SIAM J. OPTIM.
Vol. 8, No. 2, pp. 583–603, May 1998

015

QUASI-NEWTON BUNDLE-TYPE METHODS FOR NONDIFFERENTIABLE CONVEX OPTIMIZATION*

ROBERT MIFFLIN†, DEFENG SUN‡, AND LIQUN QI‡

Abstract. In this paper we provide implementable methods for solving nondifferentiable convex optimization problems. A typical method minimizes an approximate Moreau–Yosida regularization using a quasi-Newton technique with inexact function and gradient values which are generated by a finite inner bundle algorithm. For a BFGS bundle-type method global and superlinear convergence results for the outer iteration sequence are obtained.

Key words. Moreau–Yosida regularization, bundle method, quasi-Newton method, superlinear convergence

AMS subject classifications. 65K05, 90C30, 52A41, 90C25

PII. S1052623496303329

1. Introduction.

Consider the following minimization problem:

$$\min_{x\in\Re^n} f(x), \tag{1.1}$$

where $f : \Re^n \to \Re$ is a possibly nondifferentiable convex function.

Throughout this paper, we use $\|\cdot\|$ to denote the Euclidean vector norm on $\Re^n$ or its induced matrix norm on $\Re^{n\times n}$. Let M be a symmetric positive definite $n \times n$ matrix. For any $x \in \Re^n$ let

$$\|x\|_M^2 = x^T Mx.$$

We let F_M be the Moreau–Yosida [19, 27] regularization of f, associated with M, defined by

$$F_M(x) = \min_{y\in\Re^n} \left\{ f(y) + \frac{1}{2}\|y - x\|_M^2 \right\}. \tag{1.2}$$

It is well known that F_M is a continuously differentiable convex function defined on $\Re^n$ even though f may be nondifferentiable. The derivative of F_M at x is defined by

$$G_M(x) \equiv \nabla F_M(x) = M(x - p(x)) \in \partial f(p(x)),$$

where $p(x)$ is the unique minimizer in (1.2) and ∂f is the subdifferential mapping of f [25]. Here, $p(x)$ is called the proximal point of x. Furthermore, G_M is globally Lipschitz continuous with modulus $\|M\|$, the set of minimizers of (1.1) is exactly the set of minimizers of

$$\min_{x\in\Re^n} F_M(x), \tag{1.3}$$

*Received by the editors May 10, 1996; accepted for publication (in revised form) February 24, 1997. This research was supported by the Australian Research Council.

http://www.siam.org/journals/siopt/8-2/30332.html

†Department of Pure and Applied Mathematics, Washington State University, Pullman, WA 99164-3113 (mifflin@beta.math.wsu.edu). The research of this author was supported by the National Science Foundation grants DMS-9402018 and DMS-9703952.

‡School of Mathematics, University of New South Wales, Sydney 2052, Australia (sun@maths.unsw.edu.au, qi@maths.unsw.edu.au). This research was performed while Defeng Sun was on leave from the Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China.

and x^* minimizes f if and only if $G_M(x^*) = 0$ and $p(x^*) = x^*$. For additional properties, see [26, 17, 23].

In this paper we use Moreau–Yosida regularization, bundle and quasi-Newton ideas to develop a convergent minimization method for f. We do not assume that the subproblem in (1.2) is solved exactly at each outer iteration nor do we assume f is differentiable at a solution x^*. For a particular BFGS bundle method applied to an approximation of F_M we obtain global and superlinear convergence (of outer iterations) if $\nabla G_M(x^*)$ is positive definite and the directional derivative of G_M is radially Lipschitz continuous at x^*. Related work on this subject appears in [1, 5, 6, 7, 12, 16, 18]. In particular, in [1] global and superlinear convergence results for a BFGS proximal method are given by assuming that f is continuously differentiable and $p(x)$ is computed exactly. In the literature, for example [10], global convergence of particular quasi-Newton methods with inexact gradient values has been discussed. In this paper we approximate F_M in addition to G_M and these two approximations are related.

The plan of this paper is as follows. In section 2 we discuss how a bundle method can be used to satisfy our requirement for approximating $p(x)$. We give the quasi-Newton bundle-type algorithm in section 3 and discuss its global convergence in section 4. In section 5 we discuss global and superlinear convergence of a BFGS bundle-type method. Some concluding remarks are given in section 6.

2. The bundle concept.

The bundle idea plays a central role in approximating $F_M(x)$ and $\nabla F_M(x)$ as is developed in [16] and [18], for example. Let $d = y - x$ in (1.2) and minimize over d instead of y. This gives

$$F_M(x) = \min_{d \in \Re^n} \left\{ f(x+d) + \frac{1}{2} d^T M d \right\}.$$

Now we consider approximating $f(x+d)$ by a polyhedral function

$$\check{f}(x+d) = \max_{i=1,\ldots,j} \{ f(u^i) + (g^i)^T (x + d - u^i) \},$$

where the data $(u^i, f(u^i), g^i)$ with $g^i = g(u^i) \in \partial f(u^i)$ constitute a bundle generated sequentially starting from x and $g(x) \in \partial f(x)$ and, possibly, a subset of the previous set used to generate x. Since f is convex, we have

$$f(x+d) \geq \check{f}(x+d). \tag{2.1}$$

If we define a linearization error by letting

$$e(x, u^i) = f(x) - f(u^i) - (g^i)^T (x - u^i),$$

then $\check{f}(x+d)$ can be written as

$$\check{f}(x+d) = f(x) + \max_{i=1,\ldots,j} \{ (g^i)^T d - e(x, u^i) \}. \tag{2.2}$$

Let

$$\check{F}_M(x) = \min_{d \in \Re^n} \left\{ \check{f}(x+d) + \frac{1}{2} d^T M d \right\}$$

$$= f(x) + \min_{d \in \Re^n} \left\{ \max_{i=1,\ldots,j} \{ (g^i)^T d - e(x, u^i) \} + \frac{1}{2} d^T M d \right\}. \tag{2.3}$$

From (2.1) and the definition of $F_M(x)$, we have

$$\check{F}_M(x) \le F_M(x).$$

So $\check{F}_M(x)$ is an underapproximation of the unknown value $F_M(x)$. Let $d(x)$ solve the minimization problem in (2.3), and let

$$v(x) = \max_{i=1,\ldots,j} \{(g^i)^T d(x) - e(x, u^i)\}.$$

Then

$$\check{F}_M(x) = f(x) + v(x) + \frac{1}{2}(d(x))^T M d(x).$$

Let $a(x) = x + d(x)$ be an approximation of $p(x)$, and let

$$\hat{F}_M(x) = f(a(x)) + \frac{1}{2}(d(x))^T M d(x).$$

Since $p(x)$ is the unique minimizer in (1.2), we have

$$F_M(x) \le \hat{F}_M(x)$$

and equality holds if and only if $a(x) = p(x)$.

