

Two Modified Hybrid Conjugate Gradient Methods Based on a Hybrid Secant Equation*

Saman Babaie-Kafaki^{a,b} and Nezam Mahdavi-Amiri^c

^a*Semnan University, Faculty of Mathematics, Statistics and Computer Sciences*

P.O. Box: 35195-363, Semnan, Iran

^b*Institute for Research in Fundamental Sciences (IPM)*

P.O. Box: 19395-5746, Tehran, Iran

^c*Sharif University of Technology, Faculty of Mathematical Sciences*

P.O. Box: 11155-9415, Azadi St., Tehran, Iran

E-mail: sbk@semnan.ac.ir

E-mail(*corresp.*): nezamm@sharif.edu

Received November 28, 2011; revised October 16, 2012; published online February 1, 2013

Abstract. Taking advantage of the attractive features of Hestenes–Stiefel and Dai–Yuan conjugate gradient methods, we suggest two globally convergent hybridizations of these methods following Andrei’s approach of hybridizing the conjugate gradient parameters convexly and Powell’s approach of nonnegative restriction of the conjugate gradient parameters. In our methods, the hybridization parameter is obtained based on a recently proposed hybrid secant equation. Numerical results demonstrating the efficiency of the proposed methods are reported.

Keywords: unconstrained optimization, large-scale optimization, conjugate gradient method, secant equation, global convergence.

AMS Subject Classification: 49M37; 65K05; 90C53.

1 Introduction

Conjugate gradient (CG) methods comprise a class of unconstrained optimization algorithms characterized by low memory requirements and strong global convergence properties [9]. CG methods have played significant roles in solving large-scale nonlinear optimization problems.

* The first author acknowledges partial support by a grant from IPM (No. 91900051), and partial support by the Research Council of Semnan University. The second author thanks the Research Council of Sharif University of Technology for its support. The authors thank the four anonymous referees for their valuable comments and suggestions leading to an improved presentation. They are also grateful to Professor William W. Hager for providing the C++ code of the CG.Descent algorithm, and Professor Michael Navon for providing the MATLAB code of the line search.

Generally, a CG method is designed to solve the following unconstrained optimization problem,

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

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a smooth nonlinear function and its gradient is available. The iterative formula of a CG method is given by

$$\begin{aligned} x_0 &\in \mathbb{R}^n, \\ x_{k+1} &= x_k + s_k, \quad s_k = \alpha_k d_k, \quad k = 0, 1, \dots, \end{aligned} \tag{1.2}$$

where α_k is a steplength to be computed by a line search procedure and d_k is the search direction defined by

$$\begin{aligned} d_0 &= -g_0, \\ d_{k+1} &= -g_{k+1} + \beta_k d_k, \quad k = 0, 1, \dots, \end{aligned} \tag{1.3}$$

where $g_k = \nabla f(x_k)$ and β_k is a scalar called the CG (update) parameter.

The steplength α_k is usually chosen to satisfy certain line search conditions [19, 21]. Among them, the so-called (standard) Wolfe line search conditions [19, 21] have attracted special attention in the convergence analysis and the implementations of CG methods, requiring that

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k \nabla f(x_k)^T d_k, \tag{1.4}$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f(x_k)^T d_k, \tag{1.5}$$

where $0 < \delta < \sigma < 1$. Also, the strong Wolfe line search conditions consist of (1.4) and the following strengthened version of (1.5),

$$|\nabla f(x_k + \alpha_k d_k)^T d_k| \leq -\sigma \nabla f(x_k)^T d_k. \tag{1.6}$$

Different choices for the CG parameter lead to different CG methods. Let $\|\cdot\|$ denotes the Euclidean norm and define $y_k = g_{k+1} - g_k$. The essential CG methods can be divided into two categories. In the first category, all the CG parameters have the common numerator $g_{k+1}^T y_k$. These methods have been proposed by Hestenes and Stiefel (HS), Polak, Ribière and Polyak (PRP), and Liu and Storey (LS) with the following CG parameters [14],

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \quad \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}, \quad \beta_k^{LS} = -\frac{g_{k+1}^T y_k}{d_k^T g_k}.$$

In the second category, all the CG parameters have the common numerator $\|g_{k+1}\|^2$. These methods have been proposed by Fletcher and Reeves (FR), Fletcher (conjugate descent or CD), and Dai and Yuan (DY) with the following CG parameters [14],

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \quad \beta_k^{CD} = -\frac{\|g_{k+1}\|^2}{d_k^T g_k}, \quad \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}.$$

There are some advantages and disadvantages for the CG methods in each category. For example, in spite of the global convergence property of the FR

method for general functions [29], this method numerically performs much worse than the HS, PRP and LS methods [1]. The poor numerical performance of the FR method is related to the jamming phenomenon [11], that is, the algorithm could take many short steps without making significant progress to the minimum. However, the methods with the common numerator $g_{k+1}^T y_k$ possess an automatic approximate restart feature which avoids jamming. More exactly, when the step s_k is small, the factor y_k in the numerator tends to zero. Therefore, β_k becomes small and the new search direction d_{k+1} is approximately the steepest descent direction. On the other hand, Powell [20] constructed a three dimensional counter example and showed that the PRP and HS methods could cycle infinitely without convergence to a solution; that is, they lack global convergence in certain circumstances. Based on the insight gained by his counter example, Powell [20] proposed a modified PRP method with the following CG parameter,

$$\beta_k^{PRP+} = \max\{\beta_k^{PRP}, 0\}.$$

Then, Gilbert and Nocedal [11] studied the theoretical effectiveness and numerical efficiency of the PRP+ method for general objective functions. A similar study has been done on the HS+ method [11], i.e., a CG method with the following CG parameter,

$$\beta_k^{HS+} = \max\{\beta_k^{HS}, 0\}.$$

Generally, the FR, CD and DY methods have strong global convergence properties [8, 9, 11], but they may have modest computational performance [1]. Meanwhile, the methods of HS, PRP and LS may not always be convergent, but they often have better computational performance [1]. A nice review of different CG methods can be found in [14].

To attain good computational performance and to maintain the attractive feature of strong global convergence, researchers paid special attention to hybridize the CG methods of the two categories (see [14] and the references therein). Recently, Andrei proposed several efficient hybrid CG methods based on convex combinations of the CG parameters of the two categories (see [3] and the references therein). Specially, in [3], using an acceleration scheme, Andrei proposed another hybridization of HS and DY methods in which the hybridization parameter is computed based on the modified secant equation proposed by Li et al. [16]. In recent efforts following Andrei's approach, Babaie-Kafaki et al. [5, 6] proposed efficient hybridizations of HS and DY methods in which the hybridization parameters are computed based on the modified secant equations proposed by Yuan [25], and Li and Fukushima [15], and another scaled secant equation newly proposed based on the search direction of the Hager-Zhang CG method [13].

Among the different hybrid CG methods, hybridizations of HS and DY methods have shown promising numerical performance [1]. The HS method has the nice property of satisfying the conjugacy condition $d_{k+1}^T y_k = 0$, for all $k \geq 0$, independent of the line search conditions and the objective function convexity. On the other hand, the DY method has remarkable convergence properties

in contrast to the other CG methods [8]. For example, in addition to the generation of descent search directions, the DY method has been proved to have a certain self-adjusting property independent of the line search conditions and the objective function convexity. More exactly, if there exist positive constants γ_1 and γ_2 such that $\gamma_1 \leq \|g_k\| \leq \gamma_2$, for all $k \geq 0$, then, for any $p \in (0, 1)$, there exists a positive constant τ such that the sufficient descent condition $g_i^T d_i \leq -\tau \|g_i\|^2$ holds for at least $\lfloor pk \rfloor$ indices $i \in [0, k]$, where $\lfloor j \rfloor$ denotes the largest integer less than or equal to j . Also, under mild assumptions on the objective function, the DY method was shown to be globally convergent under a variety of line search conditions.

