Lanczos tridiagonalization, Krylov subspaces and the problem of moments

Zdeněk Strakoš

Institute of Computer Science AS CR, Prague

http://www.cs.cas.cz/˜strakos

Numerical Linear Algebra in Signals and Systems,

Monopoli (Bari), September 2006.

Lanczos tridiagonalization (1950, 1952)

$A \in \mathbb{R}^{N,N}$, large and sparse, symmetric, $w_1 \ (\equiv r_0/\|r_0\|, \ r_0 \equiv b - Ax_0)$,

$$AW_k \ = \ W_k T_k \ + \ \delta_{k+1} w_{k+1} e_k^T, \quad W_k^T W_k = I, \ W_k^T w_{k+1} = 0, \ k = 1, 2, \ldots \ ,$$

$$T_k \equiv \begin{pmatrix} \gamma_1 & \delta_2 & & \\ \delta_2 & \gamma_2 & \ddots & \\ & \ddots & \ddots & \delta_k \\ & & \delta_k & \gamma_k \end{pmatrix}, \quad \delta_l > 0 \, .$$

Stewart (1991): Lanczos and Linear Systems

Golub - Kahan bidiagonalization (1965), SVD

$B \in \mathbb{R}^{M,N}$, with no loss of generality $M \geq N, x_0 = 0;\ v_0 \equiv 0, u_1 \equiv b/\|b\|$,

$$B^T U_k \ = \ V_k \, L_k^T \, , \quad B V_k \ = \ [U_k, u_{k+1}] \, L_{k+} \, , \quad k = 1, 2, \ldots \, ,$$

$$L_k \equiv \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \alpha_k \end{pmatrix} , \quad L_{k+} \equiv \begin{pmatrix} L_k \\ \beta_{k+1} e_k^T \end{pmatrix} ,$$

$$U_k^T U_k = V_k^T V_k = I, \ U_k^T u_{k+1} = V_k^T v_{k+1} = 0.$$

Paige (1974), Paige and Saunders (1982), Björck (1988, 2005)

Relationship I

The Lanczos tridiagonalization applied to the augmented matrix

$$A \equiv \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$$

with the starting vector $w_1 \equiv (u_1, 0)^T$ yields in $2k$ steps the orthogonal matrix

$$W_{2k} = \begin{pmatrix} u_1 & 0 & \dots & u_k & 0 \\ 0 & v_1 & \dots & 0 & v_k \end{pmatrix}$$

and the Jacobi matrix T_{2k} with the zero main diagonal and the subdiagonals equal to $(\alpha_1, \beta_2, \dots, \alpha_{k-1}, \beta_k, \alpha_k)$.

Relationship II

$$BB^T U_k = U_k L_k L_k^T + \alpha_k \beta_{k+1} u_{k+1} e_k^T ,$$

$$L_k L_k^T = \begin{pmatrix} \alpha_1^2 & \alpha_1\beta_2 & & \\ \alpha_1\beta_2 & \alpha_2^2+\beta_2^2 & \ddots & \\ & \ddots & \ddots & \alpha_{k-1}\beta_k \\ & & \alpha_{k-1}\beta_k & \alpha_k^2+\beta_k^2 \end{pmatrix} ,$$

which represents k steps of the Lanczos tridiagonalization of the matrix BB^T with the starting vector $u_1 \equiv b/\beta_1 = b/\|b\|$.

Relationship III

$$B^T B \, V_k \; = \; V_k \, L_{k+}^T L_{k+} \; + \; \alpha_{k+1} \beta_{k+1} \, v_{k+1} e_k^T \; ,$$

$$L_{k+}^T L_{k+} = L_k^T L_k + \beta_{k+1}^2 e_k e_k^T = \begin{pmatrix} \alpha_1^2 + \beta_2^2 & \alpha_2 \beta_2 & & \\ \alpha_2 \beta_2 & \alpha_2^2 + \beta_3^2 & \ddots & \\ & \ddots & \ddots & \alpha_k \beta_k \\ & & \alpha_k \beta_k & \alpha_k^2 + \beta_{k+1}^2 \end{pmatrix} ,$$

which represents k steps of the Lanczos tridiagonalization of the matrix $B^T B$ with the starting vector $v_1 \equiv B^T u_1 / \alpha_1 = B^T b / \| B^T b \|$.

