Moments,
Model Reduction and Nonlinearity
in Solving Linear Algebraic Problems

Zdeněk Strakoš
Charles University, Prague

http://www.karlin.mff.cuni.cz/~strakos

16th ILAS Meeting, Pisa, June 2010,
French-Czech WS, Prague, July 2010.

Thanks

I wish to give my thanks to many authors and collaborators whose work and help has enabled the presented view. Related to moments, I feel more and more indebted to Gene Golub.

Iveta Hnětynková,
Petr Tichý,
Dianne O'Leary,
Gerard Meurant.

Stieltjes (1893-1894)

Given a sequence of numbers $\xi_k, \; k = 0, 1, \ldots$, a non-decreasing distribution function $\omega(\lambda)$ is sought such that

$$\int_0^\infty \lambda^k \, d\omega(\lambda) \; = \; \xi_k \, , \quad k = 0, 1, \ldots \, ,$$

where ξ_k represents the k-th (generalized) mechanical moment of the distribution of positive mass on the half line $\lambda \geq 0$. Key tool: continued fractions; cf. Stieltjes (1884, 1893-94), Chebyshev (1855, 1859, ...).

The story goes back at least to Gauss (1814) .....

Matching moments formulation

Consider a non-decreasing distribution function $\omega(\lambda)$, $\lambda \geq 0$ with the moments given by the Riemann-Stieltjes integral

$$\xi_k = \int_0^\infty \lambda^k \, d\omega(\lambda) , \quad k = 0, 1, \dots .$$

Find the distribution function $\omega^{(n)}(\lambda)$ with n points of increase $\lambda_i^{(n)}$ $i = 0, 1, \dots$, which matches the first $2n$ moments for the distribution function $\omega(\lambda)$,

$$\int_0^\infty \lambda^k \, d\omega^{(n)}(\lambda) \equiv \sum_{i=1}^{n} \omega_i^{(n)} (\lambda_i^{(n)})^k = \xi_k , \quad k = 0, 1, \dots, 2n-1 .$$

Outline

1. Matching moments model reduction

2. CG and Gauss-Christoffel quadrature

3. Non-Hermitian generalizations

4. Noise level in discrete ill-posed problems

5. Conclusions

1 : Model reduction in dynamical systems

Consider a linear dynamical system (here we assume A HPD)

$$\begin{aligned} z' &= A\,z + bu\,, \\ y &= b^* z\,. \end{aligned}$$

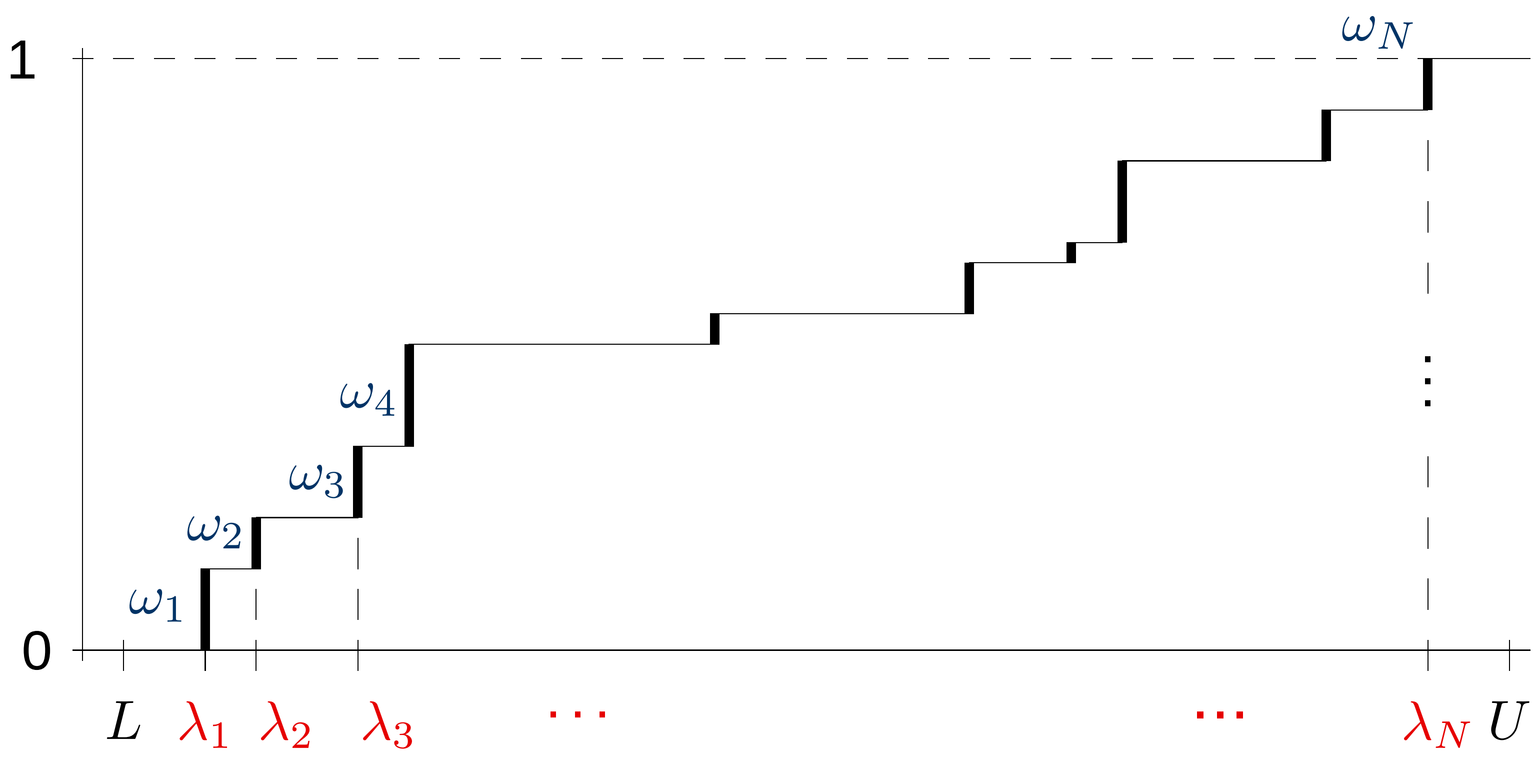
The transfer function

$$b^*(\lambda I - A)^{-1} b$$

given by the Laplace transform describes the impulse-response in the frequency domain.

