Sensitivity of Gauss-Christoffel quadrature and sensitivity of Jacobi matrices to small changes of spectral data

Zdeněk Strakoš
Academy of Sciences and Charles University, Prague
http://www.cs.cas.cz/˜strakos

A joint work with

Diane P. O'Leary and Petr Tichý

Zeuthen, June 2008.

Jacobi matrices

$$T_N \equiv \begin{pmatrix} \gamma_1 & \delta_2 & & \\ \delta_2 & \gamma_2 & \ddots & \\ & \ddots & \ddots & \delta_N \\ & & \delta_N & \gamma_N \end{pmatrix}, \quad \delta_l > 0\,.$$

Fully determined by the spectrum and the top (bottom) entries of the normalized eigenvectors.

Are the entries sensitive to small changes of the spectral data?

Outline

1. Lanczos, CG and the Gauss-Christoffel quadrature
2. Reconstruction of Jacobi matrices from the spectral data
3. Another problem
4. Gauss-Christoffel quadrature can be sensitive to small perturbations of the distribution function
5. Lanczos and CG in finite precision arithmetic

1 : Lanczos tridiagonalization (1950, 1952)

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

$$AW_n = W_n T_n + \delta_{n+1} w_{n+1} e_n^T, \quad W_n^T W_n = I, \; W_n^T w_{n+1} = 0, \; n = 1, 2, \ldots ,$$

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

Consider, for clarity, that the process does not stop until $n = N$.

1 : Lanczos, CG and Gauss-Ch. quadrature

$$
\begin{array}{ccc}
Ax = b\,,\ x_0 & \longrightarrow & \displaystyle\int_{\zeta}^{\xi} f(\lambda)\, d\omega(\lambda) \\
\uparrow & & \uparrow \\
T_n\, y_n \ = \ \|r_0\|\, e_1 & \longleftrightarrow & \displaystyle\sum_{i=1}^{n} \omega_i^{(n)} f\left(\theta_i^{(n)}\right)
\end{array}
$$

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

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

1 : Literature (beginning)

Sensitivity of the map from the nodes and weights of the computed quadrature to the recurrence coefficients of the corresponding orthonormal polynomials:

Gautschi (1968, 1970, 1978, 1982, 2004),

Nevai (1979), H. J. Fischer (1998),
Elhay, Golub, Kautsky (1991, 1992),
Beckermann and Bourreau (1998),

Laurie (1999, 2001).

1 : Sensitivity of the Lanczos recurrences

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

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

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

Strongly nonlinear relationship $T_N(A, w_1) = W_N^T(A, w_1)\, A\, W_N(A, w_1)$.

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

Is $\tilde{T}_n$ for sufficiently small perturbations of A, w_1 close to T_n?

1 : Literature (continuation)

Sensitivity of the Lanczos recurrences:

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);

Here, however, sensitivity of Krylov subspaces has to be investigated as a part of the problem!

Outline

1. Lanczos, CG and the Gauss-Christoffel quadrature
2. Reconstruction of Jacobi matrices from the spectral data
3. Another problem
4. Gauss-Christoffel quadrature can be sensitive to small perturbations of the distribution function
5. Lanczos and CG in finite precision arithmetic

2 : Literature (continuation)

Computation of the inverse eigenvalue problem - reconstruction of T_N from the nodes and weights:

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);

2 : Literature (final part)

Computation of nodes (eigenvalues) and weights (squared top elements of the normalized eigenvectors) from T_n :

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 Tichý (2007).

2 : Accurate recovery of recursion coefficients

Laurie (99): A constructive proof of the following statement:

Given the weights and the $N-1$ positive differences between the consecutive nodes, the main diagonal entries of the corresponding Jacobi matrix (shifted by the smallest node) and the off-diagonal entries can be computed in $\frac{9}{2}N^2 + O(N)$ arithmetic operations, all of which can involve only addition, multiplication and division of positive numbers.

Consequently, in finite precision arithmetic they can be computed to a relative accuracy no worse than $\frac{9}{2}N^2\varepsilon + O(N\varepsilon)$, where ε denotes machine precision.

2 : Sensitivity result

Laurie (99, 01): This result bounds also the conditioning of the problem:

If the weights and the $N-1$ positive differences between the consecutive nodes are perturbed, with the size of the relative perturbations of the individual entries bounded by some small ϵ, then such perturbation can cause a relative change of the individual entries of the shifted main diagonal and of the individual off-diagonal entries of the Jacobi matrix not larger than $\frac{9}{2}N^2\epsilon + O(N\epsilon)$.

The resulting algorithm combines ideas from earlier works from approximation theory, orthogonal polynomials, and numerical linear algebra.

