the signal or noise powers. *This is especially useful in situations where the SNR is not known a priori or changes unpredictably.*

B. Convergence Behavior

Next, we will analyze how the NLMS filter converges when $\mathbf{w}_o$ suffers an instantaneous change, both in the biased and standard configurations. The setup is as follows: $u(n)$ is a colored sequence obtained from a first-order autoregressive model with transfer function $\sqrt{1-h^2}/(1-hz^{-1})$, $h = 0.6$, fed with an i.i.d. Gaussian random process, whose variance is selected so that $\mathrm{tr}(\mathbf{R}) = 0.1$. Plant coefficients are initially selected as in the previous subsection, and then changed at $n = 4000$ and $n = 8000$. The three sets of coefficients have been scaled to obtain different SNRs in the reference signal (see Fig. 5), keeping $\sigma_0^2 = 0.1$ constant during the simulation. NLMS weights are adapted with $\mu = 0.1$, while the output factor $\alpha(n)$ follows (12) and (13), with $\mu_a = 0.1$. Emsemble-average curves are obtained from 5000 independent realizations.

The EMSE evolution depicted in Fig. 5(a) is a good example of the advantages that can be obtained with biased filters in scenarios where the SNR is time-varying. As illustrated in Fig. 5(a), both the standard and biased versions of NLMS reconverge at similar speeds after every perturbation in $\mathbf{w}_o$, but the biased version achieves lower EMSE during all the simulation. This shows that for this value of μ_a the steady-state EMSE reduction of the biased scheme is not obtained at the cost of a slower convergence. In principle, the speed of convergence could be degraded if a smaller μ_a was used to reduce the gradient noise introduced by the estimation of $\alpha(n)$. Note, however, that according to Fig. 4(b) the gradient noise that appears for $\mu_a = 0.1$ is already very small. In the right panel of Fig. 5, $\alpha(n)$ evolution is plotted. Interestingly, $\alpha(n)$ decreases towards zero at the beginning of each re-convergence, trying to cancel out the initial meaningless predictions of NLMS. It is only as NLMS starts recovering track of the filter weights that $\alpha(n)$ heads towards its steady-state value.

VI. CONCLUSION

Biasing the weights of adaptive filters can be an interesting way of reducing their MSE. In this correspondence, we have illustrated this idea with a very simple yet effective configuration, consisting in multiplying the filter output by a constant factor. Realizable schemes for adaptively learning this multiplicative factor are proposed, providing both theoretical and experimental evidence about the benefits of this approach.

We are currently working on more sophisticated biased schemes which exploit the structure of the input regressors, as well as on the extension of these ideas to filters operating in tracking situations.

REFERENCES

[1] A. H. Sayed, *Adaptive Filters*. Hoboken, NJ: Wiley, 2008.
[2] Y. C. Eldar, "Uniformly improving the Cramér-Rao bound and maximum likelihood estimation," *IEEE Trans. Signal Process.*, vol. 54, pp. 2943–2956, 2006.
[3] S. Kay and Y. C. Eldar, "Rethinking biased estimation," *IEEE Signal Process. Mag.*, pp. 133–136, 2008.
[4] S. S. Kozat and A. C. Singer, "Multi-stage adaptive signal processing algorithms," in *Proc. SAM Signal Process. Workshop*, 2000, pp. 380–384.
[5] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54, pp. 1078–1090, 2006.
[6] J. Arenas-García, M. Martínez-Ramón, A. Navia-Vázquez, and A. R. Figueiras-Vidal, "Plant identification via adaptive combination of transversal filters," *Signal Process.*, vol. 86, pp. 2430–2438, 2006.
[7] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-García, "A normalized adaptation scheme for the convex combination of two adaptive filters," in *Proc. ICASSP'08*, Las Vegas, NV, 2008, pp. 3301–3304.
[8] R. Candido, M. T. M. Silva, and V. H. Nascimento, "Affine combination of adaptive filters," in *Proc. 42nd Asilomar Conf. Signal, Syst. Comput.*, Oct. 2008, pp. 236–240.
[9] N. J. Bershad, J. C. M. Bermudez, and J.-Y. Tourneret, "An affine combination of two LMS adaptive filters—Transient mean-square analysis," *IEEE Trans. Signal Process.*, vol. 56, pp. 1853–1864, 2008.
[10] E. Trentin, "Networks with trainable amplitude of activation functions," *Neural Netw.*, vol. 14, pp. 471–493, 2001.
[11] S. L. Goh and D. P. Mandic, "Recurrent neural networks with trainable amplitude of activation functions," *Neural Netw.*, vol. 16, pp. 1095–1100, 2003.

Quaternion-Valued Stochastic Gradient-Based Adaptive IIR Filtering

Clive Cheong Took, *Member, IEEE*, and
Danilo P. Mandic, *Senior Member, IEEE*

Abstract—A learning algorithm for the training of quaternion valued adaptive infinite impulse (IIR) filters is introduced. This is achieved by taking into account specific properties of stochastic gradient approximation in the quaternion domain and the recursive nature of the sensitivities within the IIR filter updates, to give the quaternion-valued stochastic gradient algorithm for adaptive IIR filtering (QSG-IIR). Further, to reduce computational complexity, a variant of the QSG-IIR is introduced, which for small stepsizes makes better use of the available information. Stability analysis and simulations on both synthetic and real world 4D data support the approach.

Index Terms—Adaptive prediction, infinite impulse response (IIR) filters, quaternion adaptive filtering, stochastic gradient, wind modeling.

I. INTRODUCTION

Linear adaptive finite impulse response (FIR) filtering in $\mathbb{R}$ is well established, and the least mean square (LMS) algorithm has long become a standard in many practical applications. However, when modeling systems with long impulse responses and signals with long-term correlation, FIR filters may require a prohibitively large filter length. In those cases, the infinite impulse response (IIR) architecture is a more convenient choice, as due to the feedback, long time correlations can be modeled using a small-scale model [1]. Such filters have rational transfer functions, which are more general than the polynomial transfer functions of FIR filters, and have been used in a number of applications [1]–[3].

