Generalization Characteristics of Complex-Valued Feedforward Neural Networks in Relation to Signal Coherence

Akira Hirose, *Senior Member, IEEE*, and Shotaro Yoshida

***Abstract*— Applications of complex-valued neural networks (CVNNs) have expanded widely in recent years—in particular in radar and coherent imaging systems. In general, the most important merit of neural networks lies in their generalization ability. This paper compares the generalization characteristics of complex-valued and real-valued feedforward neural networks in terms of the coherence of the signals to be dealt with. We assume a task of function approximation such as interpolation of temporal signals. Simulation and real-world experiments demonstrate that CVNNs with amplitude-phase-type activation function show smaller generalization error than real-valued networks, such as bivariate and dual-univariate real-valued neural networks. Based on the results, we discuss how the generalization characteristics are influenced by the coherence of the signals depending on the degree of freedom in the learning and on the circularity in neural dynamics.**

***Index Terms*— Complex-valued neural network, function approximation, generalization, supervised learning.**

I. Introduction

COMPLEX-VALUED neural networks (CVNNs) have extended their applications into various fields, such as metal-defect imagers [1], radars (including ground penetrating radars to visualize plastic landmines) [2], [3], blur-compensation image processing [4], blind separation based on principal component analysis in sonar [6] and voice processing [7], filtering and other time-sequential signal processing [8], frequency-domain multiplexed microwave signal processing [9] and pulse beamforming in ultra-wideband communications [10], frequency-domain multiplexed neural networks and learning logic circuits using lightwave [11]–[13], theory and design of quantum computation based on superconductive devices [14], and developmental motion learning [15]. Many of these applications deal with wave-related signals, or coherent signals, in the time/space domain or Fourier domain.

Research on general CVNNs has revealed various aspects of their dynamics. It is true that a complex number is represented by a pair of real numbers, namely, real and imaginary parts, or amplitude and phase. A variety of useful neural dynamics are actually obtained by paying attention to the real and imaginary parts [16]–[18] or amplitude and phase [19], [20]. This fact sometimes leads to an assumption that a CVNN is almost equivalent to a double-dimensional real-valued neural network (RVNN). However, it has been discussed intuitively that CVNN has only a smaller degree of freedom at the synaptic weighting, because of the specific complex multiplication in the four arithmetic operations, and that this fact suggests the realization of learning with the presumption to obey phase-rotational dynamics [21].

In this paper, we compare complex- and real-valued neural networks by focusing on their generalization characteristics. Generalization is one of the features most useful and specific to neural networks. We investigate the generalization ability of feedforward CVNNs and RVNNs for function approximation, such as filtering and interpolation. We observe the characteristics by feeding signals that have various degrees of coherence.

We examine four types of neural networks. As a basic complex-valued network, we adopt a CVNN having an amplitude-phase-type activation function because of its suitability to circular (isotropic in the complex plane) signals, which are often observed in coherent imaging and other electronics. We also examine a CVNN having a real-imaginary separate-type activation function (RI-CVNN), which is another widely used CVNN. Regarding real-valued networks, basically we construct a double-input-terminal double-output-neuron RVNN with full connections between layers, which processes real and imaginary signals in a bivariate manner. We also prepare a dual-univariate RVNN (dual-RVNN) with the same numbers of input terminals and output neurons but processing the real part of the signals separately from the imaginary part.

Computer simulations as well as experiments for real-world signal data demonstrate that the generalization characteristics of the four types of the neural networks are different depending on the signal coherence. We discuss the origin of the difference in the generalization abilities based on simulation and experiment results.

This paper is organized as follows. Section II presents the construction of experiments, network structures, and learning dynamics. In Section III, we conduct computer simulations as well as electronics experiments. We find that the generalization characteristics are different among the four networks. In Section IV, we discuss how the signal properties influence the generalization characteristics. Section V concludes this paper.

Manuscript received May 8, 2011; revised January 5, 2011; accepted January 5, 2011. Date of publication January 23, 2012; date of current version March 6, 2012.

The authors are with the Department of Electrical Engineering and Information Systems, University of Tokyo, Tokyo 113-8656, Japan (e-mail: ahirose@ee.t.u-tokyo.ac.jp; yoshida@eis.t.u-tokyo.ac.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2012.2183613

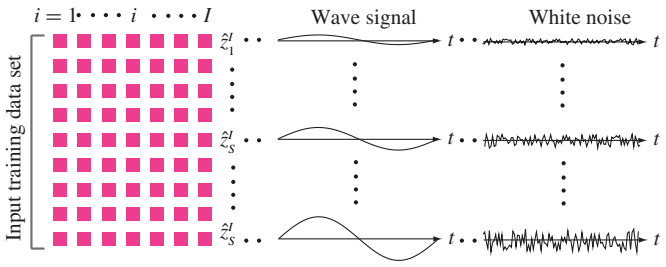

Fig. 1. Set of teacher signals.

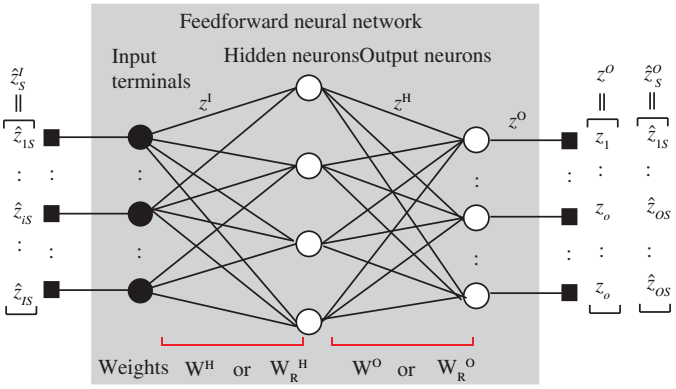

Fig. 2. Basic construction of the complex- and real-valued feedforward neural networks.

II. Construction of Experiments and Learning Dynamics

A. Coherence

We investigate numerically and experimentally the difference in the generalization characteristics of CVNNs and RVNNs. We also focus on the dependence of the generalization on coherence (degree of wave nature) of the signals. In the following computer simulation and experiments, we observe the generalization characteristics by changing the coherence of the signals with the following definition.

Fig. 1 illustrates the generation of signals having various coherence degrees. We take a sinusoidal time-sequential signal $v(t) = a\exp(\jmath\omega t)$ of time t as a completely coherent signal where a and ω are the amplitude and angular frequency, respectively. We memorize the serial data discretely, and feed it to the neural input terminals in parallel, just like in the case of a finite impulse response filter. We may add white Gaussian noise (WGN) to the sinusoid to reduce the coherence. WGN is completely incoherent. By changing the ratio of the sinusoid to the white noise, we prepare signals having various coherence degrees. Then the signal-to-noise ratio $SNR \equiv P_s/P_n$, the ratio of the sinusoidal wave power P_s to the WGN power P_n, shows the coherence degree quantitatively. When $SNR = \infty$, the signal is completely coherent, while when $SNR = 0$ (or $-\infty$dB), it is completely incoherent.

Section II-B describes the construction of the computer simulations and experiments shown in this paper. We choose a typical situation in which we can observe the dependence of generalization on the coherence. We implicitly suppose a neural network that processes signals related to lightwave, electromagnetic wave, or ultrasonic wave in, for example, a sonar imaging system.

Regarding the word "coherence," we have several closely related concepts. One of them is bandwidth, which expresses signal characteristics in the spectral domain. A high coherence signal is a narrowband signal. In the spectrum, the amount of disturbance that reduces the coherence is determined by the product of two factors, namely, bandwidth and spectral power density. Here we adopt a widely applicable WGN as the disturbance, which has infinitely wide bandwidth. In the experiment, however, we deal with discrete-time signals and, therefore, we have a limited bandwidth determined automatically by the sampling process. Then, the above definition of coherence by the signal to noise ratio (SNR) corresponds to the level of the flat WGN power density under a constant bandwidth condition. In this sense, the SNR is directly related to the coherence under the constant bandwidth condition. Another concept is circularity, which mostly means probabilistic isotropy of the signal in the complex plane. This is an important concept in, for example, prediction of wind force and direction for wind power generation. In a sinusoidal wave, circularity is related to distortion. However, the circularity does not include temporal continuity characteristics explicitly or sampling interval influence for discretization of continuous signals. Circularity and the widely linear processing will be discussed in Section IV.

B. Organization of the Experiment

The outline of our experiments is as follows.

