The Lanczos algorithms and their relations to formal orthogonal polynomials, Padé approximation, continued fractions, and the qd algorithm

+ biconjugate gradient squared methods
+ fast Hankel solvers

With a brief outlook onto:

A completed theory of the unsym. Lanczos and related algorithms

Martin H. Gutknecht

IPS*, ETH Zurich
ETH-Zentrum
CH-8092 Zurich

* Interdisciplinary Project Center for Supercomputing

AIM OF THIS PAPER

* Exposition of basic theoretical facts about
  — and relations among —
  - Lanczos tridiagonalization: BO alg., "BC alg" = BO(A)
  - Lanczos bidiagonalization: "BOBC alg."
  - LR transf. for tridiag. matrices: qd alg.
  - Application to solving linear systems

    BO ⇝ BIORES < normalized / unnormalized = Lanczos/ORTHORES

    BOBC ⇝ BIOMIN < norm. / (unnorm.) = BCG = Lanczos/ORTHOMIN

    BC ⇝ BIODIR = Lanczos/ORTHODIR
  - Breakdown conditions for these algorithms
  - Formal orthogonal polynomials and Padé approximation
  - Continued fractions
  - (Bi)conjugate squared algorithms:

    BO ⇝ BIORES^2 < normalized / unnormalized

    BOBC ⇝ BIOMIN^2 = CGS

    BC ⇝ BIODIR^2 (2 versions)
  - Fast Hankel solvers

* Emphasis is on theoretical relationships, not numerical properties or implementation.
* Preparation for "A completed theory of the unsymmetric Lanczos process and related algorithms", which treats the nongeneric situations and allows us, e.g., to overcome a "curable serious breakdown" of the BO algorithm: nongeneric BO alg, ..., nongeneric qd, ...

Others involved in nongeneric case: Gragg '74, Draux '83, Parlett/Taylor/Liu '85, Golub/G. '89, Joubert, Parlett, Boley.

1. The Lanczos biorthogonalization algorithm (BO algorithm)

Given: $A \in \mathbb{C}^{N\times N}$, $x_0, y_0 \in \mathbb{C}^N$ s.t. $y_0^H x_0 \neq 0$.

[or: H Hilbert space, $x_0, y_0 \in H$, $A \in L(H)$]

<u>Aim</u>: Generate sequences $\{x_n\}_{n=0}^{\nu}$, $\{y_n\}_{n=0}^{\nu}$ with

$$(1)\quad \begin{cases} x_n \in K_n := \operatorname{span}(x_0, Ax_0, \dots, A^n x_0), \\ y_m \in L_m := \operatorname{span}(y_0, A^H y_0, \dots, (A^H)^m y_0), \end{cases}$$

Krylov spaces

$$(2)\quad y_m^H x_n = \begin{cases} 0 & \text{if } m \neq n \\ \delta_n \neq 0 & \text{if } m = n \end{cases}$$

biorthogonal sequences

and ν as large as possible.

(1) and (2) imply:

$$(3)\quad x_n \in K_n \setminus K_{n-1}, \qquad y_n \in L_n \setminus L_{n-1}$$

$$(4)\quad K_n = \operatorname{span}(x_0, x_1, \dots, x_n), \qquad L_n = \operatorname{span}(y_0, y_1, \dots, y_n)$$

$$(5)\quad x_n \perp L_{n-1}, \qquad y_n \perp K_{n-1}$$

$$(6)\quad \begin{cases} A x_n = \overbrace{x_{n+1}\tau_{n+1,n}}^{\neq 0} + x_n \tau_{n,n} + x_{n-1}\tau_{n-1,n} + \dots + x_0 \tau_{0,n} \\ A^H y_n = \underbrace{y_{n+1}\tau'_{n+1,n}}_{\neq 0} + y_n \tau'_{n,n} + y_{n-1}\tau'_{n-1,n} + \dots + y_0 \tau'_{0,n} \end{cases} \Bigg| \begin{array}{l} y_k^\cdot \\ x_k^\cdot \\ (k \le n) \end{array}$$

$$\Rightarrow \quad \tau_{k,n} = 0, \quad \tau'_{k,n} = 0, \quad k = 0, \dots, n-2$$

$$\tau'_{n,n} = \overline{\tau_{n,n}}$$

$$(7)\quad \tau_{n-1,n}\,\tau_{n,n-1} = \overline{\tau'_{n-1,n}\,\tau'_{n,n-1}}$$

(7) can be satisfied by choosing

(8) $$\tau'_{n-1,n} = \overline{\tau_{n-1,n}}, \quad \tau'_{n,n-1} = \overline{\tau_{n,n-1}}. \qquad T' = \overline{T}$$

Another frequent choice is

(8') $$\tau'_{n-1,n} = \overline{\tau_{n,n-1}}, \quad \tau'_{n,n-1} = \overline{\tau_{n-1,n}}. \qquad T' = T^H$$

Set $\alpha_n := \tau_{n,n}$, $\beta_n := \tau_{n-1,n}$, $\gamma_n := \tau_{n+1,n}$,
then (6) becomes

(9) $$\boxed{\begin{aligned} A x_n &= \overbrace{x_{n+1}\gamma_n}^{\tilde{x}_{n+1}} + x_n\alpha_n + x_{n-1}\beta_n, \\ A^H y_n &= \underbrace{y_{n+1}\overline{\gamma_n}}_{\tilde{y}_{n+1}} + y_n\overline{\alpha_n} + y_{n-1}\overline{\beta_n}, \end{aligned}} \qquad n = 0, 1, \dots, \nu-2,$$

with $\beta_0 := 0$, $x_{-1} := y_{-1} := 0$.

Constructive process for generating $\{x_n\}, \{y_n\}, \{\alpha_n\}, \{\beta_n\}, \{\gamma_n\}$:

ALG. 1

BO algorithm [Lanczos 1950]:

For $n = 0, 1, \dots$ do:

(10)
$$\begin{aligned}
\alpha_n &:= y_n^H A x_n / \delta_n, \\
\beta_n &:= y_{n-1}^H A x_n / \delta_{n-1} = \gamma_{n-1}\delta_n / \delta_{n-1} \quad (\text{if } n > 0), \\
\tilde{x}_{n+1} &:= A x_n - x_n\alpha_n - x_{n-1}\beta_n, \\
\tilde{y}_{n+1} &:= A^H y_n - y_n\overline{\alpha_n} - y_{n-1}\overline{\beta_n}, \\
\tilde{\delta}_{n+1} &:= \tilde{y}_{n+1}^H \tilde{x}_{n+1},
\end{aligned}$$

if $\tilde{\delta}_{n+1} = 0$, set $\nu := n+1$, stop, else choose γ_n, δ_{n+1} such that $\gamma_n^2\,\delta_{n+1} = \tilde{\delta}_{n+1}$,

$x_{n+1} := \tilde{x}_{n+1}/\gamma_n$, $y_{n+1} := \tilde{y}_{n+1}/\overline{\gamma_n}$.

Typical choices for γ_n and δ_{n+1}:

(i) $\delta_{n+1} := 1 \Rightarrow \{x_m\}\{y_m\}$ biorthonormal, $\beta_n = \gamma_{n-1}$.
However, γ_n may be complex even if $A \in \mathbb{R}^{N\times N}$, unless A is Hermitian and $x_0 = y_0$ $[\Rightarrow x_n = y_n \ (\forall n)]$

(ii) $\gamma_n := 1$: Often inappropriate for numerical computations. O.K. for complex theory.

(iii) $\gamma_n := \sqrt{|\tilde{\delta}_{n+1}|} \Rightarrow |\delta_{n+1}| = 1; \Rightarrow \gamma_n, \delta_{n+1}, \alpha_n, \beta_n \in \mathbb{R}$ if $A \in \mathbb{R}^{N\times N}$, $x_0, y_0 \in \mathbb{R}^N$.

(iv) $\gamma_n := -\alpha_n - \beta_n$: for solving linear systems of equations. (BIORES)

Types of terminations / breakdowns:

(i) $\tilde{x}_\nu = 0$ or $\tilde{y}_\nu = 0$:
$\Rightarrow K_{\nu-1}$ or $L_{\nu-1}$, resp., is invariant subspace of A or A^T, resp. $[\Rightarrow \nu \le N]$

(ii) $\tilde{x}_\nu \neq 0$, $\tilde{y}_\nu \neq 0$, but $\tilde{y}_\nu^H \tilde{x}_\nu = 0$:
Serious breakdown. Cannot happen if A is Hermitian and $x_0 = y_0$. Until 1988 the theory was not able to cope with this serious breakdown. Seems to occur very rarely in practice. But may occur approximately and have a devastating effect ($\|x_{n+1}\|$ very large if $\gamma_{n+1} = 1$). There always exists x_0, y_0 so that no serious breakdown occurs [Rutishauser, Householder]. In fact, this is w.r.t. $x_0, y_0 \in \mathbb{C}^N$ the generic case [Joubert '90, Parlett '90].