The recent progress in technology, robotics, and biomedicine has highlighted the need for adaptive filtering of several classes of multidimensional signals, typically recorded from vector sensors, for instance, 3D anemometers and 3D body motion sensors [4], [5]. By pro-

cessing those data directly in the multidimensional domain where they reside, the correlation and coupling between each dimension are naturally accounted for, leading to improved accuracies. Work in this direction includes the extensions of FIR and IIR adaptive filters to the complex domain $\mathbb{C}$ [6], and more recently, for quadrivatiate data, to adaptive FIR filters operating in the quaternion domain such as the Quaternion LMS (QLMS) [4]. Within the QLMS weight update, every component of a quaternion valued weight vector is derived based on the total output error corresponding to all the four data channels, unlike the multichannel LMS (MLMS) for which the channelwise weight update is based on only on the corresponding channel error [4].

It is important to note that, unlike multichannel LMS algorithms operating in $\mathbb{R}^N$, the algorithms in the quaternion domain are not a straightforward extension of their real-valued or complex-valued counterparts [6], [7], because: i) quaternion algebra treats the axis vectors also as imaginary units, hence, e.g., the derivation of QLMS is fundamentally different from that of complex LMS or MLMS; ii) the noncommutativity of the quaternion product requires careful consideration of both the error and its conjugate, e.g., the cost functions $e^*(n)e(n)$ and $e(n)e^*(n)$ lead to different updates; iii) unlike the complex domain statistics [6] the pseudocovariance of a quaternion-valued proper (Q-proper) signal does not vanish; iv) the QLMS update includes the information related to both the covariance $\mathcal{C}_\mathbf{q}$ and the pseudocovariance $\mathcal{P}_\mathbf{q}$ and thus its update does not degenerate exactly into that of complex LMS, when only two dimensions are considered [4]. This illustrates that quaternions, although seen as ordered pairs of two complex numbers, do not admit to straightforward extensions of the corresponding complex-valued algorithms.

We here introduce a stochastic gradient algorithm for IIR adaptive filtering in $\mathbb{H}$, as rigorous adaptive IIR filtering algorithms for quaternion-valued signals are still lacking. It is shown that processing 3- and 4D signals in the quaternion domain $\mathbb{H}$ has potential advantages over the processing in $\mathbb{R}^3$ and $\mathbb{R}^4$ [8], [9]. The derivation of the stochastic gradient algorithm for adaptive IIR filters (QSG-IIR) is supported by an outline analysis and illustrative simulations.

II. PROPERTIES OF QUATERNION RANDOM VECTORS

A. Quaternion Algebra

Quaternions provide a very convenient framework for statistical signal processing of 3D and 4D signals and can be regarded as a noncommutative extension of complex numbers [10]. A quaternion variable $q \in \mathbb{H}$ comprises a real part $\Re\{\cdot\}$ (denoted with subscript a) and a vector part $\Im\{\cdot\}$ consisting of three imaginary parts (denoted with subscripts b, c, and d), and can be expressed as

$$
\begin{aligned}
q &= [\Re\{q\}, \Im\{q\}] \\
&= [q_a, (q_b, q_c, q_d)] = q_a + q_b\iota + q_c\jmath + q_d\kappa.
\end{aligned} \tag{1}
$$

The properties of the orthogonal unit vectors, ι, $\jmath$, κ (which are also imaginary numbers) describing the three vector dimensions of a quaternion are given by

$$
\iota\jmath = \kappa \quad \jmath\kappa = \iota \quad \kappa\iota = \jmath
$$

$$
\iota\jmath\kappa = \iota^2 = \jmath^2 = \kappa^2 = -1. \tag{2}
$$

The noncommutativity of the quaternion product arises because of the vector product within the quaternion multiplication; for every $q_1, q_2 \in \mathbb{H}$, quaternion multiplication is given by (3) shown at the bottom of the page, where symbols "$\times$" and "$\cdot$" denote, respectively, the cross-product and the dot-product. Thus, the quaternion domain $\mathbb{H}$ forms a noncommutative vector space, that is, $q_1q_2 \neq q_2q_1$. The conjugate of a quaternion $q^* = [q_a, \Im\{q\}]^* = [q_a, -\Im\{q\}]$, and the norm $|q| = \sqrt{qq^*}$, which obeys the relationship $|q_1||q_2| = |q_1q_2|$. Note, that quaternion conjugation is antiinvolution, and $(q_1q_2)^* = q_2^*q_1^*$.

B. Quaternion Statistics

Recent advances in the statistics of complex variables have shown that only the use of both the covariance matrix $\mathcal{C}_\mathbf{z} = E\{\mathbf{z}(n)\mathbf{z}^H(n)\}$ and the pseudocovariance $\mathcal{P}_\mathbf{z} = E\{\mathbf{z}(n)\mathbf{z}^T(n)\}$ gives a complete description of the second-order statistical properties [11]. The covariance is usually related to the power of a complex random variable, whereas the pseudocovariance allows us to express the degree of correlation between the real part $\Re\{z(n)\}$ and the imaginary part $\Im\{z(n)\}$ [6]. It is, therefore, natural to ask whether these two second-order statistical measurements have a corresponding physical interpretation in the quaternion domain. Note that, the power of a quaternion variable can be determined as

$$
q(n)q^*(n) = q_a^2(n) + q_b^2(n) + q_c^2(n) + q_d^2(n)
$$

whereas the correlation structure is reflected in the imaginary part of the quaternion product as

$$
\Im\{q(n)q(n)\} = 2q_a(n)\Big[q_b(n)\iota + q_c(n)\jmath + q_d(n)\kappa\Big].
$$

To draw a parallel with the complex domain, recall that a vanishing pseudocovariance of a complex-valued variable implies the circular shape (rotation invariance) of the joint probability density function (pdf) of the real and imaginary part. Thus, for the pseudocovariance $\mathcal{P}_z = E\{zz\}$ of a complex valued proper variable $z(n) = x(n) + \iota y(n)$ to vanish, two conditions need to be satisfied, 1) $\sigma_x^2 = \sigma_y^2$; 2) $E\{x(n)y(n)\} = 0$, that is, the real and imaginary part are of equal power and not correlated [6]. According to Theorem 2 in [12], these conditions also have to be satisfied for a quaternion-valued second-order circular or proper (Q-proper) signal, which leads to the following observation.

