Front. Math. China 2023, 18(5): 367–380
https://doi.org/10.3868/s140-DDD-023-0026-x

RESEARCH ARTICLE

Convergence analysis of an infeasible quasi-Newton bundle method for nonsmooth convex programming

Jie SHEN[1], **Fangfang GUO**[2], **Liping PANG**[2]

1 School of Mathematics, Liaoning Normal University, Dalian 116029, China
2 School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China

Abstract By utilizing the improvement function, we change the nonsmooth convex constrained optimization into an unconstrained optimization, and construct an infeasible quasi-Newton bundle method with proximal form. It should be noted that the objective function being minimized in unconstrained optimization subproblem may vary along the iterations (it does not change if the null step is made, otherwise it is updated to a new function). It is necessary to make some adjustment in order to obtain the convergence result. We employ the main idea of infeasible bundle method of Sagastizàbal and Solodov, and under the circumstances that each iteration point may be infeasible for primal problem, we prove that each cluster point of the sequence generated by the proposed algorithm is the optimal solution to the original problem. Furthermore, for BFGS quasi-Newton algorithm with strong convex objective function, we obtain the condition which guarantees the boundedness of quasi-Newton matrices and the R-linear convergence of the iteration points.

Keywords Non-smooth optimization, convex constraint, improvement function, bundle method, quasi-Newton direction

1 Introduction

In the field of operations research, it is often necessary to minimize a function that has no derivatives at some points, and such problems are called non-differentiable optimization (NDO) or nonsmooth optimization (NSO). Many practical problems are NSO problems. Let the nonnegative variable x represent income.

Translated from *Advances in Mathematics* (*China*), 2016, 45(2): 299–308
Corresponding author: Jie SHEN, E-mail: tt010725@163.com

In many economic systems, income x corresponds to a tax function $T(x)$, which usually has discontinuous derivatives. Given some threshold $0 = a_0 < a_1 < \cdots < a_m = +\infty$ and tax rate $0 = r_0, r_1, \ldots, r_{m-1}$. T is given in the following form: $T(x) = T_i + r_i x$, $a_i \leqslant x < a_{i+1}$, $i = 0, 1, \ldots, m-1$, where $T_0 = 0$, $T_i = T_{i-1} + a_i(r_{i-1} - r_i)$. It is easy to verify that if $|r_i| < 1$, $r_{i+1} \geqslant r_i \geqslant 0$, then T is a continuously increasing convex function. Further, it is easy to verify that T has the following form: $T(x) = \max\{T_i + r_i x \mid i = 0, 1, \ldots, m-1\}$. We usually want to minimize the tax rate, and thus the following optimization problem arises: $\min\{T(x) \mid -x \leqslant 0\}$, where $T(x)$ is a convex function and the set of constraints $\{x \in \mathbb{R} \mid -x \leqslant 0\}$ is a convex set, which is a convex programming, and its more general form is

$$\text{minimize} \quad f(x) \quad \text{s.t.} \quad c(x) \leqslant 0, \tag{1}$$

where $f, c : \mathbb{R}^n \to \mathbb{R}$ are convex function, which is generally not differentiable. Suppose the Slater bound norm of Problem (1) holds, i.e., there exists $x \in \mathbb{R}^n$ such that $c(x) < 0$. We also assume that there exists an oracle that for any given $x \in \mathbb{R}^n$ can compute the function value $f(x), c(x)$ and one subgradient of each function, i.e., $g_f(x) \in \partial f(x)$ and $g_c(x) \in \partial c(x)$.

In general, NSO problems are difficult to solve. Among the many solution methods, the well-known methods are the subgradient method, the cutting plane method, the analytic central tangent plane method, and the bundle method. Among them, the latter two are currently considered to be the most stable and trustworthy methods for solving NSO problems. The NSO problems with constraints are more complicated [4]. For general nonsmooth convex constrained problems, one approach is to equivalently solve the unconstrained optimization problem with an exact penalty function [7]. However, the application of penalty functions inevitably involves the problem of taking values of penalty parameters. If a large parameter value is required to guarantee the accuracy of a given penalty function, which will increase the difficulty in numerical computation. There are also some bundle methods that do not use the idea of penalty functions [12]. However, all descent steps in these methods are required to be feasible, including the initial point. We know that this requirement is likely to be as difficult as solving Problem (1). As a result, the entire computational burden of solving the problem increases substantially. Other work on bundle methods for solving constrained optimization problems can be found in the literature [5, 8, 9, 14]. One of them, [14] gives a dual stable bundle method resulting from combining the proximal bundle method and the horizontal bundle method, which combines the advantages of both, while proposing an effective stopping criterion. When the objective function is nonconvex, [5] constructed a reassignment bundle method. Under the condition of using only the approximate information of the objective and constraint functions, [8] gave a proximal bundle method, which requires the error between the approximate and real values to be fixed but can be unknown, and finally obtains the approximate optimal solution of the problem.

We give the improvement function related to Problem (1): for a given $x \in \mathbb{R}^n$, let

$$h_x(y) = \max\{f(y) - f(x), c(y)\}, \quad \forall y \in \mathbb{R}^n.$$

It is called the improvement function of Problem (1). $\overline{x}$ is the solution of Problem (1) when and only when $\bar{x}$ is the optimal solution of the unconstrained optimization problem $\min h_{\bar{x}}(\cdot)$, see Theorem 2.1 in the next section. As a theoretical tool in the convergence analysis of the bundle method, the use of $h_{\bar{x}}(\cdot)$ dates back to literature [11]. However, the method requires that the resulting sequence of descent steps is feasible. The piecewise linear model of the improvement function has been used in feasible direction methods for solving smooth problems [15], while these methods still require that the descent steps are feasible and the improvement function itself is not addressed in the algorithm. Infeasible bundle methods are less common, with the literature [6] and the "stage I–stage II" improvement of the feasible methods in the constrained bundle methods before [3]. In these methods, it is necessary to predict the values of the objective function and parameters, which is obviously not feasible for general functions. In this paper, the infeasible bundle method of Sagastizdbal and Solodov [16] and the Moreau-Yosida regularization idea [13] are employed, where the objective function changes with the generation of descent step, but the adjusted model is still an upper and lower approximation of the updated objective function.

