A Variant of the Multi-Step Bundle Method*

Rashid Yarullin

Kazan (Volga Region) Federal University, Institute of Computational Mathematics and Information Technologies, Kremlyovskaya str. 35, 420008 Kazan, Russian
{yarullinrs}@gmail.com
http://www.kpfu.ru

Abstract. A method from a class of bundle methods is proposed to solve an unconstrained optimization problem. In this method an epigraph of the objective function is approximated by the set which is formed on the basis of the convex quadratic function. This method is characterized in that iteration points are constructed in terms of information obtained in the previous steps of the minimization process. Computational aspects of the proposed method are discussed, convergence of this one is proved, and convergence rate of the iteration process is obtained.

Keywords: a bundle method, an epigraph, multi-step methods, approximation sets, convergence rate

1 Introduction

A class of bundle methods is quite wide (e.g. [2–6]). Such methods use multi-step approach of constructing iteration points to solve a convex programming problem. Namely, the next approximation is formed in term of prehistory of the solution process by minimizing an auxiliary convex quadratic function. Taking into account this feature the given methods profitably differ from one-step methods by constructing anti gully trajectory of the iteration points and good convergence rate.

In this paper the method is proposed for solving a convex programming problem which belongs to the mentioned class. The suggested method also applies multi-step technique of constructing approximations. Moreover, note that unlike the famous bundle methods the solution of the auxiliary quadratic programming problem is obtained in the proposed method by the formula, and this fact is convinient to use in practical implementations of the method.

2 Problem Setting

The method is proposed for solving the following problem:

* The research was supported by the Russian Fundamental Research Foundation within the project 16-31-60014.

In: A. Kononov et al. (eds.): DOOR 2016, Vladivostok, Russia, published at http://ceur-ws.org

$$\min\{f(x) : x \in \mathrm{R_n}\}, \tag{1}$$

where $f(x)$ is a continuously differentiable convex function defined in an n-dimensional Euclidian space $\mathrm{R_n}$, and the gradient of the function $f(x)$ satisfies Lipschitz continuous condition $|f'(x) - f'(y)| \le L||x - y||$ for all x, $y \in \mathrm{R_n}$ with the parameter $L > 0$.

Let $f^* = \min\{f(x) : x \in \mathrm{R_n}\}$, $X^* = \{x \in \mathrm{R_n} : f(x) = f^*\} \neq \emptyset$, $\mathrm{epi}(f) = \{(x, \delta) \in \mathrm{R_{n+1}} : \delta \ge f(x)\}$, $K = \{0, 1, \dots\}$. By $|B|$ denote the cardinality of the set $B \subset \mathrm{R_n}$.

3 The Bundle Method

A sequence $\{x_k\}$, $k \in K$, is constructed by the proposed method as follows.

0. Define numbers $l > 0$, $\gamma > 0$ and $r \in K$ such that $r \ge 1$. Select any point $v \in \mathrm{R_n}$. Assign $\eta_0 = l$, $k = 0$ and $B_0 = \{v\}$.

1. Choose

$$x_k = \mathrm{argmin}\{f(y) : y \in B_k\}, \tag{2}$$

and find a point z_k as a solution of the problem

$$\min\{\varphi_k(x) : x \in \mathrm{R_n}\}, \tag{3}$$

where

$$\varphi_k(x) = f(x_k) + \langle f'(x_k), x - x_k\rangle + \frac{\gamma}{2}||x - x_k||^2. \tag{4}$$

Determine $\varphi_k = \varphi_k(z_k)$.

2. Choose a number $\alpha_k(b) \in (0, 1]$ for each $b \in B_k$ such that

$$(u_k(b), \mu_k(b)) = (b, \theta_k(b)) + \alpha_k(b)\big((z_k, \varphi_k) - (b, \theta_k(b))\big) \in \mathrm{epi}(f), \tag{5}$$

where

$$\theta_k(b) = f(b) + \eta_k. \tag{6}$$

.

3. Find a point

$$w_k = \mathrm{argmin}\{f(u_k(b)) : b \in B_k\}. \tag{7}$$

4. If the inequality $|B_k| < r$ is defined, then construct the next set

$$B_{k+1} = B_k \bigcup \{w_k\}. \tag{8}$$

Otherwise find a point $v_k = \mathrm{argmax}\{f(a) : a \in B_k\}$, and determine

$$B_{k+1} = B_k \setminus \{v_k\} \bigcap \{w_k\}. \tag{9}$$

5. Assign

$$\eta_{k+1} = \frac{l}{(k+1)^2}, \tag{10}$$

increment k by one, and go to Step 1.