Here, to take advantage of the attractive features of HS and DY methods, following Andrei’s approach of hybridizing the CG parameters convexly [3] and Powell’s nonnegative restriction of the CG parameters [20], we propose two hybridizations of these methods using a hybrid secant equation proposed by Babaie-Kafaki [4]. We then discuss the global convergence property of the proposed methods and report comparative testing results. The remainder of this work is organized as follows. In Section 2, we propose our hybrid CG methods. The global convergence property of our methods is discussed in Section 3. In Section 4, we numerically compare the proposed methods with the two recently proposed hybrid CG methods of [3] and [5], as well as the efficient CG methods HS+, PRP+, DPRP-W proposed by Yu et al. [24], and CG_Descent proposed by Hager and Zhang [13]. Finally, we provide our concluding remarks.

2 Two Hybrid CG Methods

Here, after a brief discussion on the Andrei’s approach of hybridizing the CG parameters [3] and the hybrid secant equation proposed by Babaie-Kafaki [4], we propose a modified hybrid CG parameter. Then, by a nonnegative restriction of the proposed hybrid CG parameter based on the Powell’s approach [20] in order to achieve the global convergence for general objective functions, we propose another hybrid CG parameter.

As pointed out in Section 1, Andrei has recently proposed several efficient hybridizations of HS and DY methods using convex combinations of β_k^{HS} and β_k^{DY} [2, 3]. More exactly, in the hybrid CG methods proposed in [2, 3], the CG parameter, namely β_k^C , is in the following form,

$$\beta_k^C = (1 - \theta_k)\beta_k^{HS} + \theta_k\beta_k^{DY} = (1 - \theta_k)\frac{g_{k+1}^T y_k}{d_k^T y_k} + \theta_k\frac{g_{k+1}^T g_{k+1}}{d_k^T y_k}, \tag{2.1}$$

where θ_k , namely the hybridization parameter, is a scalar parameter satisfying $0 \leq \theta_k \leq 1$. Note that β_k^C can be included in the two-parameter family of CG parameters proposed by Nazareth [18]. Now, from (1.3) we have

$$d_{k+1} = -g_{k+1} + (1 - \theta_k)\frac{g_{k+1}^T y_k}{d_k^T y_k}d_k + \theta_k\frac{g_{k+1}^T g_{k+1}}{d_k^T y_k}d_k,$$

or equivalently,

$$d_{k+1} = -g_{k+1} + (1 - \theta_k) \frac{g_{k+1}^T y_k}{s_k^T y_k} s_k + \theta_k \frac{g_{k+1}^T g_{k+1}}{s_k^T y_k} s_k. \quad (2.2)$$

As known, if the point x_{k+1} is close enough to a local minimizer x^* , then a good direction to follow is the Newton direction, that is, $d_{k+1} = -\nabla^2 f(x_{k+1})^{-1} g_{k+1}$. Motivated by this, Andrei [2, 3] rewrote (2.2) as follows:

$$-\nabla^2 f(x_{k+1})^{-1} g_{k+1} = -g_{k+1} + (1 - \theta_k) \frac{g_{k+1}^T y_k}{s_k^T y_k} s_k + \theta_k \frac{g_{k+1}^T g_{k+1}}{s_k^T y_k} s_k.$$

After some algebraic manipulations one obtains

$$\theta_k = \frac{s_k^T \nabla^2 f(x_{k+1}) g_{k+1} - s_k^T g_{k+1} - \frac{g_{k+1}^T y_k}{s_k^T y_k} s_k^T \nabla^2 f(x_{k+1}) s_k}{\frac{g_{k+1}^T g_{k+1}}{s_k^T y_k} s_k^T \nabla^2 f(x_{k+1}) s_k}. \quad (2.3)$$

Due to the essential property of low memory requirements for the CG methods, Andrei applied the secant equations in order to avoid the exact computation of $\nabla^2 f(x_{k+1})$.

Secant equations play essential roles for approximation of the Hessian of an objective function in each iteration of the quasi-Newton methods, an efficient family of iterative methods for solving unconstrained optimization problems. The iterative quasi-Newton formula has the form (1.2) in which the search direction d_k is calculated by solving the following linear system of equations,

$$B_k d_k = -g_k,$$

where B_k is an approximation of $\nabla^2 f(x_k)$. The quasi-Newton methods are characterized by the fact that B_k is effectively updated to obtain a new matrix B_{k+1} as an approximation of $\nabla^2 f(x_{k+1})$ in the following form,

$$B_{k+1} = B_k + \Delta B_k,$$

where ΔB_k is a correction matrix. The matrix B_{k+1} is imposed to satisfy some suitable equations, namely secant equations, which include the second order information. The most popular equation is the standard secant equation, that is,

$$B_{k+1} s_k = y_k. \quad (2.4)$$

The standard secant equation only uses the gradient values available at the current iteration and ignores the function values. So, efforts have been made in order to modify the standard secant equation to employ more available information and consequently, to provide a better approximation for the Hessian (see [4] and the references therein). In a recent effort, Babaie-Kafaki [4] proposed a hybridization of the modified secant equations proposed by Yuan [25] and Wei et al. [22], using a convex combination of these equations. The BFGS method based on the hybrid secant equation proposed in [4] showed a promising numerical performance.

Assume that the objective function f is smooth enough. If we expand the function f at $x_k = x_{k+1} - s_k$ using Taylor's theorem, then for small values of $\|s_k\|$ we have

$$f_k = f_{k+1} - s_k^T g_{k+1} + \frac{1}{2} s_k^T \nabla^2 f(x_{k+1}) s_k - \frac{1}{6} s_k^T (T_{k+1} s_k) s_k + O(\|s_k\|^4),$$

with

$$s_k^T (T_{k+1} s_k) s_k = \sum_{i,j,l=1}^n \frac{\partial^3 f(x_{k+1})}{\partial x^i \partial x^j \partial x^l} s_k^i s_k^j s_k^l,$$

which yields the following approximation,

$$s_k^T \nabla^2 f(x_{k+1}) s_k \approx s_k^T y_k + \eta_k, \tag{2.5}$$

where

$$\eta_k = 2(f_k - f_{k+1}) + s_k^T (g_k + g_{k+1}). \tag{2.6}$$

Since in the quasi-Newton methods a matrix B_{k+1} is required to approximate $\nabla^2 f(x_{k+1})$, by considering (2.5) a modified secant equation can be proposed as follows:

$$B_{k+1} s_k = z_k, \quad z_k = y_k + \frac{\eta_k}{s_k^T u_k} u_k, \tag{2.7}$$

where $u_k \in \mathbb{R}^n$ is a vector parameter satisfying $s_k^T u_k \neq 0$. The modified secant equation (2.7) is justified by the following theorem.

Theorem 1. [22] *If f is sufficiently smooth and $\|s_k\|$ is small enough, then the following estimating relations hold,*

$$\begin{aligned} s_k^T (\nabla^2 f(x_{k+1}) s_k - y_k) &= \frac{1}{2} s_k^T (T_{k+1} s_k) s_k + O(\|s_k\|^4), \\ s_k^T (\nabla^2 f(x_{k+1}) s_k - z_k) &= \frac{1}{3} s_k^T (T_{k+1} s_k) s_k + O(\|s_k\|^4). \end{aligned}$$

For a quadratic objective function f , we have $\eta_k = 0$ and consequently, the modified secant equation (2.7) reduces to the standard secant equation (2.4). A comprehensive study has been made on the properties of the modified secant equation (2.7) by Wei et al. [22].

Different choices for the vector parameter u_k in (2.7) lead to different modified secant equations. The particular choice $u_k = s_k$ has been considered by Wei et al. [22]. Also, if the steplength α_k is computed such that the (strong) Wolfe line search conditions hold, then from (1.5) or (1.6) we have

$$s_k^T y_k = s_k^T g_{k+1} - s_k^T g_k \geq -(1 - \sigma) s_k^T g_k > 0. \tag{2.8}$$

So, using the (strong) Wolfe line search conditions, the particular choice $u_k = y_k$ was firstly proposed by Yuan [25] and then adopted by Wei et al. [22].