1) Observation 1: The pseudocovariance of a Q-proper signal does not vanish.

    Proof: The first condition of Q-properness requires equal powers for every component of a quaternion variable, that is, $\sigma_a^2 = \sigma_{b,c,d}^2$ and thus $E\{q(n)q(n)\} = \sigma^2 - 3\sigma^2 = -2\sigma^2$.

Recently, Eriksson and Koivunen proposed the strong-uncorrelating transform (SUT) to perform a simultaneous diagonalisation of both the complex-valued covariance and the pseudocovariance [13]. However, the SUT transform is not defined in $\mathbb{H}$, for it requires the pseudocovariance to be symmetric, which is not the case for general, noncircular quaternion-valued variables. Although in $\mathbb{H}$ the diagonalization of the

$$
\begin{aligned}
q_1q_2 &= [q_{a,1}, \Im\{q_1\}][q_{a,2}, \Im\{q_2\}] \\
&= [q_{a,1}q_{a,2} - \Im\{q_1\} \cdot \Im\{q_2\}, \; q_{a,1}\Im\{q_2\} + q_{a,2}\Im\{q_1\} + \Im\{q_1\} \times \Im\{q_2\}]
\end{aligned} \tag{3}
$$

pseudocovariance and the covariance are mutually exclusive, as a consequence of Observation 1, an i.i.d. Q-proper signal has a negative definite pseudocovariance $\mathcal{P}_{\mathbf{q}} = -2\sigma^2 \mathbf{I}$, whereas its corresponding positive definite covariance is given by $\mathcal{C}_{\mathbf{q}} = 4\sigma^2 \mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix, leading to the following observation.

2) Observation 2: For an i.i.d. Q-proper signal, the covariance and the pseudocovariance are diagonal matrices, and, thus, have their respective eigenvalues $\{\lambda_c\}$ and $\{\lambda_p\}$ on their main diagonals. In addition, due to Observation 1, $|\lambda_{c,i}| > |\lambda_{p,i}| \, \forall \, i$.

Notice that both $\mathcal{C}_{\mathbf{q}}$ and $\mathcal{P}_{\mathbf{q}}$ of a Q-proper signal are real-valued, symmetric, and diagonal, which is convenient for the stability analysis, as this helps to circumvent the problem of diagonalization[1] of these correlation matrices via the quaternion singular value decomposition (whose left eigenvalues remain an open problem). We will therefore refer to the eigenvalues of the quaternion singular value decomposition as the right eigenvalues [14]. These two observations will be used in the stability analysis of adaptive quaternion-valued algorithms, thus providing an upper bound, whilst ensuring mathematical tractability of the results.

III. DERIVATION OF A RECURSIVE ALGORITHM FOR ADAPTIVE QUATERNION VALUED IIR FILTERING

A. A Recursive IIR Adaptive Filtering Algorithm

The operation of a direct form IIR adaptive filter is characterized by the following relation [15]:

$$y(n) = \sum_{m=1}^{M} y^*(n-m) g_m(n) + \sum_{m=0}^{N} x^*(n-m) h_m(n) \quad (4)$$

with $x(n)$, M, N, $h_m(n)$, $g_m(n)$ denoting, respectively, the input, order of feedback, tap input length, and adaptive weights corresponding to the delayed input and feedback. This can be written in a compact form as

$$y(n) = \mathbf{u}^H(n)\mathbf{w}(n) \quad (5)$$

where

$$\mathbf{w}(n) = [g_1(n), \ldots, g_M(n), h_0(n), \ldots, h_N(n)]^T \quad (6)$$
$$\mathbf{u}(n) = [y(n-1), \ldots, y(n-M), x(n), \ldots, x(n-N)]^T. \quad (7)$$

The objective function[2] to be minimized is the instantaneous square of the modulus of the output error, given by

$$E(n) = \frac{1}{2} e^*(n) e(n)$$

---

[1]Principal Component Analysis is traditionally employed to diagonalize the covariance, and thereby achieve interchannel decorrelation. However, this is no longer adequate when processing noncircular signals.

[2]Similarly to complex-valued adaptive filters [6], there are several equivalent formulations for the operation of quaternion-valued filters. For instance, the cost function $\boldsymbol{E(n)}$ is typically formulated in terms of the error $\boldsymbol{e(n)}$ and its conjugate $\boldsymbol{e^*(n)}$. However, $\boldsymbol{E(n)}$ can also be expressed in terms of three perpendicular involutions $\boldsymbol{e^\imath(n)}$, $\boldsymbol{e^\jmath(n)}$, $\boldsymbol{e^\kappa(n)}$ as $\boldsymbol{E^v(n)} = 0.5[\boldsymbol{e^\imath(n)} + \boldsymbol{e^\jmath(n)} + \boldsymbol{e^\kappa(n)} - \boldsymbol{e(n)}]\boldsymbol{e(n)}$. Each involution, given by $\boldsymbol{e^v(n)} = -\boldsymbol{v e(n) v}$ where $\boldsymbol{v} = \boldsymbol{\imath, \jmath, \kappa}$, recovers its corresponding element of the vector/imaginary part. The equivalence of $\boldsymbol{e^*(n)} = 0.5[\boldsymbol{e^\imath(n)} + \boldsymbol{e^\jmath(n)} + \boldsymbol{e^\kappa(n)} - \boldsymbol{e(n)}]$ results in the identity $\boldsymbol{\partial E(n)/\partial w} = \boldsymbol{\partial E^v(n)/\partial w}$, however, for simplicity $\boldsymbol{E^v(n)}$ will not be considered here. All these alternative forms have identical convergence and properties. Similar observations apply to the alternative expressions for the filter output $\boldsymbol{y(n)}$, that is, $\boldsymbol{y(n)} = \boldsymbol{w^T(n)x(n)}, \boldsymbol{y(n)} = \boldsymbol{x^T(n)w(n)}$ or $\boldsymbol{y(n)} = \boldsymbol{x^H(n)w(n)}$.

