UDC 517.9

Weijun Zhou (Changsha Univ. Sci. and Technology, China)

A GLOBALLY AND R-LINEARLY CONVERGENT HYBRID HS AND PRP METHOD AND ITS INEXACT VERSION WITH APPLICATIONS *

ГЛОБАЛЬНО ТА R-ЛІНІЙНО ЗБІЖНИЙ ГІБРИДНИЙ HS ТА PRP МЕТОД ТА ЙОГО НЕТОЧНА ВЕРСІЯ З ЗАСТОСУВАННЯМИ

We present a hybrid HS- and PRP-type conjugate gradient method for smooth optimization that converges globally and R-linearly for general functions. We also introduce its inexact version for problems of this sort whose gradients or function values are unavailable or difficult to compute. Moreover, we apply the inexact method to solve a nonsmooth convex optimization problem by converting it into a one-time continuously differentiable function by the method of Moreau – Yosida regularization.

Наведено гібридний HS та PRP метод спряженого аргументу, глобально та R-лінійно збіжний для загальних функцій. Також введено неточний метод для таких проблем, в яких градієнти або значення функцій невідомі або важко визначаються. Крім того, неточний метод застосовано до негладкої опуклої проблеми оптимізації, що перетворює її в однократно неперервно диференційовну функцію за допомогою регуляризації Моро – Йосіди.

1. Introduction. Conjugate gradient methods are a class of important methods for solving the large-scale unconstrained optimization problem

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

where $f\colon R^n \to R$ is continuously differentiable and its gradient is denoted by $g(x)$. A general scheme of conjugate gradient methods is

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

where $\alpha_k > 0$ is a stepsize, and the search direction d_k is given by

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if} \quad k \geq 1, \end{cases}$$

where β_k ia a parameter and $g_k = g(x_k)$. The Fletcher – Reeves (FR) method [9], the Polak – Ribiére – Polyak (PRP) method [17, 18], the Hestenes – Stiefel (HS) method [12] and the Dai – Yuan (DY) method [7] are several well-known nonlinear conjugate gradient algorithms [16]. They are specified by

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}},$$

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2},$$

* This work was supported by the NSF (11371073 and 11461015) of China, the Project of the Scientific Research Fund (12A004 and 13B137) of the Hunan Provincial Education Department, and the NSF (13JJ4062) of Hunan Province.

where $y_{k-1} = g_k - g_{k-1}$ and $\|\cdot\|$ stands for the Euclidean norm. If exact line search is used, they are equivalent in the sense that all yield the same search directions and converge globally and R-linearly for strongly convex functions [20]. However, for a general nonlinear function with inexact line search, their behavior is markedly different.

Since 1985, many efforts have been devoted to study the global convergence properties of various conjugate gradient methods with inexact line searches for general functions. Al-Baali [1] showed that the FR method can produce sufficient descent directions and converges for nonconvex functions with the strong Wolfe line search. Dai and Yuan [7] proved that the DY method is a descent method and globally convergent in the case of the standard Wolfe line search. However, the HS method and the PRP method may generate ascent directions even with the strong Wolfe line search [4], which prevent them from global convergence. To guarantee global convergence of the PRP method, some line searches which force it to generate descent direction were proposed [4, 14]. Recently, by the use of an approximate descent backtracking line search, Zhou [25] showed that the original PRP method converges globally even for nonconvex functions whether the search direction is descent or not.

A simple way for ensuring global convergence is that of using the steepest descent direction when the sufficient descent condition is violated. However, it is not guaranteed that the resulting algorithm will differ significantly from the steepest descent method. Some other globalization techniques for conjugate gradient methods also have been proposed when solving nonconvex optimization. The most famous one is the PRP+ globalization technique [13], namely, $\beta_k^{PRP+} = \max\{\beta_k^{PRP}, 0\}$. After this, almost all existing PRP type or HS type methods have adopted the PRP+ technique to obtain global convergence for nonconvex functions such as [6, 11, 21]. But these modifed methods can not reduce to the original PRP method when the exact line search is used and they are not the standard conjugate gradient methods any more in this sense.

To improve practical computation efficiency and convergence properties of conjugate gradient methods, many hybrid methods have been proposed, please see the recent survey [4] and references therein. These hybrid methods can be divided into two classes, one is the hybrid FR and PRP type methods such as the hybrid method [13] where $\beta_k = \max\left\{-\beta_k^{FR}, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\right\}$, another is the DY and HS type methods such as that of [5] where $\beta_k = \max\left\{0, \min\{\beta_k^{HS}, \beta_k^{DY}\}\right\}$.

