On improving accuracy of the error estimates in the conjugate gradient method

Petr Tichý

Faculty of Mathematics and Physics, Charles University

based on joint work with
Gérard Meurant and Zdeněk Strakoš

PANM 19

June 24-29, 2018, Hejnice, Czech Republic

The conjugate gradient method

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

input A, b, x_0

$p_0 = r_0 = b - Ax_0$

for $k = 1, 2, \ldots$ **do**

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

end for

exact arithmetic

$\downarrow$

orthogonality

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

optimality of x_k

$$\min_{y \in \mathcal{K}_k} \|x - y\|_A \,.$$

Estimating the A-norm of the error

A brief history

- $\|x - x_k\|_A^2$ ... **measure** of the “goodness” of x_k

  [Hestenes, Stiefel 1952]

Estimating the A-norm of the error

A brief history

- $\|x - x_k\|_A^2$ . . . **measure** of the "goodness" of x_k

  [Hestenes, Stiefel 1952]

- Gauss **quadrature** error bounds (Lanczos) $\rightarrow$ **GQL**

  [Dahlquist, Golub, Nash 1978], [Golub, Meurant 1994]

Estimating the A-norm of the error

A brief history

- $\|x - x_k\|_A^2$ . . . **measure** of the "goodness" of x_k

  [Hestenes, Stiefel 1952]

- Gauss **quadrature** error bounds (Lanczos) $\rightarrow$ **GQL**

  [Dahlquist, Golub, Nash 1978], [Golub, Meurant 1994]

- Estimating errors in CG $\rightarrow$ **CGQL** $\rightarrow$ **CGQ**

  [Golub, Strakoš 1994], [Golub, Meurant, 1997], [Meurant, T. 2013]

Estimating the A-norm of the error

A brief history

- $\|x - x_k\|_A^2$ . . . **measure** of the "goodness" of x_k

  [Hestenes, Stiefel 1952]

- Gauss **quadrature** error bounds (Lanczos) $\rightarrow$ **GQL**

  [Dahlquist, Golub, Nash 1978], [Golub, Meurant 1994]

- Estimating errors in CG $\rightarrow$ **CGQL** $\rightarrow$ **CGQ**

  [Golub, Strakoš 1994], [Golub, Meurant, 1997], [Meurant, T. 2013]

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

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

Quadrature bounds

Gauss and Gauss-Radau quadrature bounds

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

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

where

$$\alpha_{k+1}^{(\mu)} = \frac{\left(\alpha_k^{(\mu)} - \alpha_k\right)}{\mu \left(\alpha_k^{(\mu)} - \alpha_k\right) + \beta_{k+1}}, \qquad \alpha_0^{(\mu)} = \frac{1}{\mu}.$$

Gauss and Gauss-Radau quadrature bounds

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

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

where

$$\alpha_{k+1}^{(\mu)} = \frac{\left(\alpha_k^{(\mu)} - \alpha_k\right)}{\mu\left(\alpha_k^{(\mu)} - \alpha_k\right) + \beta_{k+1}}, \qquad \alpha_0^{(\mu)} = \frac{1}{\mu}.$$

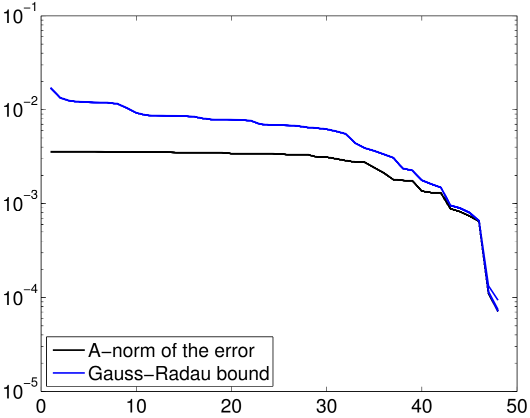
Practically relevant questions:

- How to get μ?
- **Quality** of the bound?
- Numerical **behavior**?

