This is a repository copy of A Zero-attracting Quaternion-valued Least Mean Square Algorithm for Sparse System Identification.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/94837/

Version: Accepted Version

Proceedings Paper:

https://doi.org/10.1109/CSNDSP.2014.6923898

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
A Zero-attracting Quaternion-valued Least Mean Square Algorithm for Sparse System Identification

Mengdi Jiang, Wei Liu
Communications Research Group
Department of Electronic and Electrical Engineering
University of Sheffield, UK
{mjiang3, w.liu}@sheffield.ac.uk

Yi Li
School of Mathematics and Statistics
University of Sheffield, S3 7RH UK
yili@sheffield.ac.uk

Abstract—Recently, quaternion-valued signal processing has received more and more attention. In this paper, the quaternion-valued sparse system identification problem is studied for the first time and a zero-attracting quaternion-valued least mean square (LMS) algorithm is derived by considering the $l_1$ norm of the quaternion-valued adaptive weight vector. By incorporating the sparsity information of the system into the update process, a faster convergence speed is achieved, as verified by simulation results.

Keywords: quaternion; sparsity; system identification; adaptive filtering; LMS algorithm.

I. INTRODUCTION

In adaptive filtering [1], there is a class of algorithms specifically designed for sparse system identification, where the unknown system only has a few large coefficients while the remaining ones have a very small amplitude so that they can be ignored without significant effect on the overall performance of the system. A good example of them is the zero-attracting least mean square (ZA-LMS) algorithm proposed in [2]. This algorithm can achieve a higher convergence speed, and meanwhile, reduce the steady state excess mean square error (MSE). Compared to the classic LMS algorithm [3], the ZA-LMS algorithm introduces an $l_1$ norm in its cost function, which modifies the weight vector update equation with a zero attractor term.

Recently, the hypercomplex concepts have been introduced to solve problems related to three or four-dimensional signals [4], such as vector-sensor array signal processing [5], [6], color image processing [8] and wind profile prediction [9], [10]. As quaternion-valued algorithms can be regarded as an extension of the complex-valued ones, the adaptive filtering algorithms in complex domain could be extended to the quaternion domain as well, such as the quaternion-valued LMS (QLMS) algorithm in [11].

This work is partially funded by National Grid, UK and will appear in the Proc. of the 9th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), Manchester, UK, July 2014 (submitted in March 2014 and accepted on 18 April 2014).

In this paper, we propose a novel quaternion-valued adaptive algorithm with a sparsity constraint, which is called zero-attracting QLMS (ZA-QLMS) algorithm. The additional constraint is formulated based on the $l_1$ norm. Both the QLMS and ZA-QLMS algorithms can identify an unknown sparse system effectively. However, a better performance in terms of convergence speed is achieved by the latter one.

This paper is organized as follows. A review of basic operations in the quaternion domain is provided in Section II to facilitate the following derivation of the ZA-QLMS algorithm. The proposed ZA-QLMS algorithm is derived in Section III. Simulation results are given in Section IV, and conclusions are drawn in Section V.

II. QUATERNION-VALUED ADAPTIVE FILTERING

A. Basics of Quaternion

Quaternion is a non-commutative extension of the complex number, and normally a quaternion consists of one real part and three imaginary parts, denoted by subscripts $a$, $b$, $c$ and $d$, respectively.

For a quaternion number $q$, it can be described as

$$q = q_a + (q_b i + q_c j + q_d k), \quad (1)$$

where $q_a$, $q_b$, $q_c$, and $q_d$ are real-valued [12], [13]. For a quaternion, when its real part is zero, it becomes a pure quaternion. In this paper, we consider the conjugate operator of $q$ as $q^* = q_a - q_b i - q_c j - q_d k$. The three imaginary units $i$, $j$, and $k$ satisfy

$$ij = k, \quad jk = i, \quad ki = j,$$

$$ijk = i^2 = j^2 = k^2 = -1. \quad (2)$$

As a quaternion has the noncommutativity property, in multiplication, the exchange of any two elements in their order will give a different result. For example, we have $ji = -ij$ rather than $ji = ij$. 