Thus, we have the following lemma.

LEMMA 2.1.

(i) $\check{F}_M(x) \le F_M(x) \le \hat{F}_M(x)$.

(ii) $F_M(x) = \hat{F}_M(x)$ *if and only if* $a(x) = p(x)$.

This simple lemma plays an important role in the design of our algorithm.

Let

$$\varepsilon(x) = \hat{F}_M(x) - \check{F}_M(x). \tag{2.4}$$

We base our rule for accepting $a(x)$ as an approximation of $p(x)$ on $\varepsilon(x)$ as follows: Accept if

$$\varepsilon(x) \le \delta(x) \min\{(d(x))^T M d(x), N\}, \tag{2.5}$$

where $\delta(x)$ and N are given positive numbers and $\delta(x)$ is fixed during the bundling process. If (2.5) is not satisfied then we let $u^{j+1} = x + d(x)$ and $g^{j+1} = g(u^{j+1})$, append a new piece $(g^{j+1})^T d - e(x, u^{j+1})$ to (2.2), replace j by $j+1$, and solve a new subproblem in (2.3) for a new $d(x)$ and a new $\varepsilon(x)$ to be tested in (2.5). If this process, in which $\varepsilon(x)$ and $d(x)$ vary, does not terminate we have the following result.

LEMMA 2.2. *Suppose x does not minimize f. In this subalgorithm, if* (2.5) *is never satisfied, then*

$$\varepsilon(x) \to 0.$$

Proof. Following the proof of Proposition 3 in [11] (see also [15] and [8]), we can prove that

$$\check{F}_M(x) \to F_M(x) \text{ and } \hat{F}_M(x) \to F_M(x) \text{ as } j \to \infty.$$

So the result of this lemma follows from (2.4). ∎

Let

$$\tilde{G}_M(x) = M(x - a(x)) = -Md(x).$$

The following result is a slight extension of Lemma 1 in [12]. For completeness, we give the proof.

LEMMA 2.3.

$$\|G_M(x) - \tilde{G}_M(x)\|_{M^{-1}} = \|p(x) - a(x)\|_M \leq \sqrt{2\varepsilon(x)}, \tag{2.6}$$

$$\|G_M(x) - \tilde{G}_M(x)\| \leq \sqrt{2\varepsilon(x)\|M\|}. \tag{2.7}$$

Proof. Define the function $\psi : \Re^n \to \Re$ by

$$\psi(z) = f(z) + \frac{1}{2}\|z - x\|_M^2.$$

Since f is convex and $\|z - x\|_M^2$ is a strongly convex quadratic function in z, we have the inequality

$$\psi(u) \geq \psi(z) + \omega^T(u - z) + \frac{1}{2}\|u - z\|_M^2 \quad \text{for all } u, z \in \Re^n \text{ and all } \omega \in \partial\psi(z). \tag{2.8}$$

Since $p(x)$ is the argmin in (1.2), $0 \in \partial\psi(p(x))$. Letting $u = a(x)$, $z = p(x)$, and $\omega = 0$ in (2.8) gives

$$\psi(a(x)) \geq \psi(p(x)) + \frac{1}{2}\|a(x) - p(x)\|_M^2;$$

i.e.,

$$\hat{F}_M(x) \geq F_M(x) + \frac{1}{2}\|a(x) - p(x)\|_M^2.$$

Then, from Lemma 2.1 and (2.4), (2.6) holds. Finally, we have

$$\|G_M(x) - \tilde{G}_M(x)\|^2 = \|M(p(x) - a(x))\|^2 \leq \|M\|\|p(x) - a(x)\|_M^2,$$

which when combined with (2.6) implies that (2.7) holds. ◻

LEMMA 2.4. *If x does not minimize f, then after a finite number of subproblem steps we can find a subproblem solution $d(x)$ such that* (2.5) *holds.*

Proof. If not, then $j \to \infty$, so from Lemma 2.2,

$$\varepsilon(x) \to 0.$$

Then, from Lemma 2.3, $\|\tilde{G}_M(x) - G_M(x)\| \to 0$. Since x is not an optimal solution, $G_M(x) \neq 0$. So there exists a positive number δ_0 such that $\|\tilde{G}_M(x)\| \geq \delta_0$ when j is sufficiently large. Then, since

$$(d(x))^T Md(x) = (\tilde{G}_M(x))^T M^{-1}\tilde{G}_M(x) \tag{2.9}$$

and (2.5) is not satisfied,

$$\varepsilon(x) > \delta(x) \min\left\{\frac{\delta_0^2}{\|M\|}, N\right\}$$

for all j sufficiently large. This is a contradiction, because $\varepsilon(x) \to 0$ when $j \to \infty$. ◻

Lemma 2.4 says that a bundle-type algorithm can be used to find a vector $d(x)$ such that (2.5) holds if x is not an optimal solution. This is essential for our algorithm.

A practical stopping test for the overall algorithm is to stop if the subalgorithm generates a solution with

$$|v(x)| \le \text{tol}, \tag{2.10}$$

where tol is a small positive input parameter. See, for example, Theorem 1 in [18].

3. The algorithm. Since F_M is a convex function and G_M is globally Lipschitz continuous, a natural idea is to use a quasi-Newton method, such as the BFGS method, to solve (1.3). The severe practical difficulty with this approach is that we cannot expect to calculate $F_M(x)$ and $G_M(x)$ exactly. To approximate these values appropriately the results of section 2 will be useful.

We use the notation $\varepsilon_k = \varepsilon(x^k)$, $a^k = a(x^k)$, $d^k = d(x^k)$ and so on.

Quasi-Newton bundle-type algorithm.

Step 0 (initialization). Let σ, ρ, and N be positive numbers such that $\sigma < 1/2$ and $\rho < 1$. Let $\{\delta_k\}$ be a sequence of positive numbers such that $\sum_{k=0}^{\infty} \delta_k < +\infty$. Let $x^0 \in \Re^n$ be an initial solution estimate and B_0 be an $n \times n$ symmetric positive definite matrix. Set $k := 0$ and find d^0 and ε_0 as described in section 2 such that

$$\varepsilon_0 \le \delta_0 \min\{(d^0)^T M d^0, N\},$$

for example starting the bundle process with $j = 1$ and $u^1 = x^0$.