Upper bound in exact arithmetic

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

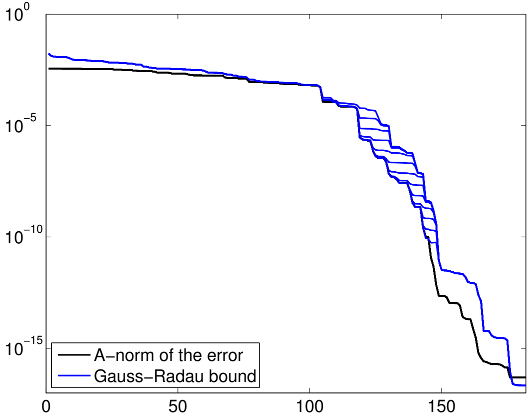
$$\mu = \frac{\lambda_{\min}}{1 + 10^{-m}}, \quad m = 2, \dots, 14$$



Upper bound in finite precision arithmetic

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

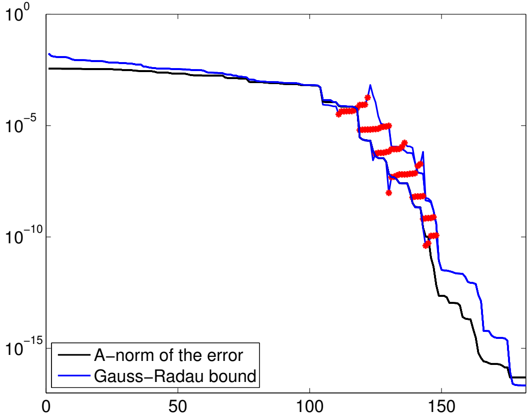
$$\mu = \frac{\lambda_{\min}}{1 + 10^{-m}}, \quad m = 2, \ldots, 14$$



Upper bound in finite precision arithmetic

$\mu > \lambda_{\min}$, bcsstk01 matrix, $n = 48$

$$\mu = \frac{\lambda_{\min}}{1 - 10^{-m}}, \quad m = 2:2:14, \quad \alpha_k^{(\mu)} < 0$$

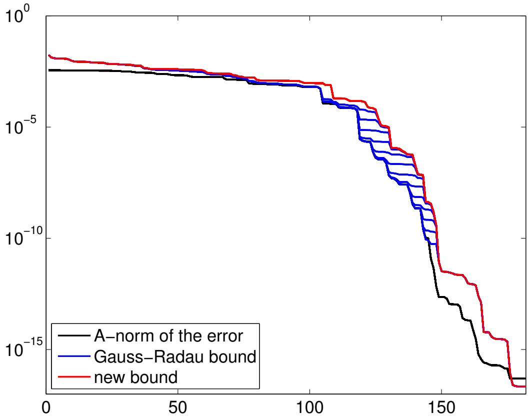


An upper bound on the upper bound

An upper bound on the Gauss-Radau bound, $\mu \leq \lambda_{\min}$

[Meurant, T. 2018?]

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



The new bound

[Meurant, T. 2018?]

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

- Having μ, we can compute it almost **for free**.
- Monotonically decreasing.
- **Not sensitive** to the choice of μ.
- As good as Gauss-Radau in many cases.
- It can be used even if $\mu > \lambda_{\min}$ (heuristics).

How to approximate μ?

The conjugate gradient method

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

input A, b, x_0
$p_0 = r_0 = b - Ax_0$
for $k = 1, 2, \ldots$ **do**

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

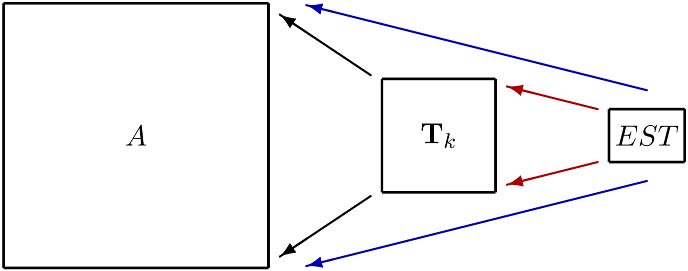
end for

$$\mathbf{U}_k$$

$$\begin{bmatrix} \frac{1}{\sqrt{\alpha_0}} & \sqrt{\frac{\beta_1}{\alpha_0}} & & \\ & \ddots & \ddots & \\ & & \ddots & \sqrt{\frac{\beta_{k-1}}{\alpha_{k-2}}} \\ & & & \frac{1}{\sqrt{\alpha_{k-1}}} \end{bmatrix}$$

$$\mathbf{T}_k = \mathbf{U}_k^T \mathbf{U}_k$$

Approximation of $\lambda_{\min}$ and $\lambda_{\max}$ in CG



$$\mathbf{T}_k = \mathbf{U}_k^T \mathbf{U}_k \quad \to \quad \lambda_{\min}(\mathbf{T}_k) \quad = \quad 1/\|\mathbf{U}_k^{-1}\|^2$$

How to approximate $\|\mathbf{U}_k^{-1}\|^2$?

