Behaviour of the Gauss-Radau upper bound

Petr Tichý
Charles University, Prague

joint work with
Gérard Meurant

GAMM Workshop, September 22, 2022, Prague

The conjugate gradient method

A is symmetric and positive definite, $Ax = b$

input A, b

$r_0 = b$, $p_0 = r_0$

for $k = 1, 2, \ldots$ until conv. **do**

$$\begin{aligned}
\gamma_{k-1} &= \frac{r_{k-1}^T r_{k-1}}{p_{k-1}^T A\, p_{k-1}} \\
x_k &= x_{k-1} + \gamma_{k-1} p_{k-1} \\
r_k &= r_{k-1} - \gamma_{k-1} A\, p_{k-1} \\
\delta_k &= \frac{r_k^T r_k}{r_{k-1}^T r_{k-1}} \\
p_k &= r_k + \delta_k p_{k-1}
\end{aligned}$$

end for

Vectors $\in \mathcal{K}_k(A, b)$

$\text{span}\{b, Ab, \ldots, A^{k-1}b\}$

Orthogonality

$r_i \perp r_j \qquad p_i \perp_A p_j$

Coefficients $\rightarrow$ T_k

How to measure quality of approximation?

... it depends on what problem we solve.

- **using residual information,**
  – normwise backward error,
  – relative residual norm.

  [Hestenes, Stiefel 1952]: "Using of the residual vector r_k as a measure of the "goodness" of the estimate x_k is not reliable"

- **using error estimates,**
  – estimate of the A-norm of the error,
  – estimate of the Euclidean norm of the error.

  [Hestenes, Stiefel 1952] : "The function $(x - x_k, A(x - x_k))$ can be used as a measure of the "goodness" of x_k as an estimate of x."

Estimating the A-norm of the error in CG

$$\|x - x_k\|_A^2$$

- An important role in **stopping criteria**:

  [Deuflhard 1994], [Arioli 2004],
  [Jiránek, Strakoš, Vohralík 2006], [Papež, Vohralík 2022]

- **Estimating errors** using quadrature approach:

  [Dahlquist, Golub, Nash 1978],
  [Golub, Meurant 1994, 1997], [Golub, Strakoš 1994],
  [Meurant 1997, 1999, 2005], [Calvetti, Morigi, Reichel, Sgallari, 2000, 2001],
  [Strakoš, T. 2002], [Meurant, T. 2013, 2019], [Meurant, Papež, T. 2021]

- Why it works in **finite precision** arithmetic?

  [Paige 1976, 1980, Greenbaum 1989],
  [Golub, Strakoš 1994], [Strakoš, T. 2002, 2005, 2011]

Quadrature bounds

Gauss quadrature lower bound

- It holds that

$$\gamma_k \|r_k\|^2 \; < \; \|x - x_k\|_A^2 \; \equiv \; \varepsilon_k \, .$$

- One can improve the accuracy of the lower bound using

$$\varepsilon_k \; = \; \underbrace{\sum_{j=k}^{\ell-1} \gamma_j \|r_j\|^2}_{\Delta_{k:\ell-1}} \; + \; \varepsilon_\ell \, .$$

[Golub, Strakoš 1994, Golub, Meurant 1997, Strakoš, T. 2002, 2005]

- How to choose $\ell > k$ such that

$$\frac{\varepsilon_k - \Delta_{k:\ell-1}}{\varepsilon_k} \; \leq \; \tau \, .$$

[Meurant, Papež, T. 2021]

Gauss-Radau upper bound

- Given $\mu \leq \lambda_{\min}$, it holds that

$$\|x - x_k\|_A^2 < \gamma_k^{(\mu)} \|r_k\|^2$$

where

$$\gamma_{k+1}^{(\mu)} = \frac{\left(\gamma_k^{(\mu)} - \gamma_k\right)}{\mu\left(\gamma_k^{(\mu)} - \gamma_k\right) + \delta_{k+1}}, \qquad \gamma_0^{(\mu)} = \frac{1}{\mu}.$$

[Meurant, T. 2013]

Practically relevant questions:

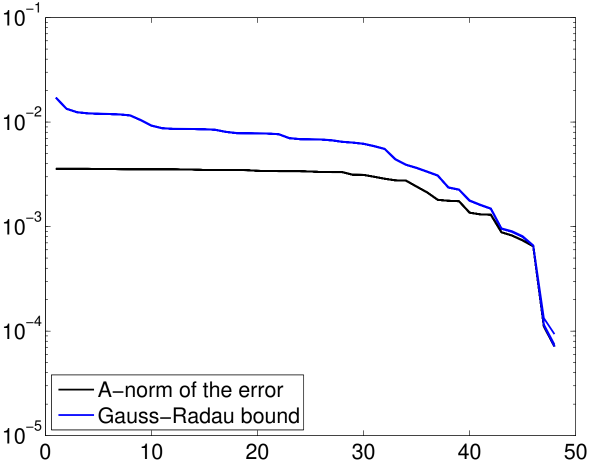
- How to get μ? [Gergelits, Mardal, Nielsen, Strakoš 2019, 2020, 2022] [Ladecký, Pultarová, Zeman 2021, 2021]
- **Quality** of the bound?
- Numerical **behavior**?

Behaviour of the upper bound

Upper bound in exact arithmetic

Gauss-Radau bound, bcsstk01 matrix, $n = 48$

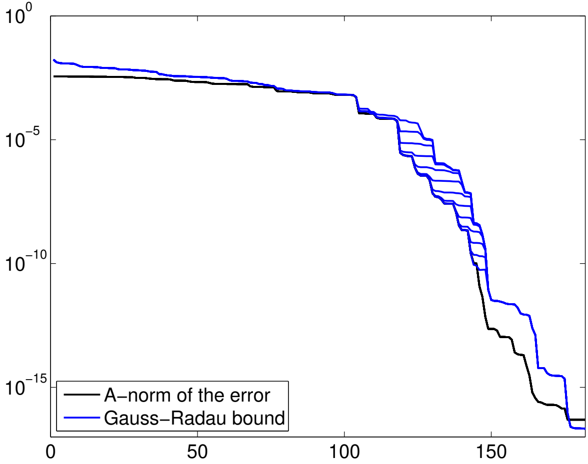
$$\mu = \frac{\lambda_{\min}}{1 + 10^{-m}}, \quad m = 2, \dots, 14$$



Upper bound in finite precision arithmetic

Gauss-Radau bound, bcsstk01 matrix, $n = 48$

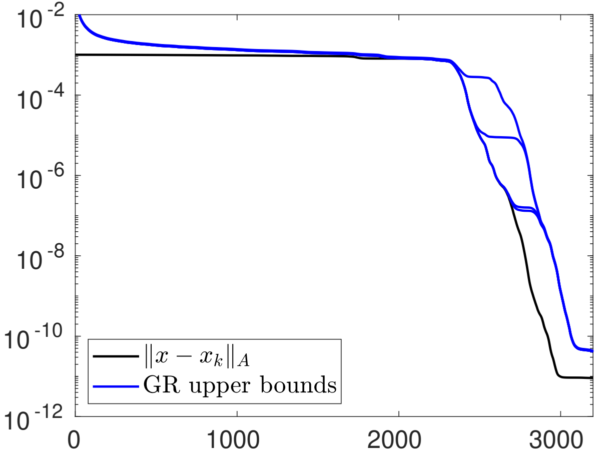
$$\mu = \frac{\lambda_{\min}}{1 + 10^{-m}}, \quad m = 2, \dots, 14$$



Upper bound in finite precision arithmetic

Gauss-Radau bound, s3dkt3m2 matrix, $n = 90449$

$$\gamma_{k+1}^{(\mu)} = \frac{\left(\gamma_k^{(\mu)} - \gamma_k\right)}{\mu\left(\gamma_k^{(\mu)} - \gamma_k\right) + \delta_{k+1}}$$

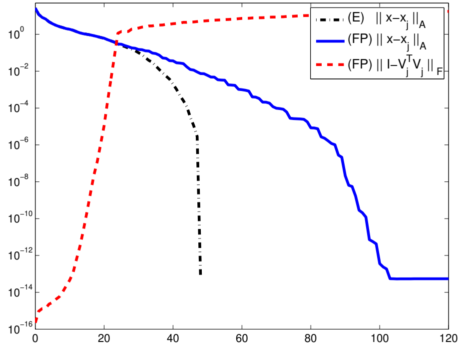


Mathematical model

of finite precision CG computations

The results of **finite precision CG** can be interpreted (up to a small inaccuracy) as the results of **exact CG** applied to a larger problem with a matrix having **clustered eigenvalues** around λ_i's.

