Quaternion VAR Modelling and Estimation

P. Ginzberg and A. T. Walden, Senior Member, IEEE

Abstract

Quaternion vector autoregression (VAR) modeling is a natural extension of real and complex VAR. We demonstrate how a quaternion VAR can be treated as a special case of structured real VAR. We show that generalized least squares and (under Gaussianity) maximum likelihood estimation of the parameters reduces to simple least squares estimation if the innovations are quaternion proper.

I. INTRODUCTION

While real-valued scalar and vector autoregressive processes are ubiquitous in science and econometrics, extensions to complex-valued and quaternion-valued processes are readily found and increasing. Complex-valued AR processes [15] have been applied to temperature forecasting [8], character recognition [14] and shape recognition and extraction [18], [19]. The fading of telecommunication signals is simulated using complex-valued vector autoregressive (VAR) processes in [1]. AR modelling has been extended naturally to the quaternion domain, following other standard complex signal processing tools, such as the polar and singular value decompositions [3], [21], partial least squares and multivariate linear regression [20]. An example quaternion-valued AR process was used in [6], and an adaptive quaternion AR filter is applied to short-term wind forecasting in [4].

An equivalence between the complex-valued or quaternion forms and structured real-valued forms provides these models with a convenient theoretical grounding. In this paper we consider the most complicated of these number fields, the quaternions, and the VAR process model, to show that generalized least squares (GLS) estimation of the model parameters reduces to least squares (LS) estimation if the process is a proper quaternion VAR process; moreover, given Gaussianity, these two estimators are also identical to the maximum likelihood estimator. If $\epsilon_t$ is any $n$-length random quaternion-valued innovations vector of the VAR process, the process is a proper quaternion VAR process if the associated $4n \times 4n$ real-valued covariance matrix has quaternion structure.

The less difficult complex case may be treated similarly and details are provided as appropriate.

II. PRELIMINARIES

A. Quaternion-Valued Matrices

A quaternion matrix $Q \in \mathbb{H}^{m \times n}$, takes the form $Q = A + Bi + Cj + Dk$, $(A, B, C, D \in \mathbb{R}^{m \times n})$, where $1, i, j, k$ are the 4 basis elements which satisfy $i^2 = j^2 = k^2 = ijk = -1$, $ij = -ji = k$.

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Paul Ginzberg and Andrew Walden are both at the Department of Mathematics, Imperial College London, 180 Queen’s Gate, London SW7 2BZ, UK. (e-mail: paul.ginzberg05@imperial.ac.uk and a.walden@imperial.ac.uk)
\( jk = -kj = i, \ ki = -ik = j \). The conjugate of \( Q \) is \( Q^* = A - Bi - Cj - Dk \). We denote the conjugate transpose by \( Q^H = (Q^*)^T \). The product of two matrices \( Q \in \mathbb{H}^{m \times n} \) and \( S \in \mathbb{H}^{n \times p} \) is \( P = QS \) with \((i, j)^{th}\) element \( p_{i, j} = \sum_{\ell=1}^n q_{i, \ell} s_{\ell, j} \). Quaternion multiplication, like matrix multiplication, is associative, distributive but not commutative. \( ||Q||^2 \) denotes the Frobenius norm given by

\[
||Q||^2 = \text{tr}\{Q^HQ\} = \sum_{i,j} |q_{i,j}|^2 \tag{1}
\]

where \( |q_{i,j}|^2 = q_{i,j}q_{i,j}^* \).

**B. Quaternion to Real Transforms**

To map quaternion matrices to real matrices, we will introduce two transformations and certain of their properties.

**Definition 1:** Let \( Q \in \mathbb{H}^{m \times n} \). Define

\[
\mathcal{V}(Q) = \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} \in \mathbb{R}^{4m \times n}
\]

and

\[
\tilde{Q} = \begin{bmatrix} \mathcal{V}(Q) & \mathcal{V}(Q_1) & \mathcal{V}(Q_2) & \mathcal{V}(Q_3) \end{bmatrix} = \begin{bmatrix} A & -B & -C & -D \\ B & A & -D & C \\ C & D & A & -B \\ D & -C & B & A \end{bmatrix} \in \mathbb{R}^{4m \times 4n}. \tag{3}
\]

Matrices of the form (4) are said to have quaternion structure. Note that for the \( n \times n \) identity matrix \( I_n \), we have \( \tilde{I}_n = I_{4n} \).

