THE OPTIMIZATION OF WEIGHT FUNCTIONS OF $R_{3,1}$ SPACE TIME BY THE ECHO STATE NETWORK

Sadataka Furui[1] and Serge Dos Santos[2]

[1]Faculty of Science and Engineering, Teikyo University, Utsunomiya, 320 Japan
[2]INSA Centre Val de Loire, Universite de Tours, INSERM, Imaging Brain & Neuropsychiatry iBraiN U1253, F-41034 Blois Cedex, France

ABSTRACT

We obtain the optimum weight function of phonetic soliton paths in Weyl fermion sea described by quaternions sitting on (3+1) D lattices. As in the previous work [5] using the Elman Recurrent Neural Network (ERNN) the action of solitonic phonons are derived from the fixed point action prepared for the quantum Chromodynamics (QCD) by DeGrand et al.[20]. We consider 7 paths in (3+1) D space with length less than or equal to 8 lattice units. Comparison of ERNN and the Echo State Network (ESN) are given. The ESN using tanh function for the nonlinear function give stable results as the simulation cycle number increases.

KEYWORDS

Time-Reversal Based Nonlinear Elastic Wave Spectroscopy, Clifford Algebra, Weyl fermion, Quaternion Field Theory, Biquaternion, Echo State Network, Elman Recurrent Neural Network

1. INTRODUCTION

In order to predict real-valued time series, the Recurrent Neural Network (RNN) is an efficient method [1,2].

Bianchi et al. [3,4] showed that the Elman RNN (ERNN) and the Echo state Network (ESN) are promising. In [5], we applied the ERNN to the solitonic phonon propagation in materials in (3+1) D lattice simulation.

The ERNN has the input x[t] ∈ R^Ni where Ni is the dimension of the input vector, and bias vector **b**_i∈ R^Nh where Nh is the number of nodes in the hidden layer.

The hidden state vector at time t becomes

$$\mathbf{h}[t] = f(W_{ih}(\mathbf{x}[t] + \mathbf{b}_i) + W_{hh}^{(t)}(\mathbf{h}[t-1] + \mathbf{b}_h)),$$

where Wih is the input weight matrix, Whh^(t) are Whh filtered by setting matrix elements equal to 0, when there is no channel mixing.

The output at time t is y[t] ∈ R^No is defined with a bias vector **b**o as

$$\mathbf{y}[t] = g(W_{ho}(\mathbf{h}[t] + \mathbf{b}_o)).$$

An ESN consists of a large, sparsely connected untrained recurrent layer of nonlinear units and a linear, memory-less read-out layer which is trained. The scheme is shown in Fig.1.

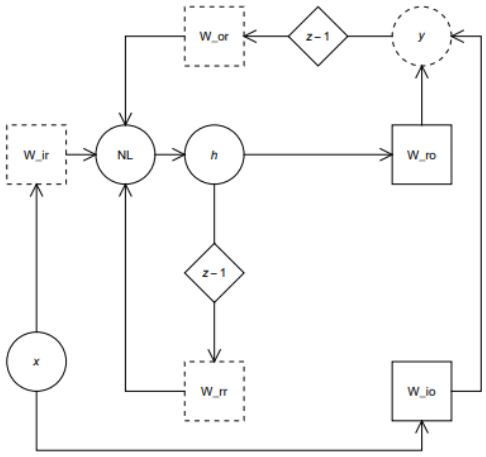

Figure 1. The standard scheme of ESN. The circles represent input **x**, state **h** and output **y**, respectively. Solid squares W_ro, W_io are the fixed matrices of the readout, while dashed squares W_rr, W_or and W_ir are randomly initialized and trainable matrices. NL means the nonlinear operation (logistic sigmoid or tanh function) and z-1 is the unit delay operator.

In our case W_ro is obtained from the that of ERNN analysis, in which the path L21' and the path L22' exists. We take the average of L21 and L21', the average of L22 and L22' in Wro and reduce the dimension of the base from 9 to 7. The weight matrix Wio is chosen to be Wir, W_ro, where Wir is chosen as a random 4 x 7 matrix, which is trained by the optimization.

The update of the vector in reservoir **h**[t] and **y**[t] are respectively

$$\begin{aligned} \mathbf{h}[t] &= f(W_{rr}\mathbf{h}[t-1] + W_{ir}\mathbf{x}[t] + W_{or}\mathbf{y}[t-1] + \epsilon) \\ \mathbf{y}[t] &= g(W_{io}\mathbf{x}[t] + W_{ro}\mathbf{h}[t]). \end{aligned}$$

where ε is a small noise term, *f* and *g* are logistic or tanh functions.

The imput signal **x**[t] produces the output **y**[t]. The vector **h**[t] describes the instantaneous state. The matrix of reservoir Wrr is required to satisfy the spectral radius ρ(Wrr) <1 by rescaling. From **x** [t], one can calculate **h**[t], and construct a matrix **S** ∈ R^[Ttr×(Ni+Nr)] .

The training sequence is taken as

$$(\mathbf{x}[1], \mathbf{y}^*[1]), \cdots, (\mathbf{x}[T_{tr}], \mathbf{y}^*[T_{tr}]),$$

where Ttr is the length of the training sequences, **y*** is the desired output.

The states are stacked in a matrix

$$\mathbf{S} = \begin{bmatrix} \mathbf{x}^T[1] & \mathbf{h}^T[1] \\ \vdots & \vdots \\ \mathbf{x}^T[T_{tr}] & \mathbf{h}^T[T_{tr}] \end{bmatrix}, \text{ and } \mathbf{y}^* = \begin{bmatrix} \mathbf{y}^*[1] \\ \vdots \\ \mathbf{y}^*[T_{tr}] \end{bmatrix}$$

is the desired output.

