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

Petr Tichý

joint work with

Gérard Meurant and Jan Papež

February 11, 2020, Prague

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

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

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

Estimating the A-norm of the error

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

- Error bounds, given $\mu \leq \lambda_{\min}$,

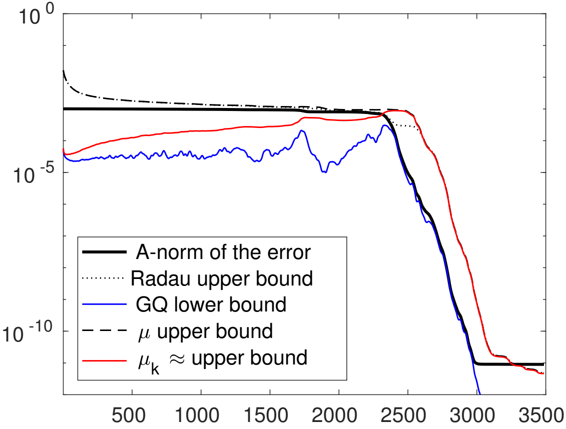
$$\alpha_k \|r_k\|^2 < \|x - x_k\|_A^2 < \alpha_k^{(\mu)} \|r_k\|^2 < \frac{\|r_k\|^2}{\mu} \frac{\|r_k\|^2}{\|p_k\|^2},$$

see, e.g., [Meurant, T. 2019] $\rightarrow$ approximate μ by μ_k.

Experiment

s3dkt3m2, $n = 90449$, `ichol`

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



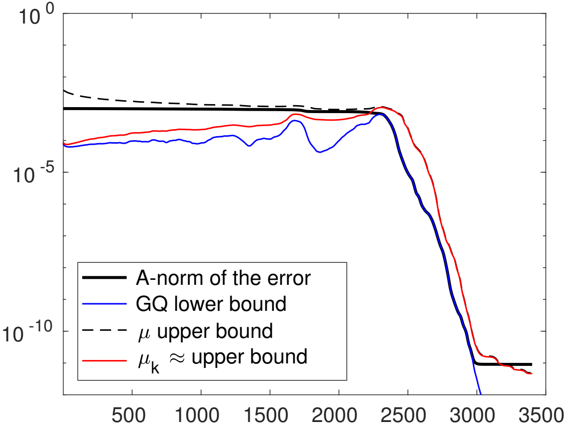
Improving accuracy of the lower bound

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

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

A constant value of d

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



How to choose d adaptively?

Lower bound

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

and

$$\nu_{k,d} \; < \; \|x - x_k\|_A^2 \; = \; \frac{\nu_{k,d}}{1 - \rho} \, .$$

Pseudo algorithm for the adaptive choice of d

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

input A, b, x_0, τ
$p_0 = r_0 = b - Ax_0$
$d = 0$
for $k = 1, \dots,$ **do**
    $[\alpha_{k-1}, \beta_k, x_k, r_k, p_k] =$ CG iteration(k)
    $d = d + 1$
    $\varrho = \frac{\varepsilon_k}{\varepsilon_{k-d}}$
    while $\varrho < \tau$ **do**
        compute lower bound at $k - d$
        $d = d - 1$
        $\varrho = \frac{\varepsilon_k}{\varepsilon_{k-d}}$
    end while
end for

Testing various approaches

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

- $$\frac{\varepsilon_j}{\varepsilon_k} \le \frac{\nu_{j,d}}{\nu_{k,d}} = \frac{\varepsilon_j - \varepsilon_{j+d}}{\varepsilon_k - \varepsilon_{k+d}} \quad \Leftrightarrow \quad \frac{\varepsilon_{j+d}}{\varepsilon_{k+d}} \le \frac{\varepsilon_j}{\varepsilon_k}\,.$$

Testing various approaches

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

- $$\frac{\varepsilon_j}{\varepsilon_k} \leq \frac{\nu_{j,d}}{\nu_{k,d}} = \frac{\varepsilon_j - \varepsilon_{j+d}}{\varepsilon_k - \varepsilon_{k+d}} \quad \Leftrightarrow \quad \frac{\varepsilon_{j+d}}{\varepsilon_{k+d}} \leq \frac{\varepsilon_j}{\varepsilon_k}\,.$$

The last inequality holds if convergence is at least **linear**.