Outline

1. Lanczos, CG and the Gauss-Christoffel quadrature
2. Reconstruction of Jacobi matrices from the spectral data
3. Another problem
4. Gauss-Christoffel quadrature can be sensitive to small perturbations of the distribution function
5. Lanczos and CG in finite precision arithmetic

3 : An example – basic problem

$A \in \mathbb{R}^{N\times N}$ is diagonal positive definite (SPD),
S (91), Greenbaum, S (92),

$$\lambda_i = \lambda_1 + \frac{i-1}{n-1}(\lambda_n - \lambda_1)\,\gamma^{n-i}, \quad i = 2,\ldots,n-1,$$

In the experiment we take $\lambda_1 = 0.1$, $\lambda_n = 100$, $n = 24$, $\gamma = 0.55$.
Starting vector $w_1 \in \mathbb{R}^N$ has been generated randomly.
Lanczos process:

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

3 : A particular larger problem

$\hat{A} \in \mathbb{R}^{2N \times 2N}$ 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 $\omega(\lambda)$.

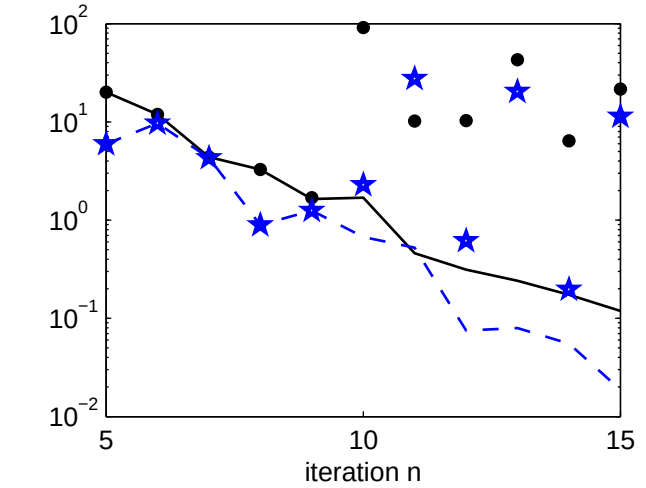
$$\hat{A}, \hat{w}_1 \quad \longrightarrow \quad \hat{T}_n \quad \longrightarrow \quad \hat{T}_{2N} = \hat{W}_{2N}^T \hat{A} \hat{W}_{2N}$$

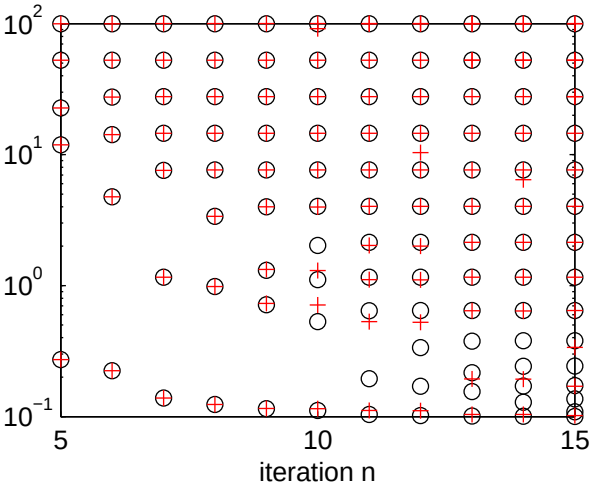
$\hat{T}_{2N}$ has all its eigenvalues close to those of A.

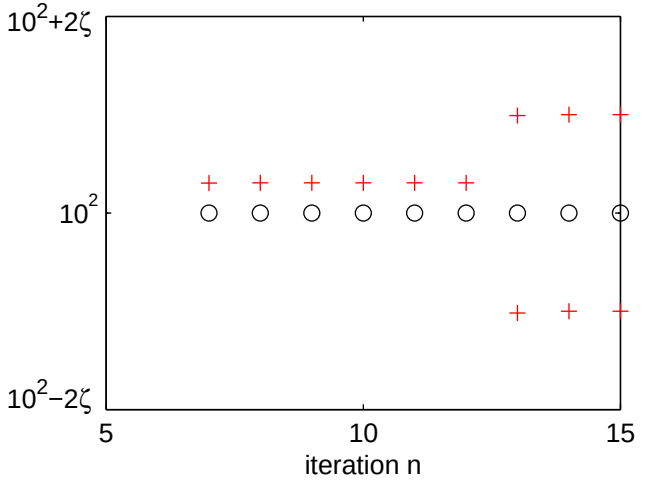
However, $\hat{T}_n$ can be for $n \leq N$ very different from T_n.

