Journal of Computational and Applied Mathematics 239 (2013) 396–405

Contents lists available at SciVerse ScienceDirect

Journal of Computational and Applied Mathematics

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

# A new class of spectral conjugate gradient methods based on a modified secant equation for unconstrained optimization

Ioannis E. Livieris *, Panagiotis Pintelas

*Department of Mathematics, University of Patras, GR 265-00, Greece*
*Educational Software Development Laboratory, Department of Mathematics, University of Patras, GR 265-00, Greece*

ARTICLE INFO

*Article history:*
Received 1 December 2010
Received in revised form 18 May 2012

*Keywords:*
Spectral conjugate gradient methods
Sufficient descent property
Modified secant equation
Line search
Global convergence

ABSTRACT

Conjugate gradient methods have played a special role for solving large scale optimization problems due to the simplicity of their iteration, convergence properties and their low memory requirements. In this work, we propose a new class of spectral conjugate gradient methods which ensures sufficient descent independent of the accuracy of the line search. Moreover, an attractive property of our proposed methods is that they achieve a high-order accuracy in approximating the second order curvature information of the objective function by utilizing the modified secant condition proposed by Babaie-Kafaki et al. [S. Babaie-Kafaki, R. Ghanbari, N. Mahdavi-Amiri, Two new conjugate gradient methods based on modified secant equations, Journal of Computational and Applied Mathematics 234 (2010) 1374–1386]. Further, a global convergence result for general functions is established provided that the line search satisfies the Wolfe conditions. Our numerical experiments indicate that our proposed methods are preferable and in general superior to the classical conjugate gradient methods in terms of efficiency and robustness.


## 1. Introduction

Let us consider the following unconstrained optimization problem:

$$\min \left\{ f(x) \mid x \in \mathbb{R}^n \right\}, \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function and its gradient is denoted by $g(x) = \nabla f(x)$.

Conjugate gradient methods constitute an excellent choice for efficiently solving the optimization problem (1.1), especially when the dimension $n$ is large due to the simplicity of their iteration, convergence properties and their low memory requirements. These methods generate a sequence of points $\{x_k\}$, starting from an initial point $x_0 \in \mathbb{R}^n$, using the iterative formula

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, \ldots, \tag{1.2}$$

where $\alpha_k > 0$ is the stepsize obtained by some line search and $d_k$ is the search direction defined by

$$d_k = \begin{cases} -g_0, & \text{if } k = 0; \\ -g_k + \beta_k d_{k-1}, & \text{otherwise}, \end{cases} \tag{1.3}$$

* Corresponding author at: Department of Mathematics, University of Patras, GR 265-00, Greece.
*E-mail addresses:* livieris@gmail.gr, livieris@upatras.gr (I.E. Livieris).

0377-0427/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2012.09.007

where $g_k = g(x_k)$ and $\beta_k$ is a scalar. Conjugate gradient methods differ in their way of defining the scalar parameter $\beta_k$. The most well-known formulas are the Hestenes–Stiefel [1], the Fletcher–Reeves [2], the Polak–Ribière [3] and the Perry [4] formulas which are specified by

$$\beta_k^{\mathrm{HS}} = \frac{g_k^T y_{k-1}}{y_{k-1}^T d_{k-1}}, \qquad \beta_k^{\mathrm{FR}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \beta_k^{\mathrm{PR}} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \qquad \beta_k^{\mathrm{P}} = \frac{g_k^T (y_{k-1} - s_{k-1})}{y_{k-1}^T d_{k-1}},$$

respectively, where $s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = g_k - g_{k-1}$ and $\|\cdot\|$ denotes the Euclidean norm. In case the objective function $f$ is strictly convex and the performed line search is exact, all choices for the update parameter are equivalent but for a general function, each choice for the update parameter leads to quite different computational efficiency and convergence properties. We refer to the books [5,6], the survey paper [7] and the references therein for more details about the numerical performance and the convergence properties of conjugate gradient methods.

Birgin and Martínez [8] introduced a more general form of conjugate gradient methods embedding the spectral gradient [9] in the conjugate gradient framework, where the search direction is determined by

$$d_k = \begin{cases} -g_0, & \text{if } k = 0; \\ -\dfrac{1}{\delta_k} g_k + \beta_k d_{k-1}, & \text{otherwise}, \end{cases} \tag{1.4}$$

where $\delta_k$ is a scalar. Using a geometric interpretation for quadratic function minimization, Birgin and Martinez suggested the following expressions for defining the update parameter $\beta_k$ in Eq. (1.4)

$$\beta_k^{\mathrm{SHS}} = \frac{g_k^T y_{k-1}}{\delta_k y_{k-1}^T d_{k-1}}, \qquad \beta_k^{\mathrm{SFR}} = \frac{\delta_{k-1}\|g_k\|^2}{\delta_k\|g_{k-1}\|^2}, \qquad \beta_k^{\mathrm{SPR}} = \frac{\delta_{k-1} g_k^T y_{k-1}}{\delta_k\|g_{k-1}\|^2}, \qquad \beta_k^{\mathrm{SP}} = \frac{g_k^T (y_{k-1} - \delta_k s_{k-1})}{\delta_k y_{k-1}^T d_{k-1}}.$$

Clearly, in case $\delta_k = \delta_{k-1} = 1$ these formulas are reduced to the respective classical formulas. Moreover, the authors presented some encouraging numerical results, in case $\delta_k$ is taken to be the spectral quotient [9], namely

$$\delta_k = \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}},$$

which lies between the minimum and the maximum eigenvalue of the average Hessian $\int_0^1 \nabla^2 f(x_{k-1} + \delta s_{k-1}) d\delta$. The motivation for this selection of $\delta_k$ constitutes in providing a two-point approximation to the secant equation underlying quasi-Newton methods [9].

