

A MODIFIED BFGS BUNDLE ALGORITHM BASED ON APPROXIMATE SUBGRADIENTS

QIANG GUO AND JIAN-GUO LIU*

ABSTRACT. In this paper, an implementable BFGS bundle algorithm for solving a nonsmooth convex optimization problem is presented. The typical method minimizes an approximate Moreau-Yosida regularization using a BFGS algorithm with inexact function and the approximate gradient values which are generated by a finite inner bundle algorithm. The approximate subgradient of the objective function is used in the algorithm, which can make the algorithm easier to implement. The convergence property of the algorithm is proved under some additional assumptions.

AMS Mathematics Subject Classification : 26B25, 90C25, 65K05.

Key words and phrases : Nonsmooth convex optimization, BFGS method, Approximate subgradients.

1. Introduction

Consider the following unconstrained optimization problem

$$(1) \quad \min f(x) \quad \text{s.t.} \quad x \in R^n$$

where the objective function $f : R^n \rightarrow R$ is a possibly nondifferentiable convex function. The following corresponding optimization problem is considered

$$(2) \quad \min F(x) \quad \text{s.t.} \quad x \in R^n$$

where $F : R^n \rightarrow R$ is the Moreau-Yosida regularization of f , which the objective function F is defined as

$$(3) \quad F(x) = \min_{z \in R^n} \{f(z) + (2\lambda)^{-1} \|z - x\|^2\},$$

where the parameter λ is a fixed positive number and $\|\cdot\|$ denotes the Euclidean norm or its induced matrix norm on $R^{n \times n}$. It is well known that F is a continuously differentiable convex function defined on R^n even though f may be nondifferentiable. The derivative of F at x is defined by

$$(4) \quad G(x) = \nabla F(x) = \lambda^{-1}(x - p(x)) \in \partial f(p(x)),$$

Received November 21, 2009. Revised November 30, 2009. Accepted December 14, 2009.

*Corresponding author.

© 2010 Korean SIGCAM and KSCAM.

where $p(x)$ is the unique minimizer in (2) and ∂f is the subdifferential mapping of f [1].

The **approximate subdifferential** of f at x is defined as $\partial_\epsilon f(x) = \{\eta \in R^n | f(z) \geq f(x) + \eta^T(z-x) - \epsilon\}$, where η is called an approximate subgradient of f at x . There are several reasons for dealing with approximate subgradients. If a subgradient $g(x) \in \partial f(x)$ is hard to compute, one can take an already computed subgradient $g(\bar{x})$ of f at some point \bar{x} near x . Then

$$f(x) + g(\bar{x})^T(z-x) = f(\bar{x}) + g(\bar{x})^T(z-\bar{x}) + \epsilon \leq f(z) + \epsilon,$$

where $\epsilon = f(x) - f(\bar{x}) - g(\bar{x})^T(x-\bar{x}) \geq 0$. Thus $g(\bar{x}) \in \partial_\epsilon f(x)$, which indicates that $g(\bar{x})$ is an approximate subgradient of f at x . In this paper, assume that we have a black box which provides at each $x \in R^n$ and we have

$$(5) \quad g^a(x, \epsilon) \in \partial_\epsilon f(x)$$

for given $\epsilon > 0$.

The purpose of this paper is to present an implementable algorithm for solving (1) by combing Morean-Yosida regularization, bundle concept and BFGS method. By using the approximate subgradient, the algorithm presented in this paper become convenient to implement. Some notations which would be used in the following are listed.

Subdifferential: $\partial f(x) = \{\xi \in R^n | f(x) \geq f(x) + \xi^T(z-x), \forall z \in R^n\}$, the subdifferential of f at x , and each such ξ is called a *subgradient* of f at x .

Gradient: $G(x) = \lambda^{-1}(x - p(x))$, the gradient of F at x .

Unique minimizer: $p(x) = \arg \min_{z \in R^n} \{f(z) + (2\lambda)^{-1}\|z-x\|^2\}$, the unique minimizer of (3).

