The QR algorithm and the QD algorithm applied to Jacobi matrices: parametrisation of all the elements of all their matrices using the elements of a single sequence of discrete variables

A.J. Macfarlane [1]

Centre for Mathematical Sciences, D.A.M.T.P.
Wilberforce Road, Cambridge CB3 0WA, UK.

Abstract

Given an $n \times n$ Jacobi matrix J_n, a sequence X_n of discrete variables, $x_r \in X_n,\ r = 0, 1, \dots$, is introduced by means of a difference equation $\mathrm{DE}(X_n)$, which together with its initial conditions $\mathrm{IC}(X_n)$, is fully determined by the characteristic equation $\mathrm{CE}(\mathrm{J}_n)$ of J_n. Using what seems to be an ideal choice of the $\mathrm{IC}(X_n)$ detailed discussion is given of the properties of the x_r, and of the derivation of expressions for them as a function of the elements of J_n. On the basis of this knowledge, the implementation of the QR and (Rutishauser) QD algorithms for calculation of the eigenvalues of J_n is studied. Explicit compact closed form algebraic expressions are presented for all of the elements of all of the matrices $\mathrm{A}_r, \mathrm{Q}_r, \mathrm{R}_r, \mathrm{B}_r, \mathrm{L}_r,\ r = 0, 1, \dots,$ of the algorithms. These are proved in a sequence of theorems to conform to all the relevant requirements. This includes proving that the initialisations $\mathrm{A}_0 = \mathrm{B}_0 = \mathrm{J}_n$ follow by setting $r = 0$ in the general $\mathrm{A}_r, \mathrm{B}_r$ expressions.

The expressions presented for each algorithm already possess a very compact form in terms of the Hankel variables h^r_s. However these simplify dramatically when written in terms of the variables q^r_s and e^r_s introduced by Rutishauser in the context of the QD algorithm. Also the equalities between the matrices A_r and R_r of the QR algorithm and the matrices B_{2r} and $(\mathrm{L}_{2r}\,\mathrm{L}_{2r+1})^T$ of the QD algorithm can be verified.

A theorem of Desnanot and Jacobi shows that the Hankel variables satisfy an identity which ensures the orthogonality of each Givens rotation factor of the matrices Q_r of the QR algorithm. It also implies an identity involving the Rutishauser variables, which expedites the harder parts of the proofs of theorems in the algorithm work.

Keywords: Jacobi matrix, characteristic equation, difference equation, Hankel matrix, QR factorisation, Cholesky factorisation.

AMS Classification: 15A23, 39A06, 47B36, 65F10, 65F30.

1 Introduction

1.1 Some definitions

Before describing the purpose and content of this paper, it is convenient to begin with a few definitions, in which n is a fixed but arbitrary integer.

Definition 1. The Jacobi matrices $\mathrm{J}_\beta, \quad \beta = 1, \dots, n$, are real symmetric positive definite tridiagonal matrices, which have non-vanishing elements given by

$$\begin{aligned} (\mathrm{J}_\beta)_{j\,j} &= b_j, \qquad j = 1, \dots, \beta \\ (\mathrm{J}_\beta)_{j\,j+1} &= a_j, \quad a_j > 0, \quad j = 1, \dots, \beta - 1. \end{aligned} \tag{1}$$

[1] *e-mail*: a.j.macfarlane@damtp.cam.ac.uk

Definition 2. The characteristic equation CE(J_β) of J_β can be written in the form

$$\begin{aligned} 0 &= \det\left(\lambda \mathrm{I}_\beta - \mathrm{J}_\beta\right) \\ &= \sum_{k=0}^{\beta} (-1)^k \lambda^{\beta-k} T_{\beta\, k}, \quad T_{\beta\, 0} = 1, \end{aligned} \tag{2}$$

where $T_{\beta\, k}$ is the k-th elementary symmetric function of the eigenvalues of J_β.

Definition 3. The set X_β, $\beta = 1, \ldots, n$, of real discrete variables $x_r^{(\beta)}$, $r = 0, 1, \ldots$, are solutions of the difference equation DE(X_β)

$$x_{r+\beta}^{(\beta)} = \sum_{k=1}^{\beta} (-1)^{k+1} T_{\beta\, k} x_{r+\beta-k}^{(\beta)}, \tag{3}$$

and initial conditions, IC(X_β), assigning suitable values to $x_r^{(\beta)}$, $r = 0. \ldots, \beta - 1$. Algebraic work involving the $x_r^{(\beta)} \in X_\beta$ is decribed below as taking place at level β.

Remark 1. The definition of each of J_n and X_n requires $(2n-1)$ independent quantities, obviously for J_n. For X_n these are the $T_{n\, k}$, $k = 1, \ldots, n$ and $x_r^{(n)}$, $r = 1, \ldots, n-1$, since $x_0^{(n)}$ is just a scale factor set equal to one here. That the $x_r^{(n)}$ can be evaluated explicitly in terms of the elements of J_n becomes obvious in the light of Definition 8 below. The inverse relations giving the b_k and a_k of Definition 1 in terms of the elements of X_n are harder to find : see (73) and (67).

Definition 4, the QR algorithm. [1–4]
Writing $\mathrm{J}_n = \mathrm{A}_0$, define the sequence of matrices A_r, $r = 0, 1, 2, \ldots$, by repetition of the following two steps

(i) QR factorisation of A_r

$$\mathrm{A}_r = \mathrm{Q}_r\, \mathrm{R}_r, \tag{4}$$

where Q_r is an orthogonal matrix with determinant $+1$, and R_r is upper triangular with positive diagonal elements,

(ii) definition of A_{r+1}

$$\mathrm{R}_r\, \mathrm{Q}_r = \mathrm{A}_{r+1}. \tag{5}$$

Definition 5, the QD algorithm. [2, 5]
Writing $\mathrm{J}_n = \mathrm{B}_0$, define the sequences of matrices $\mathrm{B}_r, \mathrm{L}_r$, $r = 0, 1, 2, \ldots$, by repetition of the following two steps

(i) Cholesky factorisation of B_r

$$\mathrm{B}_r = \mathrm{L}_r\, {\mathrm{L}_r}^T, \tag{6}$$

where L_r is lower triangular with positive diagonal elements,

(ii) definition of B_{r+1}

$${\mathrm{L}_r}^T\, \mathrm{L}_r = \mathrm{B}_{r+1}. \tag{7}$$

It is a direct consequence of (4) and (5) that all the matrices A_r of Definition 4 inherit from $\mathrm{J}_n = \mathrm{A}_0$ the properties imposed on J_n in Definition 1. The existence of the Cholseky factorisation of $\mathrm{J}_n = \mathrm{B}_0$ is ensured by the demand in Definition 1 that J_n is positive definite, and all its eigenvalues are positive. Also all its principal minors are positive, see [6], page 464, or [7], page 559, so that L_0 is well- defined and non-singular. Eqs. (6) and (7) imply the similarity relationship

$$\mathrm{B}_{r+1} = {\mathrm{L}_r}^T \mathrm{B}_r ({\mathrm{L}_r}^T)^{-1}, \tag{8}$$

the characteristic equation of B_{r+1} coincides with that of B_r, so that B_{r+1} has the same spectrum of eigenvalues as B_r. Hence it can be proved that all the matrices B_r of Definition 5 are positive definite, with well-defined Cholesky factorisations. Further it is easy to use (6) and (7) to prove that all the matrices B_r of Definition 5 are symmetric and tridiagonal.

Definition 6. Let H_s^r be the $s \times s$ Hankel matrix, $s = 1, \ldots, n$, with elements

$$[H_s^r]_{ij} = x_{r+i+j-2}, \;\; 1 \leq i, j \leq s, \;\; r = 0, 1, \ldots, \tag{9}$$

and define the Hankel variable h_s^r of X_n by

$$h_s^r = \det\, H_s^r. \tag{10}$$

1.2 The aims and content of this paper

The underlying purpose of this paper to attain suitably great understanding of the sequence X_n and its properties for use in algebraic studies in which the Jacobi matrix J_n plays a central role. Criticial in this context is the determination of a suitable set of initial conditions $\mathrm{IC}(X_n)$ for the $x_r \in X_n$, a non-trivial matter for $\mathrm{DE}(X_n)$ of general order n. The choice adopted here – see Definition 8 below – emerged naturally and compellingly from the analysis described below of the QR algorithm, from the successive treatments made early on in the programme for $n = 2, 3, 4, 5$.

