FROM QD TO LR AND QR, OR, HOW WERE THE QD AND LR ALGORITHMS DISCOVERED?

MARTIN H. GUTKNECHT AND BERESFORD N. PARLETT

Extended version of July 16, 2009

Abstract. Perhaps the most astonishing idea in eigenvalue computation is Rutishauser's idea of applying the LR transform to a matrix for generating a sequence of similar matrices that become more and more triangular. The same idea is the foundation of the ubiquitous QR algorithm. It is well known that this idea originated in Rutishauser's qd algorithm, which precedes the LR algorithm and can be understood as applying LR to a tridiagonal matrix. But how did Rutishauser discover qd, and when did he find the qd–LR connection? We checked some of the early sources and come up with an explanation.

In the year 2000 the QR algorithm was placed on the list of the top ten algorithms of the 20th century in the journal "Computers in Science and Engineering" [47]. The honor was well merited; the QR algorithm is the ubiquitous tool for computing eigenvalues of dense matrices. Its predecessor, the LR algorithm, is now largely forgotten and rarely taught to students. What we wish to say here is that, from an intellectual viewpoint, it was the LR algorithm that made the seminal contribution; QR was merely a stable version of LR. But how could anyone come up with the apparently ridiculous idea of factoring a square matrix $\mathbf{A} = \mathbf{A}_1$ into two triangular matrices, $\mathbf{A}_1 = \mathbf{L}_1\mathbf{R}_1$, and then forming a new matrix $\mathbf{A}_2 = \mathbf{R}_1\mathbf{L}_1$. This LR transform requires a lot of arithmetic and creates no zero entries in the matrix. But since $\mathbf{L}_1$, by convention, is lower triangular with ones on the diagonal, it is invertible and $\mathbf{A}_2$ has the same spectrum as $\mathbf{A}_1$. Yet it is not at all obvious that, if one is rich enough and keeps on computing $\mathbf{A}_3, \mathbf{A}_4, \mathbf{A}_5, \ldots$ by iterating the LR transform, then slowly but (nearly) surely the iterates become upper triangular and the diagonal entries converge to eigenvalues. Who would have thought of such a bizarre process?

It was Heinz Rutishauser who discovered this LR algorithm as a byproduct of his qd algorithm. In "The Algebraic Eigenvalue Problem" the eminent numerical analyst J.H. Wilkinson called Rutishauser an "algorithmic genius" [82, page vii], and regarding the LR algorithm he wrote in the same book [82, page 485]: "In my opinion its development is the most significant advance which has been made in connexion with the eigenvalue problem since the advent of automatic computers." Surely the invention of these two algorithms is evidence of genius. That still leaves open our question: How did he do it? What follows is our guess of how it happened.

Our main conclusion is that he found the qd algorithm by studying previous work of Hadamard [28], Aitken [1, 2], and Lanczos [39, Ch. VI], and by improving on it, while the LR algorithm and the many other applications he proposed for the qd algorithm resulted from his ability to quickly understand the many aspects and interrelations of a whole circle of ideas, which includes tridiagonal matrices, orthogonal polynomials, the corresponding types of continued fractions and their "partial sums" (convergents), which are Padé approximants, the Lanczos algorithm [39], and the related conjugate gradient (cg) [34] and biconjugate gradient (bicg) methods [41]. By now these connections are well known, but in the early 1950s this knowledge just emerged.

1. Historical Perspectives. After receiving a diploma in mathematics from ETH Zurich in 1942 Heinz Rutishauser (1918–1970) was first a teaching assistant for three years; then he became a high school teacher, but was at the same time still working on his PhD thesis on functions of several complex variables. Soon after defending it in 1948 — it was not published until 1950 [49] — he was hired as a research assistant by Prof. Eduard Stiefel (1909–1978), director of the newly founded Institute for Applied Mathematics at ETH Zurich. Stiefel, a former student of Heinz Hopf, had been a very successful researcher in pure mathematics before World War II, but spent much time during the war in the Swiss army, where he ultimately advanced to the rank of a colonel. After the war he decided to concentrate on the newly emerging field of numerical computing on electronic computers. His new, but still small institute made seminal contributions to hardware design, compiler techniques, and numerical algorithms. As is well known, Stiefel discovered the method of conjugate gradients (cg) at the same time as Magnus Hestenes [33, 74], and when they found out about this coincidence at a workshop organized by Olga Taussky in Los Angeles in August 1951, they decided to write a joint paper [34] that is considered now as one of the most influential ones in numerical analysis. At this workshop, Stiefel must also have learnt of Lanczos recent work on finding eigenvalues [39] and solving linear systems [41]. This work was intimately related to the conjugate gradient method and is also closely related to the topic addressed here, as we will see below. For a survey of the early contributions to the cg and Lanczos algorithms see the excellent annotated bibliography by Golub and O'Leary [21].

Rutishauser wrote a habilitation thesis in which he introduced a compiler [50], and soon after that he discovered the qd and LR algorithms [55, 54, 56, 57, 58]. This early work on qd and LR is collected in partly revised form in [62]. During the same period Rutishauser made other interesting contributions to numerical analysis, including one published 1953 on the nonsymmetric Lanczos algorithm and the reduction of a matrix to tridiagonal form [52]. This paper may have been crucial for the discovery of the LR algorithm for tridiagonal matrices and for Rutishausers' apparently small excitement about the LR algorithm for full matrices.

2. Stiefel's "assignment" for Rutishauser. On Stiefel's suggestion, around 1953, Rutishauser approached the key problem of determining the poles of a rational (or, more generally, meromorphic) function given by a power series in z^{-1},

$$f(z) := \sum_{\nu=0}^{\infty} \frac{s_\nu}{z^{\nu+1}} . \tag{2.1}$$

The application he had in mind was the following: Assume $\mathbf{A}$ is an $N \times N$ matrix and $\mathbf{x}_0$, $\mathbf{y}_0$ are two N-vectors. Then, for $s_\nu := \mathbf{y}_0^\mathsf{T} \mathbf{A}^\nu \mathbf{x}_0$, the series in (2.1) is the Taylor expansion at ∞ of

$$f(z) := \left\langle \mathbf{y}_0, (z\mathbf{I} - \mathbf{A})^{-1}\, \mathbf{x}_0 \right\rangle = \left\langle \mathbf{y}_0, \tfrac{1}{z}\, (\mathbf{I} - \tfrac{1}{z}\mathbf{A})^{-1}\, \mathbf{x}_0 \right\rangle , \tag{2.2}$$

which is a proper rational fraction of degree $n \le N$ whose poles are eigenvalues of $\mathbf{A}$. This is seen from the representation

$$f(z) = \frac{\mathbf{y}_0^\mathsf{T}\, \mathsf{adj}\, (z\mathbf{I} - \mathbf{A})\, \mathbf{x}_0}{\mathsf{det}(z\mathbf{I} - \mathbf{A})}, \tag{2.3}$$

which also reveals that only the numerator depends on $\mathbf{x}_0$ and $\mathbf{y}_0$ unless some zeros and poles cancel. This application to the matrix eigenvalue problem was the starting point and the target of Rutishauser's investigation. He called the coefficients s_ν *Schwarz constants*, but today they are referred to as *moments* in numerical linear algebra and as *Markov parameters* in systems and control theory, where the sequence of moments is the *impulse response* of the linear time-invariant discrete-time single-input single-output (SISO) control system given by the state matrix $\mathbf{A}$ and the vector $\mathbf{x}_0$ and $\mathbf{y}_0$. So, Stiefel's proposal for Rutishauser was to determine the eigenvalues of $\mathbf{A}$ given the sequence of moments. In the foreword of [62] Rutishauser wrote: "Following this suggestion the author developed an algorithm that solves the posed problem."

We next describe what was known before about this problem and how Rutishauser came up with his new solution and new insight. In view of the two dominant quantities involved he called his algorithm *quotient-difference algorithm* or, briefly, *QD algorithm*. Nowadays, the abbreviation in lower case letters, *qd algorithm*, is widely used to emphasize that q and d are not matrices (in contrast, say, to the LR and QR algorithms).

We know now that Stiefel's proposal was actually a bad one: the problem of determining the eigenvalues from the moments is typically extremely ill-conditioned. It is well-known that the elements of a symmetric tridiagonal matrix (and the nodes and weights of the corresponding Gauss-Christoffel quadrature formula) are badly determined by its moments; see, e.g., [15, 17]. Rutishauser became aware of this ill-conditioning and of a better solution of the matrix eigenvalue problem [57], namely using the Lanczos algorithm [39] for reducing the matrix to tridiagonal form and then applying the *progressive form* of his qd algorithm or, what amounts to the same, his *LR algorithm*. This approach may also run into stability problems or even break down, but this occurs only rarely, while ill-conditioning is nearly inescapable when using moments. Amazingly, in systems and control theory the matrix moments have nevertheless been used till the 1990's for exactly the task of finding a tridiagonal realization or reduced model of a system given by $\mathbf{A}$, $\mathbf{x}_0$, $\mathbf{y}_0$ as in the definition (2.2) of the transfer function. Only then the Lanczos algorithm and other alternatives became commonplace.

