A PENALTY-CONJUGATE GRADIENT ALGORITHM FOR SOLVING CONSTRAINED OPTIMIZATION PROBLEMS

Hongxiang Zhang, Zhuohua Liang & Wenling Zhao*
School of Mathematics and Statistics, Shandong University of Technology. Zibo, Shandong, China, 255049.
E-mail: 18369906817@163.com, 13287089981@163.com, zwlsdj@163.com.
*Corresponding Author E-mail: zwlsdj@163.com.

ABSTRACT
This paper presents a new algorithm for constrained optimization problems, it is called penalty-conjugate gradient method. The method apply nonlinear conjugate gradient method to penalty function, by choosing a conjugate factor β with good numerical performance and convergence, and selecting the search direction p with the descent property. The algorithm has global convergence under the Wolfe line search.

Key Words: *penalty-conjugate gradient method, global convergence, penalty function..*

1. INTRODUCTION

Constrained optimization exist in many fields, such as optimal control, parameter estimation, Nash equilibrium and sensitivity analysis. In order to solve constrained optimization problems, we usually need to deal with a large number of state variables, control variables or constrained variables. Therefore, it is very necessary to find out efficient numerical solutions quickly. For the process of finding the solutions of constrained optimization problem, we can divide it into two steps. First, the control vector can be parameterized to make the original problem become a nonlinear unconstrained optimization problem. Then the general unconstrained optimization algorithms are used to solve this problem. In these algorithms, the conjugate gradient method is a kind of practical method with good properties, simple structure and less computation. Fletcher and Reeves proposed a conjugate gradient method[1] for solving unconstrained minimization problems and it is used to solve large-scale problems. In the literature [2], AI-Baali proved that the Fletcher-Reeves method had global convergence. Literature [3] generalized the above conclusions to the case of strong Wolfe line search. However, there were always a lot of small steps, so the FR method sometimes performed poorly in numerical calculations. The PRP method proposed by Polark-Ribiere-Polyar and the HS method proposed by Hestence-Stiefel were two kinds of conjugate gradient algorithm with good numerical performance. In [4], Powell pointed out that there may be no global convergence, even the exact line was used to search for the Polark-Ribiere-Polyar method. But this method had good numerical performance(see [5]). Subsequently, many scholars had done a more profound study of the global convergence of this kind of algorithm. The DY method was proposed by Dai Y.H. and Yuan Y.X was studied detailedly. It had global convergence and intrinsic properties. In literature[6], Dai systematically introduced global convergence of the DY- method under general line search. Andrei had considered the good numerical performance of HS method and the good convergence of DY- method. The hybrid DY-HS conjugate gradient method which was proposed in literature[7,8] had good convergence, but it can't guarantee the descent of search direction.

In recent years, a lot of progress has been made in the research of various conjugate gradient methods[4, 9,10,11]. Narushima[12] and Xu Dong[13] put forward two kinds of conjugate gradient methods with sufficient descent, respectively, such that the search direction has a descent property and is not affected by the conjugate gradient factor. Zhao and Yao[14,15]proposed the geometric nonlinear conjugate gradient method and Riemann Fletcher-Reeves conjugate gradient method to solve the stochastic eigenvalue problem in literature, respectively. Liu and Ding propose two new descent conjugate gradient algorithms in the literature [16,17] was used to solve convex constrained monotone equations. On the basis of the existing research, a new penalty-conjugate gradient algorithm is presented in this paper. It combines nonlinear conjugate gradient method with penalty function. The structure of search direction p can ensure every search direction is descent, and the conjugate gradient factor β we choose can ensure the algorithm has a good convergence. And we prove that the new algorithm has global convergence under the Wolfe line search. At the end of the paper, two numerical results have been given.