Step 1 (compute a search direction). If $\|\tilde{G}(x^k)\| = 0$, stop with x^k optimal. Else, compute

$$s^k = -B_k^{-1} \tilde{G}_M(x^k). \tag{3.1}$$

Step 2 (line search). Starting with $m = 0$, let i_k be the smallest nonnegative integer m such that

$$\check{F}_M(x^k + \rho^m s^k) \le \hat{F}_M(x^k) + \sigma \rho^m (s^k)^T \tilde{G}_M(x^k), \tag{3.2}$$

where $\check{F}_M(x^k + \rho^m s^k)$ is an underapproximation of F_M at $x^k + \rho^m s^k$ and satisfies

$$\begin{aligned} &\hat{F}_M(x^k + \rho^m s^k) - \check{F}_M(x^k + \rho^m s^k) \\ &\quad \le \delta_{k+1} \min\{(d(x^k + \rho^m s^k))^T M d(x^k + \rho^m s^k), N\}. \end{aligned} \tag{3.3}$$

Set $\tau_k := \rho^{i_k}$ and $x^{k+1} := x^k + \tau_k s^k$.

Step 3 (update the quasi-Newton matrix). Let $\Delta x^k = x^{k+1} - x^k$ and $\Delta y^k = \tilde{G}_M(x^{k+1}) - \tilde{G}_M(x^k)$. If $(\Delta x^k)^T \Delta y^k > 0$, update B_k to B_{k+1} such that B_{k+1} is symmetric and positive definite and satisfies quasi-Newton equation

$$B_{k+1} \Delta x^k = \Delta y^k;$$

otherwise set $B_{k+1} := M$. Set $k := k + 1$ and go to step 1.

At Step 1 if $\|\tilde{G}_M(x^k)\| = 0$ then, from the definition of $\tilde{G}_M(x)$, $\|d(x^k)\| = 0$ and then, from (2.5), $\varepsilon(x^k) = 0$, so (2.7) implies $G_M(x^k) = 0$ and x^k is optimal.

From the discussion given in section 2, if $x^k + \rho^m s^k$ does not minimize f, we can find a vector $d(x^k + \rho^m s^k)$ satisfying (3.3) after a finite number of subproblem steps. So Step 2 proceeds as follows: First compute $d(x^k + \rho^m s^k)$ to satisfy (3.3) and then check if (3.2) is satisfied. If this is not the case, increase m by 1 and repeat with the new point $x^k+\rho^m s^k$; otherwise set $\tau_k=\rho^{i_k}$ and $x^{k+1}=x^k+\tau_k s^k$ and go to Step 3. If for some candidate nonnegative integer m used in Step 2 $x^k+\rho^m s^k$ is an optimal solution and if tol in (2.10) is zero, then the corresponding bundle subalgorithm execution may not terminate. Throughout the sequel we assume that this situation does not occur by assuming that each subalgorithm execution terminates. The next theorem shows that i_k is well defined at each iteration of the algorithm.

THEOREM 3.1. *If x^k does not minimize f, then there exists a number $\bar\tau_k>0$ such that*

$$\check F_M(x^k+\tau s^k)\le \hat F_M(x^k)+\sigma\tau(s^k)^T\tilde G_M(x^k) \tag{3.4}$$

holds for all $\tau\in(0,\bar\tau_k]$, where $\check F_M(x^k+\tau s^k)$, the underapproximation of F_M at $x^k+\tau s^k$, satisfies

$$\hat F_M(x^k+\tau s^k)-\check F_M(x^k+\tau s^k)\le \delta_{k+1}\min\{(d(x+\tau s^k))^TMd(x+\tau s^k),N\}. \tag{3.5}$$

Proof. Since x^k does not minimize f, there exists a positive number $\tilde\tau_k$ such that for any $\tau\in(0,\tilde\tau_k]$, $x^k+\tau s^k$ also does not minimize f. Then by Lemma 2.4 for each $\tau\in(0,\tilde\tau_k]$ we can find $d(x^k+\tau s^k)$ such that (3.5) holds. Next we prove this lemma by considering the following two cases.

Case 3.1. $\hat F_M(x^k)=F_M(x^k)$. Then from Lemma 2.1, we have

$$a(x^k)=p(x^k).$$

Then,

$$\tilde G_M(x^k)=M(x^k-a(x^k))=M(x^k-p(x^k))=G_M(x^k),$$

and, since x^k is not a solution, (3.1) implies

$$(s^k)^TG_M(x^k)<0.$$

Since F_M is continuously differentiable and $\sigma<1$, there exists a number $\bar\tau_k>0(\bar\tau_k\le\tilde\tau_k)$ such that for all $\tau\in(0,\bar\tau_k]$ we have

$$F_M(x^k+\tau s^k)\le F_M(x^k)+\sigma\tau(s^k)^TG_M(x^k).$$

This implies that (3.4) holds, because, by Lemma 2.1, $\check F_M(x^k+\tau s^k)\le F_M(x^k+\tau s^k)$.

Case 3.2. $\hat F_M(x^k)>F_M(x^k)$. Then when τ is sufficiently small, the right-hand side of (3.4) is greater than $F_M(x^k)+\frac12(\hat F_M(x^k)-F_M(x^k))$ and, as $\tau\to0$, the left-hand side satisfies

$$\check F_M(x^k+\tau s^k)\le F_M(x^k+\tau s^k)\to F_M(x^k).$$

So there exists a positive number $\bar\tau_k$ such that (3.4) is satisfied in this case, too. ∎

4. Global convergence. Throughout the rest of the paper we assume that the algorithm does not terminate so that $\{x^k\}$ is an infinite sequence.