**Remark 1:** We note that if \( Q_1, Q_2 \in \mathbb{H}^{m \times n} \) and \( Q_3 \in \mathbb{H}^{n \times p} \) then [10, section IV]

\[
\tilde{Q}_1 + \tilde{Q}_2 = \tilde{Q}_1 + \tilde{Q}_2 \quad \text{and} \quad \tilde{Q}_1 \tilde{Q}_3 = \tilde{Q}_1 \tilde{Q}_3. \tag{5}
\]

In particular, if \( Q_1 \) is invertible then \( \tilde{Q}_1^{-1} \tilde{Q}_1 = I_{4n} \) and hence

\[
\tilde{Q}_1^{-1} = \tilde{Q}_1^{-1}. \tag{6}
\]

Furthermore, \( \tilde{Q}_1^H = \tilde{Q}_1^T \), and in particular

\[
||\tilde{Q}_1||^2_{\mathbb{R}^{4m \times 4n}} = \text{tr} \left( \tilde{Q}_1^H \tilde{Q}_1 \right) = 4 \text{tr} \left( Q_1^HQ_1 \right) = 4 ||Q_1||^2_{\mathbb{H}^{m \times n}}, \tag{7}
\]

where we make explicit the domain on which the Frobenius norm is applied.

**Remark 2:** Let \( Q \in \mathbb{H}^{m \times n} \) and \( q \in \mathbb{H}^n \). Then

\[
\mathcal{V}(Qq) = \tilde{Q} \mathcal{V}(q). \tag{8}
\]
Definition 2: Let \( q = a + bi + cj + dk \). Define
\[
\tilde{q}^R = \begin{bmatrix}
a & -b & -c & -d \\
b & a & d & -c \\
c & -d & a & b \\
d & c & -b & a \\
\end{bmatrix}.
\]

Remark 3: \( \tilde{q}^R \) is the real matrix corresponding to multiplication by \( q \) on the right: for \( \lambda, q \in \mathbb{H} \),
\[
\tilde{q}^R \psi(\lambda) = \psi(\lambda q).
\]

Denote by \( A \otimes B \) the Kronecker product of two real matrices. Let \( Q \in \mathbb{H}^{m \times n} \) and \( \lambda \in \mathbb{H} \). Then
\[
\psi(Q \lambda) = (\lambda^R \otimes I_m) \psi(Q).
\]

Remark 4: Let \( M \in \mathbb{H}^{n \times n} \), then \( \tilde{M} \in \mathbb{R}^{4n \times 4n} \) commutes with \( \lambda^R \otimes I_n \).

Proof: For any \( q \in \mathbb{H}^n \) we have from (9) that
\[
(\lambda^R \otimes I_n) \psi(q) = \psi(q \lambda)
\]
and so, using (8), and the associativity of quaternion product,
\[
\tilde{M}(\lambda^R \otimes I_n) \psi(q) = \tilde{M} \psi(q \lambda) = \psi(M(q \lambda)) = \psi((Mq) \lambda).
\]

But using (9) again, followed by (8)
\[
\psi((Mq) \lambda) = (\lambda^R \otimes I_n) \psi(Mq) = (\lambda^R \otimes I_n) \tilde{M} \psi(q),
\]
so \( \tilde{M} \) commutes with \( \lambda^R \otimes I_n \). Note that \( \psi(q) \) is an arbitrary vector in \( \mathbb{R}^{4n} \).

Remark 5: Denote by \( \text{vec} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{mn} \) the operator which stacks the columns of a matrix. We can write \( \tilde{\Theta} : \mathbb{H}^{m \times n} \rightarrow \mathbb{R}^{4m \times 4n} \) as a linear transformation between the vector spaces \( \mathbb{R}^{4mn} \) and \( \mathbb{R}^{4m^2 - 4n} \).

\[
\text{vec}(\tilde{Q}) = \Theta^{(m,n)} \text{vec}(\psi(Q)),
\]
where
\[
\Theta^{(m,n)} = \begin{bmatrix}
I_n \otimes I_4 \otimes I_m \\
I_n \otimes i^R \otimes I_m \\
I_n \otimes j^R \otimes I_m \\
I_n \otimes k^R \otimes I_m
\end{bmatrix}.
\]