1) *Input signals*: weighted summation of the following.
   a) Sinusoid: completely coherent signal.
   b) WGN: completely incoherent data having random amplitude and phase (or real and imaginary parts).
2) *Task to learn*: identity mapping, which is expected to show the learning characteristics most clearly for the above signals with various degrees of coherence.
3) *Evaluation of generalization*: observation of the generalization error when the input signals shift in time, or when the amplitude is changed.

C. Forward Processing and Learning Dynamics

1) CVNN: Fig. 2 shows the general construction of the neural network to be considered here. It is a layered feedforward network having input terminals, hidden neurons, and output neurons. In a CVNN, we first employ a phase–amplitude-type sigmoid activation function and the teacher–signal–backpropagation learning process, [19], [22] with the following notations:

$$z^{\mathrm{I}} = [z_1, \ldots, z_i, \ldots, z_I, z_{I+1}]^T \quad \text{(Input signal vector)} \tag{1}$$

$$z^{\mathrm{H}} = [z_1, \ldots, z_h, \ldots, z_H, z_{H+1}]^T \quad \text{(Hidden} - \text{layer output signal vector)} \tag{2}$$

$$z^{\mathrm{O}} = [z_1, \ldots, z_o, \ldots, z_O]^T \quad \text{(Output} - \text{layer signal vector)} \tag{3}$$

$$\mathbf{W}^{\mathrm{H}} = [w_{hi}] \quad \text{(Hidden neuron weight matrix)} \tag{4}$$

$$\mathbf{W}^{\mathrm{O}} = [w_{oh}] \quad \text{(Output neuron weight matrix)} \tag{5}$$

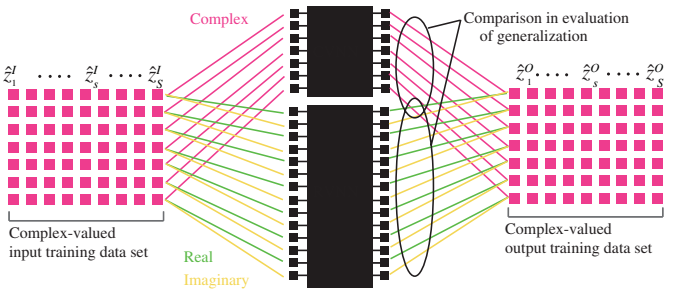

Fig. 3. Schematic diagram of the learning process for pairs of input–output teachers.

where $[\cdot]^T$ means transpose. In (4) and (5), the weight matrices include additional weights $w_{h\ I+1}$ and $w_{o\ H+1}$, equivalent to neural thresholds, where we add formal constant inputs $z_{I+1} = 1$ and $z_{H+1} = 1$ in (1) and (2), respectively. The respective signal vectors and synaptic weights are connected with one another through an activation function $f(z)$ as

$$z^{\mathrm{H}} = f\left(\mathbf{W}^{\mathrm{H}} z^{\mathrm{I}}\right) \tag{6}$$

$$z^{\mathrm{O}} = f\left(\mathbf{W}^{\mathrm{O}} z^{\mathrm{H}}\right) \tag{7}$$

where $f(z)$ is a function of each vector element z ($\in \boldsymbol{C}$) defined as

$$f(z) = \tanh(|z|)\exp(\jmath \arg z). \tag{8}$$

Fig. 3 is a diagram to explain the supervised learning process. We prepare a set of teacher signals at the input $\hat{z}_s^{\mathrm{I}} = [\hat{z}_{1s}, \ldots, \hat{z}_{is}, \ldots, \hat{z}_{Is}, \hat{z}_{I+1\ s}]^T$ and the output $\hat{z}_s^{\mathrm{O}} = [\hat{z}_{1s}, \ldots, \hat{z}_{os}, \ldots, \hat{z}_{Os}]^T$ $(s = 1, \ldots, s, \ldots S)$ for which we employ the teacher–signal backpropagation learning. We define an error function E to obtain the dynamics by referring to [9], [19], and [22] as

$$E \equiv \frac{1}{2}\sum_{s=1}^{S}\sum_{o=1}^{O}\left|z_o(\hat{z}_s^{\mathrm{I}}) - \hat{z}_{os}\right|^2 \tag{9}$$

$$\left|w_{oh}^{\mathrm{new}}\right| = \left|w_{oh}^{\mathrm{old}}\right| - K\frac{\partial E}{\partial |w_{oh}|} \tag{10}$$

$$\arg w_{oh}^{\mathrm{new}} = \arg w_{oh}^{\mathrm{old}} - K\frac{1}{|w_{oh}|}\frac{\partial E}{\partial(\arg w_{oh})} \tag{11}$$

$$\begin{aligned}\frac{\partial E}{\partial |w_{oh}|} = {} & \left(1 - |z_o|^2\right)\left(|z_o| - \left|\hat{z}_o\right|\cos\left(\arg z_o - \arg \hat{z}_o\right)\right)|z_h| \\ & \cdot \cos\left(\arg z_o - \arg \hat{z}_o - \arg w_{oh}\right) \\ & - |z_o|\left|\hat{z}_o\right|\sin\left(\arg z_o - \arg \hat{z}_o\right)\frac{|z_h|}{\tanh^{-1}|z_o|} \\ & \cdot \sin\left(\arg z_o - \arg \hat{z}_o - \arg w_{oh}\right)\end{aligned} \tag{12}$$

$$\begin{aligned}\frac{1}{|w_{oh}|}\frac{\partial E}{\partial(\arg w_{oh})} = {} & \left(1 - |z_o|^2\right)\left(|z_o| - \left|\hat{z}_o\right|\cos\left(\arg z_o - \arg \hat{z}_o\right)\right)|z_h| \\ & \cdot \sin\left(\arg z_o - \arg \hat{z}_o - \arg w_{oh}\right) \\ & + |z_o|\left|\hat{z}_o\right|\sin\left(\arg z_o - \arg \hat{z}_o\right)\frac{|z_h|}{\tanh^{-1}|z_o|} \\ & \cdot \cos\left(\arg z_o - \arg \hat{z}_o - \arg w_{oh}\right)\end{aligned} \tag{13}$$

where $(\cdot)^{\mathrm{new}}$ and $(\cdot)^{\mathrm{old}}$ indicates the update of the weights from $(\cdot)^{\mathrm{old}}$ to $(\cdot)^{\mathrm{new}}$, and K is a learning constant. The teacher signals at the hidden layer $\hat{z}^{\mathrm{H}} = [\hat{z}_1, \ldots, \hat{z}_h, \ldots, \hat{z}_H]^T$ is obtained by making the output teacher vector itself $\hat{z}^{\mathrm{O}}$ propagate backward as

$$\hat{z}^{\mathrm{H}} = \left(f\left(\left(\hat{z}^{\mathrm{O}}\right)^*\hat{\mathbf{W}}^{\mathrm{O}}\right)\right)^* \tag{14}$$

where $(\cdot)^*$ denotes the Hermite conjugate. Using $\hat{z}^{\mathrm{H}}$, the hidden layer neurons change their weights by following (10)–(13) with replacement of the suffixes o,h with h, i [11], [22].

2) CVNN Having Real–Imaginary Separate-Type Activation Function (RI-CVNN): We also investigate the characteristics of CVNNs having real–imaginary separate-type activation function. Instead of (8), a neuron has an activation function expressed as

$$f(z) = \tanh(\mathbf{Re}[z]) + \jmath\tanh(\mathbf{Im}[z]). \tag{15}$$

The structure and the dynamics of feedforward processing and backpropagation learning are those described in, for example, [18].

3) RVNN Having Double Input Terminals and Output Neurons for Bivariate Procesing (RVNN): Similarly, the forward processing and learning of a RVNN having double input terminals and output neurons are explained as follows. Fig. 3 includes also this case. We represent a complex number as a pair of real numbers as $z_i = x_{2i-1} + \jmath x_{2i}$. Then we have a double number of terminals for real and imaginary parts of input signals $z_{\mathrm{R}}^{\mathrm{I}}$ and a double number of output neurons to generate real and imaginary parts of output signals $z_{\mathrm{R}}^{\mathrm{O}}$. We also prepare a double number of hidden neurons for hidden layer signals $z_{\mathrm{R}}^{\mathrm{H}}$ so that the equivalent number of neurons is the same as that of the above CVNN.