$$= \frac{1}{2}[e_a^2(n) + e_b^2(n) + e_c^2(n) + e_d^2(n)] \quad (8)$$

where the error $e(n) = d(n) - y(n)$, with $d(n)$ the desired response. Within the steepest descent optimization, the filter weights are updated based on

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \nabla_{\mathbf{w}} E(n) \quad (9)$$

where μ is the learning rate and the gradient $\nabla_{\mathbf{w}} E(n)$ can be evaluated as

$$
\begin{aligned}
\nabla_{\mathbf{w}} E(n) &= \frac{1}{2} \nabla_{\mathbf{w}} e^*(n) e(n) \\
&= \frac{1}{2}\left[e^*(n) \nabla_{\mathbf{w}} e(n) + \nabla_{\mathbf{w}} e^*(n) e(n) \right] \\
&= -\frac{1}{2}\left[e^*(n) \nabla_{\mathbf{w}} y(n) + \nabla_{\mathbf{w}} y^*(n) e(n) \right] \\
&= -e^*(n) \boldsymbol{\Psi}(n) - \boldsymbol{\Phi}(n) e(n) \quad (10)
\end{aligned}
$$

where

$$
\boldsymbol{\Psi}(n) = \frac{1}{2}\left[\frac{\partial y(n)}{\partial g_1(n)}, \ldots, \frac{\partial y(n)}{\partial g_M(n)}, \frac{\partial y(n)}{\partial h_0(n)}, \ldots, \frac{\partial y(n)}{\partial h_N(n)} \right]
$$
$$
\boldsymbol{\Phi}(n) = \frac{1}{2}\left[\frac{\partial y^*(n)}{\partial g_1(n)}, \ldots, \frac{\partial y^*(n)}{\partial g_M(n)}, \frac{\partial y^*(n)}{\partial h_0(n)}, \ldots, \frac{\partial y^*(n)}{\partial h_N(n)} \right].
$$
$$(11)$$

To calculate the vector term $\boldsymbol{\Phi}(n)$ in (11), we need to calculate partial derivatives of the output with respect to the filter coefficients corresponding to both the feedforward and feedback part, that is

$$\phi_{g_m}(n) = \frac{1}{2}\frac{\partial y^*(n)}{\partial g_m(n)} \quad \phi_{h_m}(n) = \frac{1}{2}\frac{\partial y^*(n)}{\partial h_m(n)}. \quad (12)$$

The term $\phi_{g_m}(n)$ can be expanded as

$$
\phi_{g_m}(n) = \frac{1}{2}\left[\frac{\partial y^*(n)}{\partial g_{a_m}(n)} + \imath \frac{\partial y^*(n)}{\partial g_{b_m}(n)} + \jmath \frac{\partial y^*(n)}{\partial g_{c_m}(n)} + \kappa \frac{\partial y^*(n)}{\partial g_{d_m}(n)} \right]
$$
$$(13)$$

and the terms within the gradient (13) corresponding to the real and three imaginary dimensions of a quaternion can be found based on the expansion of quaternion product $\mathbf{y}^H(n)\mathbf{g}(n)$, as shown in (14) at the bottom of the next page.

$$\imath \frac{\partial y^*(n)}{\partial g_{b_m}(n)} = y(n-m) + \imath \sum_{l=1}^{M} g_l^*(n) \frac{\partial y(n-l)}{\partial g_{b_m}(n)} \quad (15)$$

$$\jmath \frac{\partial y^*(n)}{\partial g_{c_m}(n)} = y(n-m) + \jmath \sum_{l=1}^{M} g_l^*(n) \frac{\partial y(n-l)}{\partial g_{b_m}(n)} \quad (16)$$

$$\kappa \frac{\partial y^*(n)}{\partial g_{d_m}(n)} = y(n-m) + \kappa \sum_{l=1}^{M} g_l^*(n) \frac{\partial y(n-l)}{\partial g_{d_m}(n)}. \quad (17)$$

Finally, the gradient term $\phi_{g_m}(n)$ is obtained as

$$\phi_{g_m}(n) = 2y(n-m) + \frac{1}{2}\sum_{l=1}^{M} g_l^*(n) \frac{\partial y(n-l)}{\partial g_m(n)}. \quad (18)$$

Due to the feedback within the IIR filter, there is a recursive term on the right-hand side (RHS) of the gradient in (18), which is not possible to calculate [15], as we have a partial derivative of the output with respect to the value of the weight in the future. To make the calculation of this term mathematically feasible, for a small stepsize, we use the standard approximation that [15]

$$\mathbf{w}(n) \approx \mathbf{w}(n-1) \approx \cdots \approx \mathbf{w}(n-M). \quad (19)$$

