# Moments, Krylov subspace methods and model reduction

Zdeněk Strakoš

Academy of Sciences and Charles University, Prague

http://www.cs.cas.cz/˜strakos

8th GAMM ALA Workshop, TU Hamburg-Harburg, September 2008.

# Thanks to the collaborators

Gene H. Golub
Anne Greenbaum
Gerard Meurant
Dianne P. O’Leary
Chris Paige
Petr Tichý
Jörg Liesen

# Historical remark on iterative methods

1950 - Iterative methods for elliptic PDE
- Ph.D. Thesis by D. Young at Harvard (published in 1954)

1951, 1952 - Lanczos algorithm, conjugate gradient method
by C. Lanczos, M. Hestenes and E. Stiefel

1962 - Book Matrix Iterative Analysis by R. Varga

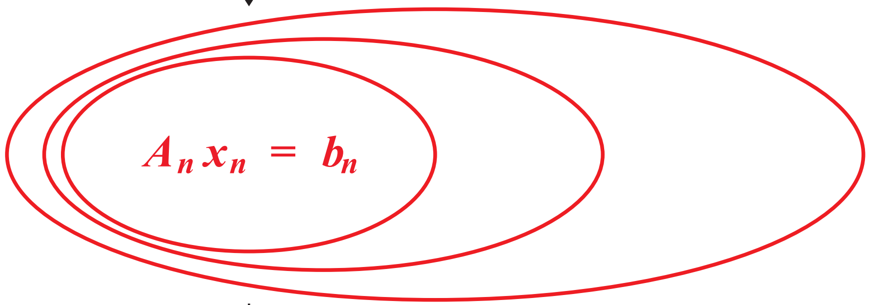
1971 - Book Iterative methods by D. Young

1971 - Lecture of J. Reid in Dundee (published in 1971)

1971 - Ph.D. Thesis of C.C. Paige at the University of London
(published in 1972, 1976 and 1980)

# Projections onto Krylov subspaces

$$A\,x \;=\; b, \quad A \in \mathbb{C}^{N \times N}, \quad r_0 = b - Ax_0$$




Here $x_n$ approximates the solution $x$ using the projection onto low dimensional subspaces

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

.

# Nonlinearity and moments

The projection process using Krylov subspaces is highly nonlinear in $A$ and it depends on $r_0$ ,

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

$\mathcal{K}_n(A, r_0)$ accumulate the dominant information of $A$ with respect to $r_0$ .

Unlike in the power method for computing the single dominant eigenspace, here all the information accumulated along the way is used, see Parlett (1980), Example 12.1.1.

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

**The story goes back to Gauss (1814).**

# Outline

1. Krylov subspace methods as  matching moments model reduction

2. Convergence of CG in the presence of close eigenvalues

3. Gauss-Christoffel quadrature can be sensitive to small perturbations of the distribution function

4. CG in finite precision arithmetic

# 1 : Matching moments

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

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

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

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

# 1 : Gauss-Christoffel quadrature

Clearly,

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

represents the $n$-point Gauss-Christoffel quadrature, see

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

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

and the description given in H. H. J. Goldstine, *A History of Numerical Analysis from the 16th through the 19th Century,* (1977).

With no loss of generality we assume $\xi_0 = 1$ .

Gauss-Christoffel quadrature formulation:

$$\int_0^\infty f(\lambda)\, d\omega(\lambda) \;\approx\; \sum_{i=1}^{n} \omega_i^{(n)} f(\lambda_i^{(n)})\,,$$

where the reduced model given by the distribution function with $n$ points of increase $\omega^{(n)}$ matches the first $2n$ moments

$$\int_0^\infty \lambda^k\, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} (\lambda_i^{(n)})^k\,, \qquad k = 0, 1, \ldots, 2n-1\,.$$

# 1 : Stieltjes recurrence

Let $p_1(\lambda) \equiv 1, p_2(\lambda), \dots, p_{n+1}(\lambda)$ be the first $n+1$ orthonormal polynomials corresponding to the distribution function $\omega(\lambda)$ .

Then, writing $P_n(\lambda) = (p_1(\lambda), \dots, p_n(\lambda))^T$ ,

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

represents the Stieltjes recurrence (1883-4), with the Jacobi matrix

$$T_n \equiv \begin{pmatrix} \gamma_1 & \delta_2 & & \\ \delta_2 & \gamma_2 & \ddots & \\ & \ddots & \ddots & \delta_n \\ & & \delta_n & \gamma_n \end{pmatrix}, \qquad \delta_l > 0 \, .$$

# 1 : Matrix computation: Lanczos $\equiv$ Stieltjes

In matrix computations, $T_n$ results from the Lanczos process (1951) applied to $T_n$ starting with $e_1$. Therefore $p_1(\lambda) \equiv 1,\, p_2(\lambda),\, \dots,\, p_n(\lambda)$ are orthonormal with respect to the inner product

$$(p_s, p_t) \;\equiv\; \sum_{i=1}^{n} |(z_i^{(n)}, e_1)|^2 \, p_s(\theta_i^{(n)}) \, p_t(\theta_i^{(n)}) \,,$$

where $z_i^{(n)}$ is the orthonormal eigenvector of $T_n$ corresponding to the eigenvalue $\theta_i^{(n)}$, and $p_{n+1}(\lambda)$ has the roots $\theta_i^{(n)},\, i = 1, \dots, n$.