During the last decade, much effort has been devoted to develop new conjugate gradient methods which not only possess strong convergence properties but are also computationally superior to the classical methods and are categorized in two classes. The first class utilizes second order information to improve the numerical efficiency of conjugate gradient methods based on modified secant equations (see [10–13]). Ford et al. [14] proposed a multi-step conjugate gradient method which is based on the multi-step quasi-Newton methods [11,12]. In more recent works, Li et al. [15] and Babaie et al. [10,16] proposed some conjugate gradient methods which are based on modified secant equations with higher order accuracy in the approximation of the curvature of the objective function. Under proper conditions, these methods are globally convergent for general functions and sometimes their numerical performance is superior to classical conjugate gradient methods. However, the main disadvantage of these methods is that they cannot ensure the generation of descent directions, therefore restarts are employed in their analysis and implementation in order to guarantee convergence.

The second class focuses on generating descent conjugate gradient methods independent of the accuracy of the line search. Similar to the spectral conjugate gradient method [8], Zhang et al. [17–19] considered to modify the search direction in order to develop conjugate gradient methods which generate descent directions using any line search. Independently, Hager and Zhang [20] considered a different approach to develop descent conjugate gradient methods by modifying the update parameter in Eq. (1.3), namely

$$\beta_k^{\mathrm{HZ}} = \beta_k^{\mathrm{HS}} - 2\frac{\|y_{k-1}\|^2}{(y_{k-1}^T d_{k-1})^2} g_k^T d_{k-1}.$$

More specifically, in their proposed method, called CG-DESCENT, the update parameter can be viewed as a modification of the HS formula by adding the extra term $-\mu g_k^T d_{k-1}$ with $\mu = 2\|y_{k-1}\|^2/(y_{k-1}^T d_{k-1})^2$ in order to ensure sufficient descent. Along this line, Yu et al. [21] by a method of undetermined coefficient, introduced a new class of spectral conjugate gradient methods in the following way

$$\beta_k^{\mathrm{DSHS}} = \beta_k^{\mathrm{SHS}} - \frac{C\|y_{k-1}\|^2}{\delta_k (y_{k-1}^T d_{k-1})^2} g_k^T d_{k-1}, \qquad \beta_k^{\mathrm{DSFR}} = \beta_k^{\mathrm{SFR}} - \frac{C\delta_{k-1}^2\|g_k\|^2}{\delta_k\|g_{k-1}\|^4} g_k^T d_{k-1},$$

$$\beta_k^{\mathrm{DSPR}} = \beta_k^{\mathrm{SPR}} - \frac{C\delta_{k-1}^2\|y_{k-1}\|^2}{\delta_k\|g_{k-1}\|^4} g_k^T d_{k-1}, \qquad \beta_k^{\mathrm{DSP}} = \beta_k^{\mathrm{SP}} - \frac{C\|y_{k-1} - \delta_k s_{k-1}\|^2}{\delta_k (y_{k-1}^T d_{k-1})^2} g_k^T d_{k-1},$$

where $C$ is a parameter which essentially controls the relative weight between conjugacy and descent and in case $C > 1/4$ then the generated directions are always descent. Notice that $\beta_k^{HZ}$ is a special case of $\beta_k^{DSHS}$ with $\delta_k = 1$ and $C = 2$. Moreover, Yu et al. proved that their proposed methods are globally convergent for uniformly convex functions under the Wolfe conditions. Recently, Yuan [22] motivated by the previous works [20,21,17] proposed a class of descent conjugate gradient methods which possess global convergence for general functions and are also computationally competitive to classical methods.

In this work, we propose some new spectral conjugate gradient methods which possess the advantages of the two previously discussed classes. More specifically, our proposed methods guarantee sufficient descent independent of the accuracy of the line search and they have the attractive property of achieving a high-order accuracy in approximating the second-order curvature information of the objective function by utilizing the modified secant condition proposed in [10]. Furthermore, we establish the global convergence of our proposed class of methods provided that the line search satisfies the Wolfe conditions.

The remainder of this paper is organized as follows. In Section 2, we present our motivation and a modification of Perry's spectral conjugate gradient method. In Section 3, we present the global convergence analysis for general functions and generalize our technique to the rest of the conjugate gradient methods. The numerical experiments are reported in Section 4 using the performance profiles of Dolan and Morè [23]. Finally, Section 5 presents our concluding remarks and our proposals for future research.

## 2. Modified spectral Perry conjugate gradient method

At this point, we recall that for quasi-Newton methods, an approximation matrix $B_{k-1}$ to the Hessian $\nabla^2 f(x_{k-1})$ is updated so that a new matrix $B_k$ satisfies the following secant condition

$$B_k s_{k-1} = y_{k-1}. \tag{2.1}$$

By expanding condition (2.1), Zhang et al. [24] and Zhang and Xu [25] proposed the following modified secant condition

$$B_{k-1} s_{k-1} = z_{k-1}, \tag{2.2}$$

with

$$\begin{cases} z_{k-1} = y_{k-1} + \dfrac{\theta_{k-1}}{s_{k-1}^T u} u, \\ \theta_{k-1} = 6(f_{k-1} - f_k) + 3(g_k + g_{k-1})^T s_{k-1}. \end{cases} \tag{2.3}$$

where $u \in \mathbb{R}^n$ is a vector parameter satisfying $s_{k-1}^T u \neq 0$ and $f_k = f(x_k)$. Notice that this new quasi-Newton equation contains not only gradient value information but also function value information at the present and the previous step. In addition, the theoretical advantages of the modified secant equation (2.2) can be seen from the following theorem.

**Theorem 2.1** (*[24]*). *Assume that the function $f$ is sufficiently smooth and $s_{k-1}$ is sufficiently small, then the following two estimates hold:*

$$s_{k-1}^T(\nabla^2 f(x_k) s_{k-1} - y_{k-1}) = O(\|s_{k-1}\|^3),$$
$$s_{k-1}^T(\nabla^2 f(x_k) s_{k-1} - z_{k-1}) = O(\|s_{k-1}\|^4).$$

Obviously, the above equations imply that the modified secant equation (2.2) is superior to the classical one (2.1) in the sense that $z_{k-1}$ better approximates $\nabla^2 f(x_k) s_{k-1}$ than $y_{k-1}$.