The full paper is structured as follows: Section 2 performs the problem transformation and gives the upper and lower approximation models of the objective function and its related properties, where the objective function is constructed by using the bundle method idea. Section 3 gives the specific infeasible quasi-Newtonian bundle method for solving the convex constrained optimization problem. Section 4 presents the global convergence analysis of the given algorithm. Section 5 gives the convergence results of the BFGS bundle method. The last section is the conclusion.

In the whole text, the following standard notation is used: $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ denotes the inner product in Euclidean space and the corresponding norm denoted by $\|\cdot\|$, $x^+ = \max\{x, 0\}$, for a subset X of $\mathbb{R}^n$, conv X represents its convex hull and $\partial h(x)$ represents the sub-differential of the convex function $h(x)$ at the point x.

2 Bundle method

Theorem 2.1 [6] *Assuming that Problem* (1) *satisfies the Slater constraint specification, the following conclusions are equivalent*:

(1) $\bar{x}$ *is the solution of Problem* (1);

(2) $\min\{h_{\bar{x}}(y) | y \in \mathbb{R}^n\} = h_{\bar{x}}(\bar{x}) = 0$;

(3) $0 \in \partial h_{\bar{x}}(\bar{x})$.

In view of the conclusion of Theorem 2.1, for a given $x \in \mathbb{R}^n$, consider the following problem:

$$\min\{h_x(y) \mid y \in \mathbb{R}^n\}. \tag{2}$$

According to the Moreau-Yosida regularization conclusion, the above problem is equivalent to

$$\min\{H_x^M(x) \mid x \in \mathbb{R}^n\}, \tag{3}$$

where

$$H_x^M(x) = \min\left\{h_x(y) + \frac{1}{2}\|y - x\|_M^2 \mid y \in \mathbb{R}^n\right\}, \tag{4}$$

M is an $n \times n$ symmetric positive definite matrix, $\|\cdot\|_M = \langle \cdot, M \cdot \rangle$. (4) is called the Moreau-Yosida regularization of $h_x(y)$ associated with M. It is well known that $H_\cdot^M(\cdot)$ is a continuous differentiable convex function defined on $\mathbb{R}^n$, although $h_x(\cdot)$ may be non-differentiable. The derivative of $H_\cdot^M(\cdot)$ at point x is $G_x^M(x) := \bigtriangledown H_x^M(x) = M(x - p(x)) \in \partial h_x(p(x))$, where $p(x)$ is the unique optimal solution to Problem (4). Further, $G_x^M(x)$ is a global Lipschitz continuous function which module is $\|M\|$. $\overline{x}$ is a minimal point of $h_x(\cdot)$ when and only when $G_x^M(\bar{x}) = 0$ and $\overline{x} = p(\bar{x})$. Let $y = x + d$. Then (4) is equivalent to

$$H_x^M(x) = \min\left\{h_x(x + d) + \frac{1}{2}\|d\|_M^2 \mid d \in \mathbb{R}^n\right\}. \tag{5}$$

The following considers the approximation of $h_x(x + d)$ using the bundle method. Assume that the information is already available in the bundle $(e_{f_i}, e_{c_i}, g_f^i \in \partial_{e_{f_i}} f(x), g_c^i \in \partial_{e_{c_i}} c(x))$. The linearization errors of f_i and c_i are defined as $e_{f_i} = f(x) - f(y^i) - \langle g_f^i, x - y^i \rangle$ and $e_{c_i} = c(x) - c(y^i) - \langle g_c^i, x - y^i \rangle$, then bundle meets $B_i^{\text{oracle}} \subseteq \cup_{i<l}\{(f_i, c_i, e_{f_i}, e_{c_i}, g_f^i, g_c^i)\}$.

Lemma 2.1 [16] *For each $i \in B_i^{\text{oracle}}$, define*

$$e_i := e_{f_i} + c^+(x), \quad g_{h_x}^i := g_f^i, \quad \textit{if} \quad f(y^i) - f(x) \geqslant c(y^i),$$
$$e_i := e_{c_i} + c^+(x) - c(x), \quad g_{h_x}^i := g_c^i, \quad \textit{if} \quad f(y^i) - f(x) < c(y^i).$$

Then $e_i \geqslant 0, g_{h_x}^i \in \partial_{e_i} h_x(x)$.

From Lemma 2.1, we know that $h_x(y) \geqslant h_x(x) + \langle g_{h_x}^i, y - x \rangle - e_i$, $\forall y \in \mathbb{R}^n$. Let $y = x + d$. Then $h_x(x + d) \geqslant c^+(x) + \max\{\langle g_{h_x}^i, d\rangle - e_i \mid i \in B_l^{\text{oracle}}\} =: \check{h}_x(x + d)$. Furthermore, make

$$\begin{aligned}\check{H}_x^M(x) &= \min\left\{\check{h}_x(x + d) + \frac{1}{2}\|d\|_M^2 \mid d \in \mathbb{R}^n\right\} \\ &= c^+(x) + \min\left\{\max\left\{-e_i + \langle g_{h_x}^i, d\rangle \mid i \in B_l^{\text{oracle}}\right\} + \frac{1}{2}d^{\mathrm{T}} M d \mid d \in \mathbb{R}^n\right\}.\end{aligned} \tag{6}$$

Let $d(x)$ be the optimal solution of (6). Let $v(x) = \max\{-e_i + \langle g^i_{h_x}, d(x)\rangle \mid i \in B_i^{\text{oracle}}\}$. Then $\check{H}^M_x(x) = c^+(x) + v(x) + \frac{1}{2}d(x)^{\mathrm{T}}Md(x)$. Let $a(x) = x + d(x)$. It is an approximation of the real solution of (5). Then let $\hat{H}^M_x(x) = h_x(a(x)) + \frac{1}{2}d(x)^{\mathrm{T}}Md(x), x \in \mathbb{R}^n$. For the upper and lower approximation functions of $H^M_x(x)$, the following conclusion holds.

