

Error estimate and stopping criteria for least-squares problems solved by CG-like algorithms CGLS and LSQR

Jan Papež, Petr Tichý

Abstract

In [2], we presented an adaptive estimate for the energy norm of the error in the conjugate gradient (CG) method. Using the notation from [2, Alg. 1], A -norm of the error between the exact solution of $Ax = b$ and the CG approximation x_ℓ given in the ℓ th step is estimated as

$$\|x - x_\ell\|_A^2 \approx \Delta_{\ell:k}^{\text{CG}} := \sum_{j=\ell}^k \alpha_j \|r_j\|^2, \quad (1)$$

where $\|v\|_A^2 \equiv v^T A v$ denotes the squared A -norm. Integrating the estimate into the existing CG code is straightforward and simple; see [4, Alg. 1]. At the current k th CG iteration, we get an estimate with the delay $d = k - \ell$ for previous approximation x_ℓ . The delay d is set adaptively by [4, Alg. 2]. From the construction, $\Delta_{\ell:k}^{\text{CG}}$ yields a lower bound

$$\|x - x_\ell\|_A^2 \geq \Delta_{\ell:k}^{\text{CG}}.$$

In [4] and in the prospective talk at Householder Symposium XXII we consider algorithms for solving least-squares problems with a general, possibly rectangular matrix

$$\min_{x \in \mathbb{R}^m} \|b - Ax\|, \quad b \in \mathbb{R}^n, A \in \mathbb{R}^{n \times m}, n \geq m,$$

that are mathematically based on applying CG to a system with a positive (semi-)definite matrix $A^T A$. We discuss CGLS based on Hestenes–Stiefel-like implementation as well as LSQR based on Golub–Kahan bidiagonalization, and both unpreconditioned and preconditioned variants. We show that the adaptive estimate used in CG can be extended for these algorithms to estimate the monotonically decreasing quantity

$$\|x - x_\ell\|_{A^T A}^2 = \|r_\ell\|^2 - \|r\|^2, \quad (2)$$

where $x = A^\dagger b$ is the minimal norm solution, x_ℓ is the approximation in the ℓ th step of CGLS or LSQR, $r_\ell = b - Ax_\ell$, and $r = b - b_{|\mathcal{R}(A)}$ with $b_{|\mathcal{R}(A)}$ being the orthogonal projection of b onto the range of A .

For example, the estimate

$$\|x - x_\ell\|_{A^T A}^2 \approx \Delta_{\ell:k}^{\text{LSQR}} := \sum_{j=k}^k \phi_{j+1}^2,$$

for estimating the quantity of interest (2) in LSQR algorithm is given, analogously to $\Delta_{\ell:k}^{\text{CG}}$, as a sum of scalar terms, which are available in the algorithm; here we use the notation from [4, Alg. 4]. Moreover, $\Delta_{\ell:k}^{\text{LSQR}}$ provides a lower bound on $\|x - x_\ell\|_{A^T A}^2$.

We emphasize the applicability of the estimates (bounds) for the computations in finite-precision arithmetic. Their derivation is only based on local orthogonality, which is typically well preserved

during computations; see [5]. We demonstrate that the estimates remain computationally inexpensive to evaluate and are numerically reliable in finite-precision arithmetic under mild assumptions. These qualities make the estimates highly suitable for stopping the iterations.

One can consider the stopping criterion requiring that

$$\frac{\|r_\ell\|^2 - \|r\|^2}{\|r\|^2} \leq \varepsilon \quad (3)$$

for a prescribed tolerance ε . It is clear from (2), that after $\|r_\ell\| \approx \|r\|$, further iterations bring no significant decrease of the residual norm $\|r_\ell\|$. Using (2), the criterion (3) is equivalent to

$$\|x - x_\ell\|_{A^T A}^2 \leq \frac{\varepsilon}{1 - \varepsilon} \|r_\ell\|^2,$$

where the estimate for $\|x - x_\ell\|_{A^T A}^2$ can be used.

Another stopping criterion based on a backward error can also be considered when applying our estimates. This criterion aims to identify the iteration at which the computed approximation can be interpreted as the least-squares solution to a perturbed system

$$\min_x \|(b + f) - (A + E)x\|,$$

with

$$\min_{f, E, \zeta} \{\zeta \text{ such that } \|E\| \leq \zeta \|A\|, \|f\| \leq \zeta \|b\|\} \leq \varepsilon.$$

This backward error for stopping LSQR iterations has been studied, e.g., in [3, 1], and can be approximated using the asymptotically tight bound

$$\frac{\|x - x_\ell\|_{A^T A}^2}{\|A\| \|x_\ell\| + \|b\|};$$

see [1].

Finally, we present a range of numerical experiments to confirm the robustness and very satisfactory behaviour of the estimates for CGLS, LSQR, and also their preconditioned variants. We hope that these estimate will prove to be useful in practical computations. They allow us to approximate, with the prescribed relative accuracy, the quantity of interest at a negligible cost.

References

- [1] X.-W. Chang, C. C. Paige, and D. Tittle-Peloquin. Stopping criteria for the iterative solution of linear least squares problems. *SIAM J. Matrix Anal. Appl.*, 31(2):831–852, 2009.
- [2] G. Meurant, J. Papež, and P. Tichý. Accurate error estimation in CG. *Numer. Algorithms*, 88(3):1337–1359, 2021.
- [3] C. C. Paige and M. A. Saunders. LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Software*, 8(1):43–71, 1982.
- [4] J. Papež and P. Tichý. Estimating error norms in CG-like algorithms for least-squares and least-norm problems. *Numer. Algorithms*, 97(1):1–28, 2024.
- [5] Z. Strakoš and P. Tichý. On error estimation in the conjugate gradient method and why it works in finite precision computations. *Electron. Trans. Numer. Anal.*, 13:56–80, 2002.