Communications to SIMAI Congress,
ISSN 1827-9015, Vol. 1 (2006)

DOI: 10.1685/CSC06076

An approximate inverse preconditioner in truncated Newton methods for large scale optimization

Giovanni Fasano
Istituto Nazionale per Studi ed Esperienze di Architettura Navale INSEAN
via di Vallerano, 139 – Roma, ITALY
g.fasano@insean.it

Massimo Roma
Dipartimento di Informatica e Sistemistica "A. Ruberti"
Università di Roma "La Sapienza"
via Buonarroti, 12 – Roma, ITALY
roma@dis.uniroma1.it

This work deals with the solution of large scale unconstrained optimization problems, namely the problems of finding a local minimizer of a real valued function $f : \mathbb{R}^n \to \mathbb{R}$, when the number of variables n is large. The function f is assumed to be twice continuously differentiable.

The growing interest in solving large scale problems is mainly due to the fact that problems with larger and larger number of variables arise very frequently in many and different contexts and real world applications (see Ref. 8 and Ref. 9).

We focus on Truncated Newton (TN) methods which represent one of the most powerful and flexible tools for solving such problems, having a sound convergence theory and showing great robustness and efficiency (see, e.g. Ref. 7). Notwithstanding these methods have been widely studied and extensively tested, two key aspects can be still considered open questions: how to handle the case with indefinite Hessian and how to formulate a general effective preconditioning strategy.

As well known, at each iteration (*outer iteration*) of a TN algorithm, a Krylov subspace method is usually used for determining the search direction d_k, by approximately solving the Newton system

$$H_k d = -g_k, \tag{0.1}$$

where $H_k = \nabla^2 f(x_k)$ and $g_k = \nabla f(x_k)$.

The Conjugate Gradient (CG) is often the method of choice, though it is a suitable solver of linear systems (0.1) in the case of positive definite Hessian. However, additional safeguard is needed in dealing with nonconvex problems; in fact, CG iterations (*inner iterations*) may break down before satisfying a termination criterion whenever the Hessian matrix is not positive definite.

On the other hand, a suited preconditioning strategy is often mandatory in dealing with large scale problems for efficiently solving the systems (0.1), thus obtaining a significative reduction of the overall number of inner CG iterations. Most of the available preconditioners are unsuited in this context, since they require the knowledge of the actual elements of the Hessian matrix. On the contrary in the framework of TN methods, the entries of the matrix H_k are supposed to be unavailable. In fact, in the large scale setting it is not possible either to store or handle any matrix, and the information

on the Hessian are usually gained by means of a routine, which provides us with the matrix–vector product of the Hessian matrix times a vector.

On the basis of these observations, in this work we propose the use of Conjugate Gradient–based schemes as a tool for facing up to both the questions. In particular, we consider new CG planar schemes recently proposed in Ref. 3 and Ref. 4, in order to overcome, in a unified framework, a couple of drawbacks of the standards CG. We are concerned with getting through the difficulties in dealing with the indefinite case and in building a suited preconditioner, which does not rely on the explicit knowledge of the Hessian matrix.

The planar CG algorithm we consider is a modification of the one proposed by Hestenes (Ref. 5). We refer to Ref. 3 for a detailed description of the algorithm. We briefly recall that, in respect to CG, the planar CG methods prevent from untimely stopping the detection of a stationary point of the quadratic function $q(d) = 1/2\ d^T H_k d + g_k^T d$. This is achieved by generating mutually conjugate directions or planes. In particular, as long as the matrix H_k is indefinite and nonsingular, by applying the standard CG, a sequence of conjugate directions $\{p_i\}$ is generated and a pivot breakdown occurs at the i-th step if $p_i^T H_k p_i = 0$. On the contrary, in the latter case, the planar CG methods generate another direction q_i and perform a search on the 2-dimensional manifold $d_k + \text{span}\{p_i, q_i\}$.

Moreover, we point out that any CG scheme provides useful information, as a by-product of the CG iterations, which can be exploited for constructing a preconditioner. In particular, we aim at defining an adaptive preconditioning technique based on the Krylov subspace information. The basic idea is to deflate the eigenvalues of the matrix of the system (0.1) associated with the invariant subspace explored by the CG.

We recall that after $h \leq n$ steps of the CG method, h orthogonal vectors, say $r_1, \ldots, r_h$, are generated such that $R_h^T H_k R_h = T_h$, where $R_h = (r_1/\|r_1\| \ \cdots \ r_h/\|r_h\|)$ and T_h is a $h \times h$ tridiagonal irreducible matrix. Moreover, if H_k is positive definite, T_h can be stably decomposed as $T_h = L_h D_h L_h^T$, where L_h is a unit lower diagonal matrix and D_h is diagonal and positive definite. Furthermore, by setting $P_h = (p_1/\|r_1\| \ \cdots \ p_h/\|r_h\|)$ it results $P_h L_h^T = R_h$.