3. Finding the poles of f from the moments: Hadamard and Aitken. If f is a proper rational function q/p of degree n with explicitly known denominator

$$p(z) = \pi_0 z^n + \pi_1 z^{n-1} + \cdots + \pi_n \,,$$

it follows from the expansion (2.1) that the moments satisfy the difference equation

$$\pi_0 s_{k+n} + \pi_1 s_{k+n-1} + \cdots + \pi_n s_k = 0 \qquad (k \geq 0). \tag{3.1}$$

The initial values $s_0, \ldots, s_{n-1}$ are determined by the numerator polynomial q, while the recursion (3.1) only depends on p. It was known to Daniel Bernoulli (1700-1782) that if p has a unique zero z_1 of maximum modulus (and hence, the series (2.1) converges for $|z| > |z_1|$), then the solution $\{s_\nu\}$ of the difference equation (3.1) satisfies

$$\lim_{\nu\to\infty} \frac{s_{\nu+1}}{s_\nu} = z_1 \,. \tag{3.2}$$

This is Bernoulli's method for finding such a greatest root [7] (see also [1]).

J. König [38] established more than 150 years later that the analogous result holds for any power series of an analytic function with a single simple pole on the boundary of the disk of convergence. Soon after that the French mathematician Jacques Hadamard (1865–1963), in his thesis [28] published in 1892, solved the problem of finding *all* the poles of f from the moments by a beautiful procedure that is very ill-suited to computer implementation, however. Now the function just had to be meromorphic in a disk around the origin and analytic at the origin, where its Taylor series was given. Here we formulate the results assuming f is analytic at ∞ and given by the series (2.1). For simplicity, we further assume that f is a proper rational function of order n.

Hadamard considered the double sequence of *Hankel determinants*

$$H_k^{(\nu)} := \begin{vmatrix} s_\nu & s_{\nu+1} & \cdots & s_{\nu+k-1} \\ s_{\nu+1} & s_{\nu+2} & \cdots & s_{\nu+k} \\ \vdots & \vdots & \ddots & \vdots \\ s_{\nu+k-1} & s_{\nu+k} & \cdots & s_{\nu+2k-2} \end{vmatrix} \qquad (k = 1, 2, \ldots;\ \nu = 0, 1, \ldots) \tag{3.3}$$

and, adapted to our situation, established the following main result.

THEOREM 1. *[Hadamard (1892)] Assume the series* (2.1) *represents a rational function whose n poles, counted including multiplicity, are ordered such that*

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_{n-1}| \geq |\lambda_n| \,. \tag{3.4}$$

If $1 \leq k < n$ and $|\lambda_{k+1}| < \Lambda < |\lambda_k|$, or if $k = n$ and $\Lambda < |\lambda_n|$, then

$$H_k^{(\nu)} = \mathsf{const} \cdot (\lambda_1 \cdots \lambda_k)^\nu \left[1 + \mathcal{O}\left(\frac{\Lambda}{|\lambda_k|} \right)^\nu \right] \qquad as \quad \nu \to \infty \,. \tag{3.5}$$

Assuming simple poles Henrici [29] gave a simpler proof of this result. Multiple poles can be treated with a technique used by Golomb [20]. New proofs of Hadamard's theorem have also been a topic in the subsequent qd literature [23, 24, 36].

Here are some obvious conclusions.

COROLLARY 2. *Under the assumptions of Theorem 1,* $H_{n+1}^{(\nu)} = 0$ $(\forall \nu)$. *Moreover, if* f *has* n *simple poles, then*

1. $H_k^{(\nu)} \neq 0$ $(k = 1, \ldots, n)$ *for large enough* ν.
2. *If* $|\lambda_k| > |\lambda_{k+1}|$, *then*

$$\frac{H_k^{(\nu+1)}}{H_k^{(\nu)}} \to \lambda_1 \lambda_2 \cdots \lambda_k \qquad \text{as} \quad \nu \to \infty . \tag{3.6}$$

3. *If* $|\lambda_{k-1}| > |\lambda_k| > |\lambda_{k+1}|$, *then*

$$q_k^{(\nu)} := \frac{H_k^{(\nu+1)}}{H_k^{(\nu)}} \cdot \frac{H_{k-1}^{(\nu)}}{H_{k-1}^{(\nu+1)}} \to \lambda_k \qquad \text{as} \quad \nu \to \infty . \tag{3.7}$$

Statement (3.7) persists for $k = 1$ if we let $H_0^{(\nu)} := 1$ $(\forall \nu)$. In view of $H_1^{(\nu)} = s_\nu$, it reduces then to Bernoulli's result (3.2).

We observe that the LU decomposition of the Hankel matrix in (3.3), if it is possible without pivoting, yields for fixed ν the Hankel determinants $H_1^{(\nu)}, \ldots, H_N^{(\nu)}$ in one factorization. Nevertheless, computing this LU decomposition for a sufficient number of indices ν is an unacceptably big effort. Obviously, Hadamard was only interested in theorems, not in practical computation. The motivation for developing an efficient algorithm must have been missing, though, in fact, Hadamard had the key in his hands: the striking nonlinear relation

$$\left(H_k^{(\nu)}\right)^2 = H_k^{(\nu-1)} H_k^{(\nu+1)} - H_{k+1}^{(\nu-1)} H_{k-1}^{(\nu+1)} \tag{3.8}$$

among neighboring Hankel determinants. It is now often called *Jacobi's identity* for Hankel determinants [30]. Note that it expresses the square of $H_k^{(\nu)}$ as a difference of products of next neighbors.

It was the New Zealander Alexander Craig Aitken (1895–1967), in Scotland, who, in 1926 came up with the now obvious algorithmic conclusion [1, 2]. He was unaware of Hadamard's work, but rediscovered Theorem 1 and knew Jacobi's identity (3.8), which he considered as a special case of a "theorem of compound determinants". He realized that it can be used to build up — from the left or from the top — the triangular table

$$\begin{array}{cccccc} 1 & & & & & \\ 1 & H_1^{(0)} & & & & \\ 1 & H_1^{(1)} & H_2^{(0)} & & & \\ 1 & H_1^{(2)} & H_2^{(1)} & H_3^{(0)} & & \\ 1 & H_1^{(3)} & H_2^{(2)} & H_3^{(1)} & H_4^{(0)} & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array} \tag{3.9}$$

Unfortunately, when one of the determinants vanishes, both the horizontal and the vertical recursions break down. But similar types of breakdown can occur in many related recursive schemes.

Another constructive tool that was available when Rutishauser solved Stiefel's problem was the *Chebyshev algorithm* [8], that allows us to compute the recurrence coefficients of a set of orthogonal polynomials if the sequence of moments of the underlying weight function is known. Specifically, we need the first $2m$ moments to construct recursively the orthogonal polynomials up to degree m. Neither Stiefel nor Rutishauser seem to ever mention this tool, despite the fact that, in its initial phase, Rutishauser qd algorithm serves the same purpose, as we will see. The Chebyshev algorithm was later revived, analyzed, and modified by Sack and Donovan [71], Wheeler [79], and, in a series of papers starting with [16] by Gautschi, who also came up with the name *modified Chebyshev algorithm* for the more stable version using *modified moments*.

4. Rutishauser's qd algorithm. According to [55], Rutishauser was aware of the work of Hadamard [28], Aitken [1, 2], and Lanczos [40] when he worked on Stiefel's problem. It seems that in the second half of 1952 or early in 1953 he took Aitken's work, improved it in a significant way, and made the connections to a number of related topics and applications. The key result was his *qd algorithm*, on which he published three papers [55, 54, 56] in 1954, the first and most fundamental of which was received by ZAMP on Aug. 5, 1953. The following year he had yet another seminal article [57] on the application of qd to the eigenvalue problem. In partly revised form this early work on qd is collected in [62], which contains also some additional material, in particular a short appendix on the LR algorithm. Another appendix contains a shortened version of [56]. A first announcement on the qd algorithm had been made by Stiefel [75]. Of great importance for the dissemination of the qd algorithm was Henrici's review article [29], the first publication on qd in English. It appeared in Volume 49 of the Applied Mathematics Series of the National Bureau of Standards (NBS). The only two other papers in that 81-page volume are Rutishauser's main publication on the LR algorithm [63] and Stiefel's paper on kernel polynomials [77], which is also related to cg and qd. The volume was issued on January 15, 1958, but it seems to have been compiled long before. In fact, the preface is dated June 26, 1956. Moreover, Rutishauser [61] cites preprints of his and Stiefel's contributions [60, 76], dated 1956 and 1955, respectively.