1 : Distribution function $\omega(\lambda)$

λ_i, s_i are the eigenpairs of A , $\omega_i = |(s_i, w_1)|^2$, $w_1 = b/\|b\|$



1 : Stieltjes recurrence

Let $p_0(\lambda) \equiv 1, p_1(\lambda), \dots, p_n(\lambda)$ be the first $n+1$ orthonormal polynomials corresponding to the distribution function $\omega(\lambda)$.

Then, writing $P_n(\lambda) = [p_0(\lambda), \dots, p_{n-1}(\lambda)]^T$,

$$\lambda \, P_n(\lambda) \;=\; T_n \, P_n(\lambda) \,+\, \delta_{n+1} \, p_n(\lambda) \, e_n$$

represents the Stieltjes recurrence (1893-4), see Chebyshev (1855), Brouncker (1655), Wallis (1656), with the Jacobi matrix

$$T_n \equiv \begin{pmatrix} \gamma_1 & \delta_2 & & \\ \delta_2 & \gamma_2 & \ddots & \\ & \ddots & \ddots & \delta_n \\ & & \delta_n & \gamma_n \end{pmatrix}, \quad \delta_l > 0\,, \ell = 2, \dots, n\,.$$

1 : Jacobi matrix and quadrature

Consider the nth Gauss-Christoffel quadrature approximation $\omega^{(n)}(\lambda)$ of the Riemann-Stieltjes distribution function $\omega(\lambda)$. Its algebraic degree is $2n-1$, i.e., it matches the first $2n$ moments

$$\xi_{\ell-1} = \int_L^U \lambda^{\ell-1}\, d\omega(\lambda) = \sum_{j=1}^{n} \omega_j^{(n)} \{\lambda_j^{(k)}\}^{\ell-1}, \quad \ell = 1, \ldots, 2n\,,$$

where

$$\omega_i^{(n)} = |(z_i^{(n)}, e^1)|^2\,, \quad \lambda_i^{(n)} = \theta_i^{(n)}\,,$$

and $z_i^{(n)}$ is the normalized eigenvector of T_n corresponding to the eigenvalue $\theta_i^{(n)}$. The (orthogonal) polynomial $p_n(\lambda)$ has the roots $\theta_i^{(n)}$, $i = 1, \ldots, n$.

1 : Continued fraction corresponding to $\omega(\lambda)$

$$\mathcal{F}_N(\lambda) \equiv \cfrac{1}{\lambda - \gamma_1 - \cfrac{\delta_2^2}{\lambda - \gamma_2 - \cfrac{\delta_3^2}{\lambda - \gamma_3 - \ldots \cfrac{\ddots}{\lambda - \gamma_{N-1} - \cfrac{\delta_N^2}{\lambda - \gamma_N}}}}}$$

The entries $\gamma_1, \ldots, \gamma_N$ and $\delta_2, \ldots, \delta_N$ represent coefficients of the Stieltjes recurrence.

1 : Partial fraction decomposition

Consider (for simplicity of notation) $\|b\| = 1$. Using the spectral decomposition,

$$b^*(\lambda I - A)^{-1} b \;=\; \int_L^U \frac{d\omega(\mu)}{\lambda - \mu} \;=\; \sum_{j=1}^{N} \frac{\omega_j}{\lambda - \lambda_j} = \frac{\mathcal{R}_N(\lambda)}{\mathcal{P}_N(\lambda)}\,,$$

$$\frac{\mathcal{R}_N(\lambda)}{\mathcal{P}_N(\lambda)} \;\equiv\; \mathcal{F}_N(\lambda)$$

The denominator $\mathcal{P}_n(\lambda)$ corresponding to the nth convergent $\mathcal{F}_n(\lambda)$ of $\mathcal{F}_N(\lambda)$, $n = 1, 2, \ldots$ is the nth orthogonal polynomial in the sequence determined by $\omega(\lambda)$; see Chebyshev (1855).

1 : Expansion at infinity

Recall ($\|b\| = 1$)

$$b^*(\lambda I - A)^{-1} b \;=\; \int_L^U \frac{d\omega(\mu)}{\lambda - \mu} \;=\; \sum_{j=1}^{N} \frac{\omega_j}{\lambda - \lambda_j} \;\equiv\; \mathcal{F}_N(\lambda)\,,$$

and consider

$$\frac{1}{\lambda - \mu} \;=\; \frac{1}{\lambda}\left(1 - \frac{\mu}{\lambda}\right)^{-1} \;=\; \frac{1}{\lambda}\left(1 + \frac{\mu}{\lambda} + \frac{\mu^2}{\lambda^2} + \ldots\right),$$

$$\frac{1}{\lambda - \lambda_j} \;=\; \frac{1}{\lambda}\left(1 - \frac{\lambda_j}{\lambda}\right)^{-1} \;=\; \frac{1}{\lambda}\left(1 + \frac{\lambda_j}{\lambda} + \frac{\lambda_j^2}{\lambda^2} + \ldots\right).$$

1 : Minimal partial realization

This gives the ninimal partial realization

$$b^*(\lambda I - A)^{-1} b \; = \; \mathcal{F}_N(\lambda) \; = \; \sum_{\ell=1}^{2n} \frac{\xi_{\ell-1}}{\lambda^\ell} \; + \; \mathcal{O}\left(\frac{1}{\lambda^{2n+1}} \right) .$$

Using the same expansion of the nth convergent $\mathcal{F}_n(\lambda)$ of $\mathcal{F}_N(\lambda)$,

$$\mathcal{F}_n(\lambda) \; = \; \sum_{\ell=1}^{2n} \frac{\xi_{\ell-1}}{\lambda^\ell} \; + \; \mathcal{O}\left(\frac{1}{\lambda^{2n+1}} \right) ,$$

Where the moments in the numerators are identical due to the Gauss quadrature property.

