AMO-Advanced Modeling and Optimization, Volume 12, Number 2, 2010

A Lagrangian Dual Spectral Projected Gradient Method for Nonconvex Constrained Optimization [1]

Jinhong Yu[a], Zhensheng Yu[b]

a. College of Foreign Language, University of Shanghai for Science and Technology, Shanghai, 200093, P.R.China

b. College of Science, University of Shanghai for Science and Technology, Shanghai, 200093, P.R.China

Abstract. Motivated by the dual problem and the spectral projected gradient method, we develop a Lagrangian dual spectral projected gradient method for nonconvex constrained nonlinear programming problem. The equivalent between the KKT point of the original problem and the projected gradient of the dual problem is considered. At each iteration, we only need to make a projected computation on a nonnegative constraint. Under certain conditions, the global convergence is obtained and numerical tests are also given to show the efficiency of the proposed method.

Keywords: Nonconvex constrained optimization, Dual problem, Spectral projected gradient, Nonmonotone line search

MSC: 65K05; 90C30

1. Introduction

The spectral gradient method was originally proposed by Barzilai and Borwein [4] and further analyzed by Raydan [18] for quadratic function. Since the method requires little computational work and greatly speeds up the convergence of gradient methods, hence it had attracted many researchers' attention. In 1997, Raydan [19] extended the spectral gradient method to unconstrained optimization and proved its global convergence.

In 2000, by combining the projected gradient method and the nonmonotone line search technique, Birgin, Martínez and Raydan [6] extended the spectral gradient method to convex constrained optimization. Since then, the method has been intensively used to many kinds of convex constrained optimization include bound constrained optimization, linear constrained optimization [1, 2, 3, 5, 7, 11, 20, 23, 25]. It has also been extended for solving non-differentiable convex constrained problem by Crema, Loreto and Raydan [9] although no convergence property was discussed. More recently, by combining the augmented lagrangian, Ehrhardt, Ruggiero, Martínez and Santos [13] extended the spectral projected gradient method to partly convex constrained optimization in the form:

$$\min f(x) \ s.t. \ h(x) = 0, x \in \Omega,$$

[1] This work was supported by Innovation Program of Shanghai Municipal Education Commission(No.10YZ99) and Shanghai Leading Discipline Project (No. S30501)

***AMO-Advanced Modeling and Optimization. ISSN: 1841-4311**

where $f(x) : R^n \to R$ and $h : R^n \to R^m,\ h(x) = (h_1(x), h_2(x), \cdots, h_m(x))^T$ have continuous first derivative and the set Ω is convex. More detailed applications of the spectral gradient method can be found in the survey paper [8] and references therein.

The aim of this paper is to extend the spectral gradient method to nonconvex constrained optimization. For convenience, we only consider the problem with inequality constraint in the form:

$$\min f(x)\ s.t.\ h(x) \leq 0. \tag{1}$$

Motivated by the work of Han and Mangasarian [16](see also[24]), we consider the dual exact differentiable penalty function(a nonnegative constrained maximum problem) and develop a spectral projected gradient method for the dual problem. The equivalent between the KKT point of the original problem and the projected gradient of the dual problem is considered. At each iteration, we only need to make a projected computation on a nonnegative constraints.

As mentioned in [6, 8, 12, 14, 23, 25], the spectral(projected) gradient direction may not be a descent direction, hence one often employ the nonmonotone line search to obtain the iteration sequence. Moreover, the nonmonotone schemes can improve convergence, in particular in the presence of a narrow curve valley, and encouraging results have been reported with nonmonotone based algorithms in [10, 15, 21, 22]. In a recent work, Zhang and Hager [26] have proposed a new nonmonotone scheme which have been shown to eliminate some of the inherent drawback of the traditional nonmonotone schemes and also computationally more economic. Based on this reason, in this paper, we consider combining the idea of by Zhang and Hager [26] with the spectral projected gradient method for solving the constrained optimization (1). Throughout this paper, we use $\langle \cdot, \cdot \rangle$ to denote the inner product of two vectors.