The main aim explicitly realised in the paper is the use of the $x_r \in X_n$ to obtain a concise parametrisation of the various matrices occurring in the development of the QR and QD algorithms for the Jacobi matrix J_n. More precisely, it states and proves correct, in a collection of theorems, explicit closed form expressions for all of the elements of all of the matrices

$$\mathrm{A}_r, \mathrm{Q}_r, \mathrm{R}_r, \mathrm{B}_r, \mathrm{L}_r, \quad r = 0, 1, \ldots\ , \tag{11}$$

in terms of the $x_r \in X_n$. Further, the equalities

$$\begin{aligned} \mathrm{A}_r &= \mathrm{B}_{2r} && (12)\\ (\mathrm{R}_r)^T &= \mathrm{L}_{2r}\,\mathrm{L}_{2r+1} && (13) \end{aligned}$$

are shown to be true for all $r = 0, 1, \ldots$.

All the desired expressions emerge in compact form already in terms of the Hankel variables of Definition 6. They can then be simplified greatly by using the Rutishauser variables of Definition 9 below.

It should be emphasised at the outset that this paper is not offered as a contribution to the practical use of the two algorithms in the determination of the eigenvalues of J_n. Rather it is concerned with the *existence* of explicit closed form expressions for important quantities in the general area, showing that such expressions can be displayed in a very compact form, which can be used to establish results such as those of (12) and (13).

The development of this paper motivated also the study of another area in which the Jacobi matrix J_n provides the starting point. This is the subject of the paper [8]. Since the $\mathrm{CE}(\mathrm{J}_n)$ of Definition 1 leads directly to the $\mathrm{DE}(X_n)$, and the discrete variables $x_r \in X_n$, an obvious question arisies: what DE is satisfied by the Hankel variables h_s^r? The answer is obtained by reference to the s-th compound matrix $C_s(\mathrm{J}_n)$ of the matrix J_n. The DE asked for bears exactly the same relationship to the characteristic equation $\mathrm{CE}(C_s(\mathrm{J}_n))$ as $\mathrm{DE}(X_n)$ bears to $\mathrm{CE}(\mathrm{J}_n)$. The article [8] reviews the properties of compound matrices in the situation just described, and some others. Some references to compound matrices are [6, 9, 10]. These provide respectively clear introductory material, a comprehensive listing of results, and a systematic exterior algebra approach.

The material of this paper can now be summarised. In Section 2, on the basis of the Definition 8 of the $\mathrm{IC}(X_\beta)$ for $\beta = 2, \ldots, n$, a body of explicit expressions of $x_r^{(\beta)}$ for increasing values of

β is presented. Various trends in the evolution of such expressions are made visible. Theorems governing these trends are proved in Section 3. Theorem 1 replaces the result for the IC(X_β) contained in Definition 8 by a stronger one, and thereby expedites not only the proofs Theorems 2 and 3, but also of the lemmas of Section 4. Theorem 1 also allows a fuller, and essential, understanding of the values of β, or the levels β, at which certain $x_r^{(\beta)}$ attain their final X_n values. These final values, valid for use in algebraic work at level n, correspond to the values of r such that $x_r^{(\alpha)} = x_r^{(\beta)}$ for all $\alpha = \beta+1, \ldots, n$. Such final X_n values can be written as x_r without superscript (n).

In Section 4, Hankel variables h_s^r and Rutishauser's variables q_s^r and e_s^r are introduced. The Desnanot-Jacobi theorem, Theorem 4, and various Lemmas involving them are proved.

In Sections 5 and 6, attention turns to the major results of this paper concerning the elements of the matrices (11) of the QR and QD algorithms. The latter, although it is rerhaps of lesser importance than the former, is taken up first, because it is very much easier to treat. Section 5 presents formulas for the elements of the matrices $\mathrm{B}_r, \mathrm{L}_r,\quad r = 0, 1, \ldots,$ and proves that they satisfy (6) and (7). Further, using the Lemmas of Section 4, it is proved that insertion of $r = 0$ into the general expressions for the elements of B_r gives agreement with the identification $\mathrm{B}_0 = \mathrm{J}_n$, and with the explicit Cholesky factorisation of J_n.

Section 6 presents formulas for the elements of the matrices $\mathrm{Q}_r, \mathrm{R}_r, \mathrm{A}_r,\quad r = 0, 1, \ldots,$ and proves that they satisfy (4) and (5). The formulas for the elements of Q_r quoted are the results of its factorisation into a product of Givens rotations performed in an obvious and natural order. Theorem 4 plays the key role of ensuring the orthogonality of each one and hence of their product Q_r. For $r = 0$ it is verified explicitly that $\mathrm{A}_0 = \mathrm{J}_n$, and that Q_0 and R_0 agree with the explicit QR factorisation of J_n.

It may also be worth emphasising that the results found in Sections 5 and 6 for the two algorithms are all given in terms of the elements of the single sequence, X_n, and can (almost) all be caste advantageously in terms of the Rutishauser variables. Further the latter fact expedites easy proof, as Theorem 9 of Sec. 7, of (12) and (13). As in the discussion in hapter 3 of [2], all the expressions displayed in Definition 11 and Theorem 5 are well-defined. Then in virtue of Theorem 9, the same applies to the espressions displayed in Definition 13 and theorem 7.

2 Some data for the sequence X_n

Here some explicit expressions for $x_r^{(\beta)}$ are displayed to indicate how the IC(X_β) evolve through increasing values of β until the IC(X_n) are reached, and how the values of β are known for which the $x_r^{(\beta)}$ attain their final values x_r, valid for use at the final X_n level.

Definition 7. Let CVk(X_β) be the k-th value, $k = 1, 2, \ldots,$ calculated in X_β using the DE(X_β) and the IC(X_β)

Deal explicitly first with $\beta = 2$ with X_2 which is fully determined by the elements of $\mathrm{J}_2 = \begin{pmatrix} b_1 & a_1 \\ a_1 & b_2 \end{pmatrix}$, writing

$$\mathrm{IC}(X_2): \quad x_0 = 1, \quad x_1 = b_1. \tag{14}$$

$$\begin{aligned} x_{r+2}^{(2)} \quad &= \quad T_{2\,1} x_{r+1}^{(2)} - T_{2\,2} x_r^{(2)} \\ T_{2\,1} \quad &= b_1 + b_2 = \mathrm{tr}\ \mathrm{J}_2 \\ T_{2\,2} \quad &= b_1 b_2 - a_1^2 = \det \mathrm{J}_2 = \Delta_2, \end{aligned} \tag{15}$$

and evaluating

$$\begin{aligned} CV1(X_2): \quad x_2 = \quad & b_1^2 + a_1^2, \\ CV2(X_2): \quad x_3 = \quad & b_1^3 + a_1^2(2b_1 + b_2) \end{aligned} \tag{16}$$

$$\begin{aligned}
CV3(X_2):\quad x_4^{(2)} &= b_1^4 + a_1^2(3b_1^2 + 2b_1b_2 + b_2^2) + a_1^4 && (17)\\
CV4(X_2):\quad x_5^{(2)} &= b_1^5 + a_1^2(4b_1^3 + 3b_1^2b_2 + 2b_2^2b_1 + b_2^3) + a_1^4(3b_1 + 2b_2) && (18)\\
CV5(X_2):\quad x_6^{(2)} &= b_1^6 + a_1^2(5b_1^4 + 4b_1^3b_2 + 3b_1^2b_2^2 + 2b_1b_2^3 + b_2^4)\\
&\quad + a_1^4(6b_1^2 + 6b_1b_2 + 3b_2^2) + a_1^6. && (19)
\end{aligned}$$

The retention of the superscript (2) for CVk(X_2), $\quad k = 3, 4, 5$, is to be noted.

Definition 8. The initial conditions IC(X_β), $\beta = 2, \ldots, n$, are fully determined by means of

$$x_r^{(\beta)} = x_r^{(\beta-1)}, \quad r = 0, \ldots, \beta - 1, \tag{20}$$

i.e. by taking for IC(X_β), the IC($X_{\beta-1}$), plus the first calculated value CV1($X_{\beta-1}$).