In order to get the desired output **y***,

$$W_{ls}^* = arg\,min_{W \in R^{N_i+N_h}} \frac{1}{2}||\mathbf{SW} - \mathbf{y}^*||^2 + \frac{\lambda_2}{2}||W||_2^2$$

is needed be solved. Where W=[Wio, Wro]^T, λ2∈ R^+ is the L2 regularization factor.
When W is fixed, a solution is expressed as

$$W_{ls}^* = (\mathbf{S}^T\mathbf{S} + \lambda_2\mathbf{I})^{-1}\mathbf{S}^T\mathbf{y}^*.$$

In our case, Wio depends on 4 x 7 matrix Wir, and we adopt another method for minimizing $||\mathbf{SW} - \mathbf{y}^*||^2$

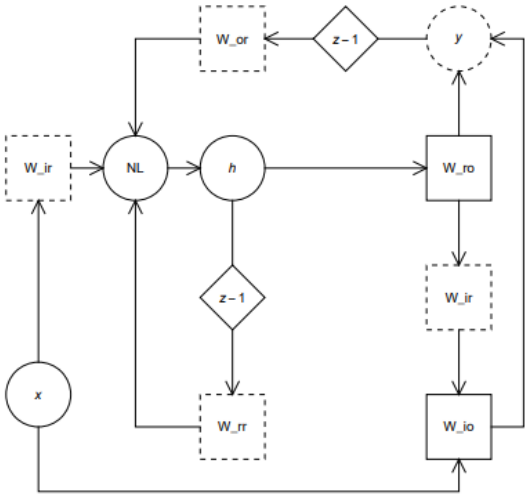

Figure 2: Our scheme of ESN. W_{io} is produced by a multiplication of W_{ir} on W_{ro}.

The structure of this work is as follows. In section.2 we present our model of ESN.

In section 3 we present numerical results of Monte Carlo (MC) simulation. Conclusion and outlook are given in section 4.

2. ESN FOR THE LATTICE MONTE CARLO SIMULATION

In [5] we analyzed the propagation of solitonic phonon in Weyl fermion sea which is described by biquaternions sitting on the (3+1) D lattice. The model was an extension of analyses of the (2+1) D lattice, in which description by quaternions obeying the Clifford algebra was possible [6,7].

Recently, application of quaternions in engineering and physics has been intensively discussed [8,9,10,11,12]. Since applications of Non Destructive Testing (NDT) using ultrasonic wave, the time reversal symmetry based nonlinear elastic wave spectroscopy (TR-NEWS) [13,14,15,16,17] are attracting attentions [18], it is important to establish the method of analysis.

In Clifford algebra, the mapping j : R3,1-> M2(**H**) proposed by Garling[19] is

$$j(\mathcal{A}^{+}_{3,1}) = \begin{pmatrix} a_1 + a_2\mathbf{k} & b_1\mathbf{i} + b_2\mathbf{j} \\ c_1\mathbf{i} + c_2\mathbf{j} & d_1 + d_2\mathbf{k} \end{pmatrix},$$

where a_i, b_i, c_i, d_i (i=1,2) are real.

In Quantum Chromo Dynamics (QCD) lattice simulation [20], 28 closed paths of the length less than or equal to 8 lattice units were considered. There are 3 types:

1) A type: path on a 2D plane spanned by **e**_1, **e**_2
2) B type: path in a 3D space spanned by **e**_1, **e**_2 and **e**_3=**e**_1∧**e**_2.
3) C type: path in a 4D space spanned by **e**_1, **e**_2, **e**_3 and **e**_0

There are 7 A-type, 13 B-type and 7 C-type paths neglecting the rectangular path rotating twice. In biquaternion bases, the C- type paths contain 2 paths that go and return along the time direction e0, which are taken into account by the Wir matrices.

In [6], we considered 7 paths in (3+1) D that contain hysteresis effects [21] following the ERNN method [4]. In this work, we adopt the ESN method to detect hysteresis effects.

The C-type paths of 16 steps are summarized in Table 1 and Table 2. In quaternion basis

$$q = q_0e_0 + q_1e_1 + q_2e_2 + q_3e_3 \text{ and } \bar{q} = q_0e_0 - q_1e_1 - q_2e_2 - q_3e_3,$$

we take the **x**=x **e**_1, **y**=y **e**_2, **z**=z **e**_3, where q_i, x, y, z ∈R.

Biquaternions are e_i e_j, i,j∈{0,1,2,3}, and when i,j ∈{1,2,3}, **e**_i **e**_j=ε^{ijk}**e**_k.

ε^{123}=-ε^{213}=1, ε^{312}=-ε^{132}=1 and ε^{231}=-ε^{321}=1.

In the Table 1 and 2, the biquaternion basis **e**i **e**j are denoted as ij, directions of the wave front along the path are x,y,z,t.

Backward propagations are -x, -y, -z, -t.

step	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
L19	x	y	z	t	-z	-t	-x	-y	-x	-y	-z	-t	z	t	x	y
	23	31	12	24	-12	-24	-23	-31	-23	-31	-12	-24	12	24	23	31
L20	x	y	z	t	-z	-y	-x	-t	-x	-y	-z	-t	z	y	x	t
	23	31	12	24	-12	-31	-23	-24	-23	-31	-12	-24	12	31	23	24
L25	x	y	z	t	-x	-y	-z	-t	-x	-y	-z	-t	x	y	z	t
	23	31	12	24	-23	-13	-12	-24	-23	-31	-12	-24	23	13	12	24

Table 1. Directions of the wave front of paths L19, L20, L25. The first line is in R^4 basis, the second line is in biquaternion basis.

step	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
L21	x	y	z	t	-z	-x	-t	-y	-x	-y	-z	-t	z	x	t	y
	23	31	12	14/24	-12	-23	-34	-13	-23	-31	-12	-14/24	12	23	34	13
L22	x	y	z	t	-z	-x	-y	-t	-x	-y	-z	-t	z	x	y	t
	23	31	12	14/24	-12	-23	-31	-34	-23	-31	-12	-14/24	12	23	31	34
L23	x	y	z	t	-y	-x	-t	-z	-x	-y	-z	-t	y	x	t	z
	23	31	12	14	-31	-23	-24	-12	-23	-31	-12	-14	31	23	24	12
L24	x	y	z	t	-y	-x	-z	-t	-x	-y	-z	-t	y	x	z	t
	23	31	12	14	-31	-23	-12	-24	-23	-31	-12	-14	31	23	12	24

Table 2. Directions of the wave front of paths $L21, L22, L23, L24$. The first line is in R^4 basis, the second line is in biquaternion basis.

