A New RNN Model with a Modified Nonlinear Activation Function Applied to Complex-Valued Linear Equations

Lei Ding, Lin Xiao, Kaiqing Zhou, Yonghong Lan, and Yongsheng Zhang

Abstract—In this paper, an improved Zhang neural network (IZNN) is proposed by using a kind of novel nonlinear activation function to solve the complex-valued systems of linear equation (CVSLE). Compared with the previous ZNN models, the convergence rate of the IZNN model has been accelerated. To do so, a kind of novel nonlinear activation function is first proposed to establish the novel recurrent neural network. Then, the corresponding maximum convergence time is given according to the randomly generated initial error vector, and the theoretical results illustrate that the new recurrent neural network using the proposed activation function has higher convergence rate than the previous neural networks using the the linear activation function or the tunable activation function.

Index Terms—Recurrent neural network, Convergence rate, Finite time, Complex-valued systems of linear equation, Novel nonlinear activation function.

I. INTRODUCTION

In recent years, the complex-valued systems of linear equation (CVSLE) problem has been used into many areas [1]–[5]. In mathematics, we can write the CVSLE problem as the following formula:

\[ SW(t) = G \in \mathbb{C}^n, \]  

(1)

where \( S \in \mathbb{C}^{n \times n} \) and \( G \in \mathbb{C}^n \) represent the coefficient matrix and vector in complex-valued domain, and \( W(t) \in \mathbb{C}^n \) represents an unknown vector in complex-valued domain. According to the complex formula, the vectors of equation (1) can be rewritten as \( S = S_{re} + jS_{im} \), \( G = G_{re} + jG_{im} \), and \( W(t) = W_{re}(t) + jW_{im}(t) \), where \( j = \sqrt{-1} \) represents an imaginary unit. So, we can further describe the equation (1) as follows:

\[ [S_{re} + jS_{im}][W_{re}(t) + jW_{im}(t)] = G_{re} + jG_{im} \in \mathbb{C}^n, \]  

(2)

where \( S_{re} \in \mathbb{R}^{n \times n}, S_{im} \in \mathbb{R}^{n \times n}, W_{re} \in \mathbb{R}^n, W_{im} \in \mathbb{R}^n, G_{re} \in \mathbb{R}^n, \) and \( G_{im} \in \mathbb{R}^n \). From the complex formula’s principle, the two sides’ real (or imaginary) part of the equation must be equal. Then we can rewrite the equation (2) as

\[
\begin{align*}
S_{re}W_{re}(t) - S_{im}W_{im}(t) &= G_{re} \in \mathbb{R}^n, \\
S_{im}W_{re}(t) + S_{re}W_{im}(t) &= G_{im} \in \mathbb{R}^n.
\end{align*}
\]  

(3)

Now the equation (3) can be described in the following compact matrix form:

\[
\begin{bmatrix}
S_{re} & -S_{im} \\
S_{im} & S_{re}
\end{bmatrix}
\begin{bmatrix}
W_{re}(t) \\
W_{im}(t)
\end{bmatrix}
= \begin{bmatrix}
G_{re} \\
G_{im}
\end{bmatrix} \in \mathbb{R}^{2n}.  
\]  

(4)

Then the equation (4) can be rewritten as the following form:

\[
DH(t) = Q \in \mathbb{R}^{2n},
\]

(5)

where \( D = \begin{bmatrix} S_{re} & -S_{im} \\
S_{im} & S_{re}\end{bmatrix} \), \( H(t) = \begin{bmatrix} W_{re}(t) \\
W_{im}(t)\end{bmatrix} \), and \( Q = \begin{bmatrix} G_{re} \\
G_{im}\end{bmatrix} \).

Now the CVSLE can be dealt with in real domain, and we can use the technique for dealing with the real-valued system of linear equation problem to deal with the CVSLE [6]–[10].

Now the recurrent neural networks have become a research hotspot [11]–[15]. As a kind of recurrent neural network, Zhang neural network (ZNN) aroused widespread concern in recent years [16]–[19]. Compared with the neural network called gradient neural network (GNN) by using the Frobenius norm as its performance indicator, the ZNN using the lagging error can exponentially converge to zero instead of converging to zero after long time [20]–[22]. It is noted that the original ZNN model using the linear activation function can’t converge to zero within finite time [23]. So, to improve the convergence rate, some nonlinear activation functions are designed for ZNN model. For example, Li et al. proposed a sign-bi-power activation function (SBPAF) to modify the performance of ZNN model [24], [25]. Based on the SBPAF, Miao et al. proposed a new nonlinear activation function called tunable activation function to further accelerate the convergence rate [26]. The tunable activation function is formulated as follows:

\[
\Psi(v) = \text{sign}(v)(a_1|v|^r + a_2|v| + a_3|v|^k),
\]

(6)
where $0 < r < 1$, $a_1 > 0$, $a_2 > 0$, $a_3 > 0$ and
\[
\text{sign}(v) = \begin{cases} 
1, & \text{if } v > 0 \\
0, & \text{if } v = 0 \\
-1, & \text{if } v < 0.
\end{cases}
\]

In [26], the study shows that the nonlinear activation function for ZNN can accelerate the convergent rate and converge to zero within finite time.