In other words, the small stepsize ensures a slow variation of the adaptive coefficients, making the approximation (19) valid [15]. This way, $\partial y^*(n-l)/\partial g_m(n)$ in (18) can be replaced with $\partial y^*(n-l)/\partial g_m(n-l)$, to yield

$$\phi_{g_m}(n) = 2y(n-m) + \sum_{l=1}^{M} g_l^*(n)\psi_{g_m}(n-l). \quad (20)$$

Recursive expressions for all the components of $\boldsymbol{\phi}(n)$ and $\boldsymbol{\psi}(n)$, that is, $\phi_{h_m}(n)$, $\psi_{g_m}(n)$ and $\psi_{h_m}(n)$, can be obtained in a similar fashion and are given

$$\phi_{h_m}(n) = 2x(n-m) + \sum_{l=1}^{M} g_l^*(n)\psi_{h_m}(n-l) \quad (21)$$

$$\psi_{g_m}(n) = -y(n-m) + \sum_{l=1}^{M} \phi_{g_m}(n-l)g_l(n) \quad (22)$$

$$\psi_{h_m}(n) = -x(n-m) + \sum_{l=1}^{M} \phi_{h_m}(n-l)g_l(n). \quad (23)$$

Recursions (20)–(23) depend on both the input $x(n-m)$ and the output $y(n-m)$, and are termed "forced difference" equations [15]. Observe that, unlike in the corresponding learning algorithm for complex valued IIR filters [15], the gradient vector (sensitivities) $\boldsymbol{\Phi}(n)$ does not vanish and it also depends on $\boldsymbol{\Psi}(n)$. The updates for the filter coefficients for quaternion valued stochastic gradient algorithm for adaptive IIR filters termed QSG-IIR, are finally given by

$$g_m(n+1) =$$
$$g_m(n) + \mu \left[e^*(n)\psi_{g_m}(n) + \phi_{g_m}(n)e(n) \right] \quad m = 1, \ldots, M$$
$$\quad (24)$$

$$h_m(n+1) =$$
$$h_m(n) + \mu \left[e^*(n)\psi_{h_m}(n) + \phi_{h_m}(n)e(n) \right] \quad m = 0, \ldots, N.$$
$$\quad (25)$$

These updates are function of both the error and its conjugate, in contrast to the gradients within the corresponding complex-valued algorithm [15].

B. Computationally Efficient Gradient Approximations

To simplify the adaptation of the QSG-IIR, a commonly used approach in adaptive IIR filtering is that, for a small value of the learning rate, the past values of $\boldsymbol{\phi}(n)$ and $\boldsymbol{\psi}(n)$ are replaced with the filtered versions of output and the input [15], to give

$$\psi_{g_m}(n) \approx y^F(n-m) \quad \phi_{g_m}(n) \approx y^{FF}(n-m)$$
$$m = 2, \ldots, M$$
$$\psi_{h_m}(n) \approx x^F(n-m) \quad \phi_{h_m}(n) \approx x^{FF}(n-m)$$
$$m = 1, \ldots, N. \quad (26)$$

The superscript F denotes the filtered versions of x and y within the update of $\psi(n)$, whereas the superscript FF corresponds to the filtered versions of x and y within the update of $\phi(n)$. This way, at every iteration, only the four updates $\phi_{g_1}(n)$, $\psi_{g_1}(n)$, $\phi_{h_0}(n)$, $\psi_{h_0}(n)$ are required, to yield

$$\phi_{g_1}(n) \approx y^{FF}(n-1) = 2y(n-1) + \sum_{l=2}^{M} g_l^*(n)y^F(n-l)$$
$$\quad (27)$$

$$\phi_{h_0}(n) \approx x^{FF}(n) = 2x(n) + \sum_{l=1}^{M} g_l^*(n)x^F(n-l) \quad (28)$$

$$\psi_{g_1}(n) \approx y^F(n-1) = -y(n-1) + \sum_{l=2}^{M} y^{FF}(n-l)g_l(n)$$
$$\quad (29)$$

$$\psi_{h_0}(n) \approx x^F(n) = -x(n) + \sum_{l=1}^{M} y^{FF}(n-l)g_l(n)$$
$$\quad (30)$$

and the computational complexity is reduced approximately by $\mathcal{O}(64M^2)$. This algorithm is referred to as the approximate QSG-IIR (AQSG-IIR).

IV. ON THE STEPSIZE BOUNDS OF QSG-IIR ALGORITHM

The main idea behind stability analysis of adaptive filters is to diagonalize the correlation matrices in the update of the weight error vector, thus directly leading to the stepsize bounds as functions of the inverse of the largest eigenvalue of the input correlation matrix. Unlike the complex domain $\mathbb{C}$, where we can use the strongly uncorrelating transform [13], the covariance and the pseudocovariance in $\mathbb{H}$ cannot be diagonalizsed simultaneously. To circumvent this problem, using the standard assumption that the input $\mathbf{x}(n)$ is an i.i.d. Q-proper random variable [1], [3], the approximate stepsize bound can be derived as

$$0 < \mu < \frac{2}{2\lambda_{c,\max} + \lambda_{p,\max}} \quad (31)$$

where $\lambda_{c,\max}$, $\lambda_{p,\max}$ denote the maximum right eigenvalue of $\mathcal{C}_\mathbf{u}$ and $\mathcal{P}_\mathbf{u}$, respectively. As shown in the Appendix, this bound is based on the largest mode of convergence and the linearization around the global

---

$$\frac{\partial y^*(n)}{\partial g_{a_m}(n)} = y_a(n-m) + \imath y_b(n-m) + \jmath y_c(n-m) + \kappa y_d(n-m) + \sum_{l=1}^{M} g_l^*(n)\frac{\partial y(n-l)}{\partial g_{a_m}(n)}$$

