,q,R}, \qquad \psi^{n,q,I} = \mathbf{H}^{n+1,q,I} - \mathbf{H}^{n,q,I}.$$

Lemma 5.2. *Suppose Assumptions (A1) and (A2) hold, then for any $1 \le q \le Q$ and $n = 0,1,2,\ldots$, one has*

$$|O^{q,R}| \le c_0, \qquad |O^{q,I}| \le c_0, \qquad \|\mathbf{H}^{n,q,R}\| \le c_0, \qquad \|\mathbf{H}^{n,q,I}\| \le c_0, \tag{5.3}$$

$$|\mu'_{qR}(t)| \le c_3, |\mu'_{qI}(t)| \le c_3, \qquad |\mu''_{qR}(t)| \le c_3, \qquad |\mu''_{qI}(t)| \le c_3, \qquad t \in \mathbb{R}^1, \tag{5.4}$$

$$\max\left\{\|\psi^{n,q,R}\|^2, \|\psi^{n,q,I}\|^2\right\} \le c_4 \sum_{m=1}^{M}(\|\Delta\mathbf{w}_m^{n,R}\|^2 + \|\Delta\mathbf{w}_m^{n,I}\|^2), \tag{5.5}$$

$$\sum_{q=1}^{Q} \left(\mu'_{qR}(S^{n,q,R}) \begin{pmatrix} \Delta\mathbf{v}^{n,R} \\ -\Delta\mathbf{v}^{n,I} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{H}^{n,q,R} \\ \mathbf{H}^{n,q,I} \end{pmatrix} + \mu'_{qI}(S^{n,q,I}) \begin{pmatrix} \Delta\mathbf{v}^{n,I} \\ \Delta\mathbf{v}^{n,R} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{H}^{n,q,R} \\ \mathbf{H}^{n,q,I} \end{pmatrix} \right) \tag{5.6}$$

$$= -\frac{1}{\eta} \left(\|\Delta\mathbf{v}^{n,R}\|^2 + \|\Delta\mathbf{v}^{n,I}\|^2 \right),$$

$$\sum_{q=1}^{Q} \mu'_{qR}(S^{n,q,R}) \begin{pmatrix} \mathbf{v}^{n,R} \\ -\mathbf{v}^{n,I} \end{pmatrix} \cdot \begin{pmatrix} \psi^{n,q,R} \\ \psi^{n,q,I} \end{pmatrix} + \mu'_{qI}(S^{n,q,I}) \begin{pmatrix} \mathbf{v}^{n,I} \\ \mathbf{v}^{n,R} \end{pmatrix} \cdot \begin{pmatrix} \psi^{n,q,R} \\ \psi^{n,q,I} \end{pmatrix} \tag{5.7}$$

$$\leq \left(c_5 - \frac{1}{\eta} \right) \sum_{m=1}^{M} \left(\|\Delta\mathbf{w}_m^{n,R}\|^2 + \|\Delta\mathbf{w}_m^{n,I}\|^2 \right),$$

$$\sum_{q=1}^{Q} \left(\mu'_{qR}(S^{n,q,R}) \begin{pmatrix} \Delta\mathbf{v}^{n,R} \\ -\Delta\mathbf{v}^{n,I} \end{pmatrix} \cdot \begin{pmatrix} \psi^{n,q,R} \\ \psi^{n,q,I} \end{pmatrix} + \mu'_{qI}(S^{n,q,I}) \begin{pmatrix} \Delta\mathbf{v}^{n,I} \\ \Delta\mathbf{v}^{n,R} \end{pmatrix} \cdot \begin{pmatrix} \psi^{n,q,R} \\ \psi^{n,q,I} \end{pmatrix} \right) \tag{5.8}$$

$$\leq c_6 \left(\sum_{m=1}^{M} (\|\Delta\mathbf{w}_m^{n,R}\|^2 + \|\Delta\mathbf{w}_m^{n,I}\|^2) + \|\Delta\mathbf{v}^{n,R}\|^2 + \|\Delta\mathbf{v}^{n,I}\|^2 \right),$$

$$\frac{1}{2} \sum_{q=1}^{Q} (\mu''_{qR}(t_1^{n,q})(S^{n+1,q,R} - S^{n,q,R})^2 + \mu''_{qI}(t_2^{n,q})(S^{n+1,q,I} - S^{n,q,I})^2) \tag{5.9}$$

$$\leq c_7 \left(\sum_{m=1}^{M} (\|\Delta\mathbf{w}_m^{n,R}\|^2 + \|\Delta\mathbf{w}_m^{n,I}\|^2) + \|\Delta\mathbf{v}^{n,R}\|^2 + \|\Delta\mathbf{v}^{n,I}\|^2 \right),$$

where c_i $(i = 0, 3, \ldots, 7)$ are constants independent of n and q, each $t_1^{n,q} \in \mathbb{R}^1$ lies on the segment between $S^{n+1,q,R}$ and $S^{n,q,R}$, and each $t_2^{n,q} \in \mathbb{R}^1$ lies on the segment between $S^{n+1,q,I}$ and $S^{n,q,I}$.