The paper is organized as follows: In Section 2, we consider the dual problem of problem (1) and establish the equivalent relationship between the KKT point of problem (1) and the projected gradient of the dual problem. In Section 3, we describe the algorithm and analyze its global convergence. The numerical tests are given in Section 4 and we conclude the paper in Section 5.

2. Dual problem and Projected gradient

In this section, we consider the dual problem of (1) and establish the equivalent relationship between the KKT point of problem (1) and the projected gradient of the dual problem. Throughout this paper, we assume that the objective function $f(x)$ and the constraint functions $h_i(x)$ are twice continuously differentiable and the feasible region $\mathcal{F} = \{\S \in \mathcal{R}^\backslash | \langle_\rangle(\S) \leq \prime,\ \rangle = \infty, \in, \cdots, \Updownarrow\} \neq \emptyset$.

Denote the lagrangian function of (1) by

$$L(x, u) = f(x) + \langle u, h(x) \rangle, \tag{2}$$

where $u \in R^m$ is the lagrangian multiplier vector. By Wolfe dual theory, the dual problem of (1) can be written as [24]:

$$\min_{x,u} F(x, u, \sigma) = -L(x, u) + \frac{1}{2}\sigma \|\nabla_x L(x, u)\|^2,\ s.t.\ u \geq 0. \tag{3}$$

where $\sigma > 0$ is a penalty parameter, $\|\cdot\|$ is the Euclidean norm and $\nabla_x L(x, u)$ denotes the gradient of $L(x, u)$ with respect to x, i.e.,

$$\nabla_x L(x, u) = \nabla f(x) + \nabla h(x) u.$$

where $\nabla h(x) \in R^{n \times m}$ is the Jacobian matrix of $h(x)$.

To study the relationship between problem (1) and (3), we first introduce the definition of the projection. Denote $\Omega = R^n \times R^m_+$, for any $z = (x^T, u^T)^T \in \Omega$, we define $P_\Omega(z)$ as the orthogonal projection on Ω. i.e. $P_\Omega(z) = (x^T, u_+^T)^T$, where $u_+ = (\max(0, u_1), \max(0, u_2), \cdots, \max(0, u_m))^T$. For convenience, we write $(x, u) = (x^T, u^T)^T$ for any $x \in R^n$ and $u \in R^m$.

The following lemma gives the Lipschitz continuity property of the projection mapping:

Lemma 1. For any $z_1,\ z_2 \in \Omega$, we have

$$\|P_\Omega(z_1) - P_\Omega(z_2)\| \le \|z_1 - z_2\|.$$

Suppose that at the kth iteration, (x_k, u_k) is a KKT point of problem (1), i.e., (x_k, u_k) satisfies:

$$\begin{cases} \nabla_x L(x_k, u_k) = \nabla f(x_k) + \nabla h(x_k) u_k = 0, \\ h(x_k) \le 0.\ u_k \ge 0,\ \langle u_k, h(x_k) \rangle = 0. \end{cases}$$

we compute the following vectors:

$$y_k = \nabla_x F(x_k, u_k, \sigma_k) = -(I - \sigma_k \nabla^2_{xx} L(x_k, u_k)) \nabla_x L(x_k, u_k) \tag{4}$$

$$w_k = \nabla_u F(x_k, \mu_k, \sigma_k) = -h(x) + \sigma_k \nabla h(x)^T \nabla_x L(x_k, u_k) \tag{5}$$

Furthermore, we denote $F(z_k, \sigma_k) = F(x_k, u_k, \sigma_k)$, $F_k = F(z_k, \sigma_k)$ and the gradient of $F(x, u, \sigma)$ by $g(z_k) = (y_k, w_k)$.

The following Theorem gives the relationship between the projected gradient and the KKT point of problem (1).

Theorem 1. If $\sigma_k < 1/\|\nabla^2_{xx} L(z_k)\|$, then $P_\Omega(z_k - g_k) - z_k = 0$ implies that z_k is a KKT point of problem (1).