Rutishauser does not clearly state how he found the qd algorithm; he only gives an indication, not a complete derivation. Henrici [29] writes that the qd algorithm "by a simple but ingenious modification of Aitken's method, entirely bypasses the computation of Hankel determinants". And that "It it is remarkable that in the computation of the q_k^n, the determinants H_k^n do not have to be used if a set of auxiliary quantities is introduced."

The details have been hinted to by Henrici [29] and worked out in [46]: first, in view of Hadamard's Theorem 1, in particular conclusion (3.7), the target of the computation are the quotients $q_k^{(\nu)}$. By multiplying Jacobi's identity (3.8) centered at $H_{k-1}^{(\nu+1)}$,

$$\left(H_{k-1}^{(\nu+1)}\right)^2 = H_{k-1}^{(\nu)}\, H_{k-1}^{(\nu+2)} - H_k^{(\nu)}\, H_{k-2}^{(\nu+2)} ,$$

with

$$\frac{H_k^{(\nu+1)}}{H_{k-1}^{(\nu+1)}\, H_k^{(\nu)}\, H_{k-1}^{(\nu+2)}}$$

we can turn the first term on the right-hand side into $q_k^{(\nu)}$:

$$\frac{H_{k-1}^{(\nu+1)}\, H_k^{(\nu+1)}}{H_k^{(\nu)}\, H_{k-1}^{(\nu+2)}} = q_k^{(\nu)} - \frac{H_{k-2}^{(\nu+2)}\, H_k^{(\nu+1)}}{H_{k-1}^{(\nu+1)}\, H_{k-1}^{(\nu+2)}} . \tag{4.1}$$

Likewise, we write down Jacobi's identity centered at $H_k^{(\nu+1)}$,

$$\left(H_k^{(\nu+1)}\right)^2 = H_k^{(\nu)}\, H_k^{(\nu+2)} - H_{k+1}^{(\nu)}\, H_{k-1}^{(\nu+2)} ,$$

and multiply it with

$$\frac{H_{k-1}^{(\nu+1)}}{H_k^{(\nu)}\, H_{k-1}^{(\nu+2)}\, H_k^{(\nu+1)}} .$$

This turns the first term on the right-hand side into $q_k^{(\nu+1)}$:

$$\frac{H_{k-1}^{(\nu+1)}\,H_k^{(\nu+1)}}{H_k^{(\nu)}\,H_{k-1}^{(\nu+2)}} = q_k^{(\nu+1)} - \frac{H_{k-1}^{(\nu+1)}\,H_{k+1}^{(\nu)}}{H_k^{(\nu)}\,H_k^{(\nu+1)}}\,. \tag{4.2}$$

Clearly the left-hand sides of (4.1) and (4.2) are identical and the second terms on the right-hand sides have the same structure. So, after introducing the auxiliary quantity

$$e_k^{(\nu)} := \frac{H_{k-1}^{(\nu+1)}\,H_{k+1}^{(\nu)}}{H_k^{(\nu)}\,H_k^{(\nu+1)}}\,, \tag{4.3}$$

we can conclude from (4.1) and (4.2) that

$$q_k^{(\nu)} + e_k^{(\nu)} = q_k^{(\nu+1)} + e_{k-1}^{(\nu+1)}\,. \tag{4.4}$$

This relation can be seen to also hold for $k = 1$ if we define $e_0^{(\nu)} := 0$ for all ν. In addition, from the definitions (3.7) of $q_k^{(\nu)}$ and (4.3) of $e_k^{(\nu)}$ it is readily verified that

$$q_{k+1}^{(\nu)}\,e_k^{(\nu)} = q_k^{(\nu+1)}\,e_k^{(\nu+1)}\,. \tag{4.5}$$

The relations (4.4) and (4.5) are the *rhombus rules* defining the qd algorithm.[1] Rutishauser [55] suggested to write down the quantities $e_k^{(\nu)}$ and $q_k^{(\nu)}$ in a triangular scheme called *qd scheme* (also known as *qd table*); see Fig. 4.1.[2] Recall that at this time, the simple computations needed to build up this scheme were normally done on a desk calculator, so a suitable scheme to write down the numbers obtained was most useful.

With the "initial values" $e_0^{(\nu)} = 0$ and $q_1^{(\nu)} = s_{\nu+1}/s_\nu$ the rhombus rules allow us to build up Rutishauser's qd scheme from the first column. Alternatively, they can be used to build up the scheme from its top diagonal, which is given by the coefficients of the formal S–fraction (5.3) of $f_0 = f$, to be discussed in Section 5. The latter application is called the *progressive qd algorithm*. It is the version that is still of importance.

In the case of a proper rational fraction f of exact degree n, as in (2.2), $e_n^{(\nu)} = 0$ holds for all ν, and thus the table is not defined beyond the nth e–column. Assuming that all the poles of f have different moduli, Rutishauser [55] could readily conclude from Aitken's work [1] that

$$\left.\begin{array}{rcl} \lim_{\nu\to\infty} q_k^{(\nu)} & = & \lambda_k \\ \lim_{\nu\to\infty} e_k^{(\nu)} & = & 0 \end{array}\right\} \qquad (k = 1, 2, \ldots, n)\,. \tag{4.6}$$

In [55] he also gave the generalization of this to meromorphic functions, referring for its proof to Hadamard ("The proof of the proposition can be extracted offhand from the above mentioned work of Hadamard [28, §§14–21]"). Among the two conclusions drawn from this generalization, but given without proof, there is one later referred to by Henrici [30] as *Rutishauser's rule*, which covers the case of poles with equal modulus. With tools from functional analysis this rule was first fully proved by Stewart [73]. A direct but complicated proof requiring an additional assumption was given by Henrici [30, pp. 642–650], who later in Part II of [32] admitted that his earlier proof in [29] is erroneous and promoted a new direct and simple proof due to his student Seewald [72].

[1] According to footnotes in [29] and [62] the name "rhombus rules" was coined by Stiefel; see [76, page 42] or [77, page 18].

[2] The first example of a qd scheme in [55] listed, for each ν, not only the two columns with $e_k^{(\nu)}$ and $q_k^{(\nu)}$, but additional two columns for $s_k^{(\nu)}$ and $d_k^{(\nu)} := q_k^{(\nu+1)} - q_k^{(\nu)}$. The rhombus rules made these two columns obsolete.

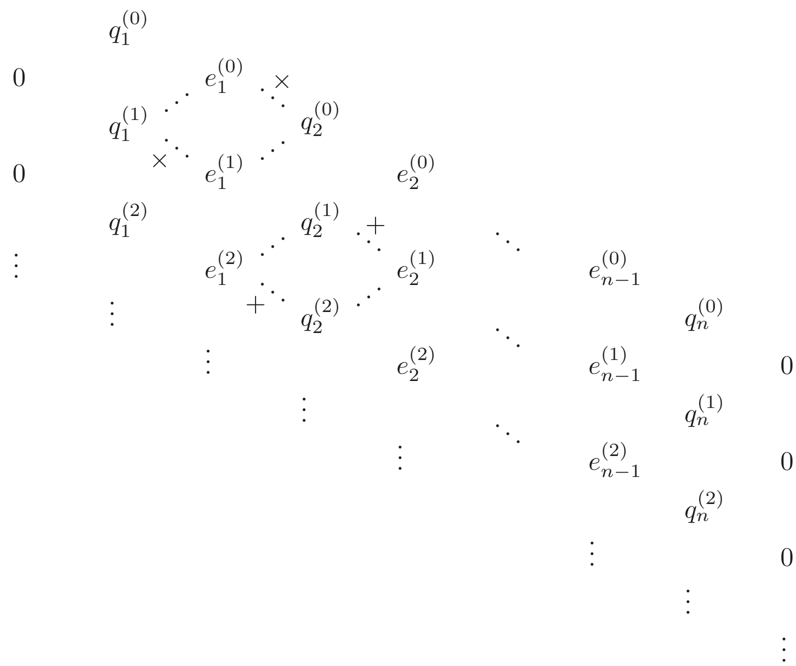

FIG. 4.1. *The qd table of a rational function f of degree n and manifestations of the two rhombus rules.*

5. Continued fractions and Padé approximants. The most elegant derivation of the qd algorithm is the algorithmic formulation of relations between various continued fractions in which the power series (2.1) can be expanded. Stiefel had an interest in continued fractions since he and Hestenes had discovered their hidden occurrence in the conjugate gradient method [34, § 18]. According to a footnote in [55] Stiefel pointed out to Rutishauser that “some proofs in the qd theory could be simplified using continued fractions”. This indicates that the following derivation is due to Stiefel — but we cannot be sure as the footnote does not specify which “proofs” it refers to. As a complex analyst of his time also Rutishauser must have known continued fractions well.[3]

