The use of inner preconditioned conjugate gradient iterations in large sparse nonlinear optimization problems

I. E. Kaporin

Computing Center of Russian Academy of Sciences, Vavilova 40, Moscow 117967, Russia

E-mail: kaporin@sci.kun.nl, kaporin@ccas.ru

In our presentation, we consider the problem of the (sub)optimum termination of inner linear Preconditioned Conjugate Gradient (PCG) iterations used within Inexact Newton nonlinear solvers [2, 4, 3].

Let F be a differentiable mapping $F : R^n \to R^m, m \geq n$, and let the nonlinear least squares problem $u_* = \arg\inf \|Fu\|$ be solved using inexact Gauss-Newton iterations. Each iteration of the method can be presented as follows: given an approximation u, one calculates the matrix $A = F'(u)^T F'(u)$, the right hand side $b = -F'(u)^T Fu$, then performs some inexact solution of the arising linear system, so that $Ax \approx b$ and then obtains the next approximation as $u_+ = u + \tau x$ choosing some proper steplength $\tau > 0$. It was shown in [4] that if the approximate Gauss-Newton direction x is properly scaled, that is, $b^T x = x^T Ax$, then the estimate

$$\|F(u_+)\| \leq \sqrt{1 - \tau\theta^2}\|Fu\|,$$

where

$$\theta = |b^T x| / \left(\|Fu\|\sqrt{x^T Ax}\right)$$

holds for some positive $\tau \leq 1$, the maximum admissible value of which is considered as the key characterization of the local nonlinearity of the mapping F. If F is a linear mapping, one can always take $\tau = 1$, and it makes sense to solve the linear problem for x to full precision. However, for the nonlinear problems the case of $\tau < 1$ always takes place, so one should make a choice for a proper termination of the PCG iterations when solving for x.

We will describe here a criterion for stopping PCG iterations which is independent of any characterization of nonlinearity and is expressed in terms of scalar coefficients involved in PCG recursions and the ratio of the PCG startup cost to a regular PCG iteration cost. Our strategy aimed *to maximization of* θ rather than to minimization of the residual $\|b - Ax\|$. It worth noting here that there exists no direct relation between the usual relative residual $\|b - Ax\|/\|b\|$ and the above quantity θ.

Let us recall the PCG algorithm. The PCG iterations [1] for the solution of the problem $Ax = b$ can be written as follows:

$$\begin{aligned} r_0 = b - Ax_0, \qquad & p_0 = Hr_0; \qquad \textbf{for } i = 0, 1, \ldots : \\ \alpha_i = r_i^T Hr_i / p_i^T Ap_i, \qquad & x_{i+1} = x_i + p_i\alpha_i, \qquad r_{i+1} = r_i - Ap_i\alpha_i, \\ \beta_i = r_{i+1}^T Hr_{i+1} / r_i^T Hr_i, \qquad & p_{i+1} = Hr_{i+1} + p_i\beta_i. \end{aligned}$$

Here H is a properly chosen SPD preconditioning matrix, which should approximate, in some sense, the matrix A^{-1}. The choice of the matrix H is subject to the requirement that a vector $w = Hr$ be easily calculated for any r. For instance, one of the best choices is the approximate Cholesky preconditioning, where $H = (U^T U)^{-1}$ and $U^T U \approx A$ with the upper triangular matrix U being much sparser than the exact Cholesky factor of A, cf.[5] and refernces therein.

