On sensitivity of Jacobi matrices to spectral data and related problems

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

Bedlewo, March 2007.

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) components of the normalized eigenvectors. Are the entries sensitive to small perturbations 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. Experiments
5. Explanation and relationships

1. Lanczos, CG and the Gauss-Christoffel quadrature

Lanczos tridiagonalization (1950, 1952)

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

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

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

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

Lanczos, CG and Gauss-Ch. quadrature

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

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

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).

Sensitivity of the Lanczos recurrences

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

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

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

The relationship $T_N(A, w_1) = W_N^T(A, w_1) A W_N(A, w_1)$ is strongly nonlinear.

$\tilde{T}_N$ 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?

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!

2. Reconstruction of Jacobi matrices from spectral data

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);

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 Tichy (2006).

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.

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.

3. Another problem

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 $\omega(\lambda)$.

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

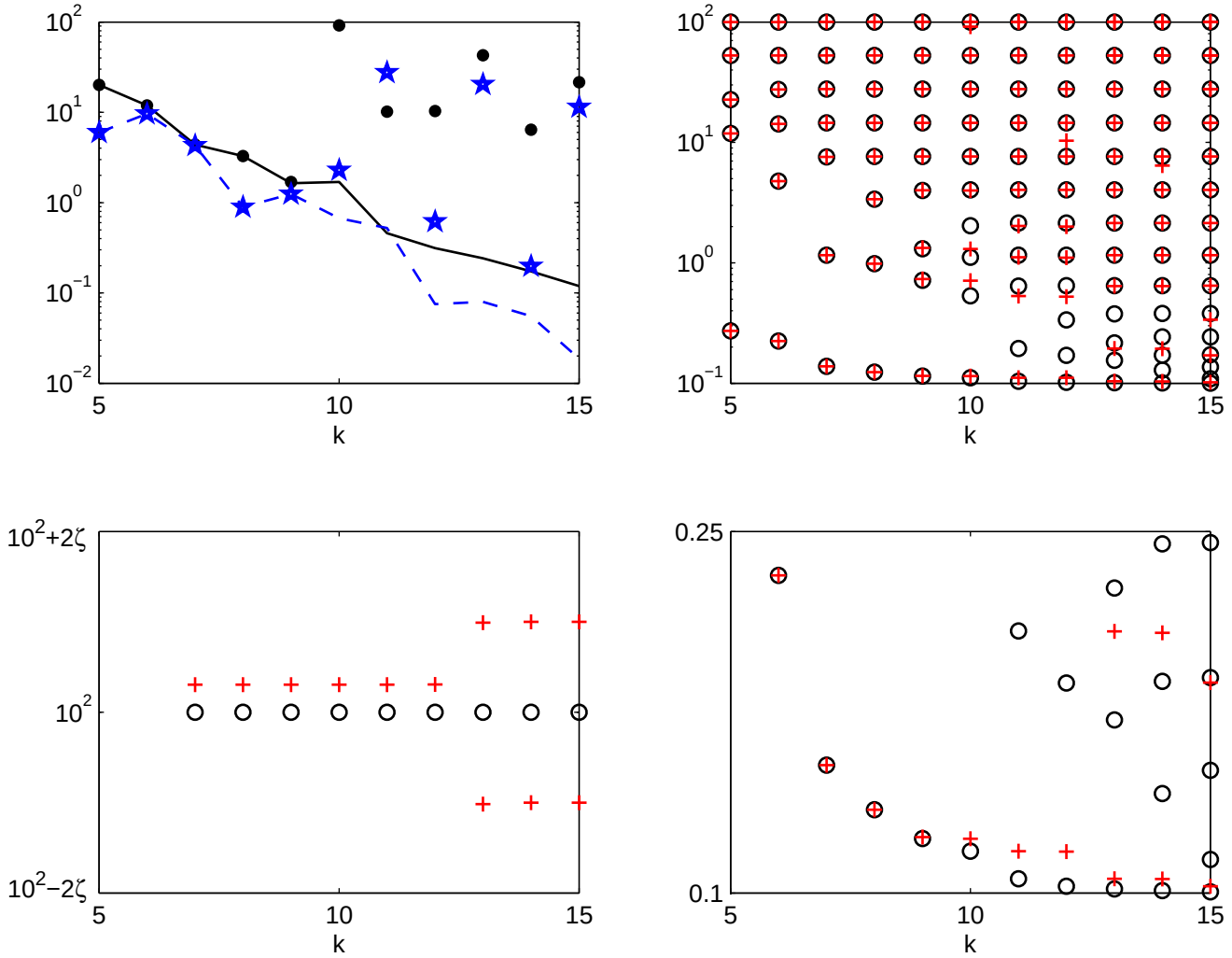
$\hat{T}_{2N}$ has all its eigenvalues close to that of A.
However, $\hat{T}_n$ can be for $n \leq N$ very different from T_n.

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!