2. PENALTY-CONJUGATE GRADIENT ALGORITHM

Now, we consider the following inequality constrained optimization problems:

$$\begin{aligned} &\min f(x) \\ &s.t\ \ c_i(x) \le 0,\ \ i = 1, \dots m. \\ &\quad x \in R^n\ . \end{aligned} \tag{2.1}$$

where $f(x)$, $c_i(x)$ are continuous differentiable convex functions. Let

$$\min \phi(x) = f(x) + \frac{\tau}{2}\sum_{i=1}^{m}\left(\max\left\{0, c_i(x)\right\}\right)^2, \tag{2.2}$$

where τ is a penalty factor. Moreover, the global optimal solution of the problem (2.2) is the optimal solution of the problem (2.1) when it satisfies certain conditions. For convenience,let

$$D(x) = \left(\max\left\{0, c_i(x)\right\}\right)^2,$$

$$\nabla\phi(x) = g(x),\ \ s_k = x_{k+1} - x_k,$$

$$y_k = g_{k+1} - g_k,\ g_k @ g(x_k).$$

Definition 2.1 It is said that the quadratic continuous differentiable function $\phi(x): R^n \to R$ is a uniform convex function (a strong convex function). If the existence of $\xi > 0$, for $\forall x, y \in R^n$, the following inequalities are established:

$$(y-x)^T\left(\nabla\phi(y) - \nabla\phi(x)\right) \ge \xi\|x-y\|^2,\ \ \forall x, y \in R^n. \tag{2.3}$$

From the literature [18], it is known that the above formula is equivalent to

$$\phi(x) \ge \phi(y) + g(y)^T(x-y) + \frac{\xi}{2}\|x-y\|,\ \ \forall x, y \in R^n\ . \tag{2.4}$$

Form (2.3), (2.4), we have the conclusion

$$s_k^{\ T} y_k \ge \xi\|s_k\|^2, \tag{2.5}$$

and

$$\phi_k - \phi_{k+1} \ge -g_{k+1}^{\ \ T} s_k + \frac{\xi}{2}\|s_k\|^2\ . \tag{2.6}$$

Next, the penalty-conjugate gradient algorithm is given.

Algorithm 2.2:

Step 1 Select the initial point $x_0 \in R^n$, give the termination error ε, let

$$p_0 = -\frac{g_0}{\|g_0\|},\ k := 0\,.$$

Step 2 Calculate p_k :

$$p_k = -(1+\beta_{k-1}\frac{g_k^T s_{k-1}}{\|g_k\|^2})g_k + \beta_{k-1}s_{k-1}, k \ge 1, \qquad (2.7)$$

where

$$\beta_{k-1} = \frac{y_{k-1}^T g_k}{s_{k-1}^T y_{k-1}} - \frac{s_{k-1}^T g_k}{s_{k-1}^T y_{k-1} + \eta_{k-1}},$$
$$\eta_{k-1} = 2\left(\phi\left(x_{k-1}\right) - \phi\left(x_k\right)\right) + \left(g_k + g_{k-1}\right)^T s_{k-1}. \qquad (2.8)$$

Step 3 Calculate the step size λ_k by Wolfe line search condition

$$\phi\left(x_k + \lambda_k p_k\right) - \phi\left(x_k\right) \le \delta\lambda_k^T g_k{}^T p_k,$$
$$g\left(x_k + \lambda_k p_k\right)^T p_k \ge \sigma g_k{}^T p_k,$$

where $0 < \delta < \sigma < 1.$

Step 4 Let $x_{k+1} = x_k + \lambda_k p_k$, calculate gradient $\left\|g\left(x_{k+1}\right)\right\|$.

If $\left\|g\left(x_{k+1}\right)\right\| \le \varepsilon$, then stop, otherwise, go back to step 2.

3. THE CONVERGENCE ANALYSIS OF THE ALGORITHM

Now, we prove that p_k is the descent direction. According to the property of the descent direction, we only need to verify $p_k^T g_k \le 0$, that is,

$$p_k^T g_k = \left(-(1+\beta_{k-1}\frac{g_k^T s_{k-1}}{\|g_k\|^2})g_k + \beta_{k-1}s_{k-1}\right)^T g_k = -\|g_k\|^2 \le 0, k \ge 1.$$

therefore, p_k is the descent direction.