Quite recently, Babaie-Kafaki et al. [10] noticed that for values of $\|s_{k-1}\|$ greater than one (i.e. $\|s_{k-1}\| > 1$), the standard secant equation (2.1) is expected to be more accurate than the modified secant equation (2.2). In order to overcome this difficulty, the authors considered an extension of the modified secant equation (2.2) as follows:

$$B_{k-1} s_{k-1} = \tilde{z}_{k-1}, \qquad \tilde{z}_{k-1} = y_{k-1} + \rho_{k-1} \frac{\max\{\theta_{k-1}, 0\}}{s_{k-1}^T u} u, \tag{2.4}$$

where parameter $\rho_{k-1} \in \{0, 1\}$ and adaptively switch between the standard secant equation (2.1) and the modified secant equation (2.4), by setting $\rho_{k-1} = 1$ if $\|s_{k-1}\| \leq 1$ and setting $\rho_{k-1} = 0$, otherwise.

By taking into consideration the theoretical advantages of the modified secant equation (2.4), the computational efficiency of the descent spectral Perry conjugate gradient method [26,21] and the strong convergence properties of the conjugate gradient methods which have the property $\beta_k \geq 0$ [27,28], we propose a modification of formula $\beta_k^{DSP}$ as follows:

$$\beta_k^{MSP} = \frac{g_k^T(\tilde{z}_{k-1} - \tilde{\delta}_k s_{k-1})}{\tilde{\delta}_k \tilde{z}_{k-1}^T d_{k-1}} - \min\left\{ \frac{g_k^T(\tilde{z}_{k-1} - \tilde{\delta}_k s_{k-1})}{\tilde{\delta}_k \tilde{z}_{k-1}^T d_{k-1}}, \frac{C\|\tilde{z}_{k-1} - \tilde{\delta}_k s_{k-1}\|^2}{\tilde{\delta}_k (\tilde{z}_{k-1}^T d_{k-1})^2} g_k^T d_{k-1} \right\} \tag{2.5}$$

with $C > 1/4$. Formula $\beta_k^{MSP}$ satisfies the following lemma whose proof is exactly the same with that of Theorem 2.1 in [21], thus we omit it.

**Lemma 2.1.** *Consider any iterative method of the form (1.2) and (1.4) in which $\tilde{\delta}_k^* > 0$ and $\beta_k \in [\beta_k^{MSP}, \max\{\beta^{MSP}, 0\}]$. If $\tilde{z}_{k-1}^T d_{k-1} \neq 0$ then*

$$g_k^T d_k = -\left(1 - \frac{1}{4C}\right)\frac{1}{\tilde{\delta}_k}\|g_k\|^2. \tag{2.6}$$

Moreover, motivated in [8,9,29,21], parameter $\tilde{\delta}_k$ is selected in a spectral manner based on the Rayleigh quotient, namely

$$\tilde{\delta}_k = \frac{s_{k-1}^T \tilde{z}_{k-1}}{s_{k-1}^T s_{k-1}}. \tag{2.7}$$

Clearly, the above spectral step size minimizes the quantity $\|\tilde{\delta} s_{k-1} - \tilde{z}_{k-1}\|$ providing a scalar approximation to the modified secant equation (2.4) [30].

Now, based on the above discussion, we present our proposed algorithm called the modified spectral Perry conjugate gradient algorithm (MSP-CG).

**Algorithm 2.1** (*MSP-CG*).

Step 1: Choose an initial point $x_0 \in \mathbb{R}^n$ and $0 < \sigma_1 < \sigma_2 < 1$; Set $k = 0$.
Step 2: If $\|g_k\| = 0$, then terminate; Otherwise go to the next step.
Step 3: Compute the descent direction by Eq. (1.4), where $\beta_k$ is defined by (2.5).
Step 4: Determine a stepsize $\alpha_k$ using the Wolfe line search:

$$f(x_k + \alpha_k d_k) - f(x_k) \le \sigma_1 \alpha_k g_k^T d_k, \tag{2.8}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma_2 g_k^T d_k. \tag{2.9}$$

Step 5: Let $x_{k+1} = x_k + \alpha_k d_k$.
Step 6: Set $k = k + 1$ and go to Step 2.

**Remarks.** In Algorithm 2.1, since the line search satisfies the Wolfe line search conditions (2.8) and (2.9), it immediately follows that $\tilde{z}_{k-1}^T d_{k-1} \ge y_{k-1}^T d_{k-1} > 0$ for all $k$, which implies that formula (2.5) is well defined.

## 3. Convergence analysis

In order to present the global convergence analysis, we make the following assumptions on the objective function $f$, which have often been used in the literature [7,5,6] to establish the global convergence of conjugate gradient methods.

**Assumption 1.** The level set $\mathcal{L} = \{x \in \mathbb{R}^n | f(x) \le f(x_0)\}$ is bounded, namely, there exists a positive constant $B > 0$ such that

$$\|x\| \le B, \quad \forall x \in \mathcal{L}. \tag{3.1}$$

**Assumption 2.** In some neighborhood $\mathcal{N}$ of $\mathcal{L}$, $f$ is differentiable and its gradient $g$ is Lipschitz continuous, i.e. there exists a positive constant $L > 0$ such that

$$\|g(x) - g(y)\| \le L\|x - y\|, \quad \forall x, y \in \mathcal{N}. \tag{3.2}$$

Since $\{f_k\}$ is a decreasing sequence, it is clear that the sequence $\{x_k\}$ generated by Algorithm MSP-CG is contained in $\mathcal{L}$. In addition, it follows directly from Assumptions 1 and 2 that there exists a positive constant $M > 0$ such that

$$\|g(x)\| \le M, \quad \forall x \in \mathcal{L}. \tag{3.3}$$

To present the convergence analysis the following lemmas are needed.