Forward signal processing connects the signal vectors as well as hidden neuron weights $\mathbf{W}_{\mathrm{R}}^{\mathrm{H}}$ and output neuron weights $\mathbf{W}_{\mathrm{R}}^{\mathrm{O}}$ through a real-valued activation function f_{R} as

$$\begin{aligned}z_{\mathrm{R}}^{\mathrm{I}} = [& \overbrace{x_1, \quad x_2}^{\text{real\&imaginary}}, \ldots, x_{2i-1}, x_{2i}, \ldots, \\ & x_{2I-1}, x_{2I}, x_{2I+1}, x_{2I+2}]^T \\ & \left(= z^{\mathrm{I}}\right) \quad \text{(Input signal vector)}\end{aligned} \tag{16}$$

$$\begin{aligned}z_{\mathrm{R}}^{\mathrm{H}} = {} & [x_1, x_2, \ldots, x_{2h-1}, x_{2h}, \ldots, \\ & x_{2H-1}, x_{2H}, x_{2H+1}, x_{2H+2}]^T \\ & \text{(Hidden}-\text{layer output signal vector)}\end{aligned} \tag{17}$$

$$\begin{aligned}z_{\mathrm{R}}^{\mathrm{O}} = {} & [x_1, x_2, \ldots, x_{2o-1}, x_{2o}, \ldots, x_{2O-1}, x_{2O}]^T \\ & \text{(Output}-\text{layer signal vector)}\end{aligned} \tag{18}$$

$$\mathbf{W}_{\mathrm{R}}^{\mathrm{H}} = [w_{\mathrm{R}hi}] \quad \text{(Hidden neuron weight matrix)} \tag{19}$$

$$\mathbf{W}_{\mathrm{R}}^{\mathrm{O}} = [w_{\mathrm{R}oh}] \quad \text{(Output neuron weight mateix)} \tag{20}$$

$$z_{\mathrm{R}}^{\mathrm{H}} = f_{\mathrm{R}}\left(\mathbf{W}_{\mathrm{R}}^{\mathrm{H}} z_{\mathrm{R}}^{\mathrm{I}}\right) \tag{21}$$

$$z_{\mathrm{R}}^{\mathrm{O}} = f_{\mathrm{R}}\left(\mathbf{W}_{\mathrm{R}}^{\mathrm{O}} z_{\mathrm{R}}^{\mathrm{H}}\right) \tag{22}$$

$$f_{\mathrm{R}}(x) = \tanh(x) \tag{23}$$

where the thresholds are $w_{\mathrm{R}\ h\ 2I+1}$, $w_{\mathrm{R}\ h\ 2I+2}$, $w_{\mathrm{R}\ h\ 2H+1}$, and $w_{\mathrm{R}\ h\ 2H+2}$ with formal additional inputs $x_{2H+1} = 1$,

TABLE I
PARAMETERS IN THE NEURAL NETWORKS

	CVNN or RI-CVNN	RVNN or dual-RVNN
Number of input neurons	$I = 16$	$2I = 32$
Number of hidden neurons	$H = 25$	$2H = 50$
Number of output neurons	$O = 16$	$2O = 32$
Learning constant	$K = 0.01$	$K = 0.01$

$x_{2H+2} = 1$, $x_{2H+1} = 1$, and $x_{2H+2} = 1$. We employ the conventional error backpropagation learning. That is, we define an error function E_{R} for a set of input and output teacher signals $(\hat{z}_s^{\mathrm{I}}, \hat{z}_s^{\mathrm{O}})$ to obtain the learning dynamics as

$$E_{\mathrm{R}} \equiv \frac{1}{2}\sum_{s=1}^{S}\sum_{o=1}^{2O}\left|x_o(\hat{z}_{\mathrm{R}s}^{\mathrm{I}}) - \hat{x}_{os}\right|^2 \quad (= E) \tag{24}$$

$$w_{\mathrm{R}oh}^{\mathrm{new}} = w_{\mathrm{R}oh}^{\mathrm{old}} - K\frac{\partial E_{\mathrm{R}}}{\partial w_{\mathrm{R}oh}} \tag{25}$$

$$w_{\mathrm{R}ji}^{\mathrm{new}} = w_{\mathrm{R}hi}^{\mathrm{old}} - K\frac{\partial E_{\mathrm{R}}}{\partial w_{\mathrm{R}hi}} \tag{26}$$

$$\frac{\partial E_{\mathrm{R}}}{\partial w_{\mathrm{R}oh}} = \left(x_o - \hat{x}_o\right)\left(1 - x_o^2\right)x_h \tag{27}$$

$$\frac{\partial E_{\mathrm{R}}}{\partial w_{\mathrm{R}hi}} = \left(\sum_{o}^{2O}\left(x_o - \hat{x}_o\right)\left(1 - x_o^2\right)w_{oh}\right)\left(1 - x_h^2\right)x_i. \tag{28}$$

4) Dual Real-Valued Neural Networks for Real-Imaginary Separate Processing (Dual-RVNN): We consider another type of RVNN in which the real and imaginary parts of input signals are processed separately. It is an extension of dual univariate RVNN having single-layer structure. We may have a variety of ways of mixing and separating the real and imaginary variables in multiple-layer networks. With this network, we examine a completely separate case where the neurons in the real-part network have no connections to those in the imaginary-part network. The learning and processing dynamics are identical to those of the above RVNN except that the numbers of input terminals and output neurons are the same as the CVNNs for the respective real and imaginary networks.

III. EXPERIMENTS

A. Experimental Setup

1) Simulation Setup: Fig. 4 shows schematically how to observe the generalization characteristics of the networks. We conducted the learning process as follows. We chose the identity mapping as the task to be learned to show the network characteristics most clearly. That is, we take a set of input and output teacher signals as $\hat{z}_s^{\mathrm{I}} = \hat{z}_s^{\mathrm{O}}$ $(s = 1, 2, \ldots, S)$ with the following conditions. For a signal set showing high coherence, we choose its wavelength in such a manner that a unit wave spans just over the neural input terminals $i = 1, \ldots, I$, and discrete I points are fed to the network evenly with a constant interval in the unit wave. In more detail, we choose multiple amplitude values between 0 and 1 evenly for

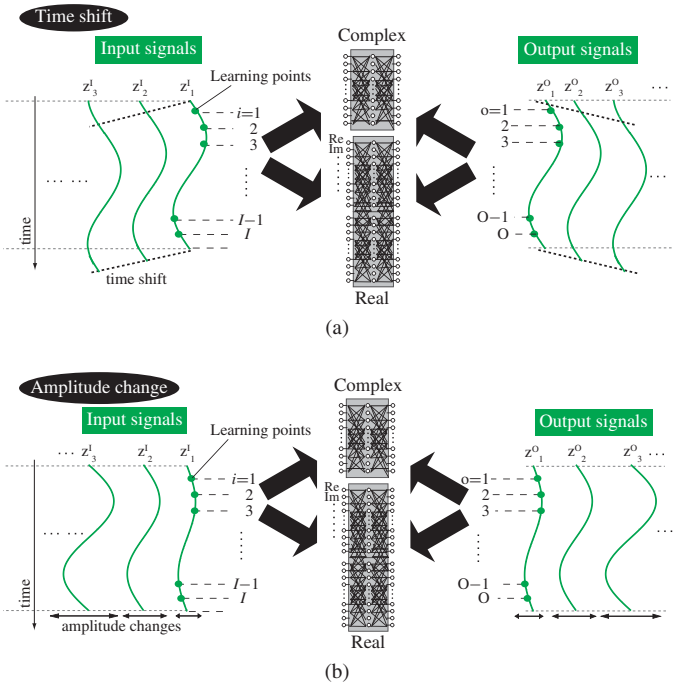

Fig. 4. Schematic diagrams showing how to feed signals to observe (a) time shift and (b) amplitude change generalizations.

$s_{\mathrm{A}} = 1, \ldots, S_{\mathrm{A}}$ teacher signals as well as multiple time shift amounts between 0 and half-wave duration (phase shift in a sinusoidal case between 0 to π) evenly for $s_{\mathrm{t}} = 1, \ldots, S_{\mathrm{t}}$ teachers. Consequently, we generate $S = S_{\mathrm{A}} \times S_{\mathrm{t}}$ points of discrete teacher–signal sets $\hat{z}_{is}$ $(s = 1, 2, \ldots, S)$ as

$$z_{is} \equiv \frac{s_{\mathrm{A}}}{S_{\mathrm{A}} + 1}\exp\left(\jmath\left(\frac{s_{\mathrm{t}}}{2S_{\mathrm{t}}} + \frac{i}{I}\right)2\pi\right). \tag{29}$$

Note that the wavelength and, thus, the signal frequency are unchanged. Fig. 1, referred to previously, includes the manner of the amplitude variation. We add WGN to the sinusoidal wave with various weightings. The noise power is adjusted depending on the signal power and the expected SNR, which is determined in each learning trial.