To our knowledge, there is little study on hybrid HS and PRP type conjugate gradient methods. One purpose of the paper is to investigate this problem. In fact, we propose a sufficient descent hybrid HS and PRP method (1.4) below. Our motivation is based on the following two methods. One is the three-term PRP method proposed by Zhang, Zhou and Li [22], whose search direction is defined by

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k^{PRP} d_{k-1} - \theta_k^{PRP} y_{k-1}, & \text{if} \quad k \geq 1, \end{cases} \tag{1.2}$$

where $\theta_k^{PRP} = \dfrac{g_k^T d_{k-1}}{\|g_{k-1}\|^2}$. Another is the three-term HS method proposed by Zhang, Zhou and Li [24], which generates the search direction

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k^{HS} d_{k-1} - \theta_k^{HS} y_{k-1}, & \text{if} \quad k \geq 1, \end{cases} \tag{1.3}$$

where $\theta_k^{HS} = \dfrac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}}$.

It is clear that if the line search is exact, both methods reduce to the standard PRP method. Extensive numerical results [22, 24] show that both methods are very efficient. The three-term PRP method (1.2) converges globally for nonconvex functions [22]. The three-term HS method (1.3) converges globally and R-linearly for strongly convex functions [24], but it has not been proved to be globally convergent for general nonconvex functions. In order to utilize advantages of both methods sufficiently, based on (1.2) and (1.3), we propose a hybrid HS and PRP method as follows, namely,

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k^{\text{hybrid}} d_{k-1} - \theta_k^{\text{hybrid}} y_{k-1}, & \text{if} \quad k \geq 1, \end{cases} \tag{1.4}$$

where

$$\beta_k^{\text{hybrid}} = \frac{g_k^T y_{k-1}}{\max\left\{d_{k-1}^T y_{k-1}, \|g_{k-1}\|^2\right\}}, \qquad \theta_k^{\text{hybrid}} = \frac{g_k^T d_{k-1}}{\max\left\{d_{k-1}^T y_{k-1}, \|g_{k-1}\|^2\right\}}. \tag{1.5}$$

From (1.4), (1.5) and by direct computation, it is easy to get

$$g_k^T d_k = -\|g_k\|^2, \tag{1.6}$$

which is independent of convexity of the objective function and the line search used. It is clear that the proposed method reduces to the standard HS or PRP method when exact line search is used since $g_k^T d_{k-1} = 0$ in this case.

In general, conjugate gradient methods use the exact gradient and function values in their convergence analysis. However, in many practical problems, the exact function value or exact gradient value can not be obtained or may be very difficult to compute [3]. In these cases, the inexact algorithms are often required. Another purpose of the paper is to present an inexact conjugate gradient method only using approximate gradient or/and function values. In fact, we extend the above exact hybrid HS and PRP method to inexact case.

The paper is organized as follows. In next section, we prove the global and R-linear convergence of the proposed method with a descent backtracking line search for nonconvex optimization. In Section 3, we present the inexact algorithm in detail and show its global convergence by the use of some approximate function value descent line search. In Section 4, we apply the inexact method to solve a nonsmooth convex optimization problem by converting it into a once continuously differentiable function by way of the Moreau – Yosida regularization technique.

2. Exact algorithm and its convergence properties. In this section, based on the above discussion, we first describe the complete hybrid HS and PRP algorithm as follows.

***Algorithm* 2.1** (Exact version)**.**

***Step* 0.** Given an initial point $x_0 \in R^n$. Choose some constants $\delta > 0$ and $\rho \in (0, 1)$. Let $k := 0$.

***Step* 1.** Compute d_k by (1.4), (1.5).

***Step* 2.** Compute the stepsize $\alpha_k = \max\{\gamma_k \rho^j,\ j = 0, 1, 2, \ldots\}$ satisfying

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \delta\|\alpha_k d_k\|^2, \tag{2.1}$$

where $\gamma_k = \dfrac{|g_k^T d_k|}{\|d_k\|^2}$.

***Step* 3.** Let $x_{k+1} = x_k + \alpha_k d_k$.

***Step* 4.** Let $k := k + 1$ and go to Step 1.

To ensure global convergence of Algorithm 2.1, we make the following standard assumption.