The 8 steps of L19, …, 25 are shown in [22]. The 16 steps of these paths are shown in Figs.3,4,5,6. At balls, time shifts occur. We assume same hysteretic effects occur stochastically in the balls.

The path of L19, L21, L22 differ by the points where time shifts occur. We selected the time shift of L19 to be 24 and consider 14 and 24 for L21 and L22, Similarly the time shift of L20 was fixed to be 24.

In the present work, the time shifts are considered in W_{rr} matrices.

3. Optimization of the Weight Function by ESN

Since the paths of L19, L20, L21, L22, L23, L24, L25 on the (3+1) D lattice are fixed [20], **x**[t] for the seven paths can be fixed, when the 2D plane of the initial path is selected. Although **x**[t] runs in the 4D space, Wir **x**[t] is a 7dimensional vector defined on the path at time t.

The output **y**[t] in the ERNN is known for t=4,5,9,10,11,12, and we use the six values for **y***.

Figure 3: The path of $L19$(left) and that of $L20$(right). Balls are the places where hysteretic time shift occurs.

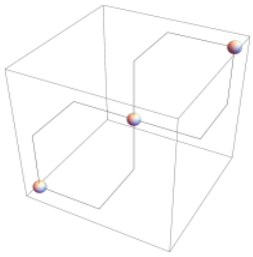

Figure 4: The path of $L25$.

Figure 5: The path of $L21$(left) and that of $L22$(right).

We prepare random matrices

1) Wro: 7 x 6 matrix, Matrix elements are fixed from ERNN inputs.
2) Wio: 4 x 6 matrix. Matrix elements are fixed from Wro and Wir.
3) Wir: 4 x 7 matrix. Matrix elements are subject to be trained.
4) Wor : 6 x 7 matrix. Matrix elements are subject to be trained.
5) Wrr^(t) 7 x 7 matrix. Matrix elements are subject to be trained. Mixing of paths at t=4,7,8,12,15,16 have specific filters.

By the criterion that at t=4,7,8,12,15,16, the paths which don't have the time shift are not disturbed, we multiply a filter to Wrr such that non disturbed matrix elements are set to 0. The interactions between reservoirs at these times are given in the following form.

$$\mathbf{W}_{rr}^{(4)} = \begin{pmatrix} * & * & * & * & 0 & 0 & * \\ * & * & * & * & 0 & 0 & * \\ * & * & * & * & 0 & 0 & * \\ * & * & * & * & 0 & 0 & * \\ 0 & 0 & * & * & * & * & 0 \\ 0 & 0 & * & * & * & * & * \\ * & * & * & * & 0 & * & * \end{pmatrix} \quad \mathbf{W}_{rr}^{(7)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{W}_{rr}^{(8)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & * \\ 0 & * & 0 & * & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * \\ 0 & * & 0 & * & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & * & * \\ * & * & * & * & 0 & * & * \end{pmatrix} \quad \mathbf{W}_{rr}^{(12)} = \begin{pmatrix} * & * & * & * & 0 & 0 & * \\ * & * & * & * & 0 & 0 & * \\ * & * & * & * & * & * & * \\ * & * & * & * & * & * & * \\ 0 & 0 & * & * & * & * & 0 \\ 0 & 0 & * & * & * & * & * \\ * & * & * & * & 0 & * & * \end{pmatrix}$$

$$\mathbf{W}_{rr}^{(15)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \mathbf{W}_{rr}^{(16)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & * & 0 & * & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & * & 0 & * & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 & * & * \\ 0 & * & 0 & 0 & 0 & * & * \end{pmatrix}$$

The low and the column are in the order L19, L20, L21, L22, L23, L24, L25 and * indicate time shifts in the own path occur or time shifts that cause mixing of paths occur.

We prepare 40 random vectors for t=4: **h**[4]=f(Wir **x**[4])=(h1[4], h2[4],…,h7[4]) corresponding to L19,L20,…, L25. Here *f* is the sigmoid function or tanh function.

The **x**[4]=(x1[4],x2[4],x3[4],x4[4]), where Wir(x1[4],x2[4],x3[4],x4[4]) is the seven dimentional vector given by the fixed point action of DeGrand et al.[20]. The matrix Wio is normalized such that

$$\sum_{i=1}^{4}\sum_{j=1}^{6}(W_{io})_{i,j}^{2} = \sum_{r=1}^{7}\sum_{j=1}^{6}(W_{ro})_{r,j}^{2}.$$

We define the loss function *L*=Sro^2, where Sro=Wro **h**-**y***. In the calculation of **h**[t], we choose

$$\mathbf{h}[t] = f(W_{io}\mathbf{x}[t] + W_{ro}(\mathbf{h}[t] + \mathbf{b}_h))$$

$$\mathbf{b}_h = -2\eta \frac{\partial \mathcal{L}}{\partial S_{ro}} \frac{\partial S_{ro}}{\partial h} = -2\eta(W_{ro}\mathbf{h} - \mathbf{y}^*)W_{ro}.$$

The weight matrix of reservoirs Wrr^(t) can also be trained to yield proper outputs. We assume that the matrix element of Wrr^(t), denoted as Wm,n^(t) is not 0, when the direction of the path m: **e**_i **e**_4 and that of n : **e**_j e_4 coincide. At t=4,7,8,12,15 and 16, there are time shift points between different paths. Time shifts occur with biquaternion base 14,24 or 34, and when the shift of different paths have the same biquaternion base, we assume that mixings of paths occur and the matrix element differs from 0.

