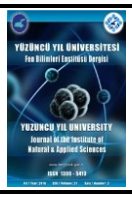
Yuzuncu Yıl University
Journal of the Institute of Natural & Applied Sciences

https://dergipark.org.tr/en/pub/yyufbed



Research Article

A Block Conjugate Gradient Method for Quaternion Linear Systems [#]

Sinem ŞİMŞEK[*], Ayça KÖRÜKÇÜ

Kırklareli University, Faculty of Art and Sciences, Department of Mathematics, 39100, Kırklareli, Türkiye
Sinem ŞİMŞEK, ORCID No: 0000-0001-5893-7080, Ayça KÖRÜKÇÜ, ORCID No: 0000-0002-0332-2107
*Corresponding author e-mail: sinem.simsek@klu.edu.tr

Article Info

Received: 31.08.2022
Accepted: 23.01.2023
Online August.2023

DOI:10.53433/yyufbed.1168844

Keywords
Block conjugate gradient method,
Krylov subspace,
Linear quaternion systems

Abstract: This study aims at the simultaneous solution of several quaternion linear systems with the same Hermitian and positive definite coefficient matrix by employing the conjugate gradient method. We consider the setting when the quaternion Hermitian positive definite coefficient matrix at hand is very large so that direct methods are not applicable. In the study, we first transform linear quaternion systems into real linear systems. The transformed real linear systems have special structure due to the fact that they are real representations of quaternion systems. Benefitting from the special structure, we further reduce the size of these linear systems. Then a block conjugate gradient method is applied to the resulting reduced real linear systems. The solution obtained after applying the conjugate gradient method is a real representation of the solution of the original quaternion problem. Thus, a conversion of this real solution to the quaternion setting is performed in the end.

Kuaterniyon Lineer Sistemler için Blok Eşlenik Gradyan Metot

Makale Bilgileri

Geliş: 31.08.2022
Kabul: 23.01.2023
Online Ağustos.2023

DOI:10.53433/yyufbed.1168844

Anahtar Kelimeler
Blok eşlenik gradyan metot,
Krylov altuzay,
Lineer kuaterniyon sistemler

Öz: Bu çalışmada katsayılar matrisi aynı ve Hermitian, pozitif tanımlı olan bir takım lineer kuaterniyon sistemlerinin eşlenik gradyan metodu kullanılarak eş zamanlı çözümü amaçlanmıştır. Kuaterniyon Hermitian pozitif tanımlı katsayılar matrisinin boyutunun çok büyük olması durumunda, lineer sistemlerin çözümü için direk metotlar kullanıma uygun değildir. Çalışmada, öncelikle lineer kuaterniyon sistemlerini reel lineer sistemlere dönüştürdük. Dönüştürülen reel lineer sistemler, kuaterniyon sistemlerin reel temsilleri olmalarından ötürü özel yapıya sahiplerdir. Bu özel yapıyı kullanarak reel lineer sistemlerin boyutlarını indirgedik. Daha sonra elde edilen indirgenmiş reel sistemlere yinelemeli bir yöntem olan blok eşlenik gradyan metodunu uyguladık. Blok eşlenik gradyan metodu uygulandıktan sonra elde edilen çözümler, orijinal kuaterniyon lineer sistemlerinin çözümlerinin bir reel temsilidir. Son olarak bu reel çözümleri original sistemin kuaterniyon çözümlerine dönüştürdük.

1. Introduction

Quaternions have important applications in quantum mechanics, image processing, and kinematics (Rodman, 2014; Wei et al., 2018). More specifically, they are useful in the theory of quantum mechanics for unifying fundamental forces concerning elementary particles (Adler, 1995). Additionally, quaternions are used to represent color video signals so as to retain correlation between the three channels of red, green, blue. Therefore, they have been widely employed in color imaging and color

[#] This paper is based on parts of the M.S. thesis of Ayça KÖRÜKÇÜ conducted under the supervision of Sinem ŞİMŞEK

videos in the literature (Sangwine, 1996; He et al., 2023). They are also suitable for describing translations, rotations of a rigid body (Arena et al., 1998; Caccavale et al., 1999). Use of quaternions in such applications stimulated interest into quaternions. On the other side, linear systems over complex numbers or quaternions are among the most common problems encountered in scientific computing. They usually arise from discretization of models for solving physical problems, linearization of nonlinear equations or problems that involve obtaining the least-squares estimates of the parameters in linear statistical models (Wendland, 2018). In this work, we focus on the solution of Hermitian positive definite linear systems over quaternions in the large-scale setting when there are many unknowns and equations. In particular, for such quaternion systems, we describe a block conjugate gradient method, an iterative subspace approach based on projections to Krylov subspaces exploiting Hermitian positive definiteness of the systems.