Assumption 2.1. (i) *The level set* $\Omega = \big\{x \in R^n |\ f(x) \le f(x_0)\big\}$ *is bounded.*

(ii) *In some neighborhood* N *of* Ω, f *is continuously differentiable and its gradient is Lipschitz continuous, namely, there is a constant* $L > 0$ *such that*

$$\big\|g(x) - g(y)\big\| \le L\|x - y\| \quad \forall x, y \in N. \tag{2.2}$$

From Assumption 2.1 and the line search (2.1), we have

$$\sum_{k=0}^{\infty} \|\alpha_k d_k\|^2 < \infty,$$

which implies

$$\lim_{k\to\infty} \alpha_k \|d_k\| = 0. \tag{2.3}$$

Lemma 2.1. *Let Assumption* 2.1 *hold and* $\{x_k\}$ *be generated by Algorithm* 2.1. *Then there exists a constant* $M_1 > 0$ *such that*

$$\|g_k\| \le \|d_k\| \le M_1\|g_k\|. \tag{2.4}$$

Proof. By (1.6), we get

$$\|g_k\|^2 = |g_k^T d_k| \le \|g_k\|\|d_k\|,$$

which shows that $\|g_k\| \le \|d_k\|$. From (1.4), (1.5), (2.2) and the line search (2.1), we obtain

$$\|d_k\| \le \|g_k\| + 2L\|g_k\|\alpha_{k-1}\frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^2} \le (1 + 2L)\|g_k\| = M_1\|g_k\|,$$

where the last inequality follows from the fact $\alpha_k \le \gamma_k = \dfrac{|g_k^T d_k|}{\|d_k\|^2}$ and (1.6).

Lemma 2.1 is proved.

The following lemma gives a bound of the stepsize α_k from below.

Lemma 2.2. *Let Assumption* 2.1 *hold and* $\{x_k\}$ *be generated by Algorithm* 2.1. *Then there exists two positive constant* $\bar{m}_1$ *and* m_1 *such that*

$$\alpha_k \ge \bar{m}_1 \frac{-g_k^T d_k}{\|d_k\|^2} \ge m_1. \tag{2.5}$$

Proof. The proof of the first inequality in (2.5) is standard, for example, see Lemma 3.1 in [23]. The second inequality in (2.5) follows from (1.6) and (2.4) directly.

Theorem 2.1. *Let Assumption* 2.1 *hold and* $\{x_k\}$ *be generated by Algorithm* 2.1. *Then*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{2.6}$$

Proof. From (2.3) and (2.5), we get

$$\lim_{k\to\infty} \|d_k\| = 0,$$

which together with (2.4) yields (2.6).

The above theorem shows the global convergence property of Algorithm 2.1 without convexity assumption on f. It only relies on the assumption that f has Lipschitz continuous gradients.

Now we turn to establishing the R-linear convergence property of Algorithm 2.1. To do this, we need the following assumption.

Assumption 2.2. (i) f is twice continuously differentiable near x^*.

(ii) The sequence $\{x_k\}$ converges to x^* where $g(x^*) = 0$ and the Hessian matrix $\nabla^2 f(x^*)$ is positive definite.

Assumption 2.2 implies that f is strongly convex in some neighborhood $N(x^*)$ of x^*, that is, there are two positive constants m and M such that

$$m\|d\|^2 \le d^T \nabla^2 f(x) d \le M\|d\|^2 \quad \forall x \in N(x^*) \quad \forall d \in R^n. \tag{2.7}$$

From (2.7), it is easy to obtain (can see [2], Theorem 3.1)

$$\frac{m}{2}\|x - x^*\|^2 \le f(x) - f(x^*) \le \frac{1}{m}\big\|g(x)\big\|^2 \quad \forall x \in N(x^*). \tag{2.8}$$

By (2.4), (2.5) and (2.1), there is a positive constant m_2 such that

$$f(x_{k+1}) \le f(x_k) - \delta m_1^2 \frac{\|g_k\|^2}{\|d_k\|^2}\|g_k\|^2 \le f(x_k) - m_2\|g_k\|^2. \tag{2.9}$$

Without loss of generality, we assume $\{x_k\} \subset N(x^*)$. From (2.9) and (2.8), we get

$$f(x_{k+1}) - f(x^*) \le (1 - mm_2)\big(f(x_k) - f(x^*)\big) \le \ldots \le (1 - mm_2)^k\big(f(x_0) - f(x^*)\big). \tag{2.10}$$

The following theorem shows the R-linear convergence property of Algorithm 2.1.

Theorem 2.2. *Let Assumption* 2.2 *hold and the sequence* $\{x_k\}$ *be generated by Algorithm* 2.1. *Then there exist three positive constants* $r \in (0,1)$, C_2 *and* C_3 *such that*

$$f(x_{k+1}) - f(x^*) \le C_2 r^k, \qquad \|x_{k+1} - x^*\| \le C_3 \sqrt{r}^k.$$

Proof. The first inequality follows from (2.10) with $r = 1 - mm_2$ and $C_2 = f(x_0) - f(x^*)$ directly. Inequalities (2.10) and (2.8) yield the second inequality with $C_3 = \sqrt{\dfrac{2(f(x_0) - f(x^*))}{m}}$.