Consequently,

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

Golub and Welsh (1969), ... ... ... ,
Meurant and S, Acta Numerica (2006).

# 1 : Linear algebraic equation

Given $Ax = b$ with an HPD $A \in \mathbb{C}^{N\times N}$, $r_0 = b - Ax_0$, $w_1 = r_0/\|r_0\|$ .
Assume, for simplicity of notation, $\dim(\mathcal{K}_n(A, r_0)) = n$ .

Consider the spectral decomposition

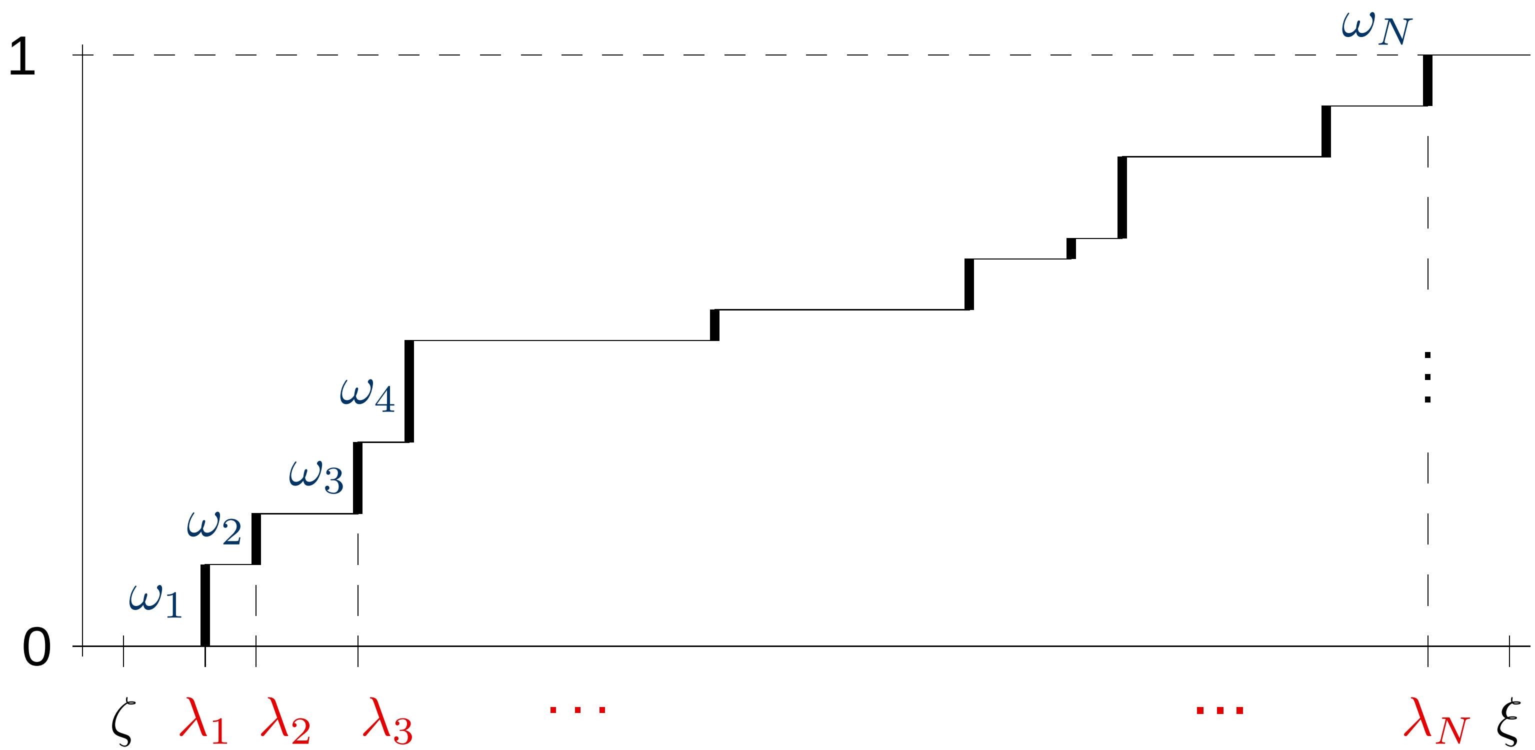
$$A = S \,\text{diag}(\lambda_i)\, S^* ,$$

where for clarity of exposition we assume that the eigenvalues are distinct,

$$0 < \lambda_1 < \ldots < \lambda_N , \quad S = [s_1, \ldots , s_N] .$$

$A$ and $w_1(b, x_0)$ determine the distribution function $\omega(\lambda)$ with
$N$ points of increase $\lambda_i$ and weights $\omega_i = |(s_i, w_1)|^2 , \; i = 1, \ldots, N$ .