Lemma 2.2 $\forall x \in \mathbb{R}^n$:
(i) $\hat{H}^M_x(x) \geqslant H^M_x(x) \geqslant \check{H}^M_x(x)$;
(ii) $\hat{H}^M_x(x) = H^M_x(x) \Leftrightarrow a(x) = p(x)$.

Proof (i) Due to the fact that for $x \in \mathbb{R}^n, h_x(x+d) \geqslant \check{h}_x(x+d), \forall d \in \mathbb{R}^n$, then $H^M_x(x) \geqslant \check{H}^M_x(x)$. By the definition of $a(x)$ and the optimal solution $d(x)$ of (6), we have $h_x(a(x)) + \frac{1}{2}d(x)^{\mathrm{T}}Md(x) \geqslant \check{h}_x(a(x)) + \frac{1}{2}d(x)^{\mathrm{T}}Md(x)$. Thus, $\hat{H}^M_x(x) \geqslant H^M_x(x)$.

(ii) Apparently holds. □

Let $\varepsilon(x) = \hat{H}^M_x(x) - \check{H}^M_x(x)$. Then $\varepsilon(x) \geqslant 0$. Let C be a given positive number and $\delta(x)$ be a given positive number during each inner loop iteration. If

$$\varepsilon(x) \leqslant \delta(x) \min\left\{d(x)^{\mathrm{T}}Md(x), C\right\}, \tag{7}$$

then $a(x)$ is considered to be an approximation of $p(x)$. If (7) does not hold, let $y^{i+1} = x + d(x)$ and compute $e_{f_{i+1}}, e_{c_{i+1}}, f_{i+1}, c_{i+1}, g_f^{i+1}, g_c^{i+1}$, and then $e_{i+1}, g^{i+1}_{h_x}$. According to Lemma 2.1, add the information of the new trial point y^{i+1} to the bundle, construct a new approximate model for $h_x(x+d)$, and then resolve (6) to find the new $d(x)$, $\varepsilon(x)$, and check it with (7). If condition (7) never holds, using a similar proof in [13], the following two lemmas can be obtained.

Lemma 2.3 *Assuming that x is not an optimal solution for $h_x(\cdot)$, and* (7) *is never satisfied in the sub-algorithm, then $\varepsilon(x) \to 0$.*

Lemma 2.4 *Let $\tilde{G}^M_x(x) = M(x - a(x)) = -Md(x)$. Then*

$$\left\|G^M_x(x) - \tilde{G}^M_x(x)\right\|_{M^{-1}} = \|p(x) - a(x)\|_M \leqslant \sqrt{2\varepsilon(x)}, \tag{8}$$

$$\left\|G^M_x(x) - \tilde{G}^M_x(x)\right\| \leqslant \sqrt{2\varepsilon(x)\|M\|}. \tag{9}$$

The following lemma shows that each inner loop iteration must stop in a finite number of steps, thus ensuring the smooth execution of the entire algorithm.

Lemma 2.5 *If x is not an optimal solution of $h_x(\cdot)$, then after a finite number of computational iterations of the sub-problem* (6), *it is always possible to find a solution $d(x)$ of the sub-problem such that* (7) *holds.*

Proof Assuming that the conclusion does not hold, we will keep adding information of the new trial point y^{i+1} to the bundle, leading to $i \to \infty$. By Lemma 2.3, we have $\varepsilon(x) \to 0$. By Lemma 2.4, $\left\|G_x^M(x) - \tilde{G}_x^M(x)\right\| \to 0$. Since x is not an optimal solution for $h_x(\cdot)$, there exists a positive δ_0 and a positive integer i_0 such that $\left\|\tilde{G}_x^M(x)\right\| \geqslant \delta_0$. This holds for all $i \geqslant i_0$. By $d(x)^{\mathrm{T}} M d(x) = \left(\tilde{G}_x^M(x)\right)^{\mathrm{T}} M^{-1} \tilde{G}_x^M(x)$ and the fact that (7) does not hold, there is $\varepsilon(x) > \delta(x) \min\{\frac{\delta_0^2}{\|M\|}, N\}, i \geqslant i_0$, which contradicts $\varepsilon(x) \to 0$. □

For the decreasing step iteration point x^{k+1} generated by the algorithm, the following lemma ensures that the new model $\check{h}_{x^{k+1}}(\cdot)$ is still a valid lower approximation to the new objective function $h_{x^{k+1}}(\cdot)$ after the objective function is updated from $\check{h}_{x^k}(\cdot)$ to $\check{h}_{x^{k+1}}(\cdot)$ in (6), thus ensuring the next iteration to proceed smoothly.

Lemma 2.6 [16] Assume that the algorithm produces descent step iteration point x^{k+1}, its linearization error associated with x^{k+1} is defined as follows:

$$\begin{aligned} e_{f_i}^{k+1} &= e_{f_i}^k + f\left(x^{k+1}\right) - f\left(x^k\right) + \langle g_f^i, x^k - x^{k+1}\rangle, \\ e_{c_i}^{k+1} &= e_{c_i}^k + c\left(x^{k+1}\right) - c\left(x^k\right) + \langle g_c^i, x^k - x^{k+1}\rangle. \end{aligned}$$

Then $g_{h_{k+1}}^i \in \partial_{e_i^{k+1}} h_{k+1}(x^{k+1})$, where $e_i^{k+1} \geqslant 0$, and $g_{h_{k+1}}^i$ is obtained by converting x to the corresponding x^{k+1} in the result of Lemma 2.1.

3 Specific algorithms

Algorithm 3.1 Infeasible quasi-Newton bundle methods for convex constrained optimization problems.

Step 1 (Initial step) σ, ρ, C are given constants and $\sigma < \frac{1}{2}, \rho < 1, \{\delta_k\}_{k=0}^{\infty}$ is a sequence of positive numbers satisfying $\sum_{k=0}^{\infty} \delta_k < +\infty$. x^0 is the initial estimated solution, B_0 is an $n \times n$ symmetric positive definite matrix. Let $k = 0$, choose d^0, ε_0 such that $\varepsilon_0 \leqslant \delta_0 \min\{(d^0)^{\mathrm{T}} M d^0, C\}$. Start the iterative process of the sub-problem with $i = 1,\ u^1 = x^0$.