Following [55], we recall that if all determinants $H_k^{(\nu)}$ and $H_k^{(\nu+1)}$ $(k = 1, 2, \dots)$ are nonzero, we can, for any nonnegative integer ν, expand the formal series[4]

$$f_\nu(z) := \sum_{k=0}^{\infty} \frac{s_{\nu+k}}{z^{k+1}} = z^\nu \left(f(z) - \sum_{k=0}^{\nu-1} \frac{s_k}{z^{k+1}} \right) \tag{5.1}$$

both into a *Jacobi fraction* or *J–fraction*

$$f_\nu(z) = \frac{s_\nu|}{|z - q_1^{(\nu)}} - \frac{e_1^{(\nu)} q_1^{(\nu)}|}{|z - q_2^{(\nu)} - e_1^{(\nu)}} - \frac{e_2^{(\nu)} q_2^{(\nu)}|}{|z - q_3^{(\nu)} - e_2^{(\nu)}} - \cdots \tag{5.2}$$

and into a formal *Stieltjes fraction*[5] or *S–fraction*

$$f_\nu(z) = \frac{s_\nu|}{|z} - \frac{q_1^{(\nu)}|}{|1} - \frac{e_1^{(\nu)}|}{|z} - \frac{q_2^{(\nu)}|}{|1} - \frac{e_2^{(\nu)}|}{|z} - \cdots . \tag{5.3}$$

[3]In fact, in the winter semester 1955/56 he gave a course “Introduction to the theory of continued fractions” (in German), but that was after he wrote [55].

[4]For readers unfamiliar with continued fractions we recommend Chapter 12 of [31], which also describes the connection to the qd algorithm.

[5]In true Stieltjes fractions the coefficients $q_k^{(\nu)}$ and $e_k^{(\nu)}$ are positive. Here this means that bfA is spd.

The J–fraction is the so-called *even part* of the S–fraction, which is obtained by merging two successive terms of the S–fraction into one. The *odd part* of the S–fraction is another J–fraction, obtained by merging the differently chosen pairs of successive terms into one,

$$f_\nu(z) = \frac{s_\nu}{z}\left\{1 + \frac{q_1^{(\nu)}}{\left| z - q_1^{(\nu)} - e_1^{(\nu)}\right.} - \frac{e_1^{(\nu)} q_2^{(\nu)}}{\left| z - q_2^{(\nu)} - e_2^{(\nu)}\right.} - \frac{e_2^{(\nu)} q_3^{(\nu)}}{\left| z - q_3^{(\nu)} - e_3^{(\nu)}\right.} - \cdots \right\}. \tag{5.4}$$

The rules for extracting the even or odd part of an S–fraction are well known; see, e.g., Jones and Thron [37, pp. 41–43]. By comparing the last J–fraction with the one for

$$f_{\nu+1}(z) = z f_\nu(z) - s_\nu \tag{5.5}$$

one readily rediscovers Rutishauser's rhombus rules (4.4) and (4.5).

With each continued fraction we can associate a sequence of "partial sums". In continued fraction theory, these are called convergents. For example, the kth *convergent* of the J–fraction (5.2) of f_ν is the proper rational fraction

$$\frac{n_k^{(\nu)}(z)}{p_k^{(\nu)}(z)} := \frac{s_\nu}{\left| z - q_1^{(\nu)}\right.} - \frac{e_1^{(\nu)} q_1^{(\nu)}}{\left| z - q_2^{(\nu)} - e_1^{(\nu)}\right.} - \cdots - \frac{e_{k-1}^{(\nu)} q_{k-1}^{(\nu)}}{\left| z - q_k^{(\nu)} - e_{k-1}^{(\nu)}\right.}. \tag{5.6}$$

It was well known [78] (and verified in [55]) that numerator and denominator satisfy the same recursion

$$\left.\begin{aligned} n_{k+1}^{(\nu)}(z) &:= \left(z - q_{k+1}^{(\nu)} - e_k^{(\nu)}\right) n_k^{(\nu)}(z) - e_k^{(\nu)} q_k^{(\nu)} n_{k-1}^{(\nu)}(z) \\ p_{k+1}^{(\nu)}(z) &:= \left(z - q_{k+1}^{(\nu)} - e_k^{(\nu)}\right) p_k^{(\nu)}(z) - e_k^{(\nu)} q_k^{(\nu)} p_{k-1}^{(\nu)}(z) \end{aligned}\right\} \quad (k = 1, 2, ...), \tag{5.7}$$

which is started with different initial conditions, however:

$$n_0^{(\nu)}(z) := 0, \quad p_0^{(\nu)}(z) := 1, \quad n_1^{(\nu)}(z) := s_\nu, \quad p_1^{(\nu)}(z) := z - q_1^{(\nu)}. \tag{5.8}$$

From the recursions it is seen that $p_k^{(\nu)}(z)$ has exact degree k and is monic, while $n_k^{(\nu)}(z)$ has exact degree $k-1$ and has leading coefficient s_ν.

Rutishauser [55] only introduced the denominators $p_k^{(\nu)}(z)$, which he displayed in a *P–scheme* (or, *P–table*) analogous to the qd scheme:

$$\begin{array}{cccccc} 1 = p_0^{(0)} & & & & & \\ & p_1^{(0)} & & & & \\ 1 = p_0^{(1)} & & p_2^{(0)} & & & \\ & p_1^{(1)} & & p_3^{(0)} & & \\ 1 = p_0^{(2)} & & p_2^{(1)} & & p_4^{(0)} & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array} \tag{5.9}$$

The recursion (5.7) links any three adjacent elements on a diagonal of this P–table, but using his rhombus rules Rutishauser [55] could also establish relations linking neighboring elements on different diagonals. Moreover, by referring to Wall [78] he noted that the nth convergent matches the first $2n$ moments of the power series for f_ν:

$$f_\nu(z) - \frac{n_k^{(\nu)}(z)}{p_k^{(\nu)}(z)} = \frac{s_{n+1}^{(\nu)}}{z^{2n+1}} + \mathcal{O}(z^{-2n-2}) \quad \text{as} \quad z \to \infty. \tag{5.10}$$

This moment matching property identifies the convergents of the J–fraction as *Padé approximants* of f_ν, though here the expansion is not at $z = 0$ but at $z = \infty$. The latter difference may be the reason why Rutishauser never seems to have mentioned

Padé approximation when referring to these convergents.[6] Later it was Gragg and Householder [22, 35, 36] who stressed the connection of the qd algorithm with Padé approximation. In fact, Householder wrote in [36]: "Moreover, the qd algorithm can hardly be understood except in terms of the Padé table." The *Padé table* is a table that contains as entries the Padé approximants of a (possibly only formal) power series. For a power series in z, the (m, n) Padé approximant is the rational function with numerator degree at most m and denominator degree at most n that matches as many terms as possible of the given series. For a series in z^{-1} it is easiest to adapt the definition by writing numerator and denominator as polynomials in z^{-1}; so here

$$\Big(z^{-k}\, n_k^{(\nu)}(z)\Big) \Big/ \Big(z^{-k}\, p_k^{(\nu)}(z)\Big)$$

can be viewed as (k, k) Padé approximant of $f_\nu(z)$, although it is only rational of type $(k-1, k)$ if the fraction is written in terms of z.

The Padé table fills the fourth quadrant of the plain, with the m–axis pointing downward and the n–axis pointing to the right. From (5.1) we have

$$f(z) \equiv f_0(z) = \frac{s_0}{z} + \frac{s_1}{z^2} + \cdots + \frac{s_{\nu-1}}{z^\nu} + \frac{f_\nu(z)}{z^\nu}\,. \tag{5.11}$$

So, expanding $f_\nu(z)$ into a series in z^{-1} leads immediately to the series for $f(z)$, and truncating the continued fraction of $f_\nu(z)$ leads to a rational function that is a Padé approximant of f. Specifically, the $(\nu + k, k)$ Padé approximant $n_{\nu+k,k}(z)/p_{k,k}(z)$ is given by

$$\frac{n_{\nu+k,k}(z)}{p_{\nu+k,k}(z)} := \frac{s_0}{z} + \frac{s_1}{z^2} + \cdots + \frac{s_{\nu-1}}{z^\nu} + \frac{1}{z^\nu}\,\frac{n_k^{(\nu)}(z)}{p_k^{(\nu)}(z)}\,. \tag{5.12}$$

So, Rutishauser's P–table contained exactly the denominators of the lower half of the Padé table of the given series f. Since he aimed at approximations for the poles of f he had no need for the numerators.[7]

An added value of the connection to Padé approximation is that the basic convergence theorem for sequences of Padé approximants along columns of the Padé table is roughly equivalent to Hadamard's theorem (Theorem 1 and Corollary 2), but easier to prove. Denote the extended complex plane by $\overline{\mathbb{C}}$ and assume the metric of the Riemann sphere on it, so that $\overline{\mathbb{C}}$ is compact. Then, adapted to our situation, this theorem published 1902 by de Montessus de Ballore [9] says that under the assumptions of Theorem 1, in particular those of (3.5),

$$\lim_{\nu\to\infty} \frac{n_k^{(\nu)}(z)}{p_k^{(\nu)}(z)} = f(z) \tag{5.13}$$

uniformly on any compact subset of

$$\big\{z \in \overline{\mathbb{C}}\,;\; |z| \geq \Lambda\,,\; z \neq \lambda_j \; (j = 1, \ldots, k)\big\}\,.$$

So, in particular, the zeros of the denominators $p_k^{(\nu)}$ converge to the k poles of f that lie outside the circle of radius Λ. An error formula in terms of a contour integral is also known. See, e.g., [6, § 6.3] for four different proofs based on Hermite's integral formula for the error in polynomial interpolation. Rutishauser nowhere refers to de Montessus de Ballore, however.