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



Matrix formulation:

$$\int_0^\infty \lambda^k \, d\omega(\lambda) \;=\; \sum_{i=1}^{N} \omega_j \, (\lambda_j)^k \;=\; w_1^* \, A^k \, w_1 \, ,$$

$$\sum_{i=1}^{n} \omega_i^{(n)} \, (\lambda_i^{(n)})^k \;=\; \sum_{i=1}^{n} \omega_i^{(n)} \, (\theta_i^{(n)})^k \;=\; e_1^T \, T_n^k \, e_1 \, .$$

matching the first $2n$ moments therefore means

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

# 1 : Conjugate gradients (CG) for $Ax = b$

The $A$-norm of the error is minimal! See Elman, Silvester and Wathen (2005), p. 71.

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

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

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

and the CG approximation given by

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

# 1 : Alternative descriptions

- Stay with $A, b, r_0, w_1$ and work with the matrix formulation using the Lanczos process (CG) applied to $A$ with $w_1$ .

- Using the basis of eigenvectors $S$ , the matrix formulation reduces to the mathematically equivalent polynomial formulation, Lanczos (CG) reduces to the Stieltjes process applied to the distribution function $\omega(\lambda)$ .

In both descriptions the $n$-th step gives the Jacobi matrix $T_n$ and the distribution function $\omega_n(\lambda)$ .

The relationship was pointed out by Hestenes and Stiefel (1952), . . .
nice Ph.D. Thesis by Kent (1989, Stanford), book by B. Fischer (1996), paper by Fischer and Freund (1992).

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

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

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

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

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

# 1 : Matching moments model reduction

CG (Lanczos) reduces for $A$ HPD at the step $n$ the original model

$$A\,x \;=\; b\,,\; r_0 = b - Ax_0$$

to

$$T_n\,y_n \;=\; \|r_0\|\,e_1\,,$$

such that the the first $2n$ moments are matched,

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

# 1 : Comments on literature

Proofs of results related to moments or model reduction are in the literature typically based on factorizations of the matrix of moments, Golub and Welsh (1969), Dahlquist, Golub and Nash (1978), . . . , Kent(1989), . . . , which is also true for Antoulas (2005).

Moment matching techniques has been used for decades in computational physics and in computational chemistry, see Gordon (1968).

Gauss quadrature formulation related to the nonsymmetric Lanczos process and to the Arnoldi process was given by Freund and Hochbruck (1993), motivated by Fischer and Freund (1992). Gauss quadrature was formally extended to the complex plane by Saylor and Smolarski (2001), with motivation from inverse scattering problems in electromagnetics by Warnick (1997), . . . , Golub, Stoll and Wathen (2008).

Here we avoid using matrix of moments, and do not need any formal generalization of the Gauss quadrature formulas to the complex plane.

# 1 : Vorobyev moment problem - 1958, 1965

Find a linear HPD operator $\mathcal{A}_n$ on $\mathcal{K}_n(A, r_0)$ such that

$$
\begin{aligned}
\mathcal{A}_n \, w_1 &= A \, w_1 \,, \\
\mathcal{A}_n \, (A \, w_1) \equiv \mathcal{A}_n^2 \, w_1 &= A^2 w_1 \,, \\
&\vdots \\
\mathcal{A}_n \, (A^{n-2} w_1) \equiv \mathcal{A}_n^{n-1} \, w_1 &= A^{n-1} w_1 \,, \\
\mathcal{A}_n \, (A^{n-1} w_1) \equiv \mathcal{A}_n^{n} \, w_1 &= Q_n \, (A^n w_1) \,,
\end{aligned}
$$

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

By construction,

$$w_1^* \, A^k \, w_1 \; = \; w_1^* \, A_n^k \, w_1 \,, \qquad k = 0, \, \ldots \, , n - 1 \, .$$

Since $\mathcal{K}_n(A, w_1) \; = \;$ span$\{w_1, \, \ldots \, , A^{n-1} w_1\} \, ,$ the projection

$$Q_n \, (A^n \, w_1) \; - \; A_n^n \, w_1 \; = \; Q_n \, (A^n \, w_1 - A_n^n \, w_1) \; = \; 0$$

gives (note that $A$ is Hermitian)

$$w_1^* \, A^k \, w_1 \; = \; w_1^* \, A_n^k \, w_1 \,, \qquad k = 0, 1, \, \ldots \, , 2n - 1 \, .$$