Remark 2. Eq. (20) assigns the IC(X_β), but a stronger result emerges in the form of Theorem 1, proved in Section 3, which states that (20) holds for all $r = 0, \ldots, 2\beta - 3$.

For $\beta = 3$, with X_3 fully determined in terms of the elements of J$_3$, write

$$\mathrm{IC}(X_3): \quad x_0 = 1, \quad x_1 = b_1, \quad x_2 = b_1^2 + a_1^2 \tag{21}$$

$$\begin{aligned}
x_{r+3}^{(3)} &= T_{3\,1}x_{r+2}^{(3)} - T_{3\,2}x_{r+1}^{(3)} + T_{3\,3}x_r^{(3)}\\
T_{3\,1} &= b_1 + b_2 + b_3 = \mathrm{tr}\ \mathrm{J}_3\\
T_{3\,2} &= b_1b_2 + b_1b_3 + b_2b_3 - a_1^2 - a_2^2\\
T_{3\,3} &= b_1b_2b_3 - b_1a_2^2 - b_3a_1^2 = \det \mathrm{J}_3 = \Delta_3. && (22)
\end{aligned}$$

Eqs. (21) and (22) lead to

$$\begin{aligned}
CV1(X_3):\quad x_3 &= CV2(X_2) && (23)\\
CV2(X_3):\quad x_4 &= CV3(X_2) + a_2^2\, a_1^2 && (24)\\
CV3(X_3):\quad x_5 &= CV4(X_2) + a_2^2\, a_1^2\, (2b_1 + 2b_2 + b_3). && (25)\\
CV4(X_3):\quad x_6^{(3)} &= CV5(X_2) + a_2^2a_1^2(3b_1^2 + 4b_1b_2 + 3b_2^2 + 2b_1b_3 + 2b_2a_3 + b_3^2)\\
&\quad + 2a_1^4a_2^2 + a_1^2a_2^4. && (26)
\end{aligned}$$

Similarly for $\beta = 4$, there follows

$$\mathrm{IC}(X_4): \quad x_r = x_r^{(3)}, \quad r = 0, \ldots, 3. \tag{27}$$

$$\begin{aligned}
CV1(X_4):\quad x_4 &= CV2(X_3)\\
CV2(X_4):\quad x_5 &= CV3(X_3)\\
CV3(X_4):\quad x_6 &= CV4(X_3) + a_3^2\, a_2^2\, a_1^2 && (28)\\
CV4(X_4):\quad x_7 &= CV5(X_3) + a_3^2\, a_2^2\, a_1^2\, (2b_1 + 2b_2 + 2b_3 + b_4). && (29)
\end{aligned}$$

Remark 2 is very important, since it shows that the final X_n values of the x_r are attained at much lower levels β than would follow if no more than Definition 8 were true. In fact the tractability found in uses here of the sequence X_n depends critically on this fact, as is clear in the proofs of the theorems of Section 3. What can be seen in the data given in this section, and proved generally true in Section 3, is encapsulated in the following table. All x_r values that have reached their final X_n values are shown for each level β and only those. The letter IC has been inserted to emphasise that the IC(X_β) set only those entries that lie to the left of them. Those to the right emerge as a result of the result proved as Theorem 1 below.

β	0	1	2	3	4	5	6	7
1	x_0 IC	x_1						
2	x_0	x_1 IC	x_2	x_3				
3	x_0	x_1	x_2 IC	x_3	x_4	x_5		
4	x_0	x_1	x$_2$	x_3 IC	x_4	x_5	x_6	x_7

Alternatively it may be stated that that the pairs

$$(x_0, x_1), \;\; (x_2, x_3), \;\; \ldots \;\; (x_{2\beta-2}, x_{2\beta-1}) \tag{30}$$

attain their final values – the ones that are valid for use in the context of X_n – within

$$X_1, X_2, \ldots, X_\beta.$$

Thus in situations in X_n work in which x_r enter only for $r \leq 2\beta - 1$, the DE(X_β) can be used.

3 Theorems for variables of $X_\beta, \;\; \beta = 2, \ldots, n$

To establish the generalisation of (20) indicated in Remark 2 begin with

Lemma 1. The dependence on $b_\beta, a_{\beta-1}$ of the invariants $T_{\beta k}, \quad k = 1, \ldots, \beta$, of the Jacobi matrix J_β is given, for $\beta = 1, \ldots, n$, by

$$\begin{aligned}
T_{\beta\, 0} &= T_{\beta-1\, 0} = 1 \\
T_{\beta\, 1} &= T_{\beta-1\, 1} + b_\beta \\
T_{\beta\, k} &= T_{\beta-1\, k} + b_\beta T_{\beta-1\, k-1} - a_{\beta-1}^2 T_{\beta-2\, k-2}, \quad k = 2, \ldots \beta - 1 \\
T_{\beta\, \beta} &= b_\beta T_{\beta-1\, \beta-1} - a_{\beta-1}^2 T_{\beta-2\, \beta-2}.
\end{aligned} \tag{31}$$

Proof. This follows from the expansion of det $(\lambda \mathrm{I} - \mathrm{J}_\beta)$ on its last row or column.

Theorem 1. For each $\beta = 2, \ldots, n$, the equalities

$$x_r^{(\beta)} = x_r^{(\beta-1)} \tag{32}$$

hold for all $r = 0, \ldots, 2\beta - 3$.

Proof. Let $\mathcal{P}_\beta$ denote the statement that has to be proved. It is empty for $\beta = 2$. In the case of $\mathcal{P}_3$, the entry CV1(X_3) of (23) can be read as $x_3^{(3)} = x_3^{(2)}$. Since $2\beta - 3 = 3$ when $\beta = 3$, this means that $\mathcal{P}_3$ is true.

To provide an inductive proof of the theorem, assume that all of $\mathcal{P}_2, \ldots, \mathcal{P}_{\beta-1}$ are true, and turn attention to $\mathcal{P}_\beta$.

Begin by considering $x_\beta^{(\beta)}$ for which DE(X_β) and (20) give

$$x_\beta^{(\beta)} = \sum_{k=1}^{\beta} T_{\beta\, k} x_{\beta-k}^{(\beta)} (-1)^{k+1} = \sum_{k=1}^{\beta} T_{\beta\, k} x_{\beta-k}^{(\beta-1)} (-1)^{k+1}. \tag{33}$$

Use of Lemma 1 now gives

$$\begin{aligned}
x_\beta^{(\beta)} \;\; = \;\; & (T_{\beta-1\, 1} + b_\beta) x_{\beta-1}^{(\beta-1)} \\
& + \sum_{k=2}^{\beta-1} (T_{\beta-1\, k} + b_\beta T_{\beta-1\, k-1} - a_{\beta-1}^2 T_{\beta-2\, k-2}) x_{\beta-k}^{(\beta-1)} (-1)^{k+1} \\
& + (b_\beta T_{\beta-1\, \beta-1} - a_{\beta-1}^2 T_{\beta-2\, \beta-2}) x_0^{(\beta-1)} (-1)^{\beta+1}.
\end{aligned} \tag{34}$$

The terms on the right of (34) that do not contain b_β or $a_{\beta-1}$ evaluate to $x_\beta^{(\beta-1)}$ using $\mathrm{DE}(X_{\beta-1})$, and this DE also shows the terms proportional to b_β evaluate to zero. In the treatment of the terms proportional to $a_{\beta-1}^2$, $\mathcal{P}_{\beta-1}$ allows use of

$$x_r^{(\beta-1)} = x_r^{(\beta-2)} \tag{35}$$

for the set of $\beta-1$ values $r = 0,\ldots,\beta-2$. Then $\mathrm{DE}(X_{\beta-2})$ shows that these terms also evaluate to zero. So proof is complete that

$$x_\beta^{(\beta)} = x_\beta^{(\beta-1)}. \tag{36}$$

It is to be observed however that $\mathcal{P}_{\beta-1}$ justifies the truth of (35) not only for the set of r values $0,\ldots,\beta-2$ but for all r values up to $r = 2\beta-5$. reaching the highest r value for which treatment by $\mathcal{P}_{\beta-1}$ is justified. □

As (30) implies, $x_{2\beta-2}^{(\beta)}$ and $x_{2\beta-1}^{(\beta)}$ are the final X_n values of $x_{2\beta-1}$ and $x_{2\beta-1}$ at level β, but Theorem 1 does not apply to them. However the data presented in Section 2 suggests the truth of the following two general results.