- $\mathbf{U}_k$ is bidiagonal,

$$\mathbf{U}_k^{-1} \quad \rightarrow \quad \mathbf{U}_{k+1}^{-1}$$

by **adding one column and one row**.

Incremental estimation

of the smallest Ritz value in CG

- $\mathbf{U}_k$ is bidiagonal,

$$\mathbf{U}_k^{-1} \quad \rightarrow \quad \mathbf{U}_{k+1}^{-1}$$

  by **adding one column and one row**.

- **Incremental norm estimation**: incrementally improve an approximation of the maximum right singular vector.
  [Bischof 1990], [Duff, Vömmel 2002], [Duintjer Tebbens, Tůma 2014].

Approximation of the smallest Ritz value in CG

Having α_k and β_k, update

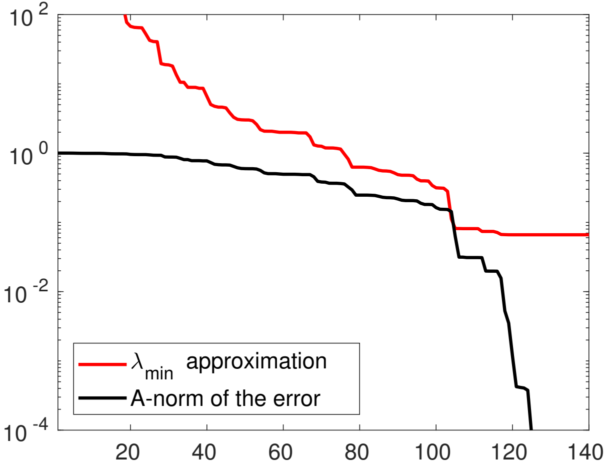
$$
\begin{aligned}
\sigma_k &= -\sqrt{\frac{\alpha_k \beta_k}{\alpha_{k-1}}}\,(s_{k-1}\sigma_{k-1} + c_{k-1}\tau_{k-1}) \\
\tau_k &= \alpha_k \left(b_k^2 \tau_{k-1} + 1\right) \\
\omega_k^2 &= (\rho_k - \tau_k)^2 + 4\sigma_k^2 \\
c_k^2 &= \frac{1}{2}\left(1 - \frac{\rho_k - \tau_k}{\omega_k}\right) \\
\rho_{k+1} &= \rho_k + \omega_k c_k^2 \\
s_k &= \sqrt{1 - c_k^2}, \\
c_k &= |c_k| \operatorname{sign}(\sigma_k) \\
\mu_{k+1} &= \rho_{k+1}^{-1}
\end{aligned}
$$

- Very cheap, no need to store vectors or coefficients,

[Meurant, T. 2018?]