Proof. The validation of (5.3) can be easily got by (2.4)–(2.6) when the set of samples are fixed and Assumptions (A1) and (A2) are satisfied. By (2.8), we have

$$\mu'_{qR}(t) = g'(t)(g(t) - O^{q,R}),$$

$$\mu'_{qI}(t) = g'(t)(g(t) - O^{q,I}),$$

$$\mu''_{qR}(t) = g''(t)(g(t) - O^{q,R}) + (g'(t))^2, \tag{5.10}$$

$$\mu''_{qI}(t) = g''(t)(g(t) - O^{q,I}) + (g'(t))^2, \quad 1 \leq q \leq Q, \ t \in \mathbb{R}^1.$$

Then (5.4) follows directly from Assumption (A1) by defining $c_3 = c_1(c_1 + c_0) + (c_1)^2$.

It follows from (5.2), Assumption (A1), the Mean-Value Theorem and the Cauchy-Schwartz Inequality that for any $1 \leq q \leq Q$ and $n = 0, 1, 2, \ldots,$

$$
\begin{aligned}
\left\| \psi^{n,q,R} \right\|^2 &= \left\| \mathbf{H}^{n+1,q,R} - \mathbf{H}^{n,q,R} \right\|^2 \\
&= \left\| \begin{pmatrix} f(U_1^{n+1,q,R}) - f(U_1^{n,q,R}) \\ \vdots \\ f(U_M^{n+1,q,R}) - f(U_M^{n,q,R}) \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} f'(s_1^{n,q})(U_1^{n+1,q,R} - U_1^{n,q,R}) \\ \vdots \\ f'(s_M^{n,q})(U_M^{n+1,q,R} - U_M^{n,q,R}) \end{pmatrix} \right\|^2 \\
&= \sum_{m=1}^{M} \left(f'(s_m^{n,q})(\Delta \mathbf{w}_m^{n,R} \cdot \mathbf{x}^q - \Delta \mathbf{w}_m^{n,I} \cdot \mathbf{y}^q) \right)^2 \\
&\leq 2c_1 \sum_{m=1}^{M} \left((\Delta \mathbf{w}_m^{n,R} \cdot \mathbf{x}^q)^2 + (\Delta \mathbf{w}_m^{n,I} \cdot \mathbf{y}^q)^2 \right) \\
&\leq 2c_1 \sum_{m=1}^{M} \left(\| \Delta \mathbf{w}_m^{n,R} \|^2 \| \mathbf{x}^q \|^2 + \| \Delta \mathbf{w}_m^{n,I} \|^2 \| \mathbf{y}^q \|^2 \right) \\
&\leq c_4 \sum_{m=1}^{M} \left(\| \Delta \mathbf{w}_m^{n,R} \|^2 + \| \Delta \mathbf{w}_m^{n,I} \|^2 \right),
\end{aligned}
\tag{5.11}
$$

where $c_4 = 2c_1 \max_{1 \leq q \leq Q} \{ \| \mathbf{x}^q \|^2, \| \mathbf{y}^q \|^2 \}$ and each $s_m^{n,q}$ is on the segment between $U_m^{n+1,q,R}$ and $U_m^{n,q,R}$ for $m = 1, \ldots, M$. Similarly we can get

$$
\left\| \psi^{n,q,I} \right\| \leq c_4 \sum_{m=1}^{M} \left(\| \Delta \mathbf{w}_m^{n,R} \|^2 + \| \Delta \mathbf{w}_m^{n,I} \|^2 \right).
\tag{5.12}
$$

Thus, we have (5.5).

By (2.10), (2.11), (2.16), and (5.2), we have

$$
\begin{aligned}
\sum_{q=1}^{Q} &\left(\mu_{qR}'(S^{n,q,R}) \begin{pmatrix} \Delta \mathbf{v}^{n,R} \\ -\Delta \mathbf{v}^{n,I} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{H}^{n,q,R} \\ \mathbf{H}^{n,q,I} \end{pmatrix} + \mu_{qI}'(S^{n,q,I}) \begin{pmatrix} \Delta \mathbf{v}^{n,I} \\ \Delta \mathbf{v}^{n,R} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{H}^{n,q,R} \\ \mathbf{H}^{n,q,I} \end{pmatrix} \right) \\
&= \sum_{q=1}^{Q} \left(\mu_{qR}'(S^{n,q,R}) \mathbf{H}^{n,q,R} \cdot \Delta \mathbf{v}^{n,R} + \mu_{qI}'(S^{n,q,I}) \mathbf{H}^{n,q,I} \cdot \Delta \mathbf{v}^{n,R} \right. \\
&\qquad \left. - \mu_{qR}'(S^{n,q,R}) \mathbf{H}^{n,q,I} \cdot \Delta \mathbf{v}^{n,I} + \mu_{qI}'(S^{n,q,I}) \mathbf{H}^{n,q,R} \cdot \Delta \mathbf{v}^{n,I} \right) \\
&= \frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{v}^R} \cdot \Delta \mathbf{v}^{n,R} + \frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{v}^I} \cdot \Delta \mathbf{v}^{n,I} \\
&= -\frac{1}{\eta} \left(\| \Delta \mathbf{v}^{n,R} \|^2 + \| \Delta \mathbf{v}^{n,I} \|^2 \right).
\end{aligned}
\tag{5.13}
$$