**Lemma 3.1** (*[10]*). *Suppose that Assumptions 1 and 2 hold. For $\theta_{k-1}$ and $\tilde{z}_{k-1}$ defined by equations (2.3) and (2.4), respectively, we have*

$$|\theta_{k-1}| \le 3L\|s_{k-1}\|^2 \quad \text{and} \quad \|\tilde{z}_{k-1}\| \le 4L\|s_{k-1}\|.$$

The following lemma is a general result of conjugate gradient methods implemented with a line search that satisfies the Wolfe conditions (2.8) and (2.9).

**Lemma 3.2.** *Suppose that Assumptions 1 and 2 hold. Consider any method of the form (1.2) where $d_k$ is a descent direction, i.e. $d_k^T g_k < 0$ and $\alpha_k$ satisfies the Wolfe line search conditions (2.8) and (2.9), then*

$$\sum_{k\geq 0} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty.$$

Obviously, it follows from Lemma 3.2 and Eq. (2.6), that

$$\sum_{k\geq 0} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty, \tag{3.4}$$

which is very useful for the global convergence analysis.

In the following, we establish that the Algorithm MSP-CG is globally convergent for general nonlinear functions. In the rest of this section, we assume that the positive sequence $\{\tilde{\delta}_k\}$ is uniformly bounded, namely there exist positive constants $\tilde{\delta}_{\mathrm{MIN}}$ and $\tilde{\delta}_{\mathrm{MAX}}$, such that $\tilde{\delta}_{\mathrm{MIN}} \leq \tilde{\delta}_k \leq \tilde{\delta}_{\mathrm{MAX}}$.

For the purpose of showing the global convergence, we state some properties for the search direction $d_k$, formula $\beta_k^{\mathrm{MSP}}$ and step $s_{k-1}$. Firstly, we present a lemma which shows that, asymptotically, the search directions change slowly.

**Lemma 3.3.** *Suppose that Assumptions 1 and 2 hold. Let $\{x_k\}$ and $\{d_k\}$ be generated by Algorithm MSP-CG, if there exists a positive constant $\mu > 0$ such that for all $k$*

$$\|g_k\| \geq \mu, \tag{3.5}$$

*then $d_k \neq 0$ and*

$$\sum_{k\geq 1} \|w_k - w_{k-1}\|^2 < \infty,$$

*where $w_k = d_k/\|d_k\|$.*

**Proof.** Firstly, note that $d_k \neq 0$, for otherwise (2.6) would imply $g_k = 0$. Therefore, $w_k$ is well defined. Now, let us define

$$r_k := -\frac{1}{\tilde{\delta}_k} \frac{g_k}{\|d_k\|} \quad \text{and} \quad \theta_k := \beta_k^{\mathrm{MSP}} \frac{\|d_{k-1}\|}{\|d_k\|}, \tag{3.6}$$

Then, by Eq. (1.4), we have

$$w_k = r_k + \theta_k w_{k-1}. \tag{3.7}$$

Using this relation with the identity $\|w_k\| = \|w_{k-1}\| = 1$, we obtain

$$\|r_k\| = \|w_k - \theta_k w_{k-1}\| = \|w_{k-1} - \theta_k w_k\|.$$

Moreover, using this with the condition $\theta_k \geq 0$ and the triangle inequality, we get

$$\|w_k - w_{k-1}\| \leq \|w_k - \theta_k w_{k-1}\| + \|w_{k-1} - \theta_k w_k\| = 2\|r_k\|. \tag{3.8}$$

Therefore, using this relation with (3.4), we obtain

$$\sum_{k\geq 1} \|w_k - w_{k-1}\|^2 = \sum_{k\geq 1} 4\|r_k\|^2 \leq \frac{4}{\tilde{\delta}_{\mathrm{MIN}}^2 \mu^2} \sum_{k\geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty,$$

which completes the proof. □

Next, we present a lemma which shows that $\beta_k^{\mathrm{MSP}}$ will be small when the step $s_{k-1}$ is small which implies that the Algorithm MSP-CG prevents the inefficient behavior of the jamming phenomenon [31] from occurring. This property is similar to but slightly different from *Property* $(*)$, which was derived in [28].

**Lemma 3.4.** *Suppose that Assumptions 1 and 2 hold. Let $\{x_k\}$ and $\{d_k\}$ be generated by Algorithm MSP-CG, if there exists a positive constant $\mu > 0$ such that Eq. (3.5) holds; then there exist constants $b > 1$ and $\lambda > 0$ such that for all $k$:*

$$|\beta_k^{MSP}| \leq b \tag{3.9}$$

*and*

$$\|s_{k-1}\| \leq \lambda \Rightarrow |\beta_k^{MSP}| \leq \frac{1}{b}. \tag{3.10}$$

**Proof.** Utilizing Lemma 3.1 together with (2.4), (2.6), (2.9) and (3.5), we obtain

$$\tilde{z}_{k-1}^T d_{k-1} \geq y_{k-1}^T d_{k-1} \geq (\sigma_2 - 1) g_{k-1}^T d_{k-1} \geq (1-\sigma_2)\left(1 - \frac{1}{4C}\right)\frac{\mu^2}{4L}. \tag{3.11}$$

Moreover, observe that

$$g_k^T d_{k-1} = y_{k-1}^T d_{k-1} + g_{k-1}^T d_{k-1} < \tilde{z}_{k-1}^T d_{k-1}. \tag{3.12}$$

Again, the Wolfe condition (2.9) gives

$$g_k^T d_{k-1} \geq \sigma_2 g_{k-1}^T d_{k-1} \geq -\sigma_2 \tilde{z}_{k-1}^T d_{k-1} + \sigma_2 g_k^T d_{k-1}. \tag{3.13}$$