C. Complex-Valued Matrices

For the complex case we have \( Q = A + Bi \) and define instead
\[
\psi(Q) = \begin{bmatrix} A \\ B \end{bmatrix},
\]
\[
\tilde{Q} = \begin{bmatrix} A & -B \\ B & A \end{bmatrix}.
\]
\[ \tilde{q}^R = \tilde{q}, \]

and

\[
\Theta^{(m,n)} = \begin{bmatrix}
I_n \otimes I_2 \otimes I_m \\
I_n \otimes \mathbb{I}_R \otimes I_m
\end{bmatrix}.
\]

So for the complex-valued case, in what follows we can simply replace “quaternion” with “complex”, \( \mathbb{H} \) with \( \mathbb{C} \), 4 with 2 and use \( V, \cdot, \cdot^R \) as detailed above in this subsection.

### III. Quaternion VAR as a Structured Real VAR

**Definition 3:** Let \( A_1, \ldots, A_p \in \mathbb{R}^{n \times n}, \mu \in \mathbb{R}^n \), and let \( \epsilon_t \in \mathbb{R}^n \) be a sequence of uncorrelated zero-mean random vectors with a common covariance matrix \( \Sigma_\epsilon = E\{\epsilon_t \epsilon_t^T\} \). The process

\[
y_t = \mu + A_1 y_{t-1} + \ldots + A_p y_{t-p} + \epsilon_t
\]

is a real VAR process \( AR^R_n(p) \).

**Definition 4:** Let \( q \in \mathbb{H}^n \) so that \( q = a + bi + cj + dk \), where \( a, b, c, d \in \mathbb{R}^n \). For \( q \) a random variable with mean zero, its second-order properties are given by the \( 4n \times 4n \) real covariance matrix \( \Sigma_{V(q)} = E\{V(q) V(q)^T\} \), namely,

\[
\Sigma_{V(q)} = \begin{bmatrix}
\Sigma_a & \Sigma_{a,b} & \Sigma_{a,c} & \Sigma_{a,d} \\
\Sigma_{b,a} & \Sigma_b & \Sigma_{b,c} & \Sigma_{b,d} \\
\Sigma_{c,a} & \Sigma_{c,b} & \Sigma_c & \Sigma_{c,d} \\
\Sigma_{d,a} & \Sigma_{d,b} & \Sigma_{d,c} & \Sigma_d
\end{bmatrix},
\]

where \( \Sigma_{a,b} = E\{ab^T\} \) and \( \Sigma_a = E\{aa^T\} \), etc. Then \( q \) is \( \mathbb{H} \)-proper, or just proper, if and only if (e.g., [7, p. 3026])

\[
\Sigma_{V(q)} = \begin{bmatrix}
\Sigma_a & -\Sigma_{b,a} & -\Sigma_{c,a} & -\Sigma_{d,a} \\
-\Sigma_{b,a} & \Sigma_a & \Sigma_{c,a} & \Sigma_{d,a} \\
-\Sigma_{c,a} & \Sigma_{d,a} & \Sigma_a & -\Sigma_{b,a} \\
-\Sigma_{d,a} & -\Sigma_{c,a} & -\Sigma_{b,a} & \Sigma_a
\end{bmatrix},
\]

i.e., if and only if \( \Sigma_{V(q)} \) has quaternion structure — see (4).

**Definition 5:** Let \( A_1, \ldots, A_p \in \mathbb{H}^{n \times n}, \mu \in \mathbb{H}^n \), and let \( \epsilon_t \in \mathbb{H}^n \) be a sequence of uncorrelated zero-mean proper random vectors with a common covariance matrix \( \Sigma_\epsilon = E\{\epsilon_t \epsilon_t^H\} \). The process

\[
q_t = \mu + A_1 q_{t-1} + \ldots + A_p q_{t-p} + \epsilon_t
\]

is a proper quaternion VAR process, i.e., proper \( AR^\mathbb{H}_n(p) \).

**Proposition 1:** Let \( q_t \) be the proper \( AR^\mathbb{H}_n(p) \) process of Definition 5. Then \( y_t = V(q_t) \) is an \( AR^\mathbb{R}_n(p) \) process

\[
y_t = V(\mu) + \tilde{A}_1 y_{t-1} + \ldots + \tilde{A}_p y_{t-p} + \tilde{V}(\epsilon_t).
\]

Furthermore, the coefficient matrices \( \tilde{A}_1, \ldots, \tilde{A}_p \) and the innovations covariance \( \Sigma_{V(\epsilon)} \in \mathbb{R}^{4n \times 4n} \) have the quaternion structure (4).
Conversely, for any AR$_{2n}^{R}(p)$ process $y_t$ with quaternion-structured coefficients and innovations covariance, $q_t = \mathbf{V}^{-1}(y_t)$ is a proper AR$_{2n}^{H}(p)$ process.