Lemma 3.1([12]) Suppose x_0 is the initial point, and $x_{k+1} = x_k + \lambda_k p_k$. Where p_k is obtained by the algorithm 2.2, and λ_k satisfies the Wolfe line search condition, then

$$\sum_{k=0}^{\infty} \frac{\left(p_k^T g_k\right)^2}{\|p_k\|^2} < \infty \qquad (3.1)$$

From (2.7), we can get $p_k^T g_k = -\|g_k\|^2$, then (3.1) is equivalent to the following formula

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|p_k\|^2} < \infty$$

.

In order to ensure the good convergence of the algorithm and prove the convergence of the algorithm, we need to assume several conditions.

Assumption 1 The level set $\Omega=\left\{x\in R^n,\phi\left(x\right)\le\phi\left(x_0\right)\right\}$ is bounded, where x_0 is the initial point.

Assumption 2 In a convex neighborhood N of the level set Ω, $\phi\left(x\right)$ is continuously differentiable, the gradient $g\left(x\right)$ satisfies the Lipschitz condition, that is, the existence constant $L>0$ satisfies

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

Assumption 3 Assume that function $\phi\left(x\right):\mathrm{R}^n\to\mathrm{R}$ is a uniformly convex function.

Theorem 3.2 Suppose the objective function satisfies the Assumption 1, 2, and 3, $\left\{x_k\right\}$ is generated by the algorithm 2.2, then we have $\lim_{k\to\infty}\left\|g_k\right\|=0$ or $\liminf_{k\to\infty}\left\|g_k\right\|=0$.

Proof Assume that for any k there is $\left\|g_k\right\|\neq 0$, then according to $\phi\left(x\right)$ is a uniform convex function, we can obtain

$$s_k{}^T y_k=s_k{}^T\left(g_k-g_{k-1}\right)\ge\xi\left\|s_k\right\|^2. \tag{3.3}$$

Because p_k is the descent direction, there is $\left\|p_k\right\|\neq 0$. From (2.6), (2.8) and (3.2), we can obtain

$$\left|\beta_k\right|\le\frac{\left\|y_k\right\|\left\|g_{k+1}\right\|}{\left|s_k^T y_k\right|}+\frac{\left\|s_k\right\|\left\|g_{k+1}\right\|}{2\left|\phi\left(x_k\right)-\phi\left(x_{k+1}\right)+s_k^T g_{k+1}\right|}$$

$$\le\frac{\left\|y_k\right\|\left\|g_{k+1}\right\|}{\left|s_k^T y_k\right|}+\frac{\left\|s_k\right\|\left\|g_{k+1}\right\|}{\xi\left\|s_k\right\|^2}$$

$$\le\frac{\left\|y_k\right\|\left\|g_{k+1}\right\|}{\xi\left\|s_k\right\|}+\frac{\left\|s_k\right\|\left\|g_{k+1}\right\|}{\xi\left\|s_k\right\|^2}$$

$$\le\frac{L\left\|s_k\right\|\left\|g_{k+1}\right\|}{\xi\left\|s_k\right\|^2}+\frac{\left\|s_k\right\|\left\|g_{k+1}\right\|}{\xi\left\|s_k\right\|^2}$$

$$\le\left(\frac{L+1}{\xi}\right)\frac{\left\|g_{k+1}\right\|}{\left\|s_k\right\|}$$

that is,

$$\left|\beta_k\right|\le\left(\frac{L+1}{\xi}\right)\frac{\left\|g_{k+1}\right\|}{\left\|s_k\right\|}$$

Therefore, form (2.7), we can obtain

$$\|p_{k+1}\| \le \|g_{k+1}\| + |\beta_k| \frac{\|g_{k+1}\|\|s_k\|}{\|g_{k+1}\|^2}\|g_{k+1}\| + \|\beta_k\|\|s_k\|$$

$$\le \|g_{k+1}\| + \left(\frac{L+1}{\xi}\right)\|g_{k+1}\| + \left(\frac{L+1}{\xi}\right)\|g_{k+1}\|$$

$$\le \left(1 + 2\left(\frac{L+1}{\xi}\right)\right)\|g_{k+1}\|.$$

Let $\sqrt{\theta} = 1 + 2\left(\frac{L+1}{\xi}\right)$, then the above formula is equivalent to

$$\|p_{k+1}\|^2 \le \theta \|g_{k+1}\|^2,$$

Therefore, form Lemma 3.1, we can obtain

$$\sum_{k=0}^{\infty} \|g_{k+1}\|^2 \le \theta \sum_{k=0}^{\infty} \frac{\|g_{k+1}\|^4}{\|p_{k+1}\|^2} < \infty,$$

So, $\liminf_{k\to\infty} \|g_k\| = 0$.