# 1 : Matching moments model reduction

Using the unitary basis $W_n$ with $Q_n = W_n W_n^*$ ,

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

$$A_n^k \quad = \quad W_n \, T_n^k \, W_n^* \, ,$$

which gives the result

$$w_1^* \, A^k \, w_1 \; = \; w_1^* \, A_n^k \, w_1 = \; e_1^T \, T_n^k \, e_1 \, , \qquad k = 0, 1, \, \ldots \, , 2n - 1 \, .$$

# 1 : Non-Hermitian Lanczos

Given a nonsingular $A \in \mathbb{C}^{N \times N}$ , $v \in \mathbb{C}^N$ , $w \in \mathbb{C}^N$ , $v^* w = 1$ .

The non-Hermitian Lanczos algorithm can be written in the form

$$\begin{aligned} A\, W_n &= W_n\, T_n + \delta_{n+1}\, w_{n+1}\, e_n^T , \\ A^*\, V_n &= V_n\, T_n^* + \beta_{n+1}^*\, v_{n+1}\, e_n^T , \end{aligned}$$

$$V_n^*\, W_n = I_n , \quad T_n = V_n^*(A, v_1, w_1)\, A\, W_n(A, v_1, w_1) .$$

We assume that the algorithm does not break down in steps $1$ through $n$ (it can break down later).

Here

$$T_n \equiv \begin{pmatrix} \gamma_1 & \beta_2 & & \\ \delta_2 & \gamma_2 & \ddots & \\ & \ddots & \ddots & \beta_n \\ & & \delta_n & \gamma_n \end{pmatrix}, \quad \delta_l > 0,\ \beta_l \neq 0\,,$$

The columns of $W_n$ form a basis of $\mathcal{K}_n(A, w_1)$ , while the columns of $V_n$ a basis of $\mathcal{K}_n(A^*, v_1)$ . Since $V_n^* W_n = I_n$ , the oblique projector onto $\mathcal{K}_n(A, w_1)$ orthogonal to $\mathcal{K}_n(A^*, v_1)$ can be written as

$$Q_n = W_n V_n^*\,.$$

# 1 : Vorobyev moment problem for N. L.

Find a linear operator $A_n$ on $\mathcal{K}_n(A, w_1)$ such that

$$\begin{aligned}
A_n\, w_1 &= A\, w_1\,, \\
A_n\, (A\, w_1) &= A^2\, w_1\,, \\
&\vdots \\
A_n\, (A^{n-2}\, w_1) &= A^{n-1}\, w_1\,, \\
\\
A_n\, (A^{n-1} w_1) &= (W_n\, V_n^*)\, (A^n w_1)\,.
\end{aligned}$$

Using ortogonality to the basis vectors $v_1, A^* v_1, \ldots, (A^*)^{n-1} v_1\,,$

$$v_1^*\, A^{k+n}\, w_1 \;=\; v_1^*\, A^k\, A_n^n\, w_1\,, \qquad k = 0, 1, \ldots, n-1\,.$$

$$\begin{aligned} A_n &= Q_n\, A\, Q_n \;=\; W_n V_n^*\, A\, W_n V_n^* \;=\; W_n\, T_n\, V_n^* \,, \\ A_n^n &= W_n\, T_n^n\, V_n^* \,, \end{aligned}$$

$$v_1^*\, A^k \;=\; e_1^T\, T_n^k\, V_n^* \,, \qquad k = 0, 1, \ldots, n-1 \,,$$

we finally get (using a simple multiplication argument for the first $n$ moments)