Next, we prove (5.7). By (2.2), (2.4), (5.2), and Taylor's formula, for any $1 \leq q \leq Q$, $1 \leq m \leq M$, and $n = 0, 1, 2, \ldots$, we have

$$
\begin{aligned}
H_m^{n+1,q,R} - H_m^{n,q,R} &= f(U_m^{n+1,q,R}) - f(U_m^{n,q,R}) \\
&= f'(U_m^{n,q,R})(U_m^{n+1,q,R} - U_m^{n,q,R}) + \frac{1}{2} f''(t_m^{n,q,R})(U_m^{n+1,q,R} - U_m^{n,q,R})^2,
\end{aligned}
\tag{5.14}
$$

$$
\begin{aligned}
H_m^{n+1,q,I} - H_m^{n,q,I} &= f(U_m^{n+1,q,I}) - f(U_m^{n,q,I}) \\
&= f'(U_m^{n,q,I})(U_m^{n+1,q,I} - U_m^{n,q,I}) + \frac{1}{2} f''(t_m^{n,q,I})(U_m^{n+1,q,I} - U_m^{n,q,I})^2,
\end{aligned}
\tag{5.15}
$$

where $t_m^{n,q,R}$ is an intermediate point on the line segment between the two points $U_m^{n+1,q,R}$ and $U_m^{n,q,R}$, and $t_m^{n,q,I}$ between the two points $U_m^{n+1,q,I}$ and $U_m^{n,q,I}$. Thus according to (2.12), (2.13), (2.16), (5.2), (5.14), and (5.15), we have

$$
\begin{aligned}
&\sum_{q=1}^{Q} \left(\mu_{qR}'(S^{n,q,R}) \begin{pmatrix} \mathbf{v}^{n,R} \\ -\mathbf{v}^{n,I} \end{pmatrix} \cdot \begin{pmatrix} \psi^{n,q,R} \\ \psi^{n,q,I} \end{pmatrix} + \mu_{qI}'(S^{n,q,I}) \begin{pmatrix} \mathbf{v}^{n,I} \\ \mathbf{v}^{n,R} \end{pmatrix} \cdot \begin{pmatrix} \psi^{n,q,R} \\ \psi^{n,q,I} \end{pmatrix} \right) \\
&= \sum_{q=1}^{Q} \sum_{m=1}^{M} \left(\mu_{qR}'(S^{n,q,R}) v_m^{n,R} f'(U_m^{n,q,R}) \begin{pmatrix} \Delta \mathbf{w}_m^{n,R} \\ -\Delta \mathbf{w}_m^{n,I} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x}^q \\ \mathbf{y}^q \end{pmatrix} \right. \\
&\qquad\qquad - \mu_{qR}'(S^{n,q,R}) v_m^{n,I} f'(U_m^{n,q,I}) \begin{pmatrix} \Delta \mathbf{w}_m^{n,I} \\ \Delta \mathbf{w}_m^{n,R} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x}^q \\ \mathbf{y}^q \end{pmatrix} \\
&\qquad\qquad + \mu_{qI}'(S^{n,q,I}) v_m^{n,I} f'(U_m^{n,q,R}) \begin{pmatrix} \Delta \mathbf{w}_m^{n,R} \\ -\Delta \mathbf{w}_m^{n,I} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x}^q \\ \mathbf{y}^q \end{pmatrix} \\
&\qquad\qquad \left. + \mu_{qI}'(S^{n,q,I}) v_m^{n,R} f'(U_m^{n,q,I}) \begin{pmatrix} \Delta \mathbf{w}_m^{n,I} \\ \Delta \mathbf{w}_m^{n,R} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x}^q \\ \mathbf{y}^q \end{pmatrix} \right) + \delta_1 \\
&= \sum_{m=1}^{M} \left(\left(\sum_{q=1}^{Q} [\mu_{qR}'(S^{n,q,R})(v_m^{n,R} f'(U_m^{n,q,R}) \mathbf{x}^q - v_m^{n,I} f'(U_m^{n,q,I}) \mathbf{y}^q) \right.\right. \\
&\qquad\qquad \left. + \mu_{qI}'(S^{n,q,I})(v_m^{n,I} f'(U_m^{n,q,R}) \mathbf{x}^q + v_m^{n,R} f'(U_m^{n,q,I}) \mathbf{y}^q)] \right) \cdot \Delta \mathbf{w}_m^{n,R} \\
&\qquad\qquad + \left(\sum_{q=1}^{Q} [\mu_{qR}'(S^{n,q,R})(-v_m^{n,R} f'(U_m^{n,q,R}) \mathbf{y}^q - v_m^{n,I} f'(U_m^{n,q,I}) \mathbf{x}^q) \right. \\
&\qquad\qquad \left.\left. + \mu_{qI}'(S^{n,q,I})(-v_m^{n,I} f'(U_m^{n,q,R}) \mathbf{y}^q + v_m^{n,R} f'(U_m^{n,q,I}) \mathbf{x}^q)] \right) \cdot \Delta \mathbf{w}_m^{n,I} \right) + \delta_1 \\
&= -\frac{1}{\eta} \sum_{m=1}^{M} (\|\Delta \mathbf{w}_m^{n,R}\|^2 + \|\Delta \mathbf{w}_m^{n,I}\|^2) + \delta_1,
\end{aligned}
\tag{5.16}
$$