Here, in order to apply the interesting features of the modified secant equations proposed in [22] and [25], we suggest a hybridization of these equations

by considering the vector parameter u_k in (2.7) as a convex combination of the vectors s_k and y_k , i.e.,

$$u_k = (1 - \lambda_k)y_k + \lambda_k s_k, \quad \lambda_k \in [0, 1]. \quad (2.9)$$

Note that the choices $\lambda_k = 1$ and $\lambda_k = 0$ in (2.9) convert equation (2.7) to the modified secant equations proposed in [22] and [25], respectively. Hence, our hybrid secant equation, i.e., the modified secant equation (2.7) with the vector parameter u_k defined by (2.9), can be considered as an extension of the modified secant equations proposed in [22] and [25]. Next, we suggest a formula for computing λ_k based on an effective modified secant equation proposed by Li and Fukushima [15].

To increase the accuracy of our hybrid secant equation, here λ_k is computed making use of the information from two previous steps. More precisely, based on the proposal of [4], using the modified secant equation proposed by Li and Fukushima [15] which is effective even for nonconvex functions, we compute λ_k such that in addition to the secant equation (2.7), at the $(k - 1)$ th iteration the following secant equation is implicitly satisfied,

$$B_{k+1}s_{k-1} = \bar{z}_{k-1}, \quad (2.10)$$

with $\bar{z}_{k-1} = y_{k-1} + h_{k-1}\|g_{k-1}\|^r s_{k-1}$, where $r > 0$, and $h_{k-1} > 0$ is defined by

$$h_{k-1} = C + \max\left\{-\frac{s_{k-1}^T y_{k-1}}{\|s_{k-1}\|^2}, 0\right\} \|g_{k-1}\|^{-r}, \quad (2.11)$$

with some positive constant C (see also [5, 28]). Taking the inner product of (2.10) with s_k and using the modified secant equation (2.7), after some algebraic manipulations, we get

$$\lambda_k = \frac{w_k^T y_k}{w_k^T (y_k - s_k)}, \quad (2.12)$$

where $w_k = s_{k-1} - \delta_k s_k$, with $\delta_k = (s_k^T \bar{z}_{k-1} - s_{k-1}^T y_k) / \eta_k$. The parameter λ_k computed by (2.12) may be outside the interval $[0, 1]$. Thus, in order to have a convex combination in (2.9), if $\lambda_k < 0$, then we let $\lambda_k = 0$, and if $\lambda_k > 1$, then we let $\lambda_k = 1$. It is remarkable that if $(w_k^T y_k)(w_k^T s_k) < 0$, then we have $\lambda_k \in (0, 1)$.

Now, in order to make the hybridization parameter θ_k defined by (2.3) more effective by avoiding the exact computation of the Hessian $\nabla^2 f(x_{k+1})$, we follow Andrei's approach using our hybrid secant equation. More precisely, we replace the expression $\nabla^2 f(x_{k+1})s_k$ in (2.3) with the approximation suggested by (2.7), that is, $B_{k+1}s_k = y_k + \frac{\eta_k}{s_k^T u_k} u_k$, with the vector parameter u_k defined by (2.9). So, after some algebraic manipulations, we have

$$\theta_k = \frac{\eta_k \left(\frac{g_{k+1}^T u_k}{s_k^T u_k} - \frac{g_{k+1}^T y_k}{s_k^T y_k} \right) - s_k^T g_{k+1}}{g_{k+1}^T g_k + \eta_k \frac{g_{k+1}^T g_k}{s_k^T y_k}}, \quad (2.13)$$

in which u_k is defined by (2.9). It is remarkable that the choices $\lambda_k = 1$ and $\lambda_k = 0$ convert θ_k given by (2.13) to the hybridization parameters proposed in [3] and [5], respectively. So, the hybrid CG method in the form of (1.2)–(1.3) with the hybrid CG parameter β_k^C defined by (2.1), in which the hybridization parameter θ_k is computed by (2.13), can be considered as an extension of the hybrid CG parameters proposed in [3, 5]. The hybridization parameter θ_k given by (2.13) may be outside the interval $[0, 1]$. Here, in order to have a convex combination in (2.1), similar to the advocacy in [3, 5] if $\theta_k < 0$, then we let $\theta_k = 0$, and if $\theta_k > 1$, then we let $\theta_k = 1$; that is, θ_k is restricted to $[0, 1]$.

If the objective function f is quadratic, then η_k , defined by (2.6), is zero. Thus, for quadratic objective functions, the hybridization parameter θ_k defined by (2.13) reduces to a hybridization parameter proposed by Andrei [2], that is,

$$\theta_k = -\frac{s_k^T g_{k+1}}{g_{k+1}^T g_k}.$$

In this case, if σ in the strong Wolfe line search condition (1.6) tends to zero, then $s_k^T g_{k+1} \rightarrow 0$, and consequently, $\theta_k \rightarrow 0$. Therefore, the numerical behavior of our hybrid CG method approaches to the numerical behavior of the HS method. More precisely, under these conditions our method has an automatic approximate restart feature which avoids jamming, as discussed in Section 1.

In the next section, we will show that our hybrid CG method with the hybridization parameter θ_k defined by (2.13) is globally convergent for uniformly convex functions. In order to achieve the global convergence for general functions, we adopt the nonnegative restriction of the CG parameters, originally proposed in [20] and then further studied in [11], and propose a new hybrid CG parameter as follows:

$$\beta_k^{C+} = (1 - \theta_k) \max\{\beta_k^{HS}, 0\} + \theta_k \beta_k^{DY}, \quad \theta_k \in [0, 1]. \tag{2.14}$$

From (2.8), the Wolfe line search conditions guarantee that $\beta_k^{C+} \geq 0$ which, as will be shown in the next section, is necessary to achieve the global convergence for general functions. It is remarkable that for the quadratic objective functions, if σ in the strong Wolfe line search condition (1.6) tends to zero, then $\beta_k^{C+} \approx \beta_k^{HS+}$. In other words, in such situations the numerical behavior of our second hybrid CG method tends to the numerical behavior of the efficient HS+ method.

3 Convergence Analysis

Here, after a brief discussion on the uniformly convex functions and some necessary results, we show that our hybrid CG method with the CG parameter β_k^C defined by (2.1), in which the hybridization parameter θ_k is computed by (2.13), is globally convergent for uniformly convex functions and also, our hybrid CG method with the CG parameter β_k^{C+} defined by (2.14) is globally convergent for general functions.

DEFINITION 1. [21] Let $S \subset \mathbb{R}^n$ be a nonempty convex set and consider the function $f : S \rightarrow \mathbb{R}$. If there exists a positive constant ρ such that, for any $x, y \in S$ and for all $\alpha \in (0, 1)$,

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \frac{1}{2}\rho\alpha(1 - \alpha)\|x - y\|^2,$$

then f is called a uniformly (or strongly) convex function on S .

We make the following basic assumptions on the objective function, commonly used in the convergence analysis of the CG methods.

ASSUMPTIONS 1.

1. The level set $\mathcal{L} = \{x \mid f(x) \leq f(x_0)\}$, with x_0 to be the starting point of the algorithm, is bounded.
2. In a neighborhood \mathcal{N} of \mathcal{L} , f is continuously differentiable and its gradient is Lipschitz continuous; that is, there exists a positive constant L such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathcal{N}. \quad (3.1)$$

DEFINITION 2. We say that the search direction d_k is a descent direction (or equivalently, satisfies the descent condition) if and only if

$$d_k^T g_k < 0. \quad (3.2)$$

Also, we say that the search directions $\{d_k\}_{k \geq 0}$ satisfy the sufficient descent condition if and only if

$$d_k^T g_k \leq -c\|g_k\|^2, \quad \forall k \geq 0, \quad (3.3)$$

where c is a positive constant.