The above study shows that the appropriate nonlinear activation functions can accelerate the convergent rate. To further accelerate the convergent rate, we propose an improved nonlinear activation function for ZNN model based on the tunable activation function. So an improved ZNN model using the improved nonlinear activation function is proposed to solve the CVSLE in this paper.

The remainder of this paper is divided into the following four parts. An improved ZNN (IZNN) model using a new nonlinear activation function to deal with the CVSLE is proposed, and the corresponding theoretical proof is given in section II. The corresponding the simulation results is given to show the superiority of this new nonlinear activation function in section III. Section IV gives the corresponding conclusions.

II. Finite Time Convergent IZNN

From the above analysis, we can calculate the CVSLE problem in real domain. For the original ZNN model, we can describe the error function $Y(t)$ as
\[
Y(t) = DH(t) - Q \in \mathbb{R}^{2n}.
\]
Then we have $\dot{Y}(t) = -\beta \Psi(Y(t))$, which denotes the design formula of ZNN. Based on this, we have
\[
\dot{H}(t) = -\beta \Psi(DH(t) - Q),
\]
which is called the ZNNL model. If the tunable activation function is used, we have
\[
\dot{Y}(t) = -\beta \text{sign}(Y(t))(a_1|Y(t)|^p + a_2|Y(t)|\dot{Y}(t) - a_3|Y(t)|),
\]
which is called the ZNNT model. Now we propose a new nonlinear activation function, which is defined as
\[
\Psi(v) = \text{sign}(v)(a_1|v|^p + a_2|v|^\frac{p}{r} - a_3|v|),
\]
where $a_1 > a_3 > 0$, $a_2 > a_3 > 0$, and $p > 1$. Then we have
\[
\dot{Y}(t) = -\beta \text{sign}(Y(t))(a_1|Y(t)|^p + a_2|Y(t)|\dot{Y}(t)) - a_3|Y(t)|.
\]

We can rewrite the formula (12) as
\[
\dot{H}(t) = -\beta \text{sign}(DH(t) - Q)(a_1|DH(t) - Q|^p + a_2|DH(t) - Q|^\frac{p}{r} - a_3|DH(t) - Q|),
\]
which is called the IZNN model for solving the CVSLE. Now to verify the IZNN model’s finite-time convergence property, the corresponding theorems are presented as follows.

Theorem 1: Regardless of what the value of the initial residual error is, the residual error $Y(t)$ of equation(12) will get to zero within $t(x_0)$ satisfies:
\[
t(x_0) = \begin{cases}
\frac{\ln \frac{2^p}{a_1|Y_0|^{p+1} - 2^p}}{\beta a_3(p-1)}, & \text{if } |Y_M(0)| \geq 1 \\
\frac{-\ln \frac{2^p}{a_1|Y_0|^{p+1} - 2^p}}{\beta a_3(p-1)}, & \text{if } |Y_M(0)| < 1
\end{cases}
\]

where $|Y_M(0)|$ represents the maximum element of the initial residual error function vector $|Y(t)|$.

Proof: According to (12), the vector $Y(t)$’s each element has the identical dynamic, then the equation (12) can be written as
\[
\dot{Y}_i(t) = -\beta \text{sign}(Y_i(t))(a_1|Y_i(t)|^p + a_2|Y_i(t)|\dot{Y}(t)) - a_3|Y_i(t)|, 
\]
where $Y_i(t)$ represents the vector $Y(t)$’s ith element. From equation (14) and based on the conditions of $a_1 > a_3 > 0$ and $a_2 > a_3 > 0$, we have
\[
(a_1|Y_i(t)|^p + a_2|Y_i(t)|\dot{Y}_i(t)) \geq 2\sqrt{a_1|Y_i(t)|^p \ast a_2|Y_i(t)|\dot{Y}_i(t)} \\
\geq 2\sqrt{a_1 \ast a_2|Y_i(t)|^2} \\
> 2\sqrt{a_3 \ast a_3|Y_i(t)|^2} \\
= 2a_3|Y_i(t)|.
\]