Matrix interpretation

Let

$$X := (x_0\ x_1\ \cdots\ x_{\nu-1}), \qquad Y := (y_0\ y_1\ \cdots\ y_{\nu-1}) \qquad N \times \nu \text{ matrices}$$

Then, by (2) and (9), resp.,

(11a) $$\boxed{Y^H X = D}$$

(11b) $$\boxed{AX = XT + \underbrace{x_\nu \gamma_{\nu-1} e_\nu^T}_{0 \text{ or rank-one}}, \quad A^H Y = Y\overline{T} + \underbrace{y_\nu \overline{\gamma_{\nu-1}} e_\nu^T}_{0 \text{ or rank-one}}}$$

where

$$T := \begin{pmatrix} \alpha_0 & \beta_1 & & & \\ \gamma_0 & \alpha_1 & \beta_2 & & \\ & \gamma_1 & \ddots & \ddots & \\ & & \ddots & \ddots & \beta_{\nu-1} \\ & & & \gamma_{\nu-2} & \alpha_{\nu-1} \end{pmatrix}$$

$x_\nu = 0 \Rightarrow X$ is invariant subspace of A

T represents A in X w.r.t. basis $\{x_i\}$

$\Rightarrow$ EVALs of T are EVALs of A

$y_\nu = 0$: analogous

$\Rightarrow$ EVAL computations : compute EVALs of T

(numerically very tricky; some eigenvalues tend to appear several times)

[C. Paige, B. Parlett, J. Cullum/R. Willoughby, H. Simon, ...]

N.B.: (11a), (11b) imply

$$\boxed{Y^H A X = DT}$$

Solution of linear system $Az = b$

"Consistency condition": Column sums $= 0$

ALG 2

Choose in BO algorithm $\gamma_n := -\alpha_n - \beta_n$ and $x_0 := b - Az_0$, where z_0 is initial approximation of solution. ($\Rightarrow$ Breakdown if $\gamma_n = 0$!)

Add to the BO algorithm the instruction

$$\boxed{(12) \qquad z_{n+1} := -(x_n + \alpha_n z_n + \beta_n z_{n-1})/\gamma_n}$$

BIORES

Then $\underline{x_{n+1} = b - Az_{n+1}}$ (residual)

Proof: $b - Az_{n+1} = b + (Ax_n + \alpha_n Az_n + \beta_n Az_{n-1})/\gamma_n$

$$\underset{\text{Induction}}{=} \cancel{b} + \{Ax_n + \alpha_n(\cancel{b} - x_n) + \beta_n(\cancel{b} - x_{n-1})\}/\gamma_n$$

$$= x_{n+1} \qquad \square$$

In the case of a Hermitian system the BO algorithm with (12) becomes the <u>3-term recurrence version of the conjugate gradient method</u> (<u>CG algorithm</u>) often called <u>ORTHORES</u> today or, briefly, <u>ORES</u>.

Matrix interpretation of (12): Let $Z := (z_0, z_1, \ldots)$, then

$$(12') \qquad \boxed{X = -ZT}$$

The consistency condition $\gamma_n = -\alpha_n - \beta_n$ of BIORES may cause a breakdown due to $\gamma_n = 0$ at $n < \nu$.

Let $\dot{\nu}$ denote the breakdown index of BIORES, i.e.,

$$\delta_{\dot{\nu}} = 0 \quad \text{or} \quad \gamma_{\dot{\nu}} = 0$$

Clearly, $\dot{\nu} \leq \nu$ (:= breakdown index of BO alg.)

Can one avoid the breakdown due to $\gamma_n = 0$?

ALG. 3

Unnormalized BIORES [G. '90 COPPER]

Same alg. as (normalized) BIORES, except that $\gamma_n \neq 0$ is chosen arbitrarily.

$\Rightarrow$ Breakdown index ν (same as BO alg)

However, what is then the meaning of approximants z_n?

LEMMA Let $\dot{\rho}_0 := 1$ and

$$\dot{\rho}_{n+1} := -(\alpha_n \dot{\rho}_n + \beta_n \dot{\rho}_{n-1}) / \gamma_n \quad , \quad n = 0, 1, \ldots, \nu$$

Then x_n and z_n are in unnormalized BIORES related by

$$x_n = b\dot{\rho}_n - Az_n$$

Proof.

$$(b\dot{\rho}_{n+1} - Az_{n+1})\gamma_n = -b(\alpha_n \dot{\rho}_n + \beta_n \dot{\rho}_{n-1}) + Ax_n + Az_n\alpha_n + Az_{n-1}\beta_n$$

$$= Ax_n - x_n\alpha_n - x_{n-1}\beta_n = x_{n+1}\gamma_n$$

$\square$

2. The Lanczos biconjugate gradient algorithm (BCG algorithm)

Just another way to generate the same sequences $\{x_n\}, \{y_n\}$ as in §1.

Lanczos [1952] called it "the complete algorithm for minimized iterations"

Main application: Solution of linear equations.

Popularized by Fletcher [1976].

ALG. 4

BOBC algorithm ($\approx$ BCG, but not adapted to linear systems)

Given $A \in \mathbb{C}^{N\times N}$, $x_0 \in \mathbb{C}^N$, $y_0 \in \mathbb{C}^N$ s.t. $y_0^H x_0 \neq 0$,

set $u_0 := x_0$, $v_0 := y_0$.

For $n = 0, 1, \ldots$ do:

$\varphi_n := y_n^H A u_n / \delta_n = v_n^H A u_n / \delta_n$,

if $\varphi_n = 0$, set $\dot{\nu} := n$ and stop; otherwise

choose $\gamma_n \neq 0$,

$x_{n+1} := (A u_n - x_n \varphi_n)/\gamma_n$

$y_{n+1} := (A^H v_n - y_n \overline{\varphi_n})/\overline{\gamma_n}$

$\delta_{n+1} := y_{n+1}^H x_{n+1}$, $\psi_{n+1} := \delta_{n+1}\gamma_n / (\delta_n \varphi_n)$

if $\delta_{n+1} = 0$ ($\Rightarrow \psi_{n+1} = 0$), set $\dot{\nu} := n+1$, stop; otherwise

$u_{n+1} := x_{n+1} - u_n \psi_{n+1}$,

$v_{n+1} := y_{n+1} - v_n \overline{\psi_{n+1}}$.

N.B.: Two different possible causes for breakdown

THM. 1: Assume $\delta_n \neq 0$, $\varphi_n \neq 0$ $(n = 0, \dots, \nu-1)$.
Then, for $m, n = 0, \dots, \nu-1$:

$$y_m^H x_n = \begin{cases} 0 & , m \neq n \\ \delta_n \neq 0 & , m = n \end{cases} \qquad \text{biorthogonal}$$

$$v_m^H A u_n = \begin{cases} 0 & , m \neq n \\ \delta_n \varphi_n \neq 0 & , m = n \end{cases} \qquad \text{biconjugate}$$

<u>Matrix interpretation:</u>

For simplicity assume again $(\nu = N)$. Set

X, Y as before, $U := (u_0, \dots, u_{N-1})$, $V := (v_0, \dots, v_{N-1})$.

Then

$$\boxed{\begin{array}{ll} AU = XL \,; & A^H V = Y\overline{L} \,; \\ X = UR \,, & Y = V\overline{R} \,, \end{array}}$$

where

$$L := \begin{pmatrix} \varphi_0 & & & & \\ \gamma_0 & \varphi_1 & & & \\ & \gamma_1 & \varphi_2 & & \\ & & \ddots & \ddots & \\ & & & \gamma_{N-2} & \varphi_{N-1} \end{pmatrix}, \quad R := \begin{pmatrix} 1 & \psi_1 & & & \\ & 1 & \psi_2 & & \\ & & 1 & \ddots & \\ & & & \ddots & \psi_{N-1} \\ & & & & 1 \end{pmatrix}$$

By eliminating X and Y or U and V, respectively:

$$\boxed{\begin{array}{ll} AU = UT_1 \,, & A^H V = V\overline{T_1} \,, \\ \hline AX = XT \,, & A^H Y = Y\overline{T} \,, \end{array}}$$

where

$$T_1 := RL \,, \qquad T = LR.$$

$\Rightarrow$ BOBC method makes implicit use of the LU decomposition of T. (Hence, it may break down earlier than BO.)

