Convergence of Split-Complex Backpropagation Algorithm with Momentum

HUISHENG ZHANG\textsuperscript{1}, WEI WU\textsuperscript{2}\textsuperscript{*}

\textsuperscript{1}Department of Mathematics, Dalian Maritime University, Dalian, China, 116026
\textsuperscript{2}Applied Mathematics Department, Dalian University of Technology, Dalian, China, 116024

Abstract

This paper investigates a split-complex backpropagation algorithm with momentum (SCBPM) for complex-valued neural networks. Some convergence results for SCBPM are proved under relaxed conditions compared with existing results. The monotonicity of the error function during the training iteration process is also guaranteed. Two numerical examples are given to support the theoretical findings.

Keywords Complex-valued neural networks; Split-complex backpropagation algorithm; Convergence.

1 Introduction

In recent years great interest has been aroused in the complex-valued neural networks (CVNN) for their powerful capability in processing complex-valued signals \cite{1, 2, 3}. CVNN are extensions of real-valued neural networks \cite{4, 5}. Fully complex BP algorithm and split-complex BP algorithm are two types of complex backpropagation algorithms for training CVNN. Different from the fully complex BP algorithm \cite{6, 7}, the operation of activation function in the split-complex BP algorithm is split into real part and imaginary part \cite{2, 4, 5, 8}. This split-complex BP algorithm avoids the occurrence of singular points in the adaptive training process. This paper considers the split-complex BP algorithms.

An important issue of a successful training algorithm for neural networks is its convergence properties. The convergence of BP algorithm for real-valued neural networks has been analyzed by many authors from different aspects\cite{9, 10, 11}. By using the contraction mapping theorem, the convergence in the mean and in the mean square for recurrent neurons has been obtained by Mandic and Goh \cite{12, 13}. For recent convergence analyses of complex-valued perceptrons and BP algorithm for complex-valued neural networks, we refer the readers to \cite{1, 14, 15}.

It is well known that the learning process of BP algorithm can be very slow. To speed up and stabilize the learning procedure, a momentum is often added to BP algorithm. BP algorithm with momentum (BPM in short) can be viewed as a memory gradient method in optimization theory \cite{16}. Starting from a point close enough to a minimum point of the objective function, the memory gradient method converges under certain conditions \cite{16, 17}. This local convergence result is also obtained in \cite{18} where the convergence of BPM for neural network is considered. However, these results cannot be applied to a more important case where the initial weights are chosen stochastically. Some convergence results for BPM are given in \cite{19}, which are of global nature in that they are valid for arbitrarily given initial values of the weights. Our contribution in this paper is to present some convergence results of

\textsuperscript{*}This work is partly supported by the National Natural Science Foundation of China (10871220).
\textsuperscript{1}Corresponding author. E-mail: wuweiw@dlut.edu.cn
the split-complex backpropagation algorithm with momentum (SCBPM). We borrow some ideas from [19], but we employ some different proof techniques resulting in a new learning rate restriction which is much relaxed and easier to check than the counterpart in [19]. Actually, our approach here can be applied to BPM in [19] to relax the learning rate restriction there. Two numerical example is given to support our theoretical findings. We also mention a recent paper on split quaternion nonlinear adaptive filtering [20], and we expect to extend our results in future to quaternion networks.

The remainder of this paper is organized as follows. A CVNN model with one hidden layer and one output neuron. Let the numbers of input and hidden neurons be $L$ and $M$, respectively. We write $w_m = w_m^R + i w_m^I = (w_{m1}, w_{m2}, \cdots, w_{mL})^T \in \mathbb{C}^L$ as the weight vector between the input neurons and $m$th hidden neuron, where $w_{ml}^R = w_{ml}^R + iw_{ml}^I$, $w_{ml}^R \in \mathbb{R}$ and $w_{ml}^I \in \mathbb{R}$, $i = \sqrt{-1}$, $m = 1, \cdots, M$, and $l = 1, \cdots, L$. Similarly, write $w_{M+1} = w_{M+1}^R + i w_{M+1}^I = (w_{M+1,1}, w_{M+1,2}, \cdots, w_{M+1,M})^T \in \mathbb{C}^M$ as the weight vector between the hidden neurons and the output neuron, where $w_{M+1,m} = w_{M+1,m}^R + iw_{M+1,m}$, $m = 1, \cdots, M$. For simplicity, all the weight vectors are incorporated into a total weight vector

$$W = ((w_1)^T, (w_2)^T, \cdots, (w_{M+1})^T)^T \in \mathbb{C}^{ML+M}. \quad (1)$$

For an input signal $z = (z_1, z_2, \cdots, z_L)^T = x + iy \in \mathbb{C}^L$, where $x = (x_1, x_2, \cdots, x_L)^T \subset \mathbb{R}^L$, and $y = (y_1, y_2, \cdots, y_L)^T \subset \mathbb{R}^L$, the input to the $m$th hidden neuron is

$$U_m = U_m^R + iU_m^I$$

$$= \sum_{l=1}^L (w_{ml}^R x_l - w_{ml}^I y_l) + i \sum_{l=1}^L (w_{ml}^I x_l + w_{ml}^R y_l)$$

$$= x^T w_m^R - y^T w_m^I + i(x^T w_m^I + y^T w_m^R). \quad (2)$$

Here "$(\cdot)^T$" is the vector transpose operator.
In order to apply SCBPM to train CVNN, we consider the following popular real-imaginary-type activation function [22]

\[ f_C(U) = f(U^R) + if(U^I) \]  