Since $\sigma_2 < 1$, by rearranging the previous inequality, we obtain

$$g_k^T d_{k-1} \geq -(\sigma_2/1-\sigma_2)\tilde{z}_{k-1}^T d_{k-1},$$

which together with (3.12), gives

$$\left|\frac{g_k^T d_{k-1}}{\tilde{z}_{k-1}^T d_{k-1}}\right| \leq \max\left\{\frac{\sigma_2}{(1-\sigma_2)}, 1\right\}. \tag{3.14}$$

From Assumption 2 together with Lemma 3.1 and relations (3.5), (3.11) and (3.14) we have

$$\begin{aligned}
|\beta_k^{\mathrm{MSP}}| &= \left|\frac{g_k^T(\tilde{z}_{k-1} - \tilde{\delta}_k s_{k-1})}{\tilde{\delta}_k \tilde{z}_{k-1}^T d_{k-1}} - \frac{C\|\tilde{z}_{k-1} - \tilde{\delta}_k s_{k-1}\|^2}{\tilde{\delta}_k (\tilde{z}_{k-1}^T d_{k-1})^2} g_k^T d_{k-1}\right| \\
&\leq \frac{\|g_k\|(\|\tilde{z}_{k-1}\| + \tilde{\delta}_k \|s_{k-1}\|)}{\tilde{\delta}_k |\tilde{z}_{k-1}^T d_{k-1}|} + C\frac{\|\tilde{z}_{k-1}\|^2 + 2\tilde{\delta}_k s_{k-1}^T \tilde{z}_{k-1} + \tilde{\delta}_k^2 \|s_{k-1}\|^2}{\tilde{\delta}_k |\tilde{z}_{k-1}^T d_{k-1}|^2} |g_k^T d_{k-1}| \\
&\leq \frac{32ML^2 + 128CBL^3 \max\{\sigma_2/(1-\sigma_2), 1\}}{\tilde{\delta}_{\mathrm{MIN}}(1-\sigma_2)(1-1/(4C))\mu^2}\|s_{k-1}\| \triangleq D\|s_{k-1}\|.
\end{aligned} \tag{3.15}$$

Therefore, by setting $b := \max\{2, 2DB\}$ and $\lambda := 1/Db$, we have that relations (3.9) and (3.10) hold. The proof is complete. □

Subsequently, making use of Lemmas 3.3 and 3.4, we establish the global convergence theorem for Algorithm MSP-CG for general functions.

**Theorem 3.1.** *Suppose that Assumptions 1 and 2 hold. If $\{x_k\}$ is obtained by Algorithm MSP-CG, then we have*

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{3.16}$$

**Proof.** Assume that the conclusion (3.16) is not true. Then there exists a positive constant $\mu > 0$ such that for all $k$, $\|g_k\| \geq \mu$. The proof is divided in the following steps:

Step I. A bound on the step $s_k$. Let $\delta$ be a positive integer, chosen large enough that $\delta \geq 4BD$, where $B$ and $D$ are defined in (3.1) and (3.15), respectively. For any $l > k \geq k_0$ with $l - k \leq \delta$, following the same proof as the case II of Theorem 3.2 in [20], we get

$$\sum_{j=k}^{l-1} \|s_j\| < 2B.$$

Step II. A bound on the directions $d_l$. It follows from the definition of $d_k$ in Eq. (1.4) with (3.1), that

$$\|d_l\|^2 \leq \left(\frac{1}{\tilde{\delta}_l^*}\|g_l\| + |\beta_l^{\mathrm{MSP}}|\,\|d_{l-1}\|\right)^2 \leq 2\left(\frac{M}{\tilde{\delta}_{\mathrm{MIN}}}\right)^2 + 2D^2\|s_{l-1}\|^2\|d_{l-1}\|^2.$$

Now, the remaining argument is standard in the same way as case III in Theorem 3.2 in [20], thus we omit it. This completes the proof. □

Using the same approach with Yu [26], Yu et al. [21] and Yuan [22], we modify the rest of the spectral conjugate gradient methods, accordingly

$$\beta_k^{\mathrm{MSHS}} = \frac{g_k^T \tilde{z}_{k-1}}{\tilde{\delta}_k \tilde{z}_{k-1}^T d_{k-1}} - \min\left\{ \frac{g_k^T \tilde{z}_{k-1}}{\tilde{\delta}_k \tilde{z}_{k-1}^T d_{k-1}}, \frac{C\|\tilde{z}_{k-1}\|^2}{\tilde{\delta}_k (\tilde{z}_{k-1}^T d_{k-1})^2} g_k^T d_{k-1} \right\}, \tag{3.17}$$

$$\beta_k^{\mathrm{MSFR}} = \frac{\tilde{\delta}_{k-1}\|g_k\|^2}{\tilde{\delta}_k \|g_{k-1}\|^2} - \min\left\{ \frac{\tilde{\delta}_{k-1}\|g_k\|^2}{\tilde{\delta}_k \|g_{k-1}\|^2}, \frac{C\tilde{\delta}_{k-1}^2 \|g_k\|^2}{\tilde{\delta}_k \|g_{k-1}\|^4} g_k^T d_{k-1} \right\}, \tag{3.18}$$

$$\beta_k^{\mathrm{MSPR}} = \frac{\tilde{\delta}_{k-1} g_k^T \tilde{z}_{k-1}}{\tilde{\delta}_k \|g_{k-1}\|^2} - \min\left\{ \frac{\tilde{\delta}_{k-1} g_k^T \tilde{z}_{k-1}}{\tilde{\delta}_k \|g_{k-1}\|^2}, \frac{C\tilde{\delta}_{k-1}^2 \|\tilde{z}_{k-1}\|^2}{\tilde{\delta}_k \|g_{k-1}\|^4} g_k^T d_{k-1} \right\}. \tag{3.19}$$

Similar to the MSP formula we can easily verify that if parameter $C > 1/4$, then all the above formulas ensure sufficient descent independent of the accuracy of the line search and under the Wolfe line search conditions (2.8) and (2.9) the corresponding methods are globally convergent.