Proof. By (4)(5) and the definition of the projection, $P_\Omega(z_k - g_k) - z_k = 0$ means $y_k = 0$ and $(u_k - w_k)_+ - u_k = 0$. Since $\sigma_k < 1/\|\nabla^2_{xx} L(z_k)\|$, so $y_k = 0$ implies

$$\nabla_x L(x_k, u_k) = \nabla f(x_k) + \nabla h(x_k) u_k = 0. \tag{6}$$

Hence (5) means $w_k = -h(x_k)$ and therefore we have

$$(u_k + h(x_k))_+ - u_k = 0. \tag{7}$$

By the definition of projection, if $u_k^i + h_i(x_k) \ge 0$, then (7) means $h_i(x_k) = 0$ and $u_k^i \ge 0$. Otherwise, if $u_k^i + h_i(x_k) < 0$, then (7) means $u_k^i = 0$ and $h_i(x_k) < 0$. Hence (7) implies that

$$h(x_k) \le 0,\ u_k \ge 0 \text{ and } \langle u_k, h(x_k) \rangle = 0. \tag{8}$$

By (6) and (8), we obtain the desired result. □

3. Algorithm and Global Convergence

In this section, we first introduce the spectral gradient method [19] for unconstrained minimization problem:

$$\min f(x), \quad x \in \mathbb{R}^{\ltimes}$$

where $f : \mathbb{R}^{\ltimes} \to \mathbb{R}$ is continuously differentiable and its gradient $\nabla f(x)$ is available. Spectral gradient method is defined by

$$x_{k+1} = x_k - \lambda_k \nabla f(x_k),$$

where the scalar λ_k is given by

$$\lambda_k = \frac{\langle s_{k-1}, s_{k-1}\rangle}{\langle s_{k-1}, \delta_{k-1}\rangle},$$

where $s_{k-1} = x_k - x_{k-1}, \delta_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$.

In what follows, we describe our nonmonotone dual lagrangian spectral projected gradient algorithm detailed:

Algorithm 1

Step 0. Given $z_0 = (x_0, u_0) \in R^n \times R^m_+$, choose $0 < \alpha_{min} < \alpha_{max}$, $0 < \eta_{min} < \eta_{max} < 1$, $\gamma \in (0, 1)$, $\sigma_0 > 0$, set $C_0 = F(z_0, \sigma_0)$, $Q_0 = 1$, $\alpha_0 \in [\alpha_{min}, \alpha_{max}]$, $k := 0$.

Step 1. Compute y_k and w_k by (4) and (5), if $\|y_k\| \geq \frac{1}{2}\|\nabla_x L(z_k)\|$ go to Step 2, otherwise, set $\sigma_k := \frac{1}{2}\sigma_k$, repeat Step 1.

Step 2. If $\|P(z_k - g_k) - z_k\| = 0$, then stop.

Step 3. Compute $d_k = P(z_k - \alpha_k g_k) - z_k$, set $\beta_k = 1$.

Step 4. Set $z_+ = z_k + \beta_k d_k$.

Step 5. If

$$F(z_+, \sigma_k) \leq C_k + \gamma\beta_k\langle d_k, g_k\rangle, \tag{9}$$

then define $z_{k+1} = z_+$, $\sigma_{k+1} = \sigma_k$, $s_k = z_{k+1} - z_k$, $\delta_k = g_{k+1} - g_k$, and go to Step 6.

If (9) does not hold, define $\beta_{new} \in [0.1\beta_k, 0.9\beta_k]$, set $\beta_k = \beta_{new}$, and go to Step 4.

Step 6. Compute $b_k = \langle s_k, \delta_k\rangle$, If $b_k \leq 0$, set $\alpha_{k+1} = \alpha_{max}$, else, compute $a_k = \langle s_k, s_k\rangle$ and

$$\alpha_{k+1} = \min\{\alpha_{max}, \max\{\alpha_{min}, a_k/b_k\}\}.$$

Choose $\eta_k \in [\eta_{min}, \eta_{max}]$ and set

$$Q_{k+1} = \eta_k Q_k + 1,\ C_{k+1} = (\eta_k Q_k C_k + F(z_{k+1}, \sigma_{k+1}))/Q_{k+1},$$

$k := k + 1$ go to Step 1.

In what follows, we analyze the global convergence properties of Algorithm 1. To this end, we introduce some basic definitions and lemmas.