1 : Minimal partial realization 1859 - 94

The nth convergent $\mathcal{F}_n(\lambda)$ of $\mathcal{F}_N(\lambda)$ matches $2n$ moments $\xi_0, \ldots, \xi_{2n-1}$, and it approximates $\mathcal{F}_N(\lambda)$ with the error proportional to $\lambda^{-(2n+1)}$. This represents the minimal partial realization; see

Chebyshev (1859), Stieltjes (1894).

The minimal partial realization was rediscovered in the engineering literature by Kalman (1979).

The works of Krylov (1931), Hestenes and Stiefel (1952), Vorobyev (1958, 1965) (see Brezinski (1991, ...)), were not studied and recalled. The links with Chebyshev and Stieltjes were pointed out by Gragg (1974), Bultheel and Van Barel (1997).

Is it of any good to recall these line of thoughts in modern NLA?

We wish to solve large systems of linear algebraic equations etc. by modern methods and algorithms and not to be bothered by some stories and the thoughts of Chebyshev or Stieltjes about some moments .... They can not be used in computations .... Are they of any use?

Whatever we think, the moments are going to get us.

Outline

1. Matching moments model reduction

2. CG and Gauss-Christoffel quadrature

3. Non-Hermitian generalizations

4. Noise level in discrete ill-posed problems

5. Conclusions

2 : Specific problem for $\lambda = 0$

Consider $b^*(\lambda I - A)^{-1} b$ with $\lambda = 0$. Then the minimal partial realization representing the expansion at infinity is not applicable. We do either

$$b^* A^{-1} b = b^* (A^{-1} b) \approx b^* x_n ,$$

with x_n being an approximation to the solution of $A x = b$, or

$$b^* A^{-1} b \approx b^* A_n^{-1} b ,$$

with A_n being an approximation to A. Mathematically, both approaches are equivalent. Computationally, however, they can give different results.

2 : Caution

A possible source of confusion with model reduction in dynamical systems.

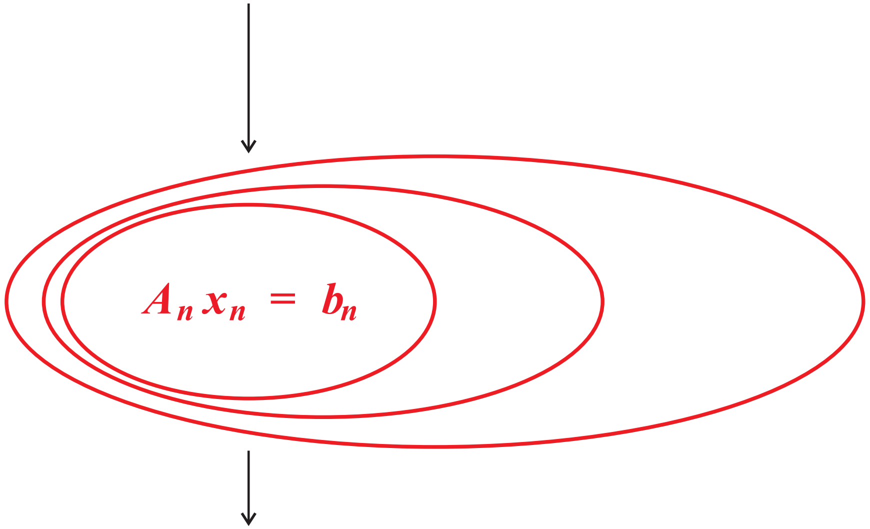
Approximation of the scalar value $b^* A^{-1} b$ will lead to matching the same moments as in the minimal partial realization given by the expansion of the function $\mathcal{F}_N(\lambda)$ at infinity.

The expansion of $\mathcal{F}_N(\lambda)$ at zero that uses matching moments with A^{-1} is in approximating the scalar $b^* A^{-1} b$ of no use.

The error of the approximation will be expressed in an elegant way with no relation to the error estimates in dynamical systems model reduction (see below).

2 : Krylov subspace methods

$$Ax = b$$



x_n approximates the solution x
using the subspace of small dimension.

$$\mathcal{S}_n \equiv \mathcal{K}_n(A, r_0) \equiv span\,\{r_0, Ar_0, \cdots, A^{n-1}r_0\} \longrightarrow \text{moments !}$$

2 : Conjugate gradients (CG): A HPD

$$\|x - x_n\|_A \;=\; \min_{u \,\in\, x_0 + \mathcal{K}_n(A, r_0)} \|x - u\|_A$$

with the formulation via the Lanczos process, $w_1 = r_0 / \|r_0\|$,

$$A\, W_n \;=\; W_n\, T_n \;+\; \delta_{n+1} w_{n+1} e_n^T \,, \qquad T_n \;=\; W_n^*(A)\, A\, W_n(A)\,,$$

and the CG approximation given by

$$T_n\, y_n \;=\; \|r_0\| e_1 \,, \qquad x_n \;=\; x_0 \;+\; W_n\, y_n \,.$$

$$A_n \;=\; Q_n\, A\, Q_n \;=\; W_n W_n^*\, A\, W_n W_n^* \;=\; W_n\, T_n\, W_n^* \,,$$

2 : Computational algorithm

Given x_0 (in approximating $b^* A^{-1} b$ we set $x_0 = 0$), $r_0 = b - Ax_0$,
$p_0 = r_0$

For $n = 1, 2, \ldots$

$$\gamma_{n-1} = (r_{n-1}, r_{n-1}) / (p_{n-1}, Ap_{n-1})$$

$$\begin{aligned} x_n &= x_{n-1} + \gamma_{n-1}\, p_{n-1} \\ r_n &= r_{n-1} - \gamma_{n-1}\, Ap_{n-1} \end{aligned}$$

$$\delta_n = (r_n, r_n) / (r_{n-1}, r_{n-1})$$

$$p_n = r_n + \delta_n\, p_{n-1} .$$

Search directions are given by the modified residuals, γ_{n-1} gives the line search minimum, δ_n ensures the local A-orthogonality of the direction vectors. No moments are visible. If we wish to get an insight, we need them.