4. Experiments

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



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,

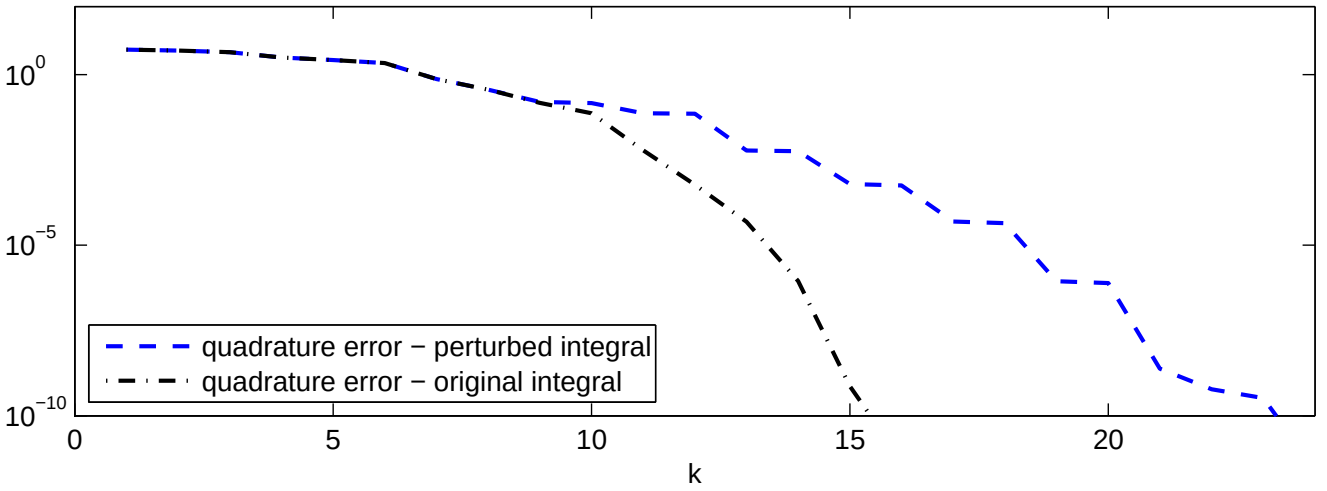
$$\int_{\zeta}^{\xi} f(\lambda)\, d\omega(\lambda) \;=\; \sum_{i=1}^{n} \omega_i^{(n)} f(\theta_i^{(n)}) \;+\; R_n(f)\,.$$

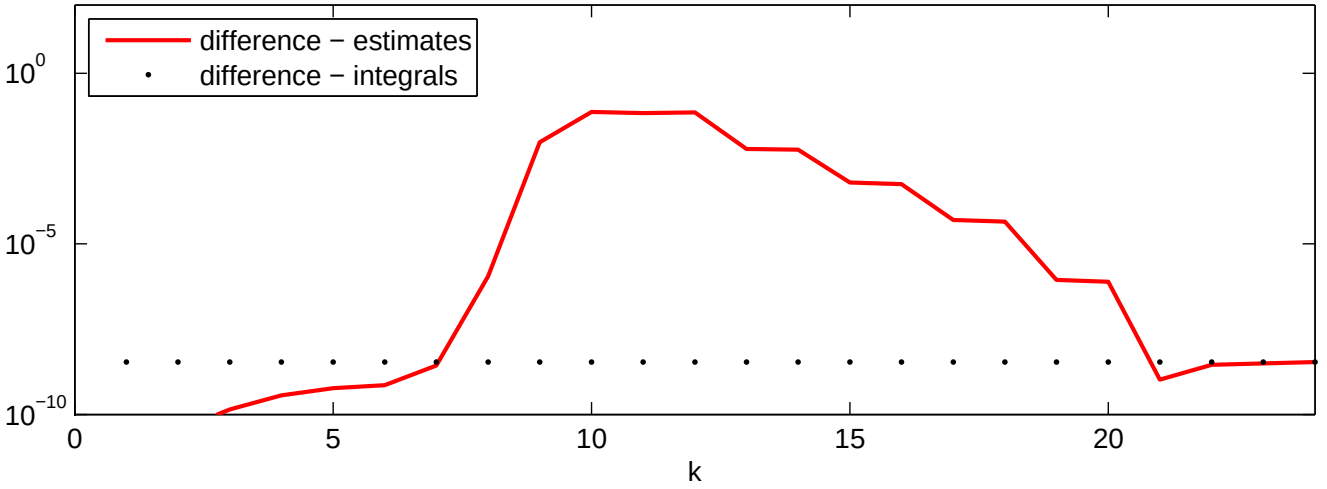
For $f(\lambda) \equiv \lambda^{-1}$ the formula takes the form