**Theorem 3.2.** *Consider any conjugate gradient method of the form (1.2)–(1.4) where the update parameter $\beta_k$ in Eq. (1.4) is defined by one of the formulas (3.17)–(3.19), then the generated search directions guarantee the sufficient descent condition. Moreover, if the line search satisfies the Wolfe conditions (2.8) and (2.9) and the assumptions of Theorem 3.1 hold, then the corresponding conjugate gradient method is globally convergent for general functions.*

## 4. Experimental results

In this section, we report some numerical results on a set of 73 unconstrained optimization problems. These test problems with the given initial points can be found on Andrei Neculai's web site.[1] Each test function is made an experiment with the number of variable 1000, 5000 and 10 000, respectively.

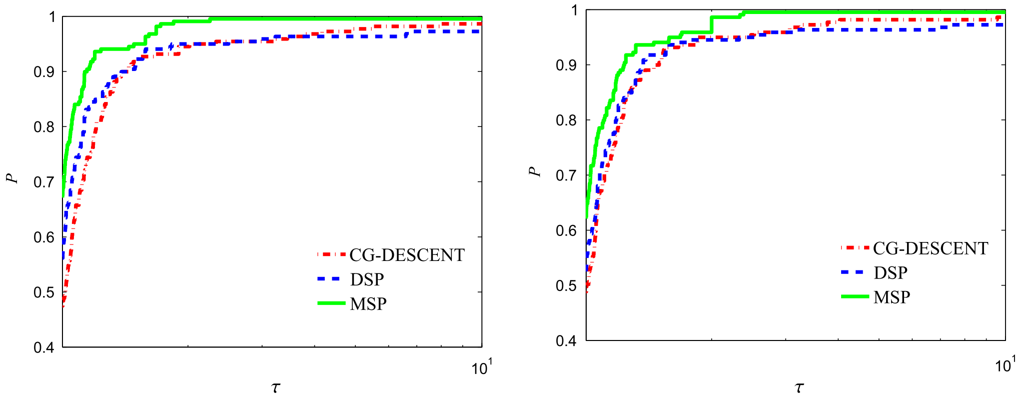
We evaluate the performance of our proposed conjugate gradient algorithm MSP-CG with that of the CG-DESCENT method [20] and the descent spectral Perry (DSP) conjugate gradient method [21]. The CG-DESCENT code is coauthored by Hager and Zhang obtained from Hager's web page.[2] We stop the iterations if the inequality $\|g_k\|_\infty \leq 10^{-6}$ is satisfied. We implemented Algorithm MSP-CG with the following parameters: $C = 0.5$, $\tilde{\delta}_{\mathrm{MIN}} = 10^{-10}$, $\tilde{\delta}_{\mathrm{MAX}} = 10^{10}$ and $u = s_{k-1}$ and in case $\tilde{\delta}_k$ does not belong to the interval $[\tilde{\delta}_{\mathrm{MIN}}, \tilde{\delta}_{\mathrm{MAX}}]$, then we set $\tilde{\delta}_k = \tilde{\delta}_{k-1}$ as in [21]. The implementation code was written in Fortran and compiled with `ifort` (with compiler settings `-O2 -double-size 128`) on a PC (2.66 GHz Quad-Core processor, 4 GB RAM) running Linux operating system. The detailed numerical results can be found in http://www.math.upatras.gr/~livieris/Results/MSCG.zip. Moreover, all methods were implemented with the same line search as CG-DESCENT.

The cumulative total for a performance metric over all simulations does not seem to be too informative, since a small number of simulations can tend to dominate these results. For this reason, we have used the performance profiles proposed in [23] to display the performance of each algorithms, in terms of function and gradient evaluations. The use of profiles provide a wealth of information such as solver efficiency, robustness and probability of success in compact form and eliminate the influence of a small number of problems on the benchmarking process and the sensitivity of results associated with the ranking of solvers [23]. The performance profile plots the fraction $P$ of problems for which any given method is within a factor $\tau$ of the best solver. The horizontal axis of the figure gives the percentage of the test problems for which a method is the fastest (efficiency), while the vertical axis gives the percentage of the test problems that were successfully solved by each method (robustness).

Fig. 1 presents the performance profiles of CG-DESCENT, DSP and MSP relative to the function and gradient evaluations. Clearly, MSP exhibits the best overall performance since it illustrates the best probability of being the optimal solver, outperforming CG-DESCENT, relative to both performance metrics. More analytically, the performance profile for function evaluations shows that MSP solves 65.3% of the test problems with the least number of function evaluations while CG-DESCENT and DSP solve about 46.5% and 53.8% of the test problems, respectively. As regards the gradient evaluations metric, the interpretation in Fig. 1(b) illustrates that MSP solves 59.8% of the test problems with the least number of gradient evaluations while CG-DESCENT and DSP solve about 48.4% and 51.6% of the test problems, respectively in the same situation. Moreover, it is worth mentioning that MSP is the only method that successfully solved all the test problems. Since all methods are implemented with the same line search, we conclude that the MSP appears to generate the best search directions, on average.

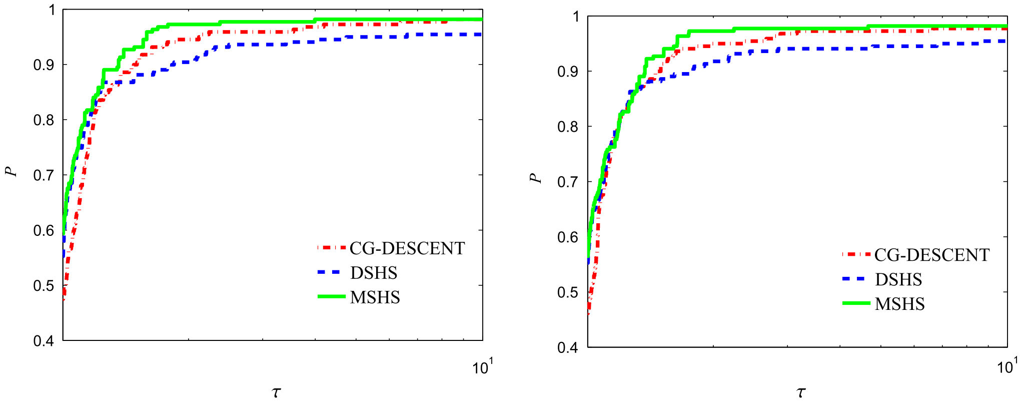
Subsequently, we continue our experimental analysis by evaluating the performance of the rest of our proposed spectral conjugate gradient with that of the CG-DESCENT and the corresponding classical descent spectral conjugate gradient

