A look-ahead modified Chebyshev algorithm

Martin H. Gutknecht

Seminar for Applied Mathematics
ETH Zurich

GAMM-ANLA Workshop
Wuppertal, 10 September 2013
Given $A \in \mathbb{R}^{N \times N}$ and $y_0, \tilde{y}_0 \in \mathbb{R}^N$, the moments = Markov parameters = Schwarz constants of $A$ are

$$\mu_k := \tilde{y}_0^T A^k y_0 \quad (k = 0, 1, 2, \ldots). \quad (1)$$

Their generating function (in terms of $z^{-1}$)

$$f(z) = \sum_{k=0}^{\infty} \frac{\mu_k}{z^{k+1}} = \frac{\mu_0}{z} + \frac{\mu_1}{z^2} + \frac{\mu_2}{z^3} + \ldots \quad (2)$$

can be written as the transfer function of a time-invariant linear single-input-single-output (SISO) system:

$$f(z) \equiv \langle \tilde{y}_0, (zI - A)^{-1} y_0 \rangle = \langle \tilde{y}_0, \frac{1}{z} (I - \frac{1}{z} A)^{-1} y_0 \rangle. \quad (3)$$

$f$ is a rational function of type at most $(N - 1, N)$ whose poles are eigenvalues of $A$. (Multiplicity is mostly lost!)
So, the eigenvalues of $A$ are determined by the moments. E.g., according to Daniel Bernoulli (1732), J. König (1884):

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

Around 1953, Prof. Eduard Stiefel of ETH asked his collaborator PD Dr. Heinz Rutishauser to devise a practical algorithm for computing all the eigenvalues of $A$ from the moments.

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.
Moments of matrices; the qd algorithm

In 1954/55 Rutishauser published the qd algorithm. Its memorable tool is the qd table that contains the moments $\mu_k = q_1^{(k)}$ in its first column and can be built up from left to right by the rhombus rules.

After $n$ steps, the information in the top row/diagonal yields

- the first $n$ recurrence coefficients of the (formal) orthogonal polynomials (FOPs) = Lanczos polynomials associated with $A$ and $y_0, \tilde{y}_0$;
- the first $n$ recurrence coefficients for the Padé approximants of $f$;
- the first $n$ terms of the continued fraction representation of $f$;
- the entries of a tridiagonal matrix $T_n^{(0)}$ that is a (Petrov-)Galerkin approximation of $A$. 

Martin H. Gutknecht
A look-ahead modified Chebyshev algorithm
The qd table and its rhombus rules

\[ \nu_0 = q_1^{(0)} \]
\[ \nu_1 = q_1^{(1)} \]
\[ \nu_2 = q_1^{(2)} \]
\[ \nu_3 = q_1^{(3)} \]
\[ \nu_4 = q_1^{(4)} \]

\[ e_1^{(0)} \times e_2^{(0)} \]
\[ e_1^{(1)} \times e_2^{(0)} \]
\[ e_1^{(2)} + e_2^{(1)} \]
\[ e_1^{(3)} + e_2^{(2)} \]
\[ \vdots + \vdots \]
\[ \vdots + \vdots \]

\[ e_1^{(0)} \cdot q_2^{(0)} = q_1^{(1)} \cdot e_1^{(1)} \]
\[ q_2^{(1)} + e_2^{(1)} = e_1^{(2)} + q_2^{(2)} \]
The **progressive qd algorithm** allows us to construct the qd table starting from the top row/diagonal. This is the same as applying the LR algorithm to

\[
T^{(0)}_n = \begin{pmatrix}
q_1^{(0)} & 1 \\
e_1^{(0)} q_1^{(0)} & e_1^{(0)} + q_2^{(0)} & 1 \\
e_2^{(0)} q_2^{(0)} & e_2^{(0)} + q_3^{(0)} & \ddots \\
& \ddots & \ddots & 1 \\
e_{n-1}^{(0)} q_{n-1}^{(0)} & e_{n-1}^{(0)} + q_n^{(0)} & \end{pmatrix}
\]

So, computing eigenvalues with qd is split up in two stages:

- apply the (left-to-right) qd algorithm to the first \(2n + 1\) moments to construct a triangle of the qd table with the coefficients of \(T^{(0)}_n\) in the top row/diagonal;

- apply progressive qd (with shifts) = tridiagonal LR (with shifts) to \(T^{(0)}_n\).
Shortcomings of the qd algorithm:

- the first stage is **severely ill-conditioned** \((\text{Rutishauser}(1955), \text{Gautschi} (1968))\); (left-to-right) qd should be replaced by the **Lanczos algorithm** \((\text{Lanczos} (1950), \text{Rutishauser}(1953,1955))\);
- unless \(A\) is spd, progressive qd may also be unstable; shifts may help, but **tridiagonal QR** \((\text{J.G.F. Francis} (1961/62))\) is preferable;
- if \(A\) is spd, only the **differential qd algorithm (dqd, dqds)** \((\text{Rutishauser} (1974), \text{Fernando and Parlett} (1994))\) can compete with QR.

