New insight on how Rutishauser discovered the qd algorithm

Martin H. Gutknecht

Seminar for Applied Mathematics
ETH Zurich

The Third International Workshop on Matrix Computations
Online, organized by Lanzhou University, PR China
April 16, 2022
Version of April 19, 2022

Numerical Linear Algebra in 1954: $\mathbf{Ax} = \mathbf{b}$

$\mathbf{Ax} = \mathbf{b}$ with full or banded $\mathbf{A}$:

Gauss elimination (in various versions)

$\mathbf{Ax} = \mathbf{b}$ with sparse spd $\mathbf{A}$:

Chebyshev iteration, SOR, Conjugate Gradients

$\mathbf{Ax} = \mathbf{b}$ with sparse nonsym. or sym. indef. $\mathbf{A}$:

various "relaxation methods" with limited applicability:
Jacobi iteration, Gauss-Seidel, Richarson's iteration, SOR, ...

$\mathbf{Ax} = \mathbf{x}\lambda$ with full sym. $\mathbf{A}$:

Jacobi's method (rotations $\rightsquigarrow$ diagonal form)
Givens' method (reduction to sym. tridiagonal + bisection)

$\mathbf{Ax} = \mathbf{x}\lambda$ with full or sparse $\mathbf{A}$:

power method (dominant eigenpair only)
"fractional", inverse iteration (any single eigenpair)

various methods for computing the charact. polynomial, including:
Krylov's method (reduction to companion form, char. pol.)
Lanczos' method (reduction to tridiagonal form, char. pol.)
Arnoldi's method (red. to Hessenberg form, char. pol.)

qd algorithm is being discovered
LR and QR algorithms are still unknown

How did Rutishauser discover the qd algorithm?

In spring 2004 Beresford Parlett approached me and suggested to investigate this question. He had a great interest in the qd algorithm as he had been working on new versions of the qd algorithm for more than a decade: the **differential qd algorithm (dqd)** due to Rutishauser (1918–1970) and the **differential qd algorithm with shifts (dqds)** due to Fernando and Parlett (1992/1994).
It was difficult for Parlett to understand Rutishauser's early qd papers because nearly all are in German.

How did Rutishauser discover the qd algorithm? (cont'd)

In the abstract of our paper Parlett wrote:

> *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 came up with an explanation.*

How did Rutishauser discover the qd algorithm? (cont'd)

However, in the resulting paper

From qd to LR, or, how were the qd and LR algorithms discovered?

(IMA J. Numer. Anal. **31**, 741–754 (2011))

we could not give a definitive answer, but only speculate.

Actually we found several possible explanations and several hints, but no explicit statement.

Since June 2006 I have given a number of talks on that topic, including two in China and one in Japan.

New insight: we know now that some hints were misleading.

Early papers on the qd algorithm

- E. Stiefel (Aug./Sep. 1953, ZAMM; Proc. GAMM Conf.): Zur Interpolation von tabellierten Funktionen durch Exponentialsummen und zur Berechnung von Eigenwerten aus den Schwarzschen Konstanten
- H. Rutishauser (1954a, ZAMP; subm. Aug. 5, 1953):
  Der Quotienten–Differenzen–Algorithmus
- H. Rutishauser (1954b, ZAMP; subm. Sep. 18, 1953):
  Anwendungen des Quotienten–Differenzen–Algorithmus
- H. Rutishauser (1954c, Arch.Math.; subm. Sep. 25, 1953):
  Ein infinitesimales Analogon zum
  Quotienten–Differenzen–Algorithmus
- H. Rutishauser (1955a, ZAMP; subm. Jul. 19, 1954):
  Bestimmung der Eigenwerte und Eigenvektoren einer Matrix mit Hilfe des Quotienten–Differenzen–Algorithmus
- H. Rutishauser (1957a, Mitt. IAM, ETH):
  Der Quotienten–Differenzen–Algorithmus *(the “qd booklet”)*
- P. Henrici (1958, NBS book): The Quotient-Difference Algorithm

Eduard Stiefel's suggestion

Stiefel's suggestion to Rutishauser: Given $\mathbf{A}$, $\mathbf{x}_0$, $\mathbf{y}_0$, use the **Schwarz constants** (= **moments** = **Markov parameters**)

$$\boxed{s_k :\equiv \left\langle \mathbf{y}_0, \mathbf{A}^k \mathbf{x}_0 \right\rangle \qquad (k = 0, 1, 2, \ldots)} \tag{1}$$

to find all eigenvalues of $\mathbf{A}$. We know by now: this was in 1951.

Recall: Daniel Bernoulli (1732), J. König (1884):

$$\frac{s_{\nu+1}}{s_\nu} \longrightarrow \lambda_1 \qquad \text{as} \quad \nu \longrightarrow \infty \qquad \text{if} \quad |\lambda_1| > |\lambda_2| \geq |\lambda_2| \geq \ldots .$$

Note: Later it turned out that for the other eigenvalues, Stiefel's proposal was a bad idea, since the dependence of the EVals from the moments is highly ill-conditioned (Gautschi (1968)).