[1] http://camo.ici.ro/neculai/SCALCG/testuo.pdf
[2] http://www.math.ufl.edu/~hager/papers/CG



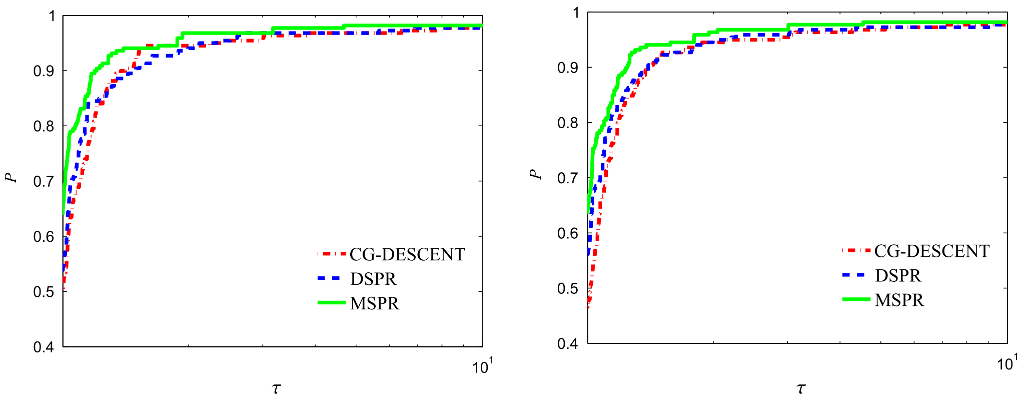

(a) Performance based on function evaluations. (b) Performance based on gradient evaluations.

**Fig. 1.** $\text{Log}_{10}$ scaled performance profiles of conjugate gradient methods CG-DESCENT, DSP and MSP.




(a) Performance based on function evaluations. (b) Performance based on gradient evaluations.

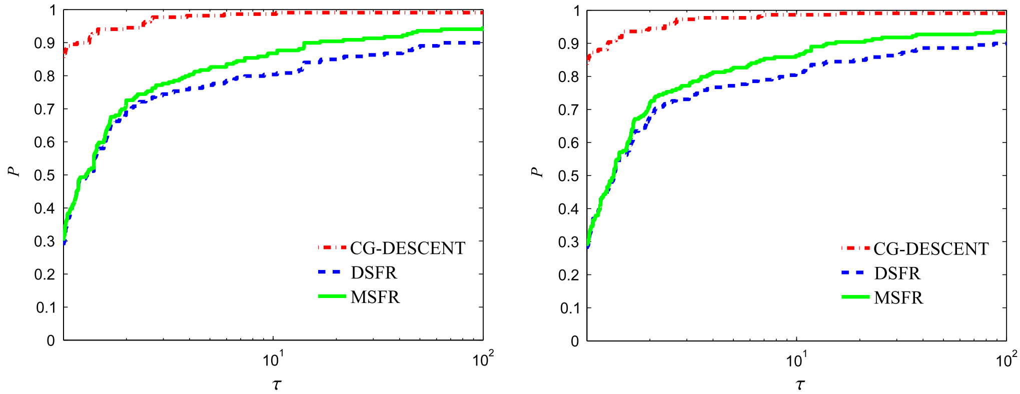
**Fig. 2.** $\text{Log}_{10}$ scaled performance profiles of conjugate gradient methods CG-DESCENT, DSHS and MSHS.




(a) Performance based on function evaluations. (b) Performance based on gradient evaluations.

**Fig. 3.** $\text{Log}_{10}$ scaled performance profiles of conjugate gradient methods CG-DESCENT, DSPR and MSPR.

methods [21]. More specifically, in Fig. 2, we compare the performance of the modified spectral Hestenes–Stiefel (MSHS) with that of the CG-DESCENT and the descent spectral Hestenes–Stiefel (DSHS), in Fig. 2 we evaluate the performance of the modified spectral Polak–Ribière (MSPR) with that of the CG-DESCENT and the descent spectral Polak–Ribière (DSPR) and in Fig. 4, we compare the performance of the modified spectral Fletcher–Reeves (MSFR) with that of the CG-DESCENT and the descent spectral Fletcher–Reeves (DSFR). The interpretation in Figs. 2–4 shows that the proposed methods outperform the classical methods relative to both performance metrics. Moreover, it is worth noticing that MSHS and MSPR outperform CG-DESCENT which implies that the proposed methods are computational efficient.




(a) Performance based on function evaluations. (a) Performance based on gradient evaluations.

**Fig. 4.** $\text{Log}_{10}$ scaled performance profiles of conjugate gradient methods CG-DESCENT, DSFR and MSFR.

## 5. Conclusions and future research

In this paper, we proposed a new class of spectral conjugate gradient methods which ensures the sufficient descent property independent of the accuracy of the line search. An important property of our proposed class of methods is that it achieves a high-order accuracy in approximating the second order curvature information of the objective function by utilizing the modified secant equation presented in [10]. Furthermore, the global convergence of our proposed methods has been established under the Wolfe line search conditions for general functions. Based on our numerical experiments, we concluded that the proposed methods are more efficient and more robust than the classical conjugate gradient methods, providing faster and more stable convergence.

Our future work is concentrated on studying the convergence properties of our proposed methods using different inexact line searches [32–36] and exploring different choices of the coefficient $\tilde{\delta}_k$.