Theorem 2. The identity

$$x_{2\beta-2} = x_{2\beta-2}^{(\beta)} = x_{2\beta-2}^{(\beta-1)} + (a_{\beta-1}\ldots a_1)^2 \tag{37}$$

holds for $\beta = 1, 2, \ldots$.

Proof. Let $\mathcal{Q}_\beta$ be the statement that (37) is true within X_β. Note that (16) shows that $\mathcal{Q}_2$ is true. Assume all of $\mathcal{Q}_2,\ldots,\mathcal{Q}_{\beta-1}$ are true, and turn attention to $\mathcal{Q}_\beta$.

The procedure followed in the proof of Theorem 1 to establish (32) for $r = 2\beta-3$ breaks down for $r = 2\beta-2$ but only for the terms proportional to ${a_{\beta-1}}^2$. Here use of the identity

$$x_{2\beta-4}^{(\beta-1)} = x_{2\beta-4}^{(\beta-2)} + (a_{\beta-2}\ldots a_1)^2 \tag{38}$$

is justified by $\mathcal{Q}_{\beta-1}$. Thus a use of $\mathrm{DE}(X_{\beta-2})$ such as was made in the proof of Theorem 1 can now be made here. The truth of $\mathcal{Q}_\beta$ follows directly, and the inductive proof of the theorem is complete. □

Theorem 3. The identity

$$x_{2\beta-1} = x_{2\beta-1}^{(\beta)} = x_{2\beta-1}^{(\beta-1)} + (a_{\beta-1}\ldots a_1)^2(2b_1+\ldots+2b_{\beta-1}+b_\beta) \tag{39}$$

is valid for $\beta = 1, 2, \ldots$.

Proof. This follows the same pattern as the proof of Theorem 2, requiring only attention to detail to find that the right side of (39) is in fact the sum of three distinct contributions.

A further result, which is quoted without proof because it is not used below, goes one stage beyond Theorem 3:

$$x_{2\beta}^{(\beta)} = x_{2\beta}^{(\beta-1)} + a_1^2\ldots a_{\beta-1}^2\left[b^T M b + 2a_1^2 + \ldots + 2a_{\beta-2}^2 + a_{\beta-1}^2\right], \tag{40}$$

where

$$\begin{aligned} b^T &= (b_1,\ldots,b_\beta) \\ M = M^T &= \begin{pmatrix} 3 & 2 & \ldots & 2 & 1 \\ 2 & 3 & \ldots & 2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 2 & 2 & \ldots & 3 & 1 \\ 1 & 1 & \ldots & 1 & 1 \end{pmatrix}. \end{aligned} \tag{41}$$

While $x^{(\beta)}_{2\beta-2}$ and $x^{(\beta)}_{2\beta-1}$ enter Theorems 2 and 3, which give their final X_n values, $x^{(\beta)}_{2\beta}$ in (40) is not a final X_n value. However Theorem 2 applies and gives

$$x^{(\beta+1)}_{2\beta} = x^{(\beta)}_{2\beta} + a_1^2 \dots a_\beta^2 = x_{2\beta}, \tag{42}$$

and this is a final X_n value.

4 Hankel and Rutishauser variables

4.1 Definitions and the Desnanot-Jacobi theorem

To present the results of theorems in Sections 5 and 6 in concise form and to complete their proofs, use of the Hankel variables naturally associated with X_n is called for. While these have been defined already, Definition 6, it is often helpful to use the notation

$$h^r_s = |x_r x_{r+1} x_{r+2} \dots x_{r+s-1}|, \tag{43}$$

displaying the first row of H^r_s within brackets $|\dots|$. Thus

$$\begin{aligned} h^r_1 &= x_r \\ h^r_2 &= |x_r x_{r+1}| = \begin{vmatrix} x_r & x_{r+1} \\ x_{r+1} & x_{r+2} \end{vmatrix} = x_r x_{r+2} - (x_{r+1})^2, \qquad (44) \\ h^r_3 &= |x_r x_{r+1} x_{r+2}| = \begin{vmatrix} x_r & x_{r+1} & x_{r+2} \\ x_{r+1} & x_{r+2} & x_{r+3} \\ x_{r+2} & x_{r+3} & x_{r+4} \end{vmatrix} \qquad (45) \end{aligned}$$

for $r = 0, 1, \dots$.

Remark 3. To present various formulas, here and similarly below, in a uniform way, without qualification in certain special cases, it is convenient to define further variables, here

$$h^r_0 = 1, \quad h^r_{-1} = 0 \tag{46}$$

for all r.

Remark 4. Use of DE(X_n) yields the results

$$h^r_n = \Delta_n\, h^{r-1}_n \quad \text{and} \quad h^r_{n+1} = 0, \tag{47}$$

for all r, where $\Delta_n = T_{n\,n} = \det \mathrm{J}_n$.

Theorem 4. The Hankel variables of Definition 6 obey the key identity

$$h^r_s h^{r+2}_{s-2} = h^r_{s-1} h^{r+2}_{s-1} - (h^{r+1}_{s-1})^2, \tag{48}$$

valid for all $r \geq 0$ and $s = 1, 2, \dots n$.

Proof. This follows from the Desnanot-Jacobi theorem, given as eq. (3.29) on page 111 of [11]. It is found also as equation (9) on page 99 of [9], and, in the form (48), on page 37 of [2]. Given the $n \times n$ matrix A, let $(\mathrm{A})_{ij\dots,kl\dots}$ denote the matrix that remains when rows $i, j, \dots$ and columns $k, l, \dots$ are deleted. Then [11] [9]

$$\begin{vmatrix} \det\,(A_{11}) & \det\,(A_{1n}) \\ \det\,(A_{n1}) & \det\,(A_{nn}) \end{vmatrix} = \det\,(\mathrm{A})\, \det\,(A_{1n,1n}), \tag{49}$$

The version of this result that is required here can be written as

$$|x_{r+2}x_{r+3}\dots x_{r+s-1}||x_r x_{r+1}\dots x_{r+s-1}|$$
$$= |x_r x_{r+1}\dots x_{r+s-2}||x_{r+2}x_{r+3}\dots x_{r+s}| - |x_{r+1}x_{r+2}\dots x_{r+s-1}|^2. \tag{50}$$

which translates directly into (48) via (43). □

It is easy to check that (48) is correct by direct manipulation of determinantal quantities, at least for low s values, *e.g.* $s = 2, 3$. Also the case $s = 1$ is trivial because of (46). The same applies to the case $s = n + 1$ because of (47).

4.2 Rutishauser's variables

In [2], page 37, it is stated that the proper variables for presentation of results in the context of the QD algorithm are those employed in [5]:

Definition 9. Rutishauser's variables are given in terms of Hankel variables by

$$q^r_s = \frac{h^{r+1}_s h^r_{s-1}}{h^r_s h^{r+1}_{s-1}} \tag{51}$$

$$e^r_s = \frac{h^{r+1}_{s-1} h^r_{s+1}}{h^r_s h^{r+1}_s}. \tag{52}$$

Two useful results involving them are those of

Lemma 2.

$$q^r_s e^r_s = q^{r-1}_{s+1} e^{r-1}_s = \frac{h^r_{s+1} h^r_{s-1}}{(h^r_s)^2} \tag{53}$$

$$q^r_s - e^{r+1}_{s-1} = q^{r+1}_s - e^r_s = \frac{h^{r+1}_s h^{r+1}_{s-1}}{h^r_s h^{r+2}_{s-1}} \tag{54}$$

Proof. The identity (53) follows directly from Definition 9. A single use of (48) for each of the first two expressions in (54) shows that they are each equal to the third one.

In Sections 5 and 6 expressions for all the matrix elements involved are given first in terms of Hankel variables, and second in terms of Rutishauser variables. In the case of the QD algorithm many of the results presented in Section 5 are to be found in [2], around page 37. The same is not true in the discussion in Section 6 of the QR algorithm. The expressions given there, and the relationships between them contained in Theorem 9, are believed to be new here whether written in terms of Hankel or of Rutishauser variables.

4.3 Lemmas involving the Hankel and Rutishauser variables

The proofs of Theorems 5 – 8 below require some lemmas, collected here.

Lemma 3. The identity

$$h^1_\beta = \Delta_\beta h^0_\beta \tag{55}$$

where $\Delta_\beta = T_{\beta\,\beta} = \det \mathrm{J}_\beta$, is valid for all $\beta = 1, \dots, n$, the case $\beta = n$ having been noted already under Remark 2.

