On error estimation in CG

Petr Tichý

Charles University, Prague

based on joint work with

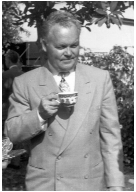
Gérard Meurant, Jan Papež, Zdeněk Strakoš

PANM 21, June 19-24, 2022, Merkur

The conjugate gradient method

Magnus Hestenes and Eduard Stiefel

1906 – 1991



1909 – 1978

Shiny, brand-new toys: SWAC (Hestenes) and Z4 (Stiefel).

Two existing classes of algorithms:

- direct methods,
- stationary iterative methods.

Need for an ideal algorithm (finite termination, if **stopped early**, would give a useful approximation) $\rightarrow$ [Hestenes, Stiefel 1952].

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

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

Orthogonality

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

Coefficients $\rightarrow R_k$

$$\begin{bmatrix} \frac{1}{\sqrt{\gamma_0}} & \sqrt{\frac{\delta_1}{\gamma_0}} & & \\ & \ddots & \ddots & \\ & & \ddots & \sqrt{\frac{\delta_{k-1}}{\gamma_{k-2}}} \\ & & & \frac{1}{\sqrt{\gamma_{k-1}}} \end{bmatrix}$$

The Lanczos algorithm

Let A be symmetric, compute orthonormal basis of $\mathcal{K}_k(A, b)$

input A, b

$v_1 = b/\|b\|$

$\beta_0 = 0,\ v_0 = 0$

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

$\quad \alpha_k = v_k^T A v_k$

$\quad w = A v_k - \alpha_k v_k - \beta_{k-1} v_{k-1}$

$\quad \beta_k = \|w\|$

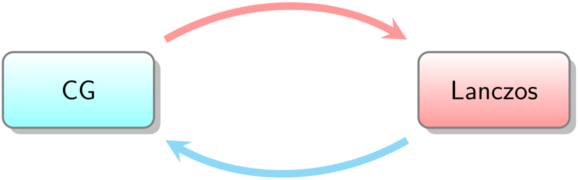
$\quad v_{k+1} = w/\beta_k$

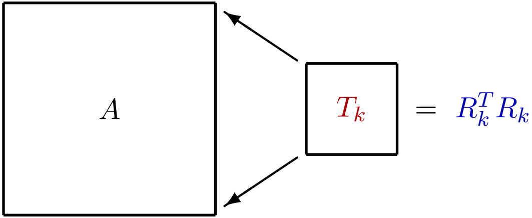
end for

$$T_k$$

$$\begin{bmatrix} \alpha_1 & \beta_1 & & \\ \beta_1 & \ddots & & \\ & & \ddots & \beta_{k-1} \\ & & \beta_{k-1} & \alpha_k \end{bmatrix}$$

$$V_k^* V_k = I$$


CG
Lanczos


A
$T_k = R_k^T R_k$

Optimality of CG

- CG as a **projection** method

$$x - x_k \perp_A \mathcal{K}_k(A, b)\,.$$

- CG as an **optimization** method

$$\mathcal{F}(y) \equiv \frac{1}{2}\, y^T A y - y^T b, \qquad \mathcal{F}(x_k) = \min_{y \in \mathcal{K}_k} \mathcal{F}(y)\,.$$

Note that

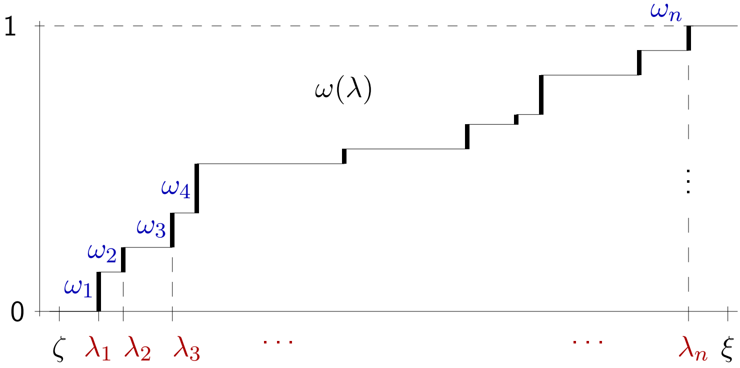
$$\mathcal{F}(y) = \frac{1}{2} \|x - y\|_A^2 + \mathcal{F}(x)\,.$$

- CG as Gauss **quadrature** .