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

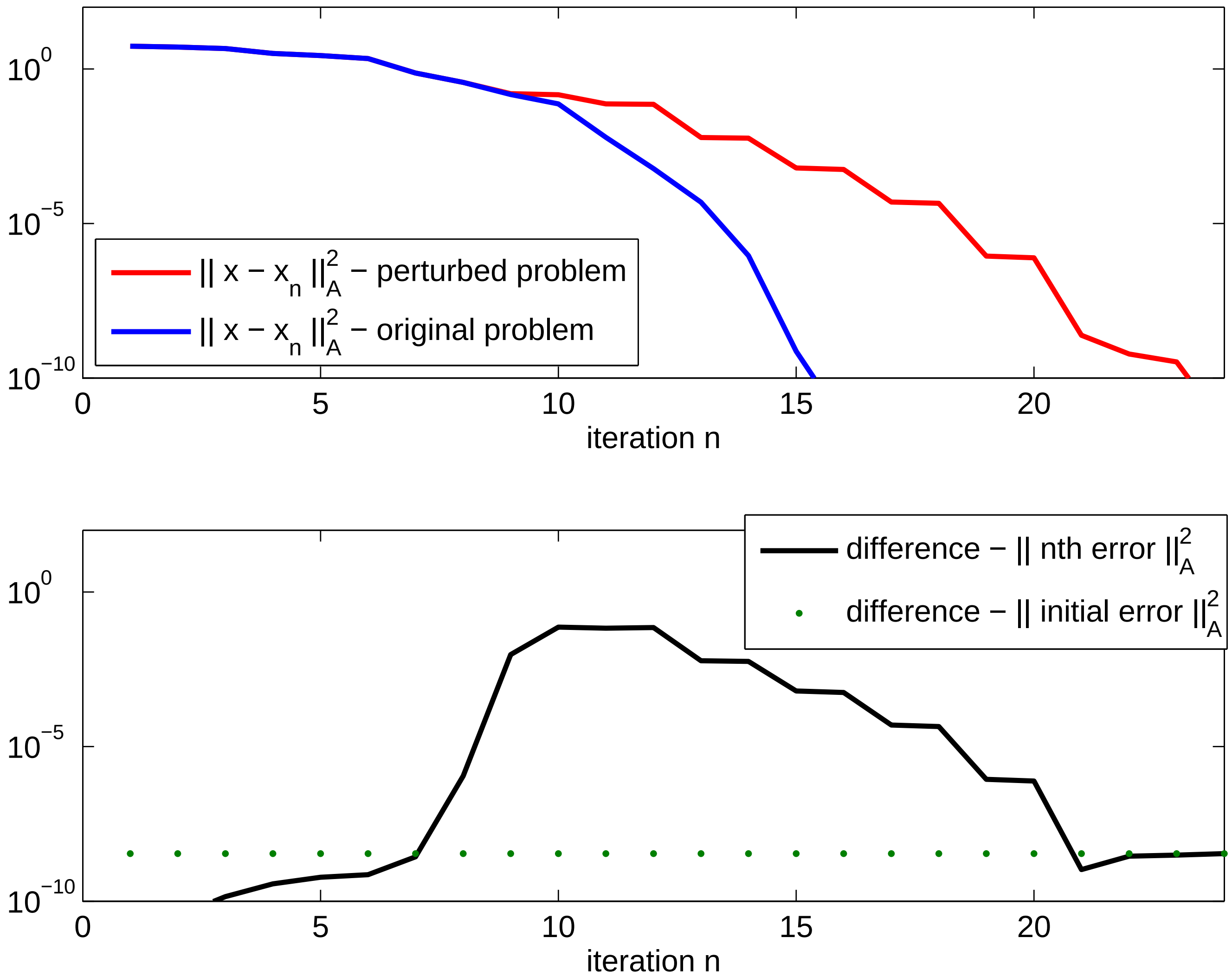
i.e., $n$ steps of the non-Hermitian Lanczos (or BiCG) represent the model reduction which matches the first $2n$ moments.

# Outline

1. Krylov subspace methods as matching moments model reduction
2. Convergence of CG in the presence of close eigenvalues
3. Gauss-Christoffel quadrature can be sensitive to small perturbations of the distribution function
4. CG in finite precision arithmetic

Exact arithmetic !

# 2 : Exact CG for $A, w_1$ and $\hat{A}, \hat{w}_1$ :



# 2 : Observations

- Replacing single eigenvalues by two close ones causes large delays. Clusters can not be replaced by single representatives! Matching moment property is responsible for the possibly large difference.

- The presence of close eigenvalues causes an irregular staircase-like behaviour.

- Local decrease of error says nothing about the total error.

- Stopping criteria must be based on the global information.

# 2 : Published explanations

The fact that the presence of close eigenvalues affects the convergence of Ritz values and therefore the rate of convergence of the conjugate gradient method is well known; see the beautiful explanation given by

van der Sluis and van der Vorst (1986, 1987).

It is closely related to the convergence of the Rayleigh quotient in the power method and to the so-called ‘misconvergence phenomenon’ in the Lanczos method, see

O’Leary, Stewart and Vandergraft (1979),
Parlett, Simon and Stringer (1982).