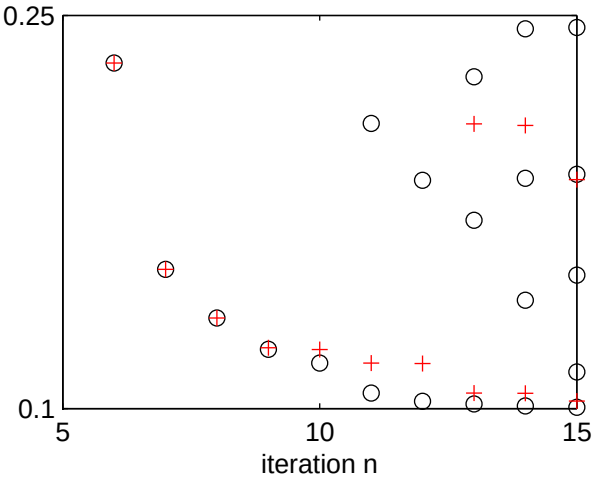
Exact arithmetic !

3 : Lanczos results for A, w_1 and $\hat{A}, \hat{w}_1$:

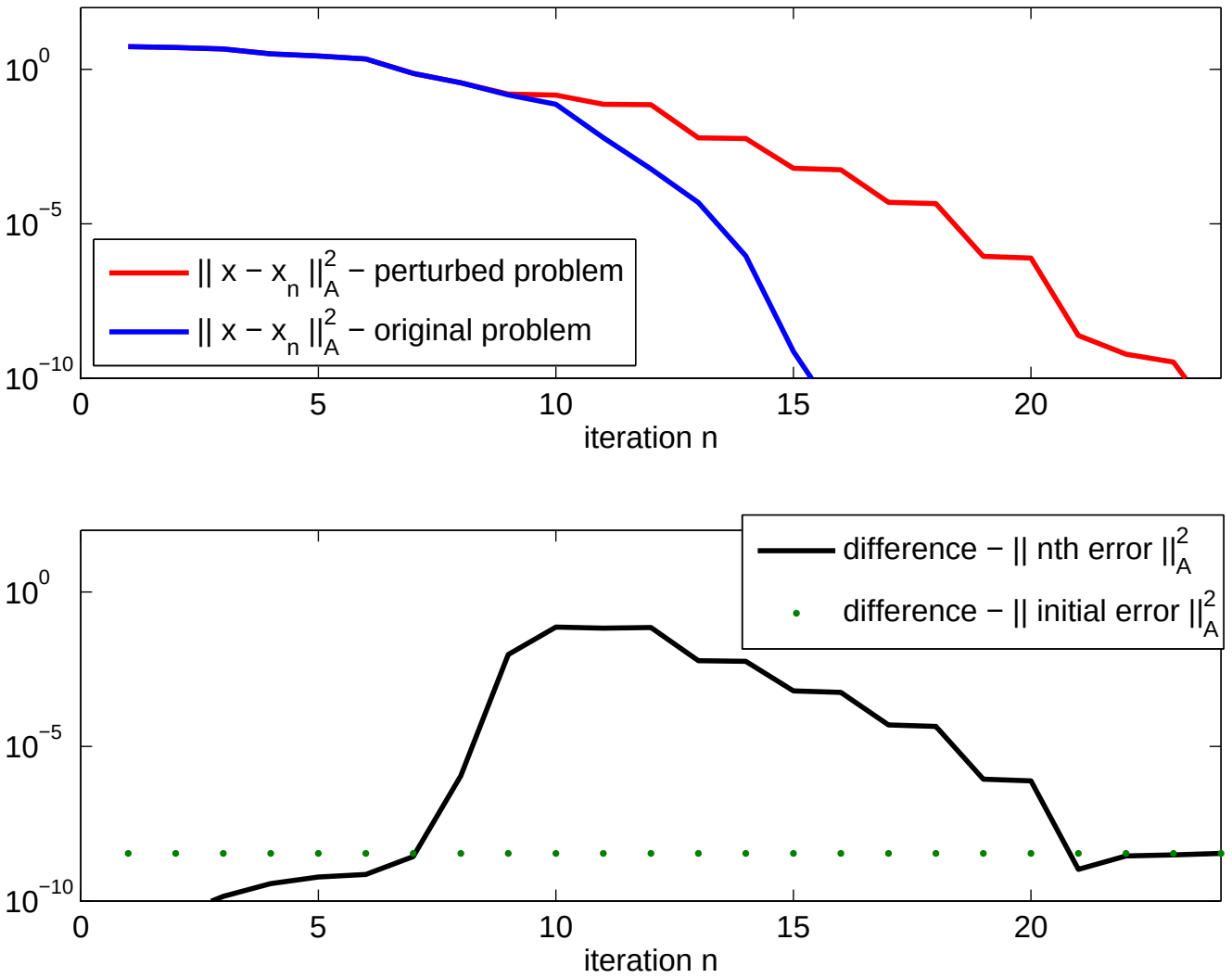








3 : CG results for A, w_1 and $\hat{A}, \hat{w}_1$:



3 : Observations

- Replacing single eigenvalues by two close ones causes large delays.
- 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.

3 : Ritz values in the presence of close eig-s

In the presence of very close eigenvalues, a Ritz value in the exact Lanczos or CG method initially converges to the cluster as fast as if the cluster were replaced by a single eigenvalue with the combined weight.