What is most important here is that Rutishauser [55] realized that the polynomials $p_k^{(0)}(z)$ in the top diagonal of the P–table are the same polynomials as those generated

[6] He must have been aware of Padé approximations as they are so closely related to J–fractions and S–fractions, and as the book of Wall [78] (referenced in [53]) contains a whole chapter on them.

[7] One cosmetic difference is that Rutishauser displayed the top "row" of his P–table with a 30° slope (as in (5.9)), while the slope of the corresponding diagonal of the Padé table is 45°.

implicitly by the *nonsymmetric Lanczos process* described in Lanczos seminal paper on eigenvalue computation [39]. Both polynomial sequences satisfy the same recurrence. The recurrence coefficients determine at the same time the top diagonal of the qd table, the continued fraction (J–fraction) for f, and the Lanczos process for $\mathbf{A}$ with initial vectors $\mathbf{x}_0$ and $\mathbf{y}_0$. We will return to these connections in Section 6.

The polynomials on the other diagonals of the P–table can also be viewed as Lanczos polynomials, provided we start the Lanczos process with suitable other initial vectors, e.g., $\mathbf{x}_0$ and $\mathbf{A}^\nu\mathbf{y}_0$, or we define it with a another (formal) inner product that contains an extra factor $\mathbf{A}^\nu$. So, on each diagonal we construct a sequence of formal orthogonal polynomials, but on each one the underlying formal weight function is different. In the case of a symmetric positive definite matrix $\mathbf{A}$, this relationship between the main diagonal of the P–table and its first subdiagonal describes the connection between the conjugate gradient method and the conjugate residual method; see [26].

6. Tridiagonal matrices. It is well known that there is an intimate, straightforward relationship between finite J–fractions

$$h(z) = \frac{s_0 \rfloor}{\lceil z-\alpha_1} - \frac{\beta_1 \rfloor}{\lceil z-\alpha_2} - \cdots - \frac{\beta_{n-1} \rfloor}{\lceil z-\alpha_n}, \tag{6.1}$$

the nonsymmetric Lanczos algorithm, and the tridiagonal matrices

$$\mathbf{T} = \begin{pmatrix} \alpha_1 & 1 & & & \\ \beta_1 & \alpha_2 & 1 & & \\ & \beta_2 & \alpha_3 & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & \beta_{n-1} & \alpha_n \end{pmatrix}. \tag{6.2}$$

For example, the denominator $p_n(z)$ of the rational function $h(z)$ is the characteristic polynomial of $\mathbf{T}$. This is quickly seen by expanding the determinant $|z\mathbf{I} - \mathbf{T}|$ along its last row and verifying that the resulting recursion is identical with the recursion for the denominators $p_k(z)$ of the convergents of the J–fraction,

$$p_{k+1}(z) := (z-\alpha_{k+1})p_k(z) - \beta_k p_{k-1}(z) \qquad (k = 0, 1, \dots, n-1), \tag{6.3}$$

started with $p_{-1}(z) := 0$, $p_0(z) := 1$, $\beta_0 := 0$. So, the eigenvalues of $\mathbf{T}$ are the zeros of p_n and the poles of h.

At the same time these recursions transform via $\mathbf{x}_k = p_k(\mathbf{A})\mathbf{x}_0$, $\mathbf{y}_k = p_k(\mathbf{A})\mathbf{y}_0$ (if we assume real data) into the recurrences for the Lanczos vectors:

$$\left.\begin{aligned} \mathbf{x}_{k+1} &:= \mathbf{A}\mathbf{x}_k - \alpha_{k+1}\,\mathbf{x}_k - \beta_k\,\mathbf{x}_{k-1} \\ \mathbf{y}_{k+1} &:= \mathbf{A}^\mathsf{T}\mathbf{y}_k - \alpha_{k+1}\,\mathbf{y}_k - \beta_k\,\mathbf{y}_{k-1} \end{aligned}\right\} \quad (k = 0, 1, \dots). \tag{6.4}$$

Therefore, the polynomials p_k are also referred to as the *Lanczos polynomials* associated with the Lanczos process for $\mathbf{A}$ with initial vectors $\mathbf{x}_0$ and $\mathbf{y}_0$.

As we know from [52, pp. 42 & 50] (submitted Sep. 8, 1952, eleven months before the first qd paper [55]), Rutishauser was aware of these connections.[8]

There are further connections between the rational function $h(z)$ and the tridiagonal matrix $\mathbf{T}$ that Rutishauser may have not yet known at this time. It can be shown that the eigenvectors of $\mathbf{T}$ are related to the partial fraction decomposition of h, but, according to Gautschi [18], this was probably observed only later; a 1962 reference is Wilf [80, Ch. 2, Exercise 9]. Another stunning fact is the magic formula

$$h(z) = s_0\,\mathbf{e}_1^\mathsf{T}(z\mathbf{I} - \mathbf{T})^{-1}\mathbf{e}_1\,, \tag{6.5}$$

[8] In [52] Rutishauser referred to tridiagonal matrices as *codiagonal*; the term "tridiagonal" was not yet generally accepted. It is also interesting to note that in books on orthogonal polynomials and continued fractions tridiagonal (or codiagonal) matrices are hardly ever mentioned, despite the close connections.

in which $\mathbf{e}_1$ is the first column of the n-by-n unit matrix. It is easily verified by making use of

$$\mathbf{e}_1^{\mathsf{T}} (z\mathbf{I} - \mathbf{T})^{-1} \mathbf{e}_1 = \frac{\mathbf{e}_1^{\mathsf{T}} \,\mathsf{adj}\,(z\mathbf{I} - \mathbf{T})\,\mathbf{e}_1}{\mathsf{det}(z\mathbf{I} - \mathbf{T})}\,. \tag{6.6}$$

The formula (6.5) has an important interpretation in systems and control theory: if f is the transfer function of the linear time-invariant SISO system $(\mathbf{A}, \mathbf{x}_0, \mathbf{y}_0)$ as in (2.2), and if, by cancelation of common factors, f reduces to a fraction h with exact denominator degree n, then $(\mathbf{T}, s_0\mathbf{e}_1, \mathbf{e}_1)$ is a tridiagonal *minimal realization* of this system. Moreover, if we replace in (6.5) $\mathbf{T}$ by its kth leading principal submatrix, we obtain a *minimal partial realization of degree* k of the given system.

Finally, when discussing the Lanczos recursions, we are now used to introduce

$$\begin{aligned}
\mathbf{p}_n^{\mathsf{T}}(z) &:= \begin{pmatrix} p_0(z) & p_1(z) & \ldots & p_{n-1}(z) \end{pmatrix},\\
\mathbf{X}_n &:= \begin{pmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \ldots & \mathbf{x}_{n-1} \end{pmatrix},\\
\mathbf{Y}_n &:= \begin{pmatrix} \mathbf{y}_0 & \mathbf{y}_1 & \ldots & \mathbf{y}_{n-1} \end{pmatrix},
\end{aligned}$$

so that we can express the polynomial recursion (6.3) for $k = 0, \ldots, n-1$ as

$$z\,\mathbf{p}_n^{\mathsf{T}}(z) = \mathbf{p}_n^{\mathsf{T}}(z)\mathbf{T} + p_n(z)\,, \tag{6.7}$$

and the corresponding Lanczos recursions (6.4) for real data as

$$\mathbf{A}\mathbf{X}_n(z) = \mathbf{X}_n(z)\mathbf{T} + \mathbf{x}_n(z)\,, \qquad \mathbf{A}^{\mathsf{T}}\mathbf{Y}_n(z) = \mathbf{Y}_n(z)\mathbf{T} + \mathbf{y}_n(z)\,. \tag{6.8}$$

Interestingly, this very useful notation was not widely used (if at all) in the 1950's, neither in the literature on orthogonal polynomials nor in the one on what we now call Krylov subspace methods. Note, e.g., that in the case where the Lanczos algorithm terminates with $\mathbf{x}_n = 0$ it manifests that the columns of $\mathbf{X}_n$ are the basis of an invariant subspace. This notation is also very useful in describing the connections between the Lanczos method, the conjugate gradient method, and the conjugate residual method; see [11, 26]. For the more general case of the Hessenberg process the notation had be introduced by Hessenberg in 1940; see [83].