# Outline

1. Krylov subspace methods as the matching moments model reduction

2. Convergence of CG in the presence of close eigenvalues

3. Gauss-Christoffel quadrature can be sensitive to small perturbations of the distribution function

4. CG in finite precision arithmetic

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

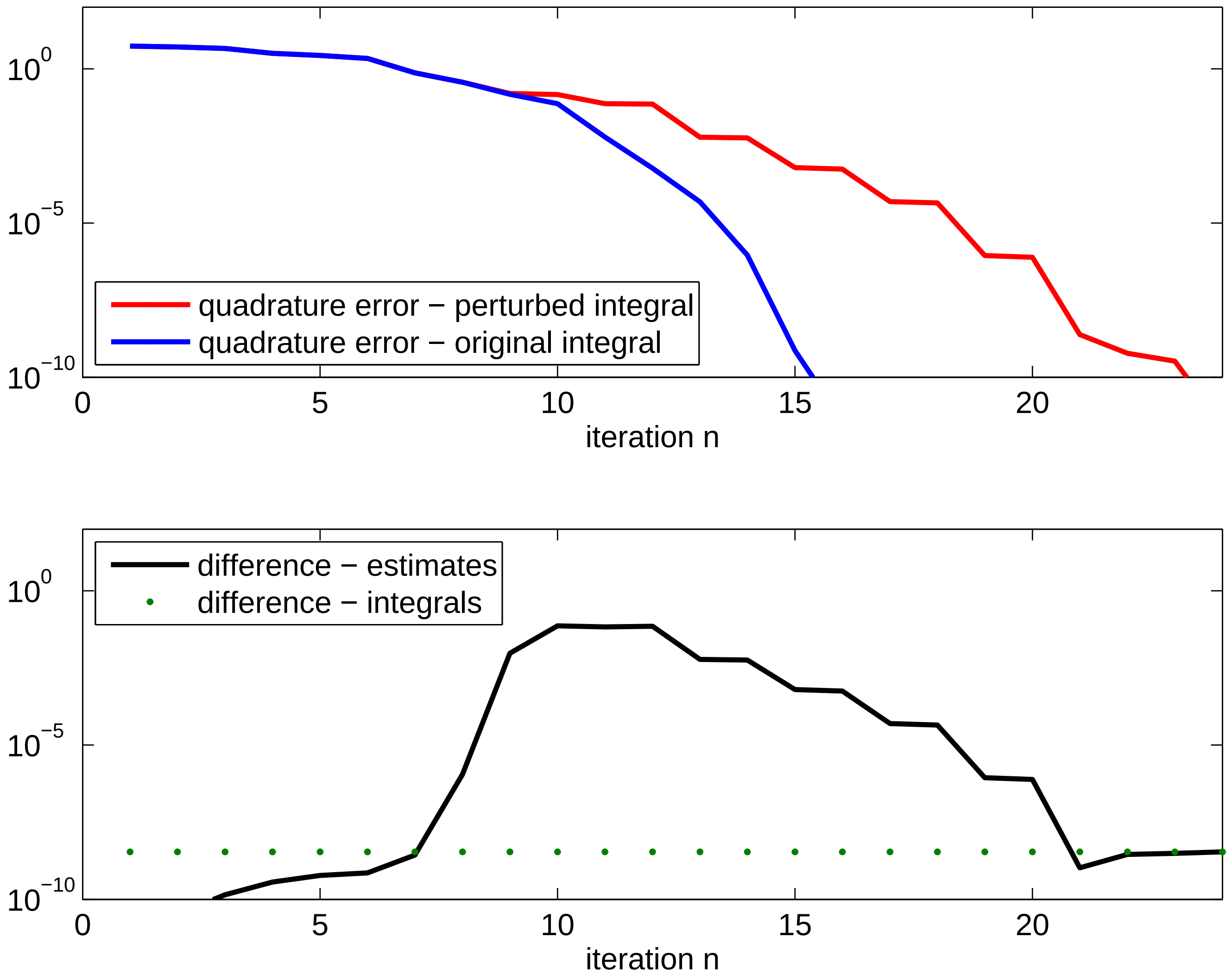
$$\int_\zeta^\xi f(\lambda)\, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} f(\theta_i^{(n)}) \;+\; R_n(f)\,.$$

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

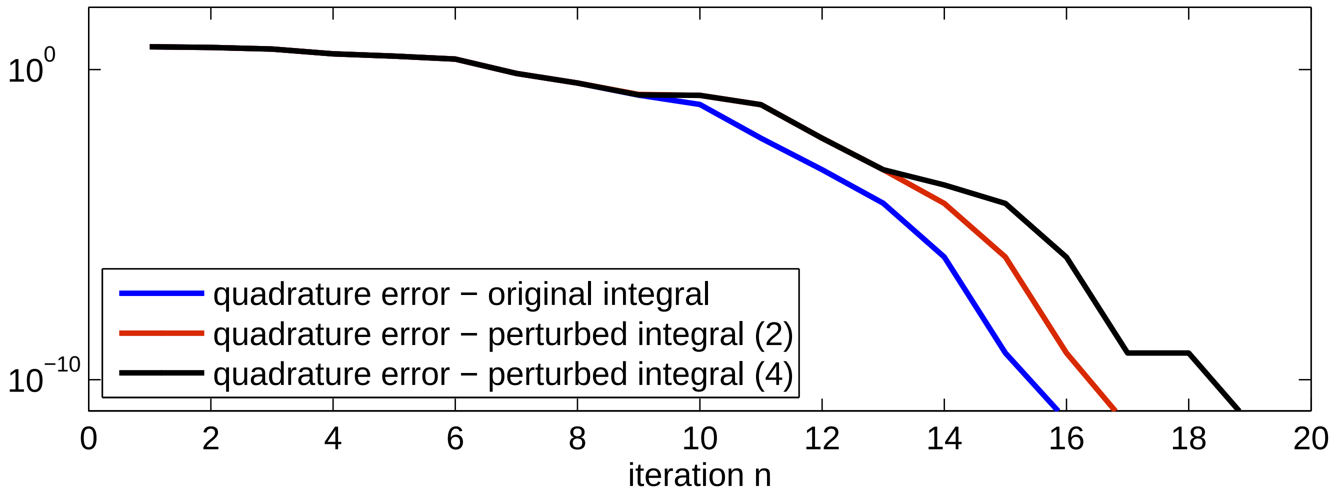
$$\frac{\|x - x_0\|_{\mathbf{A}}^2}{\|r_0\|^2} \;=\; n\text{-th Gauss quadrature} \;+\; \frac{\|x - x_n\|_{\mathbf{A}}^2}{\|r_0\|^2}\,.$$