Step 2 (Compute the search direction) If $\left\|\tilde{G}_{x^k}^M(x^k)\right\| = 0$, the algorithm stops and x^k is an optimal solution. Otherwise, calculate

$$s^k = -B_k^{-1} \tilde{G}_{x^k}^M(x^k). \tag{10}$$

Step 3 (Line search) Starting from $l = 0$, let j_k be the smallest non-negative integer l such that

$$\check{H}_{x^k+\rho^l s^k}^M\left(x^k + \rho^l s^k\right) \leqslant \hat{H}_{x^k}^M\left(x^k\right) + \sigma \rho^l \left(s^k\right)^{\mathrm{T}} \tilde{G}_{x^k}^M\left(x^k\right), \tag{11}$$

where $\check{H}^M_{x^k+\rho^l s^k}\left(x^k+\rho^l s^k\right)$ satisfies

$$\begin{aligned}&\hat{H}^M_{x^k+\rho^l s^k}\left(x^k+\rho^l s^k\right)-\check{H}^M_{x^k+\rho^l s^k}\left(x^k+\rho^l s^k\right)\\&\quad\leqslant\delta_{k+1}\min\{(d(x^k+\rho^l s^k))^{\mathrm{T}}Md(x^k+\rho^l s^k),C\}.\end{aligned}\tag{12}$$

Setting $\tau_k:=\rho^{j_k}, x^{k+1}:=x^k+\tau_k s^k$.

Step 4 (Correct the quasi-Newton matrix) Let $\Delta x^k=x^{k+1}-x^k$, $\Delta y^k=\tilde{G}^M_{x^{k+1}}(x^{k+1})-\tilde{G}^M_{x^k}(x^k)$. If $(\Delta x^k)^{\mathrm{T}}\Delta y^k>0$. Correct B_k to B_{k+1}, so that B_{k+1} is a symmetric positive definite matrix and satisfies the quasi-Newton equation $B_{k+1}\Delta x^k=\Delta y^k$. Otherwise, set $B_k=M$. Let $k=k+1$, and go back to Step 2.

Note 1 Take not of the stopping criterion in Step 2. If $\|\tilde{G}^M_{x^k}(x^k)\|=0$, then we can infer that $d(x^k)=0$ from $\tilde{G}^M_{x^k}(x^k)=-Md(x^k)$, and from (7), we know that $\varepsilon(x^k)=0$. Then we get $G^M_{x^k}(x^k)=0$ and $x^k=p(x^k), 0\in\partial h_{x^k}(x^k)$ from (9). From Theorem 2.1, we know that x^k is the optimal solution to the original Problem (1).

Note 2 In Step 3, the process of finding j_k requires repeatedly solving the sub-problem (6) to find a descent step iteration point x^{k+1} that satisfies conditions (11) and (12). During this process, the objective function of the sub-problem (6) does not change (Step 1). Once j_k is found, Step 3 is skipped and the next iteration is performed. At this time, the objective function of (6) is changed (the so-called descent step), and the original approximate model $\check{h}_{x^k}(\cdot)$ for $h_{x^k}(\cdot)$ becomes the approximate model $\check{h}_{x^{k+1}}(\cdot)$ for $h_{x^{k+1}}(\cdot)$, i.e., not only the approximate model is changed, but the original objective function that is being approximated is also changed. According to Lemma 2.1 and Lemma 2.6, $\check{h}_{x^{k+1}}(\cdot)$ is still a valid approximation of $h_{x^{k+1}}(\cdot)$, which ensures the smooth iteration of the algorithm. This is different from the previous bundle methods, in which the objective function in the sub-problem of the bundle method is fixed and does not change with the generation of descent steps.

Note 3 As the iteration proceeds, the number of elements in the bundle B_l^{oracle} increases, and it is necessary to selectively delete and compress the bundle when the number of elements in the bundle becomes large. Let the optimal solution of the dual problem of Problem (6) be $\alpha\in\mathbb{R}^{|B_l^{\text{oracle}}|}$ with $\alpha_i\geqslant 0,\sum_{i\in B_l^{\text{oracle}}}\alpha_i=1$. It is well known that the optimal solution of Problem (6) can be expressed according to the optimal solution α of its dual problem. However, in this formulation, only the sub-gradient $g^i_{h_x}$ corresponding to $\alpha_i>0$ takes effect, and this is also in effect for the linearization error e_i. Thus, when $|B_l^{\text{oracle}}|=n^{\max}$, we perform the following selection deletion and aggregation into techniques, where $n^{\max}$ represents the upper limit of the number of elements that can be stored in the bundle:

— Select the point pair $(e_i, g^i_{h_x})$ corresponding to $\alpha_i = 0$, which can be removed directly from the bundle.

— If there are still many remaining point pairs corresponding to $\alpha_i > 0$, their information is compressed into a point pair in the form of a convex combination (with α_i as the convex combination factor). A new affine function is constructed using this point pair, which is inserted into the model. After which, the point pairs corresponding to $\alpha_i > 0$ can be arbitrarily selected until the number of elements in bundle B_l^{oracle} is less than $n^{\max}$. This technique can both preserves the information of the positive point pairs in the bundle and ensures that the number of elements in the bundle is within the specified range.

4 Global convergence analysis

The following theorem is a reasonable explanation for the existence of j_k in each iteration of the algorithm.