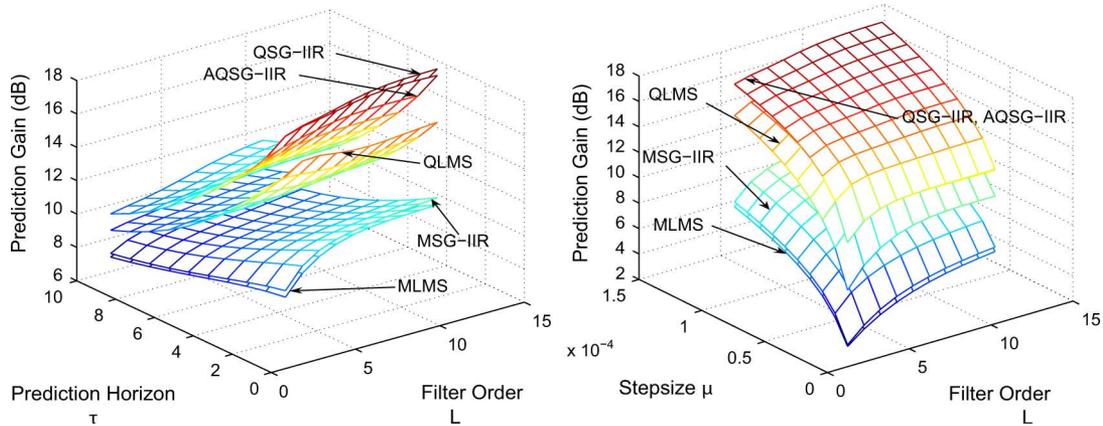$$= y(n-m) + \sum_{l=1}^{M} g_l^*(n)\frac{\partial y(n-l)}{\partial g_{a_m}(n)} \quad (14)$$

Fig. 1. Dependence of the performance on the choice of parameters of MLMS, QLMS, QSG-IIR, its approximate version (AQSG-IIR), and MSG-IIR for the nonlinear chaotic Saito's process [5].
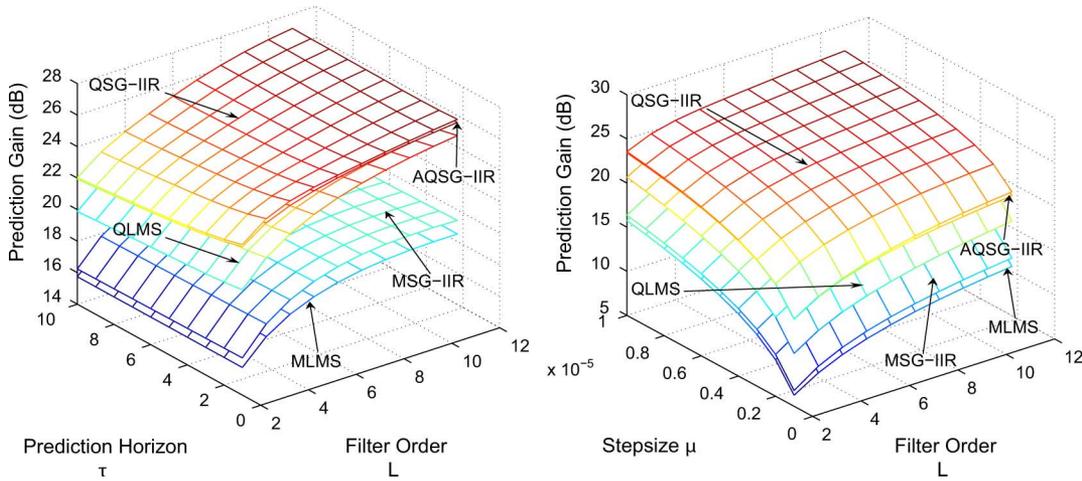

Fig. 2. Dependence of the performance of MLMS, QLMS, QSG-IIR, AQSG-IIR, and MSG-IIR on the choice of parameters. The experiments were conducted on the 4D wind recordings (3D wind speed and temperature).

minimum point (see also [1, pp. 349–351]) and simplifies into the step-size bounds of QLMS [4], when modeling moving average processes. The steady state analysis assumes stationarity of the vector $\mathbf{u}(n)$ (e.g., in the vicinity of the global minimum of the cost function). This assumption rests on the slow adaptation of the adaptive filter such that (19) is satisfied for a small stepsize [15].

V. SIMULATIONS

Three sets of simulations were conducted in a τ-step ahead prediction and a denoising setting; the datasets used were the chaotic 4D Saito's signal [5], real-world wind field, and color image (Lena). The performance index was the standard prediction gain, defined as $R_p = 10 \log \sigma_\mathbf{x}^2/\sigma_e^2$, where $\sigma_\mathbf{x}^2$ and σ_e^2 denote respectively the variances of the input signal and error (estimated from iteration $n = 1$ until convergence, see for instance [6], [7]). The filter order was chosen to be $L = M = N$. The performances of the proposed IIR algorithms were compared to those of the quaternion LMS (QLMS) [4], multichannel LMS (MLMS) [7] and a multichannel IIR architecture (MSG-IIR).

1) 4D Saito's Circuit: A 4D benchmark system considered was the Saito's process [5], characterized by four state variables (s, x, y, z) and five parameters $(\eta, \alpha_1, \alpha_2, \beta_1, \beta_2)$, which determine transitions from a torus doubling route to the area and volume of an expanding chaotic figure (for more details, see [5]). From Fig. 1, it is clear that the proposed IIR algorithms outperformed the QLMS, MLMS, and MSG-IIR, and exhibited better convergence properties (see Fig. 3) for the same filter order.

2) Wind Forecasting: Three-dimensional wind data were recorded using a 3D ultrasonic anemometer[3]. The wind speeds were taken from the north-south, east-west, and vertical directions, together with the air temperature. The 4D quaternion comprised the 3D wind speed as a vector part (pure quaternion) and air temperature as the scalar part. As illustrated in Fig. 2, the results conform with earlier observations that the class of QSG-IIR outperformed the QLMS, MLMS, and MSG-IIR algorithms, with enhanced convergence properties (see Fig. 3) for the same filter order.