## References

[1] M.R. Hestenes, E. Stiefel, Methods for conjugate gradients for solving linear systems, Journal of Research of the National Bureau of Standards 49 (1952) 409–436.
[2] R. Fletcher, C.M. Reeves, Function minimization by conjugate gradients, Computer Journal 7 (1964) 149–154.
[3] E. Polak, G. Ribière, Note sur la convergence de methods de directions conjuguees, Revue Francais d'Informatique et de Recherche Operationnelle 16 (1969) 35–43.
[4] A. Perry, A modified conjugate gradient algorithm, Operational Research 26 (1978) 1073–1078.
[5] Y.H. Dai, Y.X. Yuan, Nonlinear Conjugate Gradient Methods, Shanghai Scientific and Technical Publishers, Shanghai, 2000.
[6] J. Nocedal, S.J. Wright, Numerical Optimization, Springer-Verlag, New York, 1999.
[7] W.W. Hager, H. Zhang, A survey of nonlinear conjugate gradient methods, Pacific of Journal Optimization 2 (2006) 35–58.
[8] E.G. Birgin, J.M. Martínez, A spectral conjugate gradient method for unconstrained optimization, Applied Mathematics and Optimization 43 (1999) 117–128.
[9] J. Barzilai, J.M. Borwein, Two point step size gradient methods, IMA Journal of Numerical Analysis 8 (1988) 141–148.
[10] S. Babaie-Kafaki, R. Ghanbari, N. Mahdavi-Amiri, Two new conjugate gradient methods based on modified secant equations, Journal of Computational and Applied Mathematics 234 (2010) 1374–1386.
[11] J.A. Ford, I.A. Moghrabi, Multi-step quasi-Newton methods for optimization, Journal of Computational and Applied Mathematics 50 (1994) 305–323.
[12] J.A. Ford, I.A. Moghrabi, Using function-values multi-step quasi-Newton methods, Journal of Computational and Applied Mathematics 66 (1996) 201–211.
[13] Z. Wei, G. Yu, G. Yuan, Z. Lian, The superlinear convergence of a modified BFGS-type method for unconstrained optimization, Computational Optimization and Applications 29 (2004) 315–332.
[14] J.A. Ford, Y. Narushima, H. Yabe, Multi-step nonlinear conjugate gradient methods for unconstrained minimization, Computational Optimization and Applications 40 (2008) 191–216.
[15] G. Li, C. Tang, Z. Wei, New conjugacy condition and related new conjugate gradient methods for unconstrained optimization, Journal of Computational and Applied Mathematics 202 (2007) 523–539.
[16] S. Babaie-Kafaki, M. Fatemi, N. Mahdavi-Amiri, Two effective hybrid conjugate gradient algorithms based on modified BFGS updates, Numerical Algorithms 58 (3) (2011) 315–331.
[17] L. Zhang, W. Zhou, D. Li, Global convergence of a modified Fletcher–Reeves conjugate gradient method with Armijo-type line search, Numerische Mathematik 104 (2006) 561–572.
[18] L. Zhang, W. Zhou, D. Li, A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence, IMA Journal of Numerical Analysis 26 (4) (2006) 629–640.
[19] L. Zhang, W. Zhou, D. Li, Some descent three-term conjugate gradient methods and their global convergence, Optimization Methods and Software 22 (2007) 697–711.
[20] W.W. Hager, H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM of Journal Optimization 16 (2005) 170–192.
[21] G. Yu, L. Guan, W. Chen, Spectral conjugate gradient methods with sufficient descent property for large-scale unconstrained optimization, Optimization Methods and Software 23 (2) (2008) 275–293.
[22] G. Yuan, Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems, Optimization Letters 3 (2009) 11–21.
[23] E. Dolan, J.J. Moré, Benchmarking optimization software with performance profiles, Mathematical Programming 91 (2002) 201–213.

[24] J.Z. Zhang, N.Y. Deng, L.H. Chen, New quasi-Newton equation and related methods for unconstrained optimization, Journal of Optimization Theory and Applications 102 (1999) 147–167.
[25] J.Z. Zhang, C.X. Xu, Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations, Journal of Computational and Applied Mathematics 137 (2001) 269–278.
[26] G.H. Yu, Nonlinear self-scaling conjugate gradient methods for large-scale optimization problems, Ph.D. Thesis, Sun Yat-Sen University, 2007.
[27] Y.H. Dai, J. Han, G. Liu, D. Sun, H. Yin, Y. Yan, Convergence properties of nonlinear conjugate gradient methods, SIAM Journal of Optimization 2 (1999) 345–358.
[28] J.C. Gilbert, J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM Journal of Optimization 2 (1) (1992) 21–42.
[29] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, SIAM Journal of Optimization 7 (2007) 26–33.
[30] Y. Xiao, Q. Wang, D. Wang, Notes on the Dai–Yuan–Yuan modified spectral gradient method, Journal of Computational and Applied Mathematics 234 (10) (2010) 2986–2992.
[31] M.J.D. Powell, Restart procedures for the conjugate gradient method, Mathematical Programming 12 (1977) 241–254.
[32] L. Grippo, S. Lucidi, A globally convergent version of the Polak–Ribière conjugate gradient method, Mathematical Programming 78 (3) (1997) 375–391.
[33] J. Han, J. Sun, W. Sun, Global convergence of non-monotone descent methods for unconstrained optimization problems, Journal of Computational and Applied Mathematics 146 (2002) 89–98.
[34] Z.J. Shi, J. Guo, A new family of conjugate gradient methods, Journal of Computational and Applied Mathematics 224 (1) (2009) 444–457.
[35] Z.J. Shi, J. Shen, Convergence of nonmonotone line search method, Journal of Computational and Applied Mathematics 193 (2) (2006) 397–412.
[36] W. Sun, J. Han, J. Sun, Global convergence of nonmonotone descent methods for unconstrained optimization problems, Journal of Computational and Applied Mathematics 146 (2) (2002) 89–98.