The dots on the continuous signals in Fig. 4 indicate the discrete teacher signal points $\hat{z}_{is}$. We observe the generalization characteristics by inputting signals other than the teachers and evaluate the output errors. Fig. 4(a) illustrates the observation of outputs when the input signal is shifted in time. The continuous time signal was generated by Lagrange interpolation. Fig. 4(b) shows the observation when the amplitude is changed. We combine the time shift and the amplitude change to evaluate the generalization. In the experiment below, $S_{\mathrm{A}} = 4$, $S_{\mathrm{t}} = 4$, and the neural network parameters are listed in Table I. The learning iteration is 3000.

2) Heterodyne Signal Experiment: We process a heterodyne signal observed in a sonar imaging system. The signal has a carrier of 100 kHz with thermal noise. It is converted into 100-Hz in-phase and quadrature-phase (IQ) intermediate-frequency (IF) signals through an IQ mixer. The imbalance of the IQ mixer is less than 0.3 dB in amplitude and 3° in phase,

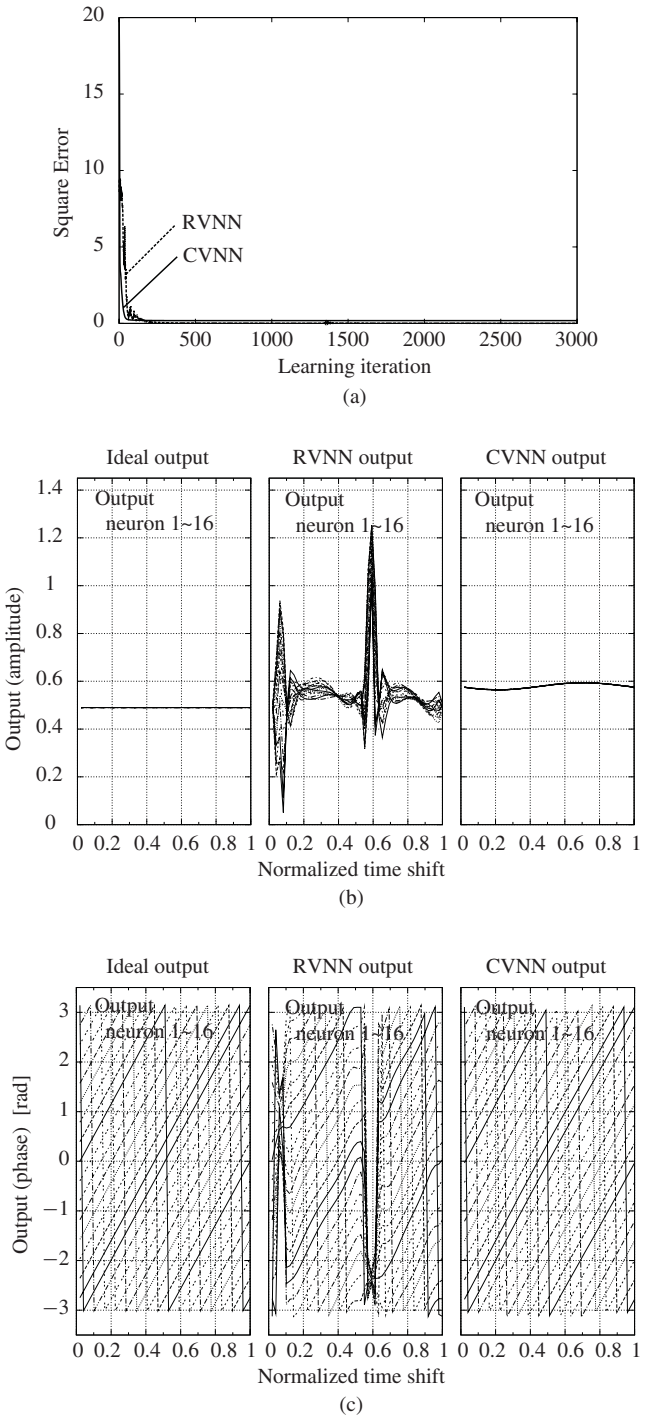

Fig. 5. Example of (a) learning curves and corresponding output (b) amplitude and (c) phase when the input signal gradually sifts in time in the real-valued and complex-valued neural networks (RVNN and CVNN) when no noise is added to sinusoidal signals ($SNR = \infty$).

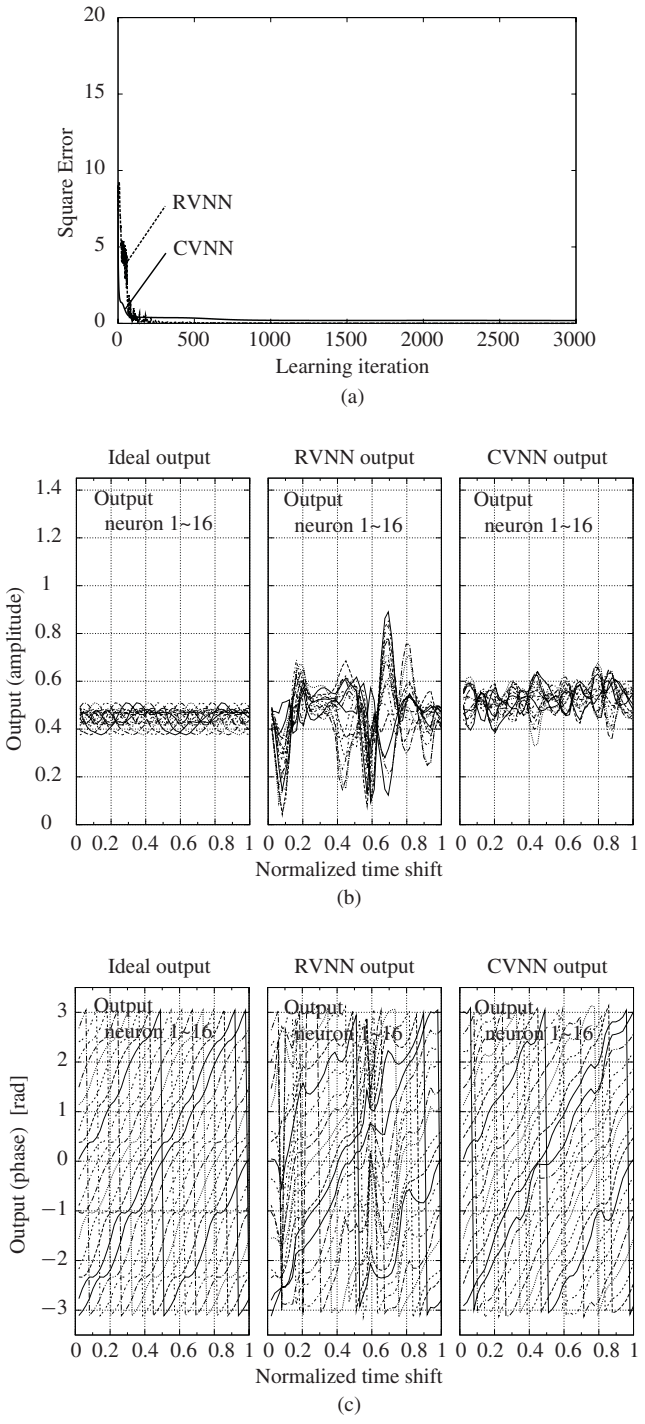

Fig. 6. Example of (a) learning curves and corresponding output (b) amplitude and (c) phase when the input signal gradually sifts in time in the real-valued and complex-valued neural networks (RVNN and CVNN) when white noise is added to sinusoidal signals ($SNR = 20$ dB).

which is common in this type of system. The IF signal is recorded by a personal computer through an analog/digital converter with a sampling frequency of 600 000 samples per second. We aim at appropriate interpolation of the signals in time and/or space domain for postprocessing to generate high-quality time-space images. When the 100-kHz carrier signal power changes, the SNR also changes for a constant noise power.