Theorem 4.1 *If x^k is not the minimal value point of $h_{x^k}(\cdot)$, then there exists $\bar{\tau}_k > 0$ such that*

$$\check{H}^M_{x^k+\tau s^k}\left(x^k+\tau s^k\right) \leqslant \hat{H}^M_{x^k}\left(x^k\right) + \sigma\tau\left(s^k\right)^{\mathrm{T}}\tilde{G}^M_{x^k}\left(x^k\right) \tag{13}$$

holds for all $\tau \in (0, \bar{\tau}_k)$, where $\check{H}^M_{x^k+\tau s^k}\left(x^k+\tau s^k\right)$ satisfies

$$\begin{aligned}&\hat{H}^M_{x^k+\tau s^k}\left(x^k+\tau s^k\right) - \check{H}^M_{x^k+\tau s^k}\left(x^k+\tau s^k\right)\\&\qquad\leqslant \delta_{k+1}\min\{(d(x^k+\tau s^k))^{\mathrm{T}}Md(x^k+\tau s^k), C\}.\end{aligned} \tag{14}$$

Proof Since x^k is not a minimal point of $h_{x^k}(\cdot)$, there exists a positive number $\tilde{\tau}_k > 0$ such that $\forall\tau \in (0, \tilde{\tau}]\,, x^k + \tau s^k$ is not a minimal point of $H^M_{\cdot}(\cdot)$ either. By Lemma 2.5, for each $\tau \in (0, \tilde{\tau}]$, there is always a $d(x^k + \tau s^k)$ such that (14) holds. If $\hat{H}^M_{x^k}(x^k) = H^M_{x^k}(x^k)$, by Lemma 2.2, we have $a(x^k) = p(x^k)$, so $\tilde{G}^M_{x^k}(x^k) = M(x^k - a(x^k)) = M(x^k - p(x^k)) = G^M_{x^k}(x^k)$. Since x^k is not an optimal solution, (10) implies that $(s^k)^{\mathrm{T}}G^M_{x^k}(x^k) < 0$. Also, since $H^M_{\cdot}(\cdot)$ is continuously differentiable and $\sigma < 1$, there is a positive number $\bar{\tau}_k, \bar{\tau}_k \leqslant \tilde{\tau}_k$ such that $\forall\tau \in (0, \bar{\tau}_k]$. While we have $H^M_{x^k+\tau s^k}\left(x^k+\tau s^k\right) \leqslant H^M_{x^k} + \sigma\tau\left(s^k\right)^{\mathrm{T}}G^M_{x^k}\left(x^k\right)$, which implies that (13) holds. If $\hat{H}^M_{x^k}(x^k) > H^M_{x^k}(x^k)$, then there also exists $\bar{\tau}_k > 0$ (sufficiently small) such that $\check{H}^M_{x^k+\tau s^k}\left(x^k+\tau s^k\right) \leqslant H^M_{x^k+\tau s^k}\left(x^k+\tau s^k\right) \xrightarrow{\tau\to 0} H^M_{x^k}\left(x^k\right) \leqslant H^M_{x^k}\left(x^k\right) + \frac{1}{2}\left(\hat{H}^M_{x^k}\left(x^k\right) - H^M_{x^k}\left(x^k\right)\right) \leqslant \hat{H}^M_{x^k}\left(x^k\right) + \sigma\tau\left(s^k\right)^{\mathrm{T}}\tilde{G}^M_{x^k}\left(x^k\right), \forall\tau \in (0, \bar{\tau}_k]$, and thus (13) also holds. □

Since it is assumed that $\sum_{k=0}^{\infty}\delta_k < +\infty$, there exists a constant $A > 0$ such that

$$\sum_{k=0}^{\infty} \delta_k \leqslant A. \tag{15}$$

Introducing the notation $D_k\{x \in \mathbb{R}^n \mid H_{x^k}^M(x) \leqslant H_{x^0}^M(x^0) + 2AC\}$, if we assume that both f and c are lower bounded functions, then $\{D_k\}_{k\geqslant 0}$ is bounded.

Lemma 4.1 *For all $k \geqslant 0$, we have*

$$H_{x^{k+1}}^M\left(x^{k+1}\right) \leqslant H_{x^k}^M\left(x^k\right) + C\left(\delta_k + \delta_{k+1}\right), \quad x^k \in D_k. \tag{16}$$

Proof According to the algorithm criterion, for all $k \geqslant 0$, there is

$$\begin{aligned} H_{x^{k+1}}^M\left(x^{k+1}\right) &\leqslant \check{H}_{x^{k+1}}^M\left(x^{k+1}\right) + C\delta_{k+1} \\ &\leqslant \hat{H}_{x^k}^M\left(x^k\right) + \sigma\rho^{j_k}\left(s^k\right)^{\mathrm{T}} \tilde{G}_{x^k}^M\left(x^k\right) + C\delta_{k+1} \\ &= \hat{H}_{x^k}^M\left(x^k\right) - \sigma\rho^{j_k}\tilde{G}_{x^k}^M\left(x^k\right)^{\mathrm{T}} B_k^{-1}\tilde{G}_{x^k}^M\left(x^k\right) + C\delta_{k+1} \\ &\leqslant \hat{H}_{x^k}^M\left(x^k\right) + C\delta_{k+1} \\ &\leqslant \check{H}_{x^k}^M\left(x^k\right) + C\delta_k + C\delta_{k+1}. \end{aligned}$$

Thus, for all $k \geqslant 0$, then $H_{x^{k+1}}^M(x^{k+1}) \leqslant H_{x^0}^M(x^0) + 2AC$. Since $x^0 \in D_0$, so $x^k \in D_k, \forall k \geqslant 0$. The conclusion is confirmed. □

The global convergence results of the algorithm are given below.

Theorem 4.2 *Assuming that f and c are lower bounded and there exist two positive constants a_1 and a_2 such that $\|B_k\| \leqslant a_1$ and $\|B_k^{-1}\| \leqslant a_2$ hold for all $k \geqslant 0$, the algorithm produces any cluster of the sequence $\{x^k\}$ that is a minimal value point for Problem* (1).