Large scale computational motivation

- Approximation of the spectral decomposition of A, of the SVD of A,
- Approximation of the solution of (possibly ill-posed) $Ax \approx b$.

The underlying principle: Model reduction by projection onto
Krylov subspaces.

A. N. Krylov, *On the numerical solution of the equations by which the frequency of small oscillations is determined in technical problems* (1931 R.),

but the story goes back to Gauss (1777-1855), Jacobi (1804-1851), Chebyshev (1821-1894), Christoffel (1829-1900), Stieltjes (1856-1894), Markov (1856-1922) and to many others not mentioned here.

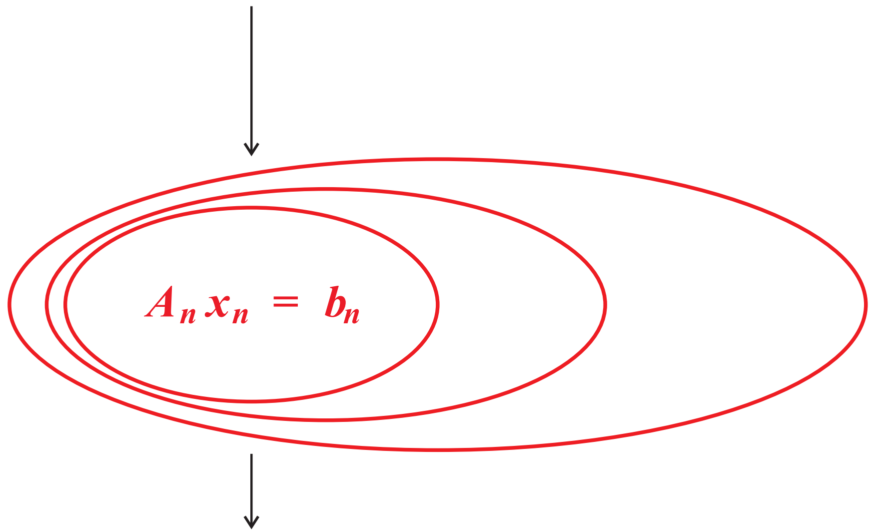
Outline

1. Krylov subspace methods
2. Stieltjes moment problem
3. Vorobyev moment problem
4. Lanczos, CG and the Gauss-Christoffel quadrature
5. Concluding remarks

Krylov subspace methods

Projections on nested subspaces

$$A\,x\ =\ b$$



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

Projection processes

$$x_n \in x_0 + \mathcal{S}_n, \quad r_0 \equiv b - Ax_0$$

where the constraints needed to determine x_n are given by

$$r_n \equiv b - Ax_n \in r_0 + A\mathcal{S}_n, \quad r_n \perp \mathcal{C}_n .$$

Here $\mathcal{S}_n$ is the search space, $\mathcal{C}_n$ is the constraint space.

Note that r_0 is decomposed to r_n + the part in $A\mathcal{S}_n$.

The projection should be called orthogonal if $\mathcal{C}_n = A\mathcal{S}_n$,
and it should be called oblique otherwise.

Krylov subspace methods

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

Krylov subspaces accumulate the dominant information of A with respect to r_0. Unlike in the power method for computing the dominant eigenspace, here all the information accumulated along the way is used

Parlett (1980), Example 12.1.1.

The idea of projections using Krylov subspaces is in a fundamental way linked with the problem of moments.