The following important result plays an essential role in proving the global convergence theorems of the CG methods.

Lemma 1. [9] *Suppose that Assumptions 1 hold. Consider any CG method in the form of (1.2)–(1.3), where for which at the k th iteration, $k = 0, 1, \dots$, the search direction d_k is a descent direction and the steplength α_k is computed so that the strong Wolfe line search conditions (1.4) and (1.6) are satisfied. If*

$$\sum_{k \geq 0} \frac{1}{\|d_k\|^2} = \infty,$$

then we have

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (3.4)$$

To prove our global convergence theorems, the following results are also needed.

Lemma 2. [5] Under Assumptions 1, there exists a positive constant γ such that

$$\|\nabla f(x)\| \leq \gamma, \quad \forall x \in \mathcal{N}. \tag{3.5}$$

Theorem 2. [21] Let $S \subset \mathbb{R}^n$ be a nonempty open convex set and consider the differentiable function $f : S \rightarrow \mathbb{R}$. The function f is uniformly convex on S if and only if its gradient is uniformly monotone on S , i.e., there exists a positive constant μ such that

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq \mu\|x - y\|^2, \quad \forall x, y \in S.$$

Theorem 3. [21] Let $S \subset \mathbb{R}^n$ be a nonempty open convex set and consider the twice continuously differentiable function $f : S \rightarrow \mathbb{R}$. The function f is uniformly convex if and only if its Hessian matrix is uniformly positive definite at each point of S , i.e., there exists a positive constant $\bar{\mu}$ such that

$$u^T \nabla^2 f(x) u \geq \bar{\mu} \|u\|^2, \quad \forall x \in S, \forall u \in \mathbb{R}^n.$$

Lemma 3. Suppose that for a CG method in the form of (1.2)–(1.3) Assumptions 1 and the descent condition (3.2) hold. If f is twice continuously differentiable and uniformly convex, then there exist positive constants m and M such that

$$m\|s_k\|^2 \leq s_k^T y_k + \eta_k \leq M\|s_k\|^2,$$

with η_k as defined by (2.6).

Proof. Using Taylor’s theorem, we can write

$$f_k = f(x_{k+1} - s_k) = f_{k+1} - s_k^T g_{k+1} + \frac{1}{2} s_k^T \nabla^2 f(\zeta x_k + (1 - \zeta)x_{k+1}) s_k,$$

or equivalently,

$$f_k - f_{k+1} + s_k^T g_{k+1} = \frac{1}{2} s_k^T \nabla^2 f(\zeta x_k + (1 - \zeta)x_{k+1}) s_k, \tag{3.6}$$

for some $\zeta \in (0, 1)$. Because of the descent condition (3.2) and (1.4) (the first condition of the strong Wolfe line search conditions), the sequence $\{x_k\}_{k \geq 0}$ generated by the considered CG method is a subset of \mathcal{L} . On the other hand, since f is a continuous and convex function, the level set \mathcal{L} is a closed convex subset of \mathbb{R}^n , and consequently $\zeta x_k + (1 - \zeta)x_{k+1} \in \mathcal{L}$. Also, because $\nabla^2 f$ is a continuous function on the closed and bounded set \mathcal{L} , for some $\bar{x} \in \mathcal{L}$ we have

$$\|\nabla^2 f(\bar{x})\| = \max_{x \in \mathcal{L}} \|\nabla^2 f(x)\|. \tag{3.7}$$

Now, from (3.6), Theorem 3 and (3.7), we have

$$\begin{aligned} \bar{\mu}\|s_k\|^2 &\leq s_k^T y_k + \eta_k = 2(f_k - f_{k+1} + s_k^T g_{k+1}) \\ &= s_k^T \nabla^2 f(\zeta x_k + (1 - \zeta)x_{k+1}) s_k \leq \|\nabla^2 f(\bar{x})\| \|s_k\|^2. \end{aligned}$$

Therefore, to complete the proof it is enough to let $m = \bar{\mu}$ and $M = \|\nabla^2 f(\bar{x})\|$. \square

Lemma 4. Under Assumptions 1, for u_k defined by (2.9) we have

$$\|u_k\| \leq (L + 1)\|s_k\|.$$

Proof. From (3.1) and since $\lambda_k \in [0, 1]$, we have

$$\|u_k\| = \|(1 - \lambda_k)y_k + \lambda_k s_k\| \leq \|y_k\| + \|s_k\| \leq L\|s_k\| + \|s_k\| = (L + 1)\|s_k\|. \quad \square$$

Lemma 5. Under Assumptions 1, if f is a uniformly convex function on \mathcal{N} , then there exists a positive constant ρ such that for u_k defined by (2.9) we have

$$s_k^T u_k \geq \rho \|s_k\|^2.$$

Proof. From Theorem 2, we have

$$\begin{aligned} s_k^T u_k &= (1 - \lambda_k)s_k^T y_k + \lambda_k s_k^T s_k \\ &\geq (1 - \lambda_k)\mu \|s_k\|^2 + \lambda_k \|s_k\|^2 = [(1 - \lambda_k)\mu + \lambda_k] \|s_k\|^2. \end{aligned} \quad (3.8)$$

So, from (3.8), to complete the proof it is enough to let $\rho = \min\{\mu, 1\}$. \square

Lemma 6. Under Assumptions 1, if f is twice continuously differentiable and uniformly convex, then we have $|\eta_k| \leq M \|s_k\|^2$, with the positive constant M as given in Lemma 3 and η_k as defined by (2.6).

Proof. From Taylor's theorem, we can write

$$\begin{aligned} f_k - f_{k+1} + g_{k+1}^T s_k &= \frac{1}{2} s_k^T \nabla^2 f(x_\varsigma) s_k, \\ f_k - f_{k+1} + g_k^T s_k &= -\frac{1}{2} s_k^T \nabla^2 f(x_\iota) s_k, \end{aligned}$$

where $x_\varsigma = \varsigma x_k + (1 - \varsigma)x_{k+1}$, $x_\iota = \iota x_k + (1 - \iota)x_{k+1}$, for some $\varsigma, \iota \in (0, 1)$. Therefore, from Lemma 3 we have

$$\begin{aligned} |\eta_k| &= |2(f_k - f_{k+1}) + (g_k + g_{k+1})^T s_k| \\ &= \frac{1}{2} |s_k^T (\nabla^2 f(x_\varsigma) - \nabla^2 f(x_\iota)) s_k| \leq \frac{1}{2} \|\nabla^2 f(x_\varsigma) - \nabla^2 f(x_\iota)\| \|s_k\|^2 \\ &\leq \frac{1}{2} (\|\nabla^2 f(x_\varsigma)\| + \|\nabla^2 f(x_\iota)\|) \|s_k\|^2 \leq M \|s_k\|^2. \quad \square \end{aligned}$$

Now, we can prove the following global convergence theorems for our hybrid CG method with the CG parameter (2.1).

Theorem 4. Suppose that Assumptions 1 hold. Consider a CG method in the form of (1.2)–(1.3) with the CG parameter β_k^C defined by (2.1) and the hybridization parameter $\theta_k \in [0, 1]$, in which, for all $k \geq 0$, the descent condition (3.2) holds and the steplength α_k is computed such that the strong Wolfe line search conditions (1.4) and (1.6) are satisfied. If f is uniformly convex and there exists a positive constant ξ such that

$$\|g_{k+1}\|^2 \leq \xi \|s_k\|, \quad (3.9)$$

then the method converges in the sense that (3.4) holds.