Moments and their generating function

Given: $N \times N$ matrix $\mathbf{A}$ and $\mathbf{x}_0, \mathbf{y}_0 \in \mathbb{R}^N$, let

$$\boxed{f(z) :\equiv \left\langle \mathbf{y}_0, (z\mathbf{I} - \mathbf{A})^{-1} \mathbf{x}_0 \right\rangle = \langle \mathbf{y}_0, \tfrac{1}{z} (\mathbf{I} - \tfrac{1}{z}\mathbf{A})^{-1} \mathbf{x}_0 \rangle} \qquad (2)$$

f is a rational function of type $(N-1, N)$, so $f(\infty) = 0$.

The poles of f are eigenvalues of A.

f can be expanded into a power series in z^{-1}:

$$\boxed{f(z) = \sum_{k=0}^{\infty} \frac{s_k}{z^{k+1}} = \frac{s_0}{z} + \frac{s_1}{z^2} + \frac{s_2}{z^3} + \dots .} \qquad (3)$$

where

$$\boxed{s_k = \left\langle \mathbf{y}_0, \mathbf{A}^k \mathbf{x}_0 \right\rangle}$$

So, f is the **generating function** of the moments.

Clearly we could also write $zf(z)$ as a function of $\zeta :\equiv z^{-1}$:

$$\boxed{\varphi(\zeta) :\equiv \zeta^{-1} f(\zeta^{-1}) = \left\langle \mathbf{y}_0, (\mathbf{I} - \zeta \mathbf{A})^{-1} \mathbf{x}_0 \right\rangle = s_0 + s_1 \zeta + s_2 \zeta^2 + \dots} \tag{4}$$

φ is also a rational function of type $(N-1, N)$.

Assume the eigenvalues λ_k of **A** are ordered such that

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_{N-1}| \geq |\lambda_N|$$

The series of f converges for $|z| > |\lambda_1|$.

The series of φ converges for $|\zeta| < |\lambda_1|^{-1}$.

Could as well look for zeros of a polynomial $b_0 + \cdots + b_n \zeta^N$:

$$\boxed{\varphi(\zeta) :\equiv \frac{1}{b_0 + b_1 \zeta + \cdots b_n \zeta^N} = s_0 + s_1 \zeta + s_2 \zeta^2 + \dots}$$

Alternative formulations of the problem

Clearly, there are several equivalent problems:

- Find eigenvalues of **A**.
- Find poles of generating (rational) function f.
- Find zeros of the denominator polynomial of f (Bernoulli).

In theory, the problem had been solved before by

- Hadamard (1892) (his PhD thesis!),
- de Montessus de Ballore (1902/1905),
- Aitken (1926/1931).

But none of them had an efficient algorithm.

Rutishauser cites Hadamard and Aitken, but never
de Montessus de Ballore, who proved the convergence of Padé approximants with fixed denominator degree.

Hadamard's theorem (1892)

Given the power series of f in z^{-1} of (3), let $H_0^{(\nu)} :\equiv 1$, and define the *Hankel determinants*

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

THEOREM

[Hadamard (1892)] If $|\lambda_{k+1}| < \Lambda < |\lambda_k|$, then, as $\nu \longrightarrow \infty$,

$$H_k^{(\nu)} = \text{const} \cdot (\lambda_1 \cdots \lambda_k)^\nu \left[1 + \mathcal{O}\left(\frac{\Lambda}{|\lambda_k|}\right)^\nu\right]$$

For a simpler proof see Henrici (1958) or Henrici (1974).

Hadamard's theorem (1892) (cont'd)

COROLLARY

If f has N simple poles, then

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

$$\boxed{\frac{H_k^{(\nu+1)}}{H_k^{(\nu)}} \longrightarrow \lambda_1 \lambda_2 \cdots \lambda_k} \qquad \textit{as} \quad \nu \longrightarrow \infty. \tag{5}$$

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

$$\boxed{q_k^{(\nu)} :\equiv \frac{H_k^{(\nu+1)}}{H_k^{(\nu)}} \cdot \frac{H_{k-1}^{(\nu)}}{H_{k-1}^{(\nu+1)}} \longrightarrow \lambda_k} \qquad \textit{as} \quad \nu \longrightarrow \infty. \tag{6}$$

Computing, for fixed ν, the Hankel determinants $H_1^{(\nu)}, \ldots, H_N^{(\nu)}$ (if nonzero) requires the LU decomposition of the matrix $H_N^{(\nu)}$.