Next, we prove that the global optimal solution of function $\phi(x)$ is the global optimal solution of objective function $f(x)$.

Theorem3.3 x^* is the global optimal solution of the constrained problem (2.1). The penalty factor $\tau \to +\infty$, if x_k is a sequence of solutions of function $\phi(x)$, then any cluster point $\tilde{x}$ of $\{x_k\}$ must be the global optimal solution of the constrained problem.

Proof Suppose $\tilde{x}$ be a cluster point of $\{x_k\}$, then there is the convergence subsequence converges to $\tilde{x}$. And we may suppose that the convergence sequence is $\{x_k\}$. According to $\tilde{x}$ is the global optimal solution of function $\phi(x)$, we can obtain

$$\phi(\tilde{x}) \le \phi(x^*) = f(x^*) + \frac{\tau}{2} D(x^*). \tag{3.4}$$

Since x^* is a feasible point for the constrained problem, therefore

$$D(x^*) = 0,$$

and

$$\phi(\tilde{x}) \le f(x^*), \tag{3.5}$$

that is

$$f(\tilde{x}) + \frac{\tau}{2} D(\tilde{x}) \le f(x^*). \tag{3.6}$$

When $\tau \to +\infty$, (3.6) means $D(\tilde{x}) = 0$, so $\tilde{x}$ is a feasible point .

Next, let's prove that $\tilde{x}$ is the global optimal solution. According to (3.6) and $D(x_k) \ge 0$, we can obtain $f(\tilde{x}) \le f(x^*)$ when $\tau \to \infty$. On the other hand, because $\tilde{x}$ is feasible solution, and x^* is global optimal solution, then there are $f(x^*) \le f(\tilde{x})$, therefore, we can obtain $f(\tilde{x}) = f(x^*)$.

4. NUMERICAL RESULTS

Next we give two examples of the above algorithm

Example 1

$$\min f(x) = x_1^2 + x_2^2 - 2x_1x_2 + 4$$

$$s.t\ x_1 + x_2 \le 0. \tag{4.1}$$

According to (2.2), there is

$$\phi(x) = x_1^2 + x_2^2 - 2x_1x_2 + 4 + \frac{\tau}{2}\max\{0, (x_1 + x_2)\}^2 \tag{4.2}$$

$$g(x) = \begin{bmatrix} 2x_1 - 2x_2 + \{0, \tau(x_1 + x_2)\} \\ 2x_2 - 2x_1 + \{0, \tau(x_1 + x_2)\} \end{bmatrix}$$

We take the initial point $x^{(0)}$ as (-1, 1), The parameters are as follows:

$$\delta = 0.5,\ \sigma = 0.75,\ \tau = 10^6,\ \varepsilon = 10^{-4}.$$

we can obtain $g\left(x^{(0)}\right) \approx \begin{bmatrix} -4 \\ 4 \end{bmatrix}$. Because $\left\| g\left(x^{(0)}\right) \right\| > 10^{-4}$, so there is $p\left(x^{(0)}\right) = \begin{bmatrix} 4 \\ -4 \end{bmatrix}$. By the Wolfe step rule and the above parameters, we can make $\lambda_0 = 0.25$, finally get $x^{(1)} = (0,0)$. Because $\left\| g\left(x^{(1)}\right) \right\| = 0$, so we get the optimal solution (0,0) of (4.2). It's clear that (0,0) is also the optimal solution for (4.1).