Proof. From (2.2) and since $\theta_k \in [0, 1]$, we have

$$\|d_{k+1}\| \leq \|g_{k+1}\| + \left(\frac{|g_{k+1}^T y_k|}{|s_k^T y_k|} + \frac{\|g_{k+1}\|^2}{|s_k^T y_k|} \right) \|s_k\|. \quad (3.10)$$

From (3.1), Lemma 2 and Theorem 2, we have

$$\|g_{k+1}\| \leq \gamma, \quad \frac{|g_{k+1}^T y_k|}{|s_k^T y_k|} \leq \frac{L\gamma\|s_k\|}{\mu\|s_k\|^2} = \frac{L\gamma}{\mu\|s_k\|}. \quad (3.11)$$

From (3.9) and Theorem 2, we have

$$\frac{\|g_{k+1}\|^2}{|s_k^T y_k|} \leq \frac{\xi\|s_k\|}{\mu\|s_k\|^2} = \frac{\xi}{\mu\|s_k\|}. \quad (3.12)$$

Now, from (3.10), (3.11) and (3.12), we have

$$\|d_{k+1}\| \leq \gamma + \frac{L\gamma}{\mu} + \frac{\xi}{\mu}. \quad (3.13)$$

So, from (3.13) and Lemma 1 the proof is complete. \square

Note that if $\|g_k\| \rightarrow 0$, then near the optimal solution we may have $\|g_{k+1}\|^2 \approx 0$. Therefore, the inequality (3.9) may be satisfied for large enough values of ξ .

Theorem 5. *Suppose that Assumptions 1 hold. Consider a CG method in the form of (1.2)–(1.3) with the CG parameter β_k^C defined by (2.1) and the hybridization parameter θ_k defined by (2.13), in which, for all $k \geq 0$, the descent condition (3.2) holds and the steplength α_k is computed such that the strong Wolfe line search conditions (1.4) and (1.6) are satisfied. If $\theta_k \in (0, 1)$ and f is twice continuously differentiable and uniformly convex, then the method converges in the sense that (3.4) holds.*

Proof. Since $\theta_k \in (0, 1)$, from (2.2) and (2.13), we have

$$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T y_k - s_k^T g_{k+1} + \eta_k \frac{g_{k+1}^T u_k}{s_k^T u_k}}{s_k^T y_k + \eta_k} s_k,$$

and hence,

$$\|d_{k+1}\| \leq \|g_{k+1}\| + \frac{|g_{k+1}^T y_k - s_k^T g_{k+1} + \eta_k \frac{g_{k+1}^T u_k}{s_k^T u_k}|}{|s_k^T y_k + \eta_k|} \|s_k\|. \quad (3.14)$$

From (3.1) and Lemma 2, we have

$$\|g_{k+1}\| \leq \gamma, \quad |g_{k+1}^T y_k| \leq \|g_{k+1}\|L\|s_k\| \leq L\gamma\|s_k\|, \quad |s_k^T g_{k+1}| \leq \gamma\|s_k\|. \quad (3.15)$$

From Lemmas 4, 5 and 6, we have

$$|\eta_k| \frac{|g_{k+1}^T u_k|}{|s_k^T u_k|} \leq M \|s_k\|^2 \frac{\|g_{k+1}\| \|u_k\|}{\rho \|s_k\|^2} \leq M \frac{(L+1)\gamma \|s_k\|}{\rho}. \quad (3.16)$$

Now, from (3.14), (3.15) and (3.16), and Lemma 3, we have

$$\begin{aligned} \|d_{k+1}\| &\leq \gamma + \frac{L\gamma \|s_k\| + \gamma \|s_k\| + M \frac{(L+1)\gamma \|s_k\|}{\rho}}{m \|s_k\|^2} \|s_k\| \\ \implies \|d_{k+1}\| &\leq \gamma \left(1 + (L+1) \frac{\rho + M}{\rho m} \right). \end{aligned} \quad (3.17)$$

So, from (3.17) and Lemma 1 the proof is complete. \square

To prove the global convergence of our hybrid CG method with the CG parameter (2.14), the following important result, established by Zoutendijk [29] and Wolfe [19, 21], is needed.

Theorem 6. *Suppose that Assumptions 1 hold and consider any iterative method in the form of (1.2) for solving the unconstrained optimization problem (1.1), in which, for all $k \geq 0$, the search direction d_k is a descent direction and the steplength α_k satisfies the Wolfe line search conditions (1.4) and (1.5). Then*

$$\sum_{k=0}^{\infty} \cos^2 \phi_k \|g_k\|^2 < \infty, \quad \cos \phi_k = -\frac{g_k^T d_k}{\|g_k\| \|d_k\|}. \quad (3.18)$$

DEFINITION 3. [11] Inequality (3.18) is called the Zoutendijk condition.

In our analysis, the following property, namely Property (*), plays a significant role. This property, which ensures that β_k is small when the step s_k is small, has been proposed by Gilbert and Nocedal [11].

Property ().* Consider a method of the form (1.2)–(1.3), and suppose that

$$0 < \bar{\gamma} \leq \|g_k\| \leq \gamma, \quad (3.19)$$

for all $k \geq 0$. Under this assumption, we say that the method has Property (*) if there exist constants $b > 1$ and $\lambda > 0$ such that for all k ,

$$|\beta_k| \leq b, \quad (3.20)$$

and

$$\|s_k\| \leq \lambda \implies |\beta_k| \leq \frac{1}{2b}. \quad (3.21)$$

Theorem 7. [11] *Suppose that Assumptions 1 hold. Consider the method (1.2)–(1.3) with the following three properties: (i) $\beta_k \geq 0$, for all $k \geq 0$; (ii) the line search satisfies $\{x_k\}_{k \geq 0} \subset \mathcal{L}$, the sufficient descent condition (3.3) and the Zoutendijk condition (3.18); (iii) Property (*) holds. Then, the method converges in the sense that (3.4) holds.*

Now, we can state the global convergence theorem of our hybrid CG method with the CG parameter (2.14).

Theorem 8. *Suppose that Assumptions 1 hold. Consider the CG method (1.2)–(1.3) with the CG parameter β_k^{C+} defined by (2.14), in which, for all $k \geq 0$, the sufficient descent condition (3.3) holds and the steplength α_k is computed such that the Wolfe line search conditions (1.4) and (1.5) are satisfied. If there exists a positive constant ϖ so that*

$$|\theta_k| \leq \varpi \|s_k\|, \tag{3.22}$$

then the method converges in the sense of establishing (3.4).

Proof. First, note that because of the (sufficient) descent condition and (1.4), the sequence $\{x_k\}_{k \geq 0}$ is a subset of the level set \mathcal{L} . Also, from Theorem 6 the Zoutendijk condition holds. Therefore, considering Theorem 7, to complete the proof it is enough to show that the method has Property (*). In this context, from (2.8), (3.3) and (3.19), we can write

$$d_k^T y_k \geq -(1 - \sigma) d_k^T g_k \geq c(1 - \sigma) \|g_k\|^2 \geq c(1 - \sigma) \bar{\gamma}^2, \tag{3.23}$$

and from (3.5), we have

$$|g_{k+1}^T y_k| \leq \|g_{k+1}\| \|g_{k+1} - g_k\| \leq 2\gamma^2. \tag{3.24}$$

Therefore, since $\theta_k \in [0, 1]$, from (2.14), (3.5), (3.23) and (3.24), we have

$$\begin{aligned} \beta_k^{C+} &= (1 - \theta_k) \max\{\beta_k^{HS}, 0\} + \theta_k \beta_k^{DY} \leq \frac{|g_{k+1}^T y_k|}{d_k^T y_k} + \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k} \\ &\leq \frac{2\gamma^2}{c(1 - \sigma)\bar{\gamma}^2} + \frac{\gamma^2}{c(1 - \sigma)\bar{\gamma}^2} = \frac{3\gamma^2}{c(1 - \sigma)\bar{\gamma}^2}. \end{aligned} \tag{3.25}$$