Solution of linear system $Az = b$

ALG. 5

Choose in BOBC algorithm $\gamma_n := -\varphi_n$ [consistency cond.] and $x_0 := b - Az_0$, where z_0 is initial approximation. Add to the BOBC algorithm the instruction

$$z_{n+1} := z_n + u_n / \varphi_n \,, \quad n = 0, 1, \ldots$$

BIOMIN (= BCG)

Then $\quad x_{n+1} = b - Az_{n+1}$ (residual)

Proof: $\quad b - Az_{n+1} = b - Az_n - Au_n/\varphi_n \underset{\text{Induction}}{=} x_n - Au_n/\varphi_n = x_{n+1}$ $\square$

In the case of a Hermitian system the BCG alg. becomes the (two-term recurrence version of) the CG algorithm: ORTHOMIN or OMIN.

Relation to LR and qd algorithms

$$T = LR \rightsquigarrow T_1 = RL$$

is one step of the LR algorithm applied to the tridiagonal matrix T (= progressive qd algorithm).

$$\Rightarrow \quad T_1 = RTR^{-1} \,, \quad \text{i.e., } T_1 \text{ is similar to } T$$

Typically, the sequence

$$T \rightsquigarrow T_1 \rightsquigarrow T_2 \rightsquigarrow T_3 \rightsquigarrow \ldots$$

converges to diagonal matrix (containing the eigenvalues of T) [König, Hadamard, de Montessus de Ballore].

BO $\Leftrightarrow$ BCG parameter relations ; qd formulas

$$T = \begin{pmatrix} \alpha_0 & \beta_1 & & \\ \gamma_0 & \alpha_1 & \beta_2 & \\ & \gamma_1 & \alpha_2 & \ddots \\ & & \ddots & \ddots \end{pmatrix}, \qquad T_1 = \begin{pmatrix} \alpha_0' & \beta_1' & & \\ \gamma_0 & \alpha_1' & \beta_2' & \\ & \gamma_1 & \alpha_2' & \ddots \\ & & \ddots & \ddots \end{pmatrix}$$

$$L_1 = \begin{pmatrix} \varphi_0' & & & \\ \gamma_0 & \varphi_1' & & \\ & \gamma_1 & \varphi_2' & \\ & & \ddots & \ddots \end{pmatrix}, \qquad R_1 = \begin{pmatrix} 1 & \psi_1' & & \\ & 1 & \psi_2' & \\ & & 1 & \ddots \\ & & & \ddots \end{pmatrix}$$

$$\boxed{T = LR} \rightsquigarrow \boxed{T_1 = RL} \rightsquigarrow \boxed{T_1 = L_1 R_1}$$

$\Downarrow$

$$\alpha_0 = \varphi_0$$
$$\beta_1 = \varphi_0 \psi_1$$
$$\vdots$$
$$\boxed{\begin{aligned} \alpha_n &= \varphi_n + \gamma_{n-1}\psi_n \\ \beta_{n+1} &= \varphi_n \psi_{n+1} \end{aligned}}$$

$n = 0, 1, \ldots$

$(\psi_0 := 0)$

$\Downarrow$

$$\varphi_0 := \alpha_0$$
$$\psi_1 := \beta_1 / \varphi_0$$
$$\vdots$$
$$\boxed{\begin{aligned} \varphi_n &:= \alpha_n - \gamma_{n-1}\psi_n \\ \psi_{n+1} &:= \beta_{n+1} / \varphi_n \end{aligned}}$$

$n = 0, 1, \ldots$

rhombus rules of qd alg.

$\downarrow$

$$\alpha_0' = \varphi_0 + \gamma_0 \psi_1$$
$$\beta_1' = \varphi_1 \psi_1$$
$$\vdots$$
$$\boxed{\begin{aligned} \alpha_n' &= \varphi_n + \gamma_n \psi_{n+1} \\ \beta_{n+1}' &= \varphi_{n+1} \psi_{n+1} \end{aligned}}$$

$n = 0, 1, \ldots$

$\Downarrow$

$$\alpha_0' = \varphi_0'$$
$$\beta_1' = \varphi_0' \psi_1'$$
$$\vdots$$
$$\boxed{\begin{aligned} \alpha_n' &= \varphi_n' + \gamma_{n-1}\psi_n' \\ \beta_{n+1}' &= \varphi_n' \psi_{n+1}' \end{aligned}}$$

$n = 0, 1, \ldots$

$(\psi_0' := 0)$

$\Downarrow \quad \Downarrow$

$$\varphi_0 + \gamma_0 \psi_1 = \alpha_0' = \varphi_0'$$
$$\varphi_1 \psi_1 = \beta_1' = \varphi_0' \psi_1'$$
$$\vdots$$
$$\boxed{\begin{aligned} \varphi_n + \gamma_n \psi_{n+1} &= \alpha_n' = \varphi_n' + \gamma_{n-1}\psi_n' \\ \varphi_{n+1} \psi_{n+1} &= \beta_{n+1}' = \varphi_n' \psi_{n+1}' \end{aligned}} \begin{aligned} &\Rightarrow \varphi_n' \\ &\Rightarrow \psi_{n+1}' \end{aligned}$$

$n = 0, 1, \ldots$

(usually $\gamma_n := 1$ (th) in qd alg.)

BIOMIN has the same breakdown index $\dot{\nu}$ as the BOBC algorithm and, as can be shown, normalized BIORES, while for unnormalized BIORES and the BO algorithm the breakdown is at $\nu \geq \dot{\nu}$.

Does it help to define an unnormalized version of BIOMIN?

ALG. 6

Unnormalized BIOMIN

Choose in BOBC algorithm $\gamma_n \neq 0$ arbitrarily, and set $x_0 := b - Az_0$, $\dot{\rho}_0 := 1$,

$$z_{n+1} := -z_n \varphi_n / \gamma_n - u_n / \gamma_n$$

$$\dot{\rho}_{n+1} := -\dot{\rho}_n \varphi_n / \gamma_n$$

Then, again, $x_n = b\dot{\rho}_n - Az_n$, $n = 0, 1, \ldots, \dot{\nu}$.
But the BOBC algorithm, and hence also unnormalized BIOMIN still break down at $\dot{\nu}$.

However, there is another possibility to reduce the chance of breakdown.

Recall:

$$v_m^H A u_n = \begin{cases} 0 & \text{if } m \neq n \\ \delta_n \varphi_n & \text{if } m = n \end{cases} \Rightarrow \begin{array}{l} \{v_m\}\{u_n\} \text{ are biconjugate} \\ \Updownarrow \\ \{v_m\}\{u_n\} \text{ are } A\text{-biorthogonal} \end{array}$$

Hence, the two sequences can be generated by the BO algorithm using the "A-norm".
($\Rightarrow$ new parameters α_n', β_n', γ_n')

Need additional formulas for x_{n+1} and z_{n+1}.

ALG. 7

"BC algorithm" (= BO algorithm with "A-norm")

Given $u_0', v_0' \in \mathbb{C}^N$ with $v_n'^H A u_n' \neq 0$, let $\beta_0' := 0$ and, for $n = 0, 1, 2, \ldots$

$$\alpha_n' := (A^H v_n')^H A u_n' / \delta_n'$$

$$\beta_n' := v_n'^H A u_n \cdot \gamma_{n-1}' / \delta_{n-1}' = \gamma_{n-1}' \delta_n' / \delta_{n-1}' \quad (\text{if } n > 0)$$

$$\tilde{u}_{n+1} := A u_n' - u_n' \alpha_n' - u_{n-1}' \beta_n'$$

$$\tilde{v}_{n+1} := A^H v_n' - v_n' \overline{\alpha_n'} - v_{n-1}' \overline{\beta_n'}$$

$$\tilde{\delta}_{n+1}' := \tilde{v}_{n+1}^H A \tilde{u}_{n+1}$$

If $\tilde{\delta}_{n+1}' = 0$, set $\nu' := n+1$ and stop; otherwise, choose γ_n' and δ_{n+1}' such that $(\gamma_n')^2 \delta_{n+1}' = \tilde{\delta}_{n+1}'$ and set

$$u_{n+1}' := \tilde{u}_{n+1} / \gamma_n', \qquad v_{n+1}' := \tilde{v}_{n+1} / \gamma_n'.$$

ALG. 8

BIODIR (="BC alg. for solving linear systems")

Choose initial approx. z_0 and set $u_0' := x_0 := b - A z_0$. Apply BC algorithm, computing additionally

$$\omega_n' := v_n'^H x_n / \delta_n'$$

$$z_{n+1} := z_n + u_n' \omega_n'$$

$$x_{n+1} := x_n - A u_n' \omega_n'$$

Breakdown only when $\delta_{n+1}' = 0$, i.e., when BC alg. breaks down. Breakdown index ν'

$\omega_n' = 0$ is no reason for breakdown, although then $z_{n+1} = z_n$ and $x_{n+1} = x_n$. But in general u_{n+1} will be nonzero and will help span a larger Krylov space, so that $z_{n+2} \neq z_n$, $x_{n+2} \neq x_n$.