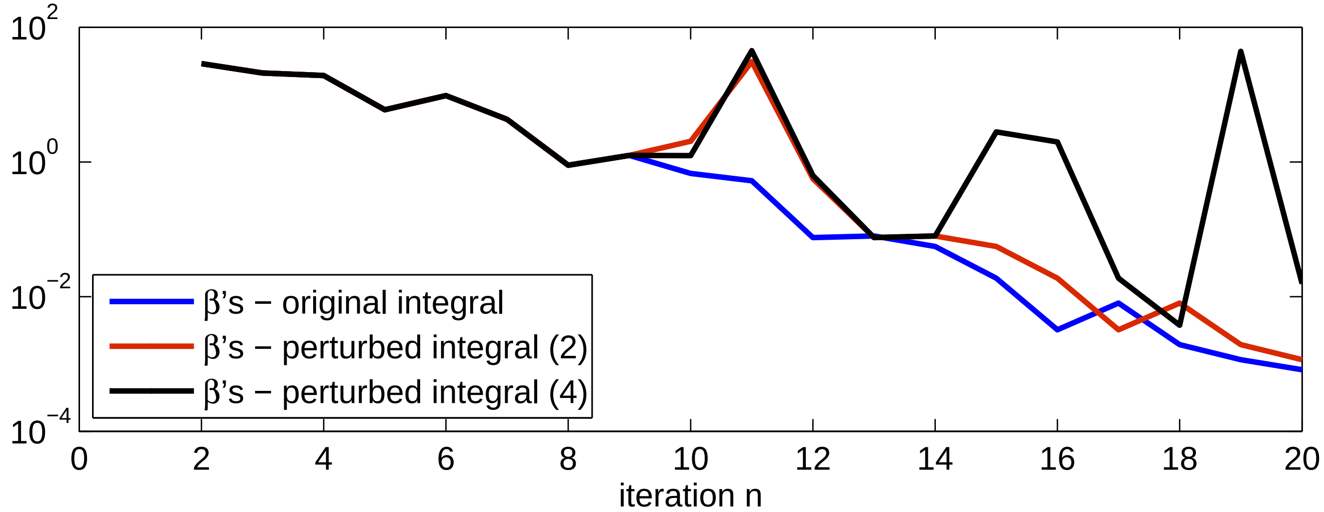
This was a base for the CG error estimation in
[DaGoNa-78, GoFi-93, GoMe-94, GoSt-94, GoMe-97, ... ]