B. Results

1) Examples of Learning Curves and Output Signals for Inputs Having Various Coherence Degrees: Figs. 5–8 display typical examples of the learning curves and output signals of the CVNN and RVNN for a single learning trial when $SNR = \infty$, 20 dB, 10 dB, and 0 dB, respectively. Fig. 5(a) shows the learning curve when $SNR = \infty$, i.e., the signal is sinusoidal.

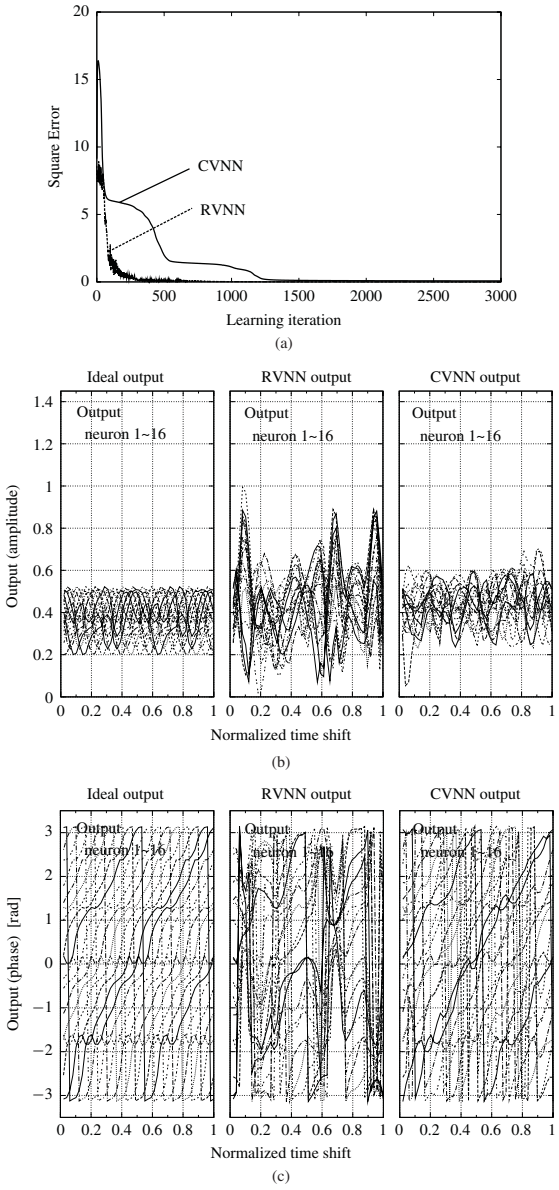

Fig. 7. Example of (a) learning curves and corresponding output (b) amplitude and (c) phase when the input signal gradually sifts in time in the real-valued and complex-valued neural networks (RVNN and CVNN) when white noise is added to sinusoidal signals ($SNR = 10$ dB).

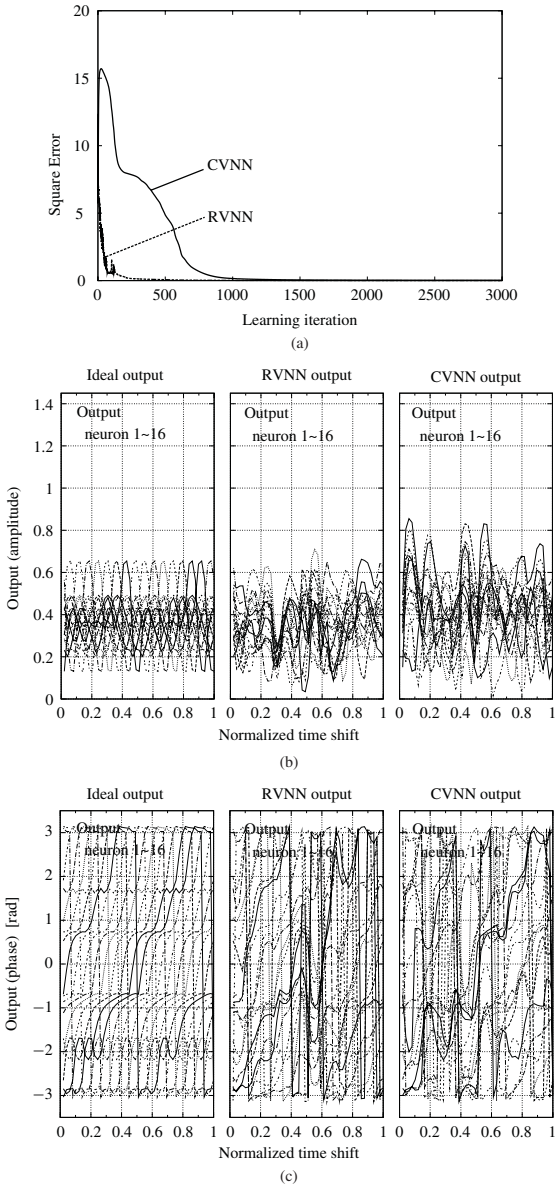

Fig. 8. Example of (a) learning curves and corresponding output (b) amplitude and (c) phase when the input signal gradually sifts in time in the real-valued and complex-valued neural networks (RVNN and CVNN) when white noise is added to sinusoidal signals ($SNR = 0$ dB).

We find that the learning is successfully completed for both CVNN and RVNN. The learning errors converge almost at zero, which means that there is only a slight residual error at the learning teacher points.

After the learning, we use other input signals to investigate the generalization. As mentioned previously, the wavelength is adjusted to span over the 16 neural input terminals. For example, we gradually move the input signal forward in time while keeping the amplitude unchanged at $a = 0.5$. Fig. 5(b) and (c) present the output amplitude and phase, respectively, showing from left-hand side to the right-hand side the ideal output of the identity mapping, the RVNN outputs, and CVNN outputs of the 16 output neurons. The horizontal axes present the time shift t normalized by the unit-wave duration.

In Fig. 5(b), we find that the output signals of the RVNN locally deviate greatly from the ideal ones. The learning points are plotted at $t = 0$ (no time shift), where the output amplitude is almost 0.5 for all the neurons. However, with the time course, the amplitude values fluctuate largely. Contrarily, the CVNN amplitude stays almost constant. At the learning point $t = 0$, the value is slightly larger than 0.5, corresponding to the slight nonzero value of the residual error in the learning curve.

In Fig. 5(c), the ideal output phase values on the left-hand side exhibit linear increase in time. In the RVNN case, though the phase values at $t = 0$ are the same as those of ideal outputs, the values sometimes swing strongly. In contrast, the CVNN output phase values increase in an orderly manner, which is almost identical with the ideal values. In summary,

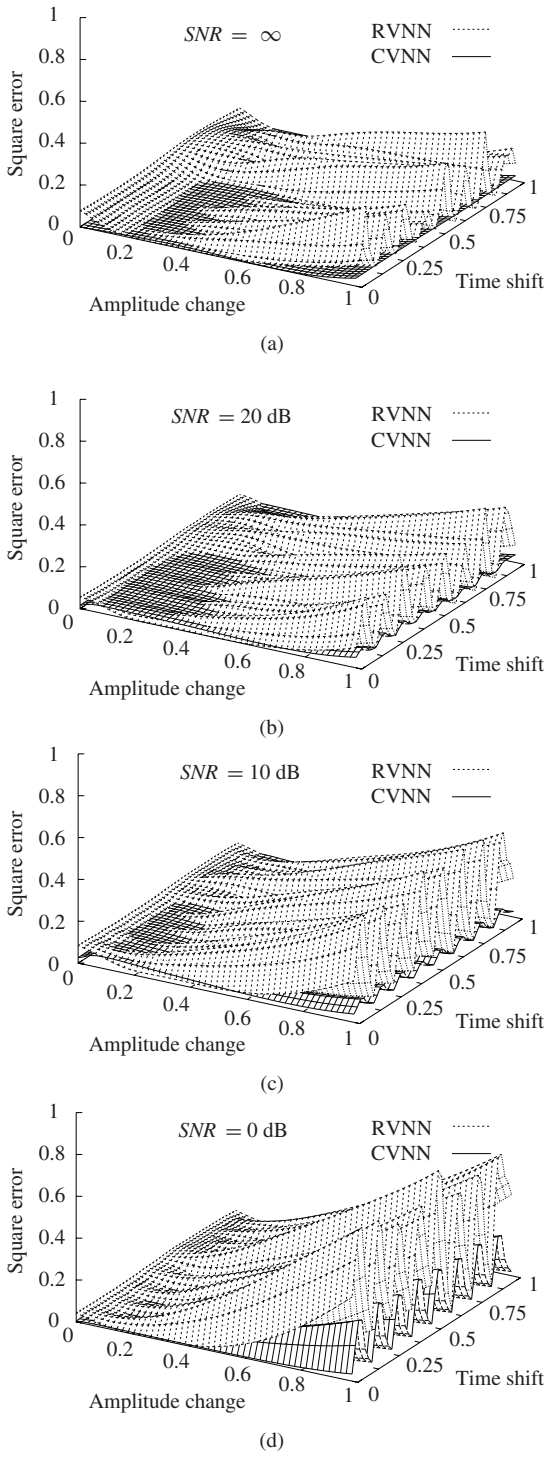

Fig. 9. Squared generalization errors averaged for 100 trials as functions of amplitude change and time shift for SNR = (a) ∞, (b) 20 dB, (c) 10 dB, and (d) 0 dB.

the CVNN presents much better generalization characteristics than the RVNN when the coherence is high, i.e., $SNR = \infty$.