Within a few further steps it converges very fast to one of the eigenvalues, with another Ritz value converging simultaneously to approximate the rest of the cluster. In the presence of more than two eigenvalues in a cluster, the story repeats until all eigenvalues in a cluster are approximated by individual Ritz values.

The ’additional’ Ritz values in the clusters are, however missing in the other part of the spectrum, and the convergence of CG is delayed, in comparison to the single eigenvalues case, by the number of steps needed to provide the ’missing’ Ritz values.

3 : 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).

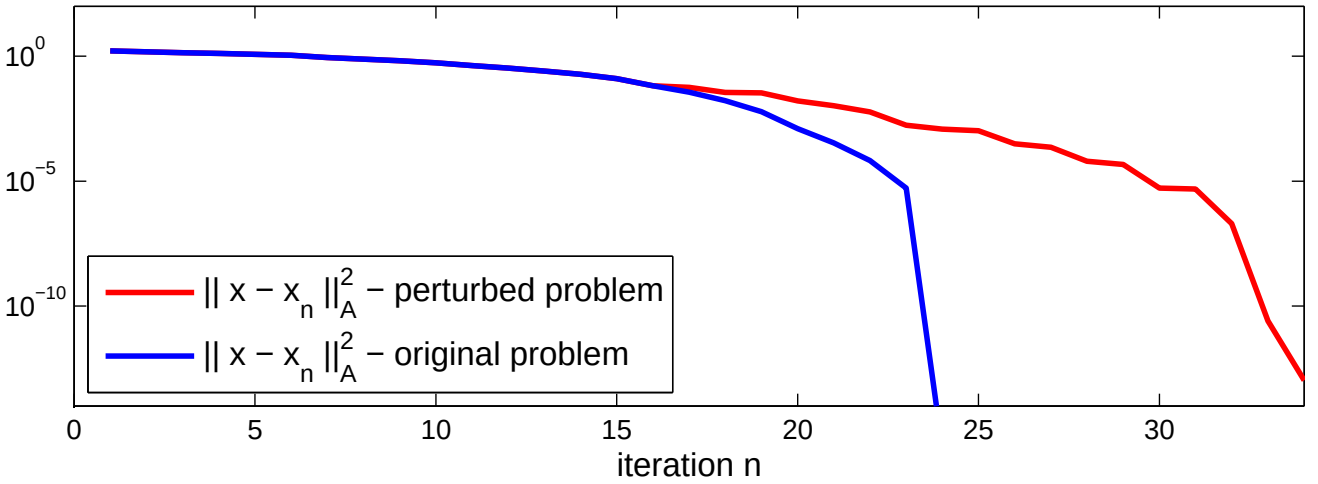
3 : Caution

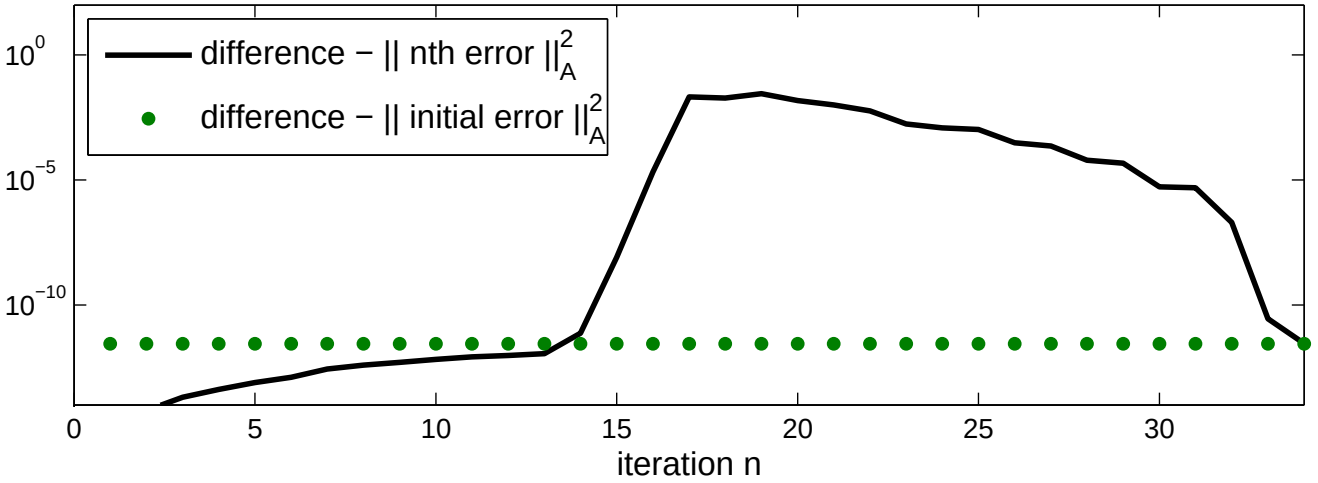
Kratzer, Parter and Steuerwalt, *Block splittings for the conjugate gradient method*, Computers and Fluids 11, (1983), pp. 255-279. The statement on p. 261, second paragraph, in our notation says:

> 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.