2 : Numerical PDE connection of CG

Find $u \equiv u(\xi_1, \xi_2)$, where ξ_1, ξ_2 denote the space variables, such that

$$-\nabla^2 u \; = \; f \quad \text{in a bounded domain } \; \Omega \subset \mathbb{R}^2 \, ,$$

$$u \; = \; g_{\mathcal{D}} \; \text{ on } \; \partial\Omega_{\mathcal{D}}, \quad \text{and} \quad \frac{\partial u}{\partial n} \; = \; g_{\mathcal{N}} \; \text{ on } \; \partial\Omega_{\mathcal{N}} \, ,$$

where $\partial\Omega_{\mathcal{D}} \cup \partial\Omega_{\mathcal{N}} \; = \; \partial\Omega$, and $\partial\Omega_{\mathcal{D}} \cap \partial\Omega_{\mathcal{N}} \; = \; \emptyset$.

For the Galerkin FEM approximation

$$\|\nabla(u - u_h^{(n)})\|^2 \; = \; \|\nabla(u - u_h)\|^2 \; + \; \|x - x_n\|_A^2 \, .$$

An insight from viewing CG through moments?

2 : CG $\equiv$ matrix formulation of the Gauss q.

$$Ax = b\,,\; x_0 \qquad \longleftrightarrow \qquad \int_\zeta^\xi (\lambda)^{-1}\, d\omega(\lambda)$$

$$\uparrow \qquad\qquad\qquad\qquad \uparrow$$

$$T_n\, y_n \;=\; \|r_0\|\, e_1 \qquad \longleftrightarrow \qquad \sum_{i=1}^{n} \omega_i^{(n)} \left(\theta_i^{(n)}\right)^{-1}$$

$$x_n \;=\; x_0 + W_n\, y_n$$

$$\boxed{\omega^{(n)}(\lambda) \longrightarrow \omega(\lambda)}$$

2 : CG and Gauss-Ch. quadrature errors

$$\int_L^U f(\lambda)\, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} f(\theta_i^{(n)}) \;+\; R_n(f)\,.$$

For $f(\lambda) \equiv \lambda^{-1}$ the formula takes the form

$$\frac{\|x - x_0\|_{\mathbf{A}}^2}{\|r_0\|^2} \;=\; n\text{-th Gauss quadrature} \;+\; \frac{\|x - x_n\|_{\mathbf{A}}^2}{\|r_0\|^2}\,.$$

or, with $x_0 = 0$,

$$b^* A^{-1} b \;=\; \sum_{j=0}^{n-1} \gamma_j \|r_j\|^2 \;+\; r_n^* A^{-1} r_n\,.$$