Proof: Eqn. (12) follows immediately from (8). From Definition 4, $e_t$ is proper if and only if $\Sigma_{\mathbf{V}(e)}$ has quaternion structure. Indeed, from [7, p. 3026], we can deduce that $\Sigma_{\mathbf{V}(e)} = \frac{1}{4} \Sigma_e$.  

The real form $\mathbf{V}(q_t)$ of a quaternion VAR allows us to effortlessly translate definitions and theoretical results for real VARs to quaternion VARs. For example, we can define $q_t$ to be stationary or Gaussian if and only if $\mathbf{V}(q_t)$ is likewise.

IV. WIDELY LINEAR AND SEMI-WIDELY LINEAR QUATERNION VAR

Similarly to how widely linear processing of complex signals deals with improper signals by allowing additional linear operations on its complex conjugate [16], a unified treatment of improper quaternion signals can be obtained by allowing additional linear operations on the three involutions $q^{(n)} = -\eta q^n$, $\eta = i, j, k$ [20].

Definition 6: Let $A_1, B_1, C_1, D_1, \ldots, A_p, B_p, C_p, D_p \in \mathbb{H}^{n \times n}$, $\mu \in \mathbb{H}^n$, and let $e_t \in \mathbb{H}^n$ be a sequence of uncorrelated zero-mean innovations with a common covariance. The process

$$q_t = \mu + A_1 q_{t-1} + B_1 q_{t-1}^{(i)} + C_1 q_{t-1}^{(j)} + D_1 q_{t-1}^{(k)} + \ldots + D_p q_{t-p}^{(k)} + e_t$$

is a widely linear quaternion VAR process, i.e., widely linear AR$_{2n}^{H}(p)$.

Widely linear quaternion AR modelling has been used for example in [9] for wind forecasting.

Definition 7: If in Definition 6 we have $B_i = C_i = 0_{n \times n}$ for $i = 1, \ldots, p$, then we say that $q_t$ is a $\mathbb{C}^k$-semi-widely linear quaternion VAR process. If furthermore the innovations $e_t$ are $\mathbb{C}^k$-proper (see [20]), then $q_t$ is a $\mathbb{C}^k$-proper quaternion VAR process i.e., $\mathbb{C}^k$-proper AR$_{2n}^{H}(p)$.

Remark 6: $q_t$ is widely linear AR$_{2n}^{H}(p)$ if and only if $y_t = \mathbf{V}(q_t)$ is AR$_{4n}^{R}(p)$. In other words, widely linear quaternion VAR processes are in 1-to-1 correspondence with unrestricted real VAR processes.

Remark 7: $q_t = a_t + b_t i + c_t j + d_t k$ is $\mathbb{C}^k$-proper AR$_{2n}^{H}(p)$ if and only if

$$y_t = \begin{bmatrix} a_t + b_t i \\ d_t + c_t i \end{bmatrix}$$

is proper AR$_{2n}^{R}(p)$. In other words, $\mathbb{C}^k$-proper quaternion VAR processes are in 1-to-1 correspondence with proper complex VAR processes.

V. EQUIVALENCE BETWEEN LS, GLS AND ML ESTIMATION

A. Background

It was noted in [5] that quaternion linear least squares estimation is equivalent to structured real linear least squares estimation. By Proposition 1, a proper quaternion VAR (proper AR$_{2n}^{H}(p)$) process, is equivalent to a real VAR (AR$_{4n}^{R}(p)$) process with linear constraints on both the regression coefficients and the innovations covariance matrix.
In econometrics a VAR model with zero constraints on the parameters is called a subset VAR model. An extensive discussion is given in [12, Sections 5 and 9], covering parameter estimation, asymptotic estimator distributions and hypothesis testing. The case of linear constraints, which applies to quaternion structure, is also treated. Although the LS and GLS estimators of the regression coefficients are equal for an unrestricted real VAR model, this is no longer true in general when linear constraints are imposed.