(3)

for any \( U = U^R + iU^I \in \mathbb{C} \), where \( f \) is a given real function (e.g., a sigmoid function). The output \( H_m \) of the hidden neuron \( m \) is given by:

\[ H_m = H_m^R + iH_m^I = f(U_m^R) + if(U_m^I). \]  

(4)

Similarly, the input to the output neuron is

\[ S = S^R + iS^I \]

\[ = \sum_{m=1}^{M} (w_{M+1,m}^R H_m^R - w_{M+1,m}^I H_m^I) + i \sum_{m=1}^{M} (w_{M+1,m}^I H_m^R + w_{M+1,m}^R H_m^I) \]

\[ = (H^R)^T w_{M+1}^R - (H^I)^T w_{M+1}^I + i ((H^R)^T w_{M+1}^I + (H^I)^T w_{M+1}^R) \]  

(5)

and the output of the network is given by

\[ O = O^R + iO^I = g(S^R) + ig(S^I), \]  

(6)

where \( H^R = (H_1^R, H_2^R, \ldots, H_M^R)^T \), \( H^I = (H_1^I, H_2^I, \ldots, H_M^I)^T \), and \( g \) is a given real function. A bias weight can be added to each neurons, but it is omitted for simplicity of the presentation and deduction.

Let the network be supplied with a given set of training examples \( \{z^j, d^j\}_{j=1}^J \subset \mathbb{C}^L \times \mathbb{C} \). For each input \( z^j = x^j + iy^j \) (1 \( \leq j \leq J \)) from the training set, we write \( U_m^j = U_m^jR + iU_m^jI \) (1 \( \leq m \leq M \)) as the input for the hidden neuron \( m \), \( H_m^j = H_m^jR + iH_m^jI \) (1 \( \leq m \leq M \)) as the output for the hidden neuron \( m \), \( S^j = S^jR + iS^jI \) as the input to the output neuron, and \( O^j = O^jR + iO^jI \) as the final output. The square error function of CVNN trained by the SCBPM can be represented as follows:

\[ E(W) = \frac{1}{2} \sum_{j=1}^{J} (O^j - d^j)(O^j - d^j)^* \]

\[ = \frac{1}{2} \sum_{j=1}^{J} [ (O^j)^R - d^j)^R \] + (O^j)^I - d^j)^I \]

\[ = \sum_{j=1}^{J} [ g_{jR}(S^jR) + g_{jI}(S^jI) ] , \]  

(7)

where “\(^*\)” denotes the complex conjugate operator, \( d^jR \) and \( d^jI \) are the real part and imaginary part of the desired output \( d^j \) respectively, and

\[ g_{jR}(t) = \frac{1}{2} (g(t) - d^j)^2, g_{jI}(t) = \frac{1}{2} (g(t) - d^j)^2, t \in \mathbb{R}, 1 \leq j \leq J. \]  

(8)

The purpose of the network training is to find \( W^* \) to minimize \( E(W) \). By writing

\[ H^j = H^jR + iH^jI \]

\[ = (H_1^jR, H_2^jR, \ldots, H_M^jR)^T + i(H_1^jI, H_2^jI, \ldots, H_M^jI)^T , \]  

(9)

we get the gradients of \( E(W) \) with respect to \( w^R_m \) and \( w^I_m \) respectively as follows

\[ \frac{\partial E(W)}{\partial w^R_{M+1}} = \sum_{j=1}^{J} [ g_{jR}(S^jR)H^jR + g_{jI}(S^jI)H^jI ] , \]  

(10)
\[
\frac{\partial E(W)}{\partial w_{M+1}^I} = \sum_{j=1}^{J} \left[-g_j R(S_{j,R})H_{j,I} + g_j I(S_{j,I})H_{j,R} \right],
\]
(11)
\[
\frac{\partial E(W)}{\partial w_m^I} = \sum_{j=1}^{J} \left[g_j R(S_{j,R})(w_{M+1,m}Rf'(U_{m}^R)x_j - w_{M+1,m}If'(U_{m}^I)y_j) + g_j I(S_{j,I})(w_{M+1,m}If'(U_{m}^I)x_j)\right], \quad 1 \leq m \leq M,
\]
(12)
\[
\frac{\partial E(W)}{\partial w_m^I} = \sum_{j=1}^{J} \left[g_j R(S_{j,R})(-w_{M+1,m}Rf'(U_{m}^R)y_j - w_{M+1,m}If'(U_{m}^I)x_j) + g_j I(S_{j,I})(-w_{M+1,m}If'(U_{m}^I)x_j)\right], \quad 1 \leq m \leq M.
\]
(13)

Write \( W^n = ((w^n_1)^T, (w^n_2)^T, \ldots, (w^n_{M+1})^T)^T \) \((n = 0, 1, \ldots)\). Let \( w^0_m = w_{0,R}^m + iw_{0,I}^m \) be arbitrarily chosen initial weights. Let \( \Delta w_{0,R}^0 = \Delta w_{0,I}^0 = 0 \). Then SCBPM algorithm updates the real part \( w_{m}^R \) and the imaginary part \( w_{m}^I \) of the weights \( w_m \) separately:
\[
\Delta w_{m+1,R} = w_{m+1,R} - w_m = -\eta \frac{\partial E(W^n)}{\partial w_m^R} + \tau_{m,R} \Delta w_{n,R},
\]
\[
\Delta w_{m+1,I} = w_{m+1,I} - w_m = -\eta \frac{\partial E(W^n)}{\partial w_m^I} + \tau_{m,I} \Delta w_{n,I},
\]
(14)

where \( \eta \in (0, 1) \) is the learning rate, \( \tau_{m,R} \) and \( \tau_{m,I} \) are the momentum factors, \( m = 1, 2, \ldots, M + 1 \), and \( n = 0, 1, \ldots \).