3) Denoising of Lena: We next assessed the AQSG-IIR and QSG-IIR algorithms in a denoising setting. The Lena image was decomposed into its three red-green-blue (RGB) channels to form a 3D time series, thus facilitating the quaternion representation. The image was corrupted with ARMA(1,2) noise $\upsilon(n)$, at a signal-to-noise SNR = 5 dB. The performance was assessed by averaging 20 independent trials, where ARMA(1,2) coefficients were randomly initialized from a Gaussian distribution. Table I shows that the approximate QSG-IIR presented in Section III-B exhibited performance similar to QSG-IIR. The proposed algorithms outperformed QLMS and MLMS.

VI. CONCLUSION

We have introduced a QSG-IIR of hypercomplex data. The forms of gradient terms within QSG-IIR reflect the properties of quaternion al-

[3]The wind data were sampled at 32 Hz and recorded by the 3D WindMaster Gill Instruments anemometer.
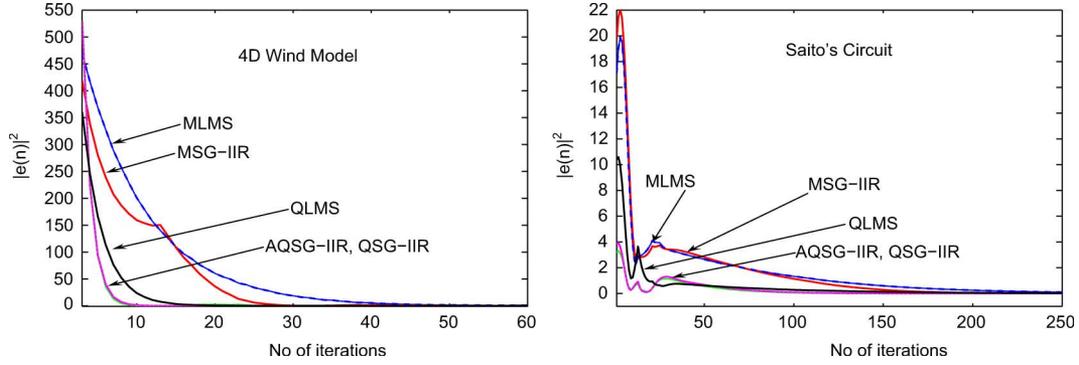
Fig. 3. Learning curves of MLMS, QLMS, QSG-IIR, AQSG-IIR, and MSG-IIR for prediction tasks.

TABLE I
DENOISING OF COLOR LENA IMAGE FOR $L = 2$

Algorithm	Performance Index R_p [dB]	Algorithm	Performance Index R_p [dB]
Multichannel LMS ($2L$)	9.62 (8.70)	MSG-IIR	13.93
Quaternion LMS ($2L$)	11.16 (11.51)	QSG-IIR	16.76
		Approximate QSG-IIR	16.34

gebra, such as the noncommutativity of the product. An approximate version of QSG-IIR with reduced computational complexity has also been introduced, and is followed by a stability analysis. Experiments have been undertaken in a prediction and denoising setting, and illustrate the benefits of QSG-IIR.

APPENDIX
DERIVATION OF THE APPROXIMATE STEPSIZE BOUNDS OF QSG-IIR

The two update (24)–(25) be written in a compact form as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \left[e^*(n) \boldsymbol{\Psi}(n) + \boldsymbol{\Phi}(n) e(n) \right] \qquad (32)$$

where (see the matrix equation at the bottom of the page). Obeying the noncommutativity of the quaternion product, we can substitute

$e^*(n)\boldsymbol{\Psi}(n) = \boldsymbol{\Psi}^*(n)e(n) - 2\Im\{\boldsymbol{\Psi}^*(n)e(n)\}$ into (32) to factor out the error $e(n)$ on the RHS and yield

$$\mathbf{w}(n+1) = \mathbf{w}(n) \\ + \mu \left(\left[\boldsymbol{\Psi}^*(n) + \boldsymbol{\Phi}(n) \right] e(n) - 2\Im\{\boldsymbol{\Psi}^*(n)e(n)\} \right). \quad (33)$$

Since $e(n) = -\mathbf{u}^H[\mathbf{w}(n) - \mathbf{w}_o] = -\mathbf{u}^H(n)\mathbf{v}(n)$, where $\mathbf{w}_o$ denotes the solution at the global minimum of the cost function and $\mathbf{u}(n)$ is an i.i.d. Q-proper signal, we can simplify (33) as shown in (34) at the bottom of the page. However, both $\mathcal{C}_\mathbf{u}$ and $\mathcal{P}_\mathbf{u}$ are not jointly diagonal even for an i.i.d. Q-proper external input $\mathbf{x}(n)$. To derive "modes of convergence," that is, to relate the dynamics of every element of the weight error vector in (31) with the corresponding eigenvalues of the correlation matrices, we need to employ Observation 2. The upper bound on the stepsize is then obtained, when the imaginary part $\Im\{\mathcal{P}_\mathbf{u}\mathbf{v}(n)\}$ is considered a full quaternion, to give $\mathbf{v}(n+1) \approx$

$$\boldsymbol{\Psi}(n) = -\mathbf{u}(n) + \boldsymbol{\zeta}(n) \quad \boldsymbol{\Phi}(n) = 2\mathbf{u}(n) + \boldsymbol{\varphi}(n)$$

$$\boldsymbol{\zeta}(n) = \left[\sum_{m=1}^{M} \phi_{g_1}(n-m), \ldots, \sum_{m=1}^{M} \phi_{g_M}(n-m), \sum_{m=1}^{M} \phi_{h_1}(n-m), \ldots, \sum_{m=1}^{M} \phi_{h_M}(n-m) \right]$$

$$\boldsymbol{\varphi}(n) = \left[\sum_{m=1}^{M} \psi_{g_1}(n-m), \ldots, \sum_{m=1}^{M} \psi_{g_M}(n-m), \sum_{m=1}^{M} \psi_{h_1}(n-m), \ldots, \sum_{m=1}^{M} \psi_{h_M}(n-m) \right].$$

$$\mathbf{v}(n+1) = \mathbf{v}(n) - \mu \left[\left(2E\{\mathbf{u}(n)\mathbf{u}^H(n)\} - E\{\mathbf{u}^*(n)\mathbf{u}^H(n)\} \right)\mathbf{v}(n) + 2\Im\left\{ E\{\mathbf{u}^*(n)\mathbf{u}^H(n)\}\mathbf{v}(n) \right\} \right]$$

$$\mathbf{v}(n+1) = \mathbf{v}(n) - \mu \left[\left(2\mathcal{C}_\mathbf{u} - \mathcal{P}_\mathbf{u} \right)\mathbf{v}(n) + 2\Im\left\{ \mathcal{P}_\mathbf{u}\mathbf{v}(n) \right\} \right]. \qquad (34)$$