The skew-field $\mathbb{H}$ of quaternions is a four-dimensional algebra over the real algebra $\mathbb{R}$ generated by the basis $\{1, i, j, k\}$ with the multiplication rules $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$. A quaternion number a can be expressed as

$$a = a_0 + a_1 i + a_2 j + a_3 k, \tag{1}$$

where a_0, a_1, a_2 and a_3 are real numbers, and the modulus of a is defined as $|a| = \sqrt{\sum_{l=0}^{3} a_l^2}$. Clearly, the multiplication of two quaternion scalars is non-commutative. Denoting by $\mathbb{H}_{m\times n}$ the set of all $m \times n$ matrices with quaternion entries, a system of quaternion linear equations can be represented as

$$Ax = b, \tag{2}$$

where $A \in \mathbb{H}_{m\times n}$ is a known matrix, $b \in \mathbb{H}_{n\times 1}$ is a known vector, and $x \in \mathbb{H}_{n\times 1}$ is the unknown vector. Iterative methods are often preferred for solving large-scale sparse linear systems. One family of iterative methods that include most widely used techniques to solve linear systems today is Krylov subspace methods, all of which seek the solution of the original problem in a lower-dimensional (Krylov) subspace. The solution of the linear system in Equation (2) over the field of real and complex fields (when A is a real and complex matrix, b and the unknown x are real and complex vectors) by Krylov subspace methods has been studied extensively. However, very few studies can be found in the literature on the quaternion linear system in Equation (2). The notable studies on the quaternion linear systems based on Krylov subspaces are as follows. Jia and Ng have developed the quaternion generalized minimum residual method (QGMRES) for solving the quaternion linear system in Equation (2) (Jia & Ng, 2021). Opfer has investigated the conjugate gradient algorithm based on Krylov subspaces for the quaternion linear systems when the coefficient matrix is a Hermitian positive definite matrix (Opfer, 2005).

Let us consider the linear systems

$$Ax_i = b_i, i = 1,2, \dots, s, \tag{3}$$

for a given Hermitian positive definite matrix $A \in \mathbb{H}_{n\times n}$, given vectors $b_i \in \mathbb{H}_{n\times 1}$, and unknown vectors $x_i \in \mathbb{H}_{n\times 1}$. We can solve these linear systems simultaneously instead of solving them separately (O'Leary, 1980). By arranging the right-hand sides and the corresponding solutions in the matrices

$$B \coloneqq [b_1|b_2| \dots b_s] \in \mathbb{H}_{n\times s} \text{ and } X \coloneqq [x_1|x_2| \dots x_s] \in \mathbb{H}_{n\times s} \tag{4}$$

the linear systems in Equation (3) can be combined into

$$AX = B. \tag{5}$$

More formally, in this work, we describe a block conjugate gradient method to solve the quaternion matrix equation in Equation (5). A block conjugate gradient method for $AX = B$ has advantages over

applying the conjugate gradient method separately to the s quaternion linear systems in Equation (3). Most notably, the solutions for all right-hand sides are estimated simultaneously. This may reduce the dimension of the Krylov subspace, and, as a result, the number of matrix-vector multiplications as well as orthogonalization costs to compute the solutions to a prescribed accuracy. Consequently, a block conjugate gradient method often results in not only computational efficiency but also a decrease in memory storage requirements (Feng et al., 1995; Ji & Li, 2017).

This paper is organized as follows. In Section 2, we first recall the preliminaries for quaternion matrices and properties of the real counterparts of quaternion matrices. The proposed approach for the solution of the matrix equation in Equation (5) is described in Section 2.2. In the proposed approach, we transform the original quaternion matrix equation into a real matrix equation by using the real representation of a quaternion matrix. Then, by exploiting the special structures of the real representations of quaternion matrices, we further reduce the sizes of the matrices involved in the real matrix equation. Subsequently, a block conjugate gradient method is applied to the resulting real matrix equation. The solution of the original quaternion matrix equation can be formed from the solution of the real matrix equation by a simple transformation. Thus in the end we apply this transformation to the solution of the real matrix equation retrieved from the block conjugate gradient method to obtain the solution of the original quaternion matrix equation. In Section 3, we confirm that the proposed approach converges to the solution as expected from a conjugate gradient method on two numerical examples. In particular, we illustrate on these examples how quickly the residuals decay depends on the number of clusters of the eigenvalues of the coefficient matrix. We conclude in Section 4 with a summary of our findings and point out future research directions.

2. Material and Methods

In this section, we first briefly give the necessary background on quaternion matrices, their real representations, spectral properties, and spell out the real counterpart of the quaternion matrix equation in Equation (5) in Section 2.1. Then, the proposed approach for the iterative solution of the quaternion matrix equation in Equation (5) based on a block conjugate gradient method is described in Section 2.2.

2.1. Background

A quaternion matrix $A \in \mathbb{H}_{m\times n}$ can be written as a sum of four real matrices in the form

$$A = A_0 + A_1 i + A_2 j + A_3 k, \tag{6}$$

where $A_0, A_1, A_2, A_3 \in \mathbb{R}_{m\times n}$. The conjugate and conjugate transpose of a quaternion matrix A is defined as $\bar{A} = A_0 - A_1 i - A_2 j - A_3 k$ and $A^* = A_0^T - A_1^T i - A_2^T j - A_3^T k$, respectively. Moreover, a matrix $A \in \mathbb{H}_{n\times n}$ is called invertible, if there exist a matrix $A^{-1} \in \mathbb{H}_{n\times n}$ satisfying $AA^{-1} = A^{-1}A = I$. For a matrix $A \in \mathbb{H}_{n\times n}$, if $A = A^*$, then A is called a Hermitian matrix. A Hermitian quaternion matrix A is called positive definite if and only if $x^*Ax > 0$ for $\forall x \in \mathbb{H}_{n\times 1}\backslash\{0\}$. The 2-norm of $x \in \mathbb{H}_{n\times 1}$ and the Frobenius norm of $A = [a_{ij}] \in \mathbb{H}_{m\times n}$ are defined by $\|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$ and $\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{i=1}^{n} |a_{ij}|^2}$, respectively. We refer the reader to (Rodman, 2014; Wei et al., 2018) for this background information and further details on quaternions.