For simplicity, let us denote
\[
p_m^{R} = \frac{\partial E(W^n)}{\partial w_m^R},
\]
\[
p_m^{I} = \frac{\partial E(W^n)}{\partial w_m^I}.
\]
(15)

Then (14) can be rewritten as
\[
\Delta w_{m+1,R} = \tau_{m,R} \Delta w_{m,R} - \eta p_m^{R},
\]
\[
\Delta w_{m+1,I} = \tau_{m,I} \Delta w_{m,I} - \eta p_m^{I}.
\]
(16)

Similar to BPM [19], we choose the adaptive momentum factor \( \tau_{m,R}^{n} \) and \( \tau_{m,I}^{n} \) as follows
\[
\tau_{m,R}^{n} = \begin{cases} \frac{\tau \|p_m^{R}\|}{\|\Delta w_m^{R}\|}, & \text{if } \|\Delta w_m^{R}\| \neq 0, \\ 0, & \text{else} \end{cases}
\]
(17)
\[
\tau_{m,I}^{n} = \begin{cases} \frac{\tau \|p_m^{I}\|}{\|\Delta w_m^{I}\|}, & \text{if } \|\Delta w_m^{I}\| \neq 0, \\ 0, & \text{else} \end{cases}
\]
(18)

where \( \tau \in (0, 1) \) is a constant parameter and \( \| \cdot \| \) is the usual Euclidian norm.

### 3 Main results

The following assumptions will be used in our discussion.

(A1): There exists a constant \( C_1 > 0 \) such that
\[
\max_{t \in \mathbb{R}} \{|f(t)|, |g(t)|, |f'(t)|, |g'(t)|, |f''(t)|, |g''(t)|\} \leq C_1.
\]
There exists a constant $C_2 > 0$ such that $\|w_{M+1}^n\| \leq C_2$ and $\|w_{M+1}^n\| \leq C_2$ for all $n = 0, 1, 2, \cdots$.

The set $\Phi_0 = \{W, \frac{\partial E(W)}{\partial w_m} = 0, \frac{\partial E(W)}{\partial w_m} = 0, m = 1, \cdots, M + 1\}$ contains only finite points.

**Theorem 1** Suppose that Assumptions (A1) and (A2) are valid, and that $\{w^n_m\}$ are the weight vector sequences generated by (14). Then, there exists a constant $C^* > 0$ such that for $0 < s < 1, \tau = s\eta$ and $\eta \leq \frac{1-s}{C^*[1+s]}$, the following results hold:

(i) $E(W^{n+1}) \leq E(W^n), \quad n = 0, 1, 2, \cdots$;

(ii) There is $E^* \geq 0$ such that $\lim_{n \to \infty} E(W^n) = E^*$;

(iii) $\lim_{n \to \infty} \left\| \frac{\partial E(W^n)}{\partial w_m} \right\| = 0$ and $\lim_{n \to \infty} \left\| \frac{\partial E(W^n)}{\partial w_m} \right\| = 0, \quad 0 \leq m \leq M + 1$.

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

(iv) $\lim_{n \to \infty} W^n = W^*$.

The monotonicity and the convergence of the error function $E(W)$ during the learning process are shown in Conclusions (i) and (ii), respectively. Conclusion (iii) indicates the convergence of $\frac{\partial E(W^n)}{\partial w_m}$ and $\frac{\partial E(W^n)}{\partial w_m}$, referred to as weak convergence. The strong convergence of $W^n$ is given in Conclusion (iv). We remark that the restriction $\eta \leq \frac{1-s}{C^*[1+s]}$ in Theorem 1 is less restrictive and easier to check than the corresponding condition in [19]. We also mention that our results are of deterministic nature compared with a related work in [1], where the convergence in the mean and in the mean square for complex-valued perceptrons is obtained.

### 4 Numerical Example

In the following subsections we illustrate the convergence behavior of the SCBPM by using two numerical examples. In both examples, we set the transfer function to be `tansig(·)` in MATLAB, which is a commonly used sigmoid function, and carry out 10 independent tests with the initial components of the weights stochastically chosen in $[-0.5, 0.5]$. The average of errors and the average of gradient norms for all the tests in each example are plotted.

#### 4.1 XOR Problem

The well-known XOR problem is a benchmark in literature of neural networks. As in [22], the training samples of the encoded XOR problem for CVNN is presented as follows:

\[
\{z^1 = -1 - i, d^1 = 1\}, \{z^2 = -1 + i, d^2 = 0\}, \\
\{z^3 = 1 - i, d^3 = 1 + i\}, \{z^4 = 1 + i, d^4 = i\}.
\]

This example uses a network with one input node, three hidden nodes, and one output node. The learning rate $\eta$ and the momentum parameter $\tau$ are set to be 0.1 and 0.01, respectively. The simulation results are shown in figure 2, which 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.