where

$$
\delta_1 = \frac{1}{2}\sum_{q=1}^{Q}\sum_{m=1}^{M}\left(\mu'_{qR}(S^{n,q,R})v_m^{n,R}f''(t_m^{n,q,R})\left(\begin{pmatrix}\Delta\mathbf{w}_m^{n,R}\\-\Delta\mathbf{w}_m^{n,I}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{x}^q\\\mathbf{y}^q\end{pmatrix}\right)^2\right.
$$
$$
-\mu'_{qR}(S^{n,q,R})v_m^{n,I}f''(t_m^{n,q,I})\left(\begin{pmatrix}\Delta\mathbf{w}_m^{n,I}\\\Delta\mathbf{w}_m^{n,R}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{x}^q\\\mathbf{y}^q\end{pmatrix}\right)^2
$$
$$
+\mu'_{qI}(S^{n,q,I})v_m^{n,I}f''(t_m^{n,q,R})\left(\begin{pmatrix}\Delta\mathbf{w}_m^{n,R}\\-\Delta\mathbf{w}_m^{n,I}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{x}^q\\\mathbf{y}^q\end{pmatrix}\right)^2
$$
$$
\left.+\mu'_{qI}(S^{n,q,I})v_m^{n,R}f''(t_m^{n,q,I})\left(\begin{pmatrix}\Delta\mathbf{w}_m^{n,I}\\\Delta\mathbf{w}_m^{n,R}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{x}^q\\\mathbf{y}^q\end{pmatrix}\right)^2\right). \tag{5.17}
$$

Using Assumptions (A1) and (A2), (5.4), and triangular inequality, we immediately get

$$
\delta_1 \le |\delta_1| \le c_5\sum_{m=1}^{M}\left(\|\Delta\mathbf{w}_m^{n,R}\|^2 + \|\Delta\mathbf{w}_m^{n,I}\|^2\right), \tag{5.18}
$$

where $c_5 = 2Qc_1c_2c_3\max_{1\le q\le Q}\{\|x^q\|^2 + \|y^q\|^2\}$. Now, (5.7) results from (5.16) and (5.18).
According to (5.2), (5.4), and (5.5), we have

$$
\sum_{q=1}^{Q}\left(\mu'_{qR}(S^{n,q,R})\begin{pmatrix}\Delta\mathbf{v}^{n,R}\\-\Delta\mathbf{v}^{n,I}\end{pmatrix}\cdot\begin{pmatrix}\psi^{n,q,R}\\\psi^{n,q,I}\end{pmatrix} + \mu'_{qI}(S^{n,q,I})\begin{pmatrix}\Delta\mathbf{v}^{n,I}\\\Delta\mathbf{v}^{n,R}\end{pmatrix}\cdot\begin{pmatrix}\psi^{n,q,R}\\\psi^{n,q,I}\end{pmatrix}\right)
$$
$$
\le c_3\sum_{q=1}^{Q}\left(\left\|\begin{pmatrix}\Delta\mathbf{v}^{n,R}\\-\Delta\mathbf{v}^{n,I}\end{pmatrix}\right\|\left\|\begin{pmatrix}\psi^{n,q,R}\\\psi^{n,q,I}\end{pmatrix}\right\| + \left\|\begin{pmatrix}\Delta\mathbf{v}^{n,I}\\\Delta\mathbf{v}^{n,R}\end{pmatrix}\right\|\left\|\begin{pmatrix}\psi^{n,q,R}\\\psi^{n,q,I}\end{pmatrix}\right\|\right)
$$
$$
\le \frac{1}{2}c_3\sum_{q=1}^{Q}\left(\left\|\begin{pmatrix}\Delta\mathbf{v}^{n,R}\\-\Delta\mathbf{v}^{n,I}\end{pmatrix}\right\|^2 + \left\|\begin{pmatrix}\psi^{n,q,R}\\\psi^{n,q,I}\end{pmatrix}\right\|^2 + \left\|\begin{pmatrix}\Delta\mathbf{v}^{n,I}\\\Delta\mathbf{v}^{n,R}\end{pmatrix}\right\|^2 + \left\|\begin{pmatrix}\psi^{n,q,R}\\\psi^{n,q,I}\end{pmatrix}\right\|^2\right) \tag{5.19}
$$
$$
= c_3\sum_{q=1}^{Q}\left(\|\Delta\mathbf{v}^{n,R}\|^2 + \|\Delta\mathbf{v}^{n,I}\|^2 + \|\psi^{n,q,R}\|^2 + \|\psi^{n,q,I}\|^2\right)
$$
$$
\le c_3\sum_{q=1}^{Q}\left(\|\Delta\mathbf{v}^{n,R}\|^2 + \|\Delta\mathbf{v}^{n,I}\|^2 + 2c_4\sum_{m=1}^{M}\left(\|\Delta\mathbf{w}_m^{n,R}\|^2 + \|\Delta\mathbf{w}_m^{n,I}\|^2\right)\right)
$$
$$
\le c_6\left(\sum_{m=1}^{M}\left(\|\Delta\mathbf{w}_m^{n,R}\|^2 + \|\Delta\mathbf{w}_m^{n,I}\|^2\right) + \|\Delta\mathbf{v}^{n,R}\|^2 + \|\Delta\mathbf{v}^{n,I}\|^2\right),
$$