Define the scaled projected gradient $g_t(z)$ as

$$g_t(z) = P[z - tg(z)] - z$$

for all $z \in \Omega$ and $t > 0$. The following lemma gives the properties about the scaled projected gradient $g_t(z)$, which can be found in [6].

Lemma 2. For all $z \in \Omega,\ t \in (0, \alpha_{max}]$,

(i)

$$\langle g(z), g_t(z)\rangle \leq -\frac{1}{t}\|g_t(z)\|_2^2 \leq -\frac{1}{\alpha_{max}}\|g_t(z)\|_2^2,$$

(ii) The vector $g_t(z^\star)$ vanishes if $z^\star$ is a stationary point of problem (3).

The Lemma 2 (ii) shows that if $d_k = 0$ at kth iteration, then z_k is a stationary point of problem (3). To prove the global convergence of Algorithm 1, we make the following assumptions:

Assumption H:

(1) The objective function $f(x)$ and the constrained functions $h_i(x)$ are twice continuously differentiable for all $x \in R^n$.

(2) The sequence generated by algorithm is contained in an open convex set S and $F(z)$ is bounded below on S.

(3) The Lagrangian gradient function $g(z)$ is continuity uniformly on S.

The following lemma shows our algorithm is well defined.

Lemma 3. Let $z_k,\ d_k$ be generated by Algorithm 1, if $d_k \neq 0$, then we have $F_k < C_k$, moreover, the algorithm is well defined.

Proof. Since $d_k \neq 0$, by Lemma 2 and the definition of d_k, we have $\langle g_k, d_k\rangle \leq -\frac{1}{\alpha_{max}}\|d_k\|^2 < 0$. Defining $D_k : R \to R$ by

$$D_k(\zeta) = \frac{\zeta C_{k-1} + F_k}{\zeta + 1},$$

we have

$$D_k'(\zeta) = \frac{C_{k-1} - F_k}{(\zeta+1)^2}.$$

Since $\langle g_k, d_k\rangle < 0$, it follows from (9) that $F_k < C_{k-1}$, which implies $D_k'(t) < 0$ for all $\zeta \geq 0$. Hence $D_k(\zeta)$ is nondecreasing and therefore $F_k = D_k(0) < D_k(\zeta)$ for all $\zeta \geq 0$. Taking $\zeta = \eta_{k-1}Q_{k-1}$, we obtain

$$F_k = D_k(0) < D_k(\eta_{k-1}Q_{k-1}) = C_k. \tag{10}$$

In what follows, we want to prove that at the kth iteration, the inner cycle Step 4-Step 5-Step 4 can be terminated after reduce the values of β many times. If this is not true, then we have $\beta_k \to 0$ and there exists a positive constant $\sigma \in [0.1, 0.9]$ such that for $\beta := \beta_k/\sigma$, (9) does not hold, i.e.,

$$F(z_k + \beta_k d_k/\sigma, \sigma_k) > C_k + \gamma\beta_k/\sigma\langle d_k, g(z_k)\rangle > F_k + \gamma\beta_k/\sigma\langle d_k, g(z_k)\rangle.$$

which implies that

$$\frac{F(z_k + \beta_k d_k/\sigma, \sigma_k) - F_k}{\beta_k/\sigma} > \gamma\langle d_k, g(z_k)\rangle.$$

Since $\beta_k \to 0$, we have

$$\langle d_k, g(z_k)\rangle > \gamma\langle d_k, g(z_k)\rangle.$$

And therefore, we get

$$(1-\gamma)\langle d_k, g(z_k)\rangle > 0.$$

This contradicts to the fact that $\langle d_k, g(z_k)\rangle < 0$, the contradiction shows the algorithm is well defined. □

Lemma 4. Under assumptions (1)-(3), for each k, the repetition $\sigma_k = \frac{1}{2}\sigma_k$ in Step 1 of Algorithm 1 terminates finitely.

Proof. Similar to Lemma 3.4 in [22]. □

By Lemma 4, without loss of generality, we assume $\sigma_k = \sigma_0$ for all k and write $F(x_k, u_k, \sigma_k) = F(x_k, u_k)$.