#### 4.2 Approximation Problem

In this example, the synthetic complex-valued function [23] defined as

\[h(z) = (z_1)^2 + (z_2)^2\] (19)
Figure 2: Convergence behavior of SCBPM for solving XOR problem. (norm of gradient = $\sum_{m=1}^{M+1} (\|p^{n,R}_m\|^2 + \|p^{n,I}_m\|^2)$)

is approximated, where $z$ is a two dimensional complex-valued vector comprised of $z_1$ and $z_2$. 10000 input points are selected from an evenly spaced $10 \times 10 \times 10 \times 10$ grid on $-0.5 \leq \text{Re}(z_1), \text{Im}(z_1), \text{Re}(z_2), \text{Im}(z_2) \leq 0.5$, where $\text{Re}()$ and $\text{Im}()$ represents the real part and imaginary part of a complex number, respectively. We use a network with 2 input neurons, 25 hidden neurons and 1 output neuron. The learning rate $\eta$ and the momentum parameter $\tau$ are set to be 0.1 and 0.01, respectively. Figure 3 illustrates the simulation results, which also supports our convergence theorem.

5 Proofs

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

Lemma 1 Suppose that $E : R^{2ML+2M} \rightarrow R$ is continuous and differentiable on a compact set $\Phi \subset R^{2ML+2M}$, and that $\Phi_1 = \{v | \frac{\partial E(v)}{\partial v} = 0\}$ contains only finite points. If a sequence $\{v^n\}_{n=1}^{\infty} \subset \Phi$ satisfies

$$\lim_{n \to \infty} \|v^{n+1} - v^n\| = 0, \quad \lim_{n \to \infty} \|\nabla E(v^n)\| = 0,$$

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

Proof This result is almost the same as Theorem 14.15 in [21], and the detail of the proof is omitted.

For any $1 \leq j < J$, $1 \leq m \leq M$ and $n = 0, 1, 2, \ldots$, the following symbols will be used in our proof later on:

$$U^{n,j}_m = U^{n,j,R}_m + iU^{n,j,I}_m = (x^j)^T w^{n,R}_m - (y^j)^T w^{n,I}_m + i ((x^j)^T w^{n,R}_m + (y^j)^T w^{n,I}_m),$$

$$H^{n,j}_m = H^{n,j,R}_m + iH^{n,j,I}_m = f(U^{n,j,R}_m) + i f(U^{n,j,I}_m),$$

$$H^{n,j,R} = (H^{n,j,R}_1, \ldots, H^{n,j,R}_M)^T, \quad H^{n,j,I} = (H^{n,j,I}_1, \ldots, H^{n,j,I}_M)^T.$$
Figure 3: Convergence behavior of SCBPM for solving approximation problem. (norm of gradient $= \sum_{m=1}^{M+1} (\|p_{m,R}^{n}\|^2 + \|p_{m,I}^{n}\|^2)$)

\[
S^{n,j} = S^{n,j,R} + iS^{n,j,I} = (H^{n,j,R})T_{M+1}^{n,R} - (H^{n,j,I})T_{M+1}^{n,I} + i \left( (H^{n,j,R})T_{M+1}^{n,I} + (H^{n,j,I})T_{M+1}^{n,R} \right),
\]

\[
\psi^{n,j,R} = H^{n+1,j,R} - H^{n,j,R}, \psi^{n,j,I} = H^{n+1,j,I} - H^{n,j,I}.
\]  

Lemma 2 Suppose Assumptions (A1) and (A2) hold, then for any $1 \leq j \leq J$ and $n = 0, 1, 2, \cdots$, we have

