Stochastic Gradient-Adaptive Complex-Valued Nonlinear Neural Adaptive Filters With a Gradient-Adaptive Step Size

Su Lee Goh and Danilo P. Mandic

Abstract—A class of variable step-size learning algorithms for complex-valued nonlinear adaptive finite impulse response (FIR) filters is proposed. To achieve this, first a general complex-valued nonlinear gradient-descent (CNGD) algorithm with a fully complex nonlinear activation function is derived. To improve the convergence and robustness of CNGD, we further introduce a gradient-adaptive step size to give a class of variable step-size CNGD (VSCNGD) algorithms. The analysis and simulations show the proposed class of algorithms exhibiting fast convergence and being able to track nonlinear and nonstationary complex-valued signals. To support the derivation, an analysis of stability and computational complexity of the proposed algorithms is provided. Simulations on colored, nonlinear, and real-world complex-valued signals support the analysis.

Index Terms—Complex nonlinear adaptive filters, complex-valued nonlinear gradient descent (CNGD), finite impulse response (FIR), variable step size (VS).

I. INTRODUCTION

Real-valued finite impulse response (FIR) adaptive filters are a *de-facto* standard in online signal processing and have found their applications in a variety of disciplines, including acoustics, communications, and seismology [1]. The least mean square (LMS) algorithm is the most frequently used algorithm for training of such filters [2]. Despite its robustness, this algorithm is relatively slow at converging, and a number of its variants has been proposed [3], [4]. It also has been recognized that a fixed step size, which governs the speed of convergence and steady-state error of stochastic gradient algorithms, is not an optimal choice for nonstationary environments, and to that end, various variable step-size least mean square (VSLMS) algorithms have been developed [3], [5], [6]. Ideally, the use of a variable step size (VS) should also help to circumvent the tradeoff between fast convergence and steady-state misadjustment, commonly experienced with the fixed step-size LMS.

Variable step-sizes (VSs) can either be made to adapt in some heuristic manner [7] or they can be gradient-adaptive. Benveniste *et al.* [5] proposed a rigorously derived general adaptive step-size algorithm based on the gradient of the instantaneous squared error with respect to the step size. Benveniste's algorithm, in fact, performs time-varying low-pass filtering of the noisy instantaneous gradients in the update of such an adaptive step size. This algorithm was derived without making the usual independence assumptions [5], which results in much improved performance, but also significantly increases computational complexity as compared to standard LMS. Attempts to reduce the computational complexity of this algorithm include the modifications by Mathews and Xie [3] and Ang and Farhang-Boroujeny [6].

For linear complex-valued adaptive filtering, the complex least mean square (CLMS) algorithm was developed [1], and subsequently, nonlinear complex-valued adaptive filters have been introduced, mostly realized as neural networks [2], [8], [9]. Despite emerging applications of complex-valued nonlinear adaptive filtering, there are relatively few results on advanced algorithms for these filters.

Our aim here is to extend the class of real-valued gradient-adaptive step-size algorithms to the case of complex-valued nonlinear adaptive filters. This is not trivial due to several open problems in the design of complex-valued nonlinear adaptive filters, such as of the choice of complex nonlinearities and their singular points [10], existence of Cauchy–Riemann conditions, and calculation of complex gradients. Complex-valued elementary transcendental functions (such as $\tanh$) proposed by Kim and Adali [11] are identified as those that can provide adequate nonlinear activation functions for this class of nonlinear filters. For generality, we focus on filters with a "fully" complex nonlinear activation function (AF)[1] of a neuron. This way, the Cauchy–Riemann[2] equations are satisfied (except for a set of isolated points) which makes it possible to use gradient descent in the update of the variable step size. A general class of gradient-adaptive step-size algorithms is then derived rigorously (termed VSCNGD1). This is followed by two other algorithms, based on VSCNGD1, but with relaxed computational complexity. The tradeoff between computational complexity and performance is analyzed, which is followed by stability analysis. The analysis is supported by simulations on complex-valued linear, nonlinear, and real-world signals.

II. CLASS OF VSCNGD ALGORITHMS

Following the standard stochastic gradient approach, the weight update of the standard CNGD algorithm for a nonlinear FIR adaptive filter (shown in Fig. 1) can be expressed as [12]

$$\mathbf{w}(k + 1) = \mathbf{w}(k) + \eta e(k)[\Phi'(\text{net}(k))]^*\mathbf{x}^*(k) \quad (1)$$

$$\text{net}(k) = \mathbf{x}^T(k)\mathbf{w}(k) \quad (2)$$

where $e(k) = d(k) - \Phi(\text{net}(k))$ denotes the instantaneous error at the output of the filter at the time instant k, $d(k)$ is the desired signal, $\mathbf{x}(k) = [x(k-1), \ldots, x(k-M+1)]^T$ is the input signal vector, M is the length of the filter, $(\cdot)^T$ is the vector transpose operator, $(\cdot)^*$ denotes the complex conjugate operator, $\mathbf{w}(k) = [w_1(k), \ldots, w_M(k)]^T$ is the filter coefficient vector, whereas symbol Φ denotes a general fully complex nonlinear activation function.[3] For simplicity, we state that

$$\Phi(\text{net}(k)) = \Phi^r(\text{net}(k)) + j\Phi^i(\text{net}(k)) = u(k) + jv(k). \quad (3)$$

The parameter η in (1) is the step size and is critical to the convergence of the learning algorithms.