$\mathbf{v}(n) - \mu[2\mathcal{C}_\mathbf{u} + \mathcal{P}_\mathbf{u}]\mathbf{v}(n)$. The stepsize bound can now be related with the inverse of the largest (right) eigenvalue of $(2\mathcal{C}_\mathbf{u} + \mathcal{P}_\mathbf{u})$, as is standard in adaptive filtering, to give $0 < \mu < 2/(2\lambda_{c,\max} + \lambda_{p,\max})$.

REFERENCES

[1] P. A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*. New York: Marcel Dekker, 1994.

[2] X. Zhang, "Maxflat fractional delay IIR filter design," *IEEE Trans. Signal Process.*, vol. 57, pp. 2950–2956, 2009.

[3] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction*. New York: Wiley, 2001.

[4] C. Cheong-Took and D. P. Mandic, "The quaternion LMS algorithm for adaptive filtering of hypercomplex real world processes," *IEEE Trans. Signal Process.*, vol. 57, pp. 1316–1327, 2009.

[5] P. Arena, L. Fortuna, G. Muscato, and M. G. Xibilia, "Neural networks in multidimensional domains," in *Lecture Note in Control and Information Sciences*. New York: Springer-Verlag, 1998, vol. 234.

[6] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. New York: Wiley, 2009.

[7] Y. Huang and J. Benesty, *Audio Signal Processing for Next Generation Multimedia Communication Systems*. Boston, MA: Kluwer Academic, 2004.

[8] C. C. Silva and R. D. A Martins, "Polar and axial vectors versus quaternions," *Amer. Assoc. Phys. Teachers*, vol. 70, no. 9, pp. 958–963, 2002.

[9] S. Miron, N. L. Bihan, and J. Mars, "Quaternion-MUSIC for vector-sensor array processing," *IEEE Trans. Signal Process.*, vol. 54, pp. 1218–1229, 2006.

[10] J. P. Ward, *Quaternions and Cayley Numbers: Algebra and Applications*. Boston, MA: Kluwer Academic, 1997.

[11] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, pp. 3473–3482, 1994.

[12] N. N. Vakhania, "Random vectors with values in quaternion Hilbert spaces," *Theory of Probabil. Appl.*, vol. 43, no. 1, pp. 99–115, 1999.

[13] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1017–1029, 2006.

[14] F. Zhang, "Quaternions and matrices of quaternions," *Linear Algebra Appl.*, vol. 251, pp. 21–57, 1997.

[15] J. J. Shynk, "A complex adaptive algorithm for IIR filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1342–1344, 1986.

Minimax Design of IIR Digital Filters Using a Sequential Constrained Least-Squares Method

Xiaoping Lai and Zhiping Lin

Abstract—The minimax design of infinite impulse response (IIR) digital filters is a nonconvex optimization problem, and thus has many local minima. It is shown in this correspondence that a sequential constrained least-squares (SCLS) method has a higher possibility of obtaining better solutions than a direct minimization method when applied to the nonconvex minimax design of IIR filters. We combine the SCLS method with a Steiglitz-McBride (SM) strategy, resulting in a practical design procedure. The positive realness stability condition proposed by Dumitrescu and Niemisto is reformulated as linear inequality constraints and then incorporated in the design procedure. Simulation examples and comparisons with several existing methods demonstrate the effectiveness of the procedure and good performances of the designed filters.

Index Terms—Infinite impulse response (IIR) digital filters, minimax design, sequential constrained least-squares, Steiglitz–McBride (SM) strategy.

I. INTRODUCTION

A major challenge in the optimal design of infinite impulse response (IIR) filters is the nonconvexity of the resulting optimization problem. For constrained least-squares (CLS) and least p-power error designs, several algorithms based on modified versions of the Steiglitz–McBride (SM) strategy [14] were presented to convert the nonconvex problems into a series of standard convex problems such as quadratic programming (QP) [9], [11], [16], and second-order conic programming (SOCP) [4] problems. A Gauss–Newton (GN) method was also presented [6] to transform the nonconvex CLS problem into a series of convex QP problems. The SM and GN methods followed by a classical decent method were adapted successively in [2], resulting in a multistage procedure that obtained better solutions than single or two-stage methods.

Linear programming (LP) was one of early attempts for the minimax design problems of IIR filters. In [13], an exact LP formulation was proposed for the minimization of the filter's maximum squared magnitude. In [1], an approximate LP formulation was obtained by neglecting the denominator of frequency response (FR) approximation error (FR error, for short) and replacing circular constraints with rectangular ones. The LP method was improved in [17] by iteratively updating the FR error's denominator with its previous estimate as in the SM strategy and approximating circular constraints with octagonal ones, resulting in an iterative LP. Iterative SOCP methods were presented in [8] and [10], where the nonconvex FR-error constraints became second-order conic ones due to the first-order approximation of the filter's FR as in the GN method. In [3], the nonconvex minimax problem was first relaxed to an SOCP, and an iterative procedure was

X. P. Lai is with the Institute of Information and Control, Hangzhou Dianzi University, Hangzhou, 310018 China (e-mail: laixp@hdu.edu.cn).

Z. P. Lin is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ezplin@ntu.edu.sg).