Also, from (3.1), (3.5), (3.22) and (3.23), we have

$$\begin{aligned} \beta_k^{C+} &\leq \frac{|g_{k+1}^T y_k|}{d_k^T y_k} + \theta_k \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k} \\ &\leq \frac{L\gamma \|s_k\|}{c(1 - \sigma)\bar{\gamma}^2} + \varpi \|s_k\| \frac{\gamma^2}{c(1 - \sigma)\bar{\gamma}^2} = \frac{L\gamma + \varpi\gamma^2}{c(1 - \sigma)\bar{\gamma}^2} \|s_k\|. \end{aligned} \tag{3.26}$$

Therefore, from (3.25) and (3.26), if we let

$$b = \frac{3\gamma^2}{c(1 - \sigma)\bar{\gamma}^2} \quad \text{and} \quad \lambda = \frac{c(1 - \sigma)\bar{\gamma}^2}{2b(L\gamma + \varpi\gamma^2)},$$

then (3.20) and (3.21) hold. Therefore, the method has Property (*). \square

Note that near the optimal solution the objective function is approximately similar to a convex quadratic function. So, as mentioned in Section 2, in some neighborhood of the optimal solution we have $\eta_k \approx 0$. Now, if $|g_{k+1}^T g_k| \geq \bar{\gamma}^2$, which seem to be reasonable when (3.19) holds [6], then from (2.13) and (3.5) we have

$$|\theta_k| \approx \left| -\frac{s_k^T g_{k+1}}{g_{k+1}^T g_k} \right| \leq \frac{\gamma}{\bar{\gamma}^2} \|s_k\|.$$

Thus, in such situation inequality (3.22) may be satisfied.

4 Numerical Experiments

Here, we present some numerical comparisons made by applying MATLAB and C++ implementations of our hybrid CG methods and several efficient nonlinear CG methods proposed in [3, 5, 11, 13, 20, 24]. All the experiments were done on a computer, Intel(R) Core(TM)2 Duo CPU 2.00 GHz, with 1 GB of RAM. In this section, a CG method in the form of (1.2)–(1.3) with $\beta_k = \beta_k^C$ defined by (2.1) and the hybridization parameter θ_k defined by (2.13) is called a hybrid CG method. Note that different hybrid CG methods are characterized by different choices for λ_k in (2.9).

Efficiency comparisons were made using the performance profile of Dolan and Moré [10] on the running time and the total number of function and gradient evaluations (N_T) defined by [13]

$$N_T = N_f + 3N_g,$$

where N_f and N_g respectively are the number of function and gradient evaluations. Performance profile gives, for every $\omega \geq 1$, the proportion $p(\omega)$ of the test problems that each considered algorithmic variant has a performance within a factor of ω of the best.

Since CG methods show their efficiency on large-scale problems, our experiments have been done on a set of 100 unconstrained optimization test problems of the CUTER collection [12] with the minimum dimension being equal to 100, as specified in Table 1.

In our MATLAB implementations, if

$$g_k^T d_k > -10^{-10} \|g_k\| \|d_k\|,$$

then we considered d_k as an uphill search direction. Although the descent condition (3.2) may not always hold, the uphill search direction seldom occurred in our numerical experiments. When encountered with an uphill search direction, we restarted the algorithm with $d_k = -g_k$ [5]. We used the strong Wolfe line search conditions (1.4) and (1.6) with $\delta = 0.01$ and $\sigma = 0.1$ [5, 16], and computed the steplength α_k using Algorithm 3.5 in [19]. It is important that the choice of the initial trial value may remarkably affect the numerical results, so that its improper setting would cause a significant increase in the number of function and gradient evaluations to achieve a required accuracy [5]. Here, for the first iteration, we set the initial trial value to $1/\|g_0\|_\infty$, and for the subsequent iterations, we set it to $\|s_{k-1}\|/\|d_k\|$. Also, all attempts to solve the test problems were limited to reaching a maximum of 10 000 iterations or achieving a solution with $\|g_k\|_\infty < 10^{-6}$.

The first part of our numerical experiments is devoted to provide a numerical support for our choice of λ_k . In this context, similar to the methodology suggested in [17], our first hybrid CG method with λ_k defined by (2.12), here called M1, is compared with the two other hybrid CG methods M1B and M1W, respectively with the best choice $\lambda_k = 0.96$ and the worst choice $\lambda_k = 0.20$ among the different constant values $\lambda_k \in \{0.01k\}_{k=1}^{99}$. For computing λ_k by (2.12) in M1, among the different values for $C \in \{10^{-k}\}_{k=0}^{10}$ in (2.11), we set $C = 10^{-8}$ because of its promising numerical results. Furthermore, similar to

Table 1. The test problems and their dimensions.

Function	n	Function	n
ARGLINA	100	DIXMAANL	3000
ARGLINA	200	DIXMAANL	9000
ARWHEAD	100	DQDRTIC	500
ARWHEAD	1000	DQRTIC	1000
BROWNAL	100	DQRTIC	5000
BROWNAL	200	EG2	1000
BROYDN7D	500	ENGVAL1	100
BROYDN7D	1000	EXTROSNB	100
BRYBND	1000	EXTROSNB	1000
BRYBND	5000	FLETCBV2	1000
BRYBND	10000	FLETCBV2	5000
CHAINWOO	100	FLETCBV2	10000
CHAINWOO	4000	FMINSRF2	5625
COSINE	100	FMINSRF2	10000
COSINE	1000	FMINSRF2	15625
COSINE	10000	GENROSE	100
CURLY10	10000	GENROSE	500
CURLY20	10000	LIARWHD	1000
CURLY30	10000	LIARWHD	5000
DIXMAANA	1500	LIARWHD	10000
DIXMAANA	3000	NONDIA	1000
DIXMAANA	9000	NONDIA	5000
DIXMAANB	1500	NONDIA	10000
DIXMAANB	3000	PENALTY1	100
DIXMAANB	9000	PENALTY1	500
DIXMAANC	1500	PENALTY1	1000
DIXMAANC	3000	QUARTC	1000
DIXMAANC	9000	QUARTC	5000
DIXMAAND	1500	QUARTC	10000
DIXMAAND	3000	SCHMVETT	100
DIXMAAND	9000	SCHMVETT	500
DIXMAANE	1500	SCHMVETT	1000
DIXMAANE	3000	SPARSQUR	1000
DIXMAANE	9000	SPARSQUR	5000
DIXMAANF	1500	SPARSQUR	10000
DIXMAANF	3000	SPMSRTL	1000
DIXMAANF	9000	SPMSRTL	4999
DIXMAANG	1500	SPMSRTL	10000
DIXMAANG	3000	SROSENBR	1000
DIXMAANG	9000	SROSENBR	5000
DIXMAANH	1500	SROSENBR	10000
DIXMAANH	3000	TOINTGSS	1000
DIXMAANH	9000	TOINTGSS	5000
DIXMAANI	1500	TOINTGSS	10000
DIXMAANI	3000	VAREIGVL	100
DIXMAANI	9000	VAREIGVL	500
DIXMAANJ	1500	VAREIGVL	1000
DIXMAANJ	3000	WOODS	1000
DIXMAANJ	9000	WOODS	4000
DIXMAANL	1500	WOODS	10000

the approach of [28], we set $r = 1$, if $\|g_{k-1}\| > \epsilon$, and $r = r_0$, otherwise, in which among the different values for $\epsilon \in \{0.01, 0.1, 0.2, \dots, 0.9, 1\}$ and $r_0 \in$

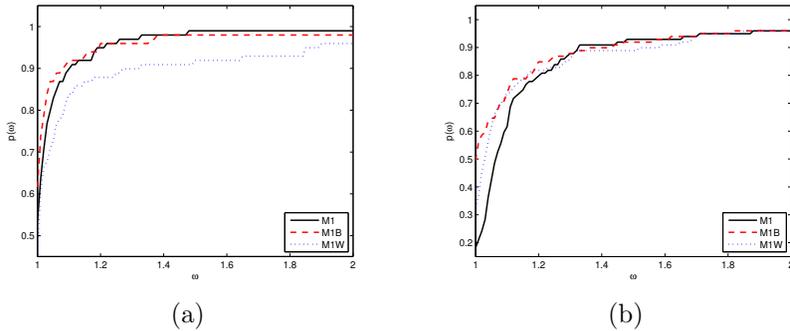


Figure 1. (a) Total number of function and gradient evaluations and (b) the CPU time performance profiles for M1, M1B and M1W.