CG and Gauss quadrature

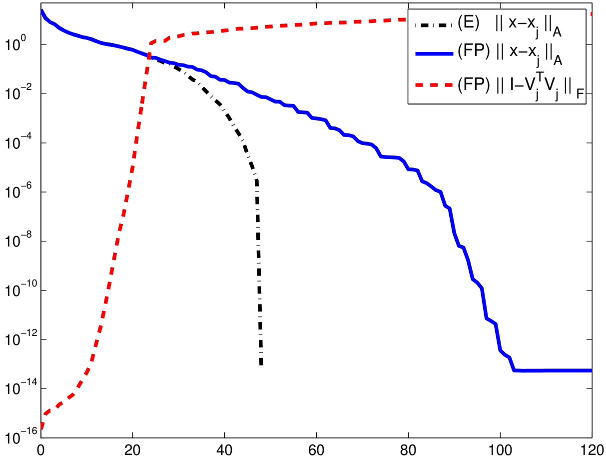
CG determines **weights** and **nodes** of the Gauss quadrature rule

$$\int_{\zeta}^{\xi} f(\lambda)\, d\omega(\lambda) \;=\; \sum_{i=1}^{k} \omega_i^{(k)} f(\theta_i^{(k)}) \;+\; \mathcal{R}_k[f]$$



CG in finite precision arithmetic

Orthogonality is lost, convergence is delayed!



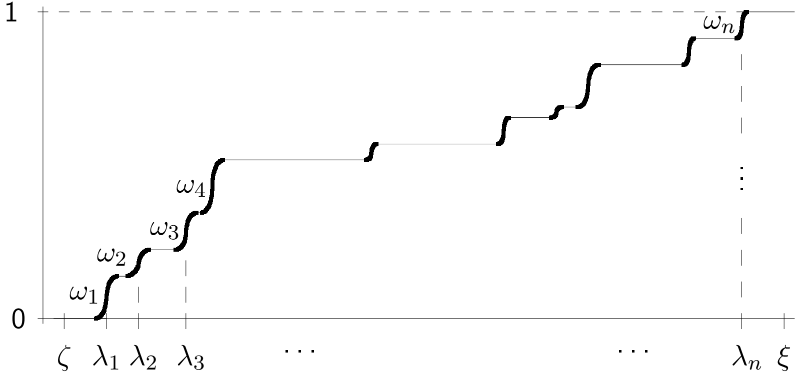
Identities need not hold in finite precision arithmetic!

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, Paige 1976, 1980]



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."

The normwise backward error

Given x_k, what are the norms of the smallest perturbations ΔA of A and Δb of b (in the relative sense) such that

$$(A + \Delta A)\, x_k = b + \Delta b\,?$$

We are interested in the quantity

$$\min\left\{\eta : \ (A + \Delta A)\, x_k = b + \Delta b, \ \frac{\|\Delta A\|}{\|A\|} \leq \eta, \ \frac{\|\Delta b\|}{\|b\|} \leq \eta\right\}$$

called the **normwise backward error**. It is given by

$$\frac{\|r_k\|}{\|A\|\|x_k\| + \|b\|}\,.$$

[Rigal, Gaches 1967]