---

[1]In a frequently used split-complex AF, the real and imaginary components of the input signal x are separated and fed through the real-valued AF $f_R(x) = f_I(x), x \in \mathbb{R}$. A split-complex AF is, therefore, given as $f(x) = f_R(Re(x)) + j f_I(Im(x))$, hence these functions are not analytic. The nonlinearity within the complex AF must be analytic and bounded almost everywhere in the complex domain, $\mathbb{C}$ [11].

[2]Cauchy–Riemann equations state that the partial derivatives of a function $f(z) = u(x, y) + jv(x, y)$ along the real and imaginary axes should be equal as follows: $f'(z) = (\partial u/\partial x) + j(\partial v/\partial x) = (\partial v/\partial y) - j(\partial u/\partial y)$. This way $(\partial u/\partial x) = (\partial v/\partial y)$ and $(\partial v/\partial x) = -(\partial u/\partial y)$.

[3]For a comprehensive account of complex-valued activation function please refer to [10].
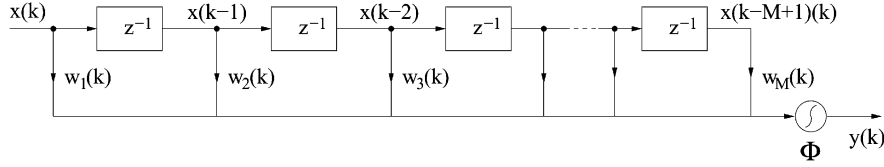
Fig. 1. Complex-valued nonlinear adaptive FIR filter (dynamical perceptron).

A. VSCNGD1 Algorithm

To make the adaptive filter from Fig. 1 capable of processing nonlinear and nonstationary signals, we propose to make the step size η in (1) gradient-adaptive, as [13]

$$\eta(k) = \eta(k-1) - \rho \nabla_\eta E(k)_{|\eta=\eta(k-1)} \tag{4}$$

where $E(k) = (1/2)e(k)e^*(k) = (1/2)|e(k)|^2$ is the cost function and ρ is some small positive constant. The real and imaginary component of the complex error $e(k)$ can be calculated from

$$e^r(k) = d^r(k) - u(k) \quad e^i(k) = d^i(k) - v(k) \tag{5}$$

which gives

$$\frac{\partial e^r(k)}{\partial u(k)} = -1 \quad \frac{\partial e^i(k)}{\partial v(k)} = -1. \tag{6}$$

The gradient $\nabla_\eta E(k)$ from (4) can now be evaluated as

$$\nabla_\eta E(k) = \frac{1}{2}\left[e(k)\frac{\partial e^*(k)}{\partial \eta(k-1)} + e^*(k)\frac{\partial e(k)}{\partial \eta(k-1)} \right]. \tag{7}$$

To derive the proposed algorithms, the main issue is the calculation of the partial derivatives $(\partial e^*(k))/(\partial \eta(k-1))$ and $(\partial e(k))/(\partial \eta(k-1))$ from (7). To calculate $(\partial e^*(k))/(\partial \eta(k-1))$, notice that

$$\frac{\partial e^*(k)}{\partial \eta(k-1)} = \frac{\partial e^r(k)}{\partial \eta(k-1)} - j\frac{\partial e^i(k)}{\partial \eta(k-1)}. \tag{8}$$

Using the Cauchy–Riemann conditions, and the previous partial derivatives from (6), we obtain the partial gradient terms from (8) as

$$\frac{\partial e^r(k)}{\partial \eta(k-1)} = \frac{\partial e^r(k)}{\partial u(k)}\left[\frac{\partial u(k)}{\partial \mathrm{net}^r(k)}\left(\frac{\partial \mathrm{net}^r(k)}{\partial \mathbf{w}^r(k)}\frac{\partial \mathbf{w}^r(k)}{\partial \eta(k-1)} \right. \right.$$
$$+ \frac{\partial \mathrm{net}^r(k)}{\partial \mathbf{w}^i(k)}\frac{\partial \mathbf{w}^i(k)}{\partial \eta(k-1)} \Bigg)$$
$$+ \frac{\partial u(k)}{\partial \mathrm{net}^i(k)}\left(\frac{\partial \mathrm{net}^i(k)}{\partial \mathbf{w}^r(k)}\frac{\partial \mathbf{w}^r(k)}{\partial \eta(k-1)} \right.$$
$$\left. \left. + \frac{\partial \mathrm{net}^i(k)}{\partial \mathbf{w}^i(k)}\frac{\partial \mathbf{w}^i(k)}{\partial \eta(k-1)} \right) \right]$$
$$= \frac{\partial u(k)}{\partial \mathrm{net}^r(k)}\left(-\mathbf{x}^r(k)\frac{\partial \mathbf{w}^r(k)}{\partial \eta(k-1)} + \mathbf{x}^i(k)\frac{\partial \mathbf{w}^i(k)}{\partial \eta(k-1)} \right)$$
$$+ \frac{\partial u(k)}{\partial \mathrm{net}^i(k)}\left(-\mathbf{x}^i(k)\frac{\partial \mathbf{w}^r(k)}{\partial \eta(k-1)} - \mathbf{x}^r(k)\frac{\partial \mathbf{w}^i(k)}{\partial \eta(k-1)} \right) \tag{9}$$