One way of dealing with a matrix A over the quaternion skew-field is to exploit its real matrix representation $\psi(A)$ (Wei et al., 2018), which is defined for $A = A_0 + A_1 i + A_2 j + A_3 k \in \mathbb{H}_{m\times n}$ with $A_0, A_1, A_2, A_3 \in \mathbb{R}_{m\times n}$ by

$$\psi(A) = \begin{bmatrix} A_1 & -A_2 & -A_3 & -A_4 \\ A_2 & A_1 & -A_4 & A_3 \\ A_3 & A_4 & A_1 & -A_2 \\ A_4 & -A_3 & A_2 & A_1 \end{bmatrix}. \tag{7}$$

This real representation of a quaternion matrix is motivated by the fact that a quaternion matrix X satisfying the quaternion linear system in Equation (5) also satisfies the real linear system below:

$$\psi(A)\psi(X) = \psi(B). \tag{8}$$

We remark that there are alternative real representations of a quaternion matrix employed in the literature to convert a quaternion linear system as in Equation (5) into a real one as in Equation (8), but there is not a notable advantage of one over the others (Jia & Ng, 2021). The following properties can be verified for the real matrix representation of a quaternion matrix above we depend on here.

Theorem 1 (Wei et al., 2018)**:** Let $A, B \in \mathbb{H}_{m\times n}$, $C \in \mathbb{H}_{n\times s}$ and $k \in \mathbb{R}$. The following assertions hold:

i. $A = B \Leftrightarrow \psi(A) = \psi(B)$.
ii. $\psi(A + B) = \psi(A) + \psi(B)$.
iii. $\psi(kA) = k\psi(A)$.
iv. $\psi(AC) = \psi(A)\psi(C)$.
v. $\psi(A^*) = \psi(A)^T$.
vi. $A \in \mathbb{H}_{n\times n}$ is an invertible matrix if and only if $\psi(A)$ is an invertible matrix, and, if $A \in \mathbb{H}_{n\times n}$ is invertible, then $\psi(A^{-1})$=$\psi(A)^{-1}$.
vii. $A \in \mathbb{H}_{n\times n}$ is a unitary matrix, i.e., $AA^* = A^*A = I$, if and only if $\psi(A)$ is an orthogonal matrix.
viii. $\|\psi(A)\|_F = 2\|A\|_F$.

Next we formally define a right eigenvalue, and eigenvector of a square quaternion matrix.

Definition 1 (Farenick & Pidkowich, 2003)**:** For $A \in \mathbb{H}_{n\times n}$, if the pair $(\lambda, x) \in \mathbb{H} \times (\mathbb{H}_{n\times 1}\backslash\{0\})$ is such that $Ax = x\lambda$, then λ is called a right eigenvalue, x is called a right eigenvector corresponding to λ, and (λ, x) is called a right eigenpair of A.

If λ is a non-real right eigenvalue such that $Ax = x\lambda$, then $Axs = xs(s^{-1}\lambda s)$ for every nonzero $s \in \mathbb{H}$. Hence, in case A has a non-real right eigenvalue, then it turns out that A has infinitely many non-real right eigenvalues. But if A is a Hermitian quaternion matrix, then all eigenvalues of A are real as stated next formally.

Theorem 2 (Farenick & Pidkowich, 2003)**:** If $A \in \mathbb{H}_{n\times n}$ is Hermitian, then every right eigenvalue of A is real and the number of right eigenvalues of A is n.

Now, we state the spectral theorem of a Hermitian quaternion matrix.

Theorem 3 (Farenick & Pidkowich, 2003)**:** If $A \in \mathbb{H}_{n\times n}$ is Hermitian, then there is a unitary matrix $U \in \mathbb{H}_{n\times n}$ such that U^*AU is a diagonal matrix, whose diagonal entries are real and correspond to the eigenvalues of A.

Hermitian positive definite quaternion matrices play an important role in this work. The next result states that their real representations are also positive definite.

Theorem 4 (Wei et al., 2018)**:** A matrix $A \in \mathbb{H}_{n\times n}$ is Hermitian positive definite if and only if $\psi(A)$ is a symmetric positive definite matrix.

The real representations of quaternion matrices possess special structure. To state the special structure possessed formally, we next introduce the notion of a JRS-symmetric matrix, and related structures.

Definition 2 (Wei et al., 2018)**:** Let J_n, R_n and S_n be the following orthogonal matrices:

$$J_n = \begin{bmatrix} 0 & 0 & -I_n & 0 \\ 0 & 0 & 0 & -I_n \\ I_n & 0 & 0 & 0 \\ 0 & I_n & 0 & 0 \end{bmatrix}, R_n = \begin{bmatrix} 0 & -I_n & 0 & 0 \\ I_n & 0 & 0 & 0 \\ 0 & 0 & 0 & I_n \\ 0 & 0 & -I_n & 0 \end{bmatrix} \text{ and } S_n = \begin{bmatrix} 0 & 0 & 0 & -I_n \\ 0 & 0 & I_n & 0 \\ 0 & -I_n & 0 & 0 \\ I_n & 0 & 0 & 0 \end{bmatrix}. \quad (9)$$

We define a JRS-symmetric, JRS-symplectic, and orthogonal JRS-symplectic matrix as follows:

i. A matrix $M \in \mathbb{R}_{4m\times 4n}$ is called JRS-symmetric if $J_m M J_n^T = M$, $R_m M R_n^T = M$ and $S_m M S_n^T = M$.
ii. A matrix $O \in \mathbb{R}_{4m\times 4n}$ is called JRS-symplectic if $O J_n O^T = J_m$, $O R_n O^T = R_m$ and $O S_n O^T = S_m$.
iii. A matrix $W \in \mathbb{R}_{4n\times 4n}$ is called orthogonal JRS-symplectic if it is orthogonal and JRS-symplectic.

