Journal of Computational Mathematics, Vol.21, No.3, 2003, 311–320.

TESTING DIFFERENT CONJUGATE GRADIENT METHODS FOR LARGE-SCALE UNCONSTRAINED OPTIMIZATION*[1)]

Yu-hong Dai
(*LSEC, ICMSEC, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing 100080, China*)

Qin Ni
(*College of Science, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*)

Abstract

In this paper we test different conjugate gradient (CG) methods for solving large-scale unconstrained optimization problems. The methods are divided in two groups: the first group includes five basic CG methods and the second five hybrid CG methods. A collection of medium-scale and large-scale test problems are drawn from a standard code of test problems, CUTE. The conjugate gradient methods are ranked according to the numerical results. Some remarks are given.

Key words: Conjugate gradient methods, Large-scale, Unconstrained optimization, Numerical tests.

1. Introduction

We consider the unconstrained optimization problem

$$\min f(x), \quad x \in \mathcal{R}^n, \tag{1}$$

where f is smooth and its gradient g is available. The line search method for solving (1) is of the form

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

where x_1 is a given initial point, d_k is a search direction, and α_k is a stepsize obtained by a 1-dimensional line search. In the steepest descent method [4], the search direction is defined as the negative gradient direction,

$$d_k = -g_k, \tag{3}$$

and the stepsize is chosen to be the 1-dimensional minimizer

$$\alpha_k = \arg\min_{\alpha>0} f(x_k + \alpha_k d_k). \tag{4}$$

In practical computations, however, the steepest descent method performs poorly, and is badly affected by ill-conditioning [2]. Another class of methods are quasi-Newton methods (see [23] for example), where

$$d_k = -B_k g_k, \tag{5}$$

and where $B_k \in R^{n\times n}$ is updated at each iteration to capture the already-obtained second derivative information. They are very efficient for medium-scale problems, but can not be used to solve large-scale problems because of its storage of matrices. The conjugate gradient (CG)

* Received May 8, 2000.
1) Research partially supported by Chinese NSF grants 19801033, 19771047 and 10171104.

method [13] uses the negative gradient direction and the previous search direction to form the current search direction, namely,

$$d_k = -g_k + \beta_k d_{k-1}, \tag{6}$$

where $d_1 = -g_1$ and β_k is a scalar. In the case when f is a strictly convex quadratic

$$f(x) = \frac{1}{2}x^T Ax + b^T x, \tag{7}$$

and α_k is obtained via an exact line search (4), the search directions generated by the CG method are conjugate to one another. As a result, the method gives the least value of (7) in at most n iterations. The CG method was extended by Fletcher and Reeves [11] to solve general nonconvex unconstrained optimization problem (1). Since it only requires storage of several vectors and is more rapid than the steepest descent method, the introduction of nonlinear CG method by Fletcher and Reeves marks the beginning of the field of large scale unconstrained optimization. Although the recent development of limited memory and discrete Newton methods have narrowed the class of problems for which CG methods are recommended, CG methods are still the best choice for solving very large problems with relatively inexpensive objective functions [16].

The purpose of this paper is to test and rank different nonlinear CG methods over a collection of standard test problems. As is known, for general nonconvex functions, there are many different choices for the scalar β_k in (6) and the properties of their corresponding CG methods may be very different. Another important reason is as follows. Usually, in the analyses and implementations of CG methods, the stepsize α_k is chosen by the strong Wolfe line search:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{8}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k, \tag{9}$$

where $0 < \delta < \sigma < 1$. Recently, however, [8] proposed a new nonlinear CG method in which β_k is given by (15). The descent property and global convergence of the method can be shown provided that the stepsize is obtained by the weak Wolfe line search, namely, (8) and

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k. \tag{10}$$

The hybrid methods related to this method are studied in [9], and the initial numerical results in [9] suggested an efficient hybrid CG algorithm that uses the weak Wolfe line search. Consequently, an overall assessment for the basic CG methods and hybrid CG methods is imperative to be done.

This paper is organized as follows. In the next section, we will give a description to the collection of test problems that are drawn from a standard code of test problems, CUTE. Other details of our numerical experiments are also provided in Section 2. In Section 3, we briefly review the five basic CG methods and report their numerical results. In Section 4, we briefly review the hybrid CG methods and report the numerical results of five hybrid CG methods. The numerical results made in Sections 3 and 4 show that the PRP, HS and DYHS2 are most efficient algorithms among all the tested CG algorithms. For the purpose of further comparisons, we draw in Section 5 some numerical results of the three efficient CG algorithms for difficult problems and listed them into a table. The table shows that one hybrid method, namely, DYHS2, outperforms the PRP and HS methods for difficult problems. Concluding remarks are given in the last section.

2. Preliminaries

Twenty-five sets of test problems are drawn from a standard code of test problems, CUTE [3]. A description of these test problems is given in Table 1, where "Name" denotes the name