$$\frac{1}{2}\sum_{q=1}^{Q}\left(\mu_{qR}''(t_1^{n,q})(S^{n+1,q,R}-S^{n,q,R})^2+\mu_{qI}''(t_2^{n,q})(S^{n+1,q,I}-S^{n,q,I})^2\right)$$

$$\leq\frac{c_3}{2}\sum_{q=1}^{Q}\left((S^{n+1,q,R}-S^{n,q,R})^2+(S^{n+1,q,I}-S^{n,q,I})^2\right)$$

$$\leq\frac{c_3}{2}\sum_{q=1}^{Q}\left(\left(\begin{pmatrix}\Delta\mathbf{v}^{n,R}\\-\Delta\mathbf{v}^{n,I}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{H}^{n+1,q,R}\\\mathbf{H}^{n+1,q,I}\end{pmatrix}+\begin{pmatrix}\mathbf{v}^{n,R}\\-\mathbf{v}^{n,I}\end{pmatrix}\cdot\begin{pmatrix}\psi^{n,q,R}\\\psi^{n,q,I}\end{pmatrix}\right)^2\right.$$

$$\left.+\left(\begin{pmatrix}\Delta\mathbf{v}^{n,I}\\\Delta\mathbf{v}^{n,R}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{H}^{n+1,q,R}\\\mathbf{H}^{n+1,q,I}\end{pmatrix}+\begin{pmatrix}\mathbf{v}^{n,I}\\\mathbf{v}^{n,R}\end{pmatrix}\cdot\begin{pmatrix}\psi^{n,q,R}\\\psi^{n,q,I}\end{pmatrix}\right)^2\right) \tag{5.20}$$

$$\leq c_3 Q\max\{(c_0)^2+(c_2)^2\}\left(\|\Delta\mathbf{v}^{n,R}\|^2+\|\Delta\mathbf{v}^{n,I}\|^2+\|\psi^{n,q,R}\|^2+\|\psi^{n,q,I}\|^2\right)$$

$$\leq c_7\left(\sum_{m=1}^{M}\left(\|\Delta\mathbf{w}_m^{n,R}\|^2+\|\Delta\mathbf{w}_m^{n,I}\|^2\right)+\|\Delta\mathbf{v}^{n,R}\|^2+\|\Delta\mathbf{v}^{n,I}\|^2\right),$$

where $c_6=Qc_3\max\{1,2c_4\}$ and $c_7=Qc_3\max\{(c_0)^2+(c_2)^2\}\max\{1,2c_4\}$. So we obtain (5.8) and (5.9). $\qquad\square$

Now, we are ready to prove Theorem 3.1 in terms of the above two lemmas.

Proof of Theorem 3.1. (i) By (5.6)–(5.9) and the Taylor's formula, we have

$$E(\mathbf{W}^{n+1})-E(\mathbf{W}^n)$$

$$=\sum_{q=1}^{Q}\left(\mu_{qR}(S^{n+1,q,R})-\mu_{qR}(S^{n,q,R})+\mu_{qI}(S^{n+1,q,I})-\mu_{qI}(S^{n,q,I})\right)$$

$$=\sum_{q=1}^{Q}\left(\mu_{qR}'(S^{n,q,R})(S^{n+1,q,R}-S^{n,q,R})+\mu_{qI}'(S^{n,q,I})(S^{n+1,q,I}-S^{n,q,I})\right.$$

$$\left.+\frac{1}{2}\mu_{qR}''(t_1^{n,q})(S^{n+1,q,R}-S^{n,q,R})^2+\frac{1}{2}\mu_{qI}''(t_2^{n,q})(S^{n+1,q,I}-S^{n,q,I})^2\right)$$

$$=\sum_{q=1}^{Q}\left(\mu_{qR}'(S^{n,q,R})\begin{pmatrix}\Delta\mathbf{v}^{n,R}\\-\Delta\mathbf{v}^{n,I}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{H}^{n,q,R}\\\mathbf{H}^{n,q,I}\end{pmatrix}+\mu_{qI}'(S^{n,q,I})\begin{pmatrix}\Delta\mathbf{v}^{n,I}\\\Delta\mathbf{v}^{n,R}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{H}^{n,q,R}\\\mathbf{H}^{n,q,I}\end{pmatrix}\right.$$

$$+\mu_{qR}'(S^{n,q,R})\begin{pmatrix}\mathbf{v}^{n,R}\\-\mathbf{v}^{n,I}\end{pmatrix}\cdot\begin{pmatrix}\psi^{n,q,R}\\\psi^{n,q,I}\end{pmatrix}+\mu_{qI}'(S^{n,q,I})\begin{pmatrix}\mathbf{v}^{n,I}\\\mathbf{v}^{n,R}\end{pmatrix}\cdot\begin{pmatrix}\psi^{n,q,R}\\\psi^{n,q,I}\end{pmatrix}$$