\[
\begin{align*}
|O^{j,R}| &\leq C_0, |O^{j,I}| \leq C_0, \|H^{n,j,R}\| \leq C_0, \|H^{n,j,I}\| \leq C_0, \\
|g_j^R(t)| &\leq C_3, |g_j^I(t)| \leq C_3, |g_j^R(t)| \leq C_3, |g_j^I(t)| \leq C_3, t \in \mathbb{R}, \\
\max\{\|\psi^{n,j,R}\|^2, \|\psi^{n,j,I}\|^2\} &\leq C_4(\tau + \eta)^2 \sum_{m=1}^{M} (\|p_{m,R}^{n}\|^2 + \|p_{m,I}^{n}\|^2), \\
\sum_{j=1}^{J} \left( g_j^R(S^{n,j,R}) \left( (H^{n,j,R})T_{M+1}^{n+1,R} - (H^{n,j,I})T_{M+1}^{n+1,I} \right) \\
+ g_j^I(S^{n,j,I}) \left( (H^{n,j,R})T_{M+1}^{n+1,I} + (H^{n,j,I})T_{M+1}^{n+1,R} \right) \right) \\
&\leq (\tau - \eta)(\|p_{M+1}^{n,R}\|^2 + \|p_{M+1}^{n,I}\|^2), \\
\sum_{j=1}^{J} \left( g_j^R(S^{n,j,R}) \left( (\psi^{n,j,R})T_{M+1}^{n,R} - (\psi^{n,j,I})T_{M+1}^{n,I} \right) \\
+ g_j^I(S^{n,j,I}) \left( (\psi^{n,j,R})T_{M+1}^{n,I} + (\psi^{n,j,I})T_{M+1}^{n,R} \right) \right) \\
&\leq (\tau - \eta + C_5(\tau + \eta)^2) \sum_{m=1}^{M} (\|p_{m,R}^{n}\|^2 + \|p_{m,I}^{n}\|^2), \\
\sum_{j=1}^{J} \left( g_j^R(S^{n,j,R}) \left( (\psi^{n,j,R})T_{M+1}^{n+1,R} - (\psi^{n,j,I})T_{M+1}^{n+1,I} \right) \right)
\end{align*}
\]
\[ + g'_{j1}(S^{n,j,I}) \left( (\psi^{n,j,R})^T \Delta w_{M+1}^{n+1,I} + (\psi^{n,j,I})^T \Delta w_{M+1}^{n+1,R} \right) \]
\[ \leq C_6(\tau + \eta)^2 \sum_{m=1}^{M+1} (\|p_m^{n,R}\|^2 + \|p_m^{n,I}\|^2), \]  
(26)
\[ \frac{1}{2} \sum_{j=1}^{J} (g''_{j1}(t_{1j}^{n,j}))(S^{n+1,j,R} - S^{n,j,R})^2 + g''_{j1}(t_{2j}^{n,j})(S^{n+1,j,I} - S^{n,j,I})^2 \]
\[ \leq C_7(\tau + \eta)^2 \sum_{m=1}^{M+1} (\|p_m^{n,R}\|^2 + \|p_m^{n,I}\|^2), \]  
(27)

where \( C_i \) (\( i = 0, 3, \cdots, 7 \)) are constants independent of \( n \) and \( j \), each \( t_{1j}^{n,j} \in \mathbb{R} \) lies on the segment between \( S^{n+1,j,R} \) and \( S^{n,j,R} \), and each \( t_{2j}^{n,j} \in \mathbb{R} \) lies on the segment between \( S^{n+1,j,I} \) and \( S^{n,j,I} \).

**Proof** The validation of (21) results easily from (4)-(6) when the set of samples are fixed and Assumptions (A1) and (A2) are satisfied. By (8), we have

\[ g_{j1}(t) = g(t)(g(t) - O^{j,R}), \]
\[ g_{j1}'(t) = g'(t)(g(t) - O^{j,I}), \]
\[ g_{j1,2}(t) = g''(t)(g(t) - O^{j,R}) + (g'(t))^2, \]
\[ g''_{j1}(t) = g''(t)(g(t) - O^{j,I}) + (g'(t))^2, \quad 1 \leq j \leq J, t \in \mathbb{R}. \]

Then (22) follows directly from Assumption (A1) by defining \( C_3 = C_1(C_1 + C_0) + (C_1)^2 \).

By (16), (17) and (18), for \( m = 1, \cdots, M+1 \),
\[ \|\Delta w_{m+1}^{n,R}\| = \|\tau_m^{n,R} \Delta w_m^{n,R} - \eta p_m^{n,R}\| \leq \tau_m^{n,R} \|\Delta w_m^{n,R}\| + \eta \|p_m^{n,R}\| \leq (\tau + \eta) \|p_m^{n,R}\|. \]  
(28)

Similarly, we have
\[ \|\Delta w_{m+1}^{n,I}\| \leq (\tau + \eta) \|p_m^{n,I}\|. \]  
(29)

If Assumption (A1) is valid, it follows from (20), (28), (29), the Cauchy-Schwartz Inequality and the Mean-Value Theorem for multivariate functions that for any \( 1 \leq j \leq J \) and \( n = 0, 1, 2, \cdots \)
\[ \|\psi^{n,j,R}\|^2 = \|H^{n+1,j,R} - H^{n,j,R}\|^2 = \left\| \left( \begin{array}{c} f(U_1^{n+1,j,R}) - f(U_1^{n,j,R}) \\ \vdots \\ f(U_M^{n+1,j,R}) - f(U_M^{n,j,R}) \end{array} \right) \right\|^2 \]
\[ = \left\| \left( \begin{array}{c} f'(s_1^{n,j})((x_j)^T \Delta w_1^{n+1,R} - (y_j)^T \Delta w_1^{n+1,I}) \\ \vdots \\ f'(s_M^{n,j})((x_j)^T \Delta w_M^{n+1,R} - (y_j)^T \Delta w_M^{n+1,I}) \end{array} \right) \right\|^2 \]
\[ = \sum_{m=1}^{M} (f'(s_m^{n,j})((x_j)^T \Delta w_m^{n+1,R} - (y_j)^T \Delta w_m^{n+1,I}))^2 \]
\[ \leq 2(C_2)^2 \sum_{m=1}^{M} ((x_j)^T \Delta w_m^{n+1,R})^2 + ((y_j)^T \Delta w_m^{n+1,I})^2 \]
\[ \leq 2(C_2)^2 \sum_{m=1}^{M} (\|\Delta w_m^{n+1,R}\|^2 ||x_j||^2 + ||\Delta w_m^{n+1,I}\|^2 ||y_j||^2) \]
and (34)-(35) we have
\[
\|\psi^{n,j,I}\| \leq C_4(\tau + \eta)^2 \sum_{m=1}^{M} (\|p_m^{n,R}\|^2 + \|p_m^{n,J}\|^2),
\]
(31)

Thus, we have (23).