Approximation of $\lambda_{\min}$

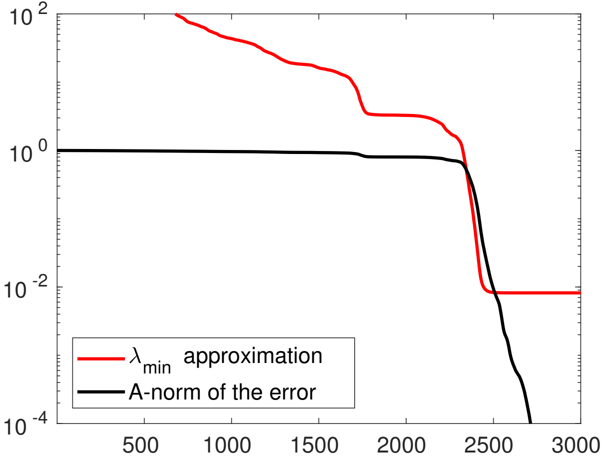
bcsstk01, $n = 48$



$$\frac{\lambda_{\min} - \mu_k}{\lambda_{\min}}$$

Approximation of $\lambda_{\min}$

s3dkt3m2, $n = 90449$, ichol

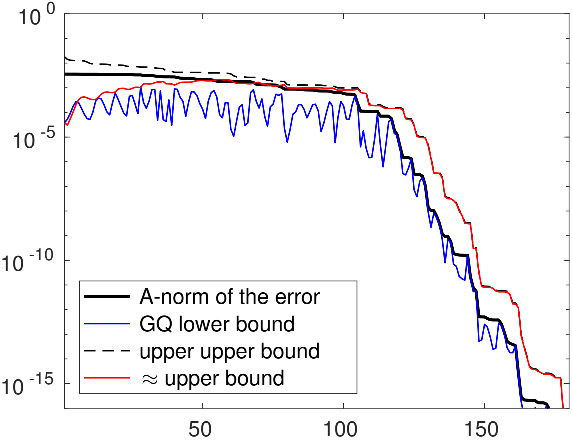


$$\frac{\lambda_{\min} - \mu_k}{\lambda_{\min}}$$

Bounds summary

bcsstk01, $n = 48$

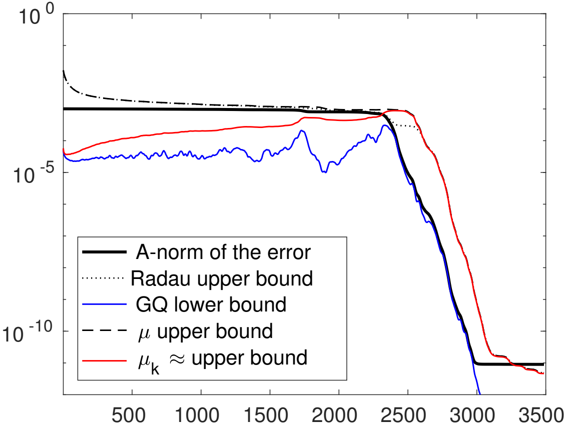
$$\alpha_k \|r_k\|^2 < \|x - x_k\|_A^2 \lesssim \frac{\|r_k\|^2}{\mu_k} \frac{\|r_k\|^2}{\|p_k\|^2}$$



Bounds summary

s3dkt3m2, $n = 90449$, ichol

$$\alpha_k \|r_k\|^2 < \|x - x_k\|_A^2 \lesssim \frac{\|r_k\|^2}{\mu_k} \frac{\|r_k\|^2}{\|p_k\|^2}$$



Improving accuracy of the estimates

$$\|x - x_k\|_A^2 \;=\; \alpha_k \|r_k\|^2 \;+\; \|x - x_{k+1}\|_A^2$$

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

Use a delay and bound $(k+d)$th error

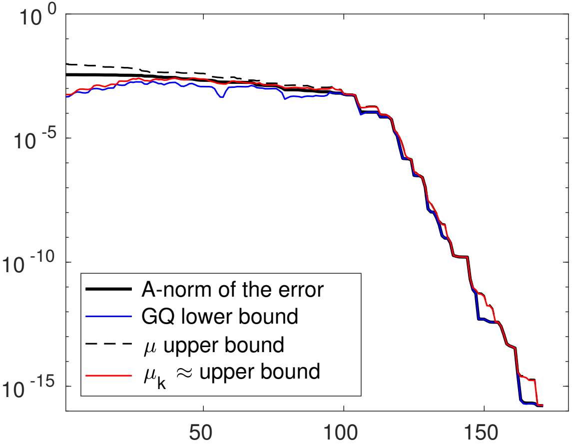
bcsstk01, $n = 48$, $d = 10$

$$\|x - x_k\|_A^2 \;=\; \sum_{j=k}^{k+d-1} \alpha_j \|r_j\|^2 \;+\; \|x - x_{k+d}\|_A^2$$

Use a delay and bound $(k+d)$th error

bcsstk01, $n = 48$, $d = 10$

$$\|x - x_k\|_A^2 = \sum_{j=k}^{k+d-1} \alpha_j \|r_j\|^2 + \|x - x_{k+d}\|_A^2$$



Use a delay and bound $(k+d)$th error

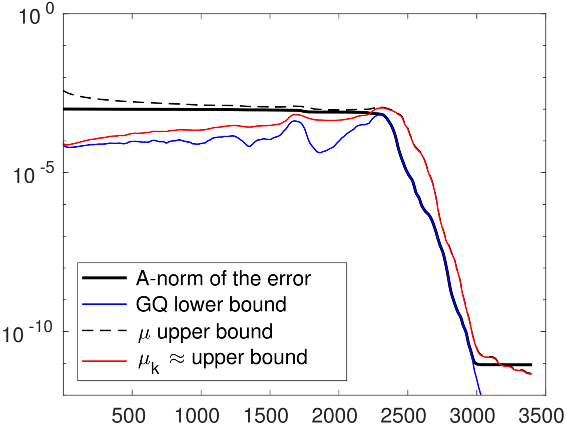
s3dkt3m2, $n = 90449$, `ichol`, $d = 100$

$$\|x - x_k\|_A^2 = \sum_{j=k}^{k+d-1} \alpha_j \|r_j\|^2 + \|x - x_{k+d}\|_A^2$$