$$+\mu_{qR}'(S^{n,q,R})\begin{pmatrix}\Delta\mathbf{v}^{n,R}\\-\Delta\mathbf{v}^{n,I}\end{pmatrix}\cdot\begin{pmatrix}\psi^{n,q,R}\\\psi^{n,q,I}\end{pmatrix}+\mu_{qI}'(S^{n,q,I})\begin{pmatrix}\Delta\mathbf{v}^{n,I}\\\Delta\mathbf{v}^{n,R}\end{pmatrix}\cdot\begin{pmatrix}\psi^{n,q,R}\\\psi^{n,q,I}\end{pmatrix}$$

$$\left.+\frac{1}{2}\mu_{qR}'(t_1^{n,q})(S^{n+1,q,R}-S^{n,q,R})^2+\frac{1}{2}\mu_{qI}''(t_2^{n,q})(S^{n+1,q,I}-S^{n,q,I})^2\right)$$

$$\leq -\frac{1}{\eta}\left(\|\Delta \mathbf{v}^{n,R}\|^2 + \|\Delta \mathbf{v}^{n,I}\|^2\right) + \left(c_5 - \frac{1}{\eta}\right)\sum_{m=1}^{M}\left(\|\Delta \mathbf{w}_m^{n,R}\|^2 + \|\Delta \mathbf{w}_m^{n,I}\|^2\right)$$

$$+ c_6\left(\sum_{m=1}^{M}\left(\|\Delta \mathbf{w}_m^{n,R}\|^2 + \|\Delta \mathbf{w}_m^{n,I}\|^2\right) + \|\Delta \mathbf{v}^{n,R}\|^2 + \|\Delta \mathbf{v}^{n,I}\|^2\right)$$

$$+ c_7\left(\sum_{m=1}^{M}\left(\|\Delta \mathbf{w}_m^{n,R}\|^2 + \|\Delta \mathbf{w}_m^{n,I}\|^2\right) + \|\Delta \mathbf{v}^{n,R}\|^2 + \|\Delta \mathbf{v}^{n,I}\|^2\right)$$

$$\leq \left(c_8 - \frac{1}{\eta}\right)\left(\sum_{m=1}^{M}\left(\|\Delta \mathbf{w}_m^{n,R}\|^2 + \|\Delta \mathbf{w}_m^{n,I}\|^2\right) + \|\Delta \mathbf{v}^{n,R}\|^2 + \|\Delta \mathbf{v}^{n,I}\|^2\right),$$

$$(5.21)$$

where $c_8 = c_5 + c_6 + c_7$, $t_1^{n,q} \in \mathbb{R}^1$ is on the segment between $S^{n+1,q,R}$ and $S^{n,q,R}$, and $t_2^{n,q} \in \mathbb{R}^1$ is on the segment between $S^{n+1,q,I}$ and $S^{n,q,I}$. Then we have

$$E(\mathbf{W}^{n+1}) \leq E(\mathbf{W}^n) - \left(\frac{1}{\eta} - c_8\right)\left(\sum_{m=1}^{M}\left(\|\Delta \mathbf{w}_m^{n,R}\|^2 + \|\Delta \mathbf{w}_m^{n,I}\|^2\right) + \|\Delta \mathbf{v}^{n,R}\|^2 + \|\Delta \mathbf{v}^{n,I}\|^2\right).$$

$$(5.22)$$

Obviously, by choosing the learning rate η to satisfy that

$$0 < \eta < \frac{1}{c_8},\tag{5.23}$$

then we have

$$E(\mathbf{W}^{n+1}) \leq E(\mathbf{W}^n), \quad n = 0, 1, 2, \ldots.\tag{5.24}$$

(ii) According to (2.16), we have

$$\sum_{m=1}^{M}\left(\|\Delta \mathbf{w}_m^{n,R}\|^2 + \|\Delta \mathbf{w}_m^{n,I}\|^2\right) + \|\Delta \mathbf{v}^{n,R}\|^2 + \|\Delta \mathbf{v}^{n,I}\|^2$$

$$= \eta^2\left(\sum_{m=1}^{M}\left(\left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{w}_m^R}\right\|^2 + \left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{w}_m^I}\right\|^2\right) + \left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{v}^R}\right\|^2 + \left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{v}^I}\right\|^2\right).$$

$$(5.25)$$

Combining with (5.21), we have

$$E(\mathbf{W}^{n+1}) \leq E(\mathbf{W}^n) - \alpha\left(\sum_{m=1}^{M}\left(\left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{w}_m^R}\right\|^2 + \left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{w}_m^I}\right\|^2\right) + \left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{v}^R}\right\|^2 + \left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{v}^I}\right\|^2\right)$$

$$\leq \cdots$$

$$\leq E(\mathbf{W}^0) - \alpha\sum_{k=0}^{n}\left(\sum_{m=1}^{M}\left(\left\|\frac{\partial E(\mathbf{W}^k)}{\partial \mathbf{w}_m^R}\right\|^2 + \left\|\frac{\partial E(\mathbf{W}^k)}{\partial \mathbf{w}_m^I}\right\|^2\right) + \left\|\frac{\partial E(\mathbf{W}^k)}{\partial \mathbf{v}^R}\right\|^2 + \left\|\frac{\partial E(\mathbf{W}^k)}{\partial \mathbf{v}^I}\right\|^2\right),$$

$$(5.26)$$

where $\alpha = (\eta - c_8\eta^2)$. Since $E(\mathbf{W}^{n+1}) \geq 0$, there holds that

$$\alpha\sum_{k=0}^{n}\left(\sum_{m=1}^{M}\left(\left\|\frac{\partial E(\mathbf{W}^k)}{\partial \mathbf{w}_m^R}\right\|^2 + \left\|\frac{\partial E(\mathbf{W}^k)}{\partial \mathbf{w}_m^I}\right\|^2\right) + \left\|\frac{\partial E(\mathbf{W}^k)}{\partial \mathbf{v}^R}\right\|^2 + \left\|\frac{\partial E(\mathbf{W}^k)}{\partial \mathbf{v}^I}\right\|^2\right) \leq E(\mathbf{W}^0). \quad (5.27)$$