3. Inexact algorithm and its global convergence. In this section, we consider the inexact version of Algorithm 2.1 with approximate gradient or/and function values. For simplicity, we denote $f^a(x,\epsilon)$ and $g^a(x,\epsilon)$ as the approximations of $f(x)$ and $g(x)$ with the possible error ϵ, respectively. More accurately, we assume that, for each $x \in R^n$, the approximations $f^a(x,\epsilon)$ and $g^a(x,\epsilon)$ can be made arbitrarily close to the exact values $f(x)$ and $g(x)$ by choosing the parameter ϵ small enough, namely,

$$|f^a(x,\epsilon) - f(x)| \le \epsilon, \tag{3.1}$$

$$\|g^a(x,\epsilon) - g(x)\| \le \epsilon. \tag{3.2}$$

With these approximations, we define the inexact method of Algorithm 2.1 as follows:

$$d_k = \begin{cases} -g^a(x_k,\epsilon_k), & \text{if} \quad k = 0, \\ -g^a(x_k,\epsilon_k) + \beta_k d_{k-1} - \theta_k y^a_{k-1}, & \text{if} \quad k \ge 1, \end{cases} \tag{3.3}$$

where $y^a_{k-1} = g^a(x_k,\epsilon_k) - g^a(x_{k-1},\epsilon_{k-1})$,

$$\beta_k = \frac{g^a(x_k,\epsilon_k)^T y^a_{k-1}}{\max\left\{d^T_{k-1} y^a_{k-1}, \|g^a(x_{k-1},\epsilon_{k-1})\|^2\right\}}, \tag{3.4}$$

$$\theta_k = \frac{g^a(x_k,\epsilon_k)^T d_{k-1}}{\max\left\{d^T_{k-1} y^a_{k-1}, \|g^a(x_{k-1},\epsilon_{k-1})\|^2\right\}}. \tag{3.5}$$

It is clear that

$$d_k^T g^a(x_k,\epsilon_k) = -\left\|g^a(x_k,\epsilon_k)\right\|^2. \tag{3.6}$$

However, the direction d_k defined by (3.3)–(3.5) with inexact gradient $g^a(x_k,\epsilon_k)$ is not necessarily a descent direction of the objective function f at x_k. Then some line search procedures such as the Wolfe (or strong Wolfe) line search and the line search given by (2.1) can not be used any more. In this case, we have to modify the line search (2.1).

Let $\{\epsilon_k\}$ and η be a given positive sequence and a positive constant satisfying

$$\sum_{k=0}^{\infty} \epsilon_k \le \eta < \infty. \tag{3.7}$$

Set $\gamma_k = \dfrac{|g^a(x_k,\epsilon_k)^T d_k|}{\|d_k\|^2}$, we determine the stepsize by the following approximate descent line search, that is, compute $\alpha_k = \max\{\gamma_k \rho^j,\ j = 0,1,2,\ldots\}$ satisfying

$$f^a(x_k + \alpha_k d_k, \rho_1\epsilon_k) \le f^a(x_k,\epsilon_k) - \delta\|\alpha_k d_k\|^2 + 2\epsilon_k, \tag{3.8}$$

where $\rho, \rho_1 \in (0,1)$ are two constants.

The following result shows that the line search (3.8) terminates finitely.

Proposition 3.1. *The line search* (3.8) *is well-defined.*

Proof. Suppose it is not true. Then for all $j \ge 0$, (3.8) does not hold, namely,

$$f^a(x_k + \gamma_k\rho^j d_k, \rho_1\epsilon_k) > f^a(x_k,\epsilon_k) - \delta\|\gamma_k\rho^j d_k\|^2 + 2\epsilon_k, \tag{3.9}$$

which together with (3.1) yields

$$f(x_k + \gamma_k\rho^j d_k) + \rho_1\epsilon_k > f(x_k) - \epsilon_k - \delta\|\gamma_k\rho^j d_k\|^2 + 2\epsilon_k.$$

This implies

$$f(x_k + \gamma_k\rho^j d_k) - f(x_k) > -\delta\|\gamma_k\rho^j d_k\|^2 + (1-\rho_1)\epsilon_k.$$

Let $j \to \infty$ in the above inequality, we have

$$0 \geq (1 - \rho_1)\epsilon_k,$$

which is a contradiction since $\rho_1 \in (0, 1)$ and $\epsilon_k > 0$.