According to (16), (17) and (18), we have
\[
(p_m^{n,R})^T \Delta w_m^{n+1,R} + (p_m^{n,J})^T \Delta w_m^{n+1,I}
= -\eta(\|p_m^{n,R}\|^2 + \|p_m^{n,J}\|^2) + \tau_m^R(p_m^{n,R})^T \Delta w_m^{n,R} + \tau_m^I(p_m^{n,J})^T \Delta w_m^{n,I}
\leq -\eta(\|p_m^{n,R}\|^2 + \|p_m^{n,J}\|^2) + \tau_m^R \|\Delta w_m^{n,R}\| \|p_m^{n,R}\| + \tau_m^I \|\Delta w_m^{n,I}\| \|p_m^{n,J}\|
\leq (\tau - \eta)(\|p_m^{n,R}\|^2 + \|p_m^{n,J}\|^2),
\]
(32)

where \(m = 1, \ldots, M, M + 1\). This together with (10) and (11) validates (24):
\[
\sum_{j=1}^{J} \left( g'_j(S_m^{n,j,R}) \left( (H_m^{n,j,R})^T \Delta w_{m+1}^{n+1,R} - (H_m^{n,j,I})^T \Delta w_{M+1}^{n+1,I} \right)
+ g'_j(S_m^{n,j,I}) \left( (H_m^{n,j,R})^T \Delta w_{M+1}^{n+1,R} + (H_m^{n,j,I})^T \Delta w_{M+1}^{n+1,I} \right) \right)
= \sum_{j=1}^{J} \left( g'_j(S_m^{n,j,R})(H_m^{n,j,R})^T \Delta w_{M+1}^{n+1,R} + g'_j(S_m^{n,j,I})(H_m^{n,j,I})^T \Delta w_{M+1}^{n+1,I}
- g'_j(S_m^{n,j,R})(H_m^{n,j,I})^T \Delta w_{M+1}^{n+1,I} + g'_j(S_m^{n,j,I})(H_m^{n,j,R})^T \Delta w_{M+1}^{n+1,R} \right)
= (p_{M+1}^{n,R})^T \Delta w_{M+1}^{n+1,R} + (p_{M+1}^{n,J})^T \Delta w_{M+1}^{n+1,I}
\leq (\tau - \eta)(\|p_{M+1}^{n,R}\|^2 + \|p_{M+1}^{n,J}\|^2).
\]
(33)

Next, we prove (25). By (2), (4), (20) and Taylor’s formula, for any \(1 \leq j \leq J, 1 \leq m \leq M\) and \(n = 0, 1, 2, \ldots\), we have
\[
H_m^{n+1,j,R} - H_m^{n,j,R} = f(U_m^{n+1,j,R}) - f(U_m^{n,j,R})
= f'(U_m^{n,j,R})(U_m^{n+1,j,R} - U_m^{n,j,R}) + \frac{1}{2} f''(U_m^{n,j,R})(U_m^{n+1,j,R} - U_m^{n,j,R})^2
\]
(34)

and
\[
H_m^{n+1,j,I} - H_m^{n,j,I} = f(U_m^{n+1,j,I}) - f(U_m^{n,j,I})
= f'(U_m^{n,j,I})(U_m^{n+1,j,I} - U_m^{n,j,I}) + \frac{1}{2} f''(U_m^{n,j,I})(U_m^{n+1,j,I} - U_m^{n,j,I})^2,
\]
(35)

where \(U_m^{n,j,R}\) is an intermediate point on the line segment between the two points \(U_m^{n+1,j,R}\) and \(U_m^{n,j,R}\), and \(U_m^{n,j,I}\) between the two points \(U_m^{n+1,j,I}\) and \(U_m^{n,j,I}\). Thus, according to (2), (12), (13), (14), (20), (32) and (34)-(35) we have
\[
\sum_{j=1}^{J} \left( g'_j(S_m^{n,j,R}) \left( (\psi_m^{n,j,R})^T w_{M+1}^{n,R} - (\psi_m^{n,j,I})^T w_{M+1}^{n,I} \right) \right)
\]
\[ \begin{align*}
&+ g_j^I(S^{n,j}_R) \left( (\psi^{n,j}_R)^T w_{M+1}^{n,j} + (\psi^{n,j}_R)^T w_{M+1}^{n,R} \right) \\
&= \sum_{j=1}^{J} \sum_{m=1}^{M} \left( g_j^R(S^{n,j}_R)w_{M+1,m}^{n,R} f'(u_{M+1,m}^{n,j}_R) \left( (x_j^i)^T \Delta w_{m+1,R}^{n+1} + (y_j^i)^T \Delta w_{m+1,I}^{n+1} \right) \\
&\quad - g_j^R(S^{n,j}_R)w_{M+1,m}^{n,I} f'(u_{M+1,m}^{n,j}_R) \left( (x_j^i)^T \Delta w_{m+1,R}^{n+1} + (y_j^i)^T \Delta w_{m+1,I}^{n+1} \right) \right) \\
&+ \left( \sum_{j=1}^{J} g_j^I(S^{n,j}_R) \left( w_{M+1,m}^{n,I} f'(u_{M+1,m}^{n,j}_R) x_j^{i} - w_{M+1,m}^{n,R} f'(u_{M+1,m}^{n,j}_R) y_j^{i} \right) \right) \Delta w_{m+1,I}^{n+1} + \delta_1 \\
&= \sum_{m=1}^{M} \left( \left( \sum_{j=1}^{J} \left( g_j^R(S^{n,j}_R) \left( w_{M+1,m}^{n,R} f'(u_{M+1,m}^{n,j}_R) x_j^{i} - w_{M+1,m}^{n,R} f'(u_{M+1,m}^{n,j}_R) y_j^{i} \right) \right) \right) \Delta w_{m+1,I}^{n+1} \right) + \delta_1 \\
&\leq (\tau - \eta) \sum_{m=1}^{M} \left( \|p_m^{n,I}\|^2 + \|p_m^{n,R}\|^2 \right) + \delta_1
\end{align*} \]