The first one of the following two result relates JRS-symmetric matrices with quaternion matrices, while the second one spells out what distinguishes two different JRS-symmetric matrices.

Theorem 5 (Wei et al., 2018)**:** A matrix $M \in \mathbb{R}_{4m\times 4n}$ is JRS-symmetric if and only if M is a real matrix representation of a quaternion matrix.

Theorem 6 (Wei et al., 2018)**:** A JRS-symmetric matrix is uniquely determined by its four submatrices on the first column block (or first row block).

In the next theorem, $\psi(A)_c$ and, $\psi(A)_r$ denote the first column block and the first row block of a quaternion matrix A, respectively.

Theorem 7 (Wei et al., 2018)**:** Letting $A, B \in \mathbb{H}_{m\times n}$, $C \in \mathbb{H}_{n\times s}$, $q \in \mathbb{H}_{n\times 1}$ and $k \in \mathbb{R}$, we have

i. $\psi(A+B)_c = \psi(A)_c + \psi(B)_c$, $\psi(A+B)_r = \psi(A)_r + \psi(B)_r$,
ii. $\psi(kA)_c = k\psi(A)_c$, $\psi(kA)_r = k\psi(A)_r$,
iii. $\psi(AC)_c = \psi(A)\psi(C)_c$, $\psi(AC)_r = \psi(A)_r\psi(C)$,
iv. $\psi(A^*)_c = (\psi(A)^T)_c$, $\psi(A^*)_r = (\psi(A)^T)_r$,
v. $\|\psi(q)_c\|_2 = \|q\|_2$, $\|\psi(q)_r\|_2 = \|q\|_2$,
vi. $\|\psi(A)_c\|_F = \|A\|_F$, $\|\psi(A)_r\|_F = \|A\|_F$.

2.2. The quaternion block conjugate gradient method

Let us consider the quaternion matrix equation $AX = B$. As the coefficient matrix is a Hermitian positive definite matrix, especially if the matrix A is sparse and large, a block conjugate gradient method is a good candidate to solve this quaternion matrix equation. As indicated in the previous section, $AX = B$ holds if and only if $\psi(A)\psi(X) = \psi(B)$ holds. Hence, instead of the quaternion matrix equation $AX = B$, we can solve the real matrix equation $\psi(A)\psi(X) = \psi(B)$. The sizes of the matrices $\psi(A)$, $\psi(X)$ and $\psi(B)$ in this real matrix equation are $4n \times 4n$, $4n \times 4s$, and $4n \times 4s$, respectively. However, it follows from Theorem 7 (iii) that the sizes of the latter two can be reduced. In particular, we can indeed solve $\psi(A)\psi(X)_c = \psi(B)_c$ instead of $\psi(A)\psi(X) = \psi(B)$; the upside of doing so is that the unknown $\psi(X)_c$ and the right-hand side $\psi(B)_c$ are both of size $4n \times s$, that is we now solve s linear systems simultaneously instead of $4s$ simultaneous linear systems needed for $\psi(A)\psi(X) = \psi(B)$. Moreover, the quaternion solution $X \in \mathbb{H}_{n\times s}$ of $AX = B$ can be formed immediately from the real solution $\psi(X)_c$ of $\psi(A)\psi(X)_c = \psi(B)_c$.

By Theorem 4, the real matrix $\psi(A)$ is symmetric and positive definite, so we can apply the block conjugate gradient method to solve the real matrix equation $\psi(A)\psi(X)_c = \psi(B)_c$ using the real floating point arithmetic. For the sake of simplifying the notation, let us denote $\psi(A)$, $\psi(X)_c$, $\psi(B)_c$

with $\mathcal{A}$, $\mathcal{X}$, and $\mathcal{B}$, respectively. Letting $\mathcal{X}_0$ be an initial guess matrix for the solution $\mathcal{X}$, and $\mathcal{R}_0$ be the corresponding residual matrix, i.e., $\mathcal{R}_0 \coloneqq \mathcal{B} - \mathcal{A}\mathcal{X}_0$, the block Krylov subspace is defined as

$$\mathcal{K}_k(\mathcal{A}, \mathcal{R}_0) = span\{\mathcal{R}_0, \mathcal{A}\mathcal{R}_0, \dots, \mathcal{A}^{k-1}\mathcal{R}_0\}. \tag{10}$$

Observe that, denoting the ith columns of $\mathcal{B}$ and $\mathcal{X}$ with $\mathcal{b}_i$ and x_i, respectively, the solution $\mathcal{X} = [x_1|x_2|\dots x_s]$ of $\mathcal{A}\mathcal{X} = \mathcal{B}$ is the unique global minimizer of

$$\Phi(\mathcal{X}) \coloneqq \sum_{i=1}^{s} \varphi_i(x_i) \coloneqq \sum_{i=1}^{s} \frac{1}{2} x_i{}^T \mathcal{A} x_i - x_i{}^T \mathcal{b}_i. \tag{11}$$