Fortunately, if $x_0 = 0$ in the above method, then one can easily see (using the A-orthogonality property of p_i) that each iterate x_i obtained in the PCG iterations applied to $Ax = b$ satisfies the above scaling condition, and the quantity θ is given by the explicit formula

$$\theta = \|Fu\|^{-1} \Big(\sum_{j=0}^{i-1} \omega_j\Big)^{1/2}, \qquad \omega_i = (r_i^T Hr_i)^2 / p_i^T Ap_i, \quad i = 0, 1, \ldots, n-1,$$

where ω_i are readily available from the scalar products calculated in the course of the PCG iterations. Using this relationship, one can try to find a proper balance between the costs of initializing and performing the inner PCG iterations and the acceleration obtained in the outer nonlinear iterations due to larger values of θ. The inner iteration stopping criterion based on maximization of θ can be constructed as follows. Let m nonlinear iterations be performed, and let τ be the lower bound for the stepsizes used. Assuming that the nonlinear convergence is measured by $\|F(u_m)\| \leq \varepsilon \|F(u_0)\|$, one can therefore find that a sufficient condition for the nonlinear iterations to be converged can be taken as $\sum_{j=1}^{m} \theta_j^2 \approx \tau^{-1} \log\left(\varepsilon^{-1}\right)$. Let then $\mathcal{P}$ be the costs of the outer iteration (including at least the generation of b, A, and the preconditioner H) and $\mathcal{I}$ be the costs per one PCG iteration (determined mainly by the costs of matrix-vector multiplications with A and H). The total computational costs is estimated as

$$\sum_{j=1}^{m} (\mathcal{P} + \mathcal{I} k_j) \approx \left(\sum_{j=1}^{m} (\mathcal{P} + \mathcal{I} k_j) / \sum_{j=1}^{m} \theta_j^2 \right) \tau^{-1} \log\left(\varepsilon^{-1}\right) \leq \left(\max_{1 \leq j \leq m} (\mathcal{P} + \mathcal{I} k_j)\, \theta_j^{-2} \right) \tau^{-1} \log\left(\varepsilon^{-1}\right),$$

where k_j is the number of inner iterations at the j-th outer iteration step. The stopping criterion can now be formulated in terms of choosing the inner iteration numbers k_j providing for a reasonably small value of each ratio $(\mathcal{P} + \mathcal{I} k_j)\, \theta_j^{-2} = \|F u_j\|^2 (\mathcal{P} + \mathcal{I} k_j) / \left(\omega_0 + \ldots + \omega_{k_j}\right)$. In our experiments we choose the value k_j for which the increase of this ratio occured for the first time, i.e. the iterations were performed until the condition

$$k \geq -(\mathcal{P}/\mathcal{I}) + (\omega_0 + \ldots + \omega_k)/\omega_k$$

holds true. In view that $\omega_0 + \ldots + \omega_k$ is bounded from above by the squared A-norm of the solution x, and ω_k tend to decrease as the PCG iterations progress, one can expect rather early termination of the inner iterations. The above inner iteration stopping criterion was successfully used in numerical experiments reported in [2]. We will also suppose to present some numerical experiments with global untangling of computational grids via continuation technique for highly nonlinear variational problem confirming the practical efficiency of the proposed PCG iteration stopping criterion.

Acknowledgement. This work has been supported by the Netherlands Technology Foundation (STW) grant NNS.4683 and by the National Science Foundation grant NWO 047.008.007.

References

[1] O. Axelsson. A class of iterative methods for finite element equations. *Computer Meth. Appl. Mech. Engrg.*, 9, 123-137, 1976.

[2] O. Axelsson and I. E. Kaporin. Minimum residual adaptive multilevel procedure for the finite element solution of nonlinear stationary problems. *SIAM J. Numer. Anal.*, 35, 1213-1229, 1998.

[3] O. Axelsson and I. E. Kaporin. Error norm estimation and stopping criteria in preconditioned conjugate gradient iterations, *Numer. Linear Algebra Appls.* 8, 265-286, 2001.

[4] I. E. Kaporin and O.Axelsson. On a class of nonlinear equation solvers based on the residual norm reduction over a sequence of affine subspaces. *SIAM J. Sci. Comput.*, 16, 228-249, 1994.

[5] I. E. Kaporin. High quality preconditioning of a general symmetric positive definite matrix based on its $U^T U + U^T R + R^T U$-decomposition. *Numer. Linear Algebra Appls.* 5, 483-509, 1998.