Let $n \to \infty$, then

$$\alpha\sum_{k=0}^{\infty}\left(\sum_{m=1}^{M}\left(\left\|\frac{\partial E(\mathbf{W}^k)}{\partial \mathbf{w}_m^R}\right\|^2 + \left\|\frac{\partial E(\mathbf{W}^k)}{\partial \mathbf{w}_m^I}\right\|^2\right) + \left\|\frac{\partial E(\mathbf{W}^k)}{\partial \mathbf{v}^R}\right\|^2 + \left\|\frac{\partial E(\mathbf{W}^k)}{\partial \mathbf{v}^I}\right\|^2\right) \leq E(\mathbf{W}^0) < \infty.$$

$$(5.28)$$

So there holds that

$$\lim_{n \to \infty}\left(\sum_{m=1}^{M}\left(\left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{w}_m^R}\right\|^2 + \left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{w}_m^I}\right\|^2\right) + \left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{v}^R}\right\|^2 + \left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{v}^I}\right\|^2\right) = 0, \quad (5.29)$$

which implies that

$$\lim_{n \to \infty}\left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{w}_m^R}\right\| = \lim_{n \to \infty}\left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{w}_m^I}\right\| = 0, \quad 0 \leq m \leq M, \quad (5.30)$$

$$\lim_{n \to \infty}\left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{v}^R}\right\| = \lim_{n \to \infty}\left\|\frac{\partial E(\mathbf{W}^n)}{\partial \mathbf{v}^I}\right\| = 0. \quad (5.31)$$

(iii) Write

$$\boldsymbol{\theta} = \left(\left(\mathbf{w}_1^R\right)^T, \ldots, \left(\mathbf{w}_M^R\right)^T, \left(\mathbf{w}_1^I\right)^T, \ldots, \left(\mathbf{w}_M^I\right)^T, \left(\mathbf{v}^R\right)^T, \left(\mathbf{v}^I\right)^T\right)^T, \quad (5.32)$$

then $E(\mathbf{W})$ can be looked as a function of $\boldsymbol{\theta}$, which is denoted as $\mathcal{E}(\boldsymbol{\theta})$. That is to say

$$E(\mathbf{W}) \equiv \mathcal{E}(\boldsymbol{\theta}). \quad (5.33)$$

Obviously, $\mathcal{E}(\boldsymbol{\theta})$ is a continuously differentiable real-valued function and

$$
\begin{aligned}
\frac{\partial \mathcal{E}(\boldsymbol{\theta})}{\partial \mathbf{v}^R} &= \frac{\partial E(\mathbf{W})}{\partial \mathbf{v}^R}, \qquad \frac{\partial \mathcal{E}(\boldsymbol{\theta})}{\partial \mathbf{v}^I} = \frac{\partial E(\mathbf{W})}{\partial \mathbf{v}^I}, \\
\frac{\partial \mathcal{E}(\boldsymbol{\theta})}{\partial \mathbf{w}_m^R} &= \frac{\partial E(\mathbf{W})}{\partial \mathbf{w}_m^R}, \quad \frac{\partial \mathcal{E}(\boldsymbol{\theta})}{\partial \mathbf{w}_m^I} = \frac{\partial E(\mathbf{W})}{\partial \mathbf{w}_m^I}, \quad m = 1, \dots, M.
\end{aligned}
\tag{5.34}
$$

Let

$$
\boldsymbol{\theta}^n = \left(\left(\mathbf{w}_1^{n,R}\right)^T, \dots, \left(\mathbf{w}_M^{n,R}\right)^T, \left(\mathbf{w}_1^{n,I}\right)^T, \dots, \left(\mathbf{w}_M^{n,I}\right)^T, \left(\mathbf{v}^{n,R}\right)^T, \left(\mathbf{v}^{n,I}\right)^T \right)^T,
\tag{5.35}
$$

then by (5.30) and (5.31), we have

$$
\lim_{n \to \infty} \left\| \frac{\partial \mathcal{E}(\boldsymbol{\theta}^n)}{\partial \mathbf{w}_m^R} \right\| = \lim_{n \to \infty} \left\| \frac{\partial \mathcal{E}(\boldsymbol{\theta}^n)}{\partial \mathbf{w}_m^I} \right\| = 0, \quad 0 \le m \le M,
\tag{5.36}
$$

$$
\lim_{n \to \infty} \left\| \frac{\partial \mathcal{E}(\boldsymbol{\theta}^n)}{\partial \mathbf{v}^R} \right\| = \lim_{n \to \infty} \left\| \frac{\partial \mathcal{E}(\boldsymbol{\theta}^n)}{\partial \mathbf{v}^I} \right\| = 0.
\tag{5.37}
$$