Indeed, since $\mathcal{A}$ is a symmetric positive definite matrix, the objective function $\Phi(\mathcal{X})$ has only one global minimizer, that is the unique solution of $\nabla\Phi(\mathcal{X}) = \mathcal{A}\mathcal{X} - \mathcal{B} = 0$. The block conjugate gradient method seeks the global minimizer $\mathcal{X}_k$ of $\Phi(\mathcal{X})$ over the affine space $\mathcal{X}_0 + \mathcal{K}_k(\mathcal{A}, \mathcal{R}_0)$. Equivalently, letting $\mathcal{X}_*$ be the solution of $\mathcal{A}\mathcal{X} = \mathcal{B}$, the matrix $\mathcal{X}_k$ is also the global minimizer $\|\mathcal{X} - \mathcal{X}_*\|_{F,\mathcal{A}}^2 \coloneqq trace\big((\mathcal{X} - \mathcal{X}_*)^T \mathcal{A}(\mathcal{X} - \mathcal{X}_*)\big)$ over the space $\mathcal{X}_0 + \mathcal{K}_k(\mathcal{A}, \mathcal{R}_0)$. Hence, the block conjugate gradient method at iteration k finds the matrix in the affine space $\mathcal{X}_0 + \mathcal{K}_k(\mathcal{A}, \mathcal{R}_0)$ that is closest to the actual solution $\mathcal{X}_*$ of $\mathcal{A}\mathcal{X} = \mathcal{B}$ with respect to the weighted Frobenius norm $\|\cdot\|_{F,\mathcal{A}}^2$. The between the iterates $\mathcal{X}_{k+1}$ and $\mathcal{X}_k$ of the conjugate method at two consecutive iterations can be represented as

$$\mathcal{X}_{k+1} = \mathcal{X}_k + \mathcal{P}_{k+1}\Lambda_{k+1}, \tag{12}$$

where $\mathcal{P}_{k+1} \in \mathbb{R}_{n\times s}$ with columns in $\mathcal{K}_k(\mathcal{A}, \mathcal{R}_0)$ corresponds to the s search directions for the solutions of the s linear systems, and $\Lambda_{k+1} \in \mathbb{R}_{s\times s}$ is a diagonal matrix with entries on the diagonal representing the step lengths that determine how big steps should be taken in search directions. As $\mathcal{X}_{k+1}$ is the minimizer over $\mathcal{X}_0 + \mathcal{K}_k(\mathcal{A}, \mathcal{R}_0)$, the diagonal step-length matrix $\Lambda_{k+1} \in \mathbb{R}_{s\times s}$ must satisfy

$$\frac{\partial\Phi(\mathcal{X}_k + \mathcal{P}_{k+1}\Lambda_{k+1})}{\partial\Lambda_{k+1}} = \mathcal{P}_{k+1}^T(\mathcal{A}(\mathcal{X}_k + \mathcal{P}_{k+1}\Lambda_{k+1}) - \mathcal{B}) = 0. \tag{13}$$

The new search directions $\mathcal{P}_{k+1} \in \mathbb{R}_{n\times s}$ can be expressed in terms of the previous search directions $\mathcal{P}_k$ as

$$\mathcal{P}_{k+1} = \mathcal{R}_k + \mathcal{P}_k\Omega_{k+1}, \tag{14}$$

where $\mathcal{R}_k = \mathcal{B} - \mathcal{A}\mathcal{X}_k$ and the parameter matrix Ω_{k+1} is given by $\Omega_{k+1} = -\big(\mathcal{P}_k^T\mathcal{A}\mathcal{P}_k\big)^{-1}\big(\mathcal{P}_k^T\mathcal{A}\mathcal{R}_k\big)$. This expression is also obtained by exploiting the optimality of $\mathcal{X}_{k+1}$, more specifically the fact that it is the minimizer of $\Phi(\mathcal{X})$ over $\mathcal{X}_0 + \mathcal{K}_{k+1}(\mathcal{A}, \mathcal{R}_0)$. As it turns out, the optimality of $\mathcal{X}_{k+1}$ implies the orthogonality of $\mathcal{X}_{k+1} - \mathcal{X}_0$ to every vector in $\mathcal{K}_k(\mathcal{A}, \mathcal{R}_0)$, i.e., $trace\left(\mathcal{Z}^T\mathcal{A}(\mathcal{X}_{k+1} - \mathcal{X}_0)\right) = 0$ for all $\mathcal{Z} \in \mathcal{K}_k(\mathcal{A}, \mathcal{R}_0)$. This in turn gives rise to the equality in Equation (14). It is straightforward to deduce $\mathcal{K}_k(\mathcal{A}, \mathcal{R}_0) = span\{\mathcal{R}_0, \mathcal{A}\mathcal{R}_0, \dots, \mathcal{A}^{k-1}\mathcal{R}_0\} = span\{\mathcal{R}_0, \mathcal{R}_1, \dots, \mathcal{R}_{k-1}\}$ and $\mathcal{R}_j^T\mathcal{R}_k = 0$ for $j = 1,2,\dots,k-1$. From this fact and $\mathcal{P}_k \in \mathcal{K}_k(\mathcal{A}, \mathcal{R}_0)$, it can also be deduced that $\mathcal{P}_k^T\mathcal{R}_k = 0$ as well as $\mathcal{P}_{k+1}^T\mathcal{A}\mathcal{P}_k = 0$. To summarize, the following properties hold among the residuals and search directions:

i. $\mathcal{R}_j^T\mathcal{R}_k = 0$ for $j \neq k$,
ii. $\mathcal{P}_k^T\mathcal{R}_k = 0$,
iii. $\mathcal{P}_{k+1}^T\mathcal{A}\mathcal{P}_k = 0$.