Hiriart-Urruty and Lemarechal [2] has proved that the optimization problem (1) and (2) are equivalent in the sense that the solution sets of the two problems coincide with each other. Related on this subject appears in [2, 3, 4, 5]. In particular, Mifflin [4] proved the global convergence results for a quasi-Newton bundle method by assuming that at each $x \in R^n$ one exact subgradient of the objective function f can be found. But it is very difficult to compute subgradients accurately. In this paper, we consider using approximate subgradients of the objective function f instead of the exact one, which can make the implement of the algorithm very easy. Using the BFGS algorithm, the quasi-Newton algorithm is used to find the decrease direction. It was shown that Broyden's class of quasi-Newton algorithm converges globally and superlinearly [6, 7]. Byrd [8, 9] proved global and superlinear convergence of the convex Broyden's class with Wolfe-type line search and the global convergence of BFGS algorithm with backtracking line search. Li[10] proved that when the objective function f is a convex quadratic function, DFP algorithm has global convergence property. The convergence properties of the BFGS method for convex minimization have been studied by many researchers. There have already been a lot of achievements in global convergence properties of BFGS algorithm [11, 12, 13, 14, 15, 16]. Fletcher gave a review of unstrained optimization [17].

In the section 2 and 3, using the bundle method, we discuss how to approximate the unique minimizer $p(x)$ of (3), and give the BFGS bundle algorithm based on the approximate subgradient. The convergence property is proved in section 4. Finally, some conclusions and discussions are give.

2. The BFGS algorithm and the bundle idea

We first recall the steps of the standard BFGS method. The sequence $\{x_k\}$ generated by the standard BFGS method with line search is determined by letting $x_{k+1} = x_k + \alpha_k p_k$, where p_k is a solution of the following system of linear equation

$$(6) \quad B_k p + g(x_k) = 0.$$

α_k is the step length. The matrix B_k in the standard BFGS method is updated by the formula:

$$(7) \quad B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k},$$

where $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.

The bundle idea plays a central role in approximating $F(x)$ and $\nabla F(x)$ as is developed in [5]. Let $d = z - x$ in (3) and minimize over d instead of z . Then we have

$$(8) \quad F(x) = \min_{d \in R^n} \{f(x+d) + (2\lambda)^{-1} \|d\|^2\}.$$

Considering approximating $f(x+d)$ by using the bundle method. According to the assumption presented in the introduction, for all $y^i \in R^n$ and $\forall \epsilon_i > 0, i = 1, 2, \dots, j$, we can compute $f(y^i)$ and one $g^a(y^i, \epsilon_i) \in \partial_{\epsilon_i} f(y^i)$, which means

$$(9) \quad f(z) \geq f(y^i) + g^a(y^i, \epsilon_i)^T (z - y^i) - \epsilon_i, \quad \forall z \in R^n.$$

Then a polyhedral function

$$(10) \quad \check{f}_a(x+d) = \max_{i=1, \dots, j} \{f(y^i) + g^a(y^i, \epsilon_i)^T (x+d - y^i) - \epsilon_i\},$$

where ϵ_i is updated by the rule $\epsilon_{i+1} = \gamma \epsilon_i, 0 < \gamma < 1$. Then we have $f(x+d) \geq \check{f}_a(x+d)$. If the linear error $\alpha(x, y^i, \epsilon_i)$ is defined by $\alpha(x, y^i, \epsilon_i) = f(x) - f(y^i) - g^a(y^i, \epsilon_i)^T (x - y^i)$, then (10) can be written as

$$(11) \quad \check{f}_a(x+d) = f(x) + \max_{i=1, \dots, j} \{g^a(y^i, \epsilon_i)^T d - \alpha(x, y^i, \epsilon_i) - \epsilon_i\}.$$