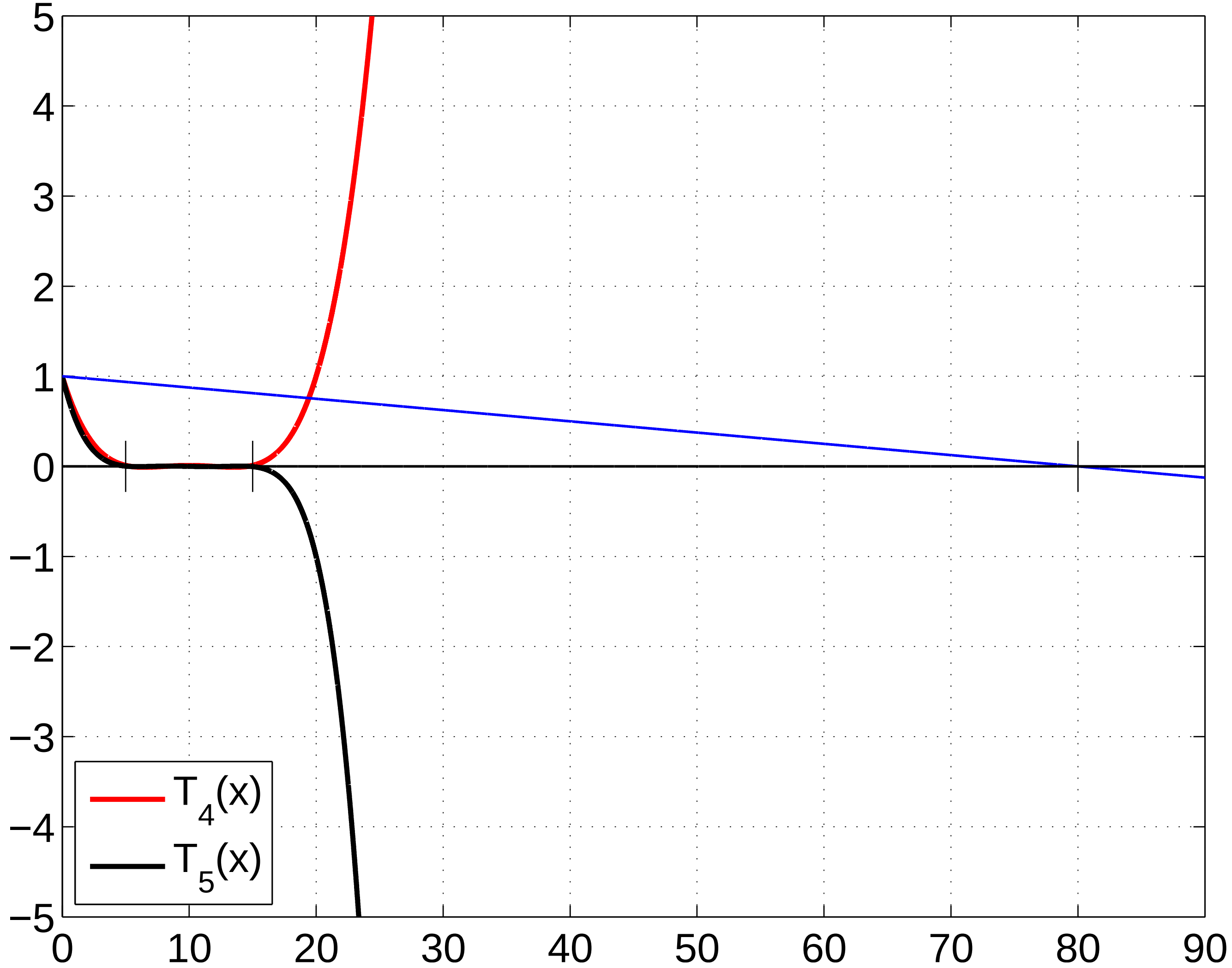
2 : Mathematical model of FP CG

CG in finite precision arithmetic can be seen as the exact arithmetic CG for the problem with the slightly modified distribution function with larger support, i.e., with single eigenvalues replaced by tight clusters.

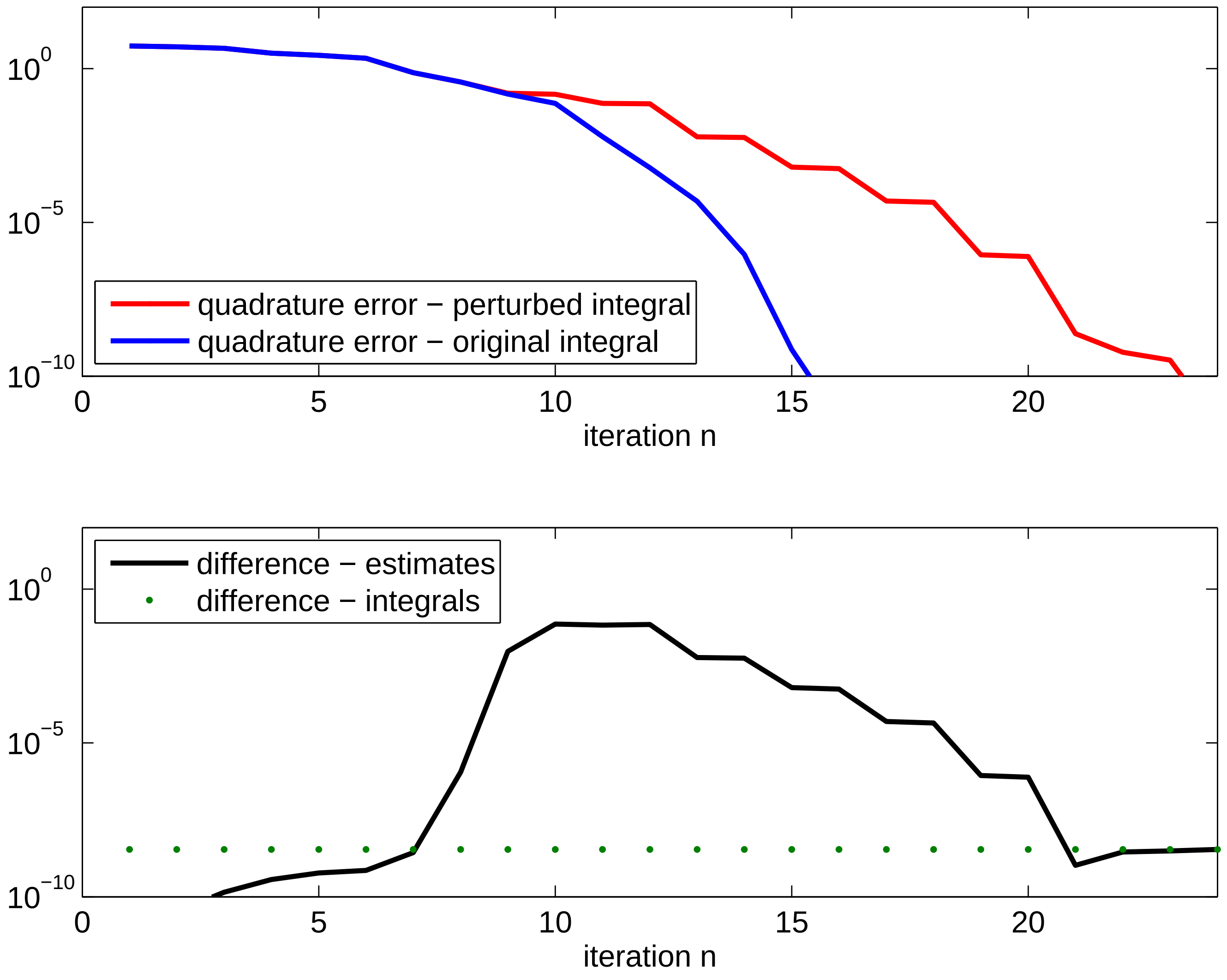
Paige (1971-80), Greenbaum (1989),
Parlett (1990), S (1991), Greenbaum and S (1992), Notay (1993), ... ,
Druskin, Kniznermann, Zemke, Wülling, Meurant, ...
Recent review and update in Meurant and S, Acta Numerica (2006).