CG with $\|A\|$ estimation

input A, b
$r_0 = b$, $p_0 = r_0$, $\delta_0 = 0$, $\gamma_{-1} = 0$, $c_1 = 1$
for $k = 1, \ldots,$ **do**
  `cgiter`(k)
  $\alpha_k = \frac{1}{\gamma_{k-1}} + \frac{\delta_{k-1}}{\gamma_{k-2}}$, $\beta_k^2 = \frac{\delta_k}{\gamma_{k-1}^2}$
  if $k = 1$ **then**
    $\rho_1 = \alpha_1$
  else
    $\omega_{k-1} = \sqrt{(\rho_{k-1} - \alpha_k)^2 + 4\beta_{k-1}^2 c_{k-1}}$
    $c_k = \frac{1}{2}\left(1 - \frac{\rho_{k-1} - \alpha_k}{\omega_{k-1}}\right)$
    $\rho_k = \rho_{k-1} + \omega_{k-1} c_k$
  end if
end for

$$\frac{\|r_k\|}{\|A\|\|x_k\| + \|b\|} \leq \frac{\|r_k\|}{\rho_k \|x_k\| + \|b\|}$$

[Meurant, T. 2019]

Estimating the A-norm of the error in CG

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

- **Estimating errors** using quadrature approach:

  [Dahlquist, Golub, Nash 1978], [Golub, Meurant 1994], [Golub, Strakoš 1994], [Golub, Meurant, 1997], [Calvetti et al. 2000], [Strakoš, T. 2002], [Meurant, T. 2013, 2019]

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

  [Golub, Strakoš 1994], [Strakoš, T. 2002, 2005, 2011]

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

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

An important role in stopping criteria

[Arioli 2004]

Find $u \in V = H_0^1(\Omega)$, such that

$$a(u, v) = f(v) \quad \forall v \in V$$

$a(\cdot, \cdot)$ is symmetric, bilinear, coercive, continuous.
Finite dimensional $V_h \subseteq V$, find $u_h \in V_h$ s.t.

$$a(u_h, v) = f(v) \quad \forall v \in V_h.$$

Considering basis functions of V_h, we get $Ax = b$, and

$$\|v\|_a^2 \equiv a(v, v) = \|y\|_A^2,$$

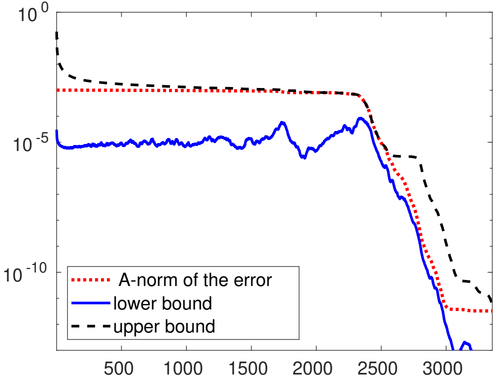
where $v \in V_h$ and y is the corresponding coordinate vector. Then

$$\underbrace{\|u - u_h^{(k)}\|_a^2}_{\text{total}} = \underbrace{\|u - u_h\|_a^2}_{\text{discretication}} + \underbrace{\|u_h - u_h^{(k)}\|_a^2}_{\text{algebraic}}.$$

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

Given $\mu \le \lambda_{\min}$,

$$\gamma_k \|r_k\|^2 < \varepsilon_k < \gamma_k^{(\mu)} \|r_k\|^2$$



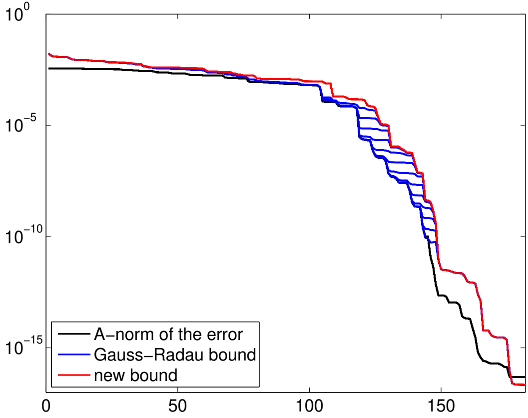
$$\varepsilon_k = \gamma_k \|r_k\|^2 + \varepsilon_{k+1}$$