Use a delay and bound $(k+d)$th error

s3dkt3m2, $n = 90449$, ichol, $d = 100$

$$\|x - x_k\|_A^2 = \sum_{j=k}^{k+d-1} \alpha_j \|r_j\|^2 + \|x - x_{k+d}\|_A^2$$



How to choose d adaptively?

$$\|x - x_k\|_A^2 = \underbrace{\sum_{j=k}^{k+d-1} \alpha_j \|r_j\|^2}_{\nu_{k,d}} + \|x - x_{k+d}\|_A^2,$$

$$= \nu_{k,d} + \frac{\|x - x_{k+d}\|_A^2}{\|x - x_k\|_A^2} \|x - x_k\|_A^2,$$

How to choose d adaptively?

$$\|x - x_k\|_A^2 = \underbrace{\sum_{j=k}^{k+d-1} \alpha_j \|r_j\|^2}_{\nu_{k,d}} + \|x - x_{k+d}\|_A^2,$$

$$= \nu_{k,d} + \frac{\|x - x_{k+d}\|_A^2}{\|x - x_k\|_A^2} \|x - x_k\|_A^2,$$

If

$$\frac{\|x - x_{k+d}\|_A}{\|x - x_k\|_A} < \varepsilon,$$

then

$$\nu_{k,d}^{1/2} < \|x - x_k\|_A < \frac{\nu_{k,d}^{1/2}}{\sqrt{1-\varepsilon^2}}, \quad \text{e.g.,} \quad \varepsilon = 0.8.$$

Pseudo algorithm - adaptive choice of d

$go = 1$
$d = d + 1$
while $((d \geq 1)$ and $(go))$ **do**
    compute $\nu_{k-d+1,d}$
    if $\left(\frac{\|x-x_k\|_A}{\|x-x_{k-d}\|_A} < \varepsilon\right)$ **then**
        compute lower or upper bounds at iteration $k-d$
        $d = d - 1$
    else
        $go = 0$
    end if
end while

Adaptive choice of d

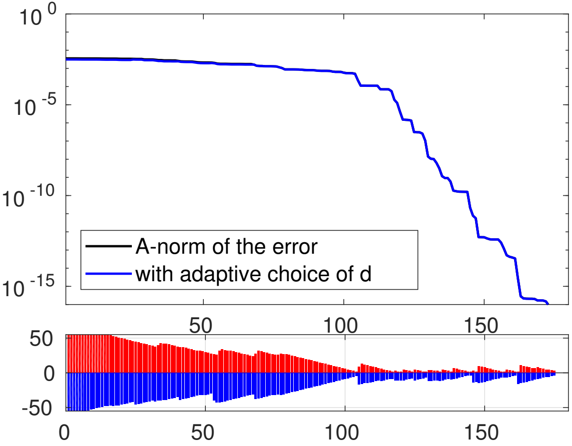
μ_k upper bound approach, bcsstk01

$$\frac{\|x - x_{k+d}\|_A}{\|x - x_k\|_A} \lesssim \frac{\frac{\|r_{k+d}\|}{\sqrt{\mu_{k+d}}} \frac{\|r_{k+d}\|}{\|p_{k+d}\|}}{\sqrt{\nu_{k,d}}} < \varepsilon = 0.5,$$

Adaptive choice of d

μ_k upper bound approach, bcsstk01

$$\frac{\|x - x_{k+d}\|_A}{\|x - x_k\|_A} \lesssim \frac{\frac{\|r_{k+d}\|}{\sqrt{\mu_{k+d}}} \frac{\|r_{k+d}\|}{\|p_{k+d}\|}}{\sqrt{\nu_{k,d}}} < \varepsilon = 0.5,$$