[Greenbaum 1989], [Greenbaum, Strakoš 1992], [Paige 1976, 1980]



A model problem

[Meurant, T. 2022]

- **Consider** $\hat{w} = m^{-1/2}(1, \ldots, 1)^T$ and $\hat{\Lambda} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_m)$,

$$\hat{\lambda}_i = \hat{\lambda}_1 + \frac{i-1}{m-1}(\hat{\lambda}_m - \hat{\lambda}_1)\rho^{m-i}, \quad i = 2, \ldots, m.$$

  see [Strakos 1991], and “blur” $\hat{\Lambda}$ and $\hat{w}$.

- **Blurred** eigenvalues and weights

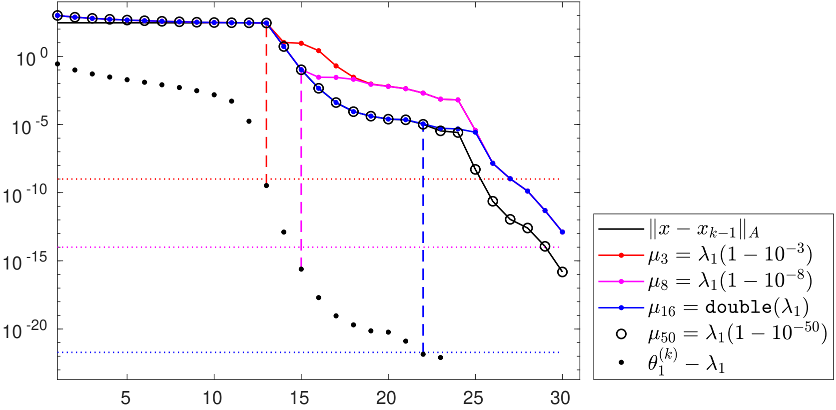
$$\lambda_j^{(i)}, \quad \omega_j^{(i)} = \frac{\hat{\omega}_i}{c_i}, \quad j = 1, \ldots, c_i,$$

  uniformly distributed in $[\hat{\lambda}_i - \delta, \hat{\lambda}_i + \delta]$, c_i grows from 1 to p.

- $m = 12$, $\hat{\lambda}_1 = 10^{-6}$, $\hat{\lambda}_m = 1$, $\rho = 0.8$, $\delta = 10^{-10}$, $p = 4$, resulting in $\Lambda x = w$ of **size** $N = 30$.

Loss of accuracy of the Gauss-Radau upper bound

Current work [Meurant, T. 2022]



Analysis

Analysis [Meurant, T. 2022]

Based on modified tridiagonal matrices

$$T_{k+1}^{(\mu)} = \left[\begin{array}{cccc|c} \alpha_1 & \beta_1 & & & \\ \beta_1 & \ddots & \ddots & & \\ & \ddots & \ddots & \beta_{k-1} & \\ & & \beta_{k-1} & \alpha_k & \beta_k \\ \hline & & & \beta_k & \alpha_{k+1}^{(\mu)} \end{array}\right]$$

that have μ as an eigenvalue,

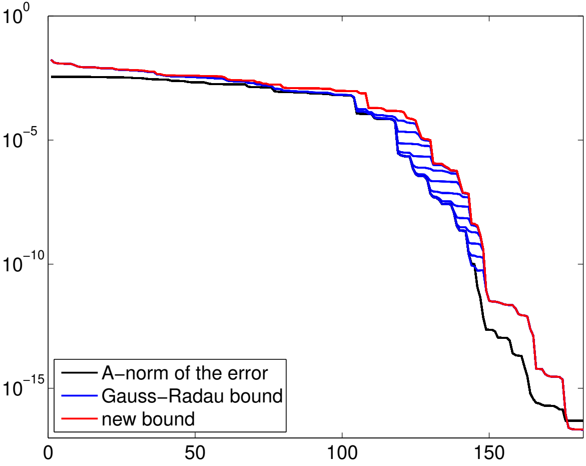
$$\alpha_{k+1}^{(\mu)} \; = \; \mu + \sum_{i=1}^{k} \eta_{i,k}^{(\mu)}, \qquad \eta_{i,k}^{(\mu)} \equiv \frac{\left(\beta_k s_{k,i}^{(k)}\right)^2}{\theta_i^{(k)} - \mu}.$$

Analysis of behaviour of the term $\eta_{1,k}^{(\mu)}$, other terms are not sensitive to small modifications of μ.