Unfortunately, Lanczos algorithm has shortcomings too:

- if \(A\) is nonsymmetric, Lanczos algorithm can have “serious” breakdowns; **look-ahead Lanczos** \((\text{Freund/G./Nachtigal} (1993), \text{G.} (1994))\) may help, but is cumbersome;
- even if \(A\) is symmetric, Lanczos algorithm is still prone to strong roundoff effects (**loss of orthogonality**).
- Lanczos algorithm is **not at all communication-avoiding**.
Further alternative: modified Chebyshev algorithm

**Problem left:** Stably compute the recurrence coefficients of a set of (formal) orthogonal polynomials (for some unknown, possibly indefinite measure).

**Recall:** Recurrence coefficients are the elements of the tridiagonal matrix $T_n$ resulting from Lanczos, on which QR can be applied.

**Earliest proposal:** Chebyshev algorithm (Chebyshev (1859)). Computes $T_n$ from moments. (Unknown to Stiefel/Rutihäuser.) But: problem highly ill-conditioned, hence algorithm useless, even when $A$ spd.


**Early application:** Use this for adapting the parameters of the Chebyshev iteration (or a similar method) for the iterative solution of large linear systems: Golub, Kent (1989).

**Many other applications:** Communic. avoiding (CA) eigenvalue computations, CA-BiCG, CA-BiCGStab, model order reduction (MOR), ...
The need for look-ahead

Like three-term Lanczos, the Chebyshev algorithm may break down if $A$ is not spd (and, hence, there exists a set of $N$ orthogonal polynomials associated with a discrete positive measure).

For the indefinite and the nonsymmetric case there is a complete theory involving FOPs and “nongeneric” recurrences that yield “reliable” or “nongeneric” algorithms.

In “nongeneric” algorithms, exact singularities are treated exactly. Near-singularities have to be treated as exact singularities, hence are approximated by exact singularities.


In contrast, “look-ahead” algorithms can treat both exact and near-singularities exactly.

Recall: given \( A \in \mathbb{R}^{N \times N} \) and \( y_0, \tilde{y}_0 \in \mathbb{R}^N \), the moments \( \mu_n \) of \( A \) are

\[
\mu_n \equiv \tilde{y}_0^T A^n y_0 \quad (n = 0, 1, 2, \ldots).
\]

(Alternative definition via measure \( \omega \) defined on subset of \( \mathbb{C} \).)

A corresponding linear functional \( \varphi \) can be defined on the set \( \mathcal{P} \) of all polynomials by

\[
\varphi(z^n) \equiv \mu_n \quad (n = 0, 1, 2, \ldots)
\]

Assume a full set \( p_0, p_1, \ldots, p_N \) of corresponding FOPs exists, so

\[
\varphi(p_m p_n) = \tilde{y}_0^T p_m(A) p_n(A) y_0 = \begin{cases} 
0 & \text{if } m \neq n, \\
\delta_n & \text{if } m = n < N.
\end{cases}
\]

Equivalently, if \( p_N \equiv (p_0 p_1 \ldots p_N) \),

\[
\varphi(p_N^T p_N) = D_N \equiv \text{diag}(\delta_0, \delta_1, \ldots, \delta_{N-1}, 0).
\]
The FOPs are a nested basis of polynomials of degree at most \( N \) that satisfy a three-term recurrence expressible as

\[
z p_{N-1}(z) = p_N(z) G_{N-1}, \quad \text{where} \quad G_{N-1} \equiv \begin{pmatrix}
\alpha_0 & \beta_1 \\
1 & \alpha_1 & \ddots \\
& 1 & \ddots & \beta_{N-1} \\
& & \ddots & \alpha_{N-1} \\
& & & 1
\end{pmatrix}
\]

(7)

with \( p_N \) the characteristic polynomial of \( A \).