Adaptive choice of d

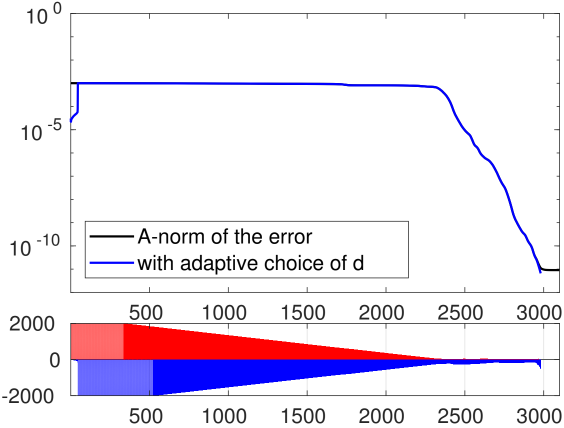
μ_k upper bound approach, s3dkt3m2

$$\frac{\|x - x_{k+d}\|_A}{\|x - x_k\|_A} \lesssim \frac{\frac{\|r_{k+d}\|}{\sqrt{\mu_{k+d}}} \frac{\|r_{k+d}\|}{\|p_{k+d}\|}}{\sqrt{\nu_{k,d}}} < \varepsilon = 0.5,$$

Adaptive choice of d

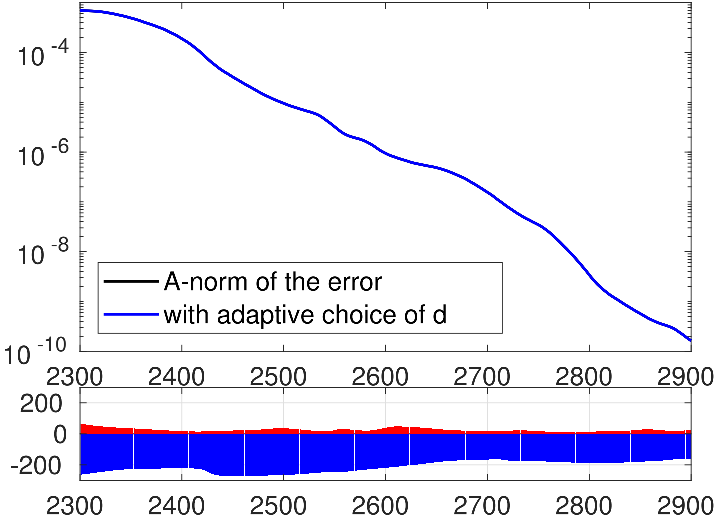
μ_k upper bound approach, s3dkt3m2

$$\frac{\|x - x_{k+d}\|_A}{\|x - x_k\|_A} \lesssim \frac{\frac{\|r_{k+d}\|}{\sqrt{\mu_{k+d}}}\frac{\|r_{k+d}\|}{\|p_{k+d}\|}}{\sqrt{\nu_{k,d}}} < \varepsilon = 0.5,$$



Adaptive choice of d

μ_k upper bound approach, s3dkt3m2



Another approach

superliner, linear, sublinear CG convergence

Let $\alpha > \beta > \gamma > 0$. Then

$$\frac{\beta - \gamma}{\alpha - \beta} > \frac{\beta}{\alpha} \quad \Leftrightarrow \quad \frac{\beta}{\alpha} > \frac{\gamma}{\beta}.$$

Another approach

superliner, linear, sublinear CG convergence

Let $\alpha > \beta > \gamma > 0$. Then

$$\frac{\beta - \gamma}{\alpha - \beta} > \frac{\beta}{\alpha} \quad \Leftrightarrow \quad \frac{\beta}{\alpha} > \frac{\gamma}{\beta}.$$

Apply to $\|x - x_{k-d}\|_A^2 > \|x - x_k\|_A^2 > \|x - x_{k+d}\|_A^2$. Then

$$\frac{\nu_{k,d}}{\nu_{k-d,d}} > \frac{\|x - x_k\|_A^2}{\|x - x_{k-d}\|_A^2} \quad \Leftrightarrow \quad \frac{\|x - x_k\|_A}{\|x - x_{k-d}\|_A} > \frac{\|x - x_{k+d}\|_A}{\|x - x_k\|_A}.$$

Another approach

superliner, linear, sublinear CG convergence

Let $\alpha > \beta > \gamma > 0$. Then

$$\frac{\beta - \gamma}{\alpha - \beta} > \frac{\beta}{\alpha} \quad \Leftrightarrow \quad \frac{\beta}{\alpha} > \frac{\gamma}{\beta}.$$