Aitken (1926, 1931) used what is now called **"Jacobi identity"** ("theorem of compound determinants")

$$\boxed{\left(H_k^{(\nu)}\right)^2 = H_k^{(\nu-1)} H_k^{(\nu+1)} + H_{k+1}^{(\nu-1)} H_{k-1}^{(\nu+1)} .} \tag{7}$$

It had also been known to Hadamard, but Aitken used it to build up — from the left or from the top — the table

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

Rutishauser's qd algorithm (QD-Algorithmus)

Rutishauser (1954a) knew Aitken's work and refers to (5),

$$\boxed{\frac{H_k^{(\nu+1)}}{H_k^{(\nu)}} \longrightarrow \lambda_1 \lambda_2 \cdots \lambda_k} \qquad \text{as} \quad \nu \longrightarrow \infty$$

as the key to computing non-dominant poles.

But instead of computing the $H_k^{(\nu)}$–table, he headed directly for recurrences for

$$\boxed{q_k^{(\nu)} :\equiv \frac{H_k^{(\nu+1)}}{H_k^{(\nu)}} \cdot \frac{H_{k-1}^{(\nu)}}{H_{k-1}^{(\nu+1)}}} \quad \text{and} \quad \boxed{e_k^{(\nu)} :\equiv \frac{H_{k+1}^{(\nu)}}{H_k^{(\nu)}} \cdot \frac{H_{k-1}^{(\nu+1)}}{H_k^{(\nu+1)}}} \tag{8}$$

In Rutishauser (1954a) he derives the formulas needed for $q_2^{(\nu)}$, and then states recursions for general k.

Rutishauser's qd algorithm (cont'd)

qd table (QD–Schema):

$$
\begin{array}{ccccccccc}
 & q_1^{(0)} & & & & & & & \\
0 & & e_1^{(0)} \;\times & & & & & & \\
 & q_1^{(1)} \;\times & & q_2^{(0)} & & & & & \\
0 & & e_1^{(1)} & & e_2^{(0)} & & & & \\
 & q_1^{(2)} & & q_2^{(1)} \;+ & & \ddots & & & \\
0 & & e_1^{(2)} \;+ & & e_2^{(1)} & & e_{N-1}^{(0)} & & \\
 & q_1^{(3)} & & q_2^{(2)} & & \ddots & & q_N^{(0)} & \\
0 & & e_1^{(3)} & & e_2^{(2)} & & e_{N-1}^{(1)} & & 0 \\
 & q_1^{(4)} & & q_2^{(3)} & & \ddots & & q_N^{(1)} & \\
0 & & e_1^{(4)} & & e_2^{(3)} & & e_{N-1}^{(2)} & & 0 \\
 & \vdots & & \vdots & & & & \vdots & \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots
\end{array}
$$

$$\boxed{e_1^{(0)} \cdot q_2^{(0)} = q_1^{(1)} \cdot e_1^{(1)}} \qquad \boxed{q_2^{(1)} + e_2^{(1)} = e_1^{(2)} + q_2^{(2)}}$$

Rhombus rules (called so by Stiefel, 1955) of **qd algorithm**:

For building up the table columnwise from left to right:

$$\left.\begin{aligned} e_k^{(\nu)} &:= e_{k-1}^{(\nu+1)} + q_k^{(\nu+1)} - q_k^{(\nu)} \\ q_{k+1}^{(\nu)} &:= q_k^{(\nu+1)} \frac{e_k^{(\nu+1)}}{e_k^{(\nu)}} \end{aligned}\right\} \qquad (k = 1, 2, \dots) \qquad (9)$$

For building up the table row-wise, from top to bottom:

$$\left.\begin{aligned} q_k^{(\nu+1)} &:= q_k^{(\nu)} + e_k^{(\nu)} - e_{k-1}^{(\nu+1)} \\ e_k^{(\nu+1)} &:= e_k^{(\nu)} \frac{q_k^{(\nu+1)}}{q_{k+1}^{(\nu)}} \end{aligned}\right\} \qquad (k = 1, 2, \dots) \qquad (10)$$

Recursions (10) are the basis of the **progressive qd algorithm** (the *relevant* version).

In Rutishauser (1954a) the correctness of the rhombus rules follows later from the connections to continued fractions ~~(probably Stiefel's argument)~~. (Not true due to new insight).

Originally, Rutishauser derived them probably from Hadamard's "Jacobi identity"

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

Henrici (1958), who was in contact with Rutishauser, pointed out that one rhombus rules ($+$) can be derived by combining two applications of this formula, the other ($\times$) just by using the definitions (10) of $q_k^{(\nu)}$ and $e_k^{(\nu)}$.