We now freely choose a second, infinite basis of monic polynomials \( t \equiv ( t_0 \ t_1 \ t_2 \ \ldots ) \) with \( \partial t_m = m \) and known recurrences expressible as

\[
z t(z) = t(z) H \quad (\forall n)
\]

(8)

where \( H \) is an infinite unit tridiagonal or Hessenberg matrix.
Further we introduce the $\infty \times (N + 1)$ matrix

$$S_N = \varphi(t^T p_N),$$  \hspace{1cm} (9)

which is unit upper triangular since $\varphi(t p_m) = 0$ for $t \in P_{m-1}$. Its elements are

$$\sigma_{m,n} \equiv \varphi(t_m p_n) = \tilde{y}_0^T t_m(A) p_n(A) y_0 \quad (m \in \mathbb{N}_0, \ n \in \{0, 1, \ldots, N\}),$$  \hspace{1cm} (10)

and its first column contains the modified moments

$$\nu_m \equiv \sigma_{m,0} = \varphi(t_m) = \tilde{y}_0^T t_m(A) y_0.$$  \hspace{1cm} (11)

If $t_m(z) = z^m \ (\forall m)$, modified moments become “ordinary” moments. If $t_m(z) = p_m(z) \ (0 \leq m \leq N)$ and $p_N$ is a factor of $t_m$ for $m \geq N$, then $S_N$ is diagonal.
The modified Chebyshev algorithm: derivation

\[
S_N G_{N-1} = \varphi(t^T p_N) G_{N-1} = \varphi(t^T p_N G_{N-1}) = \varphi(t^T z p_{N-1})
= \varphi(H^T t^T p_{N-1}) = H^T \varphi(t^T p_{N-1}) = H^T S_{N-1}.
\]

yields

\[
S_N G_{N-1} = H^T S_{N-1}. \tag{12}
\]

Both the Chebyshev algorithm and the modified Chebyshev algorithm are schemes for building up \( S_N \) from the left (i.e., from the first column with the moments or the modified moments, respectively) and, at the same time, \( G_{N-1} \) with the recurrence coefficients of the FOPs.

Everything is based on an element-by-element comparison in (12).

\[m = n - 1 \text{ yields } \beta_n = \frac{\sigma_{n,n}}{\sigma_{n-1,n-1}}.\]

\[m = n \text{ yields } \alpha_n = \eta_{n,n} + \frac{\sigma_{n+1,n}}{\sigma_{n,n}} - \frac{\sigma_{n,n-1}}{\sigma_{n-1,n-1}} \quad \text{with } \quad H = (\eta_{m,n}).\]

\[m > n \text{ yields a stencil centered at } \sigma_{m,n}.\]
The modified Chebyshev algorithm: the scheme

Stencil centered at \((m, n)\):

\[ H_{m,n} \text{ if } H \text{ tridiagonal} \]

\[ \beta_n \]

\[ 1 \]

\[ \alpha_n \]

modified Chebyshev algorithm
Chebyshev algorithm [Chebyshev (1859)]
\[ \{\mu_n\}_{0}^{2m+1} \leadsto \{\alpha_n\}_{0}^{m}, \quad \{\beta_n\}_{1}^{m}, \quad \{\sigma_{m,n}\}_{0 \leq n \leq m; n \leq m \leq 2m+1-n} \]

Modified Chebyshev algorithm (spd and generic indefinite/nonsym. case; covers exact breakdowns) [Sack/Donovan '72, Wheeler '74]
\[ \{\nu_n\}_{0}^{2m+1} \leadsto \{\alpha_n\}_{0}^{m}, \quad \{\beta_n\}_{1}^{m}, \quad \{\sigma_{m,n}\}_{0 \leq n \leq m; n \leq m \leq 2m+1-n} \]

*Breaks down if* \( \sigma_{n,n} = 0 \) *for some* \( n \).

Nongeneric modified Chebyshev algorithm (nongeneric indefinite/nonsym. case) [Golub/G. ’90]

*Inaccurate if* \( |\sigma_{n,n}| \) *very small for some* \( n \).

Look-ahead modified Chebyshev algorithm (indefinite/nonsym. case; covers also near-breakdowns)

*In both the nongeneric and look-ahead cases the matrix* \( \mathbf{G}_{N-1} \) *is Hessenberg and block-triangular with structured blocks.*

\( \mathbf{S}_N \) *can be split up into compatible blocks too, but we still have to determine them element-by-element.*
The nongeneric modified Chebyshev algorithm