The following result on the breakdown conditions of the five methods for solving a linear system can be proven:

THM. Assume that the same initial approximation z_0 and the same initial vectors $u_0 := u_0' := x_0 := b - Az_0$ and $v_0 := v_0' := y_0$ are chosen in each algorithm.

Then $\underline{\dot{\nu} = \min\{\nu, \nu'\}}$

The following four conditions are equivalent:

(i') $w_m' \neq 0$ $(m<n)$, $\underline{v_n'^H x_n = 0}$ in BIODIR $\hookrightarrow w_n' = 0$ if $\delta_n' \neq 0$

(ii') $\delta_n = 0$ in BIOMIN

(iii') $\delta_n = 0$ in BIORES and $\dot{\nu} = \nu$

(iv') $n = \dot{\nu} = \nu$

Likewise, the following four conditions are equivalent:

(i'') $\delta_n' = 0$ in BIODIR and $\nu' = \nu$

(ii'') $\varphi_n = 0$ in BIOMIN

(iii'') $\alpha_n + \beta_n = 0$ in BIORES

(iv'') $n = \dot{\nu} = \nu'$

Except (i') for BIODIR and (iii'') for unnormalized BIORES, all these conditions are breakdown conditions for the respective algorithms.

3. The relation to formal orthogonal polynomials and Padé approximation

From definitions of BO and BOBC algorithms one sees immediately that

$$(1)\qquad \boxed{\begin{aligned} x_n &= \rho_n(A)\,x_0\,, & y_n &= \overline{\rho_n}(A^H)\,y_0\,,\\ u_n &= \sigma_n(A)\,u_0\,, & v_n &= \overline{\sigma_n}(A^H)\,v_0\,. \end{aligned}}$$

where ρ_n and σ_n are polynomials satisfying the recurrences

$$(2)\qquad \boxed{\begin{aligned} \rho_0(\zeta) &:\equiv 1\,, & \rho_{n+1}(\zeta) &:= [\zeta\,\sigma_n(\zeta) - \rho_n(\zeta)\varphi_n]/\gamma_n\\ \sigma_0(\zeta) &:\equiv 1\,, & \sigma_{n+1}(\zeta) &:= \rho_{n+1}(\zeta) - \sigma_n(\zeta)\,\psi_{n+1} \end{aligned}}$$

"polynomial form of the BOBC algorithm" (BCG)

and

$$(3)\qquad \boxed{\begin{aligned} &\rho_0(\zeta) :\equiv 1\,, \quad \rho_1(\zeta) := (\zeta-\alpha_0)\,\rho_0(\zeta)/\gamma_0\,,\\ &\rho_{n+1}(\zeta) := [(\zeta-\alpha_n)\,\rho_n(\zeta) - \beta_n\,\rho_{n-1}(\zeta)]/\gamma_n\,. \end{aligned}}$$

"polynomial form of the BO algorithm"

ρ_n and σ_n have both <u>exact degree n</u> and <u>leading coefficient $(\gamma_0\gamma_1\cdots\gamma_{n-1})^{-1}$</u>.

When used for solving linear systems both algorithms are "polynomial acceleration methods" (semiiterative methods). Necessarily $\rho_n(0)=1$ ($\forall n$).

$\Longrightarrow$ Need <u>$\gamma_n+\alpha_n+\beta_n = 0$</u>, <u>$\varphi_n = -\gamma_n$</u> there.

Thm. 1 turns into:

$$y_0^H \rho_m(A)\rho_n(A)\, x_0 = \delta_{mn}\,\delta_n\,, \quad \text{biorthogonality}$$

$$y_0^H \sigma_m(A)\, A\, \sigma_n(A)\, x_0 = \delta_{mn}\,\delta_n\,\varphi_n\,. \quad \text{biconjugacy}$$

Define linear functionals Φ and Φ_1 on $\mathcal{P}$ by

$$\Phi(\zeta^k) := \mu_k\,, \qquad \Phi_1(\zeta^k) := \mu_{k+1}\,, \qquad k \in \mathbb{N}$$

where

$$\mu_k := y_0^H A^k x_0$$

Schwarz constants, moments, Markov parameters; $\{\mu_k\}$: impulse response

Then Thm. 1 turns into

$$(4) \qquad \boxed{\begin{aligned} \Phi(\rho_m \rho_n) &= \delta_{mn}\,\delta_n\,, \\ \Phi_1(\sigma_m \sigma_n) &= \delta_{mn}\,\delta_n\,\varphi_n\,, \end{aligned}} \qquad \begin{pmatrix} 0 \le m < \nu, \\ 0 \le n < \nu \end{pmatrix}$$

which, by the linearity of Φ and Φ_1 is equivalent to

$$(5) \qquad \boxed{\begin{aligned} \Phi(\zeta^m \rho_n) &= \begin{cases} 0 & \text{if } 0 \le m < n < \nu \\ \gamma_0 \cdots \gamma_{n-1}\delta_n & \text{if } m = n < \nu \end{cases} \\ \Phi_1(\zeta^m \sigma_n) &= \begin{cases} 0 & \text{if } 0 \le m < n < \nu \\ \gamma_0 \cdots \gamma_{n-1}\delta_n\varphi_n & \text{if } m = n < \nu \end{cases} \end{aligned}}$$

Hence, $\{\rho_n\}$ and $\{\sigma_n\}$ are sequences of formal orthogonal polynomials.

However, $(\rho, \sigma) \in \mathcal{P} \times \mathcal{P} \to \Phi(\rho\sigma) \in \mathbb{C}$ does not define an inner product unless $F(\zeta) := \sum_{k=0}^{\infty} \mu_k \zeta^{-k-1}$ is a Stieltjes function ($\Leftarrow$ A Hermitian pos. def. and $x_0 = y_0$).

Let

$$\varrho_n(\zeta) =: \sum_{k=0}^{n} \varrho_k^{(n)} \zeta^k \,, \qquad \sigma_n(\zeta) =: \sum_{k=0}^{n} \sigma_k^{(n)} \zeta^k$$

Then (5) implies

$$(6) \quad \begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & & & \vdots \\ \mu_{n-1} & \mu_n & \cdots & \mu_{2n-1} \end{pmatrix} \begin{pmatrix} \varrho_0^{(n)} \\ \varrho_1^{(n)} \\ \vdots \\ \varrho_n^{(n)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \mu_2 & \mu_3 & \cdots & \mu_{n+2} \\ \vdots & & & \vdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n} \end{pmatrix} \begin{pmatrix} \sigma_0^{(n)} \\ \sigma_1^{(n)} \\ \vdots \\ \sigma_n^{(n)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

These are $n \times (n+1)$ systems of linear equations, hence nontrivial solutions must exist. (We know that anyway.) Under our assumption of orthogonality of $\{\varrho_n\}_{n=0}^{\nu-1}$, $\{\sigma_n\}_{n=0}^{\nu-1}$ it is known that the solution is unique up to a scalar factor and that $\varrho_n^{(n)} \neq 0$, $\sigma_n^{(n)} \neq 0$.