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

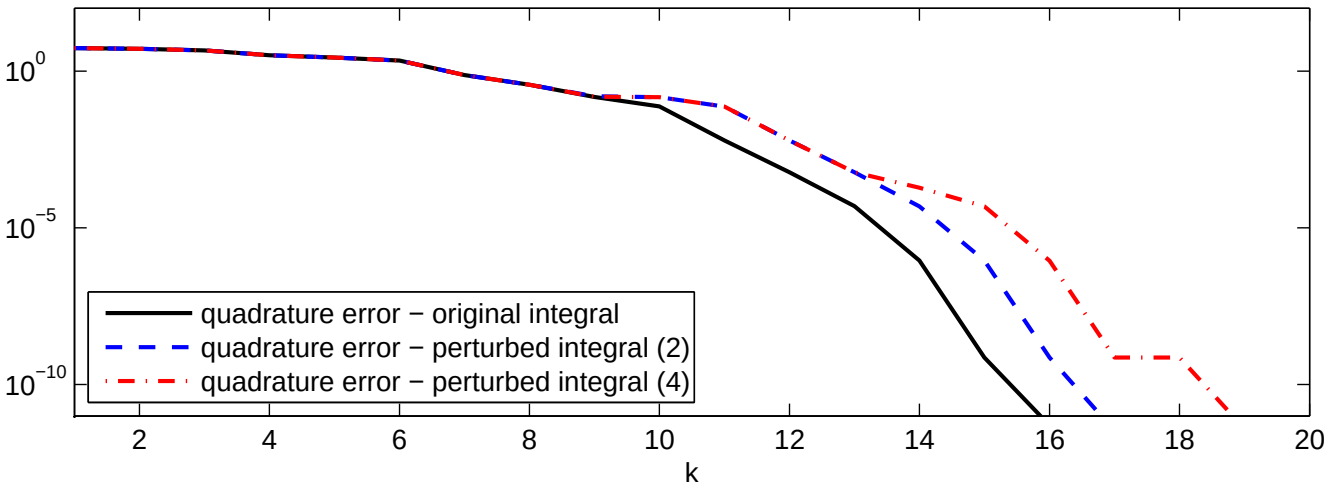
This was a base for CG error estimation in
[DaGoNa-78, GoFi-93, GoMe-94, GoSt-94, GoMe-97, . . .] .

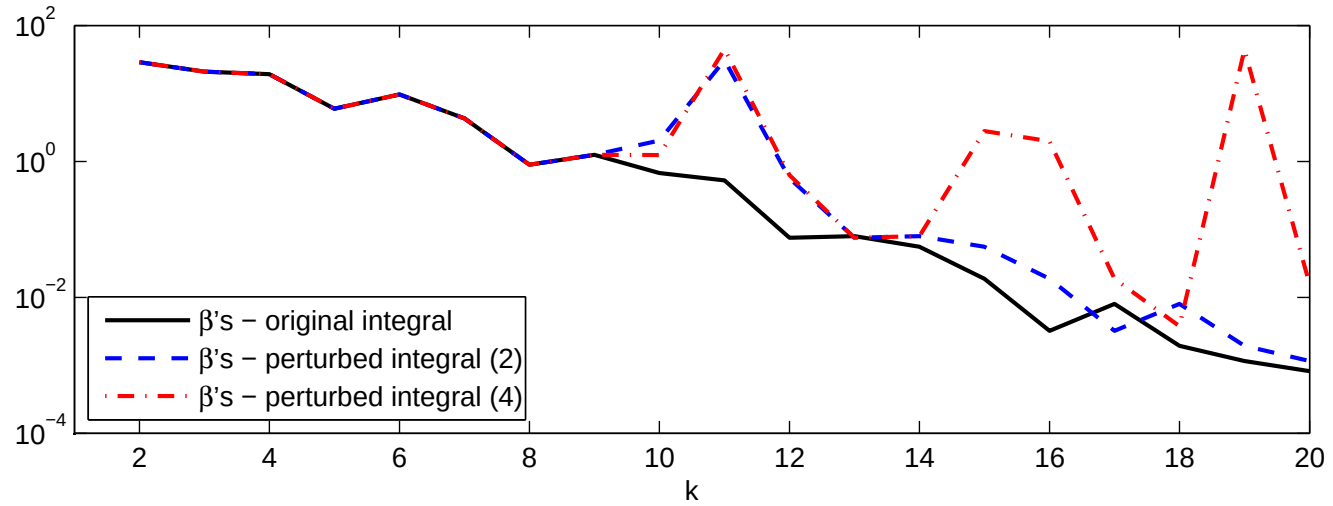
Sensitivity of the Gauss-Ch. Quadrature





Simplified problem





5. Explanation and relationships

In terms of Ritz values (quadrature nodes)

In the presence of very close eigenvalues, a Ritz value in the Lanczos and the 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 (quadrature) approximation is delayed by the number of steps needed to provide the ’missing’ Ritz values.

Agreement with published results

The fact that the presence of close eigenvalues affects 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).

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 $\hat{A}$!

Similar error can be found in several later papers and also in books.

Conclusions

- It is good to look for interdisciplinary links and for different lines of thought. Here to look at the conjugate gradient method as the Gauss-Ch. quadrature, and vice versa!

- 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.

Thank you!