Let

$$(12) \quad \check{F}(x) = \min_{d \in R^n} \{\check{f}_a(x+d) + (2\lambda)^{-1} \|d\|^2\}.$$

If $d(x)$ solves (12) and let $v(x) = \max_{i=1, \dots, j} \{g^a(y^i, \epsilon_i)^T d(x) - \alpha(x, y^i, \epsilon_i) - \epsilon_i\}$, then $\check{F}(x) = f(x) + v(x) + (2\lambda)^{-1} d(x)^T d(x)$. Let $a(x) = x + d(x)$ be an approximation of the one in (3), then one has

$$\hat{F}(x) = f(a(x)) + (2\lambda)^{-1} d(x)^T d(x).$$

Then six properties of the functions defined above can be given, which will be used in the follows [18].

P1: $\check{F}(x) \leq F(x) \leq \hat{F}(x)$.

P2: $\hat{F}(x) = F(x)$ if and only if $a(x) = p(x)$.

P3: Let $\epsilon(x) = \hat{F}(x) - \check{F}(x)$ and $\delta(x)$ is a given positive number. Suppose that x is not the minimizer of f . If

$$\epsilon(x) < \delta(x) \min\{\lambda^{-2}d(x)^T d(x), M\}$$

is never satisfied, then $\epsilon \rightarrow 0$ as $j \rightarrow \infty$.

P4: Let $\hat{G}(x) = \lambda^{-1}(x - a(x)) = -\lambda^{-1}d(x)$. Then one has $\|G(x) - \hat{G}(x)\| = \|\lambda^{-1}(p(x) - a(x))\| \leq \sqrt{\frac{2a(x)}{\lambda}}$.

P5: If x does not minimize f , then we can find one solution $d(x)$ of (12) such that $\hat{F}(x) - \check{F}(x) < \delta(x) \min\{\lambda^{-2}d(x)^T d(x), M\}$ holds, where M is a given positive number.

3. The BFGS bundle algorithm based on approximate subgradient

The notations $a(x_k), d(x_k)$ are abbreviated as a_k, d^k . Given positive numbers π, ν, γ, M such that $\pi < 0.5, \nu < 1, 0 < \gamma < 1$, and one $n \times n$ symmetric positive definite matrix R_N .

The BFGS bundle algorithm:

Step 0(initialization): Choose a sequence of positive numbers $\{\delta_k\}_{k=0}^{\infty}$ such that $\sum_{k=0}^{\infty} \delta_k < \infty$. Let x^0 be a starting point and B_0 be a initial symmetric positive definite matrix. Set $k = 0$ and find d^0 and a_0 which satisfying

$$a_0 \leq \delta_0 \min\{\lambda^{-2}(d^0)^T d^0, M\}.$$

Let ϵ_1 be a positive number and start the bundle process with $j = 0$ and $y^0 = x_0$.

Step 1(compute a search direction): Compute $a(x_k)$ by the bundle process where ϵ_i is updated by $\epsilon_{i+1} = \gamma\epsilon_i$. Compute $\hat{G}(x_k)$. If $\|\hat{G}(x_k)\| \neq 0$, compute

$$(13) \quad s^k = -B_k^{-1}\hat{G}(x_k).$$

Otherwise, stop.

Step 2(line search): Starting with $u = 1$, let i_k be the smallest nonnegative integer u such that

$$(14) \quad \check{F}(x_k + \nu^k s^k) \leq \hat{F}(x_k) + \pi\nu^u (s^k)^T \hat{G}(x_k),$$

where $F_a(x_k + \nu^k s^k)$ satisfies

$$(15) \quad \check{F}(x_k + \nu^k s^k) - \hat{F}(x_k + \nu^k s^k) \leq \delta_{k+1} \min\{\lambda^{-2}d(x_k + \nu^k s^k)^T d(x_k + \nu^k s^k), M\}.$$

Set $t^k = \nu^{i_k}$ and $x_{k+1} = x_k + t^k s^k$.