[17] show that for a proper AR\(_1\)\((p)\) process, LS and MLE estimation are equivalent under Gaussianity. We will show that the LS estimator and the GLS estimator for the regression parameter matrix

\[
B = \left[ \begin{array}{cccc}
\mu & A_1 & \ldots & A_p
\end{array} \right] \in \mathbb{H}^{n \times (np+1)}
\] (14)
of a proper AR\(_n\)\((p)\) process are equal, and that under the further assumption of Gaussian innovations, the MLE, the LS estimator and the GLS estimator are equal.

B. Equality of Least Squares and Generalized Least Squares

**Proposition 2:** Consider the standard real-valued multivariate linear regression model

\[
y = X\beta + e
\]

with error covariance matrix \(\Sigma_e\). If there exists a matrix \(S\) such that \(\Sigma_e^{-1}X = XS\), then the LS estimator

\[
\hat{\beta}_{LS} = (X^TX)^{-1}X^Ty
\] (15)

and the GLS estimator

\[
\hat{\beta}_{GLS} = (X^T\Sigma^{-1}_eX)^{-1}X^T\Sigma^{-1}_ey
\] (16)

are equal. We assume for simplicity that \(X^TX\), \(\Sigma_e\) and \(S\) are invertible.

**Proof:** From (16),

\[
\hat{\beta}_{GLS} = ((\Sigma_e^{-1}X)^TX)^{-1}(\Sigma_e^{-1}X)^Ty
\]
\[
= (S^TX^TX)^{-1}S^TX^Ty
\]
\[
= (X^TX)^{-1}(S^T)^{-1}S^TX^Ty
\]
\[
= \hat{\beta}_{LS}.
\]

As shown in [22, Theorem 2], the existence of \(S\) is also a necessary condition, and by using the Moore-Penrose generalized inverse all invertibility assumptions can be dropped. \(\blacksquare\)
C. The AR Parameter Estimation Problem

Let \( q_t \) be proper AR
\(^p\)
\(_n\) as in Definition 5. In addition to (14) define

\[
Q = \begin{bmatrix}
q_1 & \ldots & q_N \\
1 & 1 & \ldots & 1 \\
q_0 & q_1 & \ldots & q_{N-1} \\
q_{-1} & q_0 & \ldots & q_{N-2} \\
& \vdots & \ddots & \vdots \\
q_{-p+1} & q_{-p+2} & \ldots & q_{N-p} \\
\end{bmatrix} \in \mathbb{H}^{n \times N}
\]

\[
W = \begin{bmatrix}
\epsilon_1 & \ldots & \epsilon_N \\
\end{bmatrix} \in \mathbb{H}^{n \times N}
\]

so that

\[
Q = BW + \epsilon,
\]

which we could use for parameter estimation. Alternatively, we could consider instead the real-valued approach using \( V(Q) = V(BW) + V(\epsilon) \); to do this we proceed as follows. Let \( y_t = V(q_t) \) and define

\[
B^* = \begin{bmatrix}
V(\mu) & \tilde{A}_1 & \ldots & \tilde{A}_p \\
\end{bmatrix} \in \mathbb{R}^{4n \times (4np+1)}
\]

\[
Y = V(Q) = \begin{bmatrix}
y_1 & \ldots & y_N \\
1 & 1 & \ldots & 1 \\
y_0 & y_1 & \ldots & y_{N-1} \\
y_{-1} & y_0 & \ldots & y_{N-2} \\
& \vdots & \ddots & \vdots \\
y_{-p+1} & y_{-p+2} & \ldots & y_{N-p+1} \\
\end{bmatrix} \in \mathbb{R}^{(4np+1) \times N}
\]

\[
Z = \begin{bmatrix}
1 & 1 & \ldots & 1 \\
y_0 & y_1 & \ldots & y_{N-1} \\
y_{-1} & y_0 & \ldots & y_{N-2} \\
& \vdots & \ddots & \vdots \\
y_{-p+1} & y_{-p+2} & \ldots & y_{N-p+1} \\
\end{bmatrix} \in \mathbb{R}^{(4np+1) \times N}
\]

\[
E = V(\epsilon) = \begin{bmatrix}
V(\epsilon_1) & \ldots & V(\epsilon_N) \\
\end{bmatrix} \in \mathbb{R}^{4n \times N},
\]

so that \( Y = B^*Z + E \). To see this consider the case \( p = 1 \) for simplicity. Then

\[
BW = \mu + A_1\begin{bmatrix}
q_0 & \ldots & q_{N-1} \\
\end{bmatrix}
\]

so

\[
V(BW) = V(\mu) + V(A_1\begin{bmatrix}
q_0 & \ldots & q_{N-1} \\
\end{bmatrix}).
\]