and

$$\frac{\partial e^i(k)}{\partial \eta(k-1)} = \frac{\partial e^i(k)}{\partial v(k)}\left[\frac{\partial v(k)}{\partial \mathrm{net}^r(k)}\left(\frac{\partial \mathrm{net}^r(k)}{\partial \mathbf{w}^r(k)}\frac{\partial \mathbf{w}^r(k)}{\partial \eta(k-1)} \right. \right.$$
$$+ \frac{\partial \mathrm{net}^r(k)}{\partial \mathbf{w}^i(k)}\frac{\partial \mathbf{w}^i(k)}{\partial \eta(k-1)} \Bigg)$$
$$+ \frac{\partial v(k)}{\partial \mathrm{net}^i(k)}\left(\frac{\partial \mathrm{net}^i(k)}{\partial \mathbf{w}^r(k)}\frac{\partial \mathbf{w}^r(k)}{\partial \eta(k-1)} \right.$$
$$\left. \left. + \frac{\partial \mathrm{net}^i(k)}{\partial \mathbf{w}^i(k)}\frac{\partial \mathbf{w}^i(k)}{\partial \eta(k-1)} \right) \right]$$
$$= \frac{\partial v(k)}{\partial \mathrm{net}^r(k)}\left(-\mathbf{x}^r(k)\frac{\partial \mathbf{w}^r(k)}{\partial \eta(k-1)} + \mathbf{x}^i(k)\frac{\partial \mathbf{w}^i(k)}{\partial \eta(k-1)} \right)$$
$$+ \frac{\partial v(k)}{\partial \mathrm{net}^i(k)}\left(-\mathbf{x}^i(k)\frac{\partial \mathbf{w}^r(k)}{\partial \eta(k-1)} - \mathbf{x}^r(k)\frac{\partial \mathbf{w}^i(k)}{\partial \eta(k-1)} \right). \tag{10}$$

Finally, to calculate the gradient of the cost function $E(k)$ with respect to the step size η, combine (9) and (10) to yield

$$\frac{\partial e^*(k)}{\partial \eta(k-1)} = \frac{\partial u(k)}{\partial \mathrm{net}^r(k)}\left[\frac{\partial \mathbf{w}^r(k)}{\partial \eta(k-1)}\left(-\mathbf{x}^r(k) + j\mathbf{x}^i(k) \right) \right.$$
$$\left. + \frac{\partial \mathbf{w}^i(k)}{\partial \eta(k-1)}\left(j\mathbf{x}^r(k) + \mathbf{x}^i(k) \right) \right]$$
$$+ \frac{\partial u(k)}{\partial \mathrm{net}^i(k)}\left[\frac{\partial \mathbf{w}^r(k)}{\partial \eta(k-1)}\left(-j\mathbf{x}^r(k) - \mathbf{x}^i(k) \right) \right.$$
$$\left. + \frac{\partial \mathbf{w}^i(k)}{\partial \eta(k-1)}\left(-\mathbf{x}^r(k) + j\mathbf{x}^i(k) \right) \right]$$
$$= -\left[\frac{\partial u(k)}{\partial \mathrm{net}^r(k)}\left(\frac{\partial \mathbf{w}^r(k)}{\partial \eta(k-1)} - j\frac{\partial \mathbf{w}^i(k)}{\partial \eta(k-1)} \right) \right.$$
$$\left. + \frac{\partial u(k)}{\partial \mathrm{net}^r(k)}\left(j\frac{\partial \mathbf{w}^r(k)}{\partial \eta(k-1)} + \frac{\partial \mathbf{w}^i(k)}{\partial \eta(k-1)} \right) \right]$$
$$\times \mathbf{x}^*(k)$$
$$= -\mathbf{x}^H(k)\left\{ \Phi'(\mathrm{net}(k)) \right\}^* \frac{\partial \mathbf{w}^*(k)}{\partial \eta(k-1)}. \tag{11}$$

Similarly to the calculation of $(\partial e^*(k))/(\partial \eta(k-1))$, the second gradient term in (7) becomes

$$\frac{\partial e(k)}{\partial \eta(k-1)} = -\mathbf{x}^T(k)\Phi'(\mathrm{net}(k))\frac{\partial \mathbf{w}(k)}{\partial \eta(k-1)}. \tag{12}$$

TABLE I
GRADIENTS OF THE PROPOSED VSCNGD ALGORITHMS AND THE ASSOCIATED COMPLEXITY

Algorithms	Variable step-size update	No. of multiplications
VSCNGD1	$\boldsymbol{\psi}(k) = \boldsymbol{\psi}(k-1)\left[\mathbf{I} - \eta(k-1)\,\lvert\Phi'(k-1)\rvert^2\,\mathbf{x}^*(k-1)\mathbf{x}^T(k-1)\right.$ $\left. + \eta(k-1)e(k-1)\left\{\Phi''(k-1)\right\}^*\mathbf{x}^*(k-1)\mathbf{x}^H(k-1)\right]$ $+ e(k-1)\left\{\Phi'(k-1)\right\}^*\mathbf{x}^*(k-1)$	16
VSCNGD2	$\boldsymbol{\psi}(k) = \alpha\boldsymbol{\psi}(k-1) + e(k-1)\left\{\Phi'(k-1)\right\}^*\mathbf{x}^*(k-1)$	9
VSCNGD3	$\boldsymbol{\psi}(k) = e(k-1)\left\{\Phi'(k-1)\right\}^*x^*(k-1)$	8