One particular consequence (recently very relevant): In FP computations, the composed convergence bounds eliminating large outlying eigenvalues at the cost of one iteration per eigenvalue (see Axelsson (1976), Jennings (1977)) are not relevant. The recent theory of support preconditioners should be modified.

2 : Axelsson (1976), Jennings (1977)



2 : Sensitivity of the Gauss-Ch. Q.



2 : A point going back to 1814

1. Gauss-Christoffel quadrature for a small number of quadrature nodes can be highly sensitive to small changes in the distribution function that enlarge its support.
   In particular, the difference between the corresponding quadrature approximations (using the same number of quadrature nodes) can be many orders of magnitude larger than the difference between the integrals being approximated.

2. This sensitivity in Gauss-Christoffel quadrature can be observed for discontinuous, continuous, and even analytic distribution functions, and for analytic integrands uncorrelated with changes in the distribution functions, with no singularity close to the interval of integration.

O'Leary, S, Tichý (2007)

Outline

1. Matching moments model reduction

2. CG and Gauss-Christoffel quadrature

3. Non-Hermitian generalizations (S, Tichý (2007))

4. Noise level in discrete ill-posed problems

5. Conclusions

3 : Using Vorobyev method of moments (1958)

$$c^* A^{-1} b \approx c^* A_n^{-1} b ,$$

where A_n is the restriction of A to $\mathcal{K}_n(A, b)$ projected orthogonally to $\mathcal{K}_n(A^*, c)$,

$$A_n = W_n V_n^* A W_n V_n^* ,$$

with the matrix representation of the inverse

$$A_n^{-1} = W_n T_n^{-1} V_n^* .$$

Here $W_n^* V_n = I$, the columns of W_n and V_n span $\mathcal{K}_n(A, b)$ resp. $\mathcal{K}_n(A^*, c)$, and T_n matches the $2n$ moments (Lanczos - BiCG)

$$v_1^* A^k w_1 \equiv e_1^T T_n^k e_1 , \quad k = 0, 1, \ldots, 2n-1 .$$

3 : The BiCG method

Simultaneous solving of $Ax = b, \qquad A^*y = c.$

input A, b, c

$$x_0 = y_0 = 0$$
$$r_0 = p_0 = b,\ s_0 = q_0 = c$$

for $n = 0, 1, \dots$

$$\alpha_n = \frac{s_n^* r_n}{q_n^* A p_n},$$

$$x_{n+1} = x_n + \alpha_n p_n, \qquad y_{n+1} = y_n + \alpha_n{}^* q_n,$$
$$r_{n+1} = r_n - \alpha_n A p_n, \qquad s_{n+1} = s_n - \alpha_n^* A^* q_n,$$

$$\beta_{n+1} = \frac{s_{n+1}^* r_{n+1}}{s_n^* r_n},$$

$$p_{n+1} = r_{n+1} + \beta_{n+1} p_n, \qquad q_{n+1} = s_{n+1} + \beta_{n+1}^* q_n$$

end

3 : The BiCG approximation to $c^* A^{-1} b$

Using local biorthogonality

$$c^* A^{-1} b \;=\; \sum_{j=0}^{n-1} \alpha_j s_j^* r_j \;+\; s_n^* A^{-1} r_n \, .$$

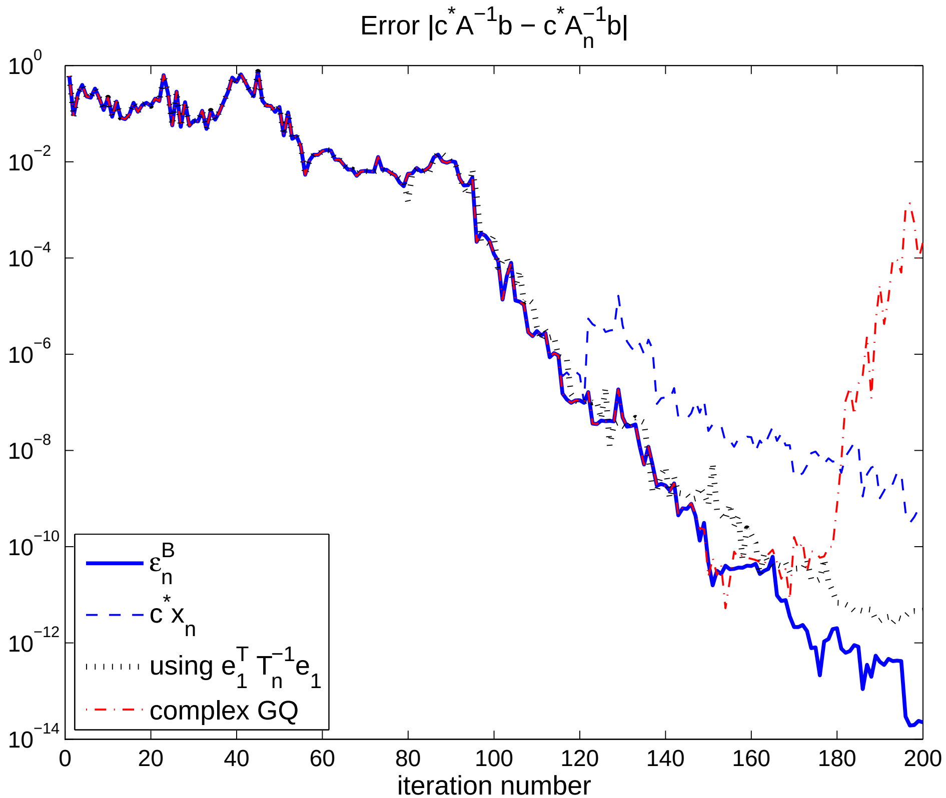
Using global biorthogonality

$$c^* A^{-1} b \;=\; c^* x_n \;+\; s_n^* A^{-1} r_n \, .$$

Finally,

$$c^* A_n^{-1} b \;=\; (c^* v_1) \, \|b\| \, (T_n^{-1})_{1,1} \;=\; c^* x_n = \sum_{j=0}^{n-1} \alpha_j s_j^* r_j \, .$$

3 : RCWA - comparison of estimates



Comparison of mathematically equivalent estimates
based on BiCG and Non-Hermitian Lanczos.