# 3 : Sensitivity of the Gauss-Ch. Quadrature



# 3 : Simplified problem



# 3 : Sensitivity statement

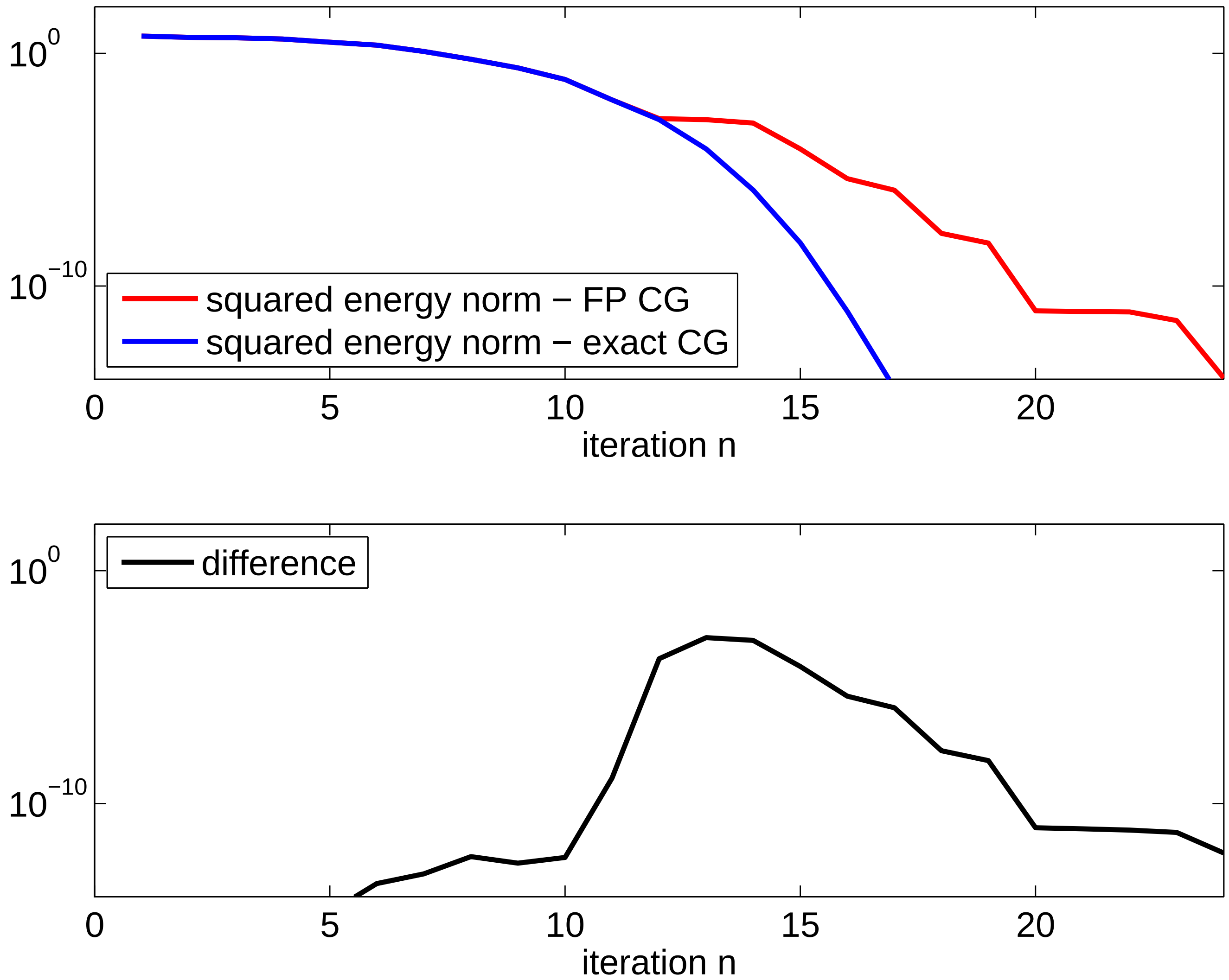
1. Gauss-Christoffel quadrature can be highly sensitive to small changes in the distribution function of the approximated integral.
   In particular, the difference between the corresponding quadrature approximations (using the same number of quadrature nodes) can be many orders of magnitude larger than 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 and with no singularity close to the interval of integration.

# Outline

1. Krylov subspace methods as matching moments model reduction
2. Convergence of CG in the presence of close eigenvalues
3. Gauss-Christoffel quadrature can be sensitive to small perturbations of the distribution function
4. CG in finite precision arithmetic

# 4 : Exact and FP CG applied to $A, w_1$



# 4 : Observations - FP CG

- Rounding errors can cause large delays.
- They may cause an irregular staircase-like behaviour.
- Local decrease of error says nothing about the total error.
- Stopping criteria must be based on global information.
- It must be justified by rigorous rounding error analysis.

Golub and S (1994),
S and Tichý (2002, 2005),

Comput. Methods Appl. Mech. Engrg. (2003).

## 4 : Close to the **exact** CG for $\hat{A}\,\hat{x} = \hat{b}$ ???

Mathematical model of finite precision Lanczos and CG computations, see

Paige (1971–80), Greenbaum (1989),
S (1991), Greenbaum and S (1992),
(also Parlett (1990)),

Recent review and update Meurant and S, Acta Numerica, (2006).

# Conclusions

- It is good to look for interdisciplinary links and for different lines of thought. Such as linking the Krylov subspace methods with model reduction and matching moments.

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

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

- Regularization effects of Krylov subspace methods viewed through matching moments model reduction?!

# Recent references


D. P. O'Leary, Z. S. and Petr Tichý, On sensitivity of Gauss-Christoffel quadrature, Numerische Mathematik 107, pp. 147–174, 2007,

Z. S., Model reduction using the Vorobyev moment problem, accepted for publication in the special issue of Numerical Algorithms, September 2008,

Z. S. and Petr Tichý, Estimation of $c^* A^{-1} b$ via matching moments, submitted to SISC, July 2008,

G. Golub, M. Stoll and A. Wathen, Approximation of the scattering amplitude and linear systems, submitted to ETNA, Dec. 2007.

Thank you!