where

\[ \delta_1 = \frac{1}{2} \sum_{j=1}^{J} \sum_{m=1}^{M} \left( g_j^R(S^{n,j}_R)w_{M+1,m}^{n,R} f'(u_{M+1,m}^{n,j}_R) \left( (x_j^i)^T \Delta w_{m+1,R}^{n+1} + (y_j^i)^T \Delta w_{m+1,I}^{n+1} \right) \right)^2 \]

Using Assumptions (A1) and (A2), (17), (18), (22) and the triangular inequality we immediately get

\[ \delta_1 \leq |\delta_1| \leq C_5 \sum_{m=1}^{M} \left( \|\Delta w_{m+1,R}^{n+1}\|^2 + \|\Delta w_{m+1,I}^{n+1}\|^2 \right) \]

\[ \leq C_5 \sum_{m=1}^{M} \left( (\tau_m^{n,R}\|\Delta w_{m+1,R}^{n+1}\| + \eta\|p_m^{n,R}\|)^2 + (\tau_m^{n,I}\|\Delta w_{m+1,I}^{n+1}\| + \eta\|p_m^{n,I}\|)^2 \right) \]

\[ \leq C_5 (\tau + \eta)^2 \sum_{m=1}^{M} \left( \|p_m^{n,R}\|^2 + \|p_m^{n,I}\|^2 \right), \]

where \( C_5 = 2JC_1C_2C_3 \max_{1 \leq j \leq J} \{|x_j|^2 + |y_j|^2\} \). Now, (25) results from (36) and (38).

According to (20), (22), (23) and (28) we have

\[ \sum_{j=1}^{J} \left( g_j^R(S^{n,j}_R) \left( (\psi^{n,j}_R)^T \Delta w_{M+1}^{n+1,R} - (\psi^{n,j}_R)^T \Delta w_{M+1}^{n+1,I} \right) \right) \]
\begin{align*}
&\quad + g_j l(S^{n,j,I}) \left( (\psi^{n,j,I})^T \Delta w^{n+1,I}_{M+1} + (\psi^{n,j,I})^T \Delta w^{n+1,R}_{M+1} \right) \\
&\leq C_3 \sum_{j=1}^J \left( \|\Delta w^{n+1,R}_{M+1}\|^2 + \|\Delta w^{n+1,I}_{M+1}\|^2 + \|\psi^{n,j,I}\|^2 + \|\psi^{n,j,R}\|^2 \right) \\
&\leq C_3 \sum_{j=1}^J \left( (\tau + \eta)^2 (\|p^{n,R}_{M+1}\|^2 + \|p^{n,I}_{M+1}\|^2) + 2C_4 (\tau + \eta)^2 \sum_{m=1}^M (\|p^{n,R}_{m}\|^2 + \|p^{n,I}_{m}\|^2) \right) \\
&\leq C_6 (\tau + \eta)^2 \sum_{m=1}^{M+1} (\|p^{n,R}_{m}\|^2 + \|p^{n,I}_{m}\|^2) \\
\end{align*}

(39)

and

\begin{align*}
&\frac{1}{2} \sum_{j=1}^J \left( g_{jR}(t^n_{1,j}) (S^{n+1,j,R} - S^{n,j,R})^2 + g_{jI}(t^n_{2,j}) (S^{n+1,j,I} - S^{n,j,I})^2 \right) \\
&\leq \frac{C_3}{2} \sum_{j=1}^J \left( (S^{n+1,j,R} - S^{n,j,R})^2 + (S^{n+1,j,I} - S^{n,j,I})^2 \right) \\
&\leq \frac{C_3}{2} \sum_{j=1}^J \left( \left( (H^{n+1,j,R})^T \Delta w^{n+1,R}_{M+1} - (H^{n+1,j,I})^T \Delta w^{n+1,I}_{M+1} + (\psi^{n,j,R})^T w^{n,R}_{M+1} - (\psi^{n,j,I})^T w^{n,I}_{M+1} \right)^2 \\
&\quad + \left( (H^{n+1,j,R})^T \Delta w^{n+1,I}_{M+1} + (H^{n+1,j,I})^T \Delta w^{n+1,R}_{M+1} + (\psi^{n,j,R})^T w^{n,I}_{M+1} + (\psi^{n,j,I})^T w^{n,R}_{M+1} \right)^2 \right) \\
&\leq 2C_3J \max \{(C_0)^2 + (C_2)^2\} (\|\Delta w^{n+1,R}_{M+1}\|^2 + \|\Delta w^{n+1,I}_{M+1}\|^2 + \|\psi^{n,j,R}\|^2 + \|\psi^{n,j,I}\|^2) \\
&\leq C_7 (\tau + \eta)^2 \sum_{m=1}^{M+1} (\|p^{n,R}_{m}\|^2 + \|p^{n,I}_{m}\|^2), \\
\end{align*}

(40)

where $C_6 = JC_3 \max\{1, 2C_4\}$ and $C_7 = 2JC_3 \max\{(C_0)^2 + (C_2)^2\} \max\{1, 2C_4\}$. Finally we obtain (26) and (27).