Proof. Recall these cases of (48)

$$\begin{aligned} h^0_\beta &= |x_0 \dots x_{\beta-1}| \\ h^1_\beta &= |x_1 \dots x_\beta| \end{aligned} \tag{56}$$

and note that the x_r values of highest r in the two cases are $x_{2\beta-2}$ and $x_{2\beta-1}$. This means that all the x_r values involved have reached their final X_n values within X_β, so that $\mathrm{DE}(X_\beta)$ can be used on all the entries of the last column of h^1_β, which leads directly to the result (55).

Lemma 4. The identity

$$h^2_\beta = \Delta_\beta h^1_\beta + (a_\beta \dots a_1)^2 h^2_{\beta-1} \tag{57}$$

is valid for all $\beta = 1, \dots, n$.

Proof. For $h^2_\beta = |x_2 \dots x_{\beta+1}|$ the variable x_r of highest r value is $x_{2\beta}$. This is the only entry which has not reached its final value within X_β. The final value within X_n for $x_{2\beta}$ is reached first within $X_{\beta+1}$, and so is given, according to Theorem 2, by

$$x^{(\beta)}_{2\beta} + (a_\beta \dots a_1)^2. \tag{58}$$

The determinantal definition of h^2_β thus can be written as the obvious sum of two determinants, in virtue of (58). Hence use of $\mathrm{DE}(X_\beta)$ on the last column of the first of these is justified and this leads to (57).

Lemma 5. The identity

$$h^0_{\beta+1} = (a_\beta \dots a_1)^2 h^0_\beta \tag{59}$$

is valid for all $\beta = 1, \dots, n-1$.

Proof. Putting $s = \beta + 1$ and $r = 0$ into (48) gives

$$h^0_\beta h^2_\beta = (h^1_\beta)^2 + h^0_{\beta+1} h^2_{\beta-1}. \tag{60}$$

Multiplying (58) by h^0_β, using (55) and comparing with (60) leads to (59).

Definition 10. Define Γ_β for $\beta = 1, \dots, n-1$, and M_β, by means of

$$\Gamma_\beta = a_1^{2\beta} a_2^{2\beta-2} \dots a_\beta^2 \tag{61}$$

$$h^2_\beta = (M_\beta)^2 \Gamma_{\beta-1}, \tag{62}$$

and

$$N_\beta = \Delta_{\beta-1}\Delta_\beta + b_{\beta+1}(M_{\beta-1})^2. \tag{63}$$

Lemma 6. This simply lists some useful results.

$$\begin{aligned} \Gamma_\beta &= (a_\beta \dots a_1)^2 \Gamma_{\beta-1} && (64) \\ h^0_\beta &= \Gamma_{\beta-1} && (65) \\ h^1_\beta &= \Delta_\beta \Gamma_{\beta-1} && (66) \\ a^2_\beta &= \frac{\Gamma_\beta \Gamma_{\beta-2}}{{\Gamma_{\beta-1}}^2}. && (67) \\ \Delta_\beta &= b_\beta \Delta_{\beta-1} - a^2_{\beta-1}\Delta_{\beta-2}. && (68) \end{aligned}$$

Lemma 7. The identity

$$(M_\beta)^2 = (\Delta_\beta)^2 + a_\beta^2 (M_{\beta-1})^2 \tag{69}$$

is valid for all $\beta = 1, \ldots, n$.

Proof. Rewrite (57), with the aid of (66) and (64), in the form

$$h_\beta^2 = (\Delta_\beta)^2 \Gamma_{\beta-1} + \frac{\Gamma_\beta}{\Gamma_{\beta-1}} h_{\beta-1}^2, \tag{70}$$

and use (62) and (67). This gives (69).

Lemma 8. Three results involving Rutishauser variables are

$$\begin{aligned} q_\beta^0 &= \frac{\Delta_\beta}{\Delta_{\beta-1}} && (71) \\ e_\beta^0 &= a_\beta^2 \frac{\Delta_{\beta-1}}{\Delta_\beta} && (72) \\ e_{\beta-1}^0 + q_\beta^0 &= b_\beta. && (73) \end{aligned}$$

Proof. The first two results follow from Definition 9 and Lemma 6. The third one follows use firstly of (71) and (72) in

$$e_{\beta-1}^0 + q_\beta^0 = \frac{1}{\Delta_{\beta-1}} \left[\Delta_\beta + a_{\beta-1}^2 \Delta_{\beta-2} \right] = b_\beta, \tag{74}$$

and then use of (68).

There is another result, not needed below, which is accessible using Theorems 2 and 3 and the method used in the proofs of Lemmas 3 and 4:

$$h_\beta^3 = \Delta_\beta^3 h_\beta^0 + (a_1 \ldots a_\beta)^2 \left[b_{\beta+1}\, h_{\beta-1}^3 + 2\Delta_\beta\, h_{\beta-1}^2 - (a_1 \ldots a_\beta)^2\, h_{\beta-2}^3 \right]. \tag{75}$$

5 The QD algorithm

Recall that the notation B is used in QD algorithm work to denote the $n \times n$ Jacobi matrix J_n of Definition 1. The algorithm has been described – see Definition 5 – above. Note also that the matrices B_r for all $r \geq 0$ share with $\mathrm{B}_0 = \mathrm{B}$ key properties, such as trace, determinant, spectrum, tridiagonal nature, and that the latter property leads to

$$(\mathrm{L}_r)_{ij} = 0 \quad \text{if either} \quad i > j+1 \quad \text{or} \quad i < j. \tag{76}$$

5.1 Statement and proof of Theorems 5 and 6

Definition 11. The elements of the matrices L_r, $\quad r = 0, 1, 2, \ldots$, are governed by (76) and

$$\begin{aligned} (\mathrm{L}_r)_{\beta\beta-1} &= \sqrt{\frac{h_\beta^r h_{\beta-2}^{r+1}}{h_{\beta-1}^r h_{\beta-1}^{r+1}}} = \sqrt{e_{\beta-1}^r}, \quad \beta = 2, 3, \ldots, n && (77) \\ (\mathrm{L}_r)_{\beta\beta} &= \sqrt{\frac{h_{\beta-1}^r h_\beta^{r+1}}{h_{\beta-1}^{r+1} h_\beta^r}} = \sqrt{q_\beta^r} \quad \beta = 1, 2, \ldots, n. && (78) \end{aligned}$$

Theorem 5. The product $\mathrm{L}_r \mathrm{L}_r{}^T$ defines a symmetric tridiagonal matrix B_r, $\quad r = 0, 1, 2, \ldots,$ with elements given by

$$
\begin{aligned}
(\mathrm{B}_r)_{\beta\gamma} &= 0 \qquad\qquad\qquad\qquad \beta > \gamma + 1 && (79)\\
(\mathrm{B}_r)_{\beta\beta-1} &= (\mathrm{L}_r)_{\beta\beta-1}(\mathrm{L}_r)_{\beta-1\beta-1} \\
&= \sqrt{\frac{h^r_\beta h^r_{\beta-2}}{(h^r_{\beta-1})^2}} \quad = \quad \sqrt{q^r_{\beta-1} e^r_{\beta-1}} && (80)\\
(\mathrm{B}_r)_{\beta\beta} &= (\mathrm{L}_r)_{\beta\beta-1}(\mathrm{L}_r)_{\beta\beta-1} + (\mathrm{L}_r)_{\beta\beta}(\mathrm{L}_r)_{\beta\beta}, \quad \beta = 2, 3, \ldots, n \\
&= \frac{h^r_\beta h^{r+1}_{\beta-2}}{h^r_{\beta-1} h^{r+1}_{\beta-1}} \quad + \quad \frac{h^r_{\beta-1} h^{r+1}_\beta}{h^r_\beta h^{r+1}_{\beta-1}} \\
&= e^r_{\beta-1} + q^r_\beta, \quad \beta = 1, 2, \ldots, n. && (81)
\end{aligned}
$$

Proof. The product clearly defines a symmetric tridiagonal matrix. Equally clearly, use of the results (77) and (78) from Definition 11 gives rise to the matrix B_r defined by (79 – 81). □

Theorem 6. The product $\mathrm{L}_r{}^T \mathrm{L}_r$ defines the symmetric tridiagonal matrix B_{r+1}, $\quad r = 0, 1, 2, \ldots,$ the elements of which arise from those of B_{r+1}, *i.e.* from (80) and (81), by the replacement of r by $r + 1$.