Outline

1. Matching moments model reduction
2. CG and Gauss-Christoffel quadrature
3. Non-Hermitian generalizations
4. Noise level in discrete ill-posed problems
5. Conclusions

4 : Discrete ill-posed problems

$$A\,x \approx b, \qquad A \in \mathbb{R}^{n\times n}, \qquad b \in \mathbb{R}^n,$$

with the right-hand side corrupted by a white noise

$$b = b^{\text{exact}} + b^{\text{noise}} \neq 0 \in \mathbb{R}^n\,, \quad \|\,b^{\text{exact}}\,\| \gg \|\,b^{\text{noise}}\,\|\,,$$

and the goal to approximate $x^{\text{exact}} \equiv A^{-1}\,b^{\text{exact}}$.

The noise level $\delta_{\text{noise}} \equiv \dfrac{\|b^{\text{noise}}\|}{\|b^{\text{exact}}\|} \ll 1\,.$

4 : Properties (assumptions)

- matrices A, A^T, AA^T have a smoothing property;
- left singular vectors u_j of A represent increasing frequencies as j increases (recall the Fredholm integral equation behind);
- the system $Ax = b^{\text{exact}}$ satisfies the discrete Picard condition.

Discrete Picard condition (DPC):

On average, the components $|(b^{\text{exact}}, u_j)|$ of the true right-hand side b^{exact} in the left singular subspaces of A decay faster than the singular values σ_j of A, $j = 1, \ldots, n$.

White noise:

The components $|(b^{\text{noise}}, u_j)|$, $j = 1, \ldots, n$ do not exhibit any trend.

4 : Hybrid methods

Golub-Kahan iterative bidiagonalization $(\, s_1 = b / \|b\| \,)$

$$A^T\, S_k = W_k\, L_k^T\,,$$

$$A\, W_k = [\, S_k,\, s_{k+1}\,]\, L_{k+} \;=\; S_{k+1}\, L_{k+}$$

is used in combination with an inner regularization of the projected bidiagonal problem.

Stopping criteria? Knowing the noise level would make a big difference.

Under the given (natural) assumptions,

the Golub-Kahan iterative bidiagonalization reveals the noise level δ_{noise} ;
see Hnětynková, Plešinger, S (2009).

4 : Main point

The GK bidiagonalization is closely related to the Lanzcos tridiagonalization and therefore to the matching moments problem. It generates at each step k the distribution function

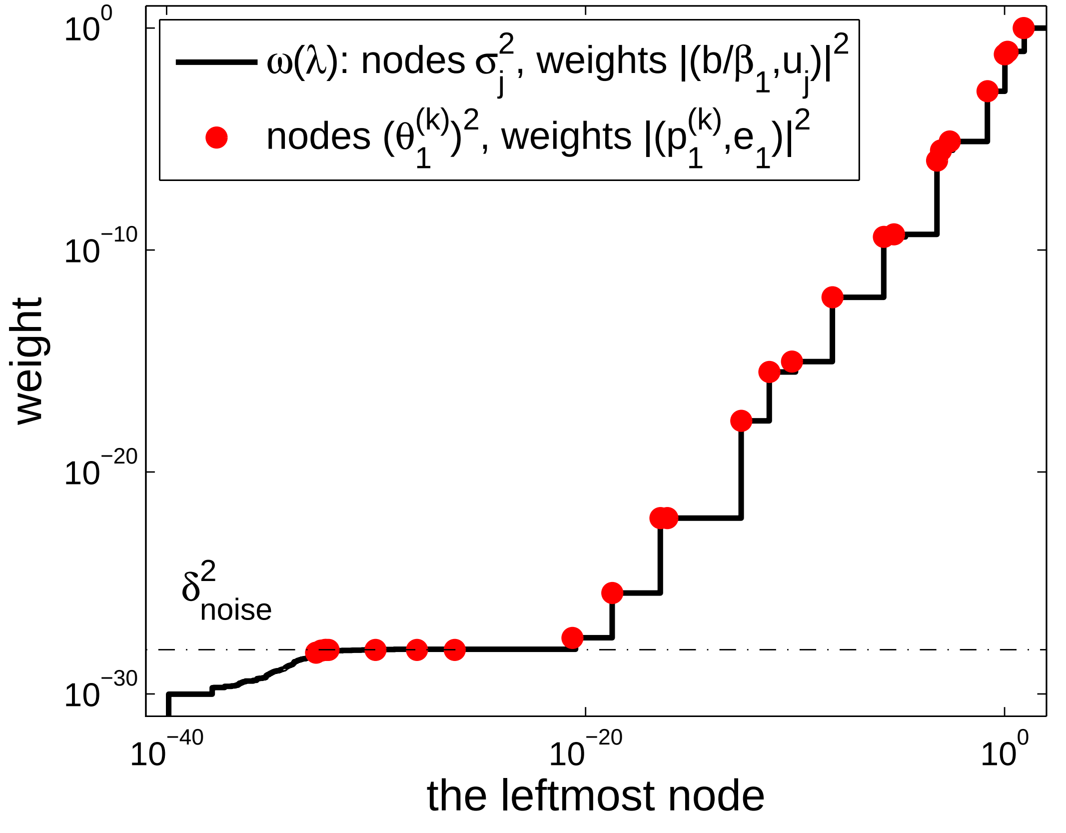
$$\omega^{(k)}(\lambda) \quad \text{with nodes} \quad (\theta_\ell^{(k)})^2 \quad \text{and weights} \quad \omega_\ell^{(k)} = |(p_\ell^{(k)}, e_1)|^2$$

that approximates the distribution function

$$\omega(\lambda) \quad \text{with nodes} \quad \sigma_j^2 \quad \text{and weights} \quad \omega_j = |(s_1, u_j)|^2 \,,$$