The details have been worked out in Parlett (1996), a TR entitled "What Hadamard missed".
See also Householder (1970): *The Numerical Treatment of a Single Non-linear Equation*.

$$\frac{\beta_1|}{|\alpha_1} + \frac{\beta_2|}{|\alpha_2} + \frac{\beta_3|}{|\alpha_3} + \ldots :\equiv \frac{\beta_1}{\alpha_1 + \frac{\beta_2}{\alpha_2 + \frac{\beta_3}{\alpha_3 + \ldots}}}. \tag{11}$$

There can be finitely or infinitely many ‘terms’.

In our case, the ‘numerators’ will be real or complex numbers, the ‘denominators’ will be either numbers or linear functions of z:

a formal **Jacobi fraction** or **J–fraction**:

$$f(z) = \frac{s|}{|z - q_1} - \frac{e_1 q_1|}{|z - q_2 - e_1} - \frac{e_2 q_2|}{|z - q_3 - e_2} - \cdots, , \tag{12}$$

a formal **Stieltjes fraction** or **S–fraction**:

$$f(z) = \frac{s|}{|z} - \frac{q_1|}{|1} - \frac{e_1|}{|z} - \frac{q_2|}{|1} - \frac{e_2|}{|z} - \cdots . \tag{13}$$

From power series to a continued fractions

By a standard operation the given power series (3) in z^{-1} of f,

$$\boxed{f(z) = \sum_{k=0}^{\infty} \frac{s_k}{z^{k+1}} = \frac{s_0}{z} + \frac{s_1}{z^2} + \frac{s_2}{z^3} + \dots .}$$

can be turned into a continued fraction (which typically converges in a much larger region). We may also write

$$f(z) = \frac{s_0}{z} + \frac{s_1}{z^2} + \cdots + \frac{s_{\nu-1}}{z^{\nu}} + \frac{f_\nu(z)}{z^{\nu}} . \tag{14}$$

and expand the remainder $f_\nu(z)$ of the power series into a continued fraction.

In each case, two different types of continued fractions can be used. So we get two whole sequences of continued fractions.

It turns out that their coefficients are related by the rhombus rules.

Continued fractions: J–fractions and S–fractions

$$f_\nu(z) :\equiv \sum_{k=0}^{\infty} \frac{s_{\nu+k}}{z^{k+1}} = z^\nu \left(f(z) - \sum_{k=0}^{\nu-1} \frac{s_k}{z^{k+1}} \right) \tag{15}$$

can be expanded both into a **Jacobi fraction** or **J–fraction**

$$f_\nu(z) = \frac{s_\nu \,|}{|\, z - q_1^{(\nu)}} - \frac{e_1^{(\nu)} q_1^{(\nu)} \,|}{|\, z - q_2^{(\nu)} - e_1^{(\nu)}} - \frac{e_2^{(\nu)} q_2^{(\nu)} \,|}{|\, z - q_3^{(\nu)} - e_2^{(\nu)}} - \cdots \tag{16}$$

and into a formal **Stieltjes fraction** or **S–fraction**

$$f_\nu(z) = \frac{s_\nu \,|}{|\, z} - \frac{q_1^{(\nu)} \,|}{|\, 1} - \frac{e_1^{(\nu)} \,|}{|\, z} - \frac{q_2^{(\nu)} \,|}{|\, 1} - \frac{e_2^{(\nu)} \,|}{|\, z} - \cdots . \tag{17}$$

The J–fraction is the so-called *even part* of the S–fraction obtained by merging two successive terms into one.

Continued fractions: J–fractions and S–fractions

The *odd part* of the S–fraction is another formal J–fraction, obtained by merging the two differently chosen successive terms into one,

$$f_\nu(z) = \frac{s_\nu}{z}\left\{1 + \frac{q_1^{(\nu)}}{\left| z - q_1^{(\nu)} - e_1^{(\nu)} \right.} - \frac{e_1^{(\nu)} q_2^{(\nu)}}{\left| z - q_2^{(\nu)} - e_2^{(\nu)} \right.} - \frac{e_2^{(\nu)} q_3^{(\nu)}}{\left| z - q_3^{(\nu)} - e_3^{(\nu)} \right.} - \cdots \right\}. \tag{18}$$

By comparing this J–fraction with the one for

$$f_{\nu+1}(z) = z f_\nu(z) - s_\nu\,, \tag{19}$$

one recovers Rutishauser's *rhombus rules* of the qd algorithm.

This is the nicest derivation of the qd algorithm, but not the original one.

Rutishauser (1954a) indicates that it was suggested to him by Stiefel. This was a misleading hint, however.

Continued fractions, Padé approximations, FOPs

The *"partial sums" = convergents = approximants* of the continued fractions are confluent rational interpolants of f.

They are *Padé approximants* (at ∞) associated with the moments $s_{k+\nu}$ ($k = 0, 1, \ldots; \ \nu$ *fixed*) of the function $f_\nu(z)$ defined by

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

For fixed ν, the denominators of the convergents (= Padé approximants) are (formal) orthogonal polynomials $p_k^{(\nu)}(z)$.