Next, let

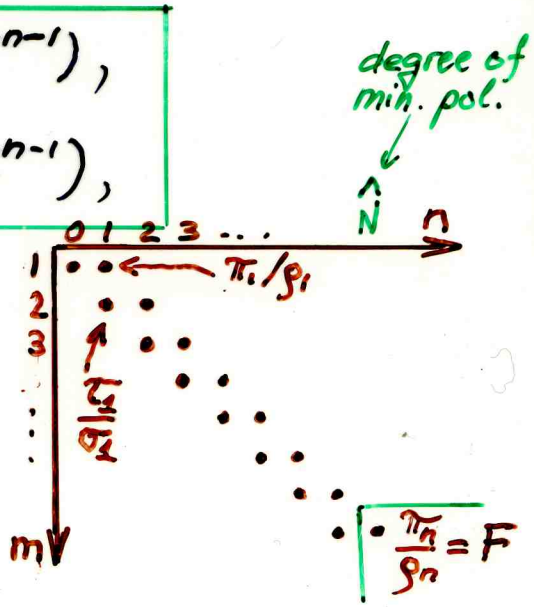
$$F(\zeta) := y_0^{\#} (\zeta I - A)^{-1} x_0 = \sum_{k=0}^{\infty} \frac{\mu_k}{\zeta^{k+1}}$$

rational Fct., order $\leq$ degree of minimal pol.

Then, by (6), there exist $\pi_n \in \mathcal{P}_{n-1}$, $\tau_n \in \mathcal{P}_n$ s.t.

$$\boxed{\begin{aligned} F(\zeta)\varrho_n(\zeta) - \pi_n(\zeta) &= O(\zeta^{-n-1}), \\ \zeta F(\zeta)\sigma_n(\zeta) - \tau_n(\zeta) &= O(\zeta^{-n-1}), \end{aligned}}$$

degree of min. pol.

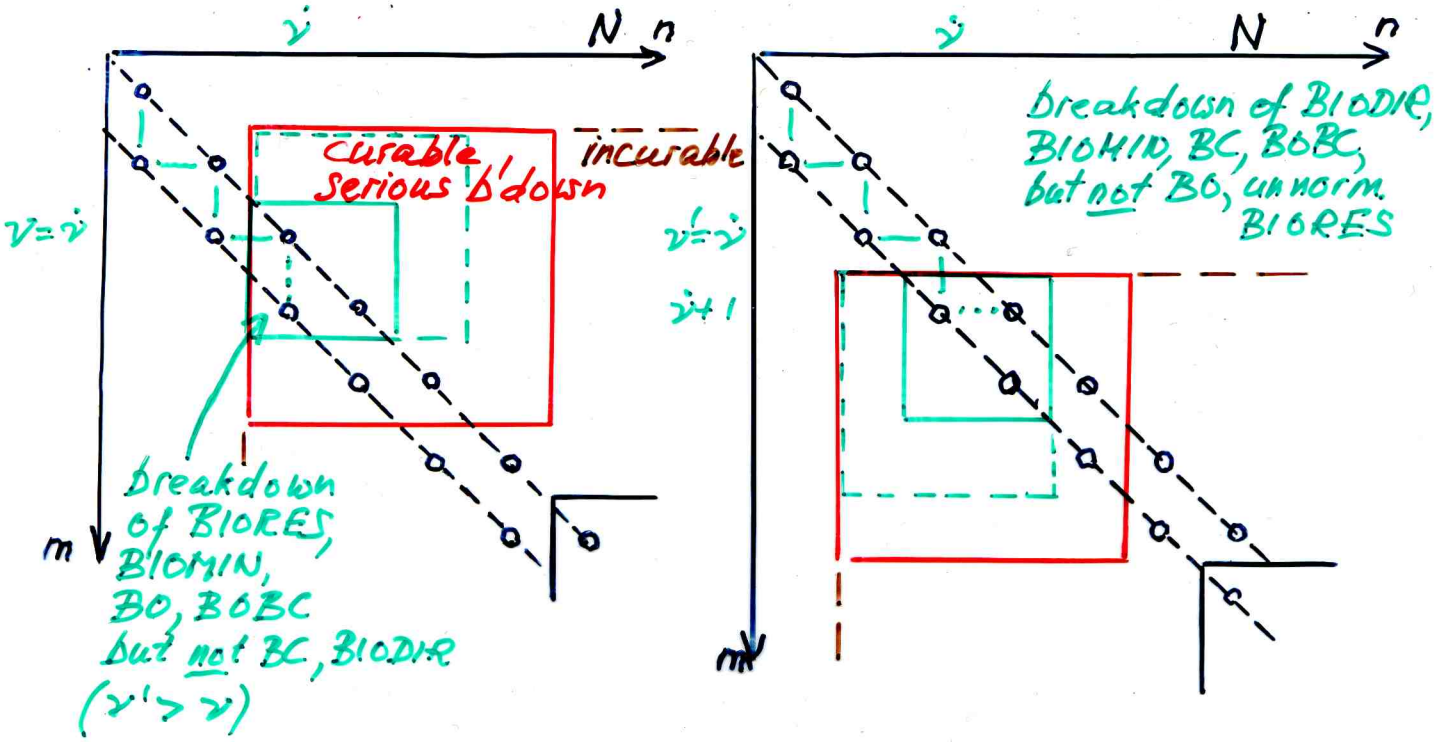


i.e.,

$$\frac{\pi_n}{\varrho_n} \quad \text{and} \quad \frac{\tau_n}{\sigma_n}$$

are the (n,n) and $(n+1,n)$ Padé approximant of F at $\zeta = \infty$. They lie on two adjacent diagonals of the Padé table.

The Padé table of F and breakdowns:



What can we say about F?

Consider EVEC decomposition of A (or transf. to Jordan form)

$$A W = W \underbrace{D_\lambda}_{\text{Jordan matrix}}, \quad W =: [w_1, w_2, \dots, w_N]$$

$$\Rightarrow A^H W^{-H} = W^{-H} D_\lambda^H, \quad W^{-H} =: [\hat{w}_1, \hat{w}_2, \dots, \hat{w}_N]$$

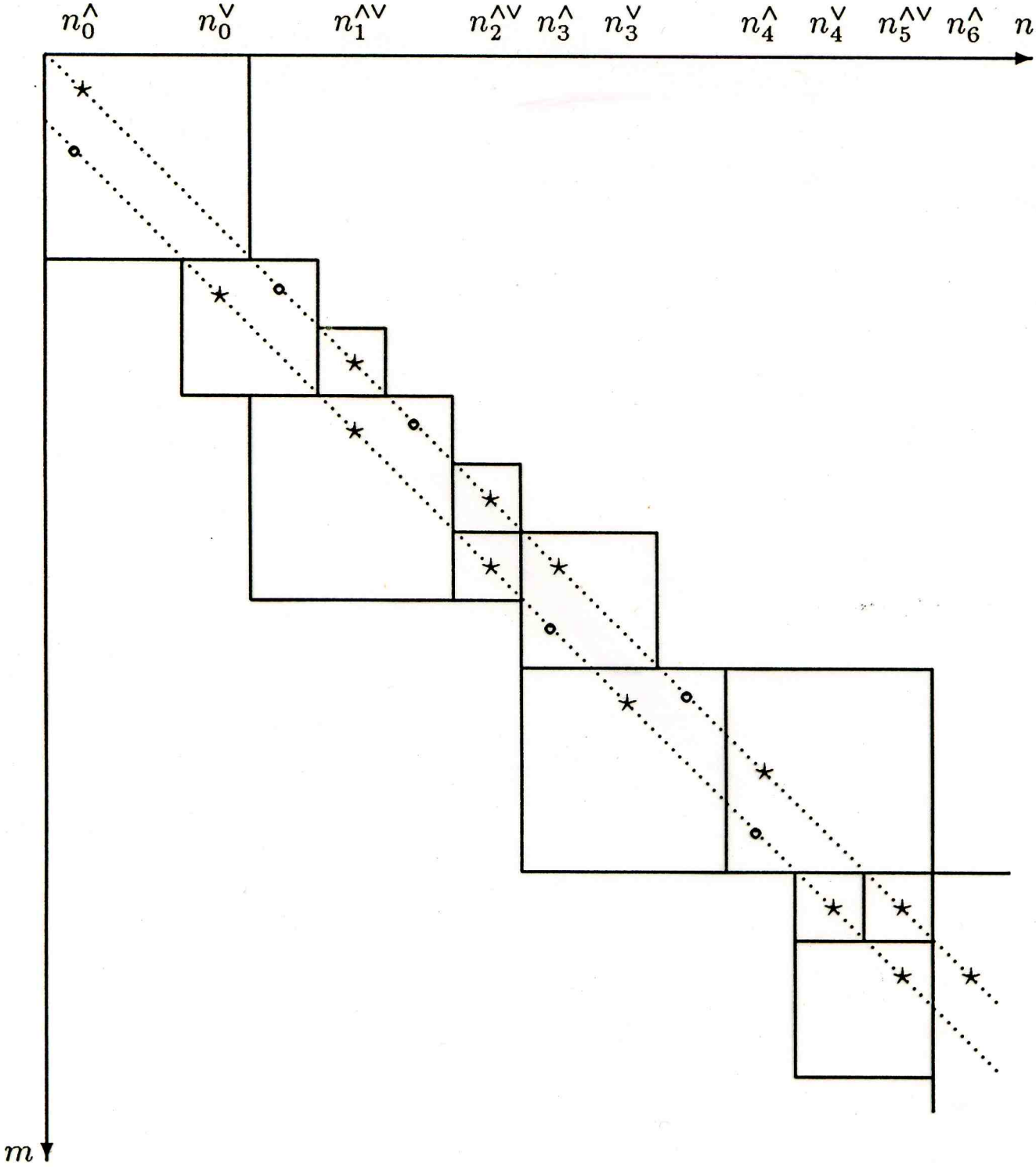
$$x_0 =: \sum_{j=1}^{N} w_j \xi_j, \quad y_0 =: \sum_{j=1}^{N} \hat{w}_j \eta_j$$