Similar statements can be found in several later papers and some books. The arguments are based on relating the CG minimizing polynomial to the minimal polynomial of A. For some distribution of eigenvalues of A, however, 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 (here at the eigenvalues of $\hat{A}$).

3 : CG results for A, w_1 and $\hat{A}, \hat{w}_1$, $\gamma = 0.8$:





Outline

1. Lanczos, CG and the Gauss-Christoffel quadrature
2. Reconstruction of Jacobi matrices from the spectral data
3. Another problem
4. Gauss-Christoffel quadrature can be sensitive to small perturbations of the distribution function
5. Lanczos and CG in finite precision arithmetic

4 : CG and Gauss-Ch. quadrature errors

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,

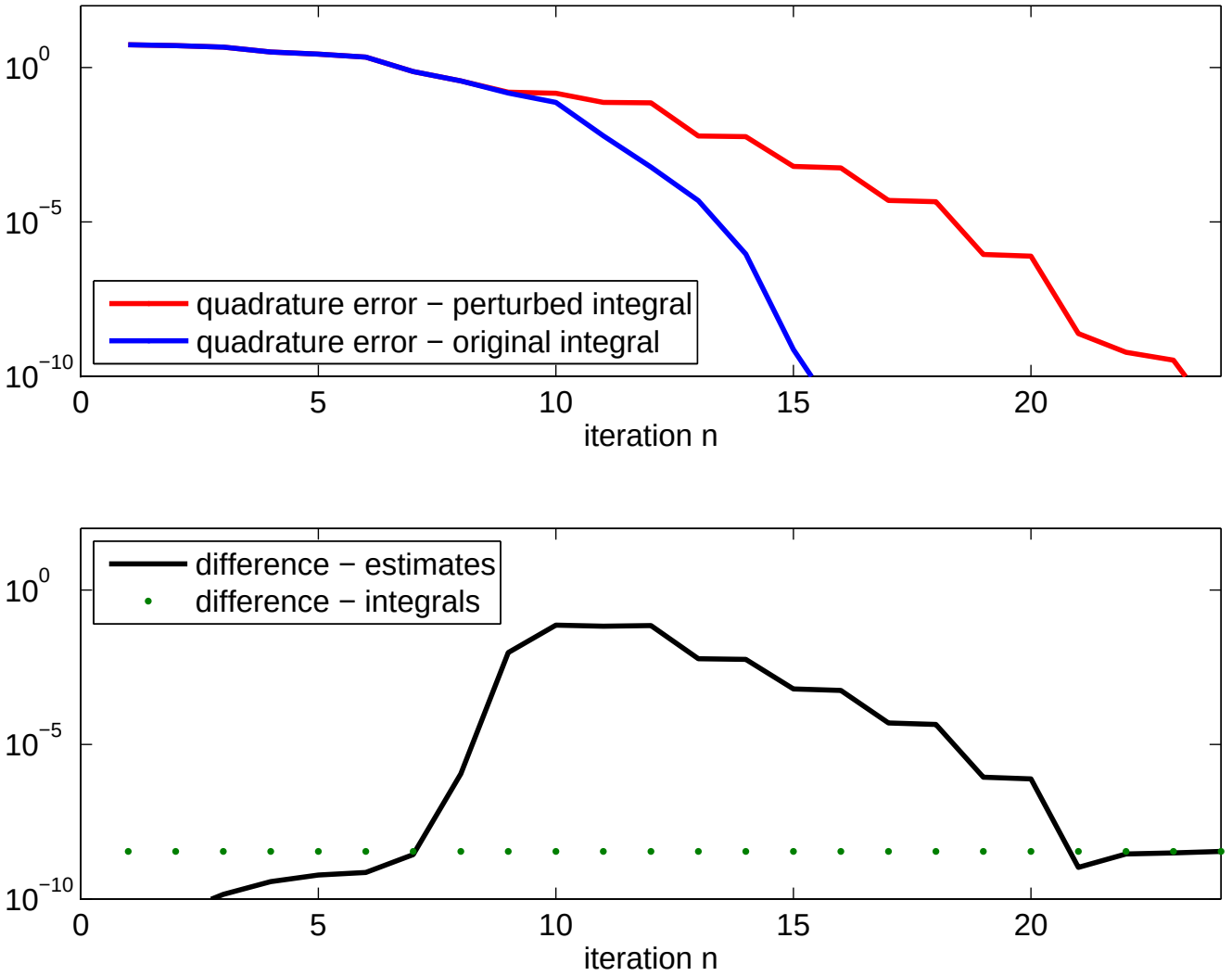
$$\int_{\zeta}^{\xi} f(\lambda)\, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} f(\theta_i^{(n)}) \;+\; R_n(f)\,.$$