They can be arranged in a table that he called **p–table**. (Analogous to the Padé table.)

p–table (P–Schema):

$$
\begin{array}{cccccc}
1 \equiv p_0^{(0)} & & & & & \\
 & p_1^{(0)} \; q_2^{(0)} & & & & \\
1 \equiv p_0^{(1)} & & z \; p_2^{(0)} & & & \\
 & p_1^{(1)} & & p_3^{(0)} & & \\
1 \equiv p_0^{(2)} & & p_2^{(1)} & & \ddots & \\
 & p_1^{(2)} & & p_3^{(1)} & & p_{N-1}^{(0)} \\
1 \equiv p_0^{(3)} & & p_2^{(2)} & & \ddots & & p_N^{(0)} \\
 & p_1^{(3)} & & p_3^{(2)} & & p_{N-1}^{(1)} \\
1 \equiv p_0^{(4)} & & \vdots & & \ddots & & p_N^{(1)} \\
 & \vdots & & \vdots & & \vdots \\
\vdots & & & & & & \vdots
\end{array}
$$

$$
\boxed{p_2^{(0)}(z) := z\, p_1^{(1)}(z) - q_2^{(0)} p_1^{(0)}(z)}
$$

In the last column, $p_N^{(0)} = p_N^{(1)} = \ldots$ is the minimal polynomial.

Continued fractions, Padé approximations, FOPs (cont'd)

Rutishauser realized that they are also the *Lanczos polynomials* of the (nonsymmetric) Lanczos algorithm (Lanczos, 1950) for **A** started with the pair $(\mathbf{y}_0, \mathbf{A}^\nu \mathbf{x}_0)$.

Rutishauser never mentions Padé approximants, but he had no need, since they are just the convergents of the J–fractions and S–fractions.

For him, actually only the FOPs in the denominator matter.

Later, 1966–74, Householder, Gragg, and Stewart stress the connection to Padé approximants in several papers.

N.B.: Hadamard's theorem (1892) $\sim$ de Montessus de Ballore's theorem (1902/1905).

When introducing the p–table Rutishauser (1954a, Sect. 4) points out that $p_k^{(0)}$ $(k = 0, 1, \dots)$ in the top diagonal appear in Lanczos (1950) in an algorithm for computing the characteristic polynomial from the moments. (This algorithm is basically the staircase recurrence for the Padé denominators.)

Later, in Sect. 8, he proved that these polynomials are equal to the *Lanczos polynomials* constructed implicitly in the BIO algorithm.

Associated polynomials: recurrences

The p–table can be built up from the initial column $p_0^{(\nu)} \equiv 1$ by the left-to-right recurrence

$$\boxed{p_k^{(\nu)}(z) := z p_{k-1}^{(\nu+1)}(z) - q_k^{(\nu)} p_{k-1}^{(\nu)}(z)\,.} \tag{20}$$

Rutishauser (1954a) derived also a top-to-bottom recurrence

$$\boxed{p_k^{(\nu+1)}(z) := p_k^{(\nu)}(z) - e_k^{(\nu)} p_{k-1}^{(\nu+1)}(z)\,.} \tag{21}$$

and the diagonal 3-term recurrence (with $e_0^{(\nu)} :\equiv 0$, $p_0^{(\nu)} :\equiv 1$)

$$\boxed{p_{k+1}^{(\nu)}(z) := \left[z - q_{k+1}^{(\nu)} - e_k^{(\nu)}\right] p_k^{(\nu)}(z) + e_k^{(\nu)} q_k^{(\nu)} p_{k-1}^{(\nu)}(z)} \tag{22}$$

Further relations and applications

So, in addition to introducing and investigating the qd algorithm Rutishauser (1954a) [rec. 5 Aug. 1953] (1954b) [rec. 18 Sep. 1953], (1955a) [rec. 19 July 1954)] explained many connections to other topics and gave many applications; *e.g.*, in (1954a):

- the connection to continued fractions,
- the connection to the Lanczos BIO algorithm,
- the connection to the CG algorithm,
- computing partial fraction decompositions of rational fcts.

In (1954b):

- summation of badly converging series,
- solving algebraic equations = computing zeros of polynomials,
- quadratic convergence by using *shifts / double shifts*.

In (1955a):

- computing EVals by *combining Lanczos'* BIO *alg. and the progressive qd algorithm*,
- computing EVecs (several new algorithms are suggested),
- EVals and EVecs of infinite matrices.

Still missing:

- *tridiagonal matrices* (except for computing shifts),
- *LU decomposition* of these tridiagonal matrices,
- *LR algorithm*.