Stieltjes moment problem

Scalar moment problem

a sequence of numbers ξ_j, $j = 0, 1, \ldots$ is given and a non-decreasing distribution function $\omega(\lambda)$, $\lambda \geq 0$ is sought such that the Riemann-Stieltjes integrals defining the moments satisfy

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

Szegö (1939), Akhiezer and Krein (1938 R., 1962 E.), Shohat and Tamarkin (1943), Gantmakher and Krein (1941 R. 1st. ed., 1950 R. 2nd. ed., 2002 E. based on the 1st. ed., *Oscillation matrices and kernels and small vibrations of mechanical systems*), Karlin, Shapley (1953), Akhiezer (1961 R., 1965 E.), Davis and Rabinowitz (1984)

An interesting historical source: Wintner, *Spektraltheorie der unendlichen Matritzen - Enführung in den Analytischen Apparat der Quantenmechanik,* (1929), thanks to Michele Benzi!

The origin in

C. F. Gauss, *Methodus nova integralium valores per approximationem inveniendi,* (1814)

C. G. J. Jacobi, *Uber Gauss' neue Methode, die Werthe der Integrale näherungsweise zu finden,* (1826)

A useful algebraic formulation:

Vorobyev moment problem

Vector moment problem (using Krylov subspaces)

Given A, r_0, find a linear operator $\mathcal{A}_n$ on $\mathcal{K}_n$ such that

$$\begin{aligned}
\mathcal{A}_n r_0 &= A r_0 \,, \\
\mathcal{A}_n (A r_0) &= A^2 r_0 \,, \\
&\vdots \\
\mathcal{A}_n (A^{n-2} r_0) &= A^{n-1} r_0 \,, \\
\mathcal{A}_n (A^{n-1} r_0) &= \mathcal{Q}_n (A^n r_0) \,,
\end{aligned}$$

where $\mathcal{Q}_n$ projects onto $\mathcal{K}_n$ orthogonally to $\mathcal{C}_n$.

Vorobyev (1958 R., 1965 E.), Brezinski (1997), Liesen and S (200?)

in the Stieltjes formulation: S(PD) case

Given the first $2n-1$ moments for the distribution function $\omega(\lambda)$, find the distribution function $\omega^{(n)}(\lambda)$ with n points of increase which matches the given moments.

Vorobyev (1958 R.), Chapter III, with references to Lanczos (1950, 1952), Hestenes and Stiefel (1952), Ljusternik (1956 R., *Solution of problems in linear algebra by the method of continued fractions*)

Though the founders were well aware of the relationship (Stiefel (1958), Rutishauser (1954, 1959),) the computational potential of the CG approach has not been by mathematicians fully realized, cf. Golub and O'Leary (1989), Saulyev (1960 R., 1964 E.) - thanks to Michele Benzi, Trefethen (2000).

Golub has emphasized the importance of moments for his whole life.

Conclusions 1, based on moments

- Information contained in the data is not processed linearly in projections using Krylov subspace methods, including Lanczos tridiagonalization and Golub-Kahan bidiagonalization,

$$T_n = W_n^T(A)\, A\, W_n(A)\,.$$

- Any linearization in description of behavior of such methods is of limited use, and it should be carefully justified.

- In order to understand the methods, it is very useful (even necessary) to combine tools from algebra and analysis.

Lanczos, CG and the Gauss-Christoffel quadrature

Lanczos, CG and orthogonal polynomials

$$AW_n = W_n T_n + \delta_{k+1} w_{k+1} e_k^T \,, \quad A \text{ SPD}$$

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

Vectors in Krylov subspaces can be viewed as matrix polynomials applied to the initial residuals. Spectral decompositions of A and T_n with projections of w_1 resp. e_1 onto invariant subspaces corresponding to individual eigenvalues lead to the scalar products in the spaces of polynomials expressed via the Riemann-Stieltjes integrals, and to the world of orthogonal polynomials, Jacobi matrices, continued fractions, Gauss-Christoffel quadrature ...

Lanczos represents matrix formulation of the Stieltjes algorithm for computing orthogonal polynomials. This fact is widely known, but its benefits are not always used in the orthogonal polynomial literature. Numerical stability analysis of the Lanczos recurrences due to Paige, Parlett, Scott, Simon, Greenbaum, Grcar, Meurant, S, Notay, Druskin, Knizhnermann, Zemke, Wülling and others is not used at all.