Optimiztion steps for Wrr are done by modifying the Wrr^(t) of t=4,7,8,12,15,16 as follows. The sigmoid function

$$\frac{1}{1+exp(-W_{rr}^{(t)}\mathbf{h}[t]-W_{ir}\mathbf{x}[t+1]-W_{or}\mathbf{y}[t])}$$

has the range (0,1] and its derivative with respect to Wrr^(t) as

$$\frac{1 - \frac{1}{1+exp(-W_{rr}^{(t)}\mathbf{h}[t]-W_{ir}\mathbf{x}[t+1]-W_{or}\mathbf{y}[t])}}{1 + exp(-W_{rr}^{(t)}\mathbf{h}[t] - W_{ir}\mathbf{x}[t+1] - W_{or}\mathbf{y}[t])}\mathbf{h}[t] = DW_{rr}^{(t)}\mathbf{h}[t].$$

The tanh function tanh(-Wrr^(t)**h**[t]-Wir **x**[t+1]-Wor **y**[t]) has the range [-1,1] and the derivative with respect to Wrr^(t) is

$$-\text{sech}^2(-W_{rr}^{(t)}\mathbf{h}[t] - W_{ir}\mathbf{x}[t+1] - W_{or}\mathbf{y}[t])\mathbf{h}[t] = DW_{rr}^{(t)}\mathbf{h}[t]$$

Wrr^(t) is modified to

$$W_{rr}^{(t)} - \eta \frac{\partial \mathcal{L}}{\partial W_{rr}^{(t)}} = W_{rr}^{(t)} - \eta DW_{rr}^{(t)}.$$

The optimization of Wrr and the state vector **h**[t] are as follows.

- Using the $\mathbf{h}[4] = f(W_{ir}\mathbf{x}[4])$, where $\mathbf{x}[4]$ is produced randomly, we calculate $\mathbf{y}[4] = g(W_{io}\mathbf{x}[4] + W_{ro}\mathbf{h}[4]) = (y_1[4], \cdots, y_6[4])$, where g and f are the logistic sigmoid function or tanh function.

  From the second cycle, calculate $\mathbf{h}[4] = f(W_{rr}^{(4)}\mathbf{h}[3] + W_{ir}\mathbf{x}[4] + W_{or}\mathbf{y}[3])$ and $\mathbf{y}[4] = g(W_{io}\mathbf{x}[4] + W_{ro}\mathbf{h}[4])$. Modify $W_{rr}^{(4)} \rightarrow W_{rr}^{(4)} - \eta DW_{rr}^{(4)}$.

- Calculate $\mathbf{h}[5] = f(W_{rr}^{(5)}\mathbf{h}[4] + W_{ir}\mathbf{x}[5] + W_{or}\mathbf{y}[4]) = (h_1[5], h_2[5], \cdots, h_7[5])$ and $\mathbf{y}[5] = g(W_{io}\mathbf{x}[5] + W_{ro}\mathbf{h}[5])$. Modify $W_{rr}^{(5)} \rightarrow W_{rr}^{(5)} - \eta DW_{rr}^{(5)}$.

- Calculate $\mathbf{h}[6] = f(W_{rr}^{(6)}\mathbf{h}[5] + W_{ir}\mathbf{x}[6] + W_{or}\mathbf{y}[5])$ and $\mathbf{y}[6] = g(W_{io}\mathbf{x}[6] + W_{ro}\mathbf{h}[6])$.

  Modify $W_{rr}^{(6)} \rightarrow W_{rr}^{(6)} - \eta DW_{rr}^{(6)}$.

- Calculate $\mathbf{h}[7] = f(W_{rr}^{(7)}\mathbf{h}[6] + W_{ir}\mathbf{x}[7] + W_{or}\mathbf{y}[6])$ and $\mathbf{y}[7] = g(W_{io}\mathbf{x}[7] + W_{ro}\mathbf{h}[7])$.

  Modify $W_{rr}^{(7)} \rightarrow W_{rr}^{(7)} - \eta DW_{rr}^{(7)}$.

- Calculate $\mathbf{h}[8] = f(W_{rr}^{(8)}\mathbf{h}[7] + W_{ir}\mathbf{x}[8] + W_{or}\mathbf{y}[7])$ and $\mathbf{y}[8] = g(W_{ro}\mathbf{h}[8])$.

  Modify $W_{rr}^{(8)} \rightarrow W_{rr}^{(8)} - \eta DW_{rr}^{(8)}$.

  - Calculate $\mathbf{h}[9] = f(W_{rr}^{(9)}\mathbf{h}[8] + W_{ir}\mathbf{x}[9] + W_{or}\mathbf{y}[8])$ and $\mathbf{y}[9] = g(W_{io}\mathbf{x}[9] + W_{ro}\mathbf{h}[9])$.

    Modify $W_{rr}^{(9)} \rightarrow W_{rr}^{(9)} - \eta DW_{rr}^{(9)}$.

  - Calculate $\mathbf{h}[10] = f(W_{rr}^{(10)}\mathbf{h}[9] + W_{ir}\mathbf{x}[10] + W_{or}\mathbf{y}[9])$ and $\mathbf{y}[10] = g(W_{io}\mathbf{x}[10] + W_{ro}\mathbf{h}[10])$.

    Modify $W_{rr}^{(10)} \rightarrow W_{rr}^{(10)} - \eta DW_{rr}^{(10)}$.