Since $\sum_{k=0}^\infty\delta_k<\infty$, there exists a constant C such that

$$\sum_{k=0}^\infty\delta_k\le C. \tag{4.1}$$

Let

$$D = \{x \in \Re^n | F_M(x) \le F_M(x^0) + 2NC\}.$$

LEMMA 4.1. *For all $k \ge 0$ we have*

$$F_M(x^{k+1}) \le F_M(x^k) + N(\delta_k + \delta_{k+1}) \tag{4.2}$$

and

$$x^k \in D.$$

Proof. By Lemma 2.1 and the algorithm rules, for $k \ge 0$

$$\begin{aligned} F_M(x^{k+1}) &\le \check{F}_M(x^{k+1}) + N\delta_{k+1} \\ &\le \hat{F}_M(x^k) + \sigma\rho^{i_k}(s^k)^T\tilde{G}_M(x^k) + N\delta_{k+1} \\ &= \hat{F}_M(x^k) - \sigma\rho^{i_k}\tilde{G}_M(x^k)^T B_k^{-1}\tilde{G}_M(x^k) + N\delta_{k+1} \\ &\le \hat{F}_M(x^k) + N\delta_{k+1} \\ &\le F_M(x^k) + N(\delta_k + \delta_{k+1}). \end{aligned}$$

Thus, for all $k \ge 0$, (4.2) holds and

$$x^{k+1} \in D.$$

The proof is completed by noting that $x^0 \in D$. □

THEOREM 4.2. *Suppose that f is bounded from below and there exist two positive numbers c_1 and c_2 such that $\|B_k\| \le c_1$ and $\|B_k^{-1}\| \le c_2$ for all k. Then any accumulation point of $\{x^k\}$ minimizes f.*

Proof. From Lemma 4.1 we know that $F_M(x^k)$ is bounded from above. On the other hand, since f is assumed to be bounded from below, F_M is also bounded from below. Suppose that $\liminf_{k\to\infty} F_M(x^k) = F_M^*$. Then, by (4.1), (4.2), and a simple $\epsilon - \delta$ argument, we have $\lim_{k\to\infty} F_M(x^k) = F_M^*$.

Since $\{\delta_k\} \to 0$, from Lemma 2.1 and the algorithm rules we have $\{\varepsilon_k\} \to 0$ and

$$\lim_{k\to\infty} \check{F}_M(x^k) = \lim_{k\to\infty} \hat{F}_M(x^k) = F_M^*.$$

Thus,

$$\lim_{k\to\infty} \tau_k (s^k)^T \tilde{G}_M(x^k) = 0,$$

which, from the assumption on $\{B_k\}$, implies that

$$\lim_{k\to\infty} \tau_k \|\tilde{G}_M(x^k)\|^2 = 0. \tag{4.3}$$

Let $\bar{x}$ be an arbitrary accumulation point of $\{x^k\}$, and let $\{x^k\}_{k\in K}$ be a subsequence converging to $\bar{x}$. By Lemma 2.3

$$\lim_{k\to\infty, k\in K} \tilde{G}_M(x^k) = G_M(\bar{x}). \tag{4.4}$$

If $\liminf_{k\to\infty, k\in K} \tau_k > 0$, then from (4.3) and (4.4) we have

$$G_M(\bar{x}) = 0.$$

On the other hand, if $\liminf_{k\to\infty, k\in K} \tau_k = 0$, then by taking a subsequence, if necessary, we can assume that $\tau_k \to 0$ for $k \in K$. From the line search stopping rule we have

$$\check{F}_M(x^k + \rho^{i_k-1}s^k) > \hat{F}_M(x^k) + \sigma\rho^{i_k-1}(s^k)^T\tilde{G}_M(x^k),$$

where $\rho^{i_k-1} = \tau_k/\rho$. So, by Lemma 2.1, we have

$$F_M(x^k + \rho^{i_k-1}s^k) > F_M(x^k) + \sigma\rho^{i_k-1}(s^k)^T\tilde{G}_M(x^k);$$

i.e.,

$$\frac{F_M(x^k + \rho^{i_k-1}s^k) - F_M(x^k)}{\rho^{i_k-1}} > \sigma(s^k)^T\tilde{G}_M(x^k). \tag{4.5}$$

By (4.4), $\{\tilde{G}_M(x^k)\}_{k\in K}$ is bounded. This, together with the assumption on $\{B_k\}$, implies that $\{s^k\}_{k\in K}$ is bounded. So, by taking a subsequence if necessary, we may assume that

$$\lim_{k\to\infty, k\in K} s^k = \bar{s}.$$

Since $\{\rho^{i_k-1}\}_{k\in K} \to 0$, by taking a limit in (4.5) on the subsequence $k \in K$, we obtain

$$\bar{s}^T G_M(\bar{x}) \geq \sigma\bar{s}^T G_M(\bar{x}). \tag{4.6}$$

Also, from the assumption on $\{B_k\}$ we have

$$\bar{s}^T G_M(\bar{x}) \leq -\frac{1}{c_2}\|\bar{s}\|^2,$$

which, combined with (4.6) and the fact that $\sigma < 1$, implies that

$$\bar{s}^T G_M(\bar{x}) = 0 \quad \text{and} \quad \bar{s} = 0.$$

Finally, this combined with the assumption on $\{B_k\}$ implies

$$G_M(\bar{x}) = 0.$$

This completes the proof. ∎

Based on the results established in [14] and [24], we could discuss local convergence of the proposed quasi-Newton bundle-type methods as in [1] by assuming that the initial point x^0 is sufficiently close to a solution x^* and the initial matrix B_0 is sufficiently close to $\nabla G_M(x^*)$. However, it should be noted that we only use an approximation of the proximal point while in [1] the exact value is used. Here we will not give such a discussion on the local convergence of the proposed methods. In the next section, we will discuss a BFGS bundle-type method for which global and superlinear convergence results are obtained.

5. A BFGS bundle-type method. For given vectors Δx and Δy, the BFGS quasi-Newton update of an $n \times n$ symmetric matrix B is the matrix

$$BFGS(B, \Delta x, \Delta y) := B - \frac{B\Delta x \Delta x^T B}{\Delta x^T B \Delta x} + \frac{\Delta y \Delta y^T}{\Delta x^T \Delta y}$$

(see [9] for instance). If B is positive definite and $\Delta x^T \Delta y > 0$, then the symmetric matrix $B^+ = BFGS(B, \Delta x, \Delta y)$ is also positive definite.

In our BFGS bundle-type method, we will assume that $B_0 = M$ and $\sum_{k=0}^{\infty} \delta_k^{1/3} < \infty$. Let

$$\Delta \bar{y}^k = G_M(x^{k+1}) - G_M(x^k).$$

At each iteration, if the following two conditions are satisfied, we will update B_k to $B_{k+1} = BFGS(B_k, \Delta x^k, \Delta y^k)$; otherwise, we let $B_{k+1} := M$. Given $c_3 \in (0, \infty)$ and $c_4 \in (0, 1)$, these two conditions are

$$\|\Delta x^k\|_M(\sqrt{2\varepsilon_k} + \sqrt{2\varepsilon_{k+1}}) \le c_3 (\Delta x^k)^T \Delta y^k \tag{5.1}$$

and

$$2\|\Delta y^k\|_M(\sqrt{2\varepsilon_k} + \sqrt{2\varepsilon_{k+1}}) \le \min\{c_4, \delta_k^{1/3} + \delta_{k+1}^{1/3}\}\|\Delta y^k\|^2. \tag{5.2}$$

In order to employ BFGS results from [3] we need the following results.