B. Differentiation with Respect to a Quaternion-valued Vector

To derive the quaternion-valued adaptive algorithm, the starting point is the general operation of differentiation with respect to a quaternion-valued vector.

At first, we need to give the definition of differentiation with respect to a quaternion \( q \) and its conjugate \( q^* \). Assume that \( f(q) \) is a function of the quaternion variable \( q \), which is expressed as

\[
    f(q) = f_a + f_i + f_j + f_k
\]

where \( f(q) \) is in general quaternion-valued. The definition of \( \frac{df(q)}{dq} \) can be expressed as

\[
    \frac{df(q)}{dq} = \frac{1}{4} \left[ \frac{\partial f(q)}{\partial q_a} - \frac{\partial f(q)}{\partial q_b} i - \frac{\partial f(q)}{\partial q_c} j - \frac{\partial f(q)}{\partial q_d} k \right].
\]

The derivative of \( f(q) \) with respect to \( q^* \) can be defined in a similar way

\[
    \frac{df(q)}{dq^*} = \frac{1}{4} \left[ \frac{\partial f(q)}{\partial q_a} + \frac{\partial f(q)}{\partial q_b} i + \frac{\partial f(q)}{\partial q_c} j + \frac{\partial f(q)}{\partial q_d} k \right].
\]

With this definition, we can easily obtain

\[
    \frac{\partial q}{\partial q} = 1, \quad \frac{\partial q}{\partial q^*} = -\frac{1}{2}.
\]

Some product rules can be obtained from above formulations, such as the differentiation of quaternion-valued functions to real variables.

Suppose \( f(q) \) and \( g(q) \) are two quaternion-valued functions of the quaternion variable \( q \), and \( q_0 \) is the real variable. Then we can have the following result

\[
    \frac{\partial f(q)g(q)}{\partial q_a} = \frac{\partial f}{\partial q_a} (f_a + i f_b + j f_c + k f_d) g
    + \frac{\partial f_b}{\partial q_a} i + \frac{\partial f_c}{\partial q_a} j + \frac{\partial f_d}{\partial q_a} k
    \]

\[
    = (f_a \frac{\partial q}{\partial q_a} + \frac{\partial f_a}{\partial q_a}) g
    + (f_b \frac{\partial q}{\partial q_a} + \frac{\partial f_b}{\partial q_a}) i
    \]

\[
    + (f_c \frac{\partial q}{\partial q_a} + \frac{\partial f_c}{\partial q_a}) j
    + (f_d \frac{\partial q}{\partial q_a} + \frac{\partial f_d}{\partial q_a}) k
    \]

\[
    = f(q) \frac{\partial g(q)}{\partial q_a} + f(q) \frac{\partial q}{\partial q_a} g(q)
\]

When the quaternion variable \( q \) is replaced by a quaternion-valued vector \( w \), given by

\[
    w = [w_1 \ w_2 \ \cdots \ w_M]^T
\]

\( w_m = a_m + b_m i + c_m j + d_m k, \ m = 1, \ldots, M \), the differentiation of the function \( f(w) \) with respect to the vector \( w \) can be derived using a combination of (4) straightforwardly in the following

\[
    \frac{\partial f}{\partial w} = \frac{1}{4} \left[ \begin{array}{c}
        \frac{\partial f}{\partial a_0} - \frac{\partial f}{\partial q_1} i - \frac{\partial f}{\partial q_2} j - \frac{\partial f}{\partial q_3} k \\
        \frac{\partial f}{\partial a_1} - \frac{\partial f}{\partial q_0} i - \frac{\partial f}{\partial q_3} j - \frac{\partial f}{\partial q_4} k \\
        \vdots \\
        \frac{\partial f}{\partial a_{M-1}} - \frac{\partial f}{\partial q_{M-2}} i - \frac{\partial f}{\partial q_{M-1}} j - \frac{\partial f}{\partial q_M} k
    \end{array} \right].
\]