Now, we are ready to prove Theorem 1 in terms of the above two lemmas.

**Proof of Theorem 1.** First we prove (i). By (24)-(27) and Taylor’s formula we have

\begin{align*}
E(W^{n+1}) - E(W^n) \\
= \sum_{j=1}^J (g_{jR}(S^{n+1,j,R}) - g_{jR}(S^{n,j,R}) + g_{jI}(S^{n+1,j,I}) - g_{jI}(S^{n,j,I})) \\
= \sum_{j=1}^J (g_{jR}(S^{n,j,I})(S^{n+1,j,R} - S^{n,j,R}) + g_{jI}(S^{n,j,I})(S^{n+1,j,I} - S^{n,j,I}) \\
\quad + \frac{1}{2} g_{jR}(t^n_{1,j})(S^{n+1,j,R} - S^{n,j,R})^2 + \frac{1}{2} g_{jI}(t^n_{2,j})(S^{n+1,j,I} - S^{n,j,I})^2) \\
= \sum_{j=1}^J \left( g_{jR}(S^{n+1,j,R}) \left( (H^{n,j,R})^T \Delta w^{n+1,R}_{M+1} - (H^{n,j,I})^T \Delta w^{n+1,I}_{M+1} \right) \\
\quad + g_{jI}(S^{n,j,I}) \left( (H^{n,j,R})^T \Delta w^{n+1,I}_{M+1} + (H^{n,j,I})^T \Delta w^{n+1,R}_{M+1} \right) \\
\quad + g_{jR}(S^{n,j,R}) \left( (\psi^{n,j,R})^T w^{n,R}_{M+1} - (\psi^{n,j,I})^T w^{n,I}_{M+1} \right) \\
\quad + g_{jI}(S^{n,j,I}) \left( (\psi^{n,j,R})^T w^{n,I}_{M+1} + (\psi^{n,j,I})^T w^{n,R}_{M+1} \right) \right)
\end{align*}
Hence, there holds
\[ n \]

Since \( E \)

Then there holds
\[ \sum_{m=1}^{M} (\|p^{n,R}_{m}\|^2 + \|p^{n,I}_{m}\|^2) \]

where \( t^{n,j}_1 \in \mathbb{R} \) is on the segment between \( S^{n+1,j,R} \) and \( S^{n,j,R} \), and \( t^{n,j}_2 \in \mathbb{R} \) is on the segment between \( S^{n+1,j,I} \) and \( S^{n,j,I} \).

Finally, we prove (iv). We use (14), (15), (28), (29), and (45) to obtain
\[ \lim_{n \to \infty} \|w^{n+1,R}_m - w^n_R\| = 0, \lim_{n \to \infty} \|w^{n+1,I}_m - w^n_I\| = 0, m = 1, \ldots, M + 1. \]
Write $\mathbf{v} = ((\mathbf{w}_1^R)^T, \cdots, (\mathbf{w}_M^R)^T, (\mathbf{w}_1^I)^T, \cdots, (\mathbf{w}_M^I)^T)^T$, then $E(\mathbf{W})$ can be viewed as a function of $\mathbf{v}$, and denoted as $\mathcal{E}(\mathbf{v})$:

$$E(\mathbf{W}) \equiv \mathcal{E}(\mathbf{v}).$$

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

$$\frac{\partial \mathcal{E}(\mathbf{v})}{\partial w_{Rm}} \equiv \frac{\partial E(\mathbf{W})}{\partial w_{Rm}}, \quad \frac{\partial \mathcal{E}(\mathbf{v})}{\partial w_{Im}} \equiv \frac{\partial E(\mathbf{W})}{\partial w_{Im}}, \quad m = 1, \cdots, M + 1.$$  

Let $\mathbf{v}^n = ((\mathbf{w}_1^{n,R})^T, \cdots, (\mathbf{w}_M^{n,R})^T, (\mathbf{w}_1^{n,I})^T, \cdots, (\mathbf{w}_M^{n,I})^T)^T$, then by (45) we have

$$\lim_{n \to \infty} \left\| \frac{\partial \mathcal{E}(\mathbf{v}^n)}{\partial w_{Rm}} \right\| = \lim_{n \to \infty} \left\| \frac{\partial \mathcal{E}(\mathbf{v}^n)}{\partial w_{Im}} \right\| = 0, \quad m = 1, \cdots, M + 1. \quad (47)$$

From Assumption (A3), (46), (47) and Lemma 1 we know that there is a $\mathbf{v}^\star$ satisfying $\lim_{n \to \infty} \mathbf{v}^n = \mathbf{v}^\star$. By considering the relationship between $\mathbf{v}^n$ and $\mathbf{W}^n$, we immediately get the desired result. We thus complete the proof.

\[ \Box \]

References