Apply to $\|x - x_{k-d}\|_A^2 > \|x - x_k\|_A^2 > \|x - x_{k+d}\|_A^2$. Then

$$\frac{\nu_{k,d}}{\nu_{k-d,d}} > \frac{\|x - x_k\|_A^2}{\|x - x_{k-d}\|_A^2} \quad \Leftrightarrow \quad \frac{\|x - x_k\|_A}{\|x - x_{k-d}\|_A} > \frac{\|x - x_{k+d}\|_A}{\|x - x_k\|_A}.$$

CG convergence:

- **linear** $\rightarrow$ **good** approximation
- **superliner** $\rightarrow$ **upper** bound
- **sublinear** $\rightarrow$ **lower** bound

Yet another approach

decrease formula

For $\ell + e < k$ and $d > 0$, one can relate

$$\frac{\|x - x_{k+d}\|_A^2}{\|x - x_k\|_A^2} \quad \text{and} \quad \frac{\|x - x_{\ell+e}\|_A^2}{\|x - x_\ell\|_A^2} \; .$$

Yet another approach

decrease formula

For $\ell + e < k$ and $d > 0$, one can relate

$$\frac{\|x - x_{k+d}\|_A^2}{\|x - x_k\|_A^2} \qquad \text{and} \qquad \frac{\|x - x_{\ell+e}\|_A^2}{\|x - x_\ell\|_A^2} \,.$$

It holds that

$$\frac{\|x - x_{k+d}\|_A^2}{\|x - x_k\|_A^2} = \left(1 - \frac{1}{\left(1 - \frac{1}{\frac{\|x - x_{\ell+e}\|_A^2}{\|x - x_\ell\|_A^2}}\right) \frac{\sum_{j=\ell+e}^{k-1} g_j}{\sum_{j=\ell}^{\ell+e-1} g_j} + 1} \right) \frac{\sum_{j=k}^{k+d-1} g_j}{\sum_{j=\ell+e}^{k-1} g_j} + 1$$

where

$$g_j = \alpha_j \|r_j\|^2 \,.$$

Conclusions

- A **new bound** on the A-norm of the error:
  - **simple**, comparable with the Gauss-Radau bound,
  - it can be used even for $\mu > \lambda_{\min}$ (heuristics).

Conclusions

- A **new bound** on the A-norm of the error:
  - **simple**, comparable with the Gauss-Radau bound,
  - it can be used even for $\mu > \lambda_{\min}$ (heuristics).
- Cheap **approximations** of extreme Ritz values
  - can be used in the new "upper" **bound** $\mu \longleftrightarrow \mu_k$,
  - using these approximations $\rightarrow$ estimate other **characteristics**.

Conclusions

- A **new bound** on the A-norm of the error:
  - **simple**, comparable with the Gauss-Radau bound,
  - it can be used even for $\mu > \lambda_{\min}$ (heuristics).
- Cheap **approximations** of extreme Ritz values
  - can be used in the new "upper" **bound** $\mu \longleftrightarrow \mu_k$,
  - using these approximations $\rightarrow$ estimate other **characteristics**.
- Strategy for **improving accuracy** of the error estimates
  - based on the **decrease** of $\|x - x_k\|_A$ in d iterations,
  - three approaches, how to choose d **adaptively**.

Conclusions

- A **new bound** on the A-norm of the error:
  - **simple**, comparable with the Gauss-Radau bound,
  - it can be used even for $\mu > \lambda_{\min}$ (heuristics).
- Cheap **approximations** of extreme Ritz values
  - can be used in the new "upper" **bound** $\mu \longleftrightarrow \mu_k$,
  - using these approximations $\rightarrow$ estimate other **characteristics**.
- Strategy for **improving accuracy** of the error estimates
  - based on the **decrease** of $\|x - x_k\|_A$ in d iterations,
  - three approaches, how to choose d **adaptively**.

Future work $\rightarrow$ **combine** approaches, numerical experiments

Related papers

- G. H. Golub and Z. Strakoš, [Estimates in quadratic formulas, Numer. Algorithms, 8 (1994), pp. 241–268.]
- G. Meurant and P. Tichý, [Practical estimation of the A-norm of the error in CG, to be submitted soon, 2018]
- 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]
- Z. Strakoš and P. Tichý, [On error estimation in CG and why it works in FP computations, Electron. Trans. Numer. Anal., 13 (2002), pp. 56–80.]

Thank you for your attention!