Loss of accuracy of the Gauss-Radau upper bound

`bcsstk01`, $n = 48$, $\mu = \frac{\lambda_{\min}}{1+10^{-m}}$, $m = 2, 4, \dots, 14$

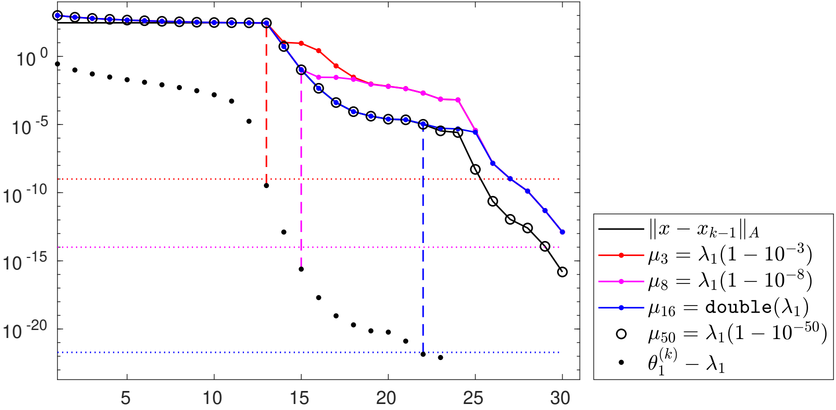
$$\|x - x_k\|_A \; < \; \sqrt{\gamma_k^{(\mu)}}\|r_k\| \; < \; \frac{\|r_k\|}{\sqrt{\mu}} \frac{\|r_k\|}{\|p_k\|}$$

[Meurant, T. 2019]



Loss of accuracy of the Gauss-Radau upper bound

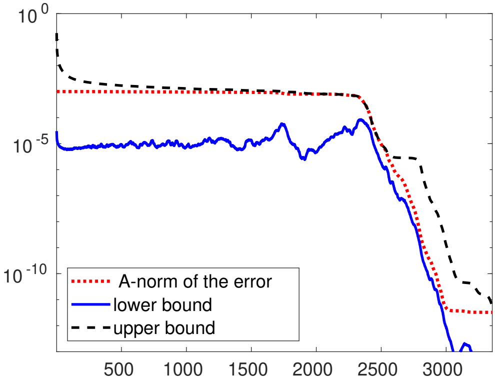
Work in progress ... [Meurant, T. 2022]



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

Given $\mu \leq \lambda_{\min}$,

$$\gamma_k \|r_k\|^2 < \varepsilon_k < \gamma_k^{(\mu)} \|r_k\|^2$$



How to **improve** and **control** the accuracy?

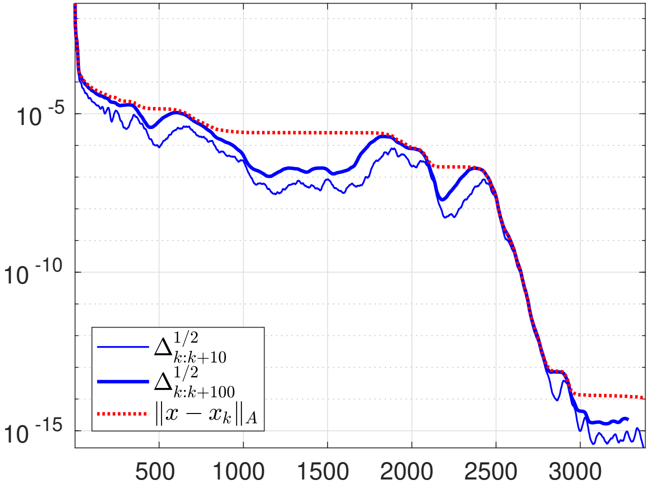
How to improve the accuracy of the estimates?