For simplicity, let us denote $\Phi(\mathbf{x}^T(k)\mathbf{w}(k)) = \Phi(k)$ and the weight gradient from (12) as $\boldsymbol{\psi}(k) = \partial\mathbf{w}(k)/\partial\eta_{\vert\eta=\eta(k-1)}$. Thus, from (1) and (12), we have

$$\boldsymbol{\psi}(k) = \boldsymbol{\psi}(k-1) + \frac{\partial\eta(k-1)}{\partial\eta(k-1)}$$
$$\times\, e(k-1)\left\{\Phi'(k-1)\right\}^*\mathbf{x}^*(k-1)$$
$$+\,\eta(k-1)\frac{\partial e(k-1)}{\partial\eta(k-1)}\left\{\Phi'(k-1)\right\}^*\mathbf{x}^*(k-1)$$
$$+\,\eta(k-1)e(k-1)\frac{\partial\left\{\Phi'(k-1)\right\}^*}{\partial\eta(k-1)}\mathbf{x}^*(k-1)$$
$$+\,\eta(k-1)e(k-1)\left\{\Phi'(k-1)\right\}^*\frac{\partial\mathbf{x}^*(k-1)}{\partial\eta(k-1)}. \quad (13)$$

Notice that the last term in (13) is zero since $(\partial\mathbf{x}^*(k-1))/(\partial\eta(k-1)) = \mathbf{0}$. We can now express the vector of weight gradients $\boldsymbol{\psi}(k)$ in a more compact form, given by

$$\boldsymbol{\psi}(k) = \boldsymbol{\psi}(k-1)\left[\mathbf{I} - \eta(k-1)\left\vert\Phi'(k-1)\right\vert^2\mathbf{x}^*(k-1)\mathbf{x}^T(k-1)\right.$$
$$+\,\eta(k-1)e(k-1)$$
$$\left.\times\left\{\Phi''(k-1)\right\}^*\mathbf{x}^*(k-1)\mathbf{x}^H(k-1)\right]$$
$$+\, e(k-1)\left\{\Phi'(k-1)\right\}^*\mathbf{x}^*(k-1). \quad (14)$$

Notice from (14) that within the gradient update of (14), we, in fact, perform low-pass filtering of local weight gradients $e(k-1)\left\{\Phi'(k-1)\right\}^*\mathbf{x}^*(k-1)$, and the term within the square brackets in (14) determines the property of this time-varying low-pass filter. Finally, we arrive at the VSCNGD1 algorithm, for which the gradient is given by

$$\nabla_\eta E(k)$$
$$= -\frac{1}{2}\left[e(k)\left\{\Phi'(k)\right\}^*\mathbf{x}^H(k)\boldsymbol{\psi}^*(k) + e^*(k)\Phi'(k)\mathbf{x}^T(k)\boldsymbol{\psi}(k)\right]. \quad (15)$$

This concludes the derivation of a general VSCNGD1 algorithm. This approach provides a general and rigorous way to obtain a gradient-adaptive step size for nonlinear complex-valued adaptive filters and corresponds to the Benveniste's algorithm for real-valued linear FIR filters. In terms of computational complexity, it requires 16 complex multiplications in the weight update. Next, we will derive modifications of this algorithm which are less computationally complex, but still maintain improved performance, compared to CNGD.

B. VSCNGD2 Algorithm

Notice that the $\boldsymbol{\psi}$ term from (14) represents a time-varying filtered version of the instantaneous gradient $e(k-1)\left\{\Phi'(k-1)\right\}^*\mathbf{x}^*(k-1)$.

To simplify the step-size update, extending the approach from [6], we replace the time-varying "filtering" term in square brackets in (14) by a constant $0 < \alpha < 1$. This leads to the VSCNGD2 algorithm given by

$$\boldsymbol{\psi}(k) = \alpha\boldsymbol{\psi}(k-1) + e(k-1)\left\{\Phi'(k-1)\right\}^*\mathbf{x}^*(k-1). \quad (16)$$

For $\alpha < 1$, (16) represents a low-pass filter which combines the present and past filtered observations of the instantaneous gradients $e(k-1)\left\{\Phi'(k-1)\right\}^*\mathbf{x}^*(k-1)$. This way, we have relaxed computational complexity as compared to VSCNGD1, as shown in Table I.

C. VSCNGD3 Algorithm

From the weight sensitivity update (16), for $\alpha = 0$, the VSCNGD3 algorithm, which is the simplest of the three proposed algorithms, is obtained. The step-size update is based only on the instantaneous estimates [3] of gradients from (4), which can exhibit noisy behavior [3]. Therefore, the VSCNGD3 algorithm does not require the knowledge of the past values of the gradient, and is computationally less demanding than VSCNGD1 and VSCNGD2. Table I summarizes the three proposed algorithms and associated computational complexity.