Proposition 3.1 is proved.

For clarity, we give the complete inexact algorithm as follows.

***Algorithm* 3.1** (Exact version)**.**

***Step* 0.** Given an initial point $x_0 \in R^n$. Choose some constants $\delta > 0$ and $\rho, \rho_1 \in (0, 1)$. Let $k := 0$.

***Step* 1.** Compute the search direction d_k by (3.3) – (3.5).

***Step* 2.** Compute the stepsize α_k by (3.8).

***Step* 3.** Let the next iterate be $x_{k+1} = x_k + \alpha_k d_k$.

***Step* 4.** Let $k := k + 1$. Go to Step 1.

We suppose that the following assumption is satisfied.

Assumption 3.1. (i) *The level set* $\Omega = \{x \in R^n |\ f(x) \leq f(x_0) + (3 + \rho_1)\eta\}$ *is bounded.*

(ii) *In some neighborhood* N *of* Ω, f *is continuously differentiable and its gradient is Lipschitz continuous*, *namely*, (2.2) *holds.*

It is clear that the sequence $\{x_k\} \subset \Omega$. In fact, from (3.1), (3.8) and (3.7), we have

$$f(x_{k+1}) \leq f^a(x_{k+1}, \rho_1\epsilon_k) + \rho_1\epsilon_k \leq f(x_k) + \epsilon_k - \delta\|\alpha_k d_k\|^2 + 2\epsilon_k + \rho_1\epsilon_k =$$

$$= f(x_k) - \delta\|\alpha_k d_k\|^2 + (3 + \rho_1)\epsilon_k \leq f(x_k) + (3 + \rho_1)\epsilon_k \leq f(x_0) + (3 + \rho_1)\eta. \tag{3.10}$$

Moreover, (3.10) and (3.7) imply that $\sum_{k=0}^{\infty} \|\alpha_k d_k\|^2 < \infty$, which shows

$$\lim_{k\to\infty} \alpha_k \|d_k\| = 0. \tag{3.11}$$

Lemma 3.1. *Let Assumption* 3.1 *hold and the sequence* $\{x_k\}$ *be generated by Algorithm* 3.1. *If* $\big\|g^a(x_k, \epsilon_k)\big\| \geq \tau_1$ *with some positive constant* τ_1 *for all* k, *then there exists a positive constant* M_3 *such that*

$$\|d_k\| \leq M_3. \tag{3.12}$$

Proof. From (2.2), (3.2), (3.7) and (3.11), we have

$$\|y_{k-1}^a\| = \|g^a(x_k, \epsilon_k) - g^a(x_{k-1}, \epsilon_{k-1})\| \leq$$

$$\leq \|g^a(x_k, \epsilon_k) - g_k\| + \|g_k - g_{k-1}\| + \|g^a(x_{k-1}, \epsilon_{k-1}) - g_{k-1}\| \leq$$

$$\leq \epsilon_k + \epsilon_{k-1} + L\|\alpha_{k-1} d_{k-1}\| \to 0. \tag{3.13}$$

This together with the assumption yield (3.12) by the same argument as that of Lemma 3.1 in [22].

Lemma 3.1 is proved.

Lemma 3.2. *Let Assumption* 3.1 *hold and the sequence* $\{x_k\}$ *be generated by Algorithm* 3.1. *Then there exist a constant* $m_3 > 0$ *such that*

$$\alpha_k \geq \frac{\|g^a(x_k, \epsilon_k)\|^2}{\|d_k\|^2} \quad \textit{or} \quad \alpha_k \geq m_3 \frac{\|g^a(x_k, \epsilon_k)\|^2}{\|d_k\|^2} - \frac{m_3\epsilon_k}{\|d_k\|}. \tag{3.14}$$

Proof. (i) If $\alpha_k = \gamma_k$, by (3.6), then the first inequality holds.