Then, we have $a_1|Y_i(t)|^p + a_2|Y_i(t)|\dot{Y}_i(t) - a_3|Y_i(t)| > 0$. If $Y_i(t) > 0$, we have
\[
\dot{Y}_i(t) = -\beta \text{sign}(Y_i(t))(a_1|Y_i(t)|^p + a_2|Y_i(t)|\dot{Y}_i(t)) - a_3|Y_i(t)|.
\]
Obviously, the equation (15) is monotone decreasing, and will finally converge to zero. If $Y_i(t) < 0$, we have
\[
\dot{Y}_i(t) = -\beta \text{sign}(Y_i(t))(a_1|Y_i(t)|^p + a_2|Y_i(t)|\dot{Y}_i(t)) - a_3|Y_i(t)|.
\]

Now suppose $Y_i(t) > 0$, and $Y_i(t) \geq 1$, then according to the equation (15), we have
\[
\dot{Y}_i(t) \leq -a_1|Y_i(t)|^p + a_2|Y_i(t)|\dot{Y}_i(t) - a_3|Y_i(t)|.
\]
From the equation (17), we can find $-a_1|Y_i(t)|^p + a_2|Y_i(t)|\dot{Y}_i(t) - a_3|Y_i(t)| < 0$, and the equation (17) is convergent. Now multiplying (17) by $e^{-a_3t}$, then we can rewrite the equation (17) as
\[
e^{-a_3t}Y_i(t) - a_3|Y_i(t)| e^{-a_3t} \leq -a_1|Y_i(t)|^p e^{-a_3t}. 
\]
Then we have
\[
\frac{d(Y_i(t)e^{-a_3t})}{Y_i(t)e^{-a_3t}} \leq -a_1 e^{-a_3(1-p)t} dt.
\]
Now integrate the inequality from 0 to \( t \),
\[
Y(t) \leq e^{\beta a_3 t}[Y(0)]^{1-p} - \frac{a_1}{a_3} + \frac{a_1}{a_3} e^{-\beta a_3 (p-1)t} Y(0)^{1-p},
\]
and let
\[
t_1 = \frac{\ln \frac{a_1}{a_3} Y(0)^{1-p}}{\beta a_3 (p-1)}.
\]
Then we can find that if \( t_1 \geq t_1 \), \( Y(t) \leq 1 \). Suppose \( Y(t) \)'s largest element, then we have \( t_1 = \max(t_1) \), and \( t_1 = \frac{\ln \frac{a_1}{a_3} Y(0)^{1-p}}{\beta a_3 (p-1)} \). Then we can conclude that if \( t_1 \geq t_1, Y(t) \leq 1 \).

When \( 0 < Y(t) < 1 \), according to the equation (15), we have
\[
Y(t) \leq -\beta a_2 (Y(t))^{1-p} + \beta a_3 Y(t).
\]
From the equation (22), we can find when \( 0 < Y(t) < 1 \), \( a_2 > a_3 \), and \( p > 1 \), we will have \((Y(t))^{1-p} > Y(t) \), and \(-a_2 (Y(t))^{1-p} + a_3 Y(t) < 0 \). So the equation (22) is convergent.

Now multiplying (22) by \( e^{-\beta a_3 t} \), then we can rewrite the equation (22) as
\[
e^{-\beta a_3 t} Y(t) - \beta a_3 Y(t) e^{-\beta a_3 t} \leq -\beta a_2 (Y(t))^{1-p} e^{-\beta a_3 t}.
\]
Then
\[
d(Y(t) e^{-\beta a_3 t}) \leq \beta a_2 e^{-\beta a_3 (1-\frac{1}{p})t}.
\]
Then, we have
\[
d(Y(t) e^{-\beta a_3 t}) \leq \beta a_2 e^{-\beta a_3 (1-\frac{1}{p})t}.
\]
Now integrate the inequality from 0 to \( t \),
\[
Y(t) \leq e^{\beta a_3 t}[Y(0)]^{1-p} - \frac{a_2}{a_3} + \frac{a_2}{a_3} e^{-\beta a_3 (p-1)t} Y(0)^{1-p},
\]
and let
\[
t_2 = \frac{-p \ln(1 - \frac{a_2}{a_3} Y(0)^{1-p})}{\beta a_3 (p-1)}.
\]

Fig. 1. Output trajectories of neural states \( X(t) \) synthesized by ZNNL model (9) in example 1

Fig. 2. Example 1: Output trajectories of neural states \( H(t) \) synthesized by ZNNT model (10) in example 1
Fig. 3. Output trajectories of neural states $H(t)$ synthesized by IZNN model (13) in example 1

Then we can find that if $t \geq t_2$, $Y_i(t) = 0$. Suppose $t_2 = \max(t_2i)$ and we have $t_2 = \frac{-p \ln(1 - \frac{\|Y(0)\|}{3})}{\beta \alpha_3(p-1)}$. Now we can find $Y_i(t) = 0$ when $t \geq t_2$.