D_λ diag: $\mu_k := y_0^H A^k x_0 = \sum_{j=1}^{N} \lambda_j^k \xi_j \bar{\eta}_j$

$$F(\zeta) := y_0^H (\zeta I - A)^{-1} x_0 = \sum_{j=1}^{N} \frac{\xi_j \bar{\eta}_j}{\zeta - \lambda_j}$$

D_λ nondiag.: $F(\zeta) = \sum_{j=1}^{N} \sum_{i=0}^{\tau(j)} \frac{\xi_{j+i} \bar{\eta}_j}{(\zeta - \lambda_j)^{i+1}}$

$\tau(j) :=$ j – index of last column in D_λ that belongs to same Jordan subblock as column j.


$n_0^\wedge$
$n_0^\vee$
$n_1^{\wedge\vee}$
$n_2^{\wedge\vee}$
$n_3^\wedge$
$n_3^\vee$
$n_4^\wedge$
$n_4^\vee$
$n_5^{\wedge\vee}$
$n_6^\wedge$
n
m

The Stieltjes procedure

Recall: with $\gamma_n := 1$ $(\forall n)$,

(S1)
$$\boxed{\begin{aligned} &\rho_0(\zeta) := 1, \qquad \rho_1(\zeta) := (\zeta - \alpha_0)\rho_0 \\ &\rho_{n+1}(\zeta) := (\zeta - \alpha_n)\rho_n(\zeta) - \beta_n \rho_{n-1}(\zeta) \end{aligned}}$$

In Lanczos setting:

$$\Phi(\rho) := \boxed{y_0^H \rho(A) x_0} = \sum_{j=1}^{N} \rho(\lambda_j)\, \xi_j \bar{\eta}_j$$

if A diagonalizable

In orthogonal polynomial setting

$$\Phi(\rho) := \boxed{\int \rho(\zeta)\, \underbrace{d\omega(\zeta)}_{\substack{\text{pos. measure} \\ \text{support} \subseteq \mathbb{R}}}}$$

If $A = A^H$, $y_0 = x_0$: $\quad d\omega(\zeta) = \sum_{j=1}^{N} |\xi_j|^2 \delta(\zeta - \lambda_j)$ positiv measure

Dirac δ fct.

From (10):

$$\alpha_n := \frac{y_n^H A x_n}{y_n^H x_n} = \frac{\Phi(\zeta \rho_n^2)}{\Phi(\rho_n^2)}, \qquad \beta_n := \frac{y_n^H x_n}{y_{n-1}^H x_{n-1}} = \frac{\Phi(\rho_n^2)}{\Phi(\rho_{n-1}^2)}$$

Hence,

(S2)
$$\boxed{\alpha_n = \frac{\int \zeta \rho_n^2(\zeta)\, d\omega}{\int \rho_n^2(\zeta)\, d\omega}, \qquad \beta_n = \frac{\int \rho_n^2(\zeta)\, d\omega}{\int \rho_{n-1}^2(\zeta)\, d\omega}}$$

(S1) + (S2) = Stieltjes procedure [1884] = Lanczos for polyns.

Factorization of the moment matrix

$$\exists \quad P := \begin{pmatrix} * & * & * & \cdots \\ & * & * & \\ & & * & \\ & & & \ddots \end{pmatrix} \quad \text{s.t.} \quad \begin{pmatrix} \rho_0(\zeta) & \rho_1(\zeta) & \cdots & \rho_{\nu-1}(\zeta) \end{pmatrix} = \begin{pmatrix} 1 & \zeta & \cdots & \zeta^{\nu-1} \end{pmatrix} P$$

coeffs. of ρ_2

or,

$$X = \begin{pmatrix} x_0 & x_1 & \cdots & x_{\nu-1} \end{pmatrix} = \begin{pmatrix} x_0 & Ax_0 & \cdots & A^{\nu-1}x_0 \end{pmatrix} P$$

$$Y = \begin{pmatrix} y_0 & y_1 & \cdots & y_{\nu-1} \end{pmatrix} = \begin{pmatrix} y_0 & A^H y_0 & \cdots & (A^H)^{\nu-1} y_0 \end{pmatrix} \overline{P}$$

Biorthogonality, i.e. $Y^H X = D$, yields

$$\underbrace{P^{-T} D P^{-1}}_{\substack{LDU \\ \text{decomp.}}} = \underbrace{P^{-T} Y^H}_{(Y\overline{P}^{-1})^H} X P^{-1} = \begin{pmatrix} y_0^H \\ y_0^H A \\ \vdots \\ y_0 A^{\nu-1} \end{pmatrix} \begin{pmatrix} x_0 & Ax_0 & \cdots & A^{\nu-1}x_0 \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_{\nu-1} \\ \mu_1 & & \iddots & \vdots \\ \vdots & \iddots & & \vdots \\ \mu_{\nu-1} & \cdots & \cdots & \mu_{2\nu-1} \end{pmatrix}}_{\substack{\nu \times \nu \text{ moment matrix } M_\nu \\ \text{(Hankel)}}}$$

moments $\mu_k := y_0^H A^k x_0$

$$\Rightarrow \quad \underbrace{P^T M_\nu P = D}_{\text{"inverse" LDU dec.}} \quad \text{or} \quad \underbrace{M_\nu^{-1} = P D^{-1} P^T}_{\text{UDL dec. of } M_\nu^{-1}}$$

Basis of some fast $O(\nu^2)$ Hankel solvers

4. The relation to continued fractions

Continued fraction:

$$(1)\qquad b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \dots}}} = \mathop{K}_{j=0}^{J \text{ or } \infty} \frac{a_j}{b_j}$$

$$= \underbrace{b_0 + \frac{a_1|}{|b_1} + \frac{a_2|}{|b_2} + \frac{a_3|}{|b_3} + \dots + \frac{a_n|}{|b_n}}_{} + \dots$$

nth convergent w_n (or approximant) := "nth partial sum"

In our applications a_n and b_n are complex numbers or polynomials. By induction one can prove easily:

LEMMA [Wallin, Euler]

Define the nth numerator u_n and the nth denominator v_n of (1) by the recurrence

$$(2)\qquad \begin{aligned} &u_{-1} := 1, \quad u_0 := b_0, \quad u_n := b_n u_{n-1} + a_n u_{n-2}, \quad n \geq 1, \\ &v_{-1} := 0, \quad v_0 := 1, \quad v_n := b_n v_{n-1} + a_n v_{n-2}, \quad n \geq 1. \end{aligned}$$

Then

$$(3)\qquad w_n := b_0 + \mathop{K}_{j=0}^{n} \frac{a_j}{b_j} = \frac{u_n}{v_n}$$

Elimination of all odd [even] indexed u_n, v_n from (2) gives a new continued fraction, called the even [odd] part of (1). [cyclic reduction]

Recall:

(4) $$F(\zeta) := y_0^H(\zeta I - A)^{-1}x_0 = \sum_{k=0}^{\infty} \frac{\mu_k}{\zeta^{k+1}}, \quad \mu_k := y_0^H A^k x_0$$

$F(\zeta)$ can be expanded into a continued fraction in several ways. For example,

$$F(\zeta) = 0 + \frac{\mu_0/\zeta}{F_1(\zeta)}, \quad \text{where } F_1(\zeta) := (1 + \mu_0'\zeta^{-1} + \mu_1'\zeta^{-2} + \dots) = \frac{\mu_0}{\zeta}(F(\zeta) - 0)^{-1}$$

$$= 0 + \frac{\mu_0/\zeta|}{|1} + \frac{\mu_0'/\zeta|}{|F_2(\zeta)}$$

etc.