Similarly, we define \( \frac{\partial f}{\partial w^*} \) as

\[
    \frac{\partial f}{\partial w^*} = \frac{1}{4} \left[ \begin{array}{c}
        \frac{\partial f}{\partial a_0} + \frac{\partial f}{\partial q_1} i + \frac{\partial f}{\partial q_2} j + \frac{\partial f}{\partial q_3} k \\
        \frac{\partial f}{\partial a_1} + \frac{\partial f}{\partial q_0} i + \frac{\partial f}{\partial q_3} j + \frac{\partial f}{\partial q_4} k \\
        \vdots \\
        \frac{\partial f}{\partial a_{M-1}} + \frac{\partial f}{\partial q_{M-2}} i + \frac{\partial f}{\partial q_{M-1}} j + \frac{\partial f}{\partial q_M} k
    \end{array} \right].
\]

Obviously, when \( M = 1 \), (9) and (10) are reduced to (4) and (5), respectively.

III. THE ZERO-ATTRACTION QLMS (ZA-QLMS) ALGORITHM

To improve the performance of the LMS algorithm for sparse system identification, the ZA-QLMS algorithm is derived in this section. To achieve this, similar to [2], in the cost function, we add an \( l_1 \) norm penalty term for the quaternion-valued weight vector \( w[n] \).

For a standard adaptive filter, the output \( y[n] \) and error \( e[n] \) can be expressed as

\[
    y[n] = w^T[n] x[n],
\]

\[
    e[n] = d[n] - w^T[n] x[n],
\]

where \( w[n] \) is the adaptive weight vector with a length of \( L \), \( d[n] \) is the reference signal, \( x[n] = [x[n-1], \ldots, x[n-L]]^T \) is the input sample vector, and \( \cdot^T \) denotes the transpose operation. Moreover, the conjugate form \( e^*[n] \) of the error signal \( e[n] \) is given by

\[
    e^*[n] = d^*[n] - x^H[n] w^*[n],
\]

where \( \gamma \) is a small constant.

Our proposed cost function with a zero attractor term is given by

\[
    J_0[n] = e[n] e^*[n] + \gamma \| w[n] \|_1,
\]

where \( \gamma \) is a small constant.

The gradient of the above cost function with respect to \( w^*[n] \) and \( w[n] \) can be respectively expressed as

\[
    \nabla_w J_0[n] = \frac{\partial J_0[n]}{\partial w^*}
\]

(15)

and

\[
    \nabla_{w^*} J_0[n] = \frac{\partial J_0[n]}{\partial w}
\]

(16)
From [14], [15], we know that the conjugate gradient gives the maximum steepness direction for the optimization surface. Therefore, the conjugate gradient $\nabla_w J_0[n]$ will be used to derive the update of the coefficient weight vector.

Expanding the cost function, we obtain

$$J_0[n] = e[n]e^*[n] + \gamma \|w[n]\|_1,$$

$$= d[n]d^*[n] - d[n]\mathbf{x}^H[n]w^*[n] - w^T[n]\mathbf{x}[n]d^*[n] + w^T[n]\mathbf{x}[n]\mathbf{x}^H[n]w^*[n] + \gamma \|w[n]\|_1.$$  \hspace{1cm} (17)

Furthermore,

$$\frac{\partial J_0[n]}{\partial w^*} = \frac{\partial (e[n]e^*[n] + \gamma \|w[n]\|_1)}{\partial w^*},$$

$$= \frac{\partial (e[n]e^*[n] + \gamma \|w[n]\|_1)}{\partial w^*}.$$  \hspace{1cm} (18)