$\{1, 2, \dots, 6\}$, we set $\epsilon = 0.1$ and $r_0 = 2$ because of their better numerical results. Figures 1(a) and 1(b) show the comparison results. As shown in Figure 1(a), M1 and M1B are competitive with respect to the total number of function and gradient evaluations. Also, Figure 1(b) shows that M1B and M1W are often preferable to M1 with respect to the running time. This seems reasonable because of the additional time spent for the computation of λ_k by (2.12) in M1. Therefore, the formula (2.12) for computation of λ_k seems to be practically effective with respect to the total number of function and gradient evaluations. Also, M1 in 24%, M1B in 31%, and M1W in 16% of the cases solved the test problems with the least total number of function and gradient evaluations. In addition, M1 in 19%, M1B in 49%, and M1W in 32% of the cases solved the test problems with the least running time.

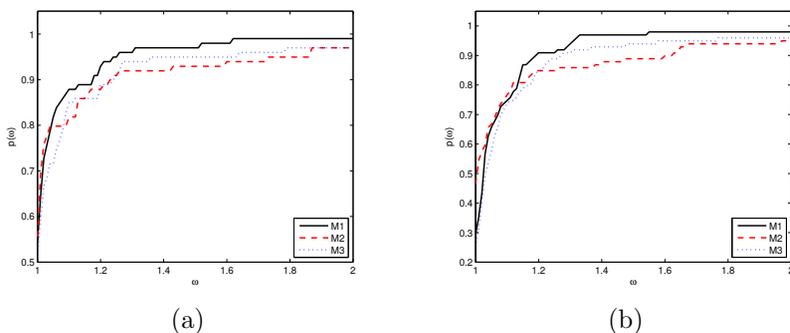


Figure 2. (a) Total number of function and gradient evaluations and (b) the CPU time performance profiles for M1, M2 and M3.

The second part of our numerical experiments is devoted to compare the CG method M1 with two other hybrid CG methods with $\lambda_k = 1$ as proposed in [3] and $\lambda_k = 0$ as proposed in [5], here respectively called M2 and M3, which

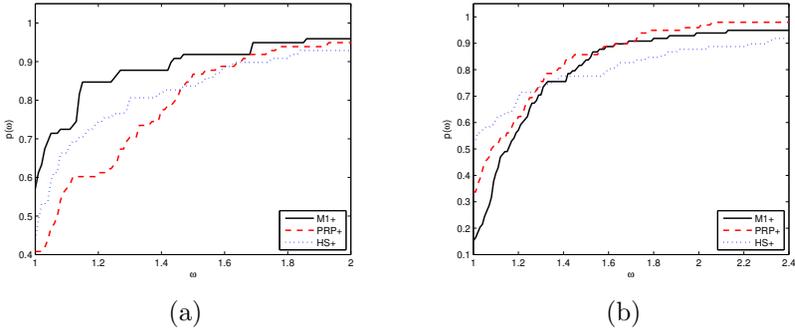


Figure 3. (a) Total number of function and gradient evaluations and (b) the CPU time performance profiles for M1+, PRP+ and HS+.

are globally convergent for uniformly convex objective functions. Figures 2(a) and 2(b) show the comparison results. As shown in these figures, although M1 outperforms M2 and M3 both with respect to the total number of function and gradient evaluations and the running time, M3 at times is preferable to M1 with respect to the running time. Also, M1 in 28%, M2 in 29%, and M3 in 25% of the cases solved the test problems with the least total number of function and gradient evaluations. In addition, M1 in 29%, M2 in 46%, and M3 in 25% of the cases solved the test problems with the least running time.

The third part of our numerical experiments is devoted to compare our second hybrid CG method with the CG parameter $\beta_k = \beta_k^C$ defined by (2.14) and the hybridization parameter θ_k defined by (2.13) with λ_k defined by (2.12), here called M1+, with the CG methods PRP+ and HS+, which are globally convergent for general objective functions. Figures 3(a) and 3(b) show the comparison results. As shown in Figure 3(a), M1+ outperforms PRP+ and HS+ with respect to the total number of function and gradient evaluations. Moreover, with respect to the running time, as shown in Figure 3(b), although PRP+ outperforms M1+, at times M1+ is preferable to HS+. Also, M1+ in 41%, PRP+ in 25%, and HS+ in 28% of the cases solved the test problems with the least total number of function and gradient evaluations. In addition, M1+ in 16%, PRP+ in 33%, and HS+ in 51% of the cases solved the test problems with the least running time.

The final part of our numerical experiments, made by a C++ implementation, is devoted to compare our second hybrid CG method M1+ with the CG methods of CG_Descent proposed by Hager and Zhang [13], and DPRP-W proposed by Yu et al. [24]. Here, for M1+, similar to the CG_Descent algorithm, we used the approximate Wolfe conditions proposed by Hager and Zhang [13] in the line search procedure, with all the parameter values suggested in [13]. Figures 4(a) and 4(b) show the comparison results. As shown in Figure 4(a), M1+ is preferable to CG_Descent and DPRP-W with respect to the total number of function and gradient evaluations. In addition, in a reasonable agreement with the numerical results of [24], DPRP-W outperforms CG_Descent with re-

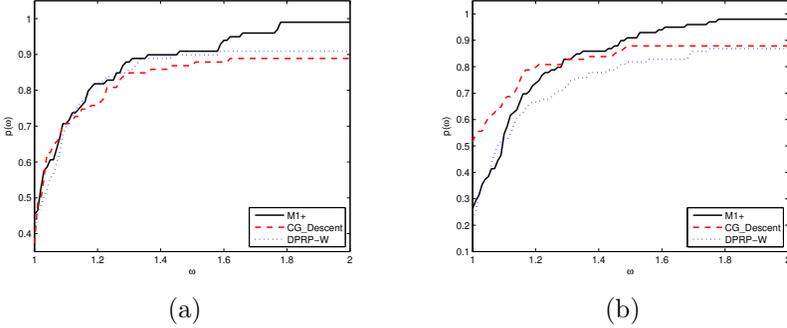


Figure 4. (a) Total number of function and gradient evaluations and (b) the CPU time performance profiles for M1+, CG_Descent and DPRP-W.

spect to the total number of function and gradient evaluations. Moreover, with respect to the running time, as shown in Figure 4(b), although M1+ outperforms DPRP-W, CG_Descent is often preferable to M1+. Also, M1+ in 36%, CG_Descent in 27%, and DPRP-W in 30% of the cases solved the test problems with the least total number of function and gradient evaluations. In addition, M1+ in 26%, CG_Descent in 51%, and DPRP-W in 23% of the cases solved the test problems with the least running time.

5 Conclusions

Based on Andrei’s approach of hybridizing the conjugate gradient (CG) methods by a convex combination of the CG parameters, we proposed a hybridization of Hestenes–Stiefel and Dai–Yuan methods which is globally convergent for uniformly convex functions. Then, in order to achieve the global convergence for general functions, we applied the Powell’s nonnegative restriction of the CG parameters [20] on our hybrid CG parameter and proposed another modified hybrid CG method. In our methods, the hybridization parameter is obtained by using a recently proposed hybrid secant equation. Numerical comparisons of the implementations of our first method (M1) with two hybrid CG methods proposed by Andrei (M2) [2] and Babaie-Kafaki et al. (M3) [5], being special cases of our first hybrid CG method and globally convergent for uniformly convex functions, and our second hybrid CG method (M1+) with the efficient CG methods of PRP+ and HS+ proposed by Gilbert and Nocedal [11], DPRP-W proposed by Yu et al. [24], and CG_Descent proposed by Hager and Zhang [13], being globally convergent for general functions, were made on a set of unconstrained optimization test problems from the CUTer collection, using the Dolan–Moré performance profile [10]. The results showed a promising numerical performance for our hybrid CG methods, specially with respect to the total number of function and gradient evaluations.