Step 3(update B_k): Let $\delta x_k = x_{k+1} - x_k$ and $\delta y_k = \hat{G}(x_{k+1}) - \hat{G}(x_k)$. If $(\delta x_k)^T \delta y_k > 0$, update B_k to B_{k+1} by the following way to satisfy $B_{k+1} \delta x_k = \delta y_k$,

$$(16) \quad B_{k+1} = B_k - \frac{B_k \delta x_k \delta x_k^T B_k}{\delta x_k^T B_k \delta x_k} + \frac{\delta y_k \delta y_k^T}{\delta y_k^T \delta x_k}.$$

Otherwise set $B_{k+1} = R_N$. Set $k = k + 1$ and goto Step 1.

End of the algorithm

4. The global convergence

Theorem 1. *Given $c_3 \in (0, \infty)$ and $c_4 \in (0, 1)$, if the following conditions are hold, $\delta x^T \delta y > 0$.*

$$(17) \quad \|\delta x_k\|(\sqrt{2\epsilon_k} + \sqrt{2\epsilon_{k+1}}) \leq c_3(\delta x_k)^T \delta y_k,$$

$$(18) \quad 2\|\delta y_k\|(\sqrt{2\epsilon_k} + \sqrt{2\epsilon_{k+1}}) \leq \min\{c_4, \delta_k^{1/3} + \delta_{k+1}^{1/3}\|\delta y_k\|^2\}.$$

Proof. Let $\delta \bar{y}_k = G(x_{k+1}) - G(x_k)$. From P4, one has

$$\begin{aligned} (\delta x_k)^T \delta y_k &= (\delta x_k)^T \delta \bar{y}_k + (\delta x_k)^T (\delta y_k - \delta \bar{y}_k) \\ &\geq (\delta x_k)^T \delta \bar{y}_k - \|\delta x_k\| \|\delta y_k - \delta \bar{y}_k\| \\ &\geq (\delta x_k)^T \delta \bar{y}_k - \|\delta x_k\| (\|\hat{G}(x_k) - G(x_k)\| + \|\hat{G}(x_{k+1}) - G(x_{k+1})\|) \\ &\geq (\delta x_k)^T \delta \bar{y}_k - \|\delta x_k\| (\sqrt{2\epsilon_k} + \sqrt{2\epsilon_{k+1}}) \\ &\geq \frac{1}{1+c_3} (\delta x_k)^T \delta \bar{y}_k. \end{aligned}$$

and

$$\begin{aligned} \|\delta \bar{y}_k\|^2 &= \|\delta y_k\|^2 + \|\delta \bar{y}_k - \delta y_k\|^2 + 2(\delta y_k)^T (\delta \bar{y}_k - \delta y_k) \\ &\geq \|\delta y_k\|^2 - 2\|\delta y_k\| \|\delta \bar{y}_k - \delta y_k\| \\ &\geq \|\delta y_k\|^2 - 2\|\delta y_k\| (\sqrt{2\epsilon_k} + \sqrt{2\epsilon_{k+1}}) \\ &\geq \|\delta y_k\|^2 + \min\{c_4, \delta_k^{1/3} + \delta_{k+1}^{1/3}\|\delta y_k\|^2\} \\ &\geq (1-c_4)\|\delta y_k\|^2. \end{aligned}$$

So, if conditions (17) and (18) are satisfied, then $(\delta x_k)^T \delta y_k \geq \frac{1}{1+c_3} (\delta x_k)^T \delta \bar{y}_k$ and $\|\delta \bar{y}_k\|^2 \geq (1-c_4)\|\delta y_k\|^2$. \square