III. STABILITY ANALYSIS

Following the approach from [14], our stability analysis is based on the contraction mapping principle and addresses the mean-square convergence of the proposed algorithms and sets bounds on the values of the step size which preserve stability. By the contraction mapping theorem (CMT), function $F : \mathbf{Z} \to \mathbf{Z}$ is a contraction if [15]

$$\left\vert F(x) - F(y)\right\vert \leq \gamma\vert x - y\vert \qquad \forall\, x, y \in \mathbf{Z} \quad (17)$$

where $0 \leq \gamma < 1$. Recall that, contractive functions have at most one fixed point, where for every $x \in \mathbf{Z}$, $\left\vert F^n(x) - F^{n+1}(x)\right\vert \to 0$ as[4] $n \to \infty$. To make use of CMT, let $\mathbf{w}_{\mathrm{opt}}$ be an optimal solution for the weight vector $\mathbf{w}$, and $\mathbf{v}(k) = \mathbf{w}(k) - \mathbf{w}_{\mathrm{opt}}$. Subtract $\mathbf{w}_{\mathrm{opt}}$ from both sides of (1), to give

$$\mathbf{v}(k+1) = \mathbf{v}(k) + \eta(k)e(k)\left[\Phi'(k)\right]^*\mathbf{x}^*(k)$$
$$= \mathbf{v}(k) - \eta(k)\left[\Phi'(k)\right]^*\mathbf{x}^*(k)$$
$$\times\left[\Phi(\mathbf{x}^T(k)\mathbf{w}(k)) - \Phi(\mathbf{x}^T(k)\mathbf{w}_{\mathrm{opt}})\right]$$
$$+\,\eta(k)e_{\mathrm{opt}}\left[\Phi'(k)\right]^*\mathbf{x}^*(k) \quad (18)$$

where $e_{\mathrm{opt}} = d(k) - \Phi(\mathbf{x}(k)\mathbf{w}_{\mathrm{opt}})$. For simplicity, let us consider a complex hyperbolic tangent, with the slope $\beta = 1$, that is $\Phi(\beta x) = \tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$ for which it is guaranteed that [14]

$$\left\vert\tanh(x) - \tanh(y)\right\vert \leq \left\vert\tanh'(\xi)\right\vert\vert x - y\vert \quad (19)$$

[4]$F^n(x) = F(F\ldots F(x))$ for n times.

TABLE II
PREDICTION GAIN R_p FOR CNGD AND PROPOSED ALGORITHMS ON VARIOUS SIGNALS

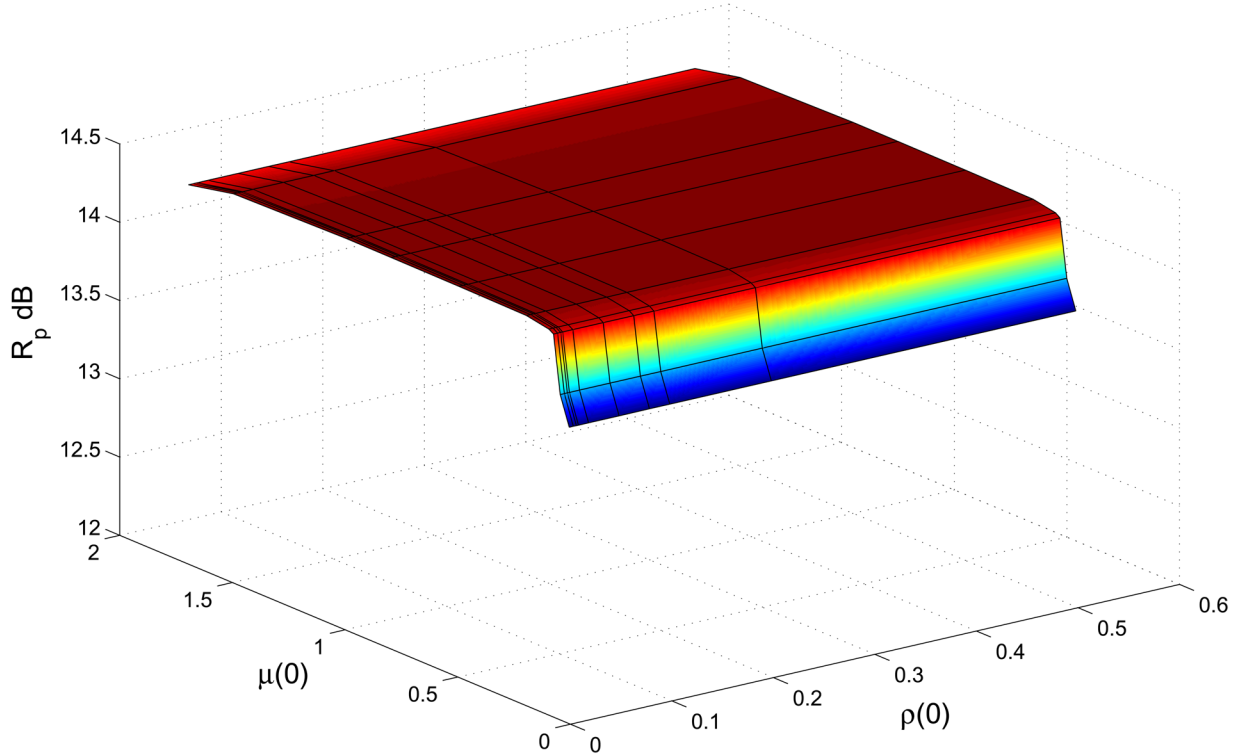| R_p [dB]	Parameters	AR(4) (23)	Nonlinear (24)	Wind (1m)	Wind (17m)	Radar
CNGD	$\eta(0)$=0.05	5.010	1.910	7.8891	9.5481	10.154
VSCNGD1	$\eta(0)$=0.05, ρ=0.0002	6.606	5.0911	17.5468	14.4142	15.510
VSCNGD2	α=0.95, $\eta(0)$=0.05, ρ=0.0002	6.271	5.0868	17.2631	14.0022	14.314
VSCNGD3	$\eta(0)$=0.05, ρ=0.0002	4.099	1.877	8.096	9.7754	13.145