Proof By Lemma 4.1 and the assumption that f and c are lower bounded, combined with (15) and (16), there exists H_*^M such that

$$\lim_{k\to\infty} H_{x^k}^M(x^k) = H_*^M.$$

Since $\delta_k \to 0$, according to Lemma 2.1 and the algorithm criterion, we have $\varepsilon_k \to 0$ and $\lim_{k\to\infty} \hat{H}_{x^k}^M\left(x^k\right) = \lim_{k\to\infty} \check{H}_{x^k}^M\left(x^k\right) = H_*^M$. Thus, according to the assumptions on B_k,

$$\lim_{k\to\infty} \tau_k \|\tilde{G}_{x^k}^M(x^k)\|^2 = 0. \tag{17}$$

Let $\overline{x}$ be any cluster-point of $\{x^k\}$, and $\{x^k\}_{k\in K}$ be a sub-sequence converging to $\overline{x}$. According to Lemma 2.3,

$$\lim_{k\to\infty, k\in K} \tilde{G}_{x^k}^M(x^k) = G_{x^k}^M(\overline{x}). \tag{18}$$

If $\liminf_{k\to\infty,\, k\in K} \tau_k > 0$, then according to (17) and (18), we have $G_{\bar{x}}^M(\overline{x}) = 0$, which means that $\bar{x}$ is the optimal solution to the original problem. Otherwise, $\liminf_{k\to\infty,\, k\in K} \tau_k = 0$. Assuming that $\tau_k \to 0 (k \to \infty, k \in K)$, according to the line search criterion and Lemma 2.1, we have

$$\frac{H^M_{x^k+\rho^{j_k-1}s^k}\left(x^k+\rho^{j_k-1}s^k\right)-H^M_{x^k}\left(x^k\right)}{\rho^{j_k-1}}>\sigma\left(s^k\right)^{\mathrm{T}}\tilde{G}^M_{x^k}\left(x^k\right). \tag{19}$$

Then by (l8), $\{\tilde{G}^M_{x^k}(x^k)\}_{k\in K}$ is bounded and with the assumption that B_k is bounded, we get $\{s^k\}_{k\in K}$ is bounded. We may assume again $\lim_{k\in K,\,k\to\infty} s^k=\bar{s}$. Since $\rho^{j_k-1}\to 0, k\in K, k\to\infty$, we get $\bar{s}^{\mathrm{T}}G^M_{\bar{x}}(\bar{x})\geqslant\sigma\bar{s}^{\mathrm{T}}G^M_{\bar{x}}(\bar{x})$. By $\sigma<\frac{1}{2}$, $\bar{s}^{\mathrm{T}}G^M_{\bar{x}}(\bar{x})\geqslant 0$ and the assumption for B_k, $\bar{s}^{\mathrm{T}}G^M_{\bar{x}}(\bar{x})\leqslant-\frac{1}{c_2}\|\bar{s}\|^2$, which implies that $\bar{s}^{\mathrm{T}}G^M_{\bar{x}}(\bar{x})=0$ as well as $\bar{s}=0$, we finally get $G^M_{\bar{x}}(\bar{x})=0$. □

Note 4 We will discuss the assumption of boundedness of $\{B_k\}$ and $\{B_k^{-1}\}$ in Theorem 4.2 in the next section.

5 Convergence analysis of infeasible BFGS bundle methods

This section mainly considers the convergence of the infeasible BFGS bundle method.

The infeasible BFGS bundle method is obtained by taking $B_0=M$ in the initial step of Algorithm 3.1 and taking the positive sequence $\{\delta_k\}_{k=1}^{\infty}$ to satisfy $\sum_{k=0}^{\infty}\delta_k^{\frac{1}{3}}<\infty$. The selection of B_k is specified in Step 4, and the infeasible BFGS bundle method is obtained as follows.

Algorithm 5.1 Infeasible BFGS bundle method for convex constrained optimization problem.

Step 4 (Correction of the quasi-Newton matrix) Let $\Delta x^k=x^{k+1}-x^k$, $\Delta y^k=\tilde{G}^M_{x^{k+1}}(x^{k+1})-\tilde{G}^M_{x^k}(x^k)$. If there exist two constants $a_3>0$, $a_4\in(0,1)$ such that the iteration of the algorithm satisfies the following two conditions:

$$\left\|\Delta x^k\right\|_M\left(\sqrt{2\varepsilon_k}+\sqrt{2\varepsilon_{k+1}}\right)\leqslant a_3\left(\Delta x^k\right)^{\mathrm{T}}\Delta y^k, \tag{20}$$

$$2\left\|\Delta y^k\right\|_M\left(\sqrt{2\varepsilon_k}+\sqrt{2\varepsilon_{k+1}}\right)\leqslant\min\left\{a_4,\delta_k^{\frac{1}{3}}+\delta_{k+1}^{\frac{1}{3}}\right\}\left\|\Delta y^k\right\|^2. \tag{21}$$

Let $B_{k+1}=B_k-\frac{B_k\Delta x^k\left(\Delta x^k\right)^{\mathrm{T}}B_k}{\left(\Delta x^k\right)^{T}B_k\Delta x^k}+\frac{\Delta y^k\left(\Delta y^k\right)^{\mathrm{T}}}{\left(\Delta x^k\right)^{\mathrm{T}}\Delta y^k}$, otherwise make $B_{k+1}=M$. Let $k=k+1$, go back to Step 2.

From Step 4, if (20) holds, then $\left(\Delta x^k\right)^{\mathrm{T}}\Delta y^k>0$; if B_k is a positive definition matrix, B_{k+1} is also a positive definition matrix and satisfies the quasi-Newton equation

$$B_{k+1}\Delta x^k=B_k\Delta x^k-\frac{B_k\Delta x^k\left(\Delta x^k\right)^{\mathrm{T}}B_k\Delta x^k}{\left(\Delta x^k\right)^{\mathrm{T}}B_k\Delta x^k}+\frac{\Delta y^k\left(\Delta y^k\right)^{\mathrm{T}}\Delta x^k}{\left(\Delta x^k\right)^{\mathrm{T}}\Delta y^k}=\Delta y^k.$$

As we all know, if $h_x(\cdot)$ is a strongly convex function on $\mathbb{R}^n$, then $H_\cdot^M(\cdot)$ is also strongly convex on $\mathbb{R}^n$ [10]. Now assume that $H_\cdot^M(\cdot)$ is strongly convex on $\{D_k\}_{k\geqslant 0}$. Then $h_x(\cdot)$ has a unique minima $\bar{x}$ on $\{D_k\}_{k\geqslant 0}$. Introduce the label $K := \{0\} \cup \{i \mid (20) \text{ or } (21) \text{ does not hold for } k = i-1\} \equiv \{k_0, k_1, \ldots, k_j, \ldots\}$. The following theorem shows that the assumption of boundedness of $\{B_k\}$ and $\{B_k^{-1}\}$ in Theorem 4.2 holds under certain conditions.