(ii) If $\alpha_k \neq \gamma_k$, then $\alpha_k' = \alpha_k/\rho$ can not satisfy the line search (3.8). This together with (3.1) yields

$$f(x_k + \alpha_k' d_k) > f(x_k) - \delta\|\alpha_k' d_k\|^2 + (1-\rho_1)\epsilon_k > f(x_k) - \delta\|\alpha_k' d_k\|^2.$$

By the mean value theorem and (2.2), it is easy to obtain that

$$f(x_k + \alpha_k' d_k) - f(x_k) \leq \alpha_k' g_k^T d_k + L\|\alpha_k' d_k\|^2.$$

Then the above two inequalities and (3.6) imply

$$\alpha_k' \geq \frac{-g_k^T d_k}{(L+\delta)\|d_k\|^2} = \frac{-g^a(x_k,\epsilon_k)^T d_k + (g^a(x_k,\epsilon_k) - g_k)^T d_k}{(L+\delta)\|d_k\|^2} \geq$$

$$\geq \frac{-g^a(x_k,\epsilon_k)^T d_k - \|g^a(x_k,\epsilon_k) - g_k\|\|d_k\|}{(L+\delta)\|d_k\|^2} \geq \frac{\|g^a(x_k,\epsilon_k)\|^2 - \epsilon_k\|d_k\|}{(L+\delta)\|d_k\|^2},$$

where the last inequality uses (3.2), which implies (3.14) with $m_3 = \dfrac{\rho}{L+\delta}$.

Lemma 3.2 is proved.

Theorem 3.1. *Let Assumption* 3.1 *hold. Then the sequence* $\{x_k\}$ *be generated by Algorithm* 3.1 *converges globally in the sense that*

$$\liminf_{k\to\infty} \big\|\nabla f(x_k)\big\| = 0. \tag{3.15}$$

Proof. Suppose it is not true. Then there exists a constant $\tau_1 > 0$ such that $\|g_k\| \geq 2\tau_1$ $\forall k \geq 0$, which together with (3.2) yields

$$\big\|g^a(x_k,\epsilon_k)\big\| \geq \tau_1 \tag{3.16}$$

for sufficiently large k. Then from Lemma 3.1 and (3.14), we have that for sufficiently large k,

$$\alpha_k \geq \frac{m_3}{M_3}\left(\frac{\tau_1^2}{M_3} - \epsilon_k\right) \geq \frac{m_3\tau_1^2}{2M_3^2},$$

which together with (3.11) means

$$\|d_k\| \to 0.$$

Then from the definition of the search direction (3.3) and (3.16), we have

$$\big\|g^a(x_k,\epsilon_k)\big\| \leq \|d_k\| + 2\frac{\|g^a(x_k,\epsilon_k)\|\|y_{k-1}^a\|\|d_{k-1}\|}{\|g^a(x_{k-1},\epsilon_{k-1})\|^2} \leq \|d_k\| + 2\frac{\|g^a(x_k,\epsilon_k)\|\|y_{k-1}^a\|\|d_{k-1}\|}{\tau_1^2} \to 0,$$

which contradicts (3.16).

Theorem 3.1 is proved.

***Remark* 3.1.** We can obtain the strong convergence of Algorithm 3.1, that is $\lim_{k\to\infty}\|g_k\| = 0$, by the same argument as that of Algorithm 2.1 if we suitably choose the positive sequence $\{\epsilon_k\}$. For example, let $\epsilon_{-1} = 1$, $\epsilon_k = \min\left\{\dfrac{\epsilon_{k-1}}{2}, \dfrac{1}{\|d_k\|}\right\}$ for $k \geq 0$, then the sequence $\{\epsilon_k\}$ satisfies $\epsilon_k < \epsilon_{k-1} \leq \dfrac{1}{\|d_{k-1}\|}$, and $\sum_{k=0}^{\infty} \epsilon_k < \infty$.

Lemma 3.3 ([8], Lemma 3.3). *Let* $\{a_k\}$ *and* $\{r_k\}$ *be positive sequences satisfying* $a_{k+1} \le (1 + r_k)a_k + r_k$ *and* $\sum_{k=0}^{\infty} r_k < \infty$. *Then* $\{a_k\}$ *converges.*

***Corollary* 3.1.** *Let Assumption* 3.1 *hold and the sequence* $\{x_k\}$ *be generated by Algorithm* 3.1. *If the function* f *is convex, then the sequence* $\{f(x_k)\}$ *converges to the minimum of* (1.1).

Proof. From Assumption 3.1 and Theorem 3.1, there exists a subsequence $\{x_{k_i}\}_{i=0}^{\infty}$ which converges to some minimizer x^* satisfying $g(x^*) = 0$. From (3.10), we have

$$f(x_{k+1}) - f(x^*) \le f(x_k) - f(x^*) + (3 + \rho_1)\epsilon_k, \tag{3.17}$$

which together with Lemma 3.3 and (3.7) shows that the sequence $\{f(x_k) - f(x^*)\}$ converges. Since the subsequence $\{f(x_{k_i})\}$ converges to $f(x^*)$, therefore $\{f(x_k)\}$ converges to $f(x^*)$.