In what follows, we prove our first global convergence result.

Theorem 2. Assume Assumptions (1)-(3) hold, then we have

$$\liminf_{k\to\infty} \|d_k\| = 0. \tag{11}$$

Proof. By contradiction, if (11) does not hold, then for all k, there exists a positive constant ε such that

$$\|d_k\| \geq \varepsilon. \tag{12}$$

Denote $\delta = \gamma/\alpha_{max}$, by the line search (9) and Lemma 2 we have

$$\begin{aligned} F_{k+1} &\leq C_k + \gamma\beta_k g_k^T d_k \\ &\leq C_k - \gamma\beta_k\|d_k\|^2/\alpha_{max} \\ &= C_k - \delta\beta_k\|d_k\|^2. \end{aligned} \tag{13}$$

Since $Q_{k+1} = \eta_k Q_k + 1$, we have

$$\begin{aligned} C_{k+1} &= (\eta_k Q_k C_k + F_{k+1})/Q_{k+1} \\ &\leq (\eta_k Q_k C_k + C_k - \delta\beta_k\|d_k\|^2)/Q_{k+1} \\ &= C_k - \delta\beta_k\|d_k\|^2/Q_{k+1} \end{aligned} \tag{14}$$

Since $F(z)$ is bounded below and $F_k \leq C_k$, we have C_k is bounded below. It follows from (14) that

$$\sum_{i=1}^{k} \frac{\delta\beta_k\|d_k\|^2}{Q_{k+1}} < C_k - C_{k+1} \tag{15}$$

Since

$$Q_{k+1} = 1 + \sum_{j=0}^{k}\prod_{i=0}^{j}\eta_{k-i} \leq 1 + \sum_{j=1}^{k}\eta_{max}^{j+1} < \sum_{j=0}^{\infty}\eta_{max}^{j} < \frac{1}{1-\eta_{max}},$$

let $k \to \infty$, (15) implies that

$$\lim_{k\to\infty} \beta_k\|d_k\|^2 = 0.$$

Since $\|d_k\| \geq \varepsilon$ for all k, we therefore have

$$\lim_{k\to\infty} \beta_k = 0 \text{ and } \lim_{k\to\infty} \beta_k\|d_k\| = 0.$$

which implies for $\beta_k = \beta_k/\sigma$, the line search (9) does not satisfied, i.e.,

$$F(z_k + \frac{\beta_k}{\sigma}d_k) > C_k + \gamma\frac{\beta_k}{\sigma}\langle g_k, d_k\rangle \geq F(z_k) + \gamma\frac{\beta_k}{\sigma}\langle g_k, d_k\rangle$$

It follows from mean value theorem that

$$\langle d_k, g_k - g(z_k + \theta_k\frac{\beta_k}{\sigma}d_k)\rangle < (1-\gamma)\langle g_k, d_k\rangle,$$

where $\theta_k \in (0, 1)$.

By the above inequality, Lemma 2 and (12), we obtain

$$\begin{aligned} 1-\gamma &< \frac{\langle d_k, g_k - g(z_k + \theta_k\beta_k d_k/\sigma\rangle}{\langle g_k, d_k\rangle} \\ &\leq \frac{\alpha_{max}\|d_k\|\|g_k - g(z_k + \theta_k\beta_k d_k/\sigma)\|}{\|d_k\|^2} \\ &\leq \frac{\alpha_{max}\|g_k - g(z_k + \theta_k\beta_k d_k/\sigma)\|}{\varepsilon} \end{aligned}$$

Since $\beta_k d_k \to 0$ and $g(z)$ is continuous uniformly, we have $1-\gamma \to 0$, which contradicts the fact that $\gamma < 1$. Hence (11) hold. □

4. Numerical tests

In this section, we give the test result of our algorithm on some typical test problems, these problems are taken from [17, 20]:

Problem 1. QQR-T1-2 [17].