Proof. The required results for the off-diagonal elements of B_{r+1} arise directly. For the diagonal elements, find

$$
\begin{aligned}
(\mathrm{L}_r^T \mathrm{L}_r)_{\beta\beta} &= (\mathrm{L}_r)_{\beta\beta}(\mathrm{L}_r)_{\beta\beta} + (\mathrm{L}_r)_{\beta+1\beta}(\mathrm{L}_r)_{\beta+1\beta} \\
&= q^r_\beta + e^r_\beta \\
&= q^{r+1}_\beta + e^{r+1}_{\beta-1} \\
&= (\mathrm{B}_{r+1})_{\beta\beta}. && (82)
\end{aligned}
$$

For the second line here input from Definition 11 is used. Then (54) is used, and comparison with (81) with r replaced by $r + 1$ allows the proof of Theorem 6 to be completed. □

5.2 Agreement with the Cholesky factorisation of J_n

Sequences of matrices $\mathrm{B}_r, \mathrm{L}_r$, $\quad r = 0, 1, \ldots,$ have been constructed and shown to satisfy eqs. (6) and (7) of the *QD* algorithm. In view of the origins of these definitions, it is necessary to verify that $\mathrm{B}_0 = \mathrm{J}_n$ and that the Cholesky factorisation of J_n is consistent with the $r = 0$ case of Definition 11.

Results from Lemma 8 give rise immediately to

$$
\begin{aligned}
(\mathrm{B}_0)_{\beta\beta-1} &= \sqrt{q^0_{\beta-1} e^0_{\beta-1}} = a_{\beta-1}, && (83)\\
(\mathrm{B}_0)_{\beta\beta} &= b_\beta, && (84)
\end{aligned}
$$

so that $\mathrm{B}_0 = \mathrm{J}_n$ follows. Similarly

$$
(\mathrm{L}_0)_{\beta\,\beta} = \sqrt{q^0_\beta} = \sqrt{\frac{\Delta_\beta}{\Delta_{\beta-1}}} \qquad (85)
$$

$$
(\mathrm{L}_0)_{\beta\,\beta-1} = \sqrt{e^0_{\beta-1}} = a_{\beta-1}\sqrt{\frac{\Delta_{\beta-2}}{\Delta_{\beta-1}}}, \qquad (86)
$$

in exact agreement with the Cholesky factorisation of J_n.

6 The QR algorithm

Recall that in QR algorithm work the notation A is used to denote the Jacobi matrix J_n of Definition 1. The algorithm has been described in above – see Definition 4. It is well-known that the matrices A_r for all $r \geq 0$ share with $\mathrm{A}_0 = \mathrm{A}$ key properties, such as trace, determinant, spectrum, and, crucially, tridiagonal nature.

Two consequences of the last property emerge. First the Q_r are upper Hessenberg, which means that below the first sub-diagonal all matrix elements are equal to zero. Second the R_r have non-zero elements only in the main diagonal and the two diagonals above it, *i.e.*

$$(\mathrm{R}_r)_{ij} = 0 \quad \text{if either} \quad i > j \quad \text{or} \quad i < j - 2. \tag{87}$$

6.1 Q_r and Givens rotations

The calculation ${\mathrm{Q}_r}^T \mathrm{A}_r = \mathrm{R}_r$, converting tridiagonal A_r to upper triangular R_r, can be undertaken by using a sequence of Givens rotations $(\mathrm{G}_r)_\alpha, \quad \alpha = 1, \dots n-1$, acting in the space of matrix rows and columns α and $\alpha + 1$. This leads to

Definition 12. The matrix Q_r for $\alpha = 1, \dots n-1$, and for all $r \in \mathbb{Z}$ is determined by

$$\mathrm{Q}_r = (\mathrm{G}_r)_1 (\mathrm{G}_r)_2 \dots (\mathrm{G}_r)_{n-1}, \tag{88}$$

where the non-trivial 2×2 submatrix

$$\begin{pmatrix} \cos(\theta_r)_\alpha & -\sin(\theta_r)_\alpha \\ \sin(\theta_r)_\alpha & \cos(\theta_r)_\alpha \end{pmatrix} \tag{89}$$

of each $(\mathrm{G}_r)_\alpha$ is defined by the formulas

$$\cos(\theta_r)_\alpha = \frac{h_\alpha^{2r+1}}{\sqrt{h_\alpha^{2r} h_\alpha^{2r+2}}} \tag{90}$$

$$\sin(\theta_r)_\alpha = \frac{\sqrt{h_{\alpha-1}^{2r+2} h_{\alpha+1}^{2r}}}{\sqrt{h_\alpha^{2r} h_\alpha^{2r+2}}} = \sqrt{\frac{e_\alpha^{2r+1}}{q_{\alpha+1}^{2r}}} \tag{91}$$

Eqs. (90) and (91) should agree with $\cos^2(\theta_r)_\alpha + \sin^2(\theta_r)_\alpha = 1$. To see that they do, refer to (48) with s and r replaced by $\alpha + 1$ and $2r$. The orthogonality of all the Givens rotation matrices $(G_r)_\alpha, \quad \alpha = 1, \dots (n-1), \quad r = 0, 1, \dots$, is ensured by (90) and (91).

Note also that (90) and (91) are well–defined also for $\alpha = 0$ and $\alpha = n$, with $\cos(\theta_r)_\alpha = 1$ and $\sin(\theta_r)_\alpha = 0$ for $\alpha = 0$ and $\alpha = n$.

Remark 5. Orthogonality considerations require the truth of (48) *only* for even values of what is denoted by r therein.

Example. For $4 \times 4 \quad \mathrm{A}_r$, the matrix

$$\mathrm{Q}_r = (\mathrm{G}_r)_1 (\mathrm{G}_r)_2 (\mathrm{G}_r)_3 \tag{92}$$

takes the form

$$\mathrm{Q}_r = \begin{pmatrix} c_1 & -s_1 c_2 & s_1 s_2 c_3 & -s_1 s_2 s_3 \\ s_1 & c_1 c_2 & -c_1 s_2 c_3 & c_1 s_2 s_3 \\ 0 & s_2 & c_2 c_3 & -c_2 s_3 \\ 0 & 0 & s_3 & c_3 \end{pmatrix}, \tag{93}$$

where $c_\alpha = \cos(\theta_r)_\alpha$ and $s_\alpha = \sin(\theta_r)_\alpha$ for $\alpha = 1, 2, 3$.

Insertion of $c_0 = 1$ and $c_4 = 1$ suitably into (93) allows it to be rewritten in a way that makes clear patterns which persist for $n > 4$

$$\mathrm{Q}_r = \begin{pmatrix} c_0c_1 & -c_0s_1c_2 & c_0s_1s_2c_3 & -c_0s_1s_2s_3c_4 \\ s_1 & c_1c_2 & -c_1s_2c_3 & c_1s_2s_3c_4 \\ 0 & s_2 & c_2c_3 & -c_2s_3c_4 \\ 0 & 0 & s_3 & c_3c_4 \end{pmatrix}. \tag{94}$$

It is easy to infer and prove two general results:

$$(\mathrm{Q}_r)_{\beta\beta-1} = \sin(\theta_r)_{\beta-1} \qquad \beta = 2,3,\ldots,n, \tag{95}$$

$$(\mathrm{Q}_r)_{\beta\beta} = \cos(\theta_r)_{\beta-1}\cos(\theta_r)_{\beta} \qquad \beta = 1,2,\ldots,n. \tag{96}$$