CG: matrix formulation of the Gauss Quadrature

$$Ax = b\,,\ x_0 \quad \longrightarrow \quad \int_\zeta^\xi f(\lambda)\, d\omega(\lambda)$$

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

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

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

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

Vast literature on the subject

Hestenes and Stiefel (1952), Golub and Welsch (1969), Dahlquist, Eisenstat and Golub (1972), Dahlquist, Golub and Nash (1978), Kautsky and Elhay (1982), Kautsky and Golub (1983), Greenbaum (1989), Golub and Meurant (1994, 1997), Golub and B. Fischer (1994), Golub and S (1994), B. Fischer and Freund (1994), B. Fischer (1996), Gutknecht (1997), Brezinski (1997), Calvetti, Morigi, Reichel and Sgallari (2000) ...

From the side of computational theory of orthogonal polynomials, see the encyclopedic work of Gautschi (1968, . . . ,1981, . . . , 2005, 2006, . . .).

Many related subjects as construction of orthogonal polynomials from modified moments, sensitivity of the map from moments to the quadrature nodes and weights, reconstruction of Jacobi matrices from the spectral data and sensitivity of this problem, sensitivity and computation of the spectral decomposition of Jacobi matrices, ...

Lines of development sometimes parallel, independent and with relationships unnoticed.

Literature (continuation)

Gautschi (1968, 1970, 1978, 1982, 2004), Nevai (1979), H. J. Fischer (1998), Elhay, Golub, Kautsky (1991, 1992), Beckermann and Bourreau (1998), Laurie (1999, 2001),

Gelfand and Levitan (1951), Burridge (1980), Natterer (1989), Xu (1993), Druskin, Borcea and Knizhnermann (2005), Carpraux, Godunov and Kuznetsov (1996), Kuznetsov (1997), Paige and van Dooren (1999);

Stieltjes (1884), de Boor and Golub (1978), Gautschi (1982, 1983, 2004, 2005), Gragg and Harrod (1984), Boley and Golub (1987), Reichel (1991), H. J. Fischer (1998), Rutishauser (1957, 1963, 1990), Fernando and Parlett (1994), Parlett (1995), Parlett and Dhillon (97), Laurie (99, 01);

Wilkinson (1965), Kahan (19??), Demmel and Kahan (1990), Demmel, Gu, Eisenstat, Slapničar, Veselič and Drmač (1999), Dhillon (1997), Li (1997), Parlett and Dhillon (2000), Laurie (2000), Dhillon and Parlett (2003, 2004), Dopico, Molera and Moro (2003), Grosser and Lang (2005), Willems, Lang and Vömel (2005).

Some summary in Meurant and S (2006), O'Leary, S and Tichy (2006).

Descriptions intentionally missing

I have resigned on including the description of the relationship with the Sturm-Liouville problem, inverse scattering problem and Gelfand-Levitan theory, as well as applications in sciences, in particular in quantum chemistry and quantum physics, engineering, statistics ...

No algorithmic developments with founding contributions of Concus, Golub, O'Leary, Axelsson, van der Vorst, Saad, Fletcher, Freund, Stoer, ...

GAMM–SIAM ALA Conference, Düsseldorf, July 2006: Golub, Meurant, Reichel, Gutknecht, Bunse-Gerstner, S, ...

Vast signal & control related literature ...

An example - sensitivity of Lanczos recurrences

$A \in \mathbb{R}^{N,N}$ diagonal SPD,

$$A,\, w_1 \quad \longrightarrow \quad T_n \quad \longrightarrow \quad T_N = W_N^T A W_N$$

$$A + E,\, w_1 + e \quad \longrightarrow \quad \tilde{T}_n \quad \longrightarrow \quad \tilde{T}_N = \tilde{W}_N^T \,(A + E)\, \tilde{W}_N$$

$\tilde{T}_n$ is, under some assumptions on the size of the perturbations relative to the separation of the eigenvalues of A, close to T_n.