Figs. 6–8 show the data for SNR = 20, 10, and 0 dB, respectively. As the coherence decreases, the generalization error increases. However, in any SNR case, both the amplitude and phase of the CVNN exhibit better generalization than the RVNN.

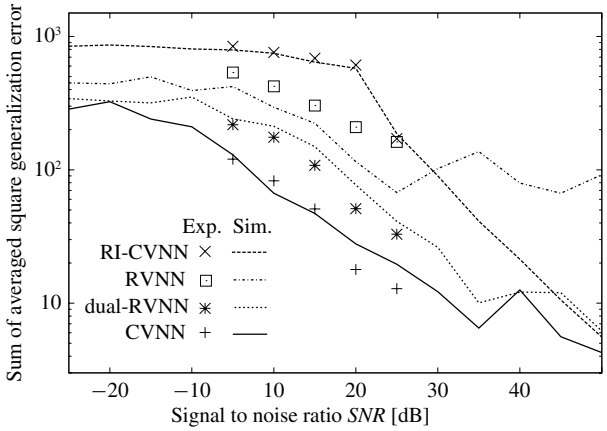

Fig. 10. Squared generalization errors summed up for all the sampling amplitude-time points shown in Fig. 9 versus signal SNR for the real-valued and complex-valued neural networks (CVNN, RI-CVNN, RVNN, and dual-RVNN, curves: simulations, marks: experiments).

The learning curves of the CVNN show that its learning speed depends on the signal coherence. The CVNN learning is completed fast when the coherence is high (about $SNR > 20$ dB), while it takes a longer time when the coherence is low ($SNR < 10$ dB). Contrarily, the learning speed of the RVNN is much less dependent on the coherence. We conducted simulations also for the RI-CVNN and the dual-RVNN. In the case of RI-CVNN, the convergence rate decreases largely in the low coherence region ($SNR < 5$ dB). In the dual-RVNN case, the learning curves are similar to those of RVNN.

2) Generalization Error and its Dependence on the Coherence: Here we present statistic results obtained by repeating the above simulations as well as the real-world experiment dealing with the heterodyne signals explained in Section III-A.

Fig. 9 is 3-D representation of the square errors as the average of 100 learning trials for various coherence degrees, namely, SNR = (a) ∞, (b) 20 dB, (c) 10 dB, and (d) 0 dB, as functions of time shift and amplitude change. The learning points exist at $t = 0$ and amplitude values of $a = 0.2$, 0.4, 0.6, and 0.8. At these points, we can find that the errors are very small, which corresponds to the almost zero residual errors in the learning curves. However, the errors at the teacher points for lower SNRs are obviously positive. This is because the learning error in some trials fails to converge at zero. As a whole, we notice in Fig. 9 that the generalization errors of the RVNN are larger than those of the CVNN, in particular in the cases of higher SNR. When SNR is low ($\sim$0 dB), the error of the CVNN also increases.

Fig. 10 compares quantitatively the generalization errors, summed up for all the sampling amplitude–phase points shown in Fig. 9, for the CVNN, RI-CVNN, RVNN, and dual-RVNN as functions of the coherence degree, i.e., SNR. The four curves show the results of the simulation, while the marks indicate experimental results. In all the neural network cases, the generalization error reduces according to the increase of the coherence (increase of SNR). The CVNN curve shows lower errors than other network curves over a wide range of SNR. The dual-RVNN also shows low errors, though at

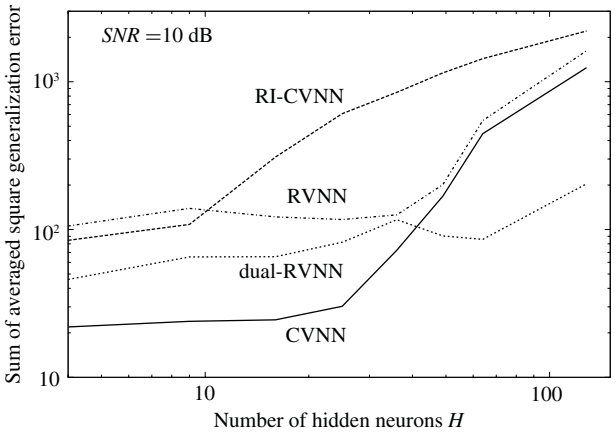

Fig. 11. Squared generalization errors summed up for all the sampling amplitude-time points shown in Fig. 9 versus the number of hidden neurons when $SNR = 10$ dB for the real-valued and complex-valued neural networks (CVNN, RI-CVNN, RVNN, and dual-RVNN).

the middle SNR ($SNR = -5$ to 15 dB) the value is 3–6 dB larger than for CVNN. The error of the simulated RVNN is about 2 dB larger than the dual-RVNN in the low and middle SNR range. The experimental results (marks) of the RVNN are a little larger. It is remarkable that, in the higher coherence region ($SNR > 10$ dB), the RVNN curve holds a floor at a nonnegligible level. The RI-CVNN shows a large generalization error in the low coherence region. This is not only because of the errors at non-teacher points but also because of the errors at teacher points. That is, the learning sometimes fails. In the high coherence region ($SNR > 20$ dB), however, the generalization error decreases and approaches the curves of the CVNN and dual-RVNN. In summary, we found that the four neural networks present different generalization characteristics among them. The experimental results have been found mostly close to the simulation results.

Fig. 11 presents the dependence of the generalization error on the number of hidden neurons H. The SNR is $SNR = 10$ dB where the coherence is high and we generally need only a small number of hidden neurons. The numbers of input terminals and output neurons are fixed as $I = 16$ and $O = 16$. The CVNN shows low generalization error over a wide region of H. However, when H is extremely large ($H > 50$), the error increases. The RVNN has larger error, but the increase tendency is almost the same as that of CVNN. The RI-CVNN has almost the same large error in small H region, and shows an earlier rise versus the increase of H. In contrast, dual-RVNN has only less increase, though the error level is not small.

IV. Discussion

The experimental results shown in Section III revealed the following characteristics of the networks.

1) Learning speed: the CVNN has a high learning speed for high coherence signals, but low speed for low coherence signals. The speed of the RVNN has a smaller dependence on the coherence.
2) Generalization error:
   a) The CVNN has 3–6 dB lower generalization error than dual-RVNN in the middle coherence region (-5 dB $< SNR < 20$ dB).
   b) The RVNN has a high error floor in the high coherence region ($SNR > 15$ dB).
   c) The RI-CVNN has higher error in the low coherence region ($SNR < 10$ dB).

These characteristics are explained by taking into consideration the degree of freedom in the learning process and the circularity of signals, as well as the neural dynamics below.

A. Degree of Freedom in Learning in CVNN, Dual-RVNN, and RVNN

What is the most specific nature of a complex number? As we focus on multiplication out of the four arithmetic operations of complex numbers, we can represent a complex number as a real 2×2 matrix. That is, with every complex number $c = a + \jmath b$, where a and b are real numbers, we associate a $\boldsymbol{C}$-linear transformation T_c of $z = x + \jmath y$ as

$$T_c : \boldsymbol{C} \to \boldsymbol{C}, \qquad z \mapsto cz = ax - by + \jmath(bx + ay). \quad (30)$$

If we identify $\boldsymbol{C}$ with $\boldsymbol{R}^2$ by

$$z = x + \jmath y \;=\; \begin{pmatrix} x \\ y \end{pmatrix} \quad (31)$$

it follows that:

$$T_c \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax - by \\ bx + ay \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (32)$$

In other words, the linear transformation T_c determined by $c = a + \jmath b$ is expressed by a matrix which means phase rotation and magnitude amplification or attenuation as

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = r \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad (33)$$

where $r \equiv \sqrt{a^2 + b^2}$ and $\theta \equiv \arctan b/a$ denote amplification or attenuation of amplitude and rotation angle applied to the complex signal z, respectively.