The following can be shown by comparing the reccurence formulas:

THM.: Assume $\nu = N$.

corresp. formal S-fraction

(i)
(5) $$\boxed{F(\zeta) = \frac{\mu_0|}{|\zeta} - \frac{\varphi_0|}{|1} - \frac{\gamma_0\psi_1|}{|\zeta} - \frac{\varphi_1|}{|1} - \frac{\gamma_1\psi_2|}{|\zeta} - \dots,}$$

where $\varphi_n, \psi_n, \gamma_n$ are the parameters of the biconjugate gradient method. For this continued fraction there holds, with $\Gamma_n := \gamma_0 \cdots \gamma_{n-1}$,

(6) $$\boxed{\begin{aligned} u_{2n} &= \Gamma_n \pi_n, & v_{2n} &= \Gamma_n \rho_n, \\ u_{2n+1} &= \Gamma_n \tau_n, & v_{2n+1} &= \Gamma_n \zeta \sigma_n, \end{aligned}}$$

and w_{2n} and w_{2n+1} are the $(n-1,n)$ and (n,n) Padé approximants of F at ∞, respectively.

(ii)

(7) $$\boxed{F(\zeta) = \frac{\mu_0|}{|\zeta-\alpha_0} - \frac{\beta_1\gamma_0|}{|\zeta-\alpha_1} - \frac{\beta_2\gamma_1|}{|\zeta-\alpha_2} - \cdots ,}$$ associated formal J-fraction

where $\alpha_n, \beta_n, \gamma_n$ are the parameters of the <u>BO algorithm</u>. For this continued fraction there holds:

(8) $$\boxed{u_n = \Gamma_n \pi_n \quad , \quad v_n = \Gamma_n \rho_n ,}$$

and w_n is the $(n-1, n)$ Padé approximant of F at $\zeta = \infty$. Hence, the continued fraction in (7) is the <u>even part</u> of the one in (5).

(iii)

(9) $$\boxed{F(\zeta) = \frac{\mu_0}{\zeta} + \frac{\mu_1\zeta^{-1}|}{|\zeta-\alpha_0'} - \frac{\beta_1'\gamma_0|}{|z-\alpha_1'} - \frac{\beta_2'\gamma_1|}{|z-\alpha_2'} - \cdots ,}$$

where

(10) $$\underline{\alpha_n' = \varphi_n + \gamma_n \psi_{n+1}} \quad , \quad \underline{\beta_{n+1}' = \varphi_{n+1}\psi_{n+1}} \quad (n \geq 0)$$

and γ_n are the elements of the matrix $T_1 = RTR^{-1}$. For this continued fraction there holds

(11) $$\boxed{u_n = \Gamma_n \tau_n \quad , \quad v_n = \Gamma_n \zeta \sigma_n ,}$$

and w_n is the (n,n) Padé approximant of F at $\zeta = \infty$. Hence, (9) is the <u>odd part</u> of (5).

(iv)

(12) $$\boxed{F(\zeta) = \frac{\mu_0}{\zeta} + \frac{\mu_1\zeta^{-1}|}{|\zeta} - \frac{\varphi_0'|}{|1} - \frac{\gamma_0\psi_1'|}{|\zeta} - \frac{\varphi_1'|}{|1} - \frac{\gamma_1\psi_2'|}{|\zeta} - \cdots}$$

where φ_n' and ψ_n' satisfy (with $\psi_{-1} := 0$)

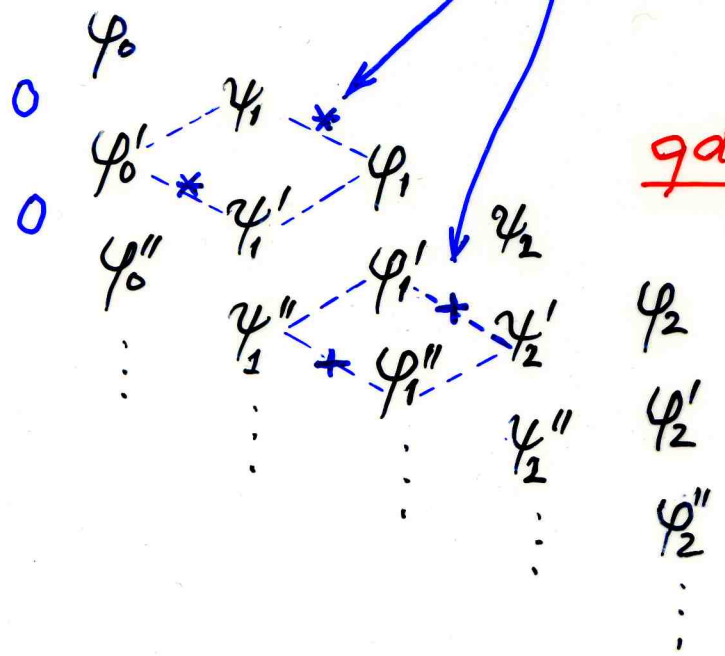
(13) $$\underline{\alpha_n' = \varphi_n' + \gamma_{n-1}\psi_n'} \quad , \quad \underline{\beta_{n+1}' = \varphi_n'\psi_{n+1}'} \quad (n \geq 0)$$

and are the element of the bidiagonal factors of the LU decomposition $T_1 = L_1 R_1$. (The diagonal elements of L_1 are γ_n, those of R are 1.) For this continued fraction w_{2n} and w_{2n+1} are the $(n, n+1)$ and $(n+1, n+1)$ Padé approximant of F, respectively.

(v) By equating (10) and (13) one has, if $\gamma_n = 1$ $(\forall n)$:

$$\boxed{\varphi_n' + \psi_n' = \varphi_n + \psi_{n+1}, \quad \varphi_n' \psi_{n+1}' = \varphi_{n+1} \psi_{n+1}} \quad (n \geq 0)$$

These are the <u>rhombus rules</u> of the qd algorithm.



qd table

Rutishauser:

$\varphi_{n-1} \rightsquigarrow q_n$

$\psi_n \rightsquigarrow e_n$

Outlook: Theory including nongeneric situations

$$F(\zeta) := y_0^H (\zeta I - A)^{-1} x_0 = \sum_{k=0}^{\infty} \frac{\mu_k}{\zeta^{k+1}}$$