But from (8) \( V(A_1q_0) = \tilde{A}_1V(q_0) = \tilde{A}_1y_0 \), so that

\[
V(BW) = V(\mu) + \tilde{A}_1\begin{bmatrix}
y_0 & \ldots & y_{N-1} \\
\end{bmatrix} = B^*Z.
\]

Since also \( V(Q) = Y \) and \( V(\epsilon) = E \) we get the desired result

\[
V(Q) = V(BW) + V(\epsilon) = \tilde{B}V(W) + V(\epsilon) = Y = B^*Z + E.
\]
We wish to estimate the quaternion parameter matrix $\mathbf{B}$ by $\hat{\mathbf{B}}$. Equivalently, we may estimate the structured parameter matrix $\mathbf{B}^*$ for the real VAR process $\mathbf{y}_t$ by $\hat{\mathbf{B}}^*$, or still equivalently, we may estimate the quaternion structured real parameter matrix $\hat{\mathbf{B}}$ by $\hat{\mathbf{B}}$. Let $\hat{\mathbf{E}} = \mathbf{Y} - \hat{\mathbf{B}}^* \mathbf{Z}$. Then the LS estimate is the choice of $\mathbf{B}^*$ minimizing the sum of squared errors $||\mathbf{E}||^2 = \text{tr}(\hat{\mathbf{E}}^T \hat{\mathbf{E}})$, whereas the GLS estimate minimizes $\text{tr}(\hat{\mathbf{E}}^T \Sigma_{\nu(e)}^{-1} \hat{\mathbf{E}})$.

Let $\hat{\epsilon} = \mathbf{Q} - \hat{\mathbf{B}} \mathbf{W}$. The LS estimation problem can be dealt with in the quaternion domain and requires minimizing $||\hat{\epsilon}||^2$. This unrestricted minimization can be summarized as $\hat{\mathbf{B}} = \text{argmin}_{\mathbf{B} \in \mathbb{HH}^{n \times (np+1)}}(||\mathbf{Q} - \mathbf{B} \mathbf{W}||^2)$. Similarly GLS estimation requires minimizing and $\text{tr}(\hat{\epsilon}^H \Sigma_{\nu(e)}^{-1} \hat{\epsilon})$.

A quaternion-domain approach will obviously lead to the same results. However such an approach raises two problems. Similarly to how $\mathbb{CR}$ calculus was used in [17] to prove the equivalence between LS and ML estimation for $\text{AR}_t^\mathbb{C} (p)$ processes, calculus-based minimisation in the quaternion domain would require the use of $\mathbb{HR}$ calculus [13]. But notably, the importance of our assumption of proper innovations would not be obvious. If the innovations are improper but the function being minimized depends only on the quaternion covariance matrix, then additional second order information is being discarded. This is equivalent to using real GLS estimation with a misspecified real covariance $\Sigma_{\nu(e)}$ with quaternion structure, as though we had mistakenly assumed that the process was proper. Such a quaternion GLS estimate would still reduce to a LS estimate, but the misspecified covariance would be expected to lead to a loss of efficiency. This latter issue can be overcome in the quaternion domain by using augmented quaternion vectors [20].

**Theorem 1:** The LS estimator and the GLS estimator for the regression parameter matrix $\mathbf{B}$ of a proper $\text{AR}_n^\mathbb{H} (p)$ process are equal.

**Proof:** We will make use of two results involving the Kronecker product [2, p. 249].

- Let $\mathbf{U} \in \mathbb{R}^{n \times m}$, $\mathbf{V} \in \mathbb{R}^{m \times \ell}$, and $\mathbf{P} \in \mathbb{R}^{\ell \times k}$, then
  \[
  \text{vec}(\mathbf{U} \mathbf{V} \mathbf{P}) = [\mathbf{P}^T \otimes \mathbf{U}] \text{vec}(\mathbf{V}).
  \]  

- Let $\mathbf{U} \in \mathbb{R}^{n \times m}$, $\mathbf{V} \in \mathbb{R}^{\ell \times k}$, $\mathbf{P} \in \mathbb{R}^{m \times q}$, and $\mathbf{M} \in \mathbb{R}^{k \times p}$, then
  \[
  [\mathbf{U} \otimes \mathbf{V}] [\mathbf{P} \otimes \mathbf{M}] = \mathbf{UP} \otimes \mathbf{VM}.
  \]  

We can write the linear constraints on $\hat{\mathbf{B}}$ in the form

\[
\text{vec}(\hat{\mathbf{B}}) = \Theta^{(n, np+1)} \gamma
\]

where $\gamma = \text{vec}(\mathbf{V}(\mathbf{B}))$ is a vector of $4n(np + 1)$ unrestricted real parameters.