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

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

$\ell = k + d$ with a constant d

s3dkq4m2, $n = 90449$, `ichol`



A need to **determine d adaptively**.

Prescribing the accuracy of the estimate

$$\varepsilon_k = \Delta_{k:\ell-1} + \varepsilon_\ell$$

Ideally, we would like to determine $\ell > k$ such that

$$\frac{\varepsilon_k - \Delta_{k:\ell-1}}{\varepsilon_k} = \frac{\varepsilon_\ell}{\varepsilon_k} \leq \tau \, ,$$

where $\tau \in (0, 1)$ is a given tolerance. Then

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

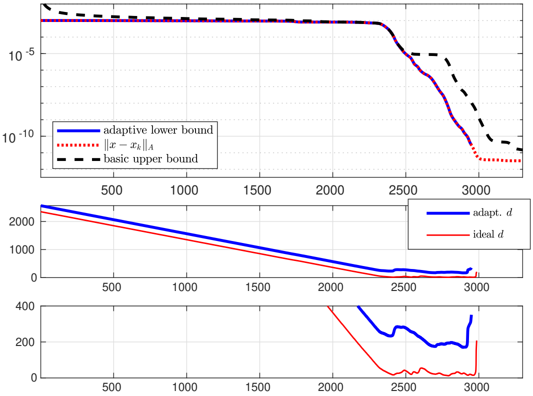
The delay $\ell - k$ should be as small as possible.

Find ℓ such that

$$\frac{\varepsilon_k - \Delta_{k:\ell-1}}{\varepsilon_k} = \frac{\varepsilon_\ell}{\varepsilon_k} \leq \tau$$

Using the upper bound

$$\frac{\varepsilon_\ell}{\varepsilon_k} \leq \frac{\gamma_\ell^{(\mu)} \|r_\ell\|^2}{\Delta_{k:\ell-1}} \leq \tau$$



Safe, but requires μ and, moreover, $\ell - k$ is far from being optimal!

Heuristic strategy

Learn from the history

It holds that [Meurant, Papež, T. 2021]

$$\Delta_j \; < \; \varepsilon_j \; < \; \kappa(A)\, \Delta_j \, .$$

Idea $\rightarrow$ find S_ℓ such that

$$\varepsilon_\ell \; \approx \; S_\ell \, \Delta_\ell .$$

Define

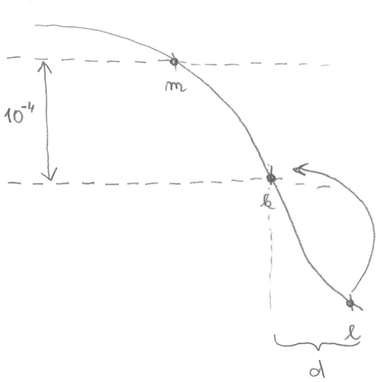
$$\widetilde{S}_j \; \equiv \; \frac{\Delta_{j:\ell}}{\Delta_j} \approx \frac{\varepsilon_j}{\Delta_j} \quad \rightarrow \quad S_\ell \; \equiv \; \max_{m \le j < \ell} \widetilde{S}_j \, .$$

Approximate

$$\frac{\varepsilon_\ell}{\varepsilon_k} \; \approx \; \frac{S_\ell \Delta_\ell}{\Delta_{k:\ell-1}} \; \le \; \tau \, .$$

How far to go into history?