Non-generic modified Chebyshev algorithm

$\nu_0, \nu_1, \nu_2, \nu_3, \ldots$

$\gamma_0 \rightarrow a_0$

$\gamma_1 \rightarrow a_1, \beta_1$

$\gamma_2 \rightarrow a_2, \beta_2$

$2m+1 \rightarrow m$
The stencils centered at \((m, n)\) are still based on an element-by-element comparison in (12),

\[ S_N G_{N-1} = H^T S_{N-1}. \]
The **Gragg matrix** $G = G_\infty$ (Gragg ’74, Draux ’83, G. ’92):

$$G \equiv \begin{pmatrix} A_0 & B_1 \\ C_0 & A_1 & \ddots \\ & \ddots & \ddots & B_J \\ & & \ddots & A_J \end{pmatrix} \quad \text{with} \quad A_i \equiv \begin{pmatrix} 0 & \ddots & 0 \\ 1 & \ddots & \ddots \\ & \ddots & \ddots & 0 \\ & & 1 \\ a_i \end{pmatrix},$$

$$a_i \equiv \begin{pmatrix} \alpha_{1,i} \\ \alpha_{2,i} \\ \vdots \\ \alpha_{h-1,i} \\ \alpha_{hi,i} \end{pmatrix}, \quad B_i \equiv \begin{pmatrix} 0 & \cdots & 0 & \beta_i \\ \vdots & \ddots & \vdots & \ddots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 0 \end{pmatrix}, \quad C_i \equiv \begin{pmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 0 \end{pmatrix},$$

If $J < \infty$ (which is true under our assumptions), $A_J$ is the infinite forward shift matrix and $B_J = C_{J-1}^T$ is the $h_{J-1} \times \infty$ zero matrix.

For $j < J$, $A_j$ is a $h_j \times h_j$ companion matrix. Actually, we could choose for it all but the last column from an irreducible Hessenberg matrix.
The lower bidiagonal and diagonal blocks of \( S_N = (\sigma_{m,n}) \) have the form

\[
S_{j,j-1} \equiv \begin{pmatrix} s' | s_j' \end{pmatrix} = \begin{pmatrix} * & \cdots & * & \sigma'_{j-1} \\ * & \cdots & \cdots & * \\ \vdots & \vdots & \vdots & \vdots \\ * & \cdots & \cdots & * \end{pmatrix}, \quad S_{j,j} \equiv \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \delta_j & \cdots & * \end{pmatrix} \begin{pmatrix} \delta_j \\ \vdots \\ \vdots \\ \delta_j \end{pmatrix}
\]

(the first component of \( s_j \) is \( \delta_j \) too) and yield ultimately

\[
\beta_j := \frac{\delta_j}{\delta_{j-1}}, \quad a_j := S_{j,j}^{-1} \left[ H_{j,j}^T s_j + e_h \sigma'_{j} - s'_j \beta_j \right]
\]

The size of \( S_{j,j} \) is determined by detecting \( \delta_j \neq 0 \).

\( S_N \) can be built up column by column or anti-diagonal by anti-diagonal.
In look-ahead Lanczos the matrix \( \mathbf{G} \) is also a block triangular Hessenberg matrix (like the Gragg matrix):

- the blocks \( C_j \) can be chosen as before;
- the blocks \( A_j \) are again irreducible Hessenberg matrices whose columns can be chosen freely except for the last one; the last column, \( a_j \) is given by a formula similar to (14);
- the blocks \( B_j \) are of rank 1:

\[
B_j = S_{j-1,j-1}^{-1} C_{j-1}^T S_{j,j} = w_{j-1} \hat{s}_j^T,
\]

where \( w_{j-1} \) is the last column of \( S_{j-1,j-1}^{-1} \) and \( \hat{s}_j^T \) is the first row of \( S_{j,j} \).

In the look-ahead modified Chebyshev algorithm \( \mathbf{S}_N \) can again be built up column by column or anti-diagonal by anti-diagonal.

The size of \( S_{j,j} \) is determined so that \( S_{j,j} \) is far enough from singular.
If we denote the last columns of the blocks $B_j$ and $A_j$ by $b_j$ and $a_j$, respectively, we get

$$b_j := S_{j-1,j-1}^{-1} e_{h_{j-1}} s_1^{(j)} ,$$  \hspace{1cm} (15) $$

where $s_1^{(j)}$ is the upper right corner element of $S_{j,j}$, that is, the first component of the last column $s_j$ of $S_{j,j}$;

$$a_j := S_{j,j}^{-1} \left[ H_{j,j} s_j - S_{j,j-1} b_j + e_{h_j} \sigma_j' \right] ,$$  \hspace{1cm} (16) $$

where $\sigma_j'$ is still the upper right corner element of $S_{j+1,j}$.  

The look-ahead Chebyshev algorithm is an interesting potential replacement for the nonsymmetric Lanczos algorithm (and even more so for the Arnoldi algorithm):

- It allows us to construct the block tridiagonal Hessenberg matrix that would result from applying look-ahead Lanczos.
- In contrast to the latter there is no need to access the matrix in every iteration and to compute inner products in every iteration, so LA-Chebyshev is communication avoiding.
- There are applications in eigenvalue computations, solving linear equations, SISO model order reduction, ...
- Of particular interests are applications to multiple systems with identical or near-by matrices.