- Calculate $\mathbf{h}[11] = f(W_{rr}^{(11)}\mathbf{h}[10] + W_{ir}\mathbf{x}[11] + W_{or}\mathbf{y}[10])$ and $\mathbf{y}[11] = g(W_{io}\mathbf{x}[11] + W_{ro}\mathbf{h}[11])$.
  Modify $W_{rr}^{(11)} \rightarrow W_{rr}^{(11)} - \eta DW_{rr}^{(11)}$.

- Calculate $\mathbf{h}[12] = f(W_{rr}^{(12)}\mathbf{h}[11] + W_{ir}\mathbf{x}[12] + W_{or}\mathbf{y}[11])$ and $\mathbf{y}[12] = g(W_{io}\mathbf{x}[12] + W_{ro}\mathbf{h}[12])$.
  Modify $W_{rr}^{(12)} \rightarrow W_{rr}^{(12)} - \eta DW_{rr}^{(12)}$.

- Calculate $\mathbf{h}[13] = f(W_{rr}^{(13)}\mathbf{h}[12] + W_{ir}\mathbf{x}[13] + W_{or}\mathbf{y}[12])$ and $\mathbf{y}[13] = g(W_{io}\mathbf{x}[13] + W_{ro}\mathbf{h}[13])$.
  Modify $W_{rr}^{(13)} \rightarrow W_{rr}^{(13)} - \eta DW_{rr}^{(13)}$.

- Calculate $\mathbf{h}[14] = f(W_{rr}\mathbf{h}[13] + W_{ir}\mathbf{x}[14] + W_{or}\mathbf{y}[13])$ and $\mathbf{y}[14] = g(W_{io}\mathbf{x}[14] + W_{ro}\mathbf{h}[14])$.
  Modify $W_{rr}^{(14)} \rightarrow W_{rr}^{(14)} - \eta DW_{rr}^{(14)}$.

- Calculate $\mathbf{h}[15] = f(W_{rr}^{(15)}\mathbf{h}[14] + W_{ir}\mathbf{x}[15] + W_{or}\mathbf{y}[14])$ and $\mathbf{y}[15] = g(W_{io}\mathbf{x}[15] + W_{ro}\mathbf{h}[15])$.
  Modify $W_{rr}^{(15)} \rightarrow W_{rr}^{(15)} - \eta DW_{rr}^{(15)}$.

- Calculate $\mathbf{h}[16] = f(W_{rr}^{(16)}\mathbf{h}[15] + W_{ir}\mathbf{x}[16] + W_{or}\mathbf{y}[15])$ and $\mathbf{y}[16] = g(W_{ro}\mathbf{h}[16])$.
  Modify $W_{rr}^{(16)} \rightarrow W_{rr}^{(16)} - \eta DW_{rr}^{(16)}$

- Calculate $\mathbf{h}[1] = f(W_{rr}^{(1)}\mathbf{h}[16] + W_{ir}\mathbf{x}[1] + W_{or}\mathbf{y}[16])$ and $\mathbf{y}[1] = g(W_{io}\mathbf{x}[1] + W_{ro}\mathbf{h}[1])$.
  Modify $W_{rr}^{(1)} \rightarrow W_{rr}^{(1)} - \eta DW_{rr}^{(12)}$

- Calculate $\mathbf{h}[2] = f(W_{rr}^{(2)}\mathbf{h}[1] + W_{ir}\mathbf{x}[2] + W_{or}\mathbf{y}[1])$ and $\mathbf{y}[2] = g(W_{io}\mathbf{x}[2] + W_{ro}\mathbf{h}[2])$.
  Modify $W_{rr}^{(2)} \rightarrow W_{rr}^{(2)} - \eta DW_{rr}^{(2)}$.

- Calculate $\mathbf{h}[3] = f(W_{rr}^{(3)}\mathbf{h}[2] + W_{ir}\mathbf{x}[3] + W_{or}\mathbf{y}[2])$ and $\mathbf{y}[3] = g(W_{io}\mathbf{x}[3] + W_{ro}\mathbf{h}[3])$.
  Modify $W_{rr}^{(3)} \rightarrow W_{rr}^{(3)} - \eta DW_{rr}^{(3)}$.

The cycle from the state vector **h**[4] to the **h**[3] continues until optimized Wir, Wrr, Wor are found.

$$\text{Store } \mathbf{S} = \begin{bmatrix} \mathbf{x}^T[4] & \mathbf{h}^T[4] \\ \mathbf{x}^T[7] & \mathbf{h}^T[7] \\ \mathbf{x}^T[8] & \mathbf{h}^T[8] \\ \mathbf{x}^T[12] & \mathbf{h}^T[12] \\ \mathbf{x}^T[15] & \mathbf{h}^T[15] \\ \mathbf{x}^T[16] & \mathbf{h}^T[16] \end{bmatrix} \text{ and } \mathbf{y}^* = \begin{bmatrix} \mathbf{y}^*[4] \\ \mathbf{y}^*[7] \\ \mathbf{y}^*[8] \\ \mathbf{y}^*[12] \\ \mathbf{y}^*[15] \\ \mathbf{y}^*[16] \end{bmatrix}.$$

As in the case of (2+1) D, we optimize the weight function of 7 paths (L19, L20, L21, L22, L23, L24, L25) in (3+1) D that minimize the loss

$$\mathcal{L} = ||\mathbf{S}\,\mathbf{W} - \mathbf{y}^*||^2, \text{ where } \mathbf{W} = [W_{io}, W_{ro}]^T.$$

$$\mathbf{SW} = \begin{bmatrix} (W_{io}\mathbf{x})^T[4] & (W_{ro}\mathbf{h})^T[4] \\ (W_{io}\mathbf{x})^T[7] & (W_{ro}\mathbf{h})^T[7] \\ (W_{io}\mathbf{x})^T[8] & (W_{ro}\mathbf{h})^T[8] \\ (W_{io}\mathbf{x})^T[12] & (W_{ro}\mathbf{h})^T[12] \\ (W_{io}\mathbf{x})^T[15] & (W_{ro}\mathbf{h})^T[15] \\ (W_{io}\mathbf{x})^T[16] & (W_{ro}\mathbf{h})^T[16] \end{bmatrix}$$

We adopt a cylindrical lattice model, such that 7 paths start from the origin of a space and returns to the origin. The total action becomes 0 when the path returns to the origin. Therefore, at t=8 and t=16, Wir **x**[t] become 0 at these epochs.