7. From qd to LR. In none of Rutishauser's early qd papers [55, 54, 56, 57] is there any hint of the LR algorithm yet. Why should there be?

To adapt the previous section to the situation of the qd algorithm we have to consider the tridiagonal matrices

$$\mathbf{T}_\nu := \begin{pmatrix} q_1^{(\nu)} & 1 & & & \\ e_1^{(\nu)}q_1^{(\nu)} & e_1^{(\nu)} + q_2^{(\nu)} & 1 & & \\ & e_2^{(\nu)}q_2^{(\nu)} & e_2^{(\nu)} + q_3^{(\nu)} & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & e_{n-1}^{(\nu)}q_{n-1}^{(\nu)} & e_{n-1}^{(\nu)} + q_n^{(\nu)} \end{pmatrix}, \tag{7.1}$$

which are associated with the nth *convergent* (5.6) of the J–fraction (5.2) of f_ν. We assume in this section that f_ν terminates there, i.e., $e_n^{(\nu)} = 0$. Note that the index does not refer to the size of the matrix, but to the index of the associated function f_ν, as we will keep the size n fixed.

The matrix $\mathbf{T}_\nu$ has the simple LU (in German: LR) decomposition

$$\mathbf{T}_\nu = \mathbf{L}_\nu\mathbf{R}_\nu \tag{7.2}$$

with

$$\mathbf{L}_\nu = \begin{pmatrix} 1 & & & & \\ e_1^{(\nu)} & 1 & & & \\ & e_2^{(\nu)} & \ddots & & \\ & & \ddots & \ddots & \\ & & & e_{n-1}^{(\nu)} & 1 \end{pmatrix}, \qquad \mathbf{R}_\nu = \begin{pmatrix} q_1^{(\nu)} & 1 & & & \\ & q_2^{(\nu)} & 1 & & \\ & & q_3^{(\nu)} & \ddots & \\ & & & \ddots & 1 \\ & & & & q_n^{(\nu)} \end{pmatrix}. \tag{7.3}$$

At some historic moment in 1954 or 1955, Rutishauser must have made the remarkable observation[9] that his rhombus rules (4.4), (4.5) for the qd table could be interpreted as computing this triangular factorization $\mathbf{T}_\nu = \mathbf{L}_\nu \mathbf{R}_\nu$ and then forming a new tridiagonal matrix $\mathbf{T}_{\nu+1}$ by multiplying together the two factors in reverse order. In fact,

$$\mathbf{R}_\nu \mathbf{L}_\nu = \begin{pmatrix} e_1^{(\nu)} + q_1^{(\nu)} & 1 & & & \\ e_1^{(\nu)} q_2^{(\nu)} & e_2^{(\nu)} + q_2^{(\nu)} & 1 & & \\ & e_2^{(\nu)} q_3^{(\nu)} & e_3^{(\nu)} + q_3^{(\nu)} & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & e_{n-1}^{(\nu)} q_n^{(\nu)} & q_n^{(\nu)} \end{pmatrix}. \tag{7.4}$$

In view of $e_n^{(\nu)} = 0$ and the rhombus rules (4.4), (4.5) this product equals exactly $\mathbf{T}_{\nu+1}$ as defined by (7.1) with ν replaced by $\nu + 1$. Thus,

$$\mathbf{R}_\nu \mathbf{L}_\nu = \mathbf{T}_{\nu+1} \qquad (\text{if} \quad e_n^{(\nu)} = 0). \tag{7.5}$$

So, the progressive qd algorithm can be viewed as performing the steps

$$\mathbf{T}_\nu = \mathbf{L}_\nu \mathbf{R}_\nu \quad \rightsquigarrow \quad \mathbf{R}_\nu \mathbf{L}_\nu = \mathbf{T}_{\nu+1}. \tag{7.6}$$

The step from $\mathbf{T}_\nu$ to $\mathbf{T}_{\nu+1}$ is called an *LR transformation*. Since

$$\mathbf{T}_{\nu+1} = \mathbf{L}_\nu^{-1} \mathbf{T}_\nu \mathbf{L}_\nu, \tag{7.7}$$

the tridiagonal matrices are similar, which is no surprise as f_ν and $f_{\nu+1}$, which are related by (5.5), have the same poles. If all the poles have distinct moduli, we know from Hadamard's theorem or from the qd algorithm that $\mathbf{L}_\nu \to \mathbf{I}$ as $\nu \to \infty$, so in the long run $\mathbf{R}_\nu$ will contain the eigenvalues in its diagonal. If there are poles that are multiple or have the same modulus, we can still apply an adaptation of what Henrici called Rutishauser's rule.

That moment in 1954 or 1955 was the birth of the LR algorithm, first only in its form for tridiagonal matrices, which is equivalent to the qd algorithm for finite J–fractions and S–fractions. But it must have taken Rutishauser only a few minutes to see that the tridiagonal form is not necessary and that one could as well start from a full matrix. From this point of view, the LR algorithm is a natural generalization of the qd algorithm. However, for a full matrix the LR algorithm is costly. But it has the most welcome feature to preserve the bandwidth of a banded matrix, in particular of a tridiagonal one. So, Rutishauser ended up with a most elegant and intriguing algorithm, but its most important use was for the tridiagonal case, where it is identical with the progressive qd algorithm. And Rutishauser knew that the reduction to tridiagonal matrices could be achieved with the Lanczos algorithm. The reduction of a symmetric matrix to tridiagonal form by orthogonal Givens rotations [19] was introduced around the same time, but was no alternative for the general case.

[9]Reportedly, Rutishauser later told his students that he found the LR algorithm by chance when testing the computer arithmetic by computing the LR factorization of test matrices and multiplying the factors together. Due to a programming error he was multiplying the factors in the wrong order — and the matrix converged to upper triangular form [W. Gander, private communication]; however, we believe that Rutishauser was joking when telling this story.

Rutishauser realized from the beginning that fast convergence requires shifting the spectrum appropriately since he had analyzed convergence and introduced spectral shifts for the progressive qd algorithm before; see [54, § 7] or [62, Ch. 2, § 8] and [57, § 4] or [62, Ch. 3, § 4]. So, in practice, (7.6) is replaced by

$$\mathbf{T}_\nu - \delta_\nu \mathbf{I} = \mathbf{L}_\nu \mathbf{R}_\nu \quad \rightsquigarrow \quad \mathbf{R}_\nu \mathbf{L}_\nu + \delta_\nu \mathbf{I} = \mathbf{T}_{\nu+1}\,, \tag{7.8}$$

where δ_ν is the shift parameter.

Rutishauser's first publication [58] on the LR algorithm is, in 1955, a two-page note in French in the *Comptes Rendus*, the premium journal for research announcement in mathematics of the time. In the following year, according to [62], he produced an mimeographed 51-page preprint in English, entitled "Report on the Solution of Eigenvalue Problems with the LR–transformation" [60][10], but only two years later this article got properly published at the NBS [63], in the volume that also contained Henrici's review of the qd algorithm [29]. Only then the qd and LR algorithms made their appearance for the English speaking world, except for the few people who had received the LR preprint before. In 1957, Rutihauser included a 5-page appendix on the LR transformation in his substantial report [62], which compiled and updated most of his previous work on qd, but was still in German.

What remains a puzzle is the following. LR is much more general than qd. Despite this, Rutishauser attached much more importance to qd than to LR. For example, in 1968, ten years after the NBS volume, Rutishauser wrote an updated script [68] of some of his research on qd with no mention of LR. One answer may be that tridiagonal matrices are much simpler and much more economical than general ones, and there was hope that by controlling the roundoff, the qd algorithm could be made more accurate than the LR algorithm for general matrices. We will come back to this in the following Section 8.

Two comments are in order to put Rutishauser's work in perspective. Firstly, starting around 1955 his main preoccupation for more than a decade was the development of the Algol60 programming language [3], the description of its usage for basic problems of numerical analysis [67], and the creation of open source high-quality programming libraries [64, 81]. At the same time he published a large number of papers on various topics in numerical analysis. Secondly, although the discovery of LR is considered to be one of the most significant moments in matrix computations, nevertheless LR was quickly eclipsed by Francis' QR algorithm [12, 13] shortly after its introduction. Francis' original papers on QR explicitly reveal the primal importance of LR for the development of QR, which was, correctly, seen as a backward stable variation of LR. Strangely enough, Rutishauser [65] presented at the IFIP Congress 1962 a paper entitled *Numerical experiments with the QD–transformation of J.G.F. Francis*. We do not know whether the "QD" in the title was a misprint or an intention (recall that he usually wrote qd in capitals). In any case, one can hardly deny that Rutishauser somehow disliked the QR algorithm: there was no room for it in his lectures [69, 70].