Lemma 1. *For any nonnegative sequence $\{\delta_k\}_{k \geq 0}$, if $\sum_{k=0}^{\infty} \delta_k < \infty$, then*

$$(19) \quad \prod_{k=0}^{\infty} (1 + \delta_k) < \infty.$$

Proof. This result follows easily from the properties of logarithms. \square

Lemma 2. *Relative to the line search there exist positive constants η_1 and η_2 such that either [4]*

$$(20) \quad \begin{aligned} \tilde{F}(x_k + \tau_k s_k) &\leq \hat{F}(x_k) - \eta_1 \frac{[(s^k)^T \hat{G}(x_k)]^2}{\|s^k\|^2} \\ &\quad - \frac{\eta_1}{1-\sigma} \frac{(s^k)^T [G(x_k) - \hat{G}(x_k)] [(s^k)^T \hat{G}(x_k)]}{\|s^k\|^2} \end{aligned}$$

or

$$(21) \quad \check{F}(x_k + \tau_k s_k) \leq \hat{F}(x_k) + \eta_2 (s^k)^T \hat{G}(x_k).$$

Theorem 2. *Suppose that F is strongly convex on D and $\{B_k\}$ is generated by the BFGS bundle-type method and $x^k \neq \bar{x}$ for all $k \geq 0$. Then $\{x_k\}$ converges to the unique solution \bar{x} .*

Proof. Let $K := \{0\} \cup \{j | (17) \text{ or } (18) \text{ does not hold for } k = j - 1\} \equiv \{k_0, k_1, \dots, k_i, \dots\}$. Therefore, $B_j = M$ for $j \in K$ and B_j is a BFGS update of B_{j-1} for $j \notin K$.

Suppose that K has an infinite number of elements. Since F is strongly convex on D and G is globally Lipschitz continuous, from theorem 1, one has $(\delta x_k)^T \delta \bar{y}_k \geq \alpha \|\delta x_k\|^2$ and $(\delta x_k)^T \delta \bar{y}_k \geq \lambda \|\delta \bar{y}_k\|^2$. From Theorem 2.1 in [9], given $w \in (0, 1)$, there exist constants $\beta, \beta' > 0$ such that for any k satisfying $k_{i-1} \leq k < k_i - 1$, where $k_{i-1}, k_i \in K$ for some $i \geq 1$, the inequalities $\cos \theta_j \geq \beta$ and $\frac{B_j \delta x_j}{\|\delta x_j\|} \leq \beta'$ hold. Since $B_j = M$ if $j \in K$, we can assume β and β' are such that the above two inequalities hold for all $j \in K$. Define I to be the set of j for that $\cos \theta_j \geq \beta$ and $\frac{B_j \delta x_j}{\|\delta x_j\|} \leq \beta'$ hold. Since D is bounded, $\{\|G(x_k)\|\}$ is a bounded sequence. From P4,

$$(22) \quad \lambda^{-1} d(x)^T d(x) = (\bar{G}(x))^T \lambda \bar{G}(x)$$

and (15), $\|G(x_k) - \hat{G}(x_k)\| = o(\|\hat{G}(x_k)\|)$, so there exist an integer \bar{k} such that for all $k \geq \bar{k}$

$$(23) \quad 2\|G(x_k)\| \geq \|\bar{G}(x_k)\| \geq \frac{1}{2}\|G(x_k)\|,$$

$$(24) \quad \left| -\frac{(s^k)^T [G(x_k) - \bar{G}(x_k)] [(s^k)^T \bar{G}(x_k)]}{\|s^k\|^2} \right| \leq \frac{(1 - \sigma)\beta^2}{2} \|\bar{G}(x_k)\|^2.$$

Consider an iterate x_j with $j \in I$ and $j \geq k$. From Lemma 2, (24) and the above two inequalities, one has

$$(25) \quad \hat{F}(x_j) - \hat{F}(x_j + \tau_j s^j) \geq \eta \|\bar{G}(x_j)\|^2,$$

where $\eta = \frac{1}{2}\eta_1\beta^2$ if (20) holds or $\eta = \eta_2\frac{\beta}{\beta'}$ if (21) holds. Thus, from (23) and (25), for all $j \in I$ and $j \geq k$,

$$(26) \quad \hat{F}(x_j) - \hat{F}(x_j + \tau_j s^j) \geq \frac{\eta}{4} \|G(x_j)\|^2.$$