Example 2

$$\min f(x) = \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 2x_1$$

$$s.t\ x_1 + x_2 - 10 \le 0$$

We take the initial point $x^{(0)}$ as (-2, 4), The parameters $\delta = 0.3,\ \sigma = 0.5,\ \tau = 10^6,\ \varepsilon = 10^{-4}$. We use MATLAB2014a to solve the above function, the result is as follows:

Table1

Solution sequence	Iterative direction	Gradient	Step size
(-2,4)	(12,-6)	(0.894,-0.447)	0.294
(1.529,2.235)	(-0.311,-0.727)	(0.353,0.706)	1.7
(1,1)	(0,0)	(0,0)	(0,0)

5. ACKNOWLEDGEMENTS
This research was supported by National Natural Science Foundations of China (No. 11771255) and Natural Science Foundation of Shandong Province (ZR2016AM07, ZR2015AL011).

5. REFERENCES

[1]. Fletcher R., Reeves. Function minimization by conjugate gradient[J]. *Computer Journal*, 1964, 7(2): 149-154.
[2]. Al-Baali M.. Descent property and global convergence of the Fletcher-Reeves method with inexact line searches[J]. *Journal of Numerical Analysis*, 1985,5(1): 121-124.
[3]. Byrd R H, Nocedal J, Yuan Y X. Global convergence of a class of Quasi-Newton methods on convex problems[J]. *Siam Journal on Numerical Analysis*, 1987, 24(5):1171-1190.
[4]. Kou C.X.. An improved nonlinear conjugate gradient method with an optimal property[J]. *Science China Mathematics*, 2014, 57(3): 635-648.
[5]. Jorge N., Stephen J.. Wright. Numerical optimization[M]. *Springer*, 2006.
[6]. Dai Y.. Yuan Y.. Nonlinear conjugate gradient method[M]. Science and Technology Press of Shanghai, 2000.
[7]. Andrei N.. Another hybrid conjugate gradient algorithm for unconstrained Optimization[J]. *Numerical Algorithms*, 2008, 47: 143-156.
[8]. Andrei N.. Accelerated hybrid conjugate gradient algorithm with modified secant condition for unconstrained optimization[J]. *Numerical Algorithms*, 2010, 54(1): 23-46.
[9]. Yao S., He D., Shi L.. An improved perry conjugate gradient method with adaptive parameter choice[J]. *Numerical Algorithms*, 2017(3): 1-15.
[10]. Babaie-Kafaki S.. A note on the global convergence of the quadratic hybridization of Polak–Ribière–Polyak and Fletcher–Reeves conjugate gradient methods[J]. *Journal of Optimization Theory & Applications*, 2013, 157(1): 297-298.
[11]. Al-Bayati A.Y.. Al-Kawaz R.Z.. A new hybrid WC-FR conjugate gradient-algorithm with modified secant condition for unconstrained optimization[J]. *J.Math. Comput. Sci*, 2012, 2(4): 937-966.
[12]. Narushima Y.. Yabe H.. Ford J. A. A.. three-term conjugate gradient method with sufficient descent property for unconstrained optimization[J]. *Computational Optimization & Applications*, 2014, 60(1): 89-110.
[13]. Dong X., He Y.. A new class of conjugate gradient methods with sufficient descent condition and strong convergence[J]. *Journal of Mathematics*, 2017, 37(2): 231-238.
[14]. Zhao Z., Jin X.Q., Bai Z.J.. A geometric nonlinear conjugate gradient method for stochastic inverse eigenvalue problems[J]. *Siam Journal on Numerical Analysis*, 2016, 54(4): 2015-2035.
[15]. Yao T.T., Bai Z.J., Zhao Z., et al. A Riemannian Fletcher--Reeves conjugate gradient method for doubly stochastic inverse eigenvalue problems[J]. 2016, 37(1): 215-234.
[16]. Liu S.Y., Huang Y.Y., Jiao H.W.. Sufficient descent conjugate gradient methods for solving convex constrained nonlinear monotone equations[J]. *Abstract & Applied Analysis*, 2014, 2014(1): 1-12.
[17]. Ding Y., Xiao Y., Li J.. A class of conjugate gradient methods for convex constrained monotone equations[J]. *Optimization*, 2017(2): 1-20.
[18]. Wang Y.J., Xiu N H. Nonlinear optimization theory and algorithm[M]. Science Publishing Company, China, 2012.