Theorem 5.1 *Suppose that* $H_\cdot^M(\cdot)$ *is strongly convex on* $\{D_k\}_{k\geqslant 0}$. *The sequences* $\{x^k\}$ *and* $\{B_k\}$ *are generated by Algorithm* 5.1. *Suppose that* $\forall k \geqslant 0, \Delta x^k$ *and* Δy^k *satisfy*

$$\frac{\left\|\Delta \bar{y}^k - M_* \Delta x^k\right\|}{\|\Delta x^k\|} \leqslant \delta_k',$$

where $\Delta \bar{y}^k = G_{x^{k+1}}^M\left(x^{k+1}\right) - G_{x^k}^M\left(x^k\right), M_*$ *is some symmetric positive definite matrix and* $\{\delta_k'\}_{k=0}^{\infty}$ *is a sequence of positive numbers satisfying* $\sum_{k=0}^{\infty} \delta_k' < \infty$. *Then* $\|B_k\|$ *and* $\left\|B_k^{-1}\right\|$ *are bounded.*

Proof If (20) and (21) hold, we have

$$\begin{aligned}
\left(\Delta x^k\right)^{\mathrm{T}} \Delta y^k &\geqslant \left(\Delta x^k\right)^{\mathrm{T}} \Delta \bar{y}^k - \left\|\Delta x^k\right\|_M \left\|\Delta y^k - \Delta \bar{y}^k\right\|_{M^{-1}} \\
&\geqslant \left(\Delta x^k\right)^{\mathrm{T}} \Delta \bar{y}^k - \left\|\Delta x^k\right\|_M \left(\left\|\tilde{G}_{x^k}^M\left(x^k\right) - G_{x^k}^M\left(x^k\right)\right\|_{M^{-1}}\right. \\
&\quad \left. + \left\|\tilde{G}_{x^{k+1}}^M\left(x^{k+1}\right) - G_{x^{k+1}}^M\left(x^{k+1}\right)\right\|_{M^{-1}}\right) \\
&\geqslant \left(\Delta x^k\right)^{\mathrm{T}} \Delta \bar{y}^k - \left\|\Delta x^k\right\|_M \left(\sqrt{2\varepsilon_k} + \sqrt{2\varepsilon_{k+1}}\right).
\end{aligned}$$

After organizing, we get

$$\left(\Delta x^k\right)^{\mathrm{T}} \Delta y^k \geqslant \frac{1}{1+a_3} \left(\Delta x^k\right)^{\mathrm{T}} \Delta \bar{y}^k. \tag{22}$$

In addition,

$$\begin{aligned}
\left\|\Delta \bar{y}^k\right\|^2 &\geqslant \left\|\Delta y^k\right\|^2 - 2\left\|\Delta y^k\right\|_M \left\|\Delta \bar{y}^k - \Delta y^k\right\|_{M^{-1}} \\
&\geqslant \left\|\Delta y^k\right\|^2 - 2\left\|\Delta y^k\right\|_M \left(\sqrt{2\varepsilon_k} + \sqrt{2\varepsilon_{k+1}}\right).
\end{aligned}$$

We reorganized and get

$$\left\|\Delta \bar{y}^k\right\|^2 \geqslant (1-a_4)\left\|\Delta y^k\right\|^2. \tag{23}$$

For $i \in K$, we have $B_i = M$. The $i \notin K$, B_i is obtained by B_{i-1} using the BFGS correction formula. Assuming that K has an infinite number of elements, for $k_{j-1} \leqslant k < k_j - 1$, where for a certain $j \geqslant 1, k_{j-1}, k_j \in K$, then (20) and (21) hold, i.e.,

$$\left(\Delta x^k\right)^{\mathrm{T}} \Delta y^k \geqslant \frac{1}{1+a_3} \left(\Delta x^k\right)^{\mathrm{T}} \Delta \bar{y}^k > 0.$$

Then, by Lemma 2.4, (20) and (21), for all $k_{j-1} \leqslant k < k_j - 1$, we have

$$\begin{aligned}\frac{\left\|\Delta y^k - M_*\Delta x^k\right\|}{\|\Delta x^k\|} &\leqslant \delta_k' + \frac{\left\|\Delta y^k - \Delta \bar{y}^k\right\|}{\|\Delta x^k\|} \\ &\leqslant \delta_k' + \frac{\|M\|\left(\sqrt{2\varepsilon_k} + \sqrt{2\varepsilon_{k+1}}\right)}{\|\Delta x^k\|} \\ &\leqslant \delta_k' + \frac{\sqrt{\|M\|\,\|M^{-1}\|}}{2\sqrt{1-a_4}}\left(\delta_k^{\frac{1}{3}} + \delta_{k+1}^{\frac{1}{3}}\right)\frac{\left\|\Delta \bar{y}^k\right\|}{\|\Delta x^k\|}\cdot \\ &\leqslant \delta_k' + \frac{\sqrt{\|M\|^3\,\|M^{-1}\|}}{2\sqrt{1-a_4}}\left(\delta_k^{\frac{1}{3}} + \delta_{k+1}^{\frac{5}{2}}\right) \\ &=: \bar{\delta}_k.\end{aligned}$$

Then for all k, where k satisfying $k_{j-1} \leqslant k < k_j - 1$, there are

$$\frac{\left\|\Delta y^k - M_*\Delta x^k\right\|}{\|\Delta x^k\|} \leqslant \bar{\delta}_k.$$

By assumption $\sum_{k=0}^{\infty}\delta_k' < \infty, \sum_{k=0}^{\infty}\delta_k^{\frac{1}{3}} < \infty$, we have $\sum_{k=0}^{\infty}\bar{\delta}_k < \infty$. Then according to [2, Theorem 3.2], it follows that for all k satisfying $k_{j-1} \leqslant k < k_j - 1$, $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded depending on $B_{k_{j-1}}$. Since $B_{k_j} = M$ for all $j \geqslant 0$, then the sequences $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded. When the number of elements in K is finite, a similar proof method can be used to obtain that $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded. □

The following results show that without the assumption of boundedness of $\|B_k\|$ and $\|B_k^{-1}\|$, it is still possible to prove that the iterative sequence generated by the algorithm still converges to the optimal solution of Problem (1) for some special problems.