Firstly, lets represent some properties of the suggested method.

Remark 1. The point v which is described at Step 0 in the algorithm is the initial iteration point. Unfortunately, there is no any general approach of constructing the initial iteration point in nonlinear programming methods. But in the process of solving practical optimization problems there are some informations to construct a region which approximate the set of solutions. Consequently, it is obviously to select the initial iteration point v as close as possible to the mentioned region.

Remark 2. Note that as well as the multi-step methods the proposed method uses prehistory of the solution process. Namely, according to (3), (5), (7) and Step 4 of the algorithm the next approximation x_{k+1}, $k > r$, is selected from the set B_{k+1} which is constructed on the basis of the last $r > 0$ iteration points of the sequence $\{x_i\}$, $i \in K$.

In the bundle methods each iteration point is obtained by minimizing the auxiliary quadratic function constructed on the basis of the model of the objective function. Since this model consists of several cutting planes, then it is nessesary to use various numerical methods for solving quadratic programming problems. In the proposed method the model of the objective function contains only one cutting plane. Thus, the solution of the auxiliary quadratic problem can be found by the formula. This result is represented in the following statement.

Lemma 1. *Suppose that sequences $\{z_k\}$, $\{\varphi_k\}$, $k \in K$, are constructed by the proposed method. Then equalities*

$$z_k = x_k - \frac{f'(x_k)}{\gamma}, \tag{11}$$

$$\varphi_k = f(x_k) - \frac{||f'(x_k)||^2}{2\gamma} \tag{12}$$

are defined for all $k \in K$.

Proof. Note that the function $\varphi_k(x)$ is differentiable and strongly convex. Consequently, problem (3) has unique solutions for all $k \in K$. Lets compute partial derivatives of the function $\varphi_k(x)$ which have the following form: $\frac{\partial \varphi_k(x)}{\partial x[i]} = f'(x_k)[i] + (x[i] - x_k[i])\gamma = 0$, $i = \overline{1,n}$. Hence, equation (11) is defined. Further, in view of (4), (11) we have $\varphi_k = \varphi_k(z_k) = f(x_k) - \frac{||f'(x_k)||^2}{\gamma} + \frac{||f'(x_k)||^2}{2\gamma} = f(x_k) - \frac{||f'(x_k)||^2}{2\gamma}$. The lemma is proved.

Lemma 2. *Suppose that the sequence $\{x_k\}$, $k \in K$, is constructed by the suggested method. Then the inequality*

$$f(x_{k+1}) \le f(u_k(x_k)) \le f(x_k) + \eta_k\big(1 - \alpha_k(x_k)\big) - \frac{\alpha_k(x_k)}{2\gamma}||f'(x_k)||^2 \tag{13}$$

is satisfied for all $k \in K$.

Proof. In accordance with ways (8), (9) of constructing the set B_{k+1} the inclusion $w_k \in B_{k+1}$ is defined for all $k \in K$. Hence, in view of (2), (7) the expression $f(x_{k+1}) \le f(w_k) \le f(u_k(x_k))$ is determined. Further, taking into account (5), (6), (12) we have $f(u_k(x_k)) \le \mu_k(x_k) = \theta_k(x_k) + \alpha_k(x_k)\big(\varphi_k - \theta_k(x_k)\big) = f(x_k) + \eta_k + \alpha_k(x_k)f(x_k) - \frac{\alpha_k(x_k)}{2\gamma}||f'(x_k)||^2 - \alpha_k(x_k)f(x_k) - \alpha_k(x_k)\eta_k = f(x_k) + \eta_k\big(1 - \alpha_k(x_k)\big) - \frac{\alpha_k(x_k)}{2\gamma}||f'(x_k)||^2$. The Lemma is proved.

Before proving convergence of the proposed method lets construct the parameter $\alpha_k(x_k)$, $k \in K$, by the following rule.

Lemma 3. *If*

$$(\varphi_k, z_k) \in \text{epi}(f), \tag{14}$$

then $\alpha_k(x_k) = 1$. *Otherwise* $\alpha_k(x_k)$ *is selected so that the equation*

$$f(u_k(x_k)) = \mu_k(x_k) \tag{15}$$

is defined. Then there exists a constant $c > 0$ *such that*

$$\alpha_k(x_k) \geq c \quad \forall k \in K. \tag{16}$$

Proof. Lets fix numbers $\varepsilon \in (1, 2)$ and $i \geq 0$ such that

$$\frac{L2^{-i}}{\gamma} \leq 2 - \varepsilon, \tag{17}$$

where $L > 0$ is Lipshitz constant of the gradient $f'(x)$.