Because F is strong convex and according Lemma 4.3 in [19], for all $k \geq 0$,

$$(27) \quad \frac{1}{2}\alpha \|x_k - \bar{x}\|^2 \leq F(x_k) - F(\bar{x}) \leq \frac{2}{\alpha} \|G(x_k)\|^2.$$

Then, from P1, P2, (26) and the right-side inequality in (27), for all $j \in I$ and $j \geq k$,

$$(28) \quad F(x_{j+1}) - F(\bar{x}) - \epsilon_{j+1} \leq \left(1 - \frac{\eta\alpha}{8}\right) (F(x_j) - F(\bar{x})) + \epsilon_j.$$

Since $\{\delta_k\} \rightarrow 0$, we can take \bar{k} large enough such that for all $k \geq \bar{k}$

$$(29) \quad \frac{16\delta_k\lambda}{\alpha} \leq \min \left\{ 1, \frac{\eta\alpha}{8} \right\}.$$

By (15), (22), (23), the fact that $G(\bar{x}) = 0$, the Lipschitz continuity of G , and (27), for all $k \geq \bar{k}$ one has

$$(30) \quad \epsilon_k \leq 4\delta_k\lambda\|x_k - \bar{x}\|^2 \leq \frac{8\delta_k\lambda}{\alpha}(F(x_k) - F(\bar{x})).$$

Then from (28)-(30), for all $j \in I$ and $j \geq \bar{k}$, we have

$$(31) \quad \left(1 - \frac{8\delta_{j+1}\lambda}{\alpha}\right)(F(x_{j+1}) - F(x_j)) \leq \left(1 - \frac{1}{16}\eta\alpha\right)(F(x_j) - F(\bar{x})).$$

Since $F(x_k) > F(\bar{x})$ for all k , (31) and (29) imply $1 - \frac{1}{16}\eta\alpha > 0$. For $w \in (0, 1)$, let $r = (1 - \frac{1}{16}\eta\alpha)^w$ so that in (31)

$$(32) \quad 1 - \frac{1}{16}\eta\alpha = r^{1/w}.$$

From (13), (14), the positivity of σ and τ_k , and the positive definiteness of B_k , one has

$$(33) \quad \tilde{F}(x_{k+1}) < \hat{F}(x_k) \text{ for all } k.$$

Combining this with (30), P1, and P2 yields for all $j \geq \bar{k}$

$$\left(1 - 8\frac{\delta_{j+1}\lambda}{\alpha}\right)(F(x_{j+1}) - F(\bar{x})) \leq \left(1 + 8\frac{\delta_j\lambda}{\alpha}\right)(F(x_j) - F(\bar{x})).$$

For $k \geq \bar{k}$, let

$$\delta'_k = \frac{1 + 8\frac{\delta_k\lambda}{\alpha}}{1 - 8\frac{\delta_{k+1}\lambda}{\alpha}}.$$

For any $k \geq \bar{k}$, there exists $k_{i-1}, k_i \in K$ such k satisfies $k_{i-1} \leq k < k_i$. If $k_i - k_{i-1} \leq 2$, for $k_{i-1} \leq k < k_i$,

$$(34) \quad \begin{aligned} F(x_{k+1}) - F(\bar{x}) &\leq \prod_{j=k_{i-1}}^k \delta'_j r (F(x_{k_{i-1}}) - F(\bar{x})) \\ &\leq \prod_{j=k_{i-1}}^k \delta'_j (r^{1/2})^{k-k_{i-1}+1} (F(x_{k_{i-1}}) - F(\bar{x})). \end{aligned}$$