Corollary 3.1 is proved.

4. Application to nonsmooth convex optimization. In this section, we consider the following convex optimization problem:

$$\min F(x), \quad x \in R^n, \tag{4.1}$$

where $F \colon R^n \to R$ is a possibly nondifferentiable convex function. Then general methods such as conjugate gradient methods for smooth optimization can not be used to solve (4.1) directly.

An efficient way is to convert the nonsmooth problem (4.1) into an equivalent smooth problem by the Moreau–Yosida regularization such as [10, 15, 19], that is,

$$\min f(x), \quad x \in R^n, \tag{4.2}$$

where f is defined by

$$f(x) = \min_{z \in R^n} \left\{ F(z) + \frac{1}{2\lambda}\|z - x\|^2 \right\} \tag{4.3}$$

and λ is a positive parameter. It is well-known that problems (4.1) and (4.2) are equivalent in the sense that the solution sets of the two problems coincide with each other. Moreover, the function f is convex and differentiable with Lipschitz continuous gradient [10, 19] given by $g(x) = \dfrac{1}{\lambda}\big(x - p(x)\big)$, which satisfies

$$\|g(x) - g(y)\| \le \frac{1}{\lambda}\|x - y\| \quad \forall x, y \in R^n, \tag{4.4}$$

where $g(x) = \nabla f(x)$ and $p(x)$ is the unique minimizer in (4.3), i.e.,

$$p(x) = \arg\min_{z \in R^n} \left\{ F(z) + \frac{1}{2\lambda}\|z - x\|^2 \right\}$$

since this is a strongly convex minimization problem.

It is clear that it is impossible in general to compute exactly the function f defined by (4.3) and its gradient g at an arbitrary point x. But for each $x \in R^n$, we may obtain approximate values of the gradient and the function by some existing methods such as [10, 19]. Therefore we can suppose that, for each $x \in R^n$, we can evaluate $f(x)$ and $g(x)$ approximately but with any desired accuracy, that is, for each $x \in R^n$ and any $\epsilon > 0$, we can find a vector $p^a(x, \epsilon) \in R^n$ such that

$$F\big(p^a(x, \epsilon)\big) + \frac{1}{2\lambda}\|p^a(x, \epsilon) - x\|^2 \le f(x) + \epsilon. \tag{4.5}$$

With this $p^a(x, \epsilon)$, we define the approximations to $f(x)$ and $g(x)$ by

$$f^a(x,\epsilon) = F\big(p^a(x,\epsilon)\big) + \frac{1}{2\lambda}\|p^a(x,\epsilon) - x\|^2 \tag{4.6}$$

and

$$g^a(x,\epsilon) = \frac{1}{\lambda}\big(x - p^a(x,\epsilon)\big), \tag{4.7}$$

respectively. The following lemma shows that the approximations $f^a(x,\epsilon)$ and $g^a(x,\epsilon)$ satisfy (3.1) and (3.2), respectively.

Lemma 4.1 ([10], Lemma 3.1)**.** *Let* $p^a(x,\epsilon)$ *be a vector satisfying* (4.5), $f^a(x,\epsilon)$ *and* $g^a(x,\epsilon)$ *be defined by* (4.6) *and* (4.7), *respectively. Then*

$$f(x) \le f^a(x,\epsilon) \le f(x) + \epsilon \qquad \textit{and} \qquad \|g^a(x,\epsilon) - g(x)\| \le \sqrt{\frac{2\epsilon}{\lambda}}.$$

From (4.4), Lemma 4.1 and Corollary 3.1, we have the following result.

***Corollary* 4.1.** *Let the problem* (4.2) *be solved by Algorithm* 3.1. *If the condition* (i) *in Assumption* 3.1 *holds*, *then the sequence* $\{f(x_k)\}$ *converges to the minimum of* (4.2).

5. Conclusions. We have proposed a hybrid HS and PRP method which converges globally and R-linearly for general optimization problems. It is also extended to inexact case which admits approximate function and gradient values. Hence this inexact method is very suitable for solving such problems whose exact gradient and function values are not available or difficult to compute. We have applied this inexact algorithm to solve nonsmooth convex problems by way of Moreau – Yosida regularization. We believe that the basic idea of this paper can be applied to other conjugate gradient methods. How to extend the proposed methods or linear conjugate gradient methods to fully derivative-free ones for solving large-scale nonlinear equations will be our further study.