Since $f(x)$ is a continuously differentiable functions, and its gradient satisfies Lipshitz condition, then in view of $\eta_k(1 - 2^{-i}) \geq 0$ for all $k \in K$ we give $f(x_k) - f(x_k - \frac{2^{-i}}{\gamma} f'(x_k)) + \eta_k(1 - 2^{-i}) \geq f(x_k) - f(x_k - \frac{2^{-i}}{\gamma} f'(x_k)) \geq \langle f'(x_k), \frac{2^{-i}}{\gamma} f'(x_k) \rangle - \frac{L}{2} || \frac{2^{-i}}{\gamma} f'(x_k)||^2 = \frac{2^{-i}}{\gamma} ||f'(x_k)||^2 - \frac{L}{\gamma^2} 2^{-2i-1} ||f'(x_k)||^2 = \frac{2^{-i-1}}{\gamma} ||f'(x_k)||^2 \varepsilon \geq \frac{2^{-i}}{2\gamma} ||f'(x_k)||^2$ for all $k \in K$.

Hence, by putting

$$\bar{u}_k = x_k - \frac{2^{-i}}{\gamma} f'(x_k) = x_k + 2^{-i}(z_k - x_k), \tag{18}$$

$$\bar{\mu}_k = f(x_k) + \eta_k(1 - 2^{-i}) - \frac{2^{-i}}{2\gamma} ||f'(x_k)||^2 = \theta_k(x_k) + 2^{-i}(\varphi_k - \theta_k(x_k)), \tag{19}$$

we have $f(\bar{u}_k) \leq \bar{\mu}_k$, consequently,

$$(\bar{u}_k, \bar{\mu}_k) \in \text{epi}(f). \tag{20}$$

Now lets prove that there exist a constant $c > 0$ such that inequality (16) is defined. Note that according to conditions of the lemma the parameter $\alpha_k(x_k)$ is constructed by 2 ways. Firstly, if inclusion (14) is determined for some $k \in K$, then $\alpha_k(x_k) = 1$. Secondly, let the parameter $\alpha_k(x_k)$ is defined in accordance with (15). In this case according to (5) the point $(u_k(x_k), \mu_k(x_k))$ is situated in the intersection of the segment $[(z_k, \varphi_k), (x_k, \theta_k(x_k))]$ with the border of the set $\text{epi}(f)$, consequently, we have

$$(u_k(x_k), \mu_k(x_k)) \notin \text{intepi}(f). \tag{21}$$

Moreover, in view of (18), (19) we get

$$(\bar{u}_k, \bar{\mu}_k) \in [(z_k, \varphi_k), (x_k, \theta_k(x_k))]. \tag{22}$$

Now taking into account (20)-(22) lets suppose that

$$2^{-i} > \alpha_k(x_k). \tag{23}$$

Then there exists a number $\beta_k > 0$ such that

$$(\bar{u}_k, \bar{\mu}_k) = (x_k, \theta_k(x_k)) + \beta_k\big((u_k(x_k), \mu_k(x_k)) - (x_k, \theta_k(x_k))\big). \tag{24}$$

Hence, in view of (18), (19), (23) and using equality (5) while $(u_k(b), \mu_k(b)) = (u_k(x_k), \mu_k(x_k))$ the expression $\beta_k = \frac{2^{-i}}{\alpha_k(x_k)} > 1$ is defined. Further, from (24) it follows that $(u_k(x_k), \mu_k(x_k)) = (x_k, \theta_k(x_k)) + \frac{1}{\beta_k}\big((\bar{u}_k, \bar{\mu}_k) - (x_k, \theta_k(x_k))\big) = (\bar{u}_k, \bar{\mu}_k) + \bar{\beta}_k\big((x_k, \theta_k(x_k)) - (\bar{u}_k, \bar{\mu}_k)\big)$, where $\bar{\beta}_k = (1 - \frac{1}{\beta_k}) \in (0, 1)$. Then from Theorem 3 [7, p. 153] it follows that $(u_k(x_k), \mu_k(x_k)) \in \operatorname{intepi}(f)$ which contradicts to condition (21). Thus, assumption (23) is wrong, consequently, $\alpha_k(x_k) \geq 2^{-i}$. Now taking into account all cases of construction of the parameter $\alpha_k(x_k)$ for all $k \in K$ we have $\alpha_k(x_k) \geq c > 0$. The theorem is proved.