We are considering paths in momentum space, the Fourier transform of the path in the position space. In the calculation of x[t] we consider the projection space RP^4 which means that we optimize the scale of the 4D vectors.

In the ERNN analysis, we observed that time shift of paths L21 and L22 can have base 14 or 24, and we distinguished the two by L21, L21' and L22, L22', and considered 9 paths.

4. MC SIMULATION RESULTS

We performed a test run of ESN of 10 cycles or 160 t, with and without dropping Wio x[8] and Wio x[16] and checked that they are indeed 0. The output y[t], which is 7dimensional vector, and t runs from t=4, …, 16, 1, 2, 3 modulus 16. We want to obtain weights of 7 bases for reproducing the 6 outputs y[t] corresponding to t=4, 5, 9, 10,11,12, which are used in the ERNN. In this test run, we take **y***[6]=y*[5](2/3), **y***[7]= **y***[5] (1/3), **y***[8]=0, **y***[13]=**y***[12](3/4), **y***[14]=**y***[12](1/2). **y***:[15]= **y***[12](1/4), **y***[16]=0, **y***[1]=**y***[4](1/4), **y***[2]=**y***[4](1/2), **y***[3] =**y***[4](3/4).

4.1. Sigmoid Function

At t=8 and t=16, the random walk of the path returns to the original in the momentum space, and the outputs are produced by biases. Therefore, we omit presentation of deviations at these t.
The figure indicates that at t=8, the path L19 has large deviation, while at t=7, the path L25 has a large deviation.

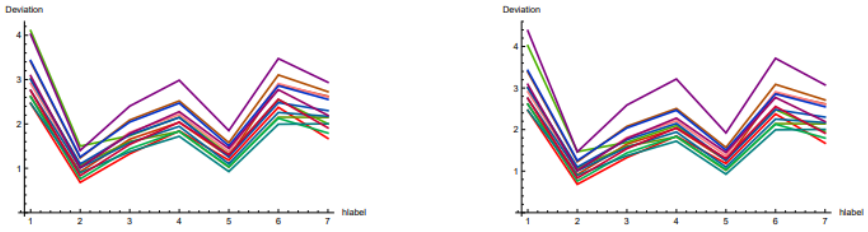

Figure 7: The deviation $||\mathbf{y}[t] - \mathbf{y}^*[t]||^2$ of 14 outputs. $t = 4$:red, $t = 7$: pink, $t = 8$: orange, $t = 12$: gray, $t = 15$: green, $t = 16$: blue. After 5000 steps (left) and after 85,000 steps (right)

We observe that the deviations of || y[t]-y*[t]||^2 are reduced in higher cycles, but reduction is slow.

The state vector h[t] is largest at t=8, but the standard deviations of **h**[t] are large for all t in the simulation using 40 samples. In order to perform the optimization effectively, it is necessary to increase the size of samples or reduce the parameter η. We took η=0.01 in general and 0.000001 for sigmoid function. The behaviors of **h**[t]+**b**h of t=8 and t=16 are defined by **b**h.

The deviation and the mean values of state vector **h**[t] are shown in Figure 8 and Figure 9, respectively.

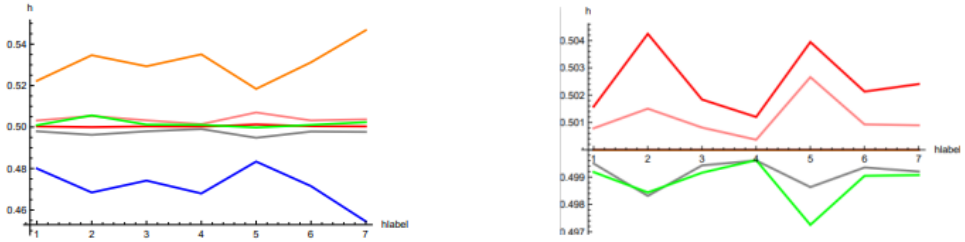

Figure 8: The mean value of $\mathbf{h}[t]$ that yields outputs at $t = 4$:red, $t = 7$: pink, $t = 8$: orange, $t = 12$: gray, $t = 15$: green, $t = 16$: blue. after 10 steps (left) and after 85,000 steps (right).

We observe that h[8] and h[16] behave opposite trends as the number of path decrease or increase at small steps.

However as the steps increase the averages approach h[8]=h[16]=0.5.

With η=0.01, effects of bias **b**h become small as the steps increase. We trained Wrr at epochs 4,7,8,12,15 and 16, where mixings of paths occur.

The slow convergence motivated training of the Wrr^(t) matrices at t=4,7,8,12,15 and 16. We modified Wrr^(t) of other epochs, but the deviation and standard deviations of **h**[t] remained large.

When cycle number increases the deviation of t=4,7,15 do not vary much, but t=8 and t=16 become closer and larger. $y_i[8]$ and $y_i[16]$ (i=1,…, 7) concentrate either nearly 1 or nearly 0.

4.2. Tanh Function

Since the deviation calculated by sigmoid functions remaind large, we tried to use tanh function for the nonlinear function as used in the Long Short Term Memory (LSTM) cells method [2].