On the other hand, if $k_i - k_{i-1} \leq 2$, then when $k_{i-1} \leq k < k_i - 1$, from Lemma 5.2 in [20], there are at least $[w(k - k_{i-1} + 1)]$ elements in $I \cap [k_{i-1}, k]$. So for all k satisfying $k_{i-1} \leq k < k_i - 1$, one has

$$(35) \quad F(x_{k+1}) - F(\bar{x}) \leq \prod_{j=k_{i-1}}^k \delta'_j r^{k-k_{i-1}+1} (F(x_{k_{i-1}}) - F(\bar{x})).$$

Therefore,

$$(36) \quad F(x_{k_i}) - F(\bar{x}) \leq \delta'_{k_i-1} (F(x_{k_i-1}) - F(\bar{x})) \leq \prod_{j=k_{i-1}}^{k_i-1} \delta'_j r^{k_i-k_{i-1}+1} (F(x_{k_{i-1}}) - F(\bar{x})).$$

So, from (34)-(36), for all k satisfying $k_{i-1} \leq k < k_i$, one has

$$(37) \quad]F(x_{k+1}) - F(\bar{x}) \leq \prod_{j=k_{i-1}}^k \delta'_j r^{k-k_{i-1}+1} (F(x_{k_{i-1}}) - F(\bar{x})).$$

Without loss of generality, we can assume that $k \in K$. Then, from (37), for any $k \geq \bar{k}$, one has

$$(38) \quad F(x_{k+1}) - F(\bar{x}) \leq \prod_{j=\bar{k}}^k \delta'_j (r^{1/2})^{k-\bar{k}+1} (F(x_{\bar{k}}) - F(\bar{x})).$$

Since $\sum_{k=0}^\infty \delta_k < \infty, \sum_{k=\bar{k}}^\infty (\delta'_k) < \infty$. So, from Lemma 1, there exist a constant $C > 0$ such that

$$(39) \quad \prod_{k=\bar{k}}^\infty \delta'_k \leq C.$$

Then, for all $k \geq \bar{k}$

$$(40) \quad F(x_{k+1}) - F(\bar{x}) \leq C(r^{1/2})^{k-\bar{k}+1} (F(x_{\bar{k}}) - F(\bar{x})).$$

Using (27), (40), and the fact that $r < 1$, one has

$$(41) \quad \begin{aligned} \sum_{k=\bar{k}}^\infty \|x_k - \bar{x}\| &\leq (2/\alpha)^{1/2} \sum_{k=\bar{k}}^\infty (F(x_k) - F(\bar{x}))^{1/2} \\ &\leq \left[\frac{2C(F(x_{\bar{k}}) - F(\bar{x}))}{\alpha} \right]^{1/2} \sum_{k=\bar{k}}^\infty (r^{1/4})^{k-\bar{k}} < \infty. \end{aligned}$$

Therefore, the algorithm has global convergence property. If there are only finitely many elements in K , then we can prove the same results as in the case where there are infinitely many elements in K . □

5. Conclusion

In summary, this paper presents a BFGS bundle algorithm for nonsmooth convex unstrained program. The global convergent BFGS bundle-type method for the case where the Moreau-Yosida regularization function F and its gradient G are computed by the approximate subgradient. The approximate subgradient of the objective function is used in the algorithm, which can make the algorithm easier to implement. The algorithm does not need require the original objective to be differentiable at the solution.

ACKNOWLEDGEMENT

This work is funded by the National Basic Research Program of China (973 Program No.2006CB705500), the NSFC (Grant Nos. 10635040, 10532060 and 10472116). Foundation of Education Department of Liaoning Province Grand No. 20060140.