As a future work, it would be interesting to develop a new formula for the hybridization parameter θ_k using the modified secant equation proposed by

Zhang et al. [26], or updating the CG parameters β_k^{HS} and β_k^{DY} based on the suggested hybrid secant equation (or its revised form which can be obtained based on the approach of [16]), and numerically compare the performance of the corresponding hybrid CG methods with the methods proposed here. Also, the approaches of [7, 17, 23, 27] can be possibly applied on the proposed hybrid CG methods in order to develop descent CG methods.

References

- [1] N. Andrei. Numerical comparison of conjugate gradient algorithms for unconstrained optimization. *Studies in Informatics and Control*, **16**(4):333–352, 2007.
- [2] N. Andrei. Another hybrid conjugate gradient algorithm for unconstrained optimization. *Numer. Algorithms*, **47**(2):143–156, 2008. <http://dx.doi.org/10.1007/s11075-007-9152-9>.
- [3] N. Andrei. Accelerated hybrid conjugate gradient algorithm with modified secant condition for unconstrained optimization. *Numer. Algorithms*, **54**(1):23–46, 2010. <http://dx.doi.org/10.1007/s11075-009-9321-0>.
- [4] S. Babaie-Kafaki. A modified BFGS algorithm based on a hybrid secant equation. *Sci. China Math.*, **54**(9):2019–2036, 2011. <http://dx.doi.org/10.1007/s11425-011-4232-7>.
- [5] S. Babaie-Kafaki, M. Fatemi and N. Mahdavi-Amiri. Two effective hybrid conjugate gradient algorithms based on modified BFGS updates. *Numer. Algorithms*, **58**(3):315–331, 2011. <http://dx.doi.org/10.1007/s11075-011-9457-6>.
- [6] S. Babaie-Kafaki and R. Ghanbari. Two hybrid nonlinear conjugate gradient methods based on a modified secant equation. *Optimization*, 2012. <http://dx.doi.org/10.1080/02331934.2012.693083>.
- [7] W. Cheng and Q. Liu. Sufficient descent nonlinear conjugate gradient methods with conjugacy condition. *Numer. Algorithms*, **53**(1):113–131, 2010. <http://dx.doi.org/10.1007/s11075-009-9318-8>.
- [8] Y.H. Dai. New properties of a nonlinear conjugate gradient method. *Numer. Math.*, **89**(1):83–98, 2001. <http://dx.doi.org/10.1007/PL00005464>.
- [9] Y.H. Dai, J.Y. Han, G.H. Liu, D.F. Sun, H.X. Yin and Y.X. Yuan. Convergence properties of nonlinear conjugate gradient methods. *SIAM J. Optim.*, **10**(2):348–358, 1999. <http://dx.doi.org/10.1137/s1052623494268443>.
- [10] E.D. Dolan and J.J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, **91**(2, Ser. A):201–213, 2002. <http://dx.doi.org/10.1007/s101070100263>.
- [11] J.C. Gilbert and J. Nocedal. Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.*, **2**(1):21–42, 1992. <http://dx.doi.org/10.1137/0802003>.
- [12] N.I.M. Gould, D. Orban and P.L. Toint. CUTER: a constrained and unconstrained testing environment, revisited. *ACM Trans. Math. Software*, **29**(4):373–394, 2003. <http://dx.doi.org/10.1145/962437.962439>.
- [13] W.W. Hager and H. Zhang. Algorithm 851: CG_Descent, a conjugate gradient method with guaranteed descent. *ACM Trans. Math. Software*, **32**(1):113–137, 2006. <http://dx.doi.org/10.1145/1132973.1132979>.

- [14] W.W. Hager and H. Zhang. A survey of nonlinear conjugate gradient methods. *Pac. J. Optim.*, **2**(1):35–58, 2006.
- [15] D.H. Li and M. Fukushima. A modified BFGS method and its global convergence in nonconvex minimization. *J. Comput. Appl. Math.*, **129**(1–2):15–35, 2001. [http://dx.doi.org/10.1016/s0377-0427\(00\)00540-9](http://dx.doi.org/10.1016/s0377-0427(00)00540-9).
- [16] G. Li, C. Tang and Z. Wei. New conjugacy condition and related new conjugate gradient methods for unconstrained optimization. *J. Comput. Appl. Math.*, **202**(2):523–539, 2007. <http://dx.doi.org/10.1016/j.cam.2006.03.005>.
- [17] I.E. Livieris and P. Pintelas. A descent Dai–Liao conjugate gradient method based on a modified secant equation and its global convergence. *ISRN Computational Mathematics*, **2012**:1–8, 2012. <http://dx.doi.org/10.5402/2012/435495>. Article ID 435495
- [18] J.L. Nazareth. Conjugate-gradient methods. In C. Floudas and P. Pardalos(Eds.), *Encyclopedia of Optimization*, volume I (A–D), pp. 319–323, Dordrecht and Boston, 2001. Kluwer Academic Publishers. <http://dx.doi.org/10.1007/0-306-48332-7-69>.
- [19] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer, New York, 2006.
- [20] M.J.D. Powell. Nonconvex minimization calculations and the conjugate gradient method. In D.F. Griffiths(Ed.), *Numerical Analysis (Dundee, 1983)*, volume 1066 of *Lecture Notes in Math.*, pp. 122–141, Berlin, 1984. Springer. <http://dx.doi.org/10.1007/BFb0099521>.
- [21] W. Sun and Y.X. Yuan. *Optimization Theory and Methods: Nonlinear Programming*. Springer, New York, 2006.
- [22] Z. Wei, G. Li and L. Qi. New quasi-Newton methods for unconstrained optimization problems. *Appl. Math. Comput.*, **175**(2):1156–1188, 2006. <http://dx.doi.org/10.1016/j.amc.2005.08.027>.
- [23] G. Yu, L. Guan and W. Chen. Spectral conjugate gradient methods with sufficient descent property for large-scale unconstrained optimization. *Optim. Methods Softw.*, **23**(2):275–293, 2008. <http://dx.doi.org/10.1080/10556780701661344>.
- [24] G. Yu, L. Guan and G. Li. Global convergence of modified Polak–Ribière–Polyak conjugate gradient methods with sufficient descent property. *J. Ind. Manag. Optim.*, **4**(3):565–579, 2008. <http://dx.doi.org/10.3934/jimo.2008.4.565>.
- [25] Y.X. Yuan. A modified BFGS algorithm for unconstrained optimization. *IMA J. Numer. Anal.*, **11**(3):325–332, 1991. <http://dx.doi.org/10.1093/imanum/11.3.325>.
- [26] J. Zhang and C. Xu. Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations. *J. Comput. Appl. Math.*, **137**(2):269–278, 2001. [http://dx.doi.org/10.1016/s0377-0427\(00\)00713-5](http://dx.doi.org/10.1016/s0377-0427(00)00713-5).
- [27] L. Zhang, W. Zhou and D. Li. A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence. *IMA J. Numer. Anal.*, **26**(4):629–640, 2006. <http://dx.doi.org/10.1093/imanum/drl016>.
- [28] W. Zhou and L. Zhang. A nonlinear conjugate gradient method based on the MBFGS secant condition. *Optim. Methods Softw.*, **21**(5):707–714, 2006. <http://dx.doi.org/10.1080/10556780500137041>.
- [29] G. Zoutendijk. Nonlinear programming, computational methods. In J. Abadie(Ed.), *Integer and Nonlinear Programming*, pp. 37–86, Amsterdam, 1970. North-Holland.