The deviations using the tanh function are much reduced from those obtained by the sigmoid function. However, near t=9 and t=16, large deviation appears. It is due to the fact that at t=8 and t=16, the random walk returns to the original position, and y*[t] is taken to be zero. The good property of this method is that from 3000steps to 4000stes, the weight is stable.

The obtained **y**[9] and **y**[16] indicates that **y***[8]= y*[16]=0 derived from the action of paths returning to the original in 3D space is inappropriate. One should consider paths that starts in the y-z plane and z-x plane. In other words, origin of large constant deviations

$$||\mathbf{y}[t] - \mathbf{y}^*[t]||^2$$

near t=8 and t=16 are known. One could consider making an average of the three initial conditions for **y**[t].

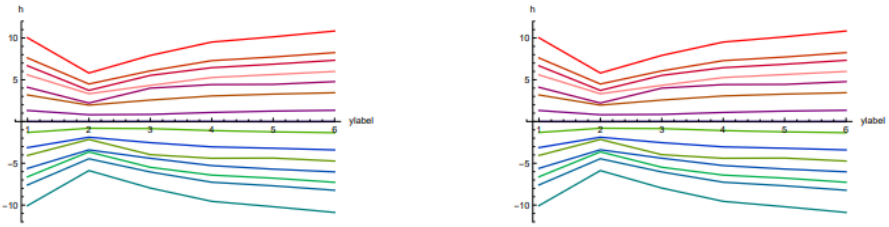

Figure 11: The mean value of $\mathbf{h}[t]$ 40 training samples that yields outputs at $t = 4$:red, to $t = 16$:plue and $t = 1$ to $t = 3$, after 3000 steps (left) and after 4000 steps (right).

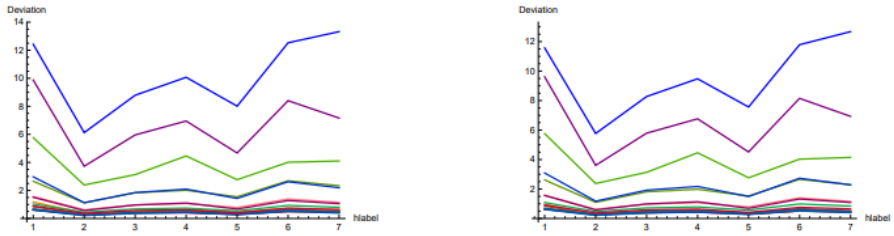

Figure 12: The variation $||\mathbf{y}[t] - \mathbf{y}^*[t]||^2$ 40 training samples after 3000 steps (left) and after 4000 steps (right). Deviation is large at $t = 16$ (blue), $t = 1$ (violet), $t = 9$ (green), $t = 15$ (pale blue) $t = 10$ (dark green). .

5. CONCLUSIONS AND OUTLOOK

We showed that the weight function of paths defined by the fixed point action can be optimized by the ESN using the tanh function for expressing the nonlinearity.

The optimal weight function of the C-type fixed point actions which contain hysteresis effect can be simulated by using the biquaternion basis. We observed stability of the action on the output

layer produced from the reservoirs. A main difference from the ERNN in our application is that there exists output from the reservoir and mixing among reservoir Wrr exist in ENS, while in ERNN, in our application output from additional 2 inputs exist and no direct output from the hidden layer. ERNN in general contains outputs from hidden layers, but in our ERNN, outputs have direct connection with values in the input layers. The ESN has more freedom of incorporating hysteresis effects.

For establishing our methods, comparison between experiments and ESN simulation results in lower dimensional system would be useful. Our method can be used for a comparison of data of TR-NEWS experiment in (2+1) D using quaternion bases which are Clifford algebra bases.

We analyzed (3+1) D ultrasonic wave propagation with hysteresis effects using biquaternions. Quaternion and biquaternion basis model can be used not only for NDT, but also for QCD lattice simulations.

Adler [23,24,25] used quaternions in generalized quantum dynmics, and in operator gauge invariant quaternionic field theory. Adler considered the total trace lagrangean and hamiltonian dynamics and asked, "Given two scalar or fermion quaternionic operator fields, is there a criterion for determining whether they are related by a bi-unitary operator gauge transformation?"[25].

Our systems are not related by a gauge transformation but related by a different choice of bases.

In signal processing, generalization of Fourier transform, Laplace transform, etc. which is called the linear canonical transformation (LCT) is utilized. Its extension to quaternion Fourier Transform and discussion on Heisenberg's uncertainty principle are given in [26] and [27]. In these works, outputs of quaternion linear time varying system

$$R(\xi, b) = \int_{R^2} H_1(\xi, b, t) f(t) dt$$

are discussed. The ξ is a vector in R^2, and b represents the time shift. An extension of quaternion windowed linear canonical transform (QWLCT) to biquaternion system is left for the future.

For getting the optimal solution of these problems, Machine Learning techniques can be applied. Nonlinearity and hysteresis could be explored in these bases. As shown in [17], the technique is applicable for an extension of dental investigation, as an example, which is restricted at present in (2+1) D system to (3+1) D system. The damaged position can be detected in the 3D space if receivers and transducers of TR waves are distributed in the 3D space and signals traveling to all directions can be detected.

ACKNOWLEDGEMENTS

S.F. thanks the Laboratory for Industrial Research (Nissanken) for the financial aid to the travel expense to INSA Centre Val de Loire in November 2023.

The numerical calculation was done using Mathematica of the Wolfram Research [28] installed on a workstation of the faculty of science and engineering of Teikyo University. S.F. is grateful to Prof. M. Arai and Prof. K. Hamada for the convenience.

REFERENCES

[1] Calin, O. (2020) "Deep Learning Architectures: A Mathematical Approach", Springer Series in the Data Science.