LEMMA 5.1. *If conditions* (5.1) *and* (5.2) *are satisfied for some* $k \ge 0$*, then*

$$(\Delta x^k)^T \Delta y^k \ge (1/(1 + c_3))(\Delta x^k)^T \Delta \bar{y}^k \text{ and } \|\Delta \bar{y}^k\|^2 \ge (1 - c_4)\|\Delta y^k\|^2. \tag{5.3}$$

Proof. From (2.6) in Lemma 2.3,

$$\begin{aligned} (\Delta x^k)^T \Delta y^k &= (\Delta x^k)^T \Delta \bar{y}^k + (\Delta x^k)^T (\Delta y^k - \Delta \bar{y}^k) \\ &\ge (\Delta x^k)^T \Delta \bar{y}^k - \|\Delta x^k\|_M \|\Delta y^k - \Delta \bar{y}^k\|_{M^{-1}} \\ &\ge (\Delta x^k)^T \Delta \bar{y}^k - \|\Delta x^k\|_M (\|\tilde{G}_M(x^k) - G_M(x^k)\|_{M^{-1}} \\ &\quad + \|\tilde{G}_M(x^{k+1}) - G_M(x^{k+1})\|_{M^{-1}}) \\ &\ge (\Delta x^k)^T \Delta \bar{y}^k - \|\Delta x^k\|_M (\sqrt{2\varepsilon_k} + \sqrt{2\varepsilon_{k+1}}) \end{aligned}$$

and

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

So, if conditions (5.1) and (5.2) are satisfied, then (5.3) holds. This competes the proof. ∎

We denote the cosine of the angle between $B_k \Delta x^k$ and Δx^k by

$$\cos\theta_k := \frac{(\Delta x^k)^T B_k \Delta x^k}{\|\Delta x^k\| \|B_k \Delta x^k\|}$$

and the corresponding Rayleigh quotient by

$$q_k := \frac{(\Delta x^k)^T B_k \Delta x^k}{(\Delta x^k)^T \Delta x^k}.$$

Let

$$K := \{0\} \cup \{j|(5.1) \text{ or } (5.2) \text{ does not hold for } k = j-1\} \equiv \{k_0, k_1, ..., k_i, ...\}.$$

This implies that $B_j = M$ for $j \in K$ and B_j is a BFGS update of B_{j-1} for $j \notin K$. Also, let $\lceil \cdot \rceil$ be the roundup operator such that $\lceil t \rceil = i$, when $i - 1 < t \leq i$ for $i \in \{1, 2, ...\}$.

LEMMA 5.2. *Let $\{B_k\}$ be generated by the BFGS bundle-type algorithm. Suppose that there exist numbers $\alpha_1 > 0$ and $\alpha_2 > 0$ such that*

$$(\Delta x^k)^T \Delta \bar{y}^k \geq \alpha_1 \|\Delta x^k\|^2 \text{ and } (\Delta x^k)^T \Delta \bar{y}^k \geq \alpha_2 \|\Delta \bar{y}^k\|^2 \tag{5.4}$$

for all $k \geq 0$. Then for any $w \in (0,1)$ there exist constants $\beta_1, \beta_2, \beta_3 > 0$ 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 relations

$$\cos\theta_j \geq \beta_1,$$

$$\beta_2 \leq q_j \leq \beta_3,$$

$$\beta_2 \leq \frac{\|B_j \Delta x^j\|}{\|\Delta x^j\|} \leq \frac{\beta_3}{\beta_1}$$

hold for at least $\lceil w(k - k_{i-1} + 1) \rceil$ values of j satisfying $k_{i-1} \leq j \leq k$.

Proof. For any k satisfying $k_{i-1} \leq k < k_i - 1$, (5.1) and (5.2) hold. Then, from (5.3) and (5.4),

$$(\Delta x^k)^T \Delta y^k \geq \bar{\alpha}_1 \|\Delta x^k\|^2 \text{ and } (\Delta x^k)^T \Delta y^k \geq \bar{\alpha}_2 \|\Delta y^k\|^2$$

hold for all k satisfying $k_{i-1} \leq k < k_i - 1$, where $\bar{\alpha}_1 = \alpha_1/(1 + c_3)$ and $\bar{\alpha}_2 = \alpha_2(1 - c_4)/(1 + c_3)$. Then the results of this lemma follow from the proof of Theorem 2.1 in [3]. □

LEMMA 5.3. *For any nonnegative sequence $\{\delta_k\}_{k \geq 0}$, if $\sum_{k=0}^{\infty} \delta_k < \infty$, then*

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

Proof. This result follows easily from the properties of logarithms. □

LEMMA 5.4. *Relative to the line search there exist positive constants η_1 and η_2 such that either*

$$\begin{aligned} \check{F}_M(x^k + \tau_k s^k) \quad &\leq \hat{F}_M(x^k) - \eta_1 \frac{((s^k)^T \tilde{G}_M(x^k))^2}{\|s^k\|^2} \\ &-\eta_1/(1-\sigma) \frac{(s^k)^T (G_M(x^k) - \tilde{G}_M(x^k))((s^k)^T \tilde{G}_M(x^k))}{\|s^k\|^2} \end{aligned} \tag{5.5}$$

or

$$\check{F}_M(x^k + \tau_k s^k) \leq \hat{F}_M(x^k) + \eta_2 (s^k)^T \tilde{G}_M(x^k). \tag{5.6}$$

Proof. If (3.2) is satisfied by the integer $m = 0$, then (5.6) holds with $\eta_2 \equiv \sigma$. Suppose that $i_k > 0$, which means that (3.2) fails to be satisfied for $m := i_k - 1$; i.e.,

$$\check{F}_M(x^k + (\tau_k/\rho)s^k) > \hat{F}_M(x^k) + \sigma(\tau_k/\rho)(s^k)^T \tilde{G}_M(x^k),$$

which together with Lemma 2.1 implies that

$$F_M(x^k + (\tau_k/\rho)s^k) > F_M(x^k) + \sigma(\tau_k/\rho)(s^k)^T \tilde{G}_M(x^k).$$