Remark 3. If inclusion (14) is not satisfied for some $k \in K$, then $(u_k(x_k), \mu_k(x_k))$ should be found as a boundary point of the set $\operatorname{epi}(f)$ by solving one-dimensional equation (15). Note that such equation is also solved in embedding methods [1] to construct cutting hyperplanes.

Theorem 1. *Suppose that the sequence $\{x_k\}$, $k \in K$, is constructed by the proposed method in accordance with conditions of Lemma 3, the set $M_\eta(x_0) = \{x \in \mathrm{R_n} : f(x) \leq f(x_0) + \eta\}$ is bounded, where $\eta = \sum_{k=0}^{\infty} \eta_k$. Then the sequence $\{x_k\}$, $k \in K$, is bounded, and the following equality takes place $\lim_{k \in K} f(x_k) = f^*$. Moreover, convergence rate*

$$f(x_k) - f^* \leq \frac{c_0}{k}, \quad k \in K, \quad k \geq 1, \tag{25}$$

is determined, where $c_0 > 0$.

Proof. In accordance with (13) and Lemma 3 the inequality

$$f(x_{k+1}) \leq f(x_k) + \eta_k - \frac{c}{2\gamma}||f'(x_k)||^2. \tag{26}$$

is defined. Hence,

$$f(x_{k+1}) \leq f(x_k) + \eta_k, \quad k \in K. \tag{27}$$

Since $f(x_k) \geq f^* > -\infty$, then from Lemma 2 [7, p. 87] and (10), (27) it follows that there exists a limit $\lim_{k \in K} f(x_k) \geq f^*$, consequently, $\lim_{k \in K} (f(x_k) - f(x_{k+1})) = 0$. Summing inequalities (27) from 0 to $m-1$ by k we have $f(x_m) \leq f(x_0) + \sum_{k=0}^{m-1} \eta_k \leq f(x_0) + \eta$. Hence, $\{x_k\} \subset M_\eta(x_0)$, and the sequence $\{x_k\}$, $k \in K$, is bounded. Further, in view of (26) we get $\lim_{k \in K} f'(x_k) = 0$, consequently, $\lim_{k \in K} f(x_k) = f^*$.

Now lets obtain convergence rate of the iteration process. For all $x^* \in X^*$ we have

$$0 \leq f(x_k) - f^* \leq ||f'(x_k)||||x_k - x^*|| \leq d||f'(x_k)||, \quad k \in K, \tag{28}$$

where $d \geq \text{diam} M_\eta(x_0)$. Suppose $a_k = f(x_k) - f^*$. Then from (26), (28) it follows that $a_k - a_{k+1} = f(x_k) - f(x_{k+1}) \geq \frac{c}{2\gamma}||f'(x_k)||^2 - \eta_k \geq \frac{c}{2\gamma d}a_k^2 - \eta_k$. Then in view of (10) and by putting $A = \max\{l, \frac{2Ld}{c}\}$ we have $a_{k+1} \leq a_k - \frac{a_k^2}{A} + \frac{A}{k^2}$ for all $k \in K$. Hence, from Lemma 5 [7, p. 89] under conditions $I_0 = K$ and $I_1 = \emptyset$ convergence rate (25) is proved.

References

1. Bulatov, V.P.: Embedding Methods in Optimization Problems. Nauka, Novosibirsk (1977) [in Russian].
2. Kiwiel, K.C.: An ellipsoid trust region bundle method for nonsmooth convex minimization. SIAM Journal on Control and Optimization 27 (4), 737–757 (1989).
3. Kiwiel, K.C.: Exact penalty functions in proximal bundle methods for constrained convex nondifferentiable minimization. Math. Program. 52 (2), 285–302 (1991).
4. Lemarechal C., Nemirovskii A., Nesterov Y. : New variants of bundle methods Math. Program. 69, 111–148 (1995).
5. Lemarechal C., Strodiot J.J., Bihain A.: On a bundle algorithm for nonsmooth optimization. Nonlinear Programming, 4, 245–281 (1981).
6. Sagastizabal, C.: Composite proximal bundle method. Math. Program. 140 (1), 189–233, (2013).
7. Vasiliev, F.P.: Optimization Methods. Factorial Press, Moscow (2002) [in Russian].