Fig. 2.   Relationship between prediction gain R_p and initial values of $\eta(0)$ and $\rho(0)$ within VSCNGD1 for nonlinear prediction of the wind signal.

for $\forall\, x, y \in [a, b]$ and $\exists\, \xi \in (a, b)$. Combining (19) and (18), we have

$$\mathbf{v}(k+1) \leq \left[\mathbf{I} - \eta(k)\left[\Phi'(k)\right]^* \Phi'(\xi)\mathbf{x}^*(k)\mathbf{x}^T(k)\right]\mathbf{v}(k) \quad (20)$$

where, in the convergence analysis, the nonhomogeneous part of (18) can be ignored. By applying the statistical expectation operator to the squared Euclidean norm of (20), we obtain

$$E\|\mathbf{v}(k+1)\|_2^2 \leq \Lambda[E[B^2]]E\|\mathbf{v}(k)\|_2^2 \quad (21)$$

where $\Lambda[E[B^2]]$ is the maximum eigenvalue of $E[B^2] = E\left[\mathbf{I} - \eta(k)\left[\Phi'(k)\right]^* \Phi'(\xi)\mathbf{x}^*(k)\mathbf{x}^T(k)\right]^2$. Combining the condition from (21) with the CMT principle (17), we obtain $\Lambda[E[B^2]] \leq \gamma^2 \leq 1$. Thus, for convergence, the step size $\eta(k)$ is bounded from before by

$$0 < \eta(k) < \frac{2}{[2\lambda + tr\,(\mathbf{R_{xx}^*})]\,|\,[\Phi'(k)]^*\,\Phi'(\xi)|} \quad (22)$$

where λ is the maximum eigenvalue of the input autocorrelation matrix $\mathbf{R_{xx}^*}$.

A. Choice of Nonlinear Activation Function

Empirical evidence shows that when using the elementary transcendental function recommended by Kim and Adali [11], the associated

singular points and discontinuities at nonzero points do not pose a problem in training when the domain of interest is bounded within a circle of radius $\pi/2$. If the domain is larger, the training process tends to become more sensitive to the choice of the learning rate and which may become prohibitive when using a fixed learning rate. Also, the initial random weights need to be bounded within a circle with small radius.

IV. SIMULATIONS

In all the experiments, the order of the nonlinear adaptive filter was chosen to be $M = 4$, with $\beta = 1$. Simulations were undertaken by averaging 200 iterations of independent trials on prediction of both complex-valued benchmark colored and nonlinear signals as well as single trial real-life signals. The colored signal was a stable complex $AR(4)$ process given by

$$r(k) = 1.79r(k-1) - 1.85r(k-2)$$
$$+ 1.27r(k-3) + 0.41r(k-4) + n(k) \quad (23)$$

with complex white Gaussian noise (CWGN) $\in \mathcal{N}(\mathbf{0}, \mathbf{C})$ where $\mathbf{C}$ is the complex-valued covariance matrix. The CWGN can be expressed as $n(k) = n^r(k) + jn^i(k)$ where $n^r(k), n^i(k) \sim \mathcal{N}(0, 1)$. The real
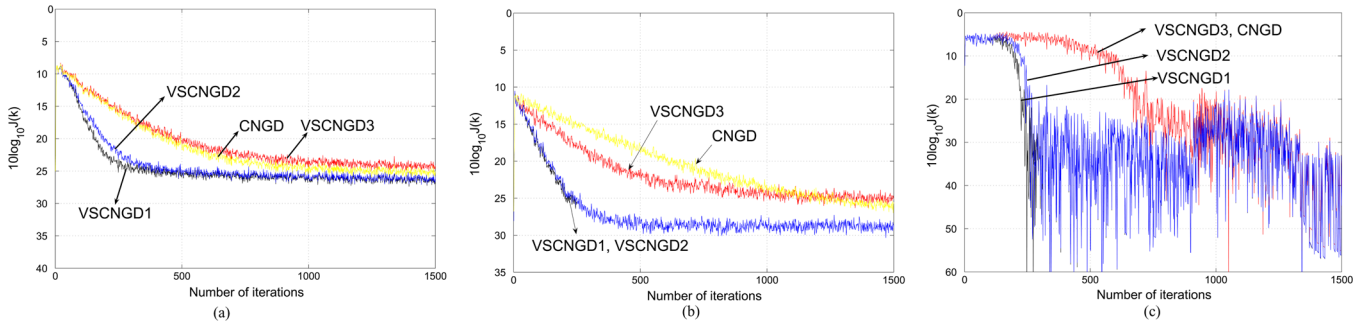
Fig. 3. Performance of CNGD, VSCNGD1, VSCNGD2, and VSCNGD3 algorithms on prediction of colored (23), nonlinear (24), and wind signals (17 m). (a) Colored input. (b) Nonlinear input. (c) Wind signal.
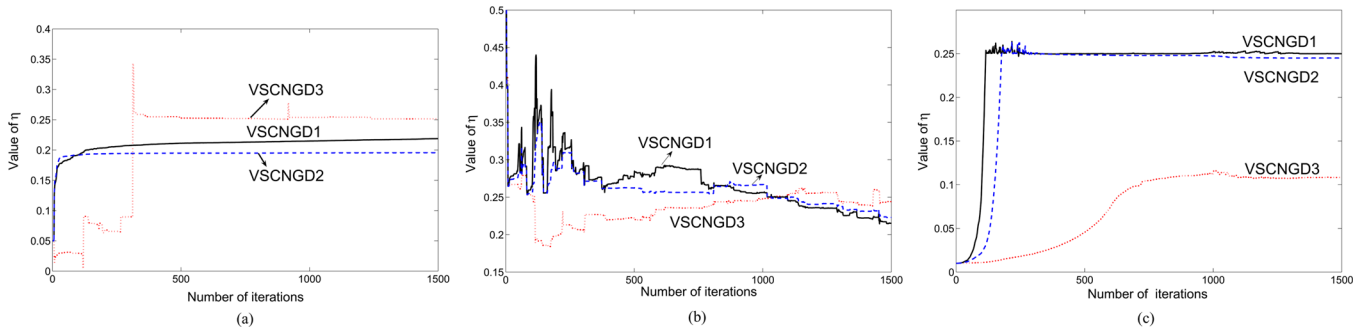