Let us consider how the above feature of the complex number emerges in neural dynamics [21]. Assume a very simple case shown in Fig. 12(a) where we have a single-layer two-input two-output feedforward neural network in the real number domain. For simplicity, we omit possible nonlinearity at the neurons, i.e., the activation function is the identity function. The task for the network here is to realize a mapping that transforms an input $\boldsymbol{x}^{\mathbf{IN}}$ to an output $\boldsymbol{x}^{\mathbf{OUT}}$ through supervised learning that adjusts the synaptic weights w_{ji}. Simply, we have only a single teacher pair of input and output signals here. Then the general input–output relationship is described by using four real numbers a, b, c, and d as

$$\begin{pmatrix} x_1^{\mathbf{OUT}} \\ x_2^{\mathbf{OUT}} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1^{\mathbf{IN}} \\ x_2^{\mathbf{IN}} \end{pmatrix}. \quad (34)$$

In the present case, we have a variety of possible mappings realized by the learning because the number of parameters to

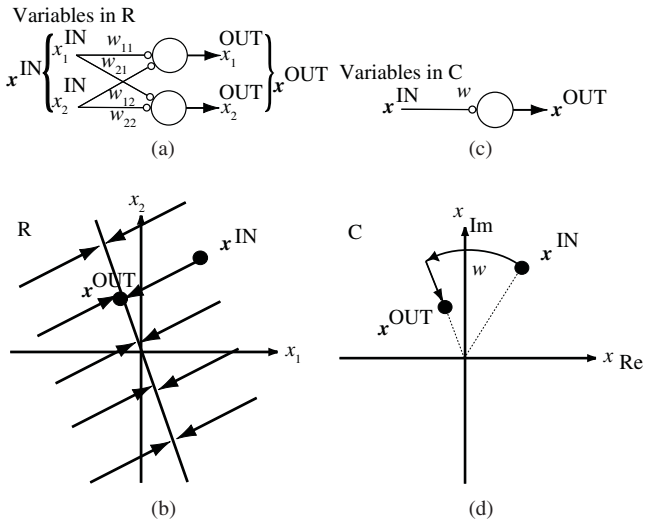

Fig. 12. Simple single-layer neural network to learn a task that maps $\boldsymbol{x}^{\mathbf{IN}}$ to $\boldsymbol{x}^{\mathbf{OUT}}$. (a) Real-valued bivariate network structure in $\boldsymbol{R}^2$. (b) Possible but degenerate solution that is often not useful. (c) Complex-valued network structure in $\boldsymbol{C}$. (d) Solution obtained in this small dgree-of-freedom case [21].

be determined is larger than the number of conditions, i.e., the learning task is an ill-posed problem. The functional difference among the possible mapping emerges as the difference in the generalization characteristics. For example, learning can result in a degenerate mapping shown in Fig. 12(c), which is often not useful in practice. Such degeneracy is suggested in the present simulation in Fig. 5(b) in the RVNN output as the bottleneck-shaped forced consistency of the amplitude curves.

In parallel, let us consider the learning of the mapping in the complex domain, which transforms a complex value $\boldsymbol{x}^{\mathbf{IN}} = (x_{\mathrm{Re}}^{\mathbf{IN}}, x_{\mathrm{Im}}^{\mathbf{IN}})$ to another complex value $\boldsymbol{x}^{\mathbf{OUT}} = (x_{\mathrm{Re}}^{\mathbf{OUT}}, x_{\mathrm{Im}}^{\mathbf{OUT}})$. Fig. 12(c) shows a complex-valued network, where the weight is a single complex value $w = |w| \exp(\jmath\theta)$. The situation is expressed just like in (34) with the constraint (33) as

$$\begin{pmatrix} x_{\mathrm{Re}}^{\mathbf{OUT}} \\ x_{\mathrm{Im}}^{\mathbf{OUT}} \end{pmatrix} = \begin{pmatrix} |w|\cos\theta & -|w|\sin\theta \\ |w|\sin\theta & |w|\cos\theta \end{pmatrix} \begin{pmatrix} x_{\mathrm{Re}}^{\mathbf{IN}} \\ x_{\mathrm{Im}}^{\mathbf{IN}} \end{pmatrix}. \tag{35}$$

The degree of freedom is reduced, and the arbitrariness of the solution is also reduced. Fig. 12(d) illustrates the result of the learning. The mapping is a combination of phase rotation and amplitude attenuation. This property can be a great advantage when we deal with coherent wave information such as electromagnetic waves, lightwaves, sonic waves, and electron waves. This property can be named *circularity* in the neural dynamics in analogy to the signal circularity for high coherence.

B. Circularity

The circularity of the signals to be processed is also an important factor. To deepen the discussion, we refer to the wide sense linear (or widely linear: WL) systems which introduce conjugate signals in addition to direct complex signals [23], [24]. WL systems well learn complex data distributed anisotropically in the complex plane, i.e., noncircular data. For example, it is useful for predicting wind strength and direction by assuming the axes of the complex number plane represent north, south, east, and west, and the distance from the origin expresses the strength. Augmented CVNNs have been proposed in such a context [25]. Wind has high anisotropy in general. The augmented complex-valued networks do not lead to the reduction of the degree of freedom. The degree is the same as that of real-valued networks, resulting in a dynamics similar to that of real-valued ones [26].

Fig. 13 is a conceptual illustration showing the bases of the respective networks. The number of bases of the augmented complex networks turns back to that of the real-valued networks, and its dynamics approaches that of real networks. This situation is analogous to the fact that the combination of positive and negative frequency spectra generates almost real-valued signals. In other words, if we compare the relationship to polarization of lightwaves, we come to the following conclusion. CVNNs deal with only right- or left-handed circular polarized light, which are suitable for circular signal processing. Note that the signal in total can be out of complete circularity, but only each frequency component has circularity. Since any waveform can be synthesized by sinusoidal components through Fourier synthesis, the signals that the complex networks can deal with are not limited to completely coherent signals. In contrast, the augmented complex-valued networks deal with both the right- and left-handed circular polarized light. They are more flexible because of the larger degree of freedom, which is too much for circular signals. Dual univariate networks have the same degree of freedom, though in this case the bases are linear polarization corresponding to the real and imaginary parts instead of the right- and left-handed circular bases in the augmented networks. In this manner, they are similar to each other.

Consequently, CVNNs are suitable for processing analytic signals, which consist of a real component and its consistent imaginary component that has the same amplitude but a phase shifted by 90°. Analytic signals are essentially circular, existing widely in electronics, for example, at the output of heterodyne or homodyne mixers and at the output of digital signal processing using the Hilbert transform. Complex-valued networks have the ability to process such analytic signals appropriately.

Noncircular signals are observed in wind processing, but less in electronics dealing with electromagnetic wave and related time-sequential signals. The reason lies in the fact that phase does not have any meaning in its absolute value, but in its difference from a certain reference. However, a few exceptions may exist. For example, signals in offset quadrature phase shift keying modulation may include nonnegligible noncircularity generated by highly unbalanced electronics.

C. Differences in Learning Speed and Generalization Errors

Based on the considerations on the degree of freedom in learning in Section IV-A and the circularity in signals and network dynamics in discussed in Section IV-B, we can summarize the experimental results as follows.

1) The CVNN shows quick learning and a small generalization error in the middle and high coherence region because of its smaller degree of freedom and the circular dynamics. In the case of an extremely large number

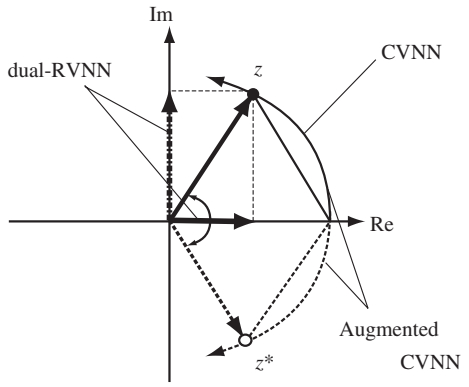

Fig. 13. Conceptual illustration of the relationship between bases in the respective neural networks to deal with complex signal z.

of hidden neurons, however, it cannot maintain the small degree of freedom, resulting in a larger generalization error. This phenomenon originates also from the circularity, i.e., the origin-point symmetry, which is similar to the degeneration of attractors in complex-valued associative memories. Because of the circularity, connections can remain excessively to express coherent signals.