$\tilde{T}_N$ has all its eigenvalues close to that of A.

A particular larger problem

$\hat{A} \in \mathbb{R}^{2N,2N}$ diagonal SPD, $\hat{w}_1 \in \mathbb{R}^{2N}$, obtained by replacing each eigenvalue of A by a pair of very close eigenvalues of $\hat{A}$ sharing the weight of the original eigenvalue. In terms of the distribution functions, $\hat{\omega}(\lambda)$ has doubled points of increase but it is very close to $\omega(\lambda)$.

$$\hat{A}, \hat{w}_1 \quad \longrightarrow \quad \hat{T}_n \quad \longrightarrow \quad \hat{T}_{2N} = \hat{W}_{2N}^T \hat{A} \hat{W}_{2N}$$

$\hat{T}_{2N}$ has all its eigenvalues close to that of A.
However, $\hat{T}_n$ can be very different from T_n.

Relationship to the mathematical model of finite precisision computation, see Greenbaum (1989), S (1991), Greenbaum and S (1992), (in some sense also Parlett (1990)), Meurant and S (2006).

Here, however, all is determined exactly!

CG and Gauss quadrature relationship

$$Ax = b\,,\; x_0 \qquad \longrightarrow \qquad \int_\zeta^\xi f(\lambda)\, d\omega(\lambda)$$

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

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

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

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

CG and Gauss quadrature errors

At any iteration step n, CG represents the matrix formulation of the n-point Gauss quadrature of the Riemann-Stieltjes integral determined by A and r_0,

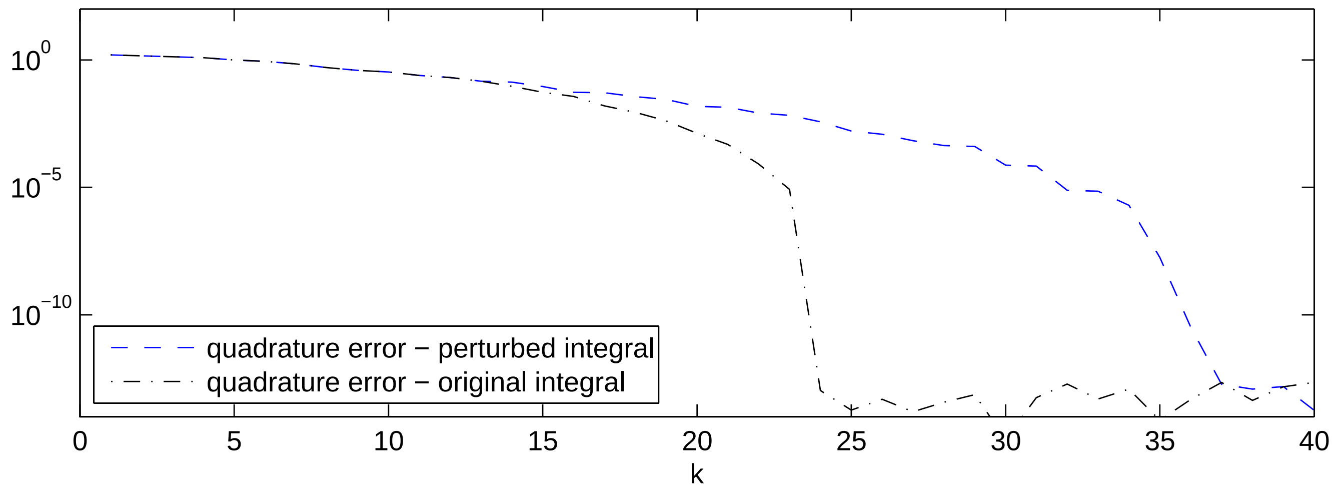
$$\int_{\zeta}^{\xi} f(\lambda)\, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} f(\theta_i^{(n)}) \;+\; R_n(f)\,.$$