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

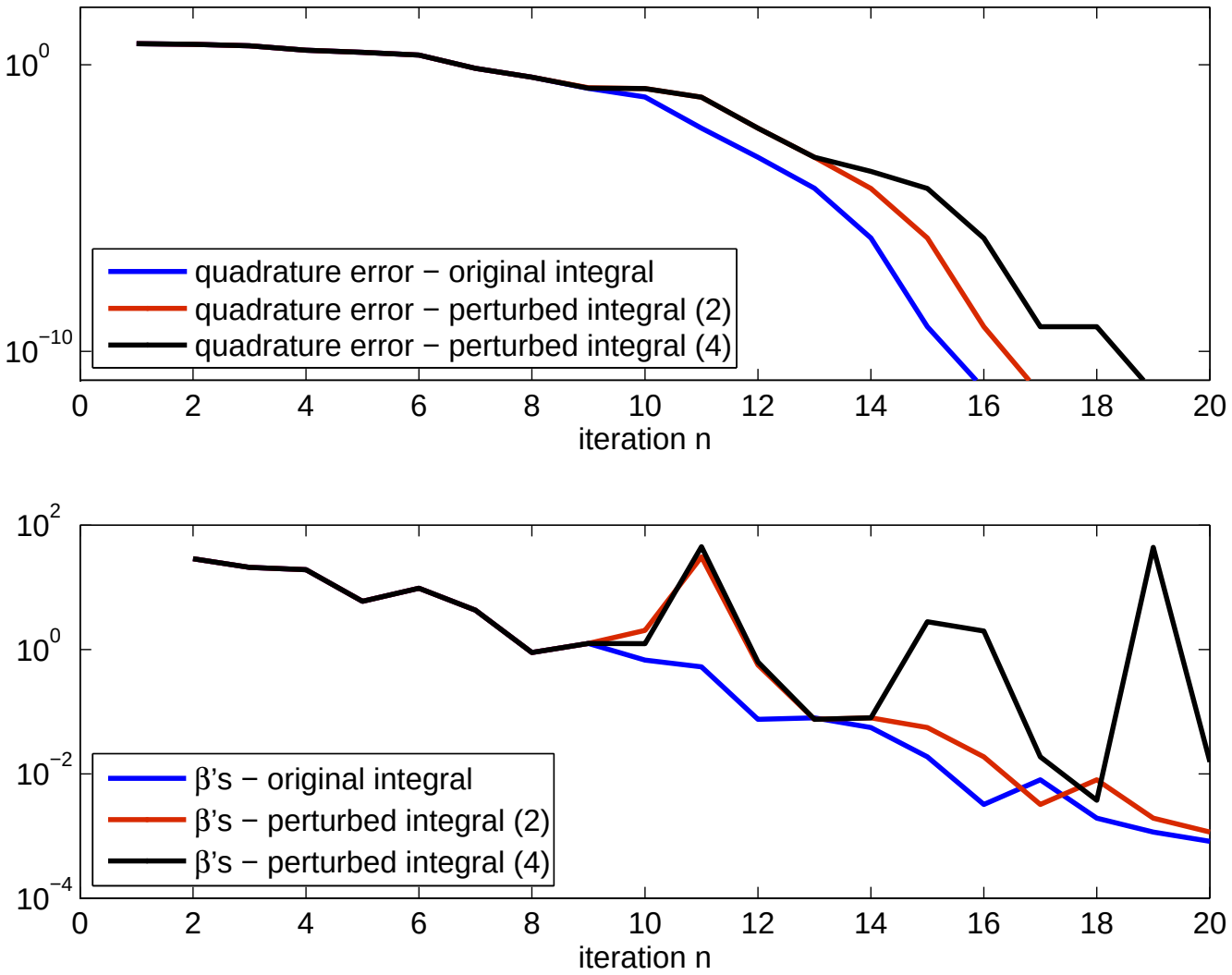
$$\frac{\|x - x_0\|_{\mathbf{A}}^2}{\|r_0\|^2} \;=\; n\text{-th Gauss quadrature} \;+\; \frac{\|x - x_n\|_{\mathbf{A}}^2}{\|r_0\|^2}\,.$$