Then the unrestricted estimation of $\gamma$ is defined by

\[
\text{vec}(\mathbf{Y}) = \text{vec}(\mathbf{V}(\mathbf{B} \mathbf{W})) + \text{vec}(\mathbf{V}(\epsilon)) = \text{vec}(\mathbf{I}_{4n} \hat{\mathbf{B}} \mathbf{V}(\mathbf{W})) + \text{vec}(\mathbf{E}) = [\mathbf{V}(\mathbf{W})^T \otimes \mathbf{I}_{4n}] \text{vec}(\hat{\mathbf{B}}) + \text{vec}(\mathbf{E}) = [\mathbf{V}(\mathbf{W})^T \otimes \mathbf{I}_{4n}] \Theta^{(n, np+1)} \gamma + \mathbf{e} = \mathbf{X} \gamma + \mathbf{e},
\]
where we have made use of (18), and define $X = [\mathcal{V}(W)^T \otimes I_{4n}] \Theta^{(n, np+1)}$, and $e = \text{vec}(E)$ with covariance matrix $\Sigma_e = I_N \otimes \Sigma_{\psi(e)}$.

Now by assumption $\Sigma_{\psi(e)}$ is invertible and has quaternion structure, so by (6) $\Sigma_{\psi(e)}^{-1}$ has quaternion structure. Hence by Remark 4 it commutes with $\mathbb{I}^R \otimes I_n$, $\mathbb{J}^R \otimes I_n$ and $\mathbb{K}^R \otimes I_n$ in $\Theta^{(n, np+1)}$ defined by (10). Hence,

$$[I_n \otimes I_{np+1} \otimes \Sigma_{\psi(e)}^{-1}] \Theta^{(n, np+1)} = \Theta^{(n, np+1)} [I_{np+1} \otimes \Sigma_{\psi(e)}^{-1}] .$$

Using the above and (19),

$$\Sigma_e^{-1} X = [I_N \otimes \Sigma_{\psi(e)}^{-1}] [\mathcal{V}(W)^T \otimes I_{4n}] \Theta^{(n, np+1)}$$

$$= [\mathcal{V}(W)^T \otimes \Sigma_{\psi(e)}^{-1}] [I_{4n}] [I_{np+1} \otimes \Sigma_{\psi(e)}^{-1}] \Theta^{(n, np+1)}$$

$$= [\mathcal{V}(W)^T \otimes I_{4n}] [I_{np+1} \otimes \Sigma_{\psi(e)}^{-1}] [I_{np+1} \otimes \Sigma_{\psi(e)}^{-1}]$$

$$= XS,$$

where $S = I_{np+1} \otimes \Sigma_{\psi(e)}^{-1}$. By Proposition 2, the LS and GLS estimates are equal. \hfill \blacksquare

**Corollary 1:** Under the assumption of Gaussian innovations the MLE, the LS estimator and the GLS estimator for the parameter $B$ of a proper $\text{AR}_n^H(p)$ process are equal.

**Proof:** Under a Gaussianity assumptions, the MLE of $B$ is equal to the GLS estimator, except that the covariance matrix of the innovations $\Sigma_{\psi(e)}$ is replaced by its maximum likelihood estimate $\hat{\Sigma}_{\psi(e)}$ [12, eqn. 5.2.17]. In this case, because we assume that $\Sigma_{\psi(e)}$ has quaternion structure, $\hat{\Sigma}_{\psi(e)}$ will be a restricted MLE with quaternion structure. Since Theorem 1 has shown that the GLS is the same as the LS estimator, and the latter does not depend on $\Sigma_{\psi(e)}$, replacing $\Sigma_{\psi(e)}$ by $\hat{\Sigma}_{\psi(e)}$ has no effect. \hfill \blacksquare

**Remark 8:** For scalar autoregressive processes, propriety of the innovations covariance is equivalent to circularity $\Sigma_{\psi(e)} = \sigma^2 I$. In this case the equivalence between LS and GLS estimation is trivial.