Theorem 5.2 *Suppose* $H_\cdot^M(\cdot)$ *is strongly convex on* $\{D_k\}_{k\geqslant 0}$, *and the sequence* $\left\{x^k\right\}, \{B_k\}$ *are generated by Algorithm* 5.1 *with* $x^k \neq \bar{x}$, $\forall k \geqslant 0$. *Then* $\left\{x^k\right\}$ *R-linearly converges to the unique solution* $\bar{x}$ *of Problem* (1) *further, with*

$$\sum_{k=0}^{\infty}\left\|x^k - \bar{x}\right\| < +\infty.$$

Proof The division K is both a finite set and an infinite set. In either case, we can use the strong convexity of $H_\cdot^M(\cdot)$, the global continuity of $G_\cdot^M(\cdot)$ modulo $\|M\|$, the positive definite correction criterion of B_k, Lemma 2.2, and $\sum_{k=0}^{\infty}\delta_k < \infty$ to obtain that there exists the constant $\theta \in (0,1)$, $\bar{N} \in (0,\infty)$ and a positive integer $\bar{k}$ such that for all $k \geqslant \bar{k}$, there is

$$H_{x^{k+1}}^M\left(x^{k+1}\right) - H_{\bar{x}}^M(\bar{x}) \leqslant \bar{N}\left(\theta^{\frac{1}{2}}\right)^{k-\bar{k}+1}\left(H_{x^k}^M\left(x^k\right) - H_{\bar{x}}^M(\bar{x})\right).$$

By the strong convexity of $H_\cdot^M(\cdot)$ and [1, Lemma 4.3], for all $k \geqslant 0$, there exists a constant $\alpha > 0$ such that

$$\frac{1}{2}\alpha \left\|x^k - \bar{x}\right\|^2 \leqslant H_{x^k}^M \left(x^k\right) - H_{\bar{x}}^M(\bar{x}) \leqslant \frac{\alpha}{2} \left\|G_{x^k}^M \left(x^k\right)\right\|^2 .$$

So, there is

$$\begin{aligned}\sum_{k=\bar{k}}^{\infty} \left\|x^k - \bar{x}\right\|^2 &\leqslant \left(\frac{2}{\alpha}\right)^{\frac{1}{2}} \sum_{k=\bar{k}}^{\infty} \left(H_{x^k}^M \left(x^k\right) - H_{\bar{x}}^M(\bar{x})\right)^{\frac{1}{2}} \\ &\leqslant \left[\frac{2\bar{N}\left(H_{x^k}^M \left(x^k\right) - H_{\bar{x}}^M(\bar{x})\right)}{\alpha}\right]^{\frac{1}{2}} \sum_{k=\bar{k}}^{\infty} \left(\theta^{\frac{1}{4}}\right)^{k-\bar{k}} \\ &\leqslant \infty.\end{aligned}$$

The conclusion is proofed. □

6 Conclusions

In this paper, we combine the infeasible bundle method of Sagastizdbal with the Moreau-Yosida regularization idea to construct a quasi-Newton bundle method for solving convex constrained optimization problems. The global convergence result of the algorithm is derived without guaranteeing that every iteration point is feasible. Although the method is essentially similar to the unconstrained bundle method, it differs from the previous methods in that the objective function of the subproblem to be solved at each iteration changes with the generation of descent steps, which requires necessary adjustments to ensure the convergence of the algorithm. For the special BFGS bundle method, we briefly discuss the conditions that guarantee the boundedness of the quasi-Newton matrix.

References

1. Bonnans J F, Gilbert J Ch, Lemaréchal C, Sagastizábal C A. A family of variable metric proximal methods. Math Program, 1995, 68(1/2/3): 15–47
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(3): 727–739
3. Fletcher R, Leyffer S. A bundle filter method for nonsmooth nonlinear optimization. Numerical Analysis Report NA/195, Dundee: Department of Mathematics, The University of Dundee, 1999
4. Frangioni A. Generalized bundle methods. SIAM J Optim, 2002, 13(1): 117–156
5. Hare W, Sagastizábal C A. A redistributed proximal bundle method for nonconvex optimization. SIAM J Optim, 2010, 20(5): 2442–2473
6. Kiwiel K C. Methods of Descent for Nondifferentiable Optimization. Lecture Notes in Math, Vol 1133. Berlin: Springer-Verlag, 1985
7. Kiwiel K C. Exact penalty functions in proximal bundle methods for constrained convex nondifferentiable minimization. Math Program, 1991, 52(1/2/3): 285–302
8. Kiwiel K C. A proximal bundle method with approximate subgradient linearizations. SIAM J Optim, 2006, 16(4): 1007–1023

9. Kiwiel K C. A proximal-projection bundle method for Lagrangian relaxation, including semidefinite programming. SIAM J Optim, 2006, 17(4): 1015–1034
10. Lemaréchal C, Sagastizábal C A. Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries. SIAM J Optim, 1997, 7(2): 367–385
11. Miffin R. An algorithm for constrained optimization with semismooth functions. Math Oper Res, 1977, 2(2): 191–207
12. Mifflin R. A modification and an extension of Lemarechal's algorithm for nonsmooth minimization. In: Nondifferential and Variational Techniques in Optimization, Math Program Stud, Vol 17. Berlin: Springer-Verlag, 1982, 77–90
13. Miflin R, Sun D F, Qi L Q. Quasi-Newton bundle type methods for nondifferentiable convex optimization. SIAM J Optim, 1998, 8(2): 583–603
14. Oliveira W, Solodov M. A doubly stabilized bundle method for nonsmooth convex optimization. 2013
15. Pironneau O, Polak E. Rate of convergence of a class of methods of feasible directions. SIAM J Numer Anal, 1973, 10(1): 161–174
16. Sagastizabal C A, Solodov M. An infeasible bundle method for nonsmooth convex constrained optimization without a penalty function or a filter. SIAM J Optim, 2005, 16(1): 146–169