Details of the derivation process for the gradient are shown in the following

$$\frac{\partial (d[n]d^*[n])}{\partial w^*[n]} = 0$$ \hspace{1cm} (19)

$$\frac{\partial (d[n]\mathbf{x}^H[n]w^*[n])}{\partial w^*[n]} = d[n]\mathbf{x}^*[n]$$ \hspace{1cm} (20)

$$\frac{\partial (w^T[n]\mathbf{x}[n]d^*[n])}{\partial w^*[n]} = -\frac{1}{2} d[n]\mathbf{x}^*[n]$$ \hspace{1cm} (21)

$$\frac{\partial (w^T[n]\mathbf{x}[n]\mathbf{x}^H[n]w^*[n])}{\partial w^*[n]} = \frac{1}{2} w^T[n]\mathbf{x}[n]\mathbf{x}^*[n]$$ \hspace{1cm} (22)

Moreover, the last part of the gradient of cost function is given by

$$\frac{\partial (\gamma \|w[n]\|_1)}{\partial w^*} = \frac{1}{4} \gamma \cdot sgn(w[n]),$$ \hspace{1cm} (23)

where the symbol $sgn$ is a component-wise sign function that is defined as

$$sgn(x) = \begin{cases} x/|x| & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Combining the above results, the final gradient can be obtained as follows

$$\nabla_w J_0[n] = -\frac{1}{2} e[n]\mathbf{x}^*[n] + \frac{1}{4} \gamma \cdot sgn(w[n]).$$ \hspace{1cm} (24)

With the general update equation for the weight vector

$$w[n+1] = w[n] - \mu \nabla_w J_0[n],$$ \hspace{1cm} (25)

where $\mu$ is the step size, we arrive at the following update equation for the proposed ZA-QLMS algorithm

$$w[n+1] = w[n] + \mu (e[n]\mathbf{x}^*[n]) - \rho \cdot sgn(w[n]),$$ \hspace{1cm} (26)

where $\rho \cdot \gamma$. The last term represents the zero attractor, which enforces the near-zero coefficients to zero and therefore accelerates the convergence process when majority of the system coefficients are nearly zero in a sparse system.

Note that equation (26) will be reduced to the normal QLMS algorithm without the zero attractor term, given by [11]

$$w[n+1] = w[n] + \mu (e[n]\mathbf{x}^*[n]).$$ \hspace{1cm} (27)

IV. SIMULATION RESULTS

In this part, simulations are performed for sparse system identification using the proposed algorithm in comparison with the QLMS algorithm. Two different sparse systems are considered corresponding to Scenario One and Scenario Two in the following. The input signal to the adaptive filter is colored and generated by passing a quaternion-valued white gaussian signal through a randomly generated filter. The noise part is quaternion-valued white Gaussian and added to the output of the unknown sparse system, with a 30dB signal to noise ratio (SNR) for both scenarios.

A. Scenario One

For the first scenario, the parameters are: the step size $\mu$ is $3 \times 10^{-7}$; the unknown sparse FIR filter length $L$ is 32, with 4 non-zero coefficients at the 2nd, 8th, 16th and 31st taps, and its magnitude of the impulse response is shown in Fig. 1. The coefficient of the zero attractor $\rho$ is $5 \times 10^{-7}$. The learning curve obtained by averaging 100 runs of the corresponding algorithm is given in Fig. 2 where we can see that the ZA-QLMS algorithm has achieved a faster convergence speed than the QLMS algorithm when they both reach a similar steady state.

B. Scenario Two

For this case, length of the unknown FIR filter is reduced to 16, still with 4 active taps. The parameters are: step size $\mu$
is $2 \times 10^{-7}$ and the value of $\rho$ is $2 \times 10^{-7}$. The results are shown in Fig. 3. Again we see that the ZA-QLMS algorithm has a faster convergence speed and has even converged to a lower steady state error in this specific scenario.

V. CONCLUSION

In this paper, a quaternion-valued adaptive algorithm has been proposed for more efficient identification of unknown sparse systems. It is derived by introducing an $l_1$ penalty term in the original cost function and the resultant zero-attracting quaternion-valued LMS algorithm can achieve a faster convergence rate by incorporating the sparsity information of the system into the update process. Simulation results have been provided to show the effectiveness of the new algorithm.

REFERENCES