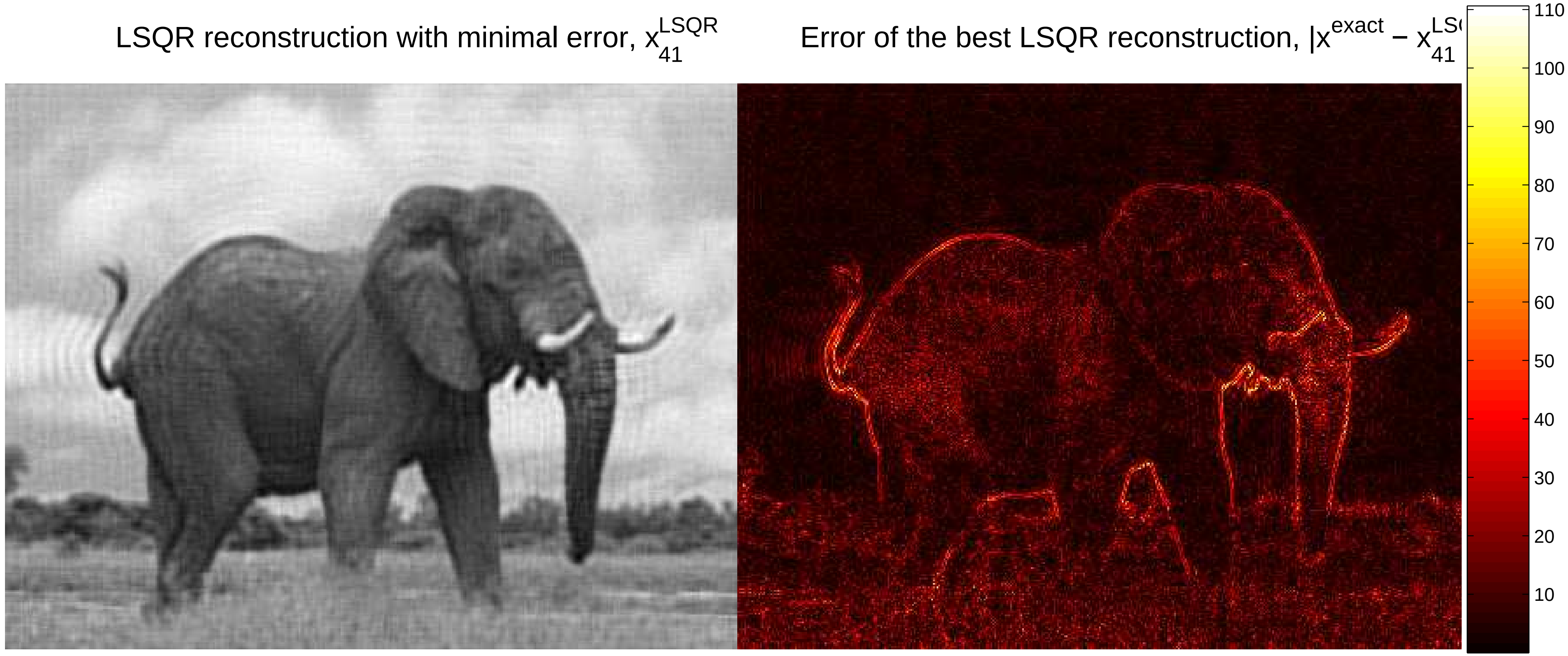
where σ_j^2, u_j are the eigenpairs of $A\,A^T$, for $j = n, \ldots, 1$, and $\theta_\ell^{(k)}$, $p_\ell^{(k)}$ are the eigenpairs of $L_k L_k^*$, $k = 1, \ldots,$, $\ell = 1, \ldots, k$.

Discrete ill-posed problem, the smallest node and weight:



Five tons large real world ill-posed problem:

4 : Reconstructed elephant and its error



Outline

1. Matching moments model reduction
2. CG and Gauss-Christoffel quadrature
3. Non-Hermitian generalizations
4. Noise level in discrete ill-posed problems
5. Conclusions

5 : Making links is good

- It is good to link the projection-based and algorithmic views of Krylov subspace methods with the model reduction and matching moments view, Gauss-Christoffel quadrature, etc.

- Rounding error analysis of Krylov subspace methods has had unexpected side effects in understanding of general mathematical phenomena independent of any numerical stability issues such as the sensitivity of the Gauss-Christoffel quadrature.

- Analysis of Krylov subspace methods for solving linear problems has to deal with highly nonlinear finite dimensional phenomena. See moments.

- Knowledge of the problem, like the properties of the distribution function corresponding to the discrete ill-posed problem, can suggest the right link.

5 : Continued fractions in mathematics and .....

- Euclid, ... , Brouncker and Wallis (1655-56): Three term recurences (for numbers)
- Euler (1748), ... , Brezinski (1991), Khrushchev (2008)
- Gauss (1814), Jacobi (1826)
- Chebyshev (1855, 1859), Markov (1884), Stieltjes (1884, 1893-94): Orthogonal polynomials
- Hilbert (190x), Von Neumann (1927, 1932), Wintner (1929)
- Krylov (1932), Lanczos (1950), Hestenes and Stiefel (1952), Vorobyev (1958, 1965), Golub and Welsh (1968), ...
- Gordon (1968), Schlesinger and Schwartz (1966), Reinhard (1979), ...
- Paige (1971), Reid (1971), Greenbaum (1989)
- Gragg (1974), Kalman (1979), Gallivan, Grimme, Van Dooren (1994), Butlheel, Van Barel (1997) $\longrightarrow$ Chebyshev, Markov, Stieltjes !

Thanks

Thank you for your kind patience!