8. Stability and positivity. The qd and LR algorithms have similar numerical properties: both are backward stable in the positive definite case and both can exhibit instability in general. The latter is the main shortcoming of these algorithms, shared by a number of related constructive methods. As Aitken's algorithm, LR and qd require that all the Hankel determinants H_k^ν ($k = 1, \dots, n$; $\nu = 0, 1, \dots$) are nonzero if we want to proceed up to the nth column of the scheme. Equivalently, all the J–fractions and S–fractions of f_ν ($\nu = 0, 1, \dots$) must exist up to the corresponding term, so that the Padé approximants up to denominator degree n are all different from each other. (In general, a proper rational function can only be expanded into a P–fraction [42, 43].) The same type of regularity plays a role in the Lanczos algorithm, but there only the existence of the top diagonal of the qd table or the distinctness of the approximants in the top diagonal of the Padé table plays role.

[10] According to the annual report 1956 of the Institute for Applied Mathematics this report was sent in that year to people interested in the subject.

Much of the theory and many of the algorithms we mentioned can be generalized to non-generic situations where exactly singular submatrices occur and are treated in exact arithmetic using special "singular" rules; see, e.g., [25, 27, 44] and references listed therein. But such methods are of little use in practice unless they are designed to deal in finite precision arithmetic appropriately with near-singular situations, as, e.g., in [48, 14] and [27, § 9 & 10].

There is the important exception where stability of the qd and LR algorithm can be established, namely when in the definition (2.2) the matrix $\mathbf{A}$ is symmetric positive definite and $\mathbf{y}_0 = \mathbf{x}_0$. Then, for all ν, the continued fraction (5.3) is a true Stieltjes fraction with positive coefficients $q_k^{(\nu)}$ and $e_k^{(\nu)}$, so the use of the product rhombus rule (4.5) will never lead to a breakdown (except one caused by round-off). The associated polynomial are then true orthogonal polynomials for a positive weight function, and the tridiagonal matrix (7.1) could be replaced by a similar, symmetric positive definite one. Incidentally, the conjugate gradient method applied to a symmetric positive definite matrix fits into this case.

In the 1960's Rutishauser realized that in this case of the qd algorithm further analysis and further improvements were possible. He spent a lot of energy analyzing the round-off behavior and developing alternative forms of the algorithm that made it stable even in finite precision arithmetic; see the remarkable paper [66] and the French report [68]. This report served as a draft of a much longer, unfinished book project, written during the last two years of his life, 1968–1970. It starts with an axiomatic treatment of a finite precision arithmetic (a very remarkable work on its own) and then exemplifies this theory with an analysis of the qd algorithm for the positive definite case. To improve the accuracy, special forms of qd for exceptional situations are developed: the *stationary qd algorithm* and the *differential qd (dqd) algorithm*. This work got published posthumously as an appendix to his "Lecture Notes on Numerical Mathematics" [69, Band 2, Anhang], [70, Appendix]. But again, there is no mention of LR in what he wrote up before being interrupted by his premature death. In the 1990's this work was rediscovered and further improved with more sophisticated shift strategies; see [10, 45].

REFERENCES

[1] A. C. Aitken, *On Bernoulli's numerical solution of algebraic equations*, Proc. Roy. Soc. Edinb., 46 (1926), pp. 289–305.

[2] ———, *Further numerical studies in algebraic equations and matrices*, Proc. Roy. Soc. Edinb., 51 (1931), pp. 80–90.

[3] J. W. Backus, F. L. Bauer, J. Green, C. Katz, J. McCarthy, P. N. (Ed.), A. J. Perlis, H. Rutishauser, K. Samelson, B. Vauquois, J. H. Wegstein, A. van Wijngaarden, and M. Woodger, *Revised report on the algorithmic language ALGOL 60*, Numer. Math., 4 (1963), pp. 420–453. See also [4].

[4] ———, *Revised report on the algorithmic language ALGOL 60*, Comm. ACM, 6 (1963), pp. 1–17. See also [5].

[5] ———, *Revised report on the algorithmic language ALGOL 60*, Computer J., 5 (1963), pp. 349–367.

[6] G. A. Baker, Jr. and P. Graves-Morris, *Padé Approximants, Second Edition*, Cambridge University Press, Cambridge, UK, 1996.

[7] D. Bernoulli, Commentarii Acad. Sc. Petropol., III (1732).

[8] P. Chebyshev, *Sur l'interpolation par la méthode des moindres carrés*, Mém. Acad. Impér. Sciences St. Pétersbourg, série VII, 1 (1859), pp. 1–24. [*Oeuvres*, vol. 1, pp. 471–498].

[9] R. de Montessus de Ballore, *Sur les fractions continues algébriques*, Bull. Soc. Math. de France, 30 (1902), pp. 28–36.

[10] K. V. Fernando and B. N. Parlett, *Accurate singular values and differential qd algorithms*, Numer. Math., 67 (1994), pp. 191–229. .

[11] R. Fletcher, *Conjugate gradient methods for indefinite systems*, in Numerical Analysis, Dundee, 1975, G. A. Watson, ed., vol. 506 of Lecture Notes in Mathematics, Springer, Berlin, 1976, pp. 73–89.

[12] J. G. F. Francis, *The QR transformation—a unitary analogue to the LR transformation, Part I*, Comput. J., 4 (1961), pp. 265–271.

[13] ———, *The QR transformation—a unitary analogue to the LR transformation, Part II*, Comput. J., 4 (1962), pp. 332–345.

[14] R. W. Freund, M. H. Gutknecht, and N. M. Nachtigal, *An implementation of the look-ahead Lanczos algorithm for non-Hermitian matrices*, SIAM J. Sci. Comput., 14 (1993), pp. 137–158. .

[15] W. Gautschi, *Construction of Gauss–Christoffel quadrature formulas*, Math. Comp., 22 (1968), pp. 251–270.

[16] ———, *On the construction of Gaussian quadrature rules from modified moments*, Math. Comp., 24 (1970), pp. 245–260.

[17] W. Gautschi, *On generating orthogonal polynomials*, SIAM J. Sci. Stat. Comput., 3 (1982), pp. 289–317.

[18] W. Gautschi, *The interplay between classical analysis and (numerical) linear algebra—a tribute to Gene H. Golub*, Electron. Trans. Numer. Anal., 13 (2002), pp. 119–147 (electronic).

[19] W. Givens, *A method of computing eigenvalues and eigenvectors suggested by classical results on symmetric matrices*, in Simultaneous linear equations and the determination of eigenvalues, National Bureau of Standards Applied Mathematics Series, No. 29, U. S. Government Printing Office, Washington, D. C., 1953, pp. 117–122.

[20] M. Golomb, *Zeros and poles of functions defined by Taylor series*, Bull. Amer. Math. Soc., 49 (1943), pp. 581–592.

[21] G. H. Golub and D. P. O'Leary, *Some history of the conjugate gradient and Lanczos algorithms: 1949–1976*, SIAM Rev., 31 (1989), pp. 50–102.

[22] W. B. Gragg, *The Padé table and its relation to certain algorithms of numerical analysis*, SIAM Rev., 14 (1972), pp. 1–62.

[23] ———, *On Hadamard's theory of polar singularities*, in Padé Approximants and Their Applications, P. Graves-Morris, ed., London, 1973, Academic Press, pp. 117–123.

[24] W. B. Gragg and A. S. Householder, *On a theorem of Koenig*, Numer. Math., 8 (1966), pp. 465–468.

[25] M. H. Gutknecht, *A completed theory of the unsymmetric Lanczos process and related algorithms, Part I*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 594–639. .

[26] ———, *Changing the norm in conjugate gradient type algorithms*, SIAM J. Numer. Anal., 30 (1993), pp. 40–56. .

[27] ———, *A completed theory of the unsymmetric Lanczos process and related algorithms, Part II*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 15–58. .

[28] J. Hadamard, *Essai sur l'étude des fonctions donnée par leurs développement de Taylor*, PhD thesis, Université de Paris, Paris, 1892. Also in J. de Math. Pures Appl., vol. 8, no. 4, pages 101-186 (1992).

[29] P. Henrici, *The quotient-difference algorithm*, in Further Contributions to the Solution of Simultaneous Linear Equations and the Determination of Eigenvalues, vol. 49 of Applied Mathematics Series, National Bureau of Standards, 1958, pp. 23–46.

[30] ———, *Applied and Computational Complex Analysis, Vol. 1*, Wiley, New York, 1974.

[31] ———, *Applied and Computational Complex Analysis, Vol. 2*, Wiley, New York, 1977.