Testing various approaches

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

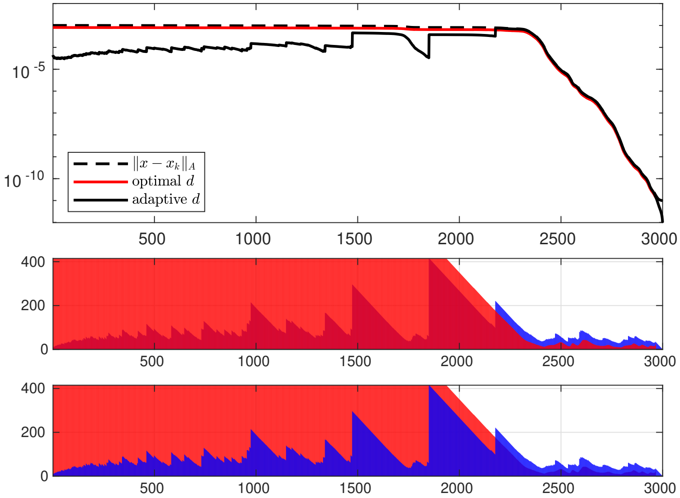
- $$\frac{\varepsilon_j}{\varepsilon_k} \leq \frac{\nu_{j,d}}{\nu_{k,d}} = \frac{\varepsilon_j - \varepsilon_{j+d}}{\varepsilon_k - \varepsilon_{k+d}} \quad \Leftrightarrow \quad \frac{\varepsilon_{j+d}}{\varepsilon_{k+d}} \leq \frac{\varepsilon_j}{\varepsilon_k}\,.$$

  The last inequality holds if convergence is at least **linear**.

- It holds that
  $$\frac{\varepsilon_j}{\varepsilon_k} \leq \frac{\kappa(A)\,\alpha_j \|r_j\|^2}{\nu_{k,d}},$$
  approximate $\kappa(A)$ [Meurant, T. 2019], heuristic $\to \kappa(A)^{1/4}$.

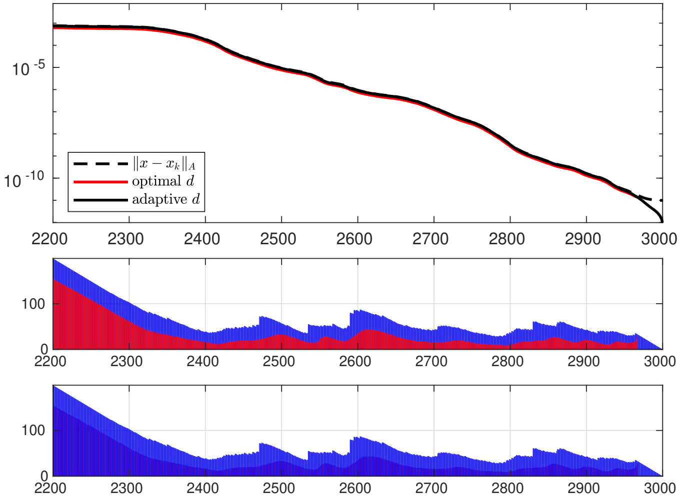
Experiment

s3dkt3m2



Experiment

s3dkt3m2



Related papers

- G. H. Golub and Z. Strakoš, [Estimates in quadratic formulas, Numer. Algorithms 8, 1994]
- G. Meurant, J. Papež and P. Tichý, [Controlling accuracy of the error estimates in CG, in preparation, 2020]
- G. Meurant and P. Tichý, [Approximating the extreme Ritz values and upper bounds for the A-norm of the error in CG, Numer. Algorithms 82, 2019]
- G. Meurant and P. Tichý, [On computing quadrature-based bounds for the A-norm of the error in CG, Numer. Algorithms 62, 2013]
- Z. Strakoš and P. Tichý, [On error estimation in CG and why it works in FP computations, Electron. Trans. Numer. Anal. 13, 2002]

Thank you for your attention!