1. *Al-Baali M.* Descent property and global convergence of the Fletcher – Reeves method with inexact line search // IMA J. Numer. Anal. – 1985. – **5**. – P. 121 – 124.
2. *Byrd R. H., Nocedal J.* A tool for the analysis of quasi-Newton methods with application to unconstrained minimization // SIAM J. Numer. Anal. – 1989. – **26**. – P. 727 – 739.
3. *Conn A. R., Scheinberg K., Toint Ph. L.* Recent progress in unconstrained nonlinear optimization without derivatives // Math. Program. – 1997. – **79**. – P. 397 – 414.
4. *Dai Y.* Nonlinear conjugate gradient methods // http://lsec.cc.ac.cn/˜ dyh/worklist.html.
5. *Dai Y., Yuan Y.* An efficient hybrid conjugate gradient method for unconstrained optimization // Ann. Oper. Res. – 2001. – **103**. – P. 33 – 47.
6. *Dai Y., Liao L. Z.* New conjugacy conditions and related nonlinear conjugate gradient methods // Appl. Math. and Optim. – 2001. – **43**. – P. 87 – 101.
7. *Dai Y., Yuan Y.* A nonlinear conjugate gradient method with a strong global convergence property // SIAM J. Optim. – 2000. – **10**. – P. 177 – 182.
8. *Dennis J. E., More J. J.* A characterization of superlinear convergence and its applications to quasi-Newton methods // Math. Comput. – 1974. – **28**. – P. 549 – 560.
9. *Fletcher R., Reeves C.* Function minimization by conjugate gradients // Comput. J. – 1964. – **7**. – P. 149 – 154.
10. *Fukushima M., Qi L.* A globally and superlinearly convergent algorithm for nonsmooth convex minimization // SIAM J. Optim. – 1996. – **6**. – P. 1106 – 1120.
11. *Hager W. W., Zhang H.* A new conjugate gradient method with guaranteed descent and an efficient line search // SIAM J. Optim. – 2005. – **16**. – P. 170 – 192.
12. *Hestenes M. R., Stiefel E. L.* Method of conjugate gradient for solving linear systems // J. Res. Nat. Bur. Stand. – 1952. – **49**. – P. 409 – 432.
13. *Gilbert J. C., Nocedal J.* Global convergence properties of conjugate gradient methods for optimization // SIAM J. Optim. – 1992. – **2**. – P. 21 – 42.

14. *Grippo L., Lucidi S.* A globally convergent version of the Polak – Ribière conjugate gradient method // Math. Program. – 1997. – **78**. – P. 375 – 391.
15. *Mifflin R., Sun D., Qi L.* Quasi-Newton bundle-type methods for nondifferentiable convex optimization // SIAM J. Optim. – 1998. – **8**. – P. 583 – 603.
16. *Nazareth L.* Conjugate-gradient methods // Encycl. Optim. / Eds C. Floudas and P. Pardalos. – Boston, USA; Dordrecht, The Netherlands: Kluwer Acad. Publ., 2001. – P. 319 – 323.
17. *Polak E., Ribière G.* Note surla convergence de méthodes de directions conjuguées // Rev. franc. inform. rech. opér. – 1969. – **16**. – P. 35 – 43.
18. *Polyak B. T.* The conjugate gradient method in extreme problems // USSR Comput. Math. and Math. Phys. – 1969. – **9**. – P. 94 – 112.
19. *Rauf A. I., Fukushima M.* Globally convergent BFGS method for nonsmooth convex optimization // J. Optim. Theory and Appl. – 2000. – **104**. – P. 539 – 558.
20. *Sun W., Yuan Y.* Optimization theory and methods. – New York: Springer Sci. and Business Media, LLC, 2006.
21. *Yabe H., Takano M.* Global convergence properties of nonlinear conjugate gradient methods with modified secant condition // Comput. Optim. and Appl. – 2004. – **28**. – P. 203 – 225.
22. *Zhang L., Zhou W., Li D.* A descent modified Polak – Ribiére – Polyak conjugate gradient method and its global convergence // IMA J. Numer. Anal. – 2006. – **26**. – P. 629 – 640.
23. *Zhang L., Zhou W., Li D.* Global convergence of a modified Fletcher – Reeves conjugate gradient method with Armijo-type line search // Numer. Math. – 2006. – **104**. – P. 561 – 572.
24. *Zhang L., Zhou W., Li D.* Some descent three-term conjugate gradient methods and their global convergence // Optim. Methods and Software. – 2007. – **22**. – P. 697 – 711.
25. *Zhou W.* A short note on the global convergence of the unmodified PRP method // Optim. Lett. – 2013. – **7**. – P. 1367 – 1372.

Received 16.03.12,
after revision — 07.12.14