[32] ———, *Topics in computational complex analysis. I: Methods of descent for polynomial equations. II: New developments on the qd algorithm. III: The asymptotic behavior of the best*

approximation to an analytic function in the unit disk. IV: The Lagrange-Bürmann formula for systems of formal power series, in Computational Complex Analysis, H. Werner, L. Wuytack, E. Ng, and H. J. Bünger, eds., D. Reidel Publishing Co., Dordrecht, 1983.

[33] M. R. Hestenes, *Iterative methods for solving linear equations*, NAML Report 52-9, National Bureau of Standards, Los Angeles, CA, July 1951. Reprinted in *J. Optim. Theory Appl.* 11, 323–334 (1973).

[34] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bureau Standards, 49 (1952), pp. 409–435.

[35] A. S. Householder, *The Padé table, the Frobenius identities, and the qd algorithm*, Linear Algebra Appl., 4 (1971), pp. 161–174.

[36] A. S. Housholder, *The Koenig-Hadamard theorem again*, in Studies in Numerical Analysis, B. K. P. Scaife, ed., London, 1974, Academic Press, pp. 235–240.

[37] W. B. Jones and W. J. Thron, *Continued Fractions. Analytic Theory and Applications*, Addison-Wesley, Reading, Mass., 1980.

[38] J. König, *Über eine Eigenschaft der Potenzreihen*, Math. Ann., 23 (1884), pp. 447–449.

[39] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bureau Standards, 45 (1950), pp. 255–281.

[40] ———, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bureau Standards, 45 (1950), pp. 255–281.

[41] ———, *Solution of systems of linear equations by minimized iterations*, J. Res. Nat. Bureau Standards, 49 (1952), pp. 33–53.

[42] A. Magnus, *Certain continued fractions associated with the Padé table*, Math. Z., 78 (1962), pp. 361–374.

[43] ———, *Expansion of power series into P-fractions*, Math. Z., 80 (1962), pp. 209–216.

[44] B. N. Parlett, *Reduction to tridiagonal form and minimal realizations*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 567–593.

[45] B. N. Parlett, *The new qd algorithms*, Acta Numerica, 4 (1995), pp. 459–491.

[46] ———, *What Hadamard missed*, Tech. Report PAM-671, Center for Pure and Applied Mathematics, University of California at Berkeley, March 1996.

[47] ———, *The QR algorithm*, Computing in Science and Engineering, (2000), pp. 38–42.

[48] B. N. Parlett, D. R. Taylor, and Z. A. Liu, *A look-ahead Lanczos algorithm for unsymmetric matrices*, Math. Comp., 44 (1985), pp. 105–124.

[49] H. Rutishauser, *Über Folgen und Scharen von analytischen und meromorphen Funktionen mehrerer Variabeln, sowie von analytischen Abbildungen*, Acta math., 83 (1950), pp. 249–325. Diss. ETH Nr. 1766.

[50] ———, *Automatische Rechenplanfertigung bei programmgesteuerten Rechenmaschinen*, Habilitationsschrift ETH, 4 (1952).

[51] ———, *Über die Instabilität von Methoden zur Integration gewöhnlicher Differentialgleichungen*, Z. Angew. Math. Phys., 3 (1952), pp. 65–74. English Translation: see [59].

[52] ———, *Beiträge zur Kenntnis des Biorthogonalisierungs-Algorithmus von Lanczos*, Z. Angew. Math. Phys., 4 (1953), pp. 35–56.

[53] ———, *Beiträge zur Kenntnis des Biorthogonalisierungs-Algorithmus von Lanczos*, Z. Angew. Math. Phys., 4 (1953), pp. 35–56.

[54] ———, *Anwendungen des Quotienten–Differenzen–Algorithmus*, Z. Angew. Math. Phys., 5 (1954), pp. 496–508.

[55] ———, *Der Quotienten–Differenzen–Algorithmus*, Z. Angew. Math. Phys., 5 (1954), pp. 233–251.

[56] ———, *Ein infinitesimales Analogon zum Quotienten–Differenzen–Algorithmus*, Arch.Math., 5 (1954), pp. 132–137.

[57] ———, *Bestimmung der Eigenwerte und Eigenvektoren einer Matrix mit Hilfe des Quotienten–Differenzen–Algorithmus*, Z. Angew. Math. Phys., 6 (1955), pp. 387–401.

[58] ———, *Une méthode pour la détermination des valeurs propres d'une matrice*, C.R. Acad. Sci. Paris, 240 (1955), pp. 34–36.

[59] ———, *On the instability of methods for the integration of ordinary differential equations*, Tech. Memo 1403, National Advisory Committee for Aeronautics (NACA), 1956. English Translation of [51].

[60] ———, *Report on the solution of eigenvalue problems with the LR–transformation*, mimeographed report, Inst. f. angew. Math. ETH Zürich, 1956.

[61] ———, *Der Quotienten-Differenzen-Algorithmus*, Mitt. Inst. angew. Math. ETH, Nr. 7, Birkhäuser, Basel, 1957.

[62] ———, *Der Quotienten–Differenzen–Algorithmus*, Mitt. Inst. f. angew. Math. ETH Zürich, Nr. 7 (1957). (Contains revised versions of [55], [54], [57].).

[63] ———, *Solution of eigenvalue problems with the LR–transformation*, in Further Contributions to the Solution of Simultaneous Linear Equations and the Determination of Eigenvalues, vol. 49 of Applied Mathematics Series, National Bureau of Standards, 1958, pp. 47–81.

[64] ———, *ALGOL–manual der ALCOR–gruppe*, Elektronische Rechenanlagen, 5 (1961). Heft 6/1961, Heft 2/1962.

[65] ———, *Numerical experiments with the QD–transformation of J.G.F. Francis*, in Information Processing 1962 (Proc. IFIP Congress, Munich, 1962), Amsterdam, 1963, North-Holland Publ. Co., pp. 202–203.

[66] ———, *Stabile Sonderfälle des Quotienten–Differenzen–Algorithmus*, Num. Math., 5 (1963),

pp. 95–112.
[67] ———, *Description of ALGOL 60 (Handbook for Automatic Computation Vol. Ia)*, Springer-Verlag, Berlin/Heidelberg/New York, 1967.
[68] ———, *Les propriétés numériques de l'algorithme quotient-différence*, Rapport EUR 4083f, Communauté Européenne de l'Energie Atomique – Euratom, 1968. Rapport des conférences données par le Prof. H. Rutishauser; rapporteur: R. - F. Gloden.
[69] ———, *Vorlesungen über numerische Mathematik (2 Bände)*, Birkhäuser Verlag, Basel/Stuttgart, 1976. Herausgegeben von M. Gutknecht unter Mitwirkung von P. Henrici, P. Läuchli, H.R. Schwarz; Engl. Transl.: [70].
[70] ———, *Lectures on Numerical Mathematics*, Birkhäuser, Boston, 1990.
[71] R. A. Sack and A. F. Donovan, *An algorithm for gaussian quadrature given modified moments*, Numer. Math., 18 (1972), pp. 465–478.
[72] W. Seewald, *Quotienten–Differenzen–Algorithmus: Beweis der Regeln von Rutishauser*, Numer. Math., 40 (1982), pp. 93–98.
[73] G. W. Stewart, *On a companion operator for analytic functions*, Numer. Math., 18 (1971), pp. 26–43.
[74] E. Stiefel, *Über einige Methoden der Relaxationsrechnung*, Z. Angew. Math. Phys., 3 (1952), pp. 1–33.
[75] ———, *Zur Interpolation von tabellierten Funktionen durch Exponentialsummen und zur Berechnung von Eigenwerten aus den Schwarzschen Konstanten*, Z. angew. Math. Mech., 33 (1953), pp. 1–3.
[76] E. L. Stiefel, *Kernel polynomials in linear algebra and their numerical applications*, progress report, National Bureau of Standards, April 1955.
[77] ———, *Kernel polynomials in linear algebra and their numerical applications*, in Further Contributions to the Solution of Simultaneous Linear Equations and the Determination of Eigenvalues, vol. 49 of Applied Mathematics Series, National Bureau of Standards, 1958, pp. 1–22.
[78] H. S. Wall, *Analytic Theory of Continued Fractions*, D. Van Nostrand Company, Inc., New York, N. Y., 1948.
[79] J. C. Wheeler, *Modified moments and gaussian quadratures*, J. Math., 4 (1974), pp. 287–295.
[80] H. S. Wilf, *Mathematics for the Physical Sciences*, Dover Publications Inc., New York, 1978. Reprinting of the 1962 original.
[81] J. Wilkinson and C. Reinsch, *Linear Algebra (Handbook for Automatic Computation)*, vol. 2, Springer-Verlag, Berlin/Heidelberg/New York, 1971.
[82] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.
[83] J.-P. Zemke, *Karl Hessenberg*. `http://www.hessenberg.de/karl1.html`.