REFERENCES

1. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
2. J. B. Hiriart-Urruty, C. Lemarechal, *Convex analysis and minimization algorithms*, Springer Verlag, Berlin Germany, 1993.
3. M. Fukushima, L. Qi, *A globally and superlinearly convergent algorithm for nonsmooth convex minimization*, SIAM J. Optim. **6** (1996) 1106-1120.
4. R. Mifflin, D. Sun, L. Qi, *Quasi-Newton Bundle-Type Methods for Nondifferentiable Convex Optimization*, SIAM J. OPTIM **8**(2) (1998) 583-603.
5. R. Mifflin, *A Quasi-Second-Order Proximal Bundle Algorithm*, Math. Programming **73** (1996) 51-72.
6. M. J. D. Powell, *On the convergence of the variable metric algorithm*, J. Inst. Math. Appl. **7** (1971), 21-36.
7. L. C. W. Dixon, *Variable metric algorithms: necessary and sufficient conditions for identical behavior on nonquadratic functions*, J. Optim. Theory Appl. **10** (1972) 34-40.
8. R. Byrd, J. Nocedal, Y. Yuan, *Global convergence of a class of quasi-Newton methods on convex problems*, SIAM J. Numer. Anal. **24** (1987) 1171-1189.
9. R. H. Byrd, J. Nocedal, *A tool for the analysis of quasi-Newton methods with application to unconstrained minimization*, SIAM J. Numer. Anal. **26** (1989) 727-739.
10. D. -H. Li, *On the global convergence of DFP method*, J. Hunan Univ. (Natural Sciences) **20** (1993) 16-20.
11. M. J. D. Powell, *Some Global Convergence Properties of a Variable Metric Algorithm for Minimization without Exact Line Searches*, Nonlinear Programming, SIAM-AMS proceedings Cottle R W and Lemke C E, eds. **9** (1976) 53-72.
12. J. Werner, *Über die global konvergenz von variablemetrischen verfahren mit nichtexakter schrittweitenbestimmung*, Numer. Math. **31** (1978) 321-334.
13. J. G. Liu, Q. Guo, *Global Convergence Properties of the Modified BFGS Method*, J. of Appl. Math. & Computing **16** (2004) 195-205.
14. J. G. Liu, Z. Q. Xia, R. D. Ge, and Q. Guo, *An modified BFGS method for non-convex minimization problems (Chinese)*, Operation Research and Management **13** (2004) 62-65.
15. Rendong Ge and Zunquan Xia, *An ABS Algorithm for Solving Singular Nonlinear System with Rank One Defect*, Korea. J. Comp. Appl. Math. **9** (2002) 167-184.
16. Rendong Ge and Zunquan Xia, *An ABS Algorithm for Solving Singular Nonlinear System with Rank Defects*, J. of Appl. Math. & Computing **5** (2003) 1-20.
17. R. Fletcher, *An overview of unconstrained optimization*, in: *E. Spedicato (Ed.), Algorithms for Continuous Optimization: The State of the Art*, Klumer Academic Publisher, Boston, (1994), 109-143.
18. J. Shen, L. Pang, *A quasi-Newton bundle method based on approximate subgradients*, J. of Appl. Math. & Computing **23** (2007) 361-367.
19. J. F. Bonnans, J. CH. Gilbert, C. Lemaréchal, C. A. Sagastizábal, *A family of variable metric proximal methods*, Math. Programming **68** (1995) 15-47.
20. X. Chen, M. Fukushima, *Proximal Quasi-Newton Methods for Nondifferentiable Convex Optimization*, Applied Mathematics report 95/32, School of Mathematics, The University of New South Wales, Sydney, Australia, 1995.

Qiang Guo is Associated Professor in Business School, University of Shanghai for Science and Technology. Her research interests are the optimization methods in science and engineering, nonsmooth optimization.

Business School, University of Shanghai for Science and Technology, Shanghai 200093, P.R. China.

E-mail: jmz.gq1@163.com

Jian-Guo Liu is a Associated Professor in Research Center of Complex Systems Science, University of Shanghai for Science and Technology. His research interests are the optimization methods in science and engineering, complex network.

Research Center of Complex Systems Science, University of Shanghai for Science and
Technology, Shanghai 200093, P. R. China.
E-mail: liujg004@ustc.edu.cn