By employing the orthogonality properties above, the expressions for Λ_{k+1} and Ω_{k+1} can be rewritten as

$$\Lambda_{k+1} = \big(\mathcal{P}_{k+1}^T\mathcal{A}\mathcal{P}_{k+1}\big)^{-1}\mathcal{R}_k^T\mathcal{R}_k \text{ and } \Omega_{k+1} = \big(\mathcal{R}_{k-1}^T\mathcal{R}_{k-1}\big)^{-1}\mathcal{R}_k^T\mathcal{R}_k. \tag{15}$$

A description of the overall block conjugate algorithm for the real matrix equation $\mathcal{A}\mathcal{X} = \mathcal{B}$ is given in Algorithm 1 below.

Algorithm 1. The block conjugate gradient algorithm

Input: $\mathcal{A} \in \mathbb{R}_{4n\times 4n}$, $\mathcal{B} \in \mathbb{R}_{4n\times s}$, maximum number of iterations κ, termination tolerance ε
Output: An approximate solution $\widehat{\mathcal{X}}$ for $\mathcal{A}\mathcal{X} = \mathcal{B}$

Chose an initial guess matrix $\mathcal{X}_0$, $\mathcal{R}_0 = \mathcal{B} - \mathcal{A}\mathcal{X}_0$, $k = 0$
while $k \leq \kappa$ and $\|\mathcal{R}_k\|_F > \varepsilon$
    If $k = 0$, then
        $\mathcal{P}_0 \coloneqq \mathcal{R}_0$
    else
        $\Omega_{k+1} = \left(\mathcal{R}_{k-1}^T \mathcal{R}_{k-1}\right)^{-1} \mathcal{R}_k^T \mathcal{R}_k$ and $\mathcal{P}_{k+1} = \mathcal{R}_k + \mathcal{P}_k \Omega_{k+1}$
    end
    $\Lambda_{k+1} = \left(\mathcal{P}_{k+1}^T \mathcal{A} \mathcal{P}_{k+1}\right)^{-1} \mathcal{R}_k^T \mathcal{R}_k$
    $\mathcal{X}_{k+1} = \mathcal{X}_k + \mathcal{P}_{k+1}\Lambda_{k+1}$
    $\mathcal{R}_{k+1} = \mathcal{R}_k - \mathcal{A}\mathcal{P}_{k+1}\Lambda_{k+1}$
    $k = k + 1$
end
$\widehat{\mathcal{X}} = \mathcal{X}_k$

3. Results

In this section, we demonstrate the accuracy and convergence of the proposed method on a quaternion matrix equation $AX = B$ with a rectangular tall and skinny coefficient matrix A. The best solution to this matrix equation in the least-squares sense is found by solving the normal equation associated with the original matrix equation. The normal equation is always consistent, that is it always has a solution. More specifically, we consider a quaternion matrix equation $AX = B$, where $A \in \mathbb{H}_{400\times 80}$, $B \in \mathbb{H}_{400\times 10}$, and $X \in \mathbb{H}_{80\times 10}$. The normal equation associated with this quaternion matrix equation is $A^*AX = A^*B$. The coefficient matrix of the normal equation A^*A is always Hermitian positive definite, so we can apply the block conjugate gradient method described in Section 2.2 to solve the normal equation $A^*AX = A^*B$. By exploiting the real matrix representations of A^*A, X and B, we transform the quaternion normal equation $A^*AX = A^*B$ into the real matrix equation

$$\psi(A^*A)\psi(X) = \psi(A^*B). \tag{16}$$

This transformation yields the real matrices $\psi(A^*A)$, $\psi(X)$ and $\psi(A^*B)$, whose dimensions are 320×320, 320×40 and 320×40, respectively. As suggested in the previous section, we can further reduce the dimensions of the matrices in Equation (16) by instead solving

$$\psi(A^*A)\psi(X)_c = \psi(A^*B)_c. \tag{17}$$

The solution X of the normal equation $A^*AX = A^*B$, hence the least-squares solution of $AX = B$, can then be retrieved from $\psi(X)_c$.

Hence, using now the short-hands $\mathcal{A}$, $\mathcal{X}$ and $\mathcal{B}$ for $\psi(A^*A)$, $\psi(X)_c$, $\psi(A^*B)_c$, respectively, we apply Algorithm 1 to $\mathcal{A}\mathcal{X} = \mathcal{B}$ by setting the maximum number of iterations equal to 40. The convergence of the algorithm is illustrated in Figure 1 and Figure 2. To be precise, the plot in Figure 1 depicts the residual norm $\|\psi(B)_c - \psi(A)(\psi(X)_c)_k\|_F$ for the real matrix equation $\psi(A)\psi(X)_c = \psi(B)_c$ as a function of number of iterations k. On the other hand, the plot in Figure 2 shows the residual norm $\|\mathcal{R}_k\|_F = \|\mathcal{B} - \mathcal{A}\mathcal{X}_k\|_F$ for the normal equations $\mathcal{A}\mathcal{X} = \mathcal{B}$ as a function of k. When the coefficient matrix is unitarily diagonalizable (normal), the convergence behavior of the conjugate gradient method is determined by the eigenvalue distribution of the coefficient matrix. In the example we consider here, the matrix $A \in \mathbb{H}_{400\times 80}$ is formed so that A^*A has 10 distinct eigenvalues. Not

surprisingly, the residual norms in Figure 1 and 2 decay initially up until the 10th iteration, but do not change much starting from 10th iteration and afterwards. In Figure 1, the residual norm becomes 114 at the 10th iteration, and remains more or less 114 after the 10th iteration. This is expected as the linear system $\psi(A)\psi(X)_c = \psi(B)_c$, equivalently the linear system $AX = B$, has no solution, but the conjugate gradient method (Algorithm 1) converges nearly to the least-squares solution minimizing $\|\psi(B)_c - \psi(A)(\psi(X)_c)_k\|_F$ over all X by the 10th iteration, as the solution to the associated normal equation is nearly converged by then. Meanwhile, in Figure 2 it can be observed that the residual norm associated with the normal equation decays to zero by the 10th iteration, and remains so afterwards, which is expected as the conjugate gradient method is applied to solve the normal equation that has a solution.

