New modification of nonlinear conjugate gradient method and application to non-parametric estimation

Mehamdia Abd Elhamid[1,2], **Chaib Yacine**[1,2]

[1]Mohamed Cherif Messaadia University, Souk Ahras, 41000, Algeria

[2]Laboratory Informatics and Mathematics (LIM)

a.mehamdia@univ-soukahras.dz,y.chaib@univ-soukahras.dz

The optimization model is a requisite mathematical problem since it has been connected to different fields such as economics, engineering and physics. Today there are many optimization algorithms, such as Newton, quasi-Newton and bundle algorithms. Note that these algorithms fail to solve large-scale optimization problems as they need to store and calculate relevant matrices. In contrast, the conjugate gradient (CG) algorithm is successful due to its simplicity of iteration and low memory requirements. In this paper, the nonlinear CG method is studied for the following unconstrained optimization problem:

$$\min\left\{f(x) : x \in \mathbb{R}^n\right\}, \tag{1}$$

where f is a smooth and nonlinear function. The CG method generates a sequence $\{x_k\}_{k\geq 0}$ such that:

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

where x_k is the current iteration point and $d_k \in \mathbb{R}^n$ is the search direction defined by the following formula:

$$d_{k+1} = -g_{k+1} + \beta_k d_k;\ d_0 = -g_0, \tag{3}$$

where g_{k+1} the gradient of f at x_{k+1} and the parameter β_k is known as the conjugate gradient coefficient. Some of the classical algorithms for β_k are Fletcher-Reeves (FR) method [2], Dai-Yuan (DY) method [1], Conjugate Descent (CD) method proposed by Fletcher [3], Polak-Ribière and Polyak (PRP) method [6, 7], Hestenes-Stiefel (HS) method [4] and Liu-Storey (LS) method [5]. The step length α_k is very important for the global convergence of conjugate gradient methods.

In this work, The novel β_k is introduced, which is defined as $\beta_k^{MHS^*}$, the conjugate gradient parameter of MHS* method is presented as follows

$$\beta_k^{MHS^*} = \frac{\left\|g_{k+1}\right\|^2 - \rho_1 \left|g_{k+1}^T g_k\right| \omega_k}{d_k^T\left(g_{k+1} - g_k\right)}, \tag{4}$$

where $\omega_k = \frac{\left|g_{k+1}^T d_k\right| g_{k+1}^T d_k}{\|g_k\| \|g_{k+1}\| \|d_k\|^2}$, $\rho_1 \in [0, 1]$.

The search direction d_k of MHS* algorithm is given by:

$$d_{k+1} = -g_{k+1} + \beta_k^{MHS^*} d_k;\ d_0 = -g_0, \tag{5}$$

The following Theorem to prove the sufficient descent direction of proposed method is needed.

The following Theorem to prove the sufficient descent direction of proposed method is needed.

Theorem 1. *Let the sequence $\{g_k\}_{k\geq 0}$ and $\{d_k\}_{k\geq 0}$ be generated by Algorithm MHS*, then for positive constant c we have*

$$g_k^T d_k \leq -c\left\|g_k\right\|^2, \ \forall\ k \geqslant 0. \tag{6}$$

To establish the global convergence of proposed methods, the following basic Assumptions on the objective function are needed.

Assumption 3.1. The level set

$$S = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\},$$

is bounded.

Assumption 1. In some open convex neighborhood $\mathcal{N}$ of S, the function f is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$, such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|, \quad \forall x, y \in \mathcal{N}. \tag{7}$$

From Assumption 2, it is deduced that for all $x \in \mathcal{N}$, there exists a positive constant $\Gamma \geq 0$, such that

$$\| \nabla f(x) \| \leq \Gamma, \text{ for all } x \in \mathcal{N}. \tag{8}$$

The following Theorem establishes the global convergence of the MHS* method with the SWLS.

Theorem 2. *Let Assumptions 1 and 2 hold. Consider any CG method in the form* (2) *and* (3)*, with the parameter $\beta_k = \beta_k^{MHS^*}$, in which the step length α_k is determined to satisfy the SWLS condition, where d_k is a descent search direction. Then, this method converges in the sense that*

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{9}$$

The authors are very grateful to the anonymous referees for their valuable comments and useful suggestions, which improved the quality of this work.

1. Dai Y. H., Yuan Y. A nonlinear conjugate gradient method with a strong global convergence property, SIAM J. Optim., 1999, 10, 177-182.
2. Fletcher R., Reeves C. Function minimization by conjugate gradients, Comput. J., 1964, 7, 149-154.
3. R. Fletcher, Practical Methods of Optimization, Second ed., Wiley, New York, 1987.
4. Hestenes M. R., Stiefel E.L. Methods of conjugate gradients for solving linear systems, J. Research Nat. Bur. Standards, 1952, 49, 409-436.
5. Liu Y., Storey C. Efficient generalized conjugate gradient algorithms, Theory JOTA, 1991, 69, 129-137.
6. Polak E., Ribière G. Note sur la convergence de directions conjuguée, Rev. Francaise Informat Recherche Operationelle, 1969, 16, 35-43.
7. Polyak B. T. The conjugate gradient method in extreme problems, USSR Comp. Math. Math. Phys., 1969, 9, 94-112.