Learn from the latest significant decrease



$$\varepsilon_k \;=\; \Delta_{k:\ell-1} \;+\; \varepsilon_\ell$$

- find m such that

$$\frac{\varepsilon_k}{\varepsilon_m} \approx \frac{\Delta_{k:\ell}}{\Delta_{m:\ell}} \leq 10^{-4}$$

- define

$$S_\ell \;=\; \max_{m \leq j < \ell} \widetilde{S}_j$$

- test

$$\frac{S_\ell \Delta_\ell}{\Delta_{k:\ell-1}} \leq \tau$$

Estimating ε_k with a prescribed accuracy τ

[Meurant, Papež, T. 2021]

input A, b, τ
$r_0 = p_0 = b$, $k = 0$
`cgiter`(0)
for $\ell = 1, \ldots,$ **do**
    `cgiter`(ℓ)
    compute $\Delta_{k:\ell-1}$ and Δ_ℓ
    determine S_ℓ
    while $\ell > k$ and $\frac{S_\ell \Delta_\ell}{\Delta_{k:\ell-1}} \leq \tau$ **do**
        accept $\Delta_{k:\ell}$
        $k = k + 1$
    end while
end for

Preconditioned CG (PCG) algorithm

$$\underbrace{L^{-1}AL^{-T}}_{\hat{A}}\underbrace{L^{T}x}_{\hat{x}} = \underbrace{L^{-1}b}_{\hat{b}}.$$

input A, b, x_0, $M = LL^T$
$r_0 = b - Ax_0$, $z_0 = M^{-1}r_0$, $p_0 = z_0$
for $k = 1, \ldots$ until convergence **do**

$$\left.\begin{aligned}
&\hat{\gamma}_{k-1} = \frac{z_{k-1}^T r_{k-1}}{p_{k-1}^T A p_{k-1}}\\
&x_k = x_{k-1} + \hat{\gamma}_{k-1} p_{k-1}\\
&r_k = r_{k-1} - \hat{\gamma}_{k-1} A p_{k-1}\\
&\text{Solve } M z_k = r_k\\
&\hat{\delta}_k = \frac{z_k^T r_k}{z_{k-1}^T r_{k-1}}\\
&p_k = z_k + \hat{\delta}_k p_{k-1}
\end{aligned}\right\} \quad \texttt{pcgiter(k)}$$

end for

$$\varepsilon_k = \sum_{j=k}^{\ell-1} \hat{\gamma}_j \, z_j^T r_j + \varepsilon_\ell$$

```
function [x,estim,delay] = pcga(A,b,tau,maxit,L,x)
r = b - A * x;
z = L\r; z = L'\z; p = z;
rr = z' * r;
k = 1;

for ell = 1:maxit+1

    RR    = rr;                    % ... begin cgiter(ell)
    Ap    = A * p;
    alpha = RR/(p' * Ap);
    x     = x + alpha * p;
    r     = r - alpha * Ap;
    z     = L \ r; z = L' \ z;
    rr    = z' * r;
    beta  = rr / RR;
    p     = z + beta * p;          % ... end cgiter(ell)

    Delta(ell) = alpha * RR;
    history(ell) = 0; history = history + Delta(ell);

    if ell > 1                     % ... adaptive choice of the delay
        S = findS(history,Delta,k);
        num = S * Delta(ell);
        den = sum(Delta(k:ell-1));
        while (ell > k) && (num/den <= tau)
            delay(k) = ell-k;
            estim(k) = den;
            k = k + 1;
            den = sum(Delta(k:ell-1));
        end
    end
end
end % of function

function [S] = findS(history,Delta,k)
ind = find((history(k)./history) <= 1e-4, 1, 'last');
if isempty(ind), ind = 1; end
S = max(history(ind:end-1)./Delta(ind:end-1));
end
```