Rutishauser knew well (see Rutishauser (1953) on the Lanczos BIO algorithm) that associated to the 3-term recurrence (22),

$$\boxed{p_{k+1}^{(\nu)}(z) := \left[z - q_{k+1}^{(\nu)} - e_k^{(\nu)}\right] p_k^{(\nu)}(z) + e_k^{(\nu)} q_k^{(\nu)} p_{k-1}^{(\nu)}(z)}$$

(with fixed ν) there is a nested set of tridiagonal matrices

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

such that $p_n^{(\nu)}(z)$ is the characteristic polynomial of $\mathbf{T}_n^{(\nu)}$.

Since he was interested in the limit of the zeros of $p_n^{(\nu)}$ as $\nu \longrightarrow \infty$ it was natural to look at $\mathbf{T}_n^{(\nu)}$.

LU (LR) decomposition of a tridiagonal matrix

Clearly, $\mathbf{T}_n^{(\nu)}$ has the *LU decomposition (LR-Zerlegung)*

$$\boxed{\mathbf{T}_n^{(\nu)} = \mathbf{L}_n^{(\nu)} \mathbf{R}_n^{(\nu)}} \qquad (23)$$

with

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

The LU (LR) transformation

At some historic moment in 1954, Rutishauser must have realized that his progressive qd algorithm (10) can be interpreted as computing this LU factorization $\mathbf{T}_n^{(\nu)} = \mathbf{L}_n^{(\nu)}\mathbf{R}_n^{(\nu)}$ and then forming

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

$$= \mathbf{T}_n^{(\nu+1)}$$

So, the qd algorithm consists of performing the step

$$\boxed{\mathbf{T}_n^{(\nu)} = \mathbf{L}_n^{(\nu)}\mathbf{R}_n^{(\nu)} \quad \rightsquigarrow \quad \mathbf{R}_n^{(\nu)}\mathbf{L}_n^{(\nu)} = \mathbf{T}_n^{(\nu+1)}}$$

called **LR transformation**, which is a similarity transformation:

$$\mathbf{T}_n^{(\nu+1)} = \mathbf{R}_n^{(\nu)}\mathbf{T}_n^{(\nu)}\left(\mathbf{R}_n^{(\nu)}\right)^{-1} .$$

The likely motivation:

- Diagonals (“rows”) of qd–table correspond to J–fractions, FOPS (Lanczos polynomials), and tridiagonal matrices.
- Rhombus rules lead us from one diagonal to the next.
- They are matched by construction with J– and S–fractions.
- There are corresponding rules for the polynomials.
- **Hence, there must be a rule for transforming one tridiagonal matrix into the next one.**

The LR algorithm

LR algorithm: succession of LR transformations (LR steps).

Convergence of $e_k^{(\nu)} \longrightarrow 0$ $(k = 1, \ldots, n)$ as $\nu \longrightarrow \infty$ means:

Convergence of $\mathbf{L}_n^{(\nu)}$ to diagonal matrix as $\nu \longrightarrow \infty$,

Convergence of $\mathbf{T}_n^{(\nu)}$ to upper bidiagonal matrix as $\nu \longrightarrow \infty$,

The diagonals of $\mathbf{T}_n^{(\nu)}$ and $\mathbf{R}_n^{(\nu)}$ ultimately contain eigenvalues of $\mathbf{A}$,

Generalization to full matrices is immediate, but unimportant.

The first two publication on the LR algorithm were in French, two two-page notes in the *Comptes Rendus*: Rutishauser (1955e) [séance du 3 janvier 1955], Rutishauser/Bauer (1955) [séance du 25 avril 1955].

1956 Rutishauser produced a mimeographed 51-page ETH research report in English, entitled “Report on the Solution of Eigenvalue Problems with the LR–transformation”. Two years later it got properly published in a National Bureau of Standards (NBS) book series (Rutishauser, 1958a).

In the same issue: Henrici’s review paper on the qd algorithm, and Stiefel’s paper on kernel polynomials in NLA.

In 1957, Rutishauser included a 5-page appendix on the LR transformation in the (German) booklet that compiled and updated most of his previous work on qd (Rutishauser, 1957a).

The qd and LR algorithms: later publications

Rutishauser kept on publishing articles on the qd and LR algorithms and their applications.

In particular, he studied the *qd algorithm in finite precision arithmetic* and could prove its stability under certain assumptions.

Before his death he was working on a long manuscript that included the finite precision results and the *differential qd algorithm*, which was later rediscovered by Fernando and Parlett (1994). The finished parts were published as an appendix of the posthumous book on his lectures (1976/1990).

The qd and LR algorithms: pros and cons

Unless **A** is spd or Hpd the LR algorithm may break down, because an LU decomposition (without pivoting) may not exist.

Using shifts a symmetric **A** can be turned into an spd matrix.

But, the LR algorithm may be unstable for nonsymmetric **A**.

Stability is gained by replacing the LU decomposition by the QR decomposition $\rightsquigarrow$ **QR algorithm**.