Similarly, when $Y_i(t) < 0$, we will have the same conclusion. Now we can rewrite $t_1$ and $t_2$ as $t_1 = \frac{\ln[1 - \frac{\|Y(0)\|}{3}]}{\beta \alpha_3(p-1)}$, and $t_2 = \frac{-p \ln(1 - \frac{|Y_M(0)|}{3})}{\beta \alpha_3(p-1)}$, respectively, where $|Y_M(0)|$ is the residual error function $|Y(t)|$'s maximum element.

This proof is successful.

**Theorem 2:** Regardless of what the randomly generated initial state matrix $H(0)$ is, the model (13)'s state matrix $H(t)$ will obtain its theoretical result within $t_u$, and

$$t_u = \left\{ \begin{array}{ll} \frac{\ln[1 - \frac{\|Y(0)\|}{3}]}{\beta \alpha_3(p-1)} + \frac{-p \ln(1 - \frac{|Y_M(0)|}{3})}{\beta \alpha_3(p-1)} & \text{if } |Y_M(0)| \geq 1 \\ \frac{-p \ln(1 - \frac{|Y_M(0)|}{3})}{\beta \alpha_3(p-1)} & \text{if } |Y_M(0)| < 1 \end{array} \right.$$  

where $|Y_M(0)|$ represents the initial residual error function matrix $|Y(t)|$'s maximum element.

**Proof:** Suppose $H_{(AZ)}(t)$ and $H_{(or)}(t)$ mean the solution and the theoretical solution of the model (13), respectively. Then $\tilde{H}(t)$ means the difference between $H_{(AZ)}(t)$ and $H_{(or)}(t)$,

$$\tilde{H}(t) = H_{(AZ)}(t) - H_{(or)}(t) \in R^{2n}. \quad (28)$$

We can rewrite the above equation as

$$H_{(AZ)}(t) = \tilde{H}(t) + H_{(or)}(t) \in R^{2n}. \quad (29)$$
According to the equation (7) and the equation (8), we have

\[ Y(t) = D(\hat{H}(t) + H_{(or)}(t)) - Q. \]

Because \( D(H(t)) \) is given, we can rewrite the above equation as

\[ \dot{Y}(t) = -\beta \text{sign}(Y(t))(a_1|Y(t)|^p + a_2|Y(t)|^q - a_3|Y(t)|). \]

The above equation is same as the equation (15). Thus the proof is successful.

III. THE COMPUTER SIMULATION

Now, we will use two digital examples to show the superiority of IZNN model (13) compared with the ZNNL model and the ZNNT model. To display the convergent rate of different models, the corresponding neural-state solutions’ output trajectories and the evolution procedure of the corresponding residual error norm \(||Y(t)||_2\) are given in this paper. Furthermore, to facilitate the comparison, we choose the same parameters \( \beta = 10, r = \frac{1}{2}, a_1 = k_1 = 0.5, a_2 = k_1 = 0.8, \) and \( a_3 = k_3 = 0.4 \) for the different models. In this section, two different examples are given.

Example 1:

\[ S_1 W_1(t) = G_1 \in \mathbb{C}^n. \]
Fig. 7. Output trajectories of neural states $H(t)$ synthesized by IZNN model (13) in example 2.

\[
\begin{bmatrix}
-2.5 \\
-1.5 \\
-0.5 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
-3.5764 \\
-1.5764 \\
0.5764 \\
0 \\
0 \\
0 \\
1
\end{bmatrix}
\]

here, $T$ means the transpose of the matrix $Q_1$.

Example 2:

\[
S_2W(t) = G_2 \in \mathbb{C}^n,
\]

where $S_2 = \frac{1}{2} (3.5345 + 4.5345j, 2.5342 + 7.4532j, 5.5344 + 6.7534j, 3.2312 - 4.7543j)$.

and

\[ Q_2^T = \begin{bmatrix} 5.3545 & 2.5432 & 5.5334 & 3.2312 & 4.5345 & 7.4532 & -6.7853 & -4.7543 \end{bmatrix}, \]

where T means the transpose of the matrix \( Q_2 \).

From the neural-state solutions’ output trajectories shown in the Figs. 1-3, and Figs. 5-7, we can find that compared with the ZNNL model and the ZNNT model, this IZNN model has the fastest convergence rate for solving the CVSE problem. Furthermore, from the evolution of the corresponding residual error norm \(|Y(t)|\) displayed in Fig. 4 and Fig. 8, this IZNN model has the fastest convergence rate than the ZNNL model and the ZNNT model.

IV. CONCLUSIONS

To accelerate the convergence rate for solving the CVSE in complex domain, an improved ZNN model with new activation function is presented and the corresponding theoretical proof is given in detail in this paper. It is the first time to present this novel nonlinear activation function to accelerate the convergence rate and even reach the finite-time convergence. The simulation results display that the IZNN model presented in this paper has the fastest convergence rate, as compared with the ZNNL model and the ZNNT model.

REFERENCES