Simple upper bound for $\mu < \lambda_{\min}$

bcsstk01, $n = 48$, [Meurant, T. 2019]

$$\|x - x_k\|_A^2 \; < \; \gamma_k^{(\mu)} \|r_k\|^2 \; < \; \frac{\|r_k\|^2}{\mu \|p_k\|^2} \|r_k\|^2$$



Closeness of the Gauss-Radau upper bound

and the simple upper bound

We observe that if the upper bound is delayed, then

$$\gamma_k^{(\mu)} \approx \frac{\|r_k\|^2}{\mu\|p_k\|^2}.$$

Explanation based on the formula

$$\gamma_k^{(\mu)} = \left(\frac{\mu\|p_k\|^2}{\|r_k\|^2} + \sum_{i=1}^{k} \left(\frac{\mu}{\theta_i^{(k)}} \right)^2 \eta_{i,k}^{(\mu)} \right)^{-1}.$$

[Meurant, T. 2022]

Improving accuracy of the upper bound

Idea

$$\varepsilon_k = \Delta_{k:\ell-1} + \varepsilon_\ell \leq \underbrace{\Delta_{k:\ell-1} + \gamma_\ell^{(\mu)} \left\| r_\ell \right\|^2}_{\Omega_{k:\ell}^{(\mu)}}$$

Choose ℓ such that

$$\frac{\Omega_{k:\ell}^{(\mu)} - \varepsilon_k}{\varepsilon_k} \leq \tau \,.$$

Since

$$\frac{\Omega_{k:\ell}^{(\mu)} - \varepsilon_k}{\varepsilon_k} < \frac{\Omega_{k:\ell}^{(\mu)} - \Delta_{k:\ell}}{\Delta_{k:\ell}} = \frac{\left\| r_\ell \right\|^2 \left(\gamma_\ell^{(\mu)} - \gamma_\ell \right)}{\Delta_{k:\ell}},$$

we can require

$$\frac{\left\| r_\ell \right\|^2 \left(\gamma_\ell^{(\mu)} - \gamma_\ell \right)}{\Delta_{k:\ell}} \leq \tau.$$

Algorithm

input A, b, μ, τ

$r_0 = b$, $p_0 = r_0$

$k = 0$, $\gamma_0^{(\mu)} = \frac{1}{\mu}$

for $\ell = 0, \ldots,$ **do**

    `cgiter(`ℓ`)`

    while $\ell \geq k$ and $\frac{\|r_\ell\|^2 \left(\gamma_\ell^{(\mu)} - \gamma_\ell\right)}{\Delta_{k:\ell}} \leq \tau$ **do**

        accept $\Omega_{k:\ell}^{(\mu)}$

        $k = k + 1$

    end while

    $\gamma_{\ell+1}^{(\mu)} = \frac{\gamma_\ell^{(\mu)} - \gamma_\ell}{\mu(\gamma_\ell^{(\mu)} - \gamma_\ell) + \delta_{\ell+1}}$

end for

- The upper bound is **delayed** if

$$|\theta_1^{(k)} - \lambda_1| < |\mu - \lambda_1|.$$

  Then, it visually coincides with the simple upper bound.

- This can happen if A has **clustered** eigenvalues (or whenever **orthogonality** is lost during computations).

- We can improve the **accuracy** of the upper bound using the information from the forthcoming CG iterations.

- The improved **Gauss lower bound** is often more **economical** than the improved Gauss-Radau upper bound.

Related papers

G. Meurant and P. Tichý, [The behaviour of the Gauss-Radau upper bound of the error norm in CG, submitted to Numer. Algorithms.]

- G. Meurant, J. Papež, and P. Tichý, [Accurate error estimation in CG, Numer. Algorithms, 88 (2021), pp. 1337-1359.]
- G. Meurant and P. Tichý, [Approximating the extreme Ritz values and upper bounds in CG, Numer. Algorithms, 82 (2019), pp. 937-968]
- G. Meurant and P. Tichý, [On computing quadrature-based bounds for the A-norm of the error in CG, Numer. Algorithms, 62 (2013), pp. 163-191]

Thank you for your attention!