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

4 : Sensitivity of the Gauss-Ch. Quadrature



4 : Simplified problem



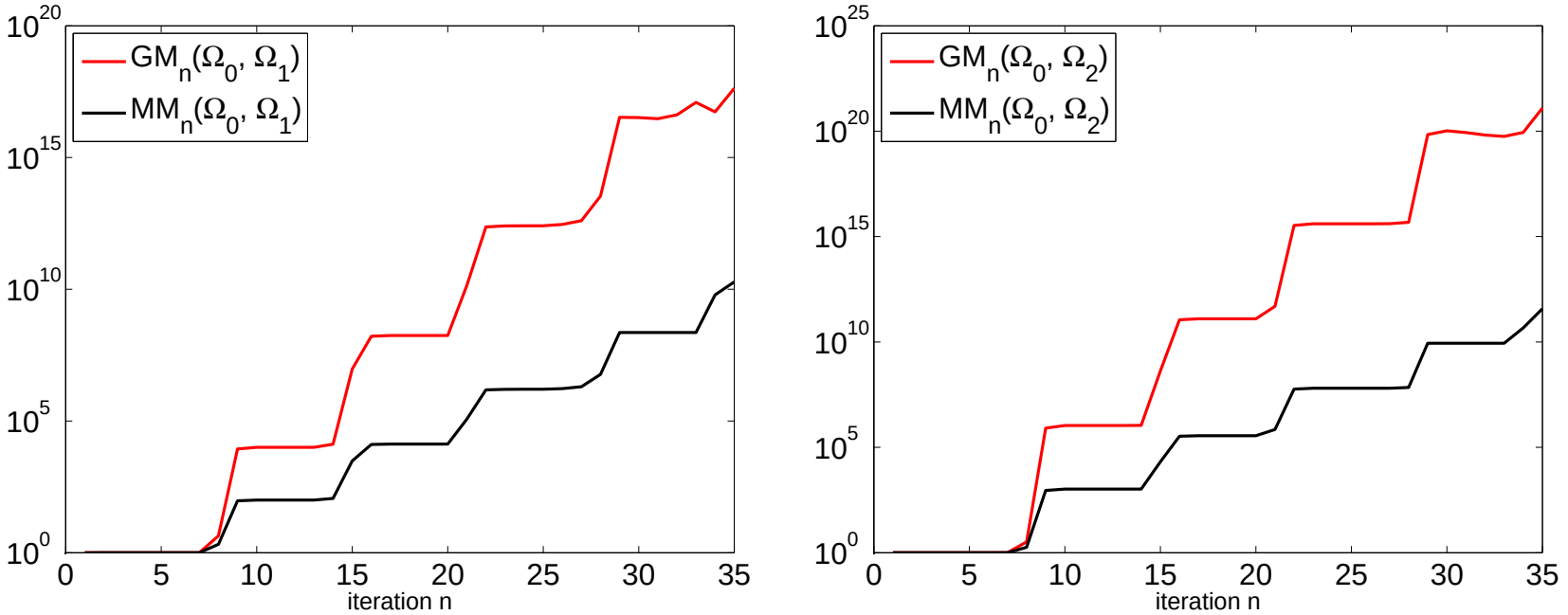
4 : Theorem - O'Leary, S, Tichý (2007)

Consider distribution functions $\omega(x)$ and $\tilde{\omega}(x)$ on $[a,b]$. Let $p_n(x) = (x-x_1)\dots(x-x_n)$ and $\tilde{p}_n(x) = (x-\tilde{x}_1)\dots(x-\tilde{x}_n)$ be the nth orthogonal polynomials corresponding to ω and $\tilde{\omega}$ respectively, with $\hat{p}_s(x) = (x-\xi_1)\dots(x-\xi_s)$ their least common multiple.

If f'' is continuous on $[a,b]$, then the difference $\Delta^n_{\omega,\tilde{\omega}}$ between the approximation I^n_ω to I_ω and the approximation $I^n_{\tilde{\omega}}$ to $I_{\tilde{\omega}}$, obtained from the k-point Gauss-Christoffel quadrature, is bounded as

$$
\begin{aligned}
|\Delta^n_{\omega,\tilde{\omega}}| \quad &\leq \quad \left| \int_a^b \hat{p}_s(x) f[\xi_1,\dots,\xi_s,x]\, d\omega(x) \;-\; \int_a^b \hat{p}_s(x) f[\xi_1,\dots,\xi_s,x]\, d\tilde{\omega}(x) \right| \\
&+ \quad \left| \int_a^b f(x)\, d\omega(x) \;-\; \int_a^b f(x)\, d\tilde{\omega}(x) \right| .
\end{aligned}
$$

4 : Modified moments do not help



Condition numbers of the matrix of the modified moments (GM) and the matrix of the mixed moments (MM). Left - enlarged supports, right - shifted supports.