Thus we have

$$
\lim_{n \to \infty} \left\| \frac{\partial \mathcal{E}(\boldsymbol{\theta}^n)}{\partial \boldsymbol{\theta}} \right\| = 0.
\tag{5.38}
$$

We use (2.16), (5.30), and (5.31) to obtain

$$
\lim_{n \to \infty} \left\| \mathbf{w}_m^{n+1,R} - \mathbf{w}_m^{n,R} \right\| = 0, \quad \lim_{n \to \infty} \left\| \mathbf{w}_m^{n+1,I} - \mathbf{w}_m^{n,I} \right\| = 0, \quad m = 1, \dots, M,
\tag{5.39}
$$

$$
\lim_{n \to \infty} \left\| \mathbf{v}^{n+1,R} - \mathbf{v}^{n,R} \right\| = 0, \quad \lim_{n \to \infty} \left\| \mathbf{v}^{n+1,I} - \mathbf{v}^{n,I} \right\| = 0.
\tag{5.40}
$$

This leads to

$$
\lim_{n \to \infty} \left\| \boldsymbol{\theta}^{n+1} - \boldsymbol{\theta}^n \right\| = 0.
\tag{5.41}
$$

Furthermore, from Assumption (A3) we know that the set $\{ \boldsymbol{\theta} \mid (\partial \mathcal{E}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}) = 0 \}$ contains only finite points. Thus, the sequence $\{\boldsymbol{\theta}^n\}_{n=1}^{\infty}$ here satisfies all the conditions needed in Lemma 5.1. As a result, there is a $\theta^\star$ which satisfies that $\lim_{n \to \infty} \boldsymbol{\theta}^n = \boldsymbol{\theta}^\star$. Since $\boldsymbol{\theta}^n$ consists of the real and imaginary parts of $\mathbf{W}^n$, we know that there is a $\mathbf{W}^\star$ such that $\lim_{n \to \infty} \mathbf{W}^n = \mathbf{W}^\star$. We thus complete the proof. □

6. Conclusion

In this paper, some convergence results of BSCBP algorithm for CVNN are presented. An up-bound of the learning rate η is given to guarantee both the monotonicity of the error

function and the convergence of the gradients for the error function. It is also proved that the network weights vector tends to a local minimum if there are only finite stable points for the error function. A numerical example is given to support the theoretical findings. Our work can help the neural network researchers to choose the appropriate activation function and learning rate to guarantee the convergence of the algorithm when they use BSCBP algorithm to train CVNN. We mention that the convergence results can be extended to a more general case that the networks have several outputs and hidden layers.

Acknowledgments

References

[1] J. Ma and L. Liu, "Multivariate nonlinear analysis and prediction of Shanghai stock market," *Discrete Dynamics in Nature and Society*, vol. 2008, Article ID 526734, 8 pages, 2008.

[2] G. M. Georgiou and C. Koutsougeras, "Complex domain backpropagation," *IEEE Transactions on Circuits and Systems II*, vol. 39, no. 5, pp. 330–334, 1992.

[3] N. Benvenuto and F. Piazza, "On the complex backpropagation algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 4, pp. 967–969, 1992.

[4] A. Hirose, *Complex-Valued Neural Networks*, Springer, New York, NY, USA, 2006.

[5] T. Nitta, "Orthogonality of decision boundaries in complex-valued neural networks," *Neural Computation*, vol. 16, no. 1, pp. 73–97, 2004.

[6] T. Kim and T. Adali, "Fully complex backpropagation for constant envelop signal processing," in *Proceedings of the IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing X*, vol. 1, pp. 231–240, Sydney, Australia, December 2000.

[7] S.-S. Yang, S. Siu, and C.-L. Ho, "Analysis of the initial values in split-complex backpropagation algorithm," *IEEE Transactions on Neural Networks*, vol. 19, no. 9, pp. 1564–1573, 2008.

[8] Y. Chen, W. Bi, and Y. Wu, "Delay-dependent exponential stability for discrete-time BAM neural networks with time-varying delays," *Discrete Dynamics in Nature and Society*, vol. 2008, Article ID 421614, 14 pages, 2008.

[9] Q. Zhang, X. Wei, and J. Xu, "On global exponential stability of discrete-time Hopfield neural networks with variable delays," *Discrete Dynamics in Nature and Society*, vol. 2007, Article ID 67675, 9 pages, 2007.

[10] A. I. Hanna and D. P. Mandic, "A data-reusing nonlinear gradient descent algorithm for a class of complex-valued neural adaptive filters," *Neural Processing Letters*, vol. 17, no. 1, pp. 85–91, 2003.

[11] S. L. Goh and D. P. Mandic, "Stochastic gradient-adaptive complex-valued nonlinear neural adaptive filters with a gradient-adaptive step size," *IEEE Transactions on Neural Networks*, vol. 18, no. 5, pp. 1511–1516, 2007.

[12] T. Nitta, "An extension of the back-propagation algorithm to complex numbers," *Neural Networks*, vol. 10, no. 8, pp. 1391–1415, 1997.

[13] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, NY, USA, 1970.