2) The dual-RVNN has a larger degree of freedom and a larger generalization error than the CVNN. In this interpolation task, however, the network can get a solution by looking up the unit-wavelength past signal (or half-wavelength past signal with reversing its sign) when the signal has very high coherence and therefore the time delay directly corresponds to the phase delay. At other points, the neural output has no correlation with the input, for which the neural connections decay. In other words, the signal degeneration does not occur on the real–imaginary bases. Then the generalization error stays almost constant against the increase of the number of hidden neurons.
3) The RVNN has further larger degrees of freedom, resulting in a larger generalization error.
4) The RI-CVNN includes two properties inconsistent with each other, namely, the circular process of the weight multiplication and the anisotropy of activation function, which lead to a large generalization error. The error is enhanced when the signal has smaller relationship between the time delay and the phase shift because of the noise. However, for high coherence signals, the anisotropy works to reduce the harmful excess connections just like in the RVNN case, and makes the generalization error small.

V. Conclusion

This paper investigated numerically the generalization characteristics in the feedforward CVNN and RVNN with their variations. We focused on their dependence on the coherence degree. We compared CVNN (with amplitude–phase-type activation function), RI-CVNN (with real–imaginary-type activation function), RVNN (bivariate), and dual-RVNN (dual univariate) in a case where the networks approximate functions. Computer experiments demonstrated that the CVNN exhibits smaller generalization error in particular for signals having high coherence. This result is attributed to the smaller degree of freedom and the coherence-compatible circularity in the neural dynamics. We discussed the relationship between the generalization characteristics and the dynamics of the various CVNNs and RVNNs in general by focusing on the degree of freedom and the circularity.

Acknowledgment

The authors would like to acknowledge the insightful suggestions of the anonymous reviewers.

References

[1] D. L. Birx and S. J. Pipenberg, "A complex mapping network for phase sensitive classification," *IEEE Trans. Neural Netw.*, vol. 4, no. 1, pp. 127–135, Jan. 1993.
[2] T. Hara and A. Hirose, "Plastic mine detecting radar system using complex-valued self-organizing map that deals with multiple-frequency interferometric images," *Neural Netw.*, vol. 17, nos. 8–9, pp. 1201–1210, 2004.
[3] S. Masuyama and A. Hirose, "Walled LTSA array for rapid, high spatial resolution, and phase-sensitive imaging to visualize plastic landmines," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 8, pp. 2536–2543, Aug. 2007.
[4] I. Aizenberg, D. V. Paliy, J. M. Zurada, and J. T. Astola, "Blur identification by multilayer neural network based on multivalued neurons," *IEEE Trans. Neural Netw.*, vol. 19, no. 5, pp. 883–898, May 2008.
[5] G. Tanaka and K. Aihara, "Complex-valued multistate associative memory with nonlinear multilevel functions for gray-level image reconstruction," *IEEE Trans. Neural Netw.*, vol. 20, no. 9, pp. 1463–1473, Sep. 2009.
[6] Y. Zhang and Y. Ma, "CGHA for principal component extraction in the complex domain," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 1031–1036, Sep. 1997.
[7] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency-domain blind source separation," *IEICE Trans. Fundam. Electron., Commun., Comput. Sci.*, vol. E86A, no. 3, pp. 590–596, 2003.
[8] S. Goh and D. P. Mandic, "Nonlinear adaptive prediction of complex-valued signals by complex-valued PRNN," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1827–1836, May 2005.
[9] A. Hirose and R. Eckmiller, "Behavior control of coherent-type neural networks by carrier-frequency modulation," *IEEE Trans. Neural Netw.*, vol. 7, no. 4, pp. 1032–1034, Jul. 1996.
[10] A. B. Suksmono and A. Hirose, "Beamforming of ultrawideband pulses by a complex-valued spatio-temporal multilayer neural network," *Int. J. Neural Syst.*, vol. 15, no. 1, pp. 1–7, 2005.
[11] A. Hirose, "Applications of complex-valued neural networks to coherent optical computing using phase-sensitive detection scheme," *Inf. Sci. - Appl.*, vol. 2, no. 2, pp. 103–117, Sep. 1994.
[12] S. Kawata and A. Hirose, "A coherent optical neural network that learns desirable phase values in frequency domain by using multiple optical-path differences," *Opt. Lett.*, vol. 28, no. 24, pp. 2524–2526, 2003.
[13] A. Hirose, T. Higo, and K. Tanizawa, "Efficient generation of holographic movies with frame interpolation using a coherent neural network," *IEICE Electron. Exp.*, vol. 3, no. 19, pp. 417–423, 2006.
[14] Y. Nakamiya, M. Kinjo, O. Takahashi, S. Sato, and K. Nakajima, "Quantum neural network composed of Kane's qubits," *Jpn. J. Appl. Phys.*, vol. 45, no. 10A, pp. 8030–8034, Oct. 2006.
[15] A. Hirose, Y. Asano, and T. Hamano, "Developmental learning with behavioral mode tuning by carrier-frequency modulation in coherent neural networks," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1532–1543, Nov. 2006.
[16] H. Leung and S. Haykin, "The complex backpropagation algorithm," *IEEE Trans. Signal Process.*, vol. 39, no. 9, pp. 2101–2104, Sep. 1991.
[17] N. Benvenuto and F. Piazza, "On the complex backpropagation algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 967–969, Apr. 1992.

[18] T. Nitta, "An extension of the back-propagation algorithm to complex numbers," *Neural Netw.*, vol. 10, no. 8, pp. 1391–1415, Nov. 1997.
[19] A. Hirose, "Continuous complex-valued back-propagation learning," *Electron. Lett.*, vol. 28, no. 20, pp. 1854–1855, 1992.
[20] G. M. Georgiou and C. Koutsougeras, "Complex domain backpropagation," *IEEE Trans. Circuits Syst. II*, vol. 39, no. 5, pp. 330–334, May 1992.
[21] A. Hirose, "Nature of complex number and complex-valued neural networks," *Front. Electr. Electron. Eng. China*, vol. 6, no. 1, pp. 171–180, 2011.
[22] A. Hirose, *Complex-Valued Neural Networks*. New York: Springer-Verlag, 2006.
[23] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.
[24] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters – Noncircularity, Widely Linear and Neural Models*. New York: Wiley, Apr. 2009.
[25] Y. Xia, B. Jelfs, M. M. Van Hulle, J. C. Principe, and D. P. Mandic, "An augumented echo state network for nonlinear adaptive filetering of complex noncircular signals," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 74–83, Jan. 2011.
[26] D. P. Mandic, S. Still, and S. C. Douglas, "Duality between widely linear and dual channel adaptive filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Taipei, Taiwan, Apr. 2009, pp. 1729–1732.

Akira Hirose (S'85–M'88–SM'08) received the Ph.D. degree in electronic engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He joined the Research Center for Advanced Science and Technology (RCAST), University of Tokyo, in 1987, as a Research Associate. In 1991, he became an Instructor at RCAST. From 1993 to 1995, he joined the Institute for Neuroinformatics, University of Bonn, Bonn, Germany. He is currently a Professor with the Department of Electrical Engineering and Information Systems, University of Tokyo. His current research interests include neural networks and wireless electronics.

Prof. Hirose is a senior member of the Society of Information and Communication Engineers (IEICE) and a member of the Japanese Neural Networks Society (JNNS). He served as the Chair of the Neurocomputing Technical Group, Institute of Electronics, IEICE, and is currently the Vice-President of JNNS, a member of the IEEE Computational Intelligence Society (CIS) Neural Networks Technical Committee, a Steering Committee Member of the IEEE CIS Japan Chapter, and a Governing Board Member of the Asia-Pacific Neural Network Assembly. He serves as Editor-in-Chief of the IEICE *Transactions on Electronics*, and as an Associate Editor of journals such as the IEEE TRANSACTIONS ON NEURAL NETWORKS and the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER.

Shotaro Yoshida was born in Oita, Japan, on November 5, 1988. He received the B.S. degree in electronic and information engineering from the University of Tokyo, Tokyo, Japan, in 2011. He is currently pursuing the M.S. degree.

His current research interests include neural networks and bioelectronics.