Numerical experiments

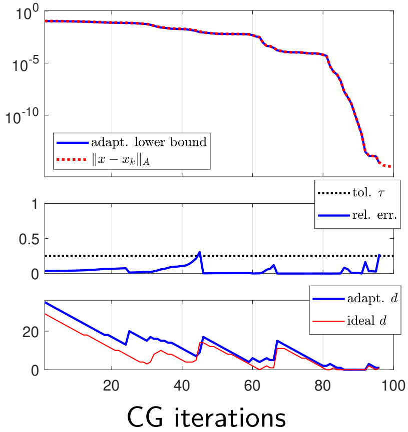
Test problems

SuiteSparse Matrix collection

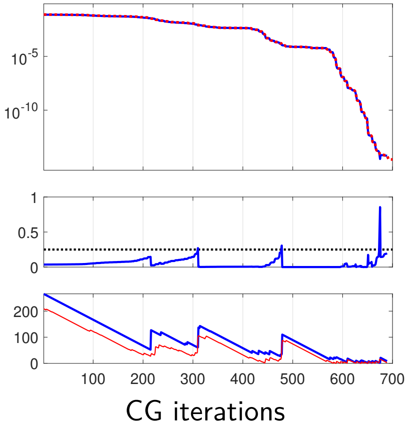
name	size	rhs b	$M = LL^T$
bcsstk02	66	equal components	—
bcsstk04	132	equal components	—
bcsstk09	1083	equal components	ict(1e-3, 1e-2)
s3dkt3m2	90 449	comes with the matrix	ict(1e-5, 1e-2)
s3dkq4m2	90 449	$\text{rand}(-1, 1)$	ict(1e-5, 1e-2)
pwtk	217 918	$\text{rand}(-1, 1)$	ict(1e-5, 1e-1)
af_shell3	504 855	$\text{rand}(-1, 1)$	zero-fill
tmt_sym	726 713	$\text{rand}(-1, 1)$	zero-fill
ldoor	952 203	$\text{rand}(-1, 1)$	zero-fill