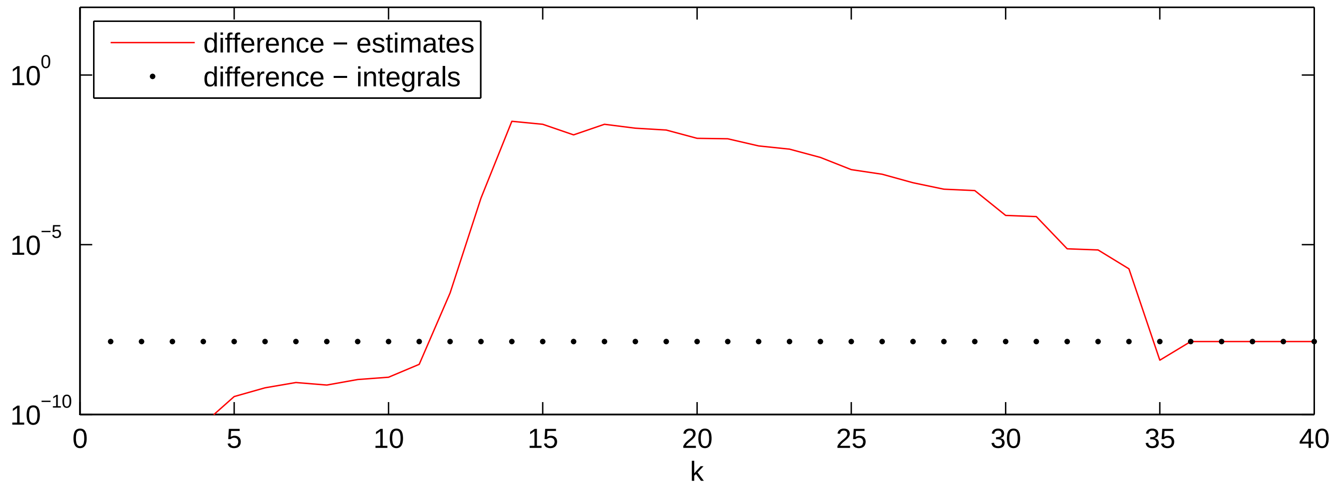
For $f(\lambda) \equiv \lambda^{-1}$ the formula takes the form

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

This was a base for CG error estimation in
[DaGoNa-78, GoFi-93, GoMe-94, GoSt-94, GoMe-97, . . .] .

Results for A, w_1 and $\hat{A}, \hat{w}_1$:





A contradiction to published results

Kratzer, Parter and Steuerwalt, *Block splittings for the conjugate gradient method*, Computers and Fluids 11, (1983), pp. 255-279. The statement on p. 261, second paragraph, in our notation (falsely) means:

The convergence of CG for A, w_1 and $\hat{A}, \hat{w}_1$ ought to be similar; at least $\|\hat{x} - \hat{x}_N\|_{\hat{A}}$ should be small.

The argument in the paper is based on relating the CG minimizing polynomial to the minimal polynomial of A. It has been underestimated, however, that for some distribution of eigenvalues of A its minimal polynomial (normalized to one at zero) can have extremely large gradients and therefore it can be very large at points even very close to its roots. That happens for the points equal to the eigenvalues of $\hat{A}$!

Remarkable related papers O'Leary, Stewart and Vandergraft (1979), Parlett, Simon and Stringer (1982), van der Sluis, van der Vorst (1986, 1987).

Conclusions 2, based on the rich matter

- It is good to look for interdisciplinary links and for different lines of thought. An overemphasized specialization together with malign deformation of the *publish or perish* policy is counterproductive. It leads to vasting of energy and to a dissipative loss of information.

- Rounding error analysis of iterative methods is not a (perhaps useful but obscure) discipline for a few strangers. It has an impact not restricted to development of methods and algorithms. Through its wide methodology and questions it can lead to understanding of general mathematical phenomena independent of any numerical issues.

Concluding remarks

- Krylov subspace methods provide a highly nonlinear model reduction.
- Their success or failure is determined by the properties of the underlying moment problems.
- Rounding error analysis should always be a part of real world computations.

Thank you!