An Efficient Conjugate Directions Method Without Linear Searches

Edouard Boudinov[1] and Arkadiy I. Manevich[2]

[1] Information Bank, FORTIS Bank, P.O. Box 243, NL-1000AE Amsterdam, The Netherlands
[2] Dniepropetrovsk National University, Dniepropetrovsk, Ukraine

Abstract. New conjugate directions algorithms are proposed, which are based on an orthogonalization procedure and do not perform line searches. The orthogonalization procedure prevents the conjugate vectors set from the degeneracy, eliminates high sensitivity to computation errors pertinent to methods of conjugate directions, and thus enable us to solve large-scale minimization problems without preconditioning. Numerical experiments have confirmed high efficiency of the algorithms for minimizing large-scale quadratic functions.

1 INTRODUCTION

Conjugate directions (CD) methods are widely used for solving large-scale linear algebra and optimization problems because of minimum storage requirements. But their theoretical property of quadratic termination holds only in precise arithmetic. If the number of variables N is large then even for quadratic objective functions the number of iterations can exceed $N+1$ by several times because of high sensitivity to computational errors. The cause is due to the usual procedure of conjugate vector construction based on the assumption that exact linear minima have been found on all preceding iterations (only under this premise the conjugacy conditions with respect to all conjugate vectors are satisfied). But even very small errors in locating the linear minima are accumulated yielding to violation of the conjugacy conditions and gradual derangement of the conjugate vector set.

In last years a clear trend to using inexact linear search is manifested aiming to reduce the number of functions or/and gradients calls and running time. Moreover, already in 70–80's there were proposed Quasi-Newton and conjugate directions algorithms without linear searches [4], [5], [8], [11]. They generate conjugate directions (or/and sequence of matrices) without accurate line minimization, retaining the quadratic termination property. However until now these algorithms were not able to solve large-scale minimization problems.

In this work a new CD - method without linear minimizations is proposed, which employs another scheme of construction of conjugate vectors based on an orthogonalization procedure: in constructing the next conjugate vector the component of the function gradient is used that is orthogonal to the subspace of preceding conjugate vectors (instead of the gradient itself). Such an

approach allows attaining very high accuracy in constructing conjugate directions and thus eliminates high sensitivity to computation errors pertinent to known CD-methods. As this procedure does not require the current point to be the minimum point along a given conjugate vector, any function evaluations for linear minimization on each iteration can be excluded ensuring considerable decrease of running time.

At each iteration only one step is done, which includes displacements along the last two conjugate vectors, so the movement occurs in the hyperplane determined by these conjugate vectors. Along each conjugate vector two successive steps are made: a preliminary step on k-th iteration and the Newton-like step on the next iteration.

The proposed algorithm and its modification have been realized in JAVA programming language. They have been tested on several quadratic test-functions with the number of variables N up to 1'000'000. The numerical experiments have confirmed that these algorithms locate the minimum for quadratic functions of very many variables in at most $N+1$-th steps, without perceptible losses in precision. The numerical experiments have shown that the running time for solving all test - problems was less by several times in comparison with known efficient algorithms for large-scale problems.

2 DESCRIPTION OF THE METHOD

2.1 Basic algorithm

Consider the basic algorithm as applied to minimization of quadratic functions (we do not discuss here modifications needed for non-quadratic functions).

At the first iteration $k = 1$ vectors d_1, n_1, and the step are as follows:

$$d_1 = n_1 = -\frac{g_1}{\|g_1\|}, \qquad x_2 = x_1 + \delta_1^{(1)} d_1, \tag{1}$$

where $g_k = g(x_k)$ is the gradient of the objective function $f(x)$, and $\delta_1^{(1)}$ is an arbitrary step.

Let, after $k-1$-th step, $\{n_i\}$ $(i = 1, \dots, k-1)$ be an orthonormalized vector set in the subspace Ω_{k-1}, which is determined by the conjugate vectors $\{d_i\}$ $(i = 1, \dots, k-1)$. Vector n_k (a normalized vector, orthogonal to the subspace Ω_{k-1}), is

$$n_k = \frac{n_k^*}{\|n_k^*\|}, \qquad n_k^* = -g_k + \gamma_{k-1}^{(k)} n_{k-1}, \qquad \gamma_{k-1}^{(k)} = (g_k, n_{k-1}) \tag{2}$$

((x, y) denotes the scalar product of vectors x and y). A new conjugate vector d_k is sought as a linear combination of the preceding conjugate vector d_{k-1} and the vector n_k $(\Delta g_{k-1} = g_k - g_{k-1})$:

$$d_k = \frac{d_k^*}{\|d_k^*\|}, \qquad d_k^* = n_k + \beta_{k-1}^{(k)} d_{k-1}, \qquad \beta_{k-1}^{k} = -\frac{(n_k, \Delta g_{k-1})}{(d_{k-1}, \Delta g_{k-1})} \tag{3}$$