Problems without preconditioning

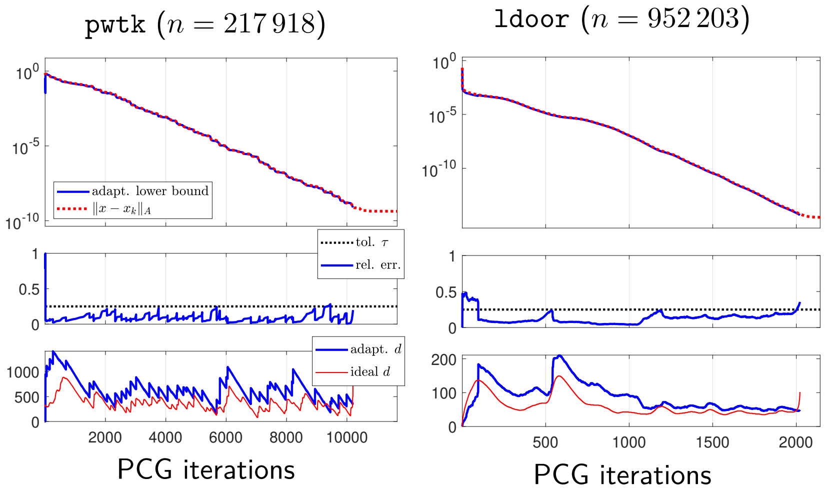
bcsstk02 $(n = 66)$



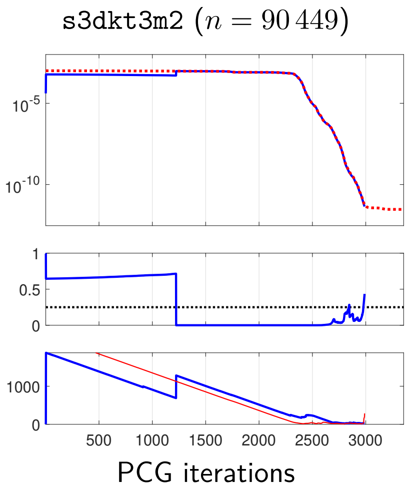
bcsstk04 $(n = 132)$

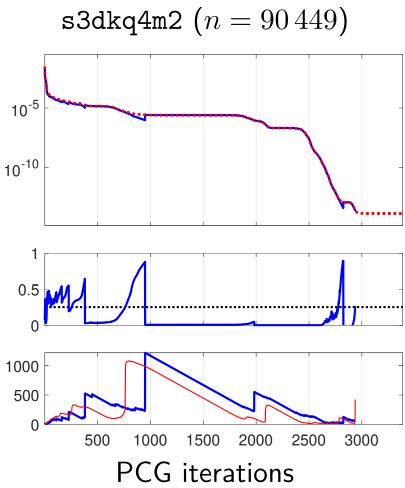


Problems with preconditioning



Difficult problems





Improved Gauss-Radau upper bound

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

Choose ℓ such that

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

Since

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

we can require

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

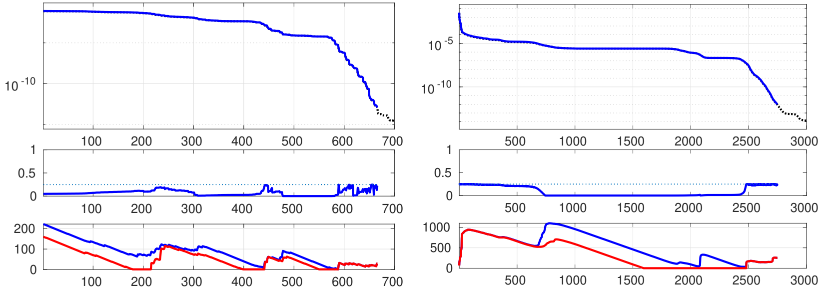
CG with the improved Gauss-Radau upper bound

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

Testing the improved upper bound

$$\frac{\Omega_{k:\ell}^{(\mu)} - \varepsilon_k}{\varepsilon_k} \leq \tau\,, \qquad \frac{\|r_\ell\|^2\,(\gamma_\ell^{(\mu)} - \gamma_\ell)}{\Delta_{k:\ell}} \leq \tau$$

`bcsstk04` $(n = 132)$ `s3dkq4m2` $(n = 90\,449)$



CG iterations PCG iterations

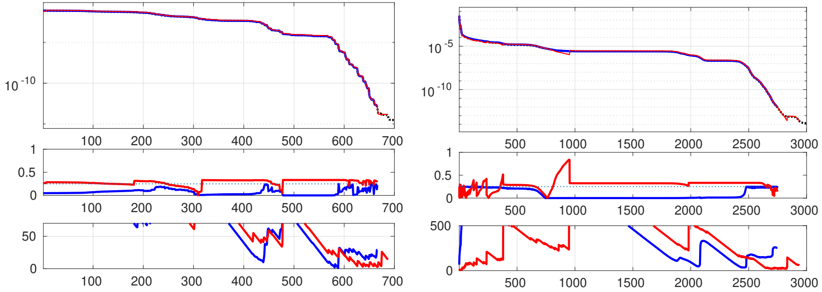
Comparison of upper bounds

$$\Omega_{k:\ell}^{(\mu)} \qquad \text{versus} \qquad \frac{\Delta_{k:\ell-1}}{1-\tau}$$

Comparison of upper bounds

$\Omega_{k:\ell}^{(\mu)}$ versus $\dfrac{\Delta_{k:\ell-1}}{1-\tau}$

`bcsstk04` `s3dkq4m2`



CG iterations PCG iterations

Conclusions

- One can **improve** the accuracy of estimates of ε_k using the information from the forthcoming CG iterations.
- We can **control** the accuracy of the estimates using:
  - Gauss-Radau **upper bound** $\rightarrow$ reliable, often not optimal.
  - A **heuristic strategy** $\rightarrow$ robust, often almost optimal.
- Generalization is possible for other CG-like methods.

Related papers

G. Meurant, J. Papež, and P. Tichý,

[Accurate error estimation in CG, Numer. Algorithms, 88 (2021), pp. 1337-1359.]

- G. H. Golub and Z. Strakoš, [Estimates in quadratic formulas, Numer. Algorithms, 8 (1994), pp. 241–268.]
- 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]
- 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!