6.2 Statement and proof of Theorems 8 and 9

Definition 13. The elements of the matrices R_r, $\quad r = 0,1,2,\ldots,$ are governed by (87) and

$$(R_r)_{\beta\beta} = \sqrt{\frac{h^{2r}_{\beta-1}h^{2r+2}_{\beta}}{h^{2r+2}_{\beta-1}h^{2r}_{\beta}}} = \sqrt{q^{2r}_{\beta}\, q^{2r+1}_{\beta}}, \quad \beta = 1,2,\ldots,n, \tag{97}$$

$$\begin{aligned}(R_r)_{\beta-1\beta} &= \frac{h^{2r+1}_{\beta-2}}{h^{2r+1}_{\beta-1}}\sqrt{\frac{h^{2r+2}_{\beta-1}h^{2r}_{\beta}}{h^{2r+2}_{\beta-2}h^{2r}_{\beta-1}}} + \frac{h^{2r+1}_{\beta}}{h^{2r+1}_{\beta-1}}\sqrt{\frac{h^{2r+2}_{\beta-2}h^{2r}_{\beta-1}}{h^{2r+2}_{\beta-1}h^{2r}_{\beta}}} \\ &= \sqrt{q^{2r+1}_{\beta-1}\, e^{2r}_{\beta-1}} + \sqrt{q^{2r}_{\beta}\, e^{2r+1}_{\beta-1}}, \qquad \beta = 2,3,\ldots,n, \end{aligned} \tag{98}$$

$$(R_r)_{\beta-2\beta} = \sqrt{\frac{h^{2r+2}_{\beta-3}h^{2r}_{\beta}}{h^{2r+2}_{\beta-2}h^{2r}_{\beta-1}}} = \sqrt{e^{2r}_{\beta-1}\, e^{2r+1}_{\beta-2}} \quad \beta = 3,4,\ldots,n. \tag{99}$$

Theorem 7. The product $\mathrm{Q}_r\mathrm{R}_r$ defines a symmetric tridiagonal matrix A_r, $\quad r = 0,1,2,\ldots,$ with elements given by

$$(\mathrm{A}_r)_{\beta\gamma} = 0, \qquad \beta > \gamma + 1 \tag{100}$$

$$\begin{aligned}(\mathrm{A}_r)_{\beta\beta-1} &= (\mathrm{Q}_r)_{\beta\beta-1}(\mathrm{R}_r)_{\beta-1\beta-1} \\ &= \frac{\sqrt{h^{2r}_{\beta-2}h^{2r}_{\beta}}}{h^{2r}_{\beta-1}} = \sqrt{q^{2r}_{\beta-1}\, e^{2r}_{\beta-1}} \quad \beta = 2,3,\ldots,n,\end{aligned} \tag{101}$$

$$\begin{aligned}(\mathrm{A}_r)_{\beta\beta} &= (\mathrm{Q}_r)_{\beta\beta-1}(\mathrm{R}_r)_{\beta-1\beta} + (\mathrm{Q}_r)_{\beta\beta}(\mathrm{R}_r)_{\beta\beta} \\ &= \frac{h^{2r+1}_{\beta-2}h^{2r}_{\beta}}{h^{2r+1}_{\beta-1}h^{2r}_{\beta-1}} + \frac{h^{2r+2}_{\beta-2}h^{2r+1}_{\beta}}{h^{2r+2}_{\beta-1}h^{2r+1}_{\beta-1}} + \frac{h^{2r+1}_{\beta}h^{2r+1}_{\beta-1}}{h^{2r}_{\beta}h^{2r+2}_{\beta-1}} \\ &= e^{2r}_{\beta-1} + e^{2r+1}_{\beta-1} + (q^{2r}_{\beta} - e^{2r+1}_{\beta-1}) \end{aligned} \tag{102}$$

$$= e^{2r}_{\beta-1} + q^{2r}_{\beta}, \quad \beta = 1,2,\ldots,n. \tag{103}$$

Remark 6. The dependence of (97–103) on $2r$ rather than r is once more to be noted.

Proof. Since Q_r is upper Hessenberg and R_r obeys (87), there are no non-zero contributions to the product $(\mathrm{Q}_r\mathrm{R}_r)_{\beta\gamma}$ for $\beta > \gamma + 1$, so that (100) is proved.

The non-trivial expressions displayed in (101) and (103) in terms of Hankel variables arise directly from the formation of the product $\mathrm{Q}_r\mathrm{R}_r$, with (96) providing the needed elements of Q_r. Evaluating the third term of (103) in terms of Rutishauser's variables takes some extra work, but produces a result seen to be consistent with (81).

The structure of the QR algorithm requires that the matrices A_r are symmetric and tridiagonal. Do the definitions of Q_r and R_r just presented respect this?

The proof that $(\mathrm{A}_r)_{\beta-s\beta} = 0$ for $s \geq 2$ is greatly epedited by noting that

$$(\mathrm{Q}_r)_{\beta-s\beta} = (-1)^s F_{r\beta} G_{r\beta-s}, \tag{104}$$

where

$$F_{r\beta} = \frac{h_\beta^{2r+1}}{\sqrt{h_\beta^{2r+2} h_{\beta-1}^{2r+2}}}, \quad G_{r\beta} = \frac{h_{\beta-1}^{2r+1}}{\sqrt{h_\beta^{2r} h_{\beta-1}^{2r}}}, \tag{105}$$

for $s = 0, 1, \ldots$. This enables the writing of

$$\begin{aligned} (\mathrm{A}_r)_{\beta-s\beta} &= (\mathrm{Q}_r)_{\beta-s\beta-2}(\mathrm{R}_r)_{\beta-2\beta} + (\mathrm{Q}_r)_{\beta-s\beta-1}(\mathrm{R}_r)_{\beta-1\beta} + (\mathrm{Q}_r)_{\beta-s\beta}(\mathrm{R}_r)_{\beta\beta} \\ &= (-1)^s G_{r\beta-s}\left[F_{r\beta-2}(\mathrm{R}_r)_{\beta-2\beta} + F_{r\beta-1}(\mathrm{R}_r)_{\beta-1\beta} + F_{r\beta}(\mathrm{R}_r)_{\beta\beta}\right]. \end{aligned} \tag{106}$$

Eq. (98) shows that the second term within the square brackets of (106) itself is the sum of two terms. Each one of these cancels one of the two other terms within the square brackets of (106). This gives $(\mathrm{A}_r)_{\beta-s\beta} = 0$ for each $s = 2, \ldots, \beta-1$. The upper triangular and hence tridiagonal nature of A_r has thus been shown.

To prove that A_r is symmetric requires the evaluation of

$$(\mathrm{A}_r)_{\beta-1\beta} = (\mathrm{Q}_r)_{\beta-1\beta-2}(\mathrm{R}_r)_{\beta-2\beta} + (\mathrm{Q}_r)_{\beta-1\beta-1}(\mathrm{R}_r)_{\beta-1\beta} + (\mathrm{Q}_r)_{\beta-1\beta}(\mathrm{R}_r)_{\beta\beta}, \tag{107}$$

which contains the quantity

$$(\mathrm{Q}_r)_{\beta-1\beta-2} = \sin(\theta_r)_{\beta-2} \tag{108}$$

to which (104) is seen not to apply. However (104) does apply in the remaining terms, showing that they reduce, by a cancellation of the same sort as found in the previous manipulation, to a single term. Now, with the aid of a single use of (48) with r, s replaced by $2r, \beta-1$, it is found that this single term combines with the first term of (107) to produce an expression for $(\mathrm{A}_r)_{\beta-1\beta}$ that agrees with (101).

Thus, the product $\mathrm{Q}_r\mathrm{R}_r$ has been shown to define a symmetric tridiagonal matrix A_r, in accordance with the factorisation step (4) of the QR algorithm. □

Theorem 8. The product $\mathrm{R}_r\mathrm{Q}_r$ defines the symmetric tridiagonal matrix A_{r+1}, $\quad r = 0, 1, 2, \ldots,$ the elements of which arise from those of A_{r+1}, *i.e.* from (100–103), by the replacement of r by $r+1$.

Proof. Since $\mathrm{A}_{r+1} = \mathrm{R}_r\mathrm{Q}_r = \mathrm{Q}_r^T\mathrm{A}_r\mathrm{Q}_r$, the product $\mathrm{R}_r\mathrm{Q}_r$ defines a symmetric matrix. Since $(\mathrm{R}_r\mathrm{Q}_r)_{\beta\gamma} = 0$ for $\beta > \gamma+1$, obviously, it follows that $(\mathrm{A}_{r+1})_{\beta\gamma} = 0$ for $\gamma > \beta+1$, so that A_{r+1} is tridiagonal. It may however be found either amusing or instructive to verify that $(\mathrm{A}_{r+1})_{\beta\beta+s} = 0$, for $s = 2, 3, \ldots$, is true, alternatively, using

$$(\mathrm{A}_{r+1})_{\beta\beta+s} = (-1)^s F_{r\beta+s}\left[(\mathrm{R}_r)_{\beta\beta}G_{r\beta} - (\mathrm{R}_r)_{\beta\beta+1}G_{r\beta+1} + (\mathrm{R}_r)_{\beta\beta+2}G_{r\beta+2}\right] \tag{109}$$

for $s = 2, 3, \ldots$. The square bracket evaluates to zero as above, but this time the roles of F and G are interchanged.

