Journal of Computational and Applied Mathematics 332 (2018) 101–106

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

A note on the global convergence theorem of accelerated adaptive Perry conjugate gradient methods[☆]

Yigui Ou

Department of Mathematics, Hainan University, Haikou 570228, China

ARTICLE INFO

Article history:
Received 23 July 2017
Received in revised form 13 October 2017

MSC 2010:
90C30
65K05
49M37

Keywords:
Unconstrained optimization
Conjugate gradient algorithm
Accelerated scheme
Self-scaling memoryless BFGS update
Convergence analysis

ABSTRACT

In Andrei (2017), a class of efficient conjugate gradient algorithms (ACGSSV) is proposed for solving large-scale unconstrained optimization problems. However, due to a wrong inequality and an incorrect reasoning used in analyzing the global convergence property for the proposed algorithm, the proof of Theorem 4.2, the global convergence theorem, is incorrect. In this paper, the necessary corrections are made. Under common assumptions, it is shown that Algorithm ACGSSV converges linearly to the unique minimizer.

1. Introduction

Due to simple computation and low memory requirements, conjugate gradient (**CG**) methods are particularly effective for solving large-scale optimization problems in the following form,

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

where $f : R^n \to R$ is a continuously differentiable function. For a detailed survey of conjugate gradient methods, see [1]. Throughout the paper, $\|\cdot\|$ denotes the Euclidean norm on R^n. For convenience, we denote $f(x_k)$ by f_k, and $\nabla f(x_k)$ by g_k.

Recently, based on the symmetrical scaled Perry CG direction matrix and the self-scaling memoryless BFGS update, Andrei [2] suggested an accelerated adaptive class of nonlinear conjugate gradient algorithms, namely ACGSSV. More precisely, given an initial point $x_0 \in R^n$, it generates a sequence $\{x_k\}$ as

$$x_{k+1} = x_k + \xi_k \alpha_k d_k, \tag{1.2}$$

where the stepsize α_k is computed by the standard Wolfe line search conditions (see [3] for details):

$$\begin{cases} f(x_k+\alpha_k d_k) - f_k \ \le\ \rho\alpha_k g_k^T d_k, \\ \nabla f(x_k+\alpha_k d_k)^T d_k \ \ge\ \sigma g_k^T d_k, \end{cases} \tag{1.3}$$

☆ This work is partially supported by NNSF of China (Nos. 11261015, 11761025) and NSF of Hainan Province (Nos. 111001, 2016CXTD004).
E-mail address: ouyigui@126.com.

https://doi.org/10.1016/j.cam.2017.10.024
0377-0427/

with $0 < \rho < \sigma < 1$, d_k is the search direction defined by

$$\begin{aligned} d_0 &= -g_0, \\ d_{k+1} &= -P_{k+1}g_{k+1}, \ \ k = 0, 1, 2, \ldots, \end{aligned} \tag{1.4}$$

where the symmetrical scaled Perry CG direction matrix $P_{k+1} \in R^{n\times n}$ is computed as

$$P_{k+1} = I - \frac{s_k y_k^T + y_k s_k^T}{y_k^T s_k} + \eta_k \frac{s_k s_k^T}{y_k^T s_k}, \tag{1.5}$$

in which $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$, and the positive scalar parameter η_k is given by

$$\eta_k = \begin{cases} \overline{\eta}_k, & \text{if } \overline{\eta}_k > 2\dfrac{\|y_k\|^2}{y_k^T s_k}, \\ 2\dfrac{\|y_k\|^2}{y_k^T s_k}, & \text{otherwise} \end{cases} \tag{1.6}$$

with

$$\overline{\eta}_k = 1 + \theta_k\Big(\frac{\|y_k\|^2}{y_k^T s_k} - \frac{y_k^T s_k}{\|s_k\|^2}\Big) + \frac{y_k^T s_k}{\|s_k\|^2}, \tag{1.7}$$

where the scaling parameter θ_k is determined by

$$\theta_k = \frac{\|s_k\|^2}{y_k^T s_k}, \tag{1.8}$$

or

$$\theta_k = \frac{y_k^T s_k}{\|y_k\|^2}, \tag{1.9}$$

and $\xi_k > 0$ is a parameter computed as

$$\xi_k = \begin{cases} 1, & \text{if } b_k = 0, \\ -\dfrac{a_k}{b_k}, & \text{otherwise} \end{cases} \tag{1.10}$$

in which $a_k = \alpha_k g_k^T d_k$ and $b_k = -\alpha_k(g_k - \nabla f(x_k + \alpha_k d_k))^T d_k$.

Although ACGSSV is more efficient and more robust than the well-known conjugate gradient algorithms, such as SCALCG [4], CG-DESCENT [5] and CONMIN [6], a part of its global convergence analysis is incorrect, due to a wrong inequality and an incorrect reasoning used. In what follows, we first describe these errors that occurred in proving the global convergence Theorem 4.2 in [2], and then make the necessary corrections. Under common assumptions, it is proven that Algorithm ACGSSV converges linearly to the unique minimizer.

2. Errors in the convergence analysis of ACGSSV

In this section, we state the wrong inequality and the incorrect reasoning used in proving the global convergence Theorem 4.2 in [2]. To this end, we need the following basic assumptions [2]:

A1. The level set $\mathcal{L}_0 = \{x \in R^n | f(x) \le f(x_0)\}$ is bounded, where x_0 is an available initial point.

A2. The gradient function $g(x)$ of $f(x)$ is Lipschitz continuous in a neighborhood $\mathcal{N}$ of $\mathcal{L}_0$, i.e., there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \le L\|x - y\|, \ \ \forall x, y \in \mathcal{N}.$$

Since $\{f_k\}$ is a monotonically decreasing sequence, it is clear that the sequence $\{x_k\}$ generated by Algorithm ACGSSV is contained in $\mathcal{L}_0$. This fact together with the above assumptions implied that there exists a constant Γ such that

$$\|\nabla f(x)\| \le \Gamma, \forall x \in \mathcal{L}_0. \tag{2.1}$$

To present our results, we also need the following definition (see [3]).

Definition 2.1. A differentiable function f is said to be uniformly convex on $\mathcal{N}$, if there exists a positive constant $\mu > 0$ such that

$$(\nabla f(x) - \nabla f(y))^T(x - y) \ge \mu\|x - y\|^2, \ \ \forall x, y \in \mathcal{N}. \tag{2.2}$$