If H_k is indefinite, the factorization of the tridiagonal matrix T_h may fail, in the sense that it may not exist or may be very unstable. Planar–CG schemes enable to overcome this drawback by means of planar steps. Indeed, the tridiagonal matrix T_h is now indefinite and can be decomposed as $L_h^T D_h L_h$, where D_h is now a 1×1 or 2×2 block diagonal matrix.

In order to define the preconditioner for the indefinite case, we recall that the columns of R_h span the Krylov subspace $\mathcal{K}_h(H_k, r_1)$. Then, introducing the matrix R_{n-h}, whose columns span $\mathbb{R}^n \setminus \text{span}\{R_h\}$, with $R_{n-h}^T R_{n-h} = I_{n-h}$, let us define the following matrix

$$M_h = R_h\, |T_h|\, R_h^T + R_{n-k} R_{n-h}^T = R_h\, |T_h|\, R_h^T + I - R_h R_h^T,$$

where $|T_h| = L_h |D_h| L_h^T$ and $|D_h|$ is a 1×1 or 2×2 positive definite block diagonal matrix.

We use this matrix (with $h \ll n$) as preconditioner. Assuming that the matrix H_k is nonsingular, then M_h is nonsingular too, and it results $M_h^{-1} = R_h |T_h|^{-1} R_h^T + R_{n-h} R_{n-h}^T$. Moreover, the spectrum of the preconditioned matrix $M_h^{-1} H_k$ contains the eigenvalue 1 and -1 with overall multiplicity at least h.

This is not the first attempt in the literature to "recycle" Krylov information generated during the iterations, in order to construct a preconditioner. In fact, a similar approach

was considered in the context of GMRES algorithms (see Ref. 1, Ref. 2, Ref. 6). However, in this approach the GMRES information are given in the form of a Hessenberg decomposition of the matrix H_k and not as tridiagonal one. This is a distinguishing important feature of our approach. In fact, by using the tridiagonal decomposition we can rewrite M_h^{-1} as

$$M_h^{-1} = P_h |D_h|^{-1} P_h^T + I - P_h L_h^T L_h P_h^T.$$

Therefore, in computing the preconditioned residuals $M_h^{-1} r_i$, it suffices to store the h vectors of P_h, the block diagonal elements of $|D_h|^{-1}$ and the subdiagonal elements of L_h, all of them available as by product from the CG iterates. This implies that no $n \times n$ matrix is stored and no explicit matrix inversion is needed to compute the preconditioned residuals.

The resulting algorithm can be easily embedded in a truncated Newton method in order to efficiently determine a good Newton type direction at each outer iteration k.

We performed some preliminary numerical experiments, by implementing a linesearch based truncated Newton method, which carries out the preconditioning strategy we propose. The results obtained, show that this strategy may lead to an improvement of the overall efficiency of the method, even though no definitive conclusions can be drawn.

REFERENCES

1. J. Baglama, D. Calvetti, G.H. Golub, and L. Reichel. Adaptively preconditioned GMRES algorithms. *SIAM Journal on Scientific Computing*, 20:243–269, 1998.

2. J. Erhel, K. Burrage, and B. Pohl. Restarted GMRES preconditioned by deflation. *Journal of Computational and Applied Mathematics*, 69:303–318, 1996.

3. G. Fasano. Planar–conjugate gradient algorithm for large–scale unconstrained optimization, Part 1: Theory. *Journal of Optimization Theory and Applications*, 125:523–541, 2005.

4. G. Fasano. Planar–conjugate gradient algorithm for large–scale unconstrained optimization, Part 2: Application. *Journal of Optimization Theory and Applications*, 125:543–558, 2005.

5. M.R. Hestenes. *Conjugate Direction Methods in Optimization*. Springer Verlag, New York, 1980.

6. L. Loghin, D. Ruiz, and A. Touhami. Adaptively preconditioners for nonlinear systems of equations. *Journal of Computational and Applied Mathematics*, 189:362–374, 2006.

7. S.G. Nash. A survey of truncated-Newton methods. *Journal of Computational and Applied Mathematics*, 124:45–59, 2000.

8. J. Nocedal. Large scale unconstrained optimization. In A. Watson and I. Duff, editors, *The state of the art in Numerical Analysis*, pages 311–338. Oxford University Press, 1997.

9. M. Roma. Large scale unconstrained optimization. In C. Floudas and P. Pardalos, editors, *Encyclopedia of Optimization*, volume III, pages 143–150. Kluwer Academic Publishers, 2001.