It therefore remains to show that

$$\begin{aligned} (\mathrm{R}_r\mathrm{Q}_r)_{\beta\beta-1} &= (\mathrm{R}_r)_{\beta\beta}(\mathrm{Q}_r)_{\beta\beta-1} = (\mathrm{A}_{r+1})_{\beta\beta-1} \qquad (110) \\ (\mathrm{R}_r\mathrm{Q}_r)_{\beta\beta} &= (\mathrm{R}_r)_{\beta\beta}(\mathrm{Q}_r)_{\beta\beta} + (\mathrm{R}_r)_{\beta\beta+1}(\mathrm{Q}_r)_{\beta+1\beta} = (\mathrm{A}_{r+1})_{\beta\beta}. \qquad (111) \end{aligned}$$

Evaluating the left side of (110) with the aid of (97) and (96), gives immediate agreement with the result of replacing r by $r+1$ in (101). But more effort is needed to prove that (111) is true.

For the first term of $(\mathrm{R}_r\mathrm{Q}_r)_{\beta\beta}$, (97) and (95) are needed. One factor of the expression so obtained is given by

$$\sqrt{q_\beta^{2r}}\ \cos(\theta_r)_{\beta-1} = \sqrt{q_\beta^{2r} - e_{\beta-1}^{2r+1}}, \tag{112}$$

upon use of (90) and (91). The other factor requires one use of (54) to show that

$$\sqrt{q_\beta^{2r+1}}\ \cos(\theta_r)_\beta = \sqrt{q_\beta^{2r+1} - e_\beta^{2r}}. \tag{113}$$

The two square roots shown in (112) and (113) are equal because of (54).

For the second term of $(\mathrm{R}_r\mathrm{Q}_r)_{\beta\beta}$, (98) and (95) are needed, and one use of (54) yields its value so that

$$\begin{aligned}(\mathrm{R}_r\mathrm{Q}_r)_{\beta\beta} &= (q_\beta^{2r+1} - e_\beta^{2r}) + (e_\beta^{2r} + e_\beta^{2r+1}) \\ &= q_\beta^{2r+1} + e_\beta^{2r+1} = q_\beta^{2r+2} + e_{\beta-1}^{2r+2} \\ &= (\mathrm{A}_{r+1})_{\beta\beta}.\end{aligned} \tag{114}$$

One more use of (54) has been made, then a comparison with (103).

The product $\mathrm{R}_r\mathrm{Q}_r$ has been shown to define a symmetric tridiagonal matrix A_{r+1}, in accordance with the definition step (5) of the QR algorithm, and consistently with the definition of A_r in its factorisation step. □

Remark 7. The proof here contrasts the essential nature of the dependence on $2r$ of the definitions made in this section, with the fact that (48) is valid, as a purely algebraic statement for all r, not only even values.

6.3 Agreement with the QR-factorisation of J_n

Sequences of matrices $\mathrm{A}_r, \mathrm{Q}_r, \mathrm{R}_r \quad r = 0, 1, \ldots$, have been constructed and shown to satisfy eqs. (4) and (5) of the QR algorithm. In view of the origins of these definitions, it is necessary to verify that the sequences constructed do indeed relate correctly to the matrix J_n which is the starting point of the algorithm.

The proof that $\mathrm{A}_0 = \mathrm{J}_n$ is the same as given in Section 5.2.

Use of the Lemmas of Section 4.3 leads to various results. First

$$\cos(\theta_0)_\beta = \frac{\Delta_\beta}{M_\beta}, \quad \sin(\theta_0)_\beta = a_\beta \frac{M_{\beta-1}}{M_\beta}, \tag{115}$$

consistently in view of (69). Second

$$\begin{aligned}(\mathrm{Q}_0)_{\beta\beta} &= \frac{\Delta_{\beta-1}\Delta_\beta}{M_\beta M_{\beta-1}} \\ -(\mathrm{Q}_0)_{\beta-1\beta} &= a_{\beta-1}\frac{\Delta_{\beta-2}\Delta_\beta}{M_\beta M_{\beta-1}} \\ (\mathrm{Q}_0)_{\beta-2\beta} &= a_{\beta-2}a_{\beta-1}\frac{\Delta_{\beta-3}\Delta_\beta}{M_\beta M_{\beta-1}},\end{aligned} \tag{116}$$

generalising obviously to $(\mathrm{Q}_0)_{\beta-s\beta}$, and compatibly with (104). Third

$$(\mathrm{R}_0)_{\beta-2\beta} = a_{\beta-1}a_{\beta-2}\frac{M_{\beta-3}}{M_{\beta-2}}, \quad (\mathrm{R}_0)_{\beta-1\beta} = \frac{a_{\beta-1}N_{\beta-1}}{M_{\beta-1}M_{\beta-2}}, \quad (\mathrm{R}_0)_{\beta\beta} = \frac{M_\beta}{M_{\beta-1}}, \tag{117}$$

where N_β is defined by (63). Some routine work on the QR factorisation of J_n confirms the correctness of these results.

7 The identities $\mathrm{A}_r = \mathrm{B}_{2r}$ and $(\mathrm{R}_r)^T = \mathrm{L}_{2r}\mathrm{L}_{2r+1}$

Theorem 9. The identities (12) and (13) hold for $r = 0, 1, \ldots$.

Proof. Inspection of the expressions found in Theorems 7 and 5 for the matrix elements of A_r and B_r imply immediately that $\mathrm{A}_r = \mathrm{B}_{2r}$ holds for all $r = 0, 1, \ldots$.

It can be shown easily that

$$\mathrm{A}_{r+1} = \mathrm{R}_r\, \mathrm{A}_r\, (\mathrm{R}_r)^{-1}, \tag{118}$$

and that (8) implies the result

$$\mathrm{B}_{2r+2} = \mathrm{S}_r \mathrm{B}_{2r} (\mathrm{S}_r)^{-1}, \tag{119}$$

or in virtue of (12), already proved true,

$$\mathrm{A}_{r+1} = \mathrm{S}_r \mathrm{A}_r (\mathrm{S}_r)^{-1}, \tag{120}$$

where

$$\mathrm{S}_r = (\mathrm{L}_{2r}\mathrm{L}_{2r+1})^T. \tag{121}$$

The elements of S_r can be calculated from Definition 11. The identity $\mathrm{R}_r = \mathrm{S}_r$ needed to complete the proof of (12) is immediately seen to hold for all $r = 0, 1, \ldots$.

References

[1] G. H. Golub and C. F. van Loan, Matrix computation, Johns Hopkins University Press, 1984.

[2] G. H. Golub and G. Meurant, Matrices, moments and quadratures with applications, Princeton University Press, Princeton 2010.

[3] P. Olver and C. Shakiban, Applied linear algebra, Pearson Education Inc., London, 2006.

[4] D. S. Watkins, The algebraic eigenvalue problem: *GR* and Krylov subspace methods, CUP, Cambridge, 2008.

[5] H. Rutishauser, *Solution of eigenvalue problems with the LR transformation*, Nat. Bur. Standards Appl. Math. Ser. **49** 47-81 (1958).

[6] D. S. Bernstein, *Matrix Math: Theory, Facts and Formulas*, second edition, Princeton University Press, Princeton, 2009.

[7] C. D. Meyer, *Matrix Analysis and Appied Linear Algebra*, SIAM, 2000.

[8] A. J. Macfarlane, *On the compound matrices and symmetric matrix functions of a given symmetric matrix*. Visible under this title at www.damtp.cam.ac.uk/user/ajm/

[9] A. C. Aitken, Determinants and matrices, Oliver and Boyd, 9th edition, Edinburgh, 1956.

[10] M. Marcus, *Finite dimensional multilinear algebra, part 1*, M. Dekker, New York, 1975.

[11] D. M. Bressoud, Proofs and confirmations: the alternating sign matrix conjecture, MAA: Spectrum Series, CUP, 1999.