Journal of Computational Mathematics, Vol.24, No.2, 2006, 149–156.

A NEW STEPSIZE FOR THE STEEPEST DESCENT METHOD *1)

Ya-xiang Yuan
(*LSEC, ICMSEC, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing 100080, China*)

Abstract

The steepest descent method is the simplest gradient method for optimization. It is well known that exact line searches along each steepest descent direction may converge very slowly. An important result was given by Barzilar and Borwein, which is proved to be superlinearly convergent for convex quadratic in two dimensional space, and performs quite well for high dimensional problems. The BB method is not monotone, thus it is not easy to be generalized for general nonlinear functions unless certain non-monotone techniques being applied. Therefore, it is very desirable to find stepsize formulae which enable fast convergence and possess the monotone property. Such a stepsize α_k for the steepest descent method is suggested in this paper. An algorithm with this new stepsize in even iterations and exact line search in odd iterations is proposed. Numerical results are presented, which confirm that the new method can find the exact solution within 3 iteration for two dimensional problems. The new method is very efficient for small scale problems. A modified version of the new method is also presented, where the new technique for selecting the stepsize is used after every two exact line searches. The modified algorithm is comparable to the Barzilar-Borwein method for large scale problems and better for small scale problems.

Mathematics subject classification: 65L05, 65F10.
Key words: Steepest descent, Line search, Unconstrained optimization, Convergence.

1. Introduction

The steepest descent method, which can be traced back to Cauchy (1847), is the simplest gradient method for unconstrained optimization:

$$\min_{x \in R^n} f(x), \tag{1.1}$$

where $f(x)$ is a continuous differential function in R^n. The method has the following form:

$$x_{k+1} = x_k + \alpha_k(-g_k), \tag{1.2}$$

where $g_k = g(x_k) = \nabla f(x_k)$ is the gradient vector of $f(x)$ at the current iterate point x_k and $\alpha_k > 0$ is the stepsize. Because the search direction in the method is the opposite of the gradient direction, it is the steepest descent direction locally, which gives the name of the method. Locally the steepest descent direction is the best direction in the sense that it reduces the objective function as much as possible.

The stepsize α_k can be obtained by exact line search:

$$\alpha_k^* = argmin\{f(x_k + \alpha(-g_k))\}, \tag{1.3}$$

or by some line search conditions, such as Goldstein conditions or Wolfe conditions (see Fletcher, 1987). It is easy to show that the steepest descent method is always convergent. That is, theoretically the method will not terminate unless a stationary point is found.

* Received September 30, 2004.
1) This work is partially supported by Chinese NSF grant 10231060.

However, even for the simplest case when the objective function $f(x)$ is a strictly convex quadratic, namely

$$f(x) = g^T x + \frac{1}{2} x^T H x, \tag{1.4}$$

where $g \in \Re^n$, $H \in \Re^{n \times n}$ symmetric and positive definite, the steepest descent method may not be very efficient. Assume that we are using exact line searches. Though we can show that the method converges linearly (see Akaike, 1959), the convergence rate can be very slow, especially when the condition number of the Hessian matrix H is very large, as the Q-linear fact of the convergence is

$$\frac{\lambda_1(H) - \lambda_n(H)}{\lambda_1(H) + \lambda_n(H)} \tag{1.5}$$

where $\lambda_1(H)$ and $\lambda_n(H)$ are the largest and smallest eigenvalues of H respectively. Forsythe(1968) gives an interesting analysis to show that the gradients $g(x_k)$ will approach zero eventually along two direction alternatively.

An surprising result was given by Barzilai and Borwein (1988), where gives formulae for the stepsize α_k which lead to superlinear convergence. The main idea of Barzilai and Borwein's approach is to use the information in the previous interation to decide the stepsize in the current iteration. The iteration (1.2) is viewed as

$$x_{k+1} = x_k - D_k g_k, \tag{1.6}$$

where $D_k = \alpha_k I$. In order to force the matrix D_k having certain quasi-Newton property, it is reasonable to require either

$$\min \|s_{k-1} - D_k y_{k-1}\|_2 \tag{1.7}$$

or

$$\min \|D_k^{-1} s_{k-1} - y_{k-1}\|_2, \tag{1.8}$$

where $s_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = g_k - g_{k-1}$, because in a quasi-Newton method we have that $x_{k+1} = x_k - B_k^{-1} g_k$ and the quasi-Newton matrix B_k satisfies the condition

$$B_k s_{k-1} = y_{k-1}. \tag{1.9}$$

Now, from $D_k = \alpha_k I$ and relations (1.7)-(1.8) we can obtain two stepsizes:

$$\alpha_k = \frac{s_{k-1}^T y_{k-1}}{\|y_{k-1}\|_2^2}, \tag{1.10}$$

and

$$\alpha_k = \frac{\|s_{k-1}\|_2^2}{s_{k-1}^T y_{k-1}} \tag{1.11}$$

respectively. For convex quadratic function in two variables, Barzilar and Borwein (1988) shows that the gradient method (1.2) with α_k given by (1.10) converges R-superlinearly and R-order is $\sqrt{2}$.

The result of Barzila and Borwein(1988) has triggered off many researches on the steepest descent method. For example, see Dai(2001), Dai et al(2002), Dai and Yuan(2003), Dai and Zhang(2001), Fletcher (2001), Friedlander et al. (1999), Nocedal et al(2000) and Raydon(1993, 1997).

The BB method performs quite well for high dimensional problems. The BB method is not monotone, and it is not easy to generalized to general nonlinear functions unless certain non-monotone techniques being applied. Therefore, it is very desirable to find stepsize formula which enables fast convergence and possesses the monotone property.

This paper tries to propose such a stepsize α_k for the steepest descent method. Due to the results of Forsythe(1968), the behave of steepest descent method for higher dimensional