We have also perturbed the coefficient matrix so that for the resulting $A \in \mathbb{H}_{400\times 80}$ the matrix A^*A has no more repeated eigenvalues, rather it has eigenvalues clustered around 10 real points. The plots of the residual norms $\|\psi(B)_c - \psi(A)(\psi(X)_c)_k\|_F$ and $\|\mathcal{R}_k\|_F = \|\mathcal{B} - \mathcal{A}\mathcal{X}_k\|_F$ for this slightly perturbed matrix equation are given in Figure 3 and Figure 4, respectively. Notice that, as before, the residual norms decay rapidly in the first 10 iterations. Though the decay in the residual norms slows down gradually after the 10th iteration, the decays in the residuals continue even after the 10th iteration, in particular the residuals do not remain constant after the 10th iteration anymore. This behavior in the residuals is expected, as the matrix A^*A has now all distinct eigenvalues.

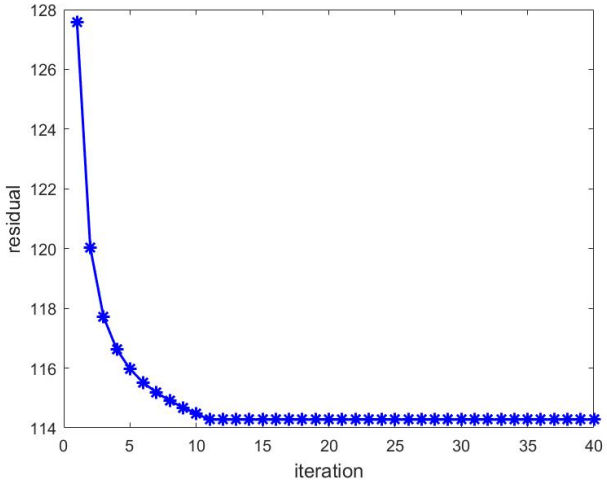


Figure 1. The decay of the residual norm $\|\psi(B)_c - \psi(A)(\psi(X)_c)_k\|_F$ as a function of k for the quaternion matrix equation $AX = B$ in Section 3.

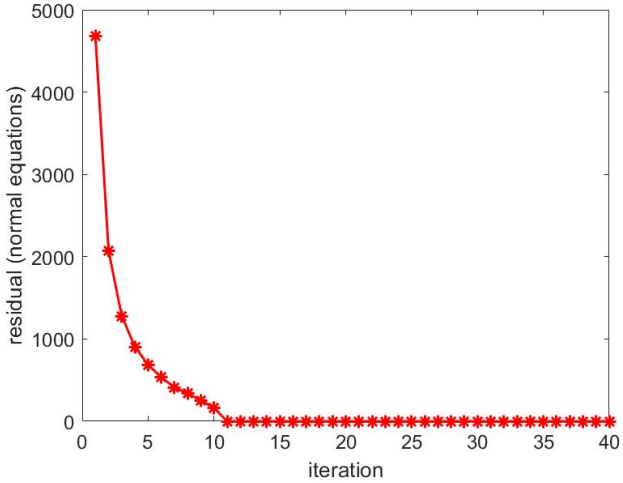


Figure 2. The decay of the residual norm $\|\mathcal{B} - \mathcal{A}\mathcal{X}_k\|_F$ associated with the normal equation $\mathcal{A}\mathcal{X} = \mathcal{B}$ with $\mathcal{A} = \psi(A^*A)$, $\mathcal{X} = \psi(X)_c$ and $\mathcal{B} = \psi(A^*B)_c$ as a function of k for the quaternion matrix equation $AX = B$ in Section 3.

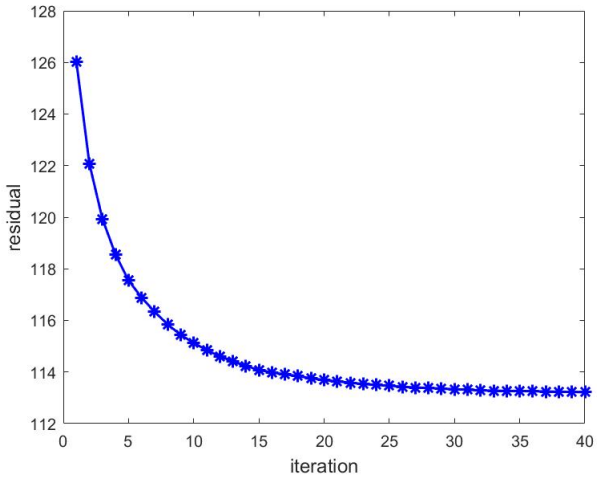


Figure 3. This is a plot of the residual norms as in Figure 1, but only now the coefficient matrix A is perturbed so that it has distinct eigenvalues clustered around 10 points on the real axis.