$$f(x) = (x_1 - 5)^2 + x_2^2 - 25,$$

$$h_1(x) = x_1^2 - x_2,$$

$$x_0 = [4.9, 0.1]^T, \quad u_0 = 1, \quad x^* = [(a - \frac{1}{a})/\sqrt{6}, \quad (a^2 - 2 + a^{-2})/6)]^T,$$

where

$$a = 7.5\sqrt{6} + \sqrt{338.5}, \quad f(x^*) = -8.498464223.$$

Problem 2. QQR-T1-3 [17].

$$f(x) = 0.5x_1^2 + x_2^2 - x_1x_2 - 7x_1 - 7x_2,$$

$$h_1(x) = -25 + 4x_1^2 + x_2^2,$$

$$x_0 = [0, 0]^T, \quad u_0 = 1, \quad x^* = [2, 3], \quad f(x^*) = -30.$$

Problem 3. QQR-T1-6 [17].

$$f(x) = (x_1 - 2)^2 + (x_2 - 1)^2,$$

$$h_1(x) = x_1 + x_2 - 2,$$

$$h_2(x) = x_1^2 - x_2,$$

$$x_0 = [2, 2]^T, \quad u_0 = [1, 1]^T, \quad x^* = [1, 1], \quad f(x^*) = 1.$$

Problem 4. QQR-T1-11 [17].

$$f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4,$$

$$h_1(x) = -8 + x_1^2 + x_2^2 + x_3^2 + x_4^4 + x_1 - x_2 + x_3 - x_4,$$

$$h_2(x) = -10 + x_1^2 + 2x_2^2 + x_3^2 + 2x_4^4 - x_1 - x_4,$$

$$h_3(x) = -5 + 2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4,$$

$$x_0 = [0, 0, 0, 0]^T, \quad u_0 = [1, 1, 1]^T, \quad x^* = [0, 1, 2, -1], \quad f(x^*) = -44.$$

Problem 5. QQR-T1-3 [20].

$$f(x) = (x_1 - 2)^2 + (x_2 - 1)^2,$$

$$h_1(x) = x_1^2 - x_2,$$

$$h_2(x) = -x_1 + x_2^2,$$

$$x_0 = [0.5, 0.5]^T, \quad u_0 = [1, 1]^T, \quad x^* = [1, 1], \quad f(x^*) = 1.$$

For the numerical experiments we set following initial parameters: $\alpha_{min} = 10^{-30}$, $\alpha_{max} = 10^{30}$, $\eta_{min} = \eta_{max} = 0.85$, $\gamma = 10^{-4}$, $\sigma_0 = 1$, $\alpha_0 = 1$. To decide when to stop the execution of the algorithms declaring convergence we used the criterion $\|z_k - P(z_k - g_k)\| \leq 10^{-5}$.

The numerical results are shown in Table 1, where the abbreviations in the table are the following ones:

No: Number of the test problem.

ET: Execution time in seconds.

NF: Number of objective function evaluations.

NG: Number of restriction function evaluations(each restriction counted).

NDF: Number of gradient evaluations of the objective function.

NDG: Number of gradient evaluations of the constraints
(each restriction counted).

TABLE 1. Results for Algorithm 1

No	ET	NF	NG	NDF	NDG
1	0.040942	88	78	39	165
2	0.058078	234	114	57	347
3	0.043604	71	132	33	272
4	0.249295	774	1830	305	4149
5	0.049099	122	232	58	474

5. Conclusion

In this paper, we extend the spectral projected gradient method the the non-convex constrained optimization and obtain the global convergence by using a nonmonotone line search technique. From the numerical tests, we can see our algorithm is effectiveness. Whether the algorithm is effectiveness for the large scale test problems deserves further studying.

References

[1] *R. Andreani, E. G. Birgin, J. M. Martínez and J. Yuan:* Spectral projected gradient and variable metric methods for optimization with linear inequalities, IMA J. Numer. Anal. 25(2005), 221-252.

[2] *M. Andretta, E. G. Birgin and J. M. Martínez:* Practical active-set Euclidian trust-region method with spectral projected gradients for bound-constrained minimization, Optimization. 54(2005), 305-325.

[3] *L. Bello and M. Raydan:* Preconditioned spectral projected gradient method on convex sets, J. Comput. Math. 23(2005), 225-232.