Fig. 4. Time variation of $\eta(k)$ for VSCNGD1, VSCNGD2, and VSCNGD3 algorithms on prediction of colored (23), nonlinear (24), and wind signals (17 m).

and imaginary components of CWGN were mutually statistically independent sequences having equal variances so that $\sigma_n^2 = \sigma_{nr}^2 + \sigma_{ni}^2$. The nonlinear benchmark input signal [16] used in simulations is given by

$$ z(k) = \frac{z(k-1)}{1 + z^2(k-1)} + r^3(k). \qquad (24) $$

The real-world signals used were the complex-valued radar[5] and two sets of wind[6] data. The measurement used to assess the performance was the prediction gain $R_p = 10 \log_{10} \left(\sigma_x^2 / \hat{\sigma}_e^2 \right) [dB]$ [17], where σ_x^2 denotes the variance of the input signal $x(k)$, and $\hat{\sigma}_e^2$ denotes the estimated variance of the forward prediction error $e(k)$.

Table II compares prediction gains of the CNGD and the proposed variable step-size algorithms for the signals considered and also shows the selection of initial parameters used in the simulations. In all the cases, VSCNGD1 gave best performance, followed by VSCNGD2, VSCNGD3, and standard CNGD. The improvement in the performance when using the proposed class of algorithms was especially visible for real-world data (columns 5–7 in Table II). To illustrate the sensitivity of the proposed variable step-size algorithms to the choice of their parameters, Fig. 2 shows the variation in R_p for a range of initial values of $\eta(0)$ and $\rho(0)$. As shown in Fig. 2, one virtue of the proposed algorithms is that the choice of initial values of $\eta(0)$ and $\rho(0)$ does not have significant effect on the performance.

In the next experiment, we compared the learning curves for the proposed algorithms. Fig. 3 shows the performance of CNGD, VSCNGD1,

VSCNGD2, and VSCNGD3 algorithms on the colored (23), nonlinear (24), and wind signal (high). Observe that for all the signals considered, VSCNGD1 exhibited fastest convergence. Depending on the signal in hand, VSCNGD3 typically exhibited similar or slightly worse performance than CNGD, whereas the performance of VSCNGD2 was in between the two. To further illustrate the convergence of the VSSCNGD algorithms, Fig. 4 illustrates the time variation of the step size $\eta(k)$ for the experiment from Fig. 3. Notice that for all the proposed algorithms, step size η exhibited fast convergence, and its behavior reflected the nonlinearity and nonstationarity within the signal in hand.

V. CONCLUSION

To speed up convergence of complex-valued nonlinear adaptive filters and to improve their operation in statistically nonstationary environments, the step size in the CNGD algorithm has been made adaptive. This has been achieved based on stochastic gradient, to give a class of VSCNGD algorithms. For generality, the derivation has been conducted for a general complex nonlinear activation function of a neuron. The derivation has been rigorous and has provided a general and computationally complex VSCNDG1 algorithm and two of its less computationally demanding approximations. A comprehensive set of simulations on both (linear and nonlinear) benchmark and real-life complex signals illustrates the validity of the proposed approach.

---

[5]Publicly available from http://soma.ece.mcmaster.ca/ipix/.

[6]The wind measurements used in simulations come from an AM ultrasonic anemometer and were sampled at 50 Hz for 1-h-long interval. The measurements were recorded at 1 and 17 m high from the ground level. The difference in height during measurements gave different wind dynamics. The wind vector can be expressed in the complex domain $\mathbb{C}$ as $v(t)e^{j\theta(t)} = v_E(t) + j\,v_N(t)$. Here, the two wind components, the speed v and direction θ, which are of different natures, are modeled as a single quantity in a complex representation space.