Then, using the mean value theorem, we obtain

$$(\tau_k/\rho)(s^k)^T G_M(x^k + \theta(\tau_k/\rho)s^k) > \sigma(\tau_k/\rho)(s^k)^T \tilde{G}_M(x^k),$$

where $\theta \in (0, 1)$. Thus, from the Lipschitz continuity of G_M,

$$\begin{aligned}
&(\tau_k/\rho)(\sigma(s^k)^T \tilde{G}_M(x^k) - (s^k)^T G_M(x^k)) \\
&\quad < (\tau_k/\rho)(s^k)^T (G_M(x^k + \theta(\tau_k/\rho)s^k) - G_M(x^k)) \\
&\quad \le \|M\|((\tau_k/\rho)\|s^k\|)^2,
\end{aligned}$$

which implies that

$$\tau_k > \rho \frac{-((s^k)^T G_M(x^k) - \sigma(s^k)^T \tilde{G}_M(x^k))}{\|M\|\|s^k\|^2}.$$

Substituting this into (3.2) gives

$$\check{F}_M(x^k + \tau_k s^k) \le \hat{F}_M(x^k) - \frac{\rho\sigma}{\|M\|} \frac{((s^k)^T G_M(x^k) - \sigma(s^k)^T \tilde{G}_M(x^k))((s^k)^T \tilde{G}_M(x^k))}{\|s^k\|^2},$$

which gives (5.5) with $\eta_1 = \frac{\rho\sigma(1-\sigma)}{\|M\|}$. □

It was proved in [17] that f is strongly convex on $\Re^n$ if and only if F_M is strongly convex on $\Re^n$. From now on we assume that F_M is strongly convex on D. Then there exists an $\alpha > 0$ such that

$$F_M(z) \ge F_M(x) + G_M(x)^T(z - x) + \frac{\alpha}{2}\|z - x\|^2 \quad \text{for all } x, z \in D,$$

$$(G_M(z) - G_M(x))^T(z - x) \ge \alpha\|z - x\|^2 \quad \text{for all } x, z \in D.$$

This implies that there is a unique minimizer of f in D and that D is bounded. Let $\bar{x}$ be the unique solution. The next result gives R-linear convergence of $\{x^k\}$ to $\bar{x}$.

THEOREM 5.5. *Suppose that F_M is strongly convex on D and $\{B_k\}$ is generated by the BFGS bundle-type method and $x^k \ne \bar{x}$ for all $k \ge 0$. Then $\{x^k\}$ converges to the unique solution $\bar{x}$; moreover,*

$$\sum_{k=0}^{\infty} \|x^k - \bar{x}\| < \infty \tag{5.7}$$

and there are constants $r \in [0, 1)$ and $\bar{C} \in (0, \infty)$ and a positive integer $\bar{k}$ such that for all $k \ge \bar{k}$ we have

$$F_M(x^{k+1}) - F_M(\bar{x}) \le \bar{C}(r^{1/2})^{k-\bar{k}+1}(F_M(x^{\bar{k}}) - F_M(\bar{x})).$$

Proof. First suppose that K has an infinite number of elements. Since F_M is strongly convex on D and G_M is globally Lipschitz continuous, from [21] or Theorem X.4.2.2 of [13], (5.4) holds for $\alpha_1 = \alpha$ and $\alpha_2 = 1/\|M\|$. So, given $w \in (0,1)$, from Lemma 5.2 there exist constants $\beta, \beta' > 0$ such that for any k satisfying $k_{i-1} \le k < k_i - 1$, where $k_{i-1}, k_i \in K$ for some $i \ge 1$, the inequalities

$$\cos\theta_j \ge \beta \tag{5.8}$$

and

$$\frac{\|B_j \Delta x^j\|}{\|\Delta x^j\|} \le \beta' \tag{5.9}$$

hold for at least $\lceil w(k - k_{i-1} + 1)\rceil$ values of j satisfying $k_{i-1} \le j \le k$. Since $B_j = M$ if $j \in K$, we can assume β and β' are such that (5.8) and (5.9) hold for all $j \in K$. We define I to be the set of indices j for which (5.8) and (5.9) hold. Since D is bounded, $\{\|G_M(x^k)\|\}$ is a bounded sequence. From (2.7), (3.3), and (2.9), $\|G_M(x^k) - \tilde{G}_M(x^k)\| = o(\|\tilde{G}_M(x^k)\|)$, so there exists an integer $\bar{k}$ such that for all $k \ge \bar{k}$

$$2\|G_M(x^k)\| \ge \|\tilde{G}_M(x^k)\| \ge \frac{1}{2}\|G_M(x^k)\| \tag{5.10}$$

and

$$\left| -\frac{(s^k)^T(G_M(x^k) - \tilde{G}_M(x^k))((s^k)^T\tilde{G}_M(x^k))}{\|s^k\|^2}\right| \le \frac{(1-\sigma)\beta^2}{2}\|\tilde{G}_M(x^k)\|^2. \tag{5.11}$$

Consider an iterate x^j with $j \in I$ and $j \ge \bar{k}$. From Lemma 5.4, (5.8), (5.9), and (5.11), we have that

$$\hat{F}_M(x^j) - \check{F}_M(x^j + \tau_j s^j) \ge \eta\|\tilde{G}_M(x^j)\|^2, \tag{5.12}$$

where $\eta = \frac{1}{2}\eta_1\beta^2$ if (5.5) holds or $\eta = \eta_2\beta/\beta'$ if (5.6) holds. So, from (5.12) and (5.10), for all $j \in I$ and $j \ge \bar{k}$,

$$\hat{F}_M(x^j) - \check{F}_M(x^j + \tau_j s^j) \ge \frac{\eta}{4}\|G_M(x^j)\|^2. \tag{5.13}$$

By strong convexity of F_M and Lemma 4.3 in [1], for all $k \ge 0$,

$$\frac{1}{2}\alpha\|x^k - \bar{x}\|^2 \le F_M(x^k) - F_M(\bar{x}) \le \frac{2}{\alpha}\|G_M(x^k)\|^2. \tag{5.14}$$

Then, from Lemma 2.1, (5.13), and the right-side inequality in (5.14), for all $j \in I$ and $j \ge \bar{k}$,

$$F_M(x^{j+1}) - F_M(\bar{x}) - \varepsilon_{j+1} \le \left(1 - \frac{\eta\alpha}{8}\right)(F_M(x^j) - F_M(\bar{x})) + \varepsilon_j. \tag{5.15}$$