[2] Raschka, S.,Liu,Y. and Mirjalili,V. (2022) "Machine Learning with Pytorch and Scikit-Learn", Packt Publishing

[3] Bianchi, E.M. et al. (2017) "Recurrent Neural Networks for Short-Term Load Forecasting", Springer Briefs in Computer Science, https://doi.org/10.1007/978-3-319-70338-1

[4] Bianchi, F.M., Scardapane, S., Uncini, A., Rizzi, A. and Sadeghian, A. (2015) "Prediction of telephone calls load using Echo State Network with exogeneous variables", Neural Networks, Elsevier **71**, 204-213.

[5] Furui, S. and Dos Santos, S. (2024) "The optimization of paths In the R^{3,1} space-time by Markov Chain Monte Carlos", arXiv:2406.14569 v1

[6] Furui, S. and Dos Santos, S. (2023) "Clifford Fourier Transforms in (2+1) D Lattice Simulations of Soliton Propagations", PoS Lattice22, The 39th International Symposium on Lattice Field Theory, 8th-13th August, 2022, Rheinische Friedrich-Wilhelms-Universitaet Bonn, Bonn, Germany, arXiv:[hep-lat physics.comp-ph]

[7] Furui, S. and Dos Santos, S. (2023) "Application of Quaternion Neural Network to Time Reversal Based Nonlinear Elastic Wave Spectroscopy", INAE, **8** 183-199.

[8] Miran, S., Flamant, J. , Le Bihan, N. , Chainais,P. and Brie, D.(2023) "Quaternion in Signal and Image Processing, A comprehensive and objective overview", IEEE Signal Processing Magazine **40** (6).

[9] Felsberg, M. and Sommer, G. (2001) "The Monogenic Signal", IEEE. Trans. Signal Proc.**49** (12) 3136-3144.

[10] Unser, M., Sage, D. and Van De Ville, D. (2009) "Multiresolutoion Monogenic Signal Analysis Using the Riesz-Laplace Wavelet Transform", IEEE Trans. Image Proc. **18** (11).

[11] Unser, M. and Van De Ville, D. (2010) "Wavelet Steerability and the Higher-Order Riesz Transform", IEEE Trans. Image Proc. **19** (3) 636-652.

[12] Bridge, C.P. (2017) "An introduction to the monogenic signal", arXiv:1703.09199v1 [cs.CV]

[13] Dvorakova, Z., Dos Santos, S., Kus V. and Prevolovsky, Z. (2023) "Localization and Classification of scattered nonlinear ultrasonic signatures in bio-mechanical media using time reversal approach", J. Acoust. Soc. Am. **154** (3) pp.1684-1695.

[14] Lopez, A., Bacelar, R., Pires, I., Santos, T. G., Sousa,J. P. and Quintino, L. (2018) "Non-destrctive testing application of radiography and ultrasound for wire and arc additive manufacturing", Additive Manufacturing, **21**, 298-306, Elsevier.

[15] Goursolle, T., Calle, S., Dos Santos, S. and Bou Matar, O. (2007) "A two-dimensional pseudospectral model for time reversal and nonlinear elastic wave spectroscopy", J. Accoust. Soc. Am **122**(6) pp.3220-3229.

[16] Dos Santos, S. and Plag, C., (2008) "Excitation Symmetry Analysis Method (ESAM) for Calculatiion of Higher Order Nonlinearities", Int. J. Non-Linear Mech. **43** 104-119.

[17] Dos Santos S. and Prevorovsky, Z. (2010) "Ultrasonographie dentaire par techniques TR-NEWS", 10eme Congres Francais d'Acoustique, Lyon

[18] Nelson, T.R. (2000) "Three-Dimensional Imaging", Ultrasound in Med & Biol **26** Suppl. 1, Elsevier pp. S35-S38.

[19] Garling,D.J.H. (2011) "Clifford Algebras: An Introduction", Cambridge University Press.

[20] DeGrand, T., Hasenfratz, A., Hasenfratz, P. and Niedermayer, F. (1995) "Non-perturbative tests of the fixed point action for SU(3) gauge theory", Nucl. Phys. **B454** 615-637: arXiv:9506031[hep-lat].

[21] Mayergoyz, I. (2003) "Mathematical Models of Hysteresis and their Applications", Elsevier, Amsterdam.

[22] Furui,S. (2023) "On the Quadratic Phase Quaternion Domain Fourier Transform and on the Clifford algebra of R^{3,1}", arXiv:2310.10680 v4.

[23] Adler, S.L. (1985) "Quaternionic Quantum Field Theory", Phys. Rev. Lett. **55** (8) 783-786. Errata (1985) **55** (13) 1430

[24] Adler, S.L. (1986) "Quaternionic Quantum Field Theory", Commun. Math. Phys, **104**, 611-656.

[25] Adler, S.L. (1994) "Generalized quantum dynamics", Nuclear Physics, **B415**, 195-242.

[26] Bahri, M. and Ashino, R. (2016) “A Simplified Proof of Uncertainty Principle for Quaternion Linear Canonical Transform”, Hindwari Abstract and Applied Analysis, Volume 2016, Article I 5879430.

[27] Prasad, A. and Kundu, M. (2023) “Uncertainty principles and Applications of quaternion windowed linear canonical transform,” Elsevier, Optik-International Journal for Light and Electron Optics, **272** 170222.

[28] Wolfram Research (2023) “Mathematica14”

AUTHORS

Sadataka Furui, Research Collaborator at Faculty of Science and Engineering of Teikyo University, Formerly Professor of Graduate School of Science and Engineering of Teikyo University, Utsunomiya, Japan

Serge Dos Santos, Associate Professor at INSA Centre Val de Loire at Blois, Universite de Tours, INSERM, Imaging Brain & Neuropsychiatry iBraiN U1253 F-411034 Blois, France