REFERENCES

[1] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *Proc. IEEE*, vol. 63, no. 4, pp. 712–720, Apr. 1975.
[2] H. Leung and S. Haykin, "The complex backpropagation algorithm," *IEEE Trans. Signal Process.*, vol. 39, no. 9, pp. 2101–2104, Sep. 1991.

[3] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2075–2087, Jun. 1993.

[4] D. P. Mandic, "A generalized normalized gradient descent algorithm," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 115–118, Feb. 2004.

[5] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximation*.   New York: Springer-Verlag, 1990.

[6] W. P. Ang and B. Farhang-Boroujeny, "A new class of gradient adaptive step-size LMS algorithms," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 805–810, Apr. 2001.

[7] J. B. Evans, P. Xue, and B. Liu, "Analysis and implementation of variable step size adaptive algorithms," *IEEE Trans. Signal Process.*, vol. 41, no. 8, pp. 2517–2535, Aug. 1993.

[8] N. Benvenuto and F. Piazza, "On the complex backpropagation algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 967–969, Apr. 1992.

[9] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: A Neural Network Approach*.   New York: Wiley, 2007.

[10] T. Kim and T. Adali, "Approximation by fully complex multilayer perceptrons," *Neural Comput.*, vol. 15, no. 7, pp. 1641–1666, 2003.

[11] T. Kim and T. Adali, "Approximation by fully complex MLP using elementary transcendental activation functions," in *Proc. 11th IEEE Workshop Neural Netw. Signal Process.*, 2001, pp. 203–212.

[12] A. I. Hanna and D. P. Mandic, "A complex-valued nonlinear neural adaptive filter with a gradient adaptive amplitude of the activation function," *Neural Netw.*, vol. 16, no. 2, pp. 155–159, 2003.

[13] S. L. Goh and D. P. Mandic, "A class of gradient-adaptive step size algorithms for complex-valued nonlinear neural adaptive filters," in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, vol. 5, pp. 253–256.

[14] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*.   New York: Wiley, 2001.

[15] L. Gorniewicz, S. A. Marano, and M. Slosarski, "Fixed points of contractive multivalued maps," *Proc. Amer. Math. Soc.*, vol. 124, no. 9, pp. 2675–2683, 1996.

[16] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Netw.*, vol. 1, no. 1, pp. 4–27, Mar. 1990.

[17] S. Haykin and L. Li, "Nonlinear adaptive prediction of nonstationary signals," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 526–535, Feb. 1995.

Blind Source Extraction Using Generalized Autocorrelations

Zhenwei Shi and Changshui Zhang

Abstract—This letter addresses blind (semiblind) source extraction (BSE) problem when a desired source signal has temporal structures, such as linear or nonlinear autocorrelations. Using the temporal characteristics of sources, we develop objective functions based on the generalized autocorrelations of primary sources. Maximizing the objective functions, we propose simple fixed-point source extraction algorithms. We give the stability analysis and prove convergence properties of the algorithms as the generalized autocorrelation function is linear or nonlinear. Especially, as the generalized autocorrelation function is linear, the algorithm has interesting character of "one-iteration" convergence under some conditions. Computer simulations and real-data application experiments show that the algorithms are appealing BSE methods for temporal signals of interest by capturing the linear or nonlinear autocorrelations of the desired sources.

Index Terms—Blind source extraction (BSE), blind source separation (BSS), fetal electrocardiogram (FECG), independent component analysis (ICA).

I. INTRODUCTION

THE problem of blind (semiblind) source extraction (BSE) has received wide attention in various fields such as biomedical signal processing and analysis, data mining, speech and image processing, and so on [2], [3], [6]. The BSE learning algorithms can extract a single source signal from linear mixture of source signals and may have several advantages over simultaneous blind source separation (BSS) [2]. For example, only "interesting" source signals need to be extracted; signals can be extracted in a specific order according to some features of source signals; lots of computing time and resources can be saved.

Many source extraction algorithms [2] can extract a specific signal as the first output, by using some *a priori* information, such as non-Gaussianity [5], smoothness or linear predictability [1], [2], sparseness [8], etc. However, the work reported in this letter has been motivated by generalized autocorrelations (linear or nonlinear autocorrelations) of the desired source signals. By capturing the generalized autocorrelation characteristics of the signals of interest, we obtain new BSE algorithms.

The structure of this letter is as follows. The objective functions based on the generalized autocorrelation functions of the sources and the fixed-point algorithms for optimizing the objective functions are proposed in Section II. Furthermore, we prove the theoretical properties of the fixed-point algorithms. In Section III, experiments on artificial data and real-world electrocardiogram (ECG) data are presented. Some discussion and conclusions are drawn in Section IV.

Z. Shi was with the State Key Laboratory of Intelligent Technology and Systems, Department of Automation, Tsinghua University, Beijing 100084, P.R. China. He is now with the Image Processing Center, School of Astronautics, Beijing University of Aeronautics and Astronautics, Beijing 100083, P.R. China (e-mail: shizhenwei@mail.tsinghua.edu.cn).

C. Zhang is with the State Key Laboratory of Intelligent Technology and Systems, Department of Automation, Tsinghua University, Beijing 100084, P.R. China.