4 : Summary

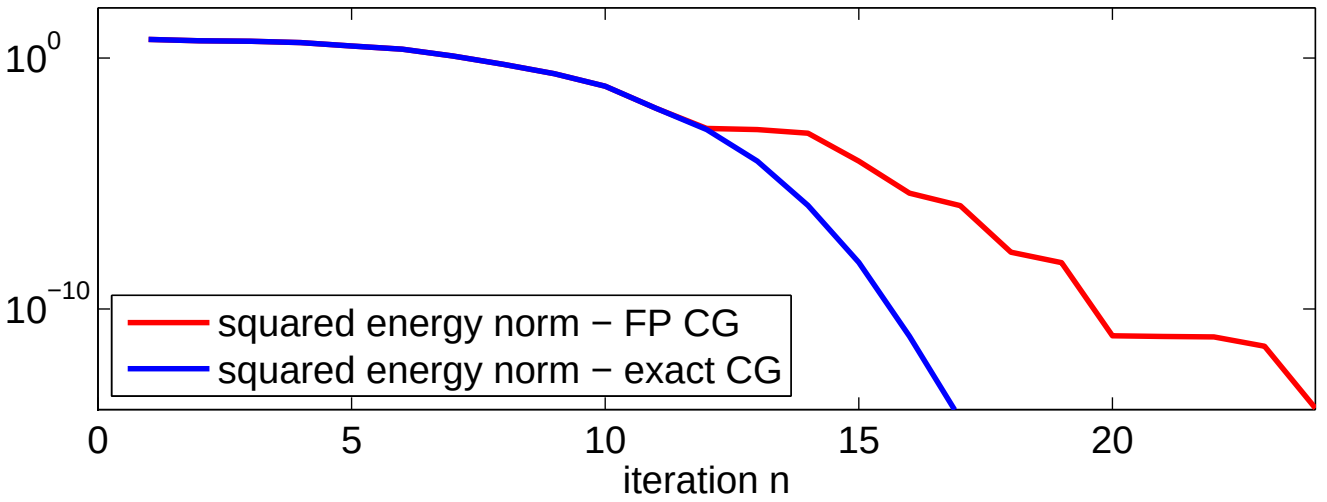
1. Gauss-Christoffel quadrature for a small number of quadrature nodes can be highly sensitive to small changes in the distribution function. In particular, the difference between the corresponding quadrature approximations (using the same number of quadrature nodes) can be many orders of magnitude larger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

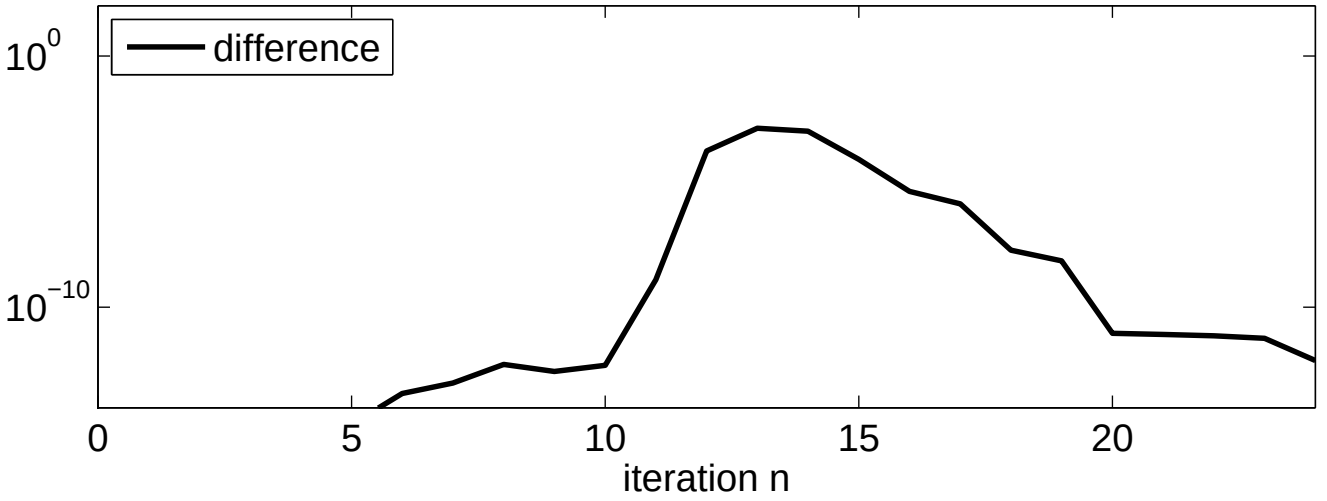
2. This sensitivity in Gauss-Christoffel quadrature can be observed for discontinuous, continuous, and even analytic distribution functions, and for analytic integrands uncorrelated with changes in the distribution functions and with no singularity close to the interval of integration.

Outline

1. Lanczos, CG and the Gauss-Christoffel quadrature
2. Reconstruction of Jacobi matrices from the spectral data
3. Another problem
4. Gauss-Christoffel quadrature can be sensitive to small perturbations of the distribution function
5. Lanczos and CG in finite precision arithmetic

4 : CG applied to the basic problem





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 $\hat{A}\,\hat{x} = \hat{b}$???

Finite precision Lanczos and CG computations as exact computations for a different problem

Paige (71–80), Greenbaum (89).

Parlett, Scott, Simon, Grcar, Cullum, S, Golub, Notay, Druskin, Knizhnerman, Meurant, Tichý, Wúlling, Zemke ...

Recent review and update Meurant and S, Acta Numerica, (06), Meurant (06).

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.
- The pieces of the mosaic fit together.

Thank you!