LR conserves the symmetry and the band structure of **A**; *e.g.,* tridiagonal and Hessenberg matrices.

QR conserves the symmetry and the Hessenberg structure of **A**; *e.g.,* Hermitian tridiagonal and non-Hermitian Hessenberg matrices.

In symmetric (Hermitian) case: two steps of LR yield the same $\mathbf{R}_n^{(2\nu)}$ as one step of QR.

The QR algorithm: publications

The QR algorithm is due to J.G.F. Francis (1961) [rec. 29 Oct. 1959; resubmitted with Part 2 on 6 June 1961], (1962). Promoted mainly for nonsymmetric case.

Independently, it was also discovered by V.N. Kublanovskaya (1961).

Francis' papers contain a full theory, including the double-shift for approaching complex pairs of EVals.
Kublanovskaya's paper is less complete.

Let us return to our question

How did Rutishauser discover qd?

What do we know? Which hints did we get?

What did Rutishauser write and reveal?

Let us return to our question: conclusions in 2009

The discovery of the qd and the LR algorithms probably evolved in the following steps:

- Generalizing Aitken's work $\rightsquigarrow$ qd table / algorithm.
- Considering the corresponding p–table (generalizing Lanczos' work) and finding the 3-term recurrence for this table.
- Making the connection to continued fractions and Lanczos polynomials (and as well to many other topics).
- Making the connection to tridiagonal matrices.
- Noticing their extremely simple LU decomposition.
- Noticing that

  qd algorithm = LR algorithm for tridiagonal matrices

- Generalizing the LR algorithm to full matrices.

Early papers on the qd algorithm

- E. Stiefel (Aug./Sep. 1953, ZAMM; Proc. GAMM Conf.): Zur Interpolation von tabellierten Funktionen durch Exponentialsummen und zur Berechnung von Eigenwerten aus den Schwarzschen Konstanten
- H. Rutishauser (1954a, ZAMP; subm. Aug. 5, 1953): Der Quotienten–Differenzen–Algorithmus
- H. Rutishauser (1954b, ZAMP; subm. Sep. 18, 1953): Anwendungen des Quotienten–Differenzen–Algorithmus
- H. Rutishauser (1954c, Arch.Math.; subm. Sep. 25, 1953): Ein infinitesimales Analogon zum Quotienten–Differenzen–Algorithmus
- H. Rutishauser (1955a, ZAMP; subm. Jul. 19, 1954): Bestimmung der Eigenwerte und Eigenvektoren einer Matrix mit Hilfe des Quotienten–Differenzen–Algorithmus
- H. Rutishauser (1957a, Mitt. IAM, ETH): Der Quotienten–Differenzen–Algorithmus *(the “qd booklet”)*
- P. Henrici (1958, NBS book): The Quotient-Difference Algorithm

The intro of the first qd publication (1954a, ZAMP) reads:

Im Anschluss an eine praktische Anwendung des BO-Algorithmus (Biorthogonalisierungs-Algorithmus von C. LANCZOS machte mich Herr Prof. E. STIEFEL, ETH, auf das Problem aufmerksam, die höheren Eigenwerte direkt aus den sogenannten Schwarzschen Konstanten zu bestimmen, das heisst ohne den Umweg über die Orthogonalisierung. Auf diese Anregung hin entwickelte der Verfasser einen Algorithmus, der die gestellte Aufgabe löst.

Subsequently to a practical application of the BO algorithm (BiOrthogonalisation algorithm of C. LANCZOS) Prof. E. STIEFEL, ETH, pointed out to me the problem of determining the higher eigenvalues directly from the so-called Schwarz constants, that is, without taking the detour around orthogonalisation. Following this proposal the author devised an algorithm that solved the posed problem.

Allerdings gab bereits A. C. AITKEN eine Methode an, welche hauptsächlich zur Auflösung algebraischer Gleichungen gedacht war, aber auch die Bestimmung höherer Eigenwerte aus Schwarzschen Konstanten gestattet. Ferner stammt von C. LANCZOS ein Algorithmus[1] zur Bestimmung des charakteristischen Polynoms einer Matrix aus Schwarzschen Konstanten.

However, A. C. AITKEN already proposed a method that was primarily targeted at solving algebraic equations, but also allowed the determination of higher eigenvalues from Schwarz constants. Moreover, C. LANCZOS had proposed an algorithm[2] for the determination of the characteristic polynomial of a matrix from Schwarz constants.

[1] Es handelt sich nicht um den BO-Algorithmus, vgl. vielmehr Kapitel VI bei [4] oder S. 173–179 bei [5].

[2] This does not refer to the BO algorithm; instead see chapter VI of [4] or pp. 173–179 of [5].