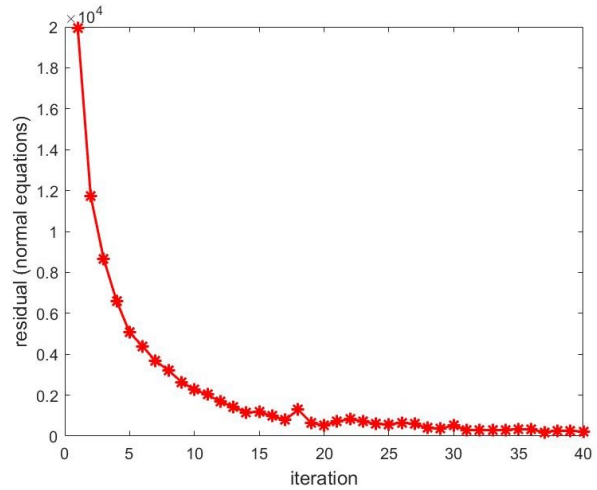


Figure 4. Plot of residual norms as in Figure 2, but the coefficient matrix A is perturbed so that it has distinct eigenvalues clustered around 10 points.

4. Discussion and Conclusion

Problems involving quaternion objects, defined over the skew-field of quaternions, arise from a wide variety of applications. Especially, solving quaternion linear systems has attracted substantial attention, recently. The analytical tools such as the Moore-Penrose Inverse, Kronecker product are very expensive computationally, and not useful if the linear system has a large and sparse quaternion coefficient matrix. Iterative methods, on the other hand, may be applicable to such large-scale sparse quaternion linear systems.

In this study, we have considered several quaternion linear systems with the same Hermitian quaternion coefficient matrix. One way of dealing with quaternion linear systems is to use the real representations of quaternion matrices. We have first applied the transformation that converts the quaternion coefficient matrix, the right-hand sides and the solution to their real representations. The resulting linear systems are real, all with a common symmetric positive definite coefficient matrix. We have taken into account the special structure of the real representation of a quaternion matrix to reduce the number of resulting real linear systems to be solved. Subsequently, a block conjugate gradient method is applied to solve the real linear systems simultaneously in real arithmetic; solving systems simultaneously by a block method potentially reduces the number of iterations, hence improves the efficiency and storage requirements. The conversion of the real representation of the solution of the

original quaternion system returned by the block conjugate gradient method to the actual quaternion solution is performed in the end without any computational burden. The proposed method is implemented in Matlab, and its convergence to the actual solution is verified on two numerical examples. These numerical experiments illustrate that, as expected, how quickly the block conjugate gradient method converges depends on the number of clusters of the eigenvalues of the Hermitian quaternion coefficient matrix; the smaller is the number of clusters, the faster is the convergence.

As future work, we intend to focus on the solutions of quaternion matrix equations such as Sylvester, Lyapunov, and Stein matrix equations by block conjugate gradient methods. Computing special type of solutions (e.g., symmetric, centrosymmetric) of these quaternion matrix equations may be of interest. Moreover, the solutions of these matrix equations can be tried to be retrieved with applications of alternative iterative methods such as a block GMRES method in case Hermiticity and/or positive-definiteness features do not exist.

References

Adler, S. L. (1995). *Quaternionic Quantum Mechanics and Quantum Fields*. New York, USA: Oxford University Press.

Arena, P., Fortuna, L., Muscato, G., & Xibilia, M. G. (1998). *Neural Networks in Multidimensional Domains*. London, UK: Springer.

Caccavale, F., Natale, C., Siciliano, B., & Villani, L. (1999). Six-DOF impedance control based on angle/axis representations. *IEEE Transactions on Robotics and Automation*, 15(2), 289-300. doi:10.1109/70.760350

Farenick, D. R., & Pidkowich, B. A. F. (2003). The spectral theorem in quaternions. *Linear Algebra and its Applications*, 371, 75-102. doi:10.1016/S0024-3795(03)00420-8

Feng, Y. T., Owen, D. R. J., & Peric, D. (1995). A block conjugate gradient method applied to linear systems with multiple right-hand sides. *Computer Methods in Applied Mechanics and Engineering*, 127, 203-215. doi:10.1016/0045-7825(95)00832-2

He, Z. H., Wang, X. X., & Zhao, Y. F. (2023). Eigenvalues of quaternion tensors with applications to color video processing. *Journal of Scientific Computing*, 94, 1-15. doi:10.1007/s10915-022-02058-5

Ji, H., & Li, Y. (2017) Block conjugate gradient algorithms for least squares problems. *Journal of Computational and Applied Mathematics*, 317, 203-217. doi:10.1016/j.cam.2016.11.031

Jia, Z., & Ng, M. K. (2021). Structure preserving quaternion generalized minimal residual method. *Journal on Matrix Analysis and Applications*, 42(2), 616-634. doi:10.1137/20M133751X

O'Leary, D. P. (1980). The block conjugate gradient algorithm and related methods. *Linear Algebra and Its Applications*, 29, 293-322. doi:10.1016/0024-3795(80)90247-5

Opfer, G. (2005). The conjugate gradient algorithm applied to quaternion valued matrices. *Journal of Applied Mathematics and Mechanics*, 85(9), 660-672. doi:10.1002/zamm.200410191

Rodman, L. (2014). *Topics in Quaternion Linear Algebra*. New Jersey, USA: Princeton University Press.

Sangwine, S. J. (1996). Fourier transforms of colour images using quaternion or hypercomplex, numbers. *Electronics Letters*, 32(21), 1979-1980. doi:10.1049/el:19961331

Wei, M., Li, Y., Zhang, F., & Zhao, J. (2018). *Quaternion Matrix computations*. New York, USA: Nova Science Publishers.

Wendland, H. (2018). *Numerical Linear Algebra: An Introduction*. Cambridge, UK: Cambridge University Press.