linear functional Φ, $\Phi(\zeta^k) = \mu_k$ } as before

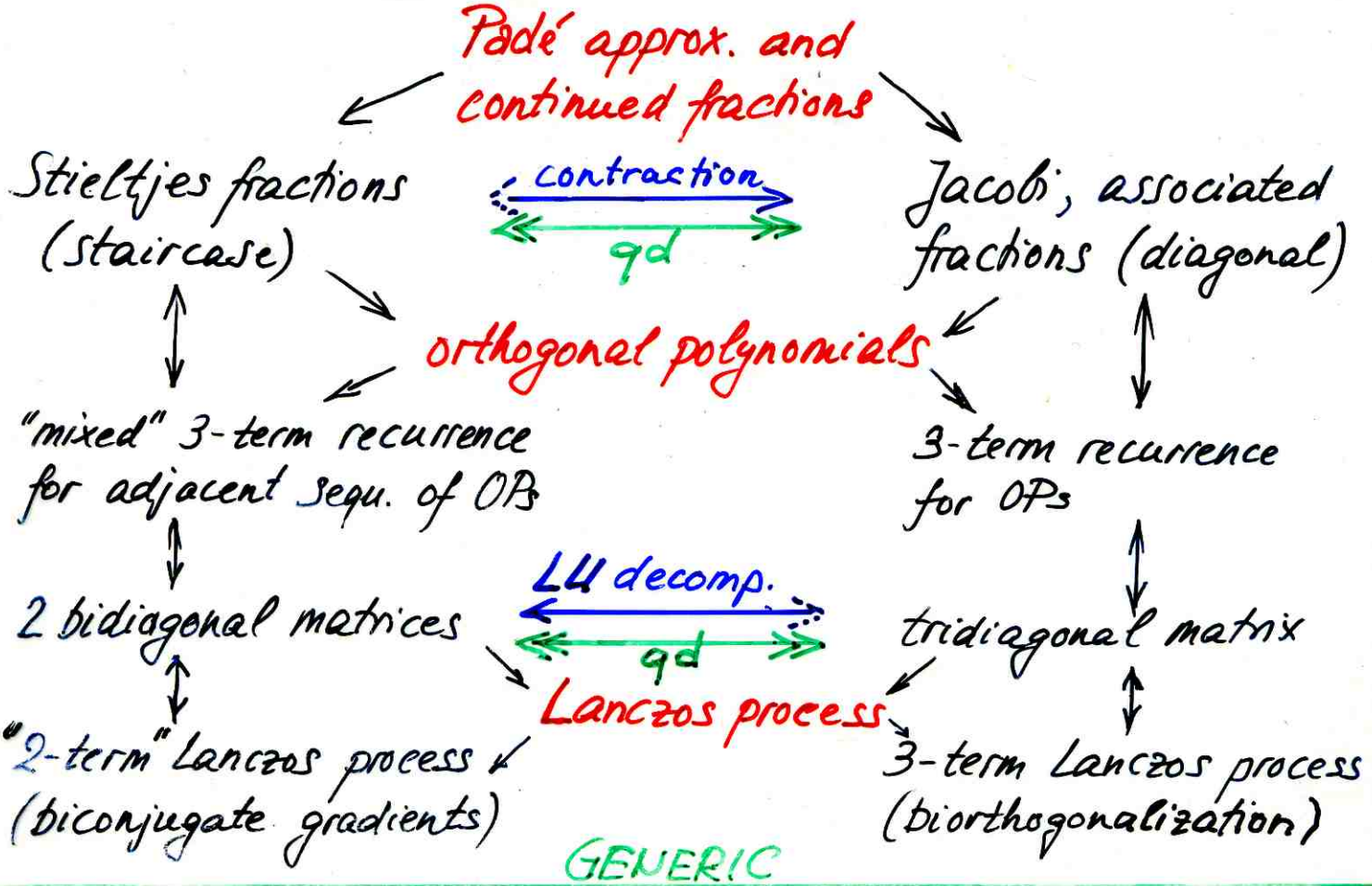
- Formal orthogonal polynomials w.r.t. Φ — G. Struble '63, A. Draux '83
  - definition, properties, conn. w/ Padé — (600 pages!)
  - recurrence relation — → simplify
- Matrix interpretation of recurrence — W.B. Gragg '74, A. Draux '83
  - $AX = XT$, T block tridiag.

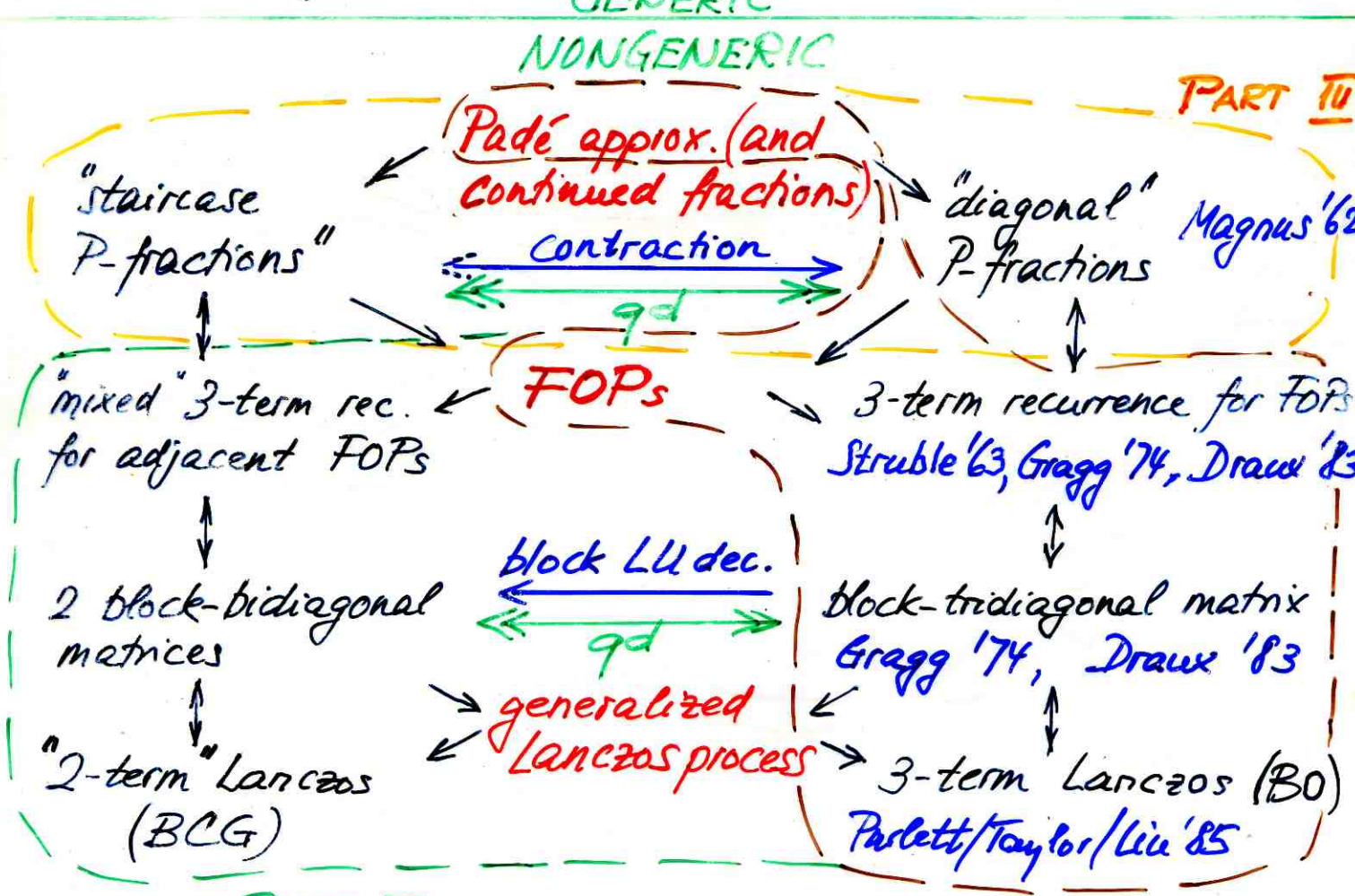
$$\alpha_n = \begin{pmatrix} 0 & & & & * \\ 1 & 0 & & & * \\ & 1 & \ddots & 0 & * \\ & & & 1 & * \end{pmatrix}, \quad \beta_n = \begin{pmatrix} 0 & * \end{pmatrix}, \quad \gamma_n = \begin{pmatrix} 0 & * \end{pmatrix}$$

- Continued fractions
  - "block diagonal": P-fractions — Chebyshev, Arne Magnus '62
  - "block staircase": staircase P-fractions — new
  - contraction of staircase P-fractions — new
    (generalization of odd/even part)
- qd algorithm (incl. deg. cases) — A. Draux, new [without polynom. messy]
- NGBO algorithm, NGBCG algorithm, NGBCS algorithm — new
- Matrix interpretation of relations between
  - block diagonal – block staircase ⇒ block LU decomp.
  - NGBO – NGBCO ⇒ " UL "
  - diagonals of qd parameters ⇒ " LR algorithm

key chapter

LP overview


Padé approx. and continued fractions
Stieltjes fractions (staircase)
contraction
qd
Jacobi, associated fractions (diagonal)
orthogonal polynomials
"mixed" 3-term recurrence for adjacent sequ. of OPs
3-term recurrence for OPs
LU decomp.
qd
2 bidiagonal matrices
tridiagonal matrix
Lanczos process
"2-term" Lanczos process (biconjugate gradients)
3-term Lanczos process (biorthogonalization)


GENERIC
NONGENERIC
PART III
Padé approx. (and continued fractions)
"staircase P-fractions"
contraction
qd
"diagonal" P-fractions
Magnus '62
FOPs
"mixed" 3-term rec. for adjacent FOPs
3-term recurrence for FOPs
Struble '63, Gragg '74, Draux '83
block LU dec.
qd
2 block-bidiagonal matrices
block-tridiagonal matrix
Gragg '74, Draux '83
generalized Lanczos process
"2-term" Lanczos (BCG)
3-term Lanczos (BO)
Parlett/Taylor/Liu '85
PART II
SIMAX 15, Jan. '94
PART I
SIMAX 13, Apr. '92