**Remark 9:** In practice, it is both simpler and faster to compute the LS estimator of $B$ using a standard real least squares algorithm on the real representation (such as $\hat{B} = \hat{Q}/\hat{W}$ in Matlab) rather than using (15).

Indeed, since the unrestricted solution to

$$\text{argmin}_{B \in \mathbb{R}^{4n \times (np+1)}}(||\hat{Q} - \hat{B}\hat{W}||^2)$$

(22)

is given by $\hat{B} = \hat{Q}\hat{W}^T [\hat{W}\hat{W}^T]^{-1}$ [12, p. 72], it has quaternion structure by Remark 1, and hence it is also the restricted solution with quaternion structure.

Alternatively, the complex representation of quaternion matrices and complex least squares can be used, as in [11].

**Remark 10:** Let $q_t$ be an $\text{AR}_n^H(p)$ process and consider the matrix-valued time series $\tilde{q}_t$. By (3), the four columns of this matrix form can be interpreted as an ensemble of four $\text{AR}_4^R(p)$ time series.
\( \mathcal{V}(q_l), \mathcal{V}(q_i), \mathcal{V}(q_j), \mathcal{V}(q_k) \), having shared regression parameters. The constant term is different for each of the four time series, and is given by \( \mathcal{V}(\mu), \mathcal{V}(\mu_i), \mathcal{V}(\mu_j), \mathcal{V}(\mu_k) \) respectively. It can be shown that (22) gives the unrestricted ensemble LS estimate of the parameters for these four real time series by noting that the first \( N \) columns of \( \tilde{Q} - \hat{B}W \) are given by

\[
\mathcal{V}(Q) - \hat{B}\mathcal{V}(W) = Y - B^*Z,
\]

and the following three blocks of \( N \) columns are given by

\[
\mathcal{V}(Qi) - \hat{B}\mathcal{V}(Wi), \mathcal{V}(Qj) - \hat{B}\mathcal{V}(Wj), \mathcal{V}(Qk) - \hat{B}\mathcal{V}(Wk),
\]

respectively, which are the corresponding matrices for the remaining three time series in the ensemble.

This ensemble-based approach can be generalized to any process whose regression parameters are invariant under the action of a finite group.

VI. Numerical simulation example

As an example, we will consider the improper AR\(_1\) (1) process

\[
q_t = Aq_{t-1} + \epsilon_{t-1},
\]

where \( A = 0.8 - 0.3i + 0.2j + 0.1k \) and \( \epsilon_t \) is an improper Gaussian quaternion white noise with covariance matrix

\[
\Sigma_{\mathcal{V}(\epsilon)} = \begin{bmatrix}
4 & 0 & 0 & 0 \\
0 & 4 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix}.
\]

For various \( N \), we generated 100 sample time-series of length \( N \), allowing for a long burn-in period to avoid initialization effects. We then estimate \( \hat{A} \) by \( \hat{A} \) using both quaternion LS and GLS estimation. The estimation error is given by the average value of \( ||\hat{A} - A|| \) over the 100 samples. Fig. 1 shows that, as expected, GLS outperforms LS for improper processes. We may also consider the corresponding AR\(_1\) (1) process given by (12) or the corresponding proper AR\(_2\) (1) process given by (13). The estimates obtained by unconstrained real LS estimation and complex LS estimation are equal to their respective GLS estimates. For this example Fig. 1 shows that correct specification of any coefficient structure which allows us to reduce the number of real parameters estimated is significantly more important than knowledge of the error covariance structure, and in particular whether or not the noise is proper.

VII. Summary

A VAR model with quaternion-valued parameters and driven by proper quaternion-valued innovations vectors is a proper quaternion VAR process. For such a process the GLS and LS parameter estimators are identical, and under Gaussianity of the innovations, also equal to the maximum likelihood estimator.
Fig. 1. Error (average value of $||\hat{A} - \tilde{A}||$) in the estimation of $\hat{A}$. From top to bottom the methods used are real LS (squares), complex LS (circles), quaternion LS (line) and quaternion GLS (dashed). $N$ is the number of samples in the series.

ACKNOWLEDGEMENTS

Paul Ginzberg thanks the EPSRC (UK) for financial support. Helpful comments by the referees were much appreciated.

REFERENCES