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

$$\frac{16\delta_k\|M^{-1}\|\|M\|^2}{\alpha} \le \min\left\{1, \frac{\eta\alpha}{8}\right\}. \tag{5.16}$$

By (3.3), (2.9), (5.10), the fact that $G_M(\bar{x}) = 0$, the Lipschitz continuity of G_M with modulus $\|M\|$, and (5.14), for all $k \geq \bar{k}$ we have

$$\begin{aligned} \varepsilon_k &\leq \delta_k \|M^{-1}\| \|\tilde{G}_M(x^k)\|^2 \\ &\leq 4\delta_k \|M^{-1}\| \|G_M(x^k)\|^2 \\ &\leq 4\delta_k \|M^{-1}\| \|M\|^2 \|x^k - \bar{x}\|^2 \\ &\leq \frac{8\delta_k \|M^{-1}\| \|M\|^2}{\alpha} (F_M(x^k) - F_M(\bar{x})). \end{aligned} \tag{5.17}$$

Then from (5.15)–(5.17), for all $j \in I$ and $j \geq \bar{k}$, we have

$$\begin{aligned} &\left(1 - 8\frac{\delta_{j+1}\|M^{-1}\|\|M\|^2}{\alpha}\right)(F_M(x^{j+1}) - F_M(\bar{x})) \\ &\leq \left(1 - \frac{1}{16}\eta\alpha\right)(F_M(x^j) - F_M(\bar{x})). \end{aligned} \tag{5.18}$$

Since $F_M(x^k) > F_M(\bar{x})$ for all k, (5.18) and (5.16) 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 (5.18)

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

From (3.1), (3.2), the positivity of σ and τ_k, and the positive definiteness of B_k we have

$$\check{F}_M(x^{k+1}) < \hat{F}_M(x^k) \quad \text{for all } k.$$

Combining this with (5.17) and Lemma 2.1 yields for all $j \geq \bar{k}$

$$\begin{aligned} &\left(1 - 8\frac{\delta_{j+1}\|M^{-1}\|\|M\|^2}{\alpha}\right)(F_M(x^{j+1}) - F_M(\bar{x})) \\ &\leq \left(1 + 8\frac{\delta_j\|M^{-1}\|\|M\|^2}{\alpha}\right)(F_M(x^j) - F_M(\bar{x})). \end{aligned}$$

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

$$\delta'_k = \frac{1 + 8\frac{\delta_k\|M^{-1}\|\|M\|^2}{\alpha}}{1 - 8\frac{\delta_{k+1}\|M^{-1}\|\|M\|^2}{\alpha}}.$$

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

$$\frac{r^{1/w}}{1 - 8\frac{\delta_{j+1}\|M^{-1}\|\|M\|^2}{\alpha}} < \delta'_j r^{1/w} \quad \text{for all } j \geq \bar{k}$$

and

$$r^{1/w} < r < r^{1/2},$$

we have for k satisfying $k_{i-1} \le k < k_i$,

$$
\begin{aligned}
F_M(x^{k+1}) - F_M(\bar{x}) &\le \prod_{j=k_{i-1}}^{k} \delta'_j r(F_M(x^{k_{i-1}}) - F_M(\bar{x})) \\
&\le \prod_{j=k_{i-1}}^{k} \delta'_j (r^{1/2})^{k-k_{i-1}+1}(F_M(x^{k_{i-1}}) - F_M(\bar{x})). \qquad (5.19)
\end{aligned}
$$

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

$$
F_M(x^{k+1}) - F_M(\bar{x}) \le \prod_{j=k_{i-1}}^{k} \delta'_j r^{k-k_{i-1}+1}(F_M(x^{k_{i-1}}) - F_M(\bar{x})). \qquad (5.20)
$$

Therefore,

$$
\begin{aligned}
F_M(x^{k_i}) - F_M(\bar{x}) &\le \delta'_{k_i-1}(F_M(x^{k_i-1}) - F_M(\bar{x})) \\
&\le \prod_{j=k_{i-1}}^{k_i-1} \delta'_j r^{k_i-k_{i-1}-1}(F_M(x^{k_{i-1}}) - F_M(\bar{x})) \\
&\le \prod_{j=k_{i-1}}^{k_i-1} \delta'_j (r^{1/2})^{k_i-k_{i-1}+1}(F_M(x^{k_{i-1}}) - F_M(\bar{x})). \qquad (5.21)
\end{aligned}
$$

So, from (5.19)–(5.21), for all k satisfying $k_{i-1} \le k < k_i$ we have

$$
F_M(x^{k+1}) - F_M(\bar{x}) \le \prod_{j=k_{i-1}}^{k} \delta'_j (r^{1/2})^{k-k_{i-1}+1}(F_M(x^{k_{i-1}}) - F_M(\bar{x})). \qquad (5.22)
$$

Without loss of generality, we can assume that $\bar{k} \in K$. Then, from (5.22), for any $k \ge \bar{k}$ we have

$$
F_M(x^{k+1}) - F_M(\bar{x}) \le \prod_{j=\bar{k}}^{k} \delta'_j (r^{1/2})^{k-\bar{k}+1}(F_M(x^{\bar{k}}) - F_M(\bar{x})).
$$

Since $\sum_{k=0}^{\infty} \delta_k < \infty$, $\sum_{k=\bar{k}}^{\infty}(\delta'_k - 1) < \infty$. So, from Lemma 5.3, there exists a constant $\bar{C} > 0$ such that

$$
\prod_{k=\bar{k}}^{\infty} \delta'_k \le \bar{C}.
$$

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

$$
F_M(x^{k+1}) - F_M(\bar{x}) \le \bar{C}(r^{1/2})^{k-\bar{k}+1}(F_M(x^{\bar{k}}) - F_M(\bar{x})). \qquad (5.23)
$$

Using (5.14), (5.23), and the fact that $r < 1$, we have

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

If there are only finitely many elements in K, then by following the above proof we can prove the same results as in the case where there are infinitely many elements in K. ◻

In the next lemma we discuss the boundedness of $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ which was assumed for convergence in Theorem 4.2.

LEMMA 5.6. *Suppose that F_M is strongly convex on D and $\{B_k\}$ is generated by the BFGS bundle-type method. Furthermore, assume that $\{\Delta x^k\}$ and $\{\Delta \bar{y}^k\}$ are such that for all $k \geq 0$*

$$\frac{\|\Delta \bar{y}^k - H_* \Delta x^k\|}{\|\Delta x^k\|} \leq \varepsilon'_k$$

for some symmetric positive definite matrix H_* and for some sequence $\{\varepsilon'_k\}$ with the property that $\sum_{k=0}^{\infty} \varepsilon'_k < \infty$. Then the sequences $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are bounded.