Überdies entwickelte J. HADAMARD in seiner Dissertation [2] eine Methode zur Bestimmung der Pole einer durch ihre Potenztreihe gegebenen Funktion. [...] Wenn hier das schon gelöste Problem nochmals aufgegriffen wird, so geschieht dies deshalb, weil der entwickelte Algorithmus eine Reihe von weiteren Anwendungen gestattet und insbesondere auch wertvolle Beziehungen zur Kettenbruchtheorie vermittelt[3]

Moreover J. HADAMARDdeveloped in his dissertation [2] a method for the determination of the poles of a function given by its power series. [...] If we pick up here the already solved problem, then this is because the algorithm we developed allows a series of additional applications and, in particular, also conveys valuable relations to the theory of continued fractions[4]

[3] Herrn Prof. STIEFEL verdanke ich [...]

[4] I owe to Prof. STIEFEL [...]

What do we know? What can we suspect? (cont'd)

So, on the first page of the first qd publication (1954a, ZAMP) as well as in the qd booklet (Mitt. IAM ETH) there are footnotes, one of which says:

Herrn Prof. Stiefel verdanke ich auch die Anregung zur Vereinfachung einiger Beweise mit Hilfe der Kettenbruchtheorie.

I owe Prof. Stiefel the suggestion to simplify some proofs with the help of the theory of continued fractions.

But: which proofs?

How did Rutishauser find qd? The answer!

The ETH library holds large collections of manuscript, notes, and correspondence of former ETH professors. In the last few years Hanna Rutishauser, the older daughter of Heinz, has screened the collection of her father's documents. Three years ago she came across a 2-page document that her father had sent to J.F. Traub. It was entitled *Report on a paper by Gragg*, and it was immediately clear to me that this was a referee report for W.B. Gragg's famous 1972 SIAM Review article

The Padé table and its relation to certain algorithms of numerical analysis

(SIAM Review **14**, pp. 1–62 (1972)). Traub had sent the manuscript to Rutishauser on May 20, 1970, the latter had asked for more time on Aug. 25, and on Sep. 17, 1970, he had sent his 2-page referee report to Traub. That is less than two months before Rutishauser's death on Nov. 10, 1970, at age 52.

Here is *Rutishauser's referee report* (retyped) [Link to pdf]
The answer is found in item *Forth*.

Report on a paper by Gragg[5]

First: I consider this a very useful survey on the Padé table and related topics, although to my taste certain notations make the reading uneasy. I acknowledge the great care the author has taken in establishing the connections to the true sources.

[5] A few misprints have been corrected.

Second: I agree that some of the authors (especially P. Wynn and myself) have been cited fairly often. P. Wynn has indeed written a great number of papers on continued fractions and related topics, in fact many more than are cited by Gragg. I have checked all entries of Wynn in the reference list, except [120], and found that just those of Wynn's papers are cited which are somehow relevant for Gragg's article. In some cases there is not much connection, however, but may it be accidental or a clever strategy—hardly could two entries be replaced by one. Thus, if anything, I would doubt the need for citing [101] since there P. Wynn proves a fact which in a certain sense is trivial.

Of my own papers [71], [72], [74] are practically contained in [77]. On the other hand many readers have access to the ZAMP in which [71], [72], [74] appeared, but [77] is a booklet which despite its low price (2 $) may not be easily accessible. If this latter is not a point of consideration, then [71], [72], [74] might be omitted. In addition [79] seems not relevant for the purpose of Mr. Gragg's article. It may be cited because of the interpretation of continued fractions as $(1, 1)$–elements of matrix inverses, but this seems to be Whittacker's idea, while the main theorem mentioned in [79] was already proved by Wintner (although not under the heading "generalized continued fractions"). If W. Gragg had really the intension of including a reference to generalized continued fractions, then Shenton (see Proc. Edinburgh Math. Soc., 1953, If W. Gragg had really the intension of including a reference to generalized continued fractions, then Shenton (see Proc. Edinburgh Math. Soc., 1953, 1954 and later) might be mentioned.

Third: I enclose a paper of Bandemer, which may have connections with Gragg's article, and which the author might have overlooked. However, this is not a proposal to include the reference but just to inform the author. In the same spirit a copy of a paper of Bauer and Frank is enclosed.

Fourth: A final remark concerns pages 3, 4. It is not true, that Prof. Stiefel suggested the use of continued fractions, but he actually initiated that I began to work on the subject. The true history, which W. Gragg of course could not know, was as follows: One day in 1951, Prof. Stiefel asked me whether I could obtain nondominant eigenvalues from Schwarz constants in a way similar to the computation of the dominant eigenvalue as the limit of s_{k+1}/s_k. On the spot I gave the answer that this problem certainly had something to do with the changes of the quotients s_{k+1}/s_k, but it took me much longer to find the algorithm which I first established by aid of determinantal identities. Only some months later I found the connections with continued fractions.

In view of these facts I would suggest the connections of pages 3, 4 as given on the accompanying sheets.

H. Rutishauser