[4] *J. Barzilai and J. M.Borwein:* Two point step size gradient method, IMA J. Numer. Anal. (8)(1988), 174-184.

[5] *E. G. Birgin and J. M. Martínez:* Large scale active-set box-constrained optimization method with spectral projected gradients, Comput. Optim. Appl. 23(2002),
101-125.

[6]*E. G. Birgin , J. M. Martínez and Raydan:* Nonmonotone spectral projected gradient methods on convex sets, SIAM J. Optim. 10(2000), 1196-1211.

[7]*E. G. Birgin , J. M. Martínez and Raydan:* Inexact spectral projected gradient methods on convex sets, IMA J. Numer. Anal. 23(2003), 539C559.

[8]*E. G. Birgin, J. M. Martínez and M. Raydan:* Spectral Projected gradient methods, in: C.A. Floudas, P.M. Pardalos (Eds.), Encyclopedia of Optimization, second ed., Springer, 2009, pp. 3652-3659.

[9] *A. Crema, M. Loreto and M. Raydan:* Spectral projected subgradient with a momentum term for the Lagrangian dual approach, Comput. Oper. Res. 34(2007), 3174-3186.

[10] *Y. Dai:* On the non-monotone line search, J. Optim. Theory Appl. 112(2001), 315-330.

[11] *Y. H. Dai and R. Fletcher:* Projected Barzilai-Borwein methods for large-scale box- constrained quadratic programming, Numerische Mathematik. 100(2005), 21-47.

[12] *Y. H. Dai and H. C. Zhang:* Adaptive two-point stepsize gradient algorithm, Numer. Algorithms. 27(2001), 377-385.

[13] *M.A. Diniz-Ehrhardt, M.A. Gomes-Ruggiero, J.M. Martínez and S.A. Santos:* Augmented Lagrangian algorithm based on the spectral projected gradient method for solving nonlinear programming problems, J. Optim. Theory Appl. 123(2004), 497-517.

[14] *R. Fletcher:* On the Barzilai-Borwein method, Department of Mathematics, University of Dundee NA/207, 2001, Dundee, Scotland.

[15] *L. Grippo, F. Lampariello and S. Lucidi:* A Nonmonotone Line Search Technique for Newton's Methods, SIAM J. Numer. Anal. 23(1986), 707-716.

[16] *S.P. Han and O.L. Mangsarian:* A dual differentiable exact penalty function, Math. Program. 25(1983), 293-306.

[17] *W. Hock and K. Schittkowski:* Test Examples for Nonlinear Programming Codes, Berlin: Springer-Verlag,1981.

[18] *M. Raydan:* On the Barzilai and Borwein choice of steplength for the gradient method, IMA J. Numer. Anal. 13(1993), 321-326.

[19] *M. Raydan:* The Barzilai-Borwein gradient method for the large scale unconstrained optimization problem, SIAM J. Optim. 7(1997), 26-33.

[20] *K. Schittkowski:* More test examples for nonlinear programming codes, Berlin: Springer-Verlag,1987.

[21] *Z.J. Shi, J. Shen:* Convergence of nonmonotone line search method, J. Comput. Appl. Math. 193(2006), 397-412.

[22] *W.Y. Sun, J.Y.Han and J.Sun:* Global convergence of non-monotone descent methods for unconstrained optimization problems, J. Comput. Appl. Math. 146(2002), 89-98.

[23] *C. Y. Wang, Q. Liu and X. M. Yang:* Convergence properties of nonmonotone spectral projected gradient methods, J. Comput. Appl. Math. 182(2006), 51-66.

[24] *H.X. Yin, J.Y.Han and Z.W.Chen"* Global convergence of a trust region algorithm for inequality constrained optimization problems, Numer. Funct. Anal. Optim. 25(2004), 571-592.

[25] *Z.S. Yu:* Solving bounded constrained optimization via a new nonmonotone spectral projected gradient method, Appl. Numer. Mathematics, 58(2008), 1340-1348.

[26] *H.C. Zhang and W.W. Hager:* A nonmonotone line search technique and its applications to unconstrained optimization, SIAM J. Optim. 14(2004), 1043-1056.