Proof. First suppose K has an infinite number of elements. For k satisfying $k_{i-1} \leq k < k_i - 1$, where $k_{i-1}, k_i \in K$ for some $i \geq 1$, (5.1) and (5.2) hold, and by (5.3)

$$(\Delta x^k)^T \Delta y^k \geq \frac{1}{(1 + c_3)} (\Delta x^k)^T \Delta \bar{y}^k > 0.$$

From Lemma 2.3, (5.2), and (5.3), for all k satisfying $k_{i-1} \leq k < k_i - 1$,

$$\begin{aligned}\frac{\|\Delta y^k - H_* \Delta x^k\|}{\|\Delta x^k\|} &\leq \varepsilon'_k + \frac{\|\Delta y^k - \Delta \bar{y}^k\|}{\|\Delta x^k\|} \\ &\leq \varepsilon'_k + \frac{\sqrt{2\varepsilon_k \|M\|} + \sqrt{2\varepsilon_{k+1}\|M\|}}{\|\Delta x^k\|} \\ &\leq \varepsilon'_k + \frac{\sqrt{\|M\|\|M^{-1}\|}}{2\sqrt{1 - c_4}} (\delta_k^{1/3} + \delta_{k+1}^{1/3}) \frac{\|\Delta \bar{y}^k\|}{\|\Delta x^k\|} \\ &\leq \varepsilon'_k + \frac{\sqrt{\|M\|^3 \|M^{-1}\|}}{2\sqrt{1 - c_4}} (\delta_k^{1/3} + \delta_{k+1}^{1/3}).\end{aligned}$$

Let $\bar{\varepsilon}_k = \varepsilon'_k + \frac{\sqrt{\|M\|^3\|M^{-1}\|}}{2\sqrt{1-c_4}} (\delta_k^{1/3} + \delta_{k+1}^{1/3})$. Then for all k satisfying $k_{i-1} \leq k < k_i - 1$, we have

$$\frac{\|\Delta y^k - H_* \Delta x^k\|}{\|\Delta x^k\|} \leq \bar{\varepsilon}_k.$$

From the assumptions that $\sum_{k=0}^{\infty} \varepsilon'_k < \infty$ and $\sum_{k=0}^{\infty} \delta_k^{1/3} < \infty$, it follows that

$$\sum_{k=0}^{\infty} \bar{\varepsilon}_k < \infty.$$

Then, from the proof of Theorem 3.2 in [3], it follows that for all k satisfying $k_{i-1} \leq k < k_i$, $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded with the bound depending on $B_{k_{i-1}}$. Finally, since $B_{k_i} = M$ for all $i \geq 0$, the entire sequences $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are bounded. The proof is completed by noting that the case where K has a finite number of elements follows in a similar manner from Theorem 3.2 in [3]. ◻

LEMMA 5.7. *Suppose that F_M is strongly convex on D and B_k is generated by the BFGS bundle-type method. If the sequences $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are bounded, then conditions* (5.1) *and* (5.2) *are satisfied for all sufficiently large k, and*

$$\|x^k - \bar{x}\| = O(\|\Delta x^k\|), \quad \|x^{k+1} - \bar{x}\| = O(\|\Delta x^k\|). \tag{5.24}$$

Proof. We first prove that τ_k is bounded away from zero. From the proof of Lemma 5.4, we have

$$\tau_k \geq \min\left\{1, \rho \frac{-((s^k)^T G_M(x^k) - \sigma(s^k)^T \tilde{G}_M(x^k))}{\|M\|\|s^k\|^2}\right\}.$$

But, since $\|G_M(x^k) - \tilde{G}_M(x^k)\| = o(\|\tilde{G}_M(x^k)\|)$, $\tilde{G}_M(x^k) = -B_k s^k$, the sequences $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are bounded, and $\sigma < 1$, it is not difficult to prove that there exists an integer $\bar{k}$ and a positive constant $\bar{\tau}$ such that for any $k \geq \bar{k}$

$$\tau_k \geq \bar{\tau}.$$

Thus, for all k, τ_k is bounded away from zero.

Since $\Delta x^k = x^{k+1} - x^k = \tau_k s^k = -\tau_k B_k^{-1} \tilde{G}_M(x^k)$, this bound on τ_k and the boundedness of $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ imply that

$$\|\tilde{G}_M(x^k)\| = O(\|\Delta x^k\|).$$

Then, by the strong convexity of F_M,

$$\|\tilde{G}_M(x^k)\| \geq \|G_M(x^k)\| - \|G_M(x^k) - \tilde{G}_M(x^k)\| \geq \alpha\|x^k - \bar{x}\| - o(\|\Delta x^k\|);$$

so, the first equality of (5.24) holds. Since $x^{k+1} = x^k + \tau_k s^k$, the first equality of (5.24) and the boundedness of $\{\|B_k^{-1}\|\}$ imply that the second equality of (5.24) holds. From Lemma 2.3, the first inequality in (5.17), and (5.24), we have

$$\begin{aligned} \|\Delta y^k - \Delta \bar{y}^k\|_{M^{-1}} &\leq \sqrt{2\varepsilon_k} + \sqrt{2\varepsilon_{k+1}} \\ &\leq \sqrt{2\delta_k\|M^{-1}\|}\|\tilde{G}_M(x^k)\| + \sqrt{2\delta_{k+1}\|M^{-1}\|}\|\tilde{G}_M(x^{k+1})\| \\ &= \sqrt{2\delta_k\|M^{-1}\|}O(\|G_M(x^k)\|) + \sqrt{2\delta_{k+1}\|M^{-1}\|}O(\|G_M(x^{k+1})\|) \\ &\leq (\sqrt{2\delta_k\|M^{-1}\|} + \sqrt{2\delta_{k+1}\|M^{-1}\|})O(\|\Delta x^k\|). \end{aligned} \tag{5.25}$$

Therefore, by strong convexity and (5.25),

$$\begin{aligned}(\Delta x^k)^T \Delta y^k &\geq (\Delta x^k)^T \Delta \bar{y}^k - \|\Delta x^k\|_M \|\Delta y^k - \Delta \bar{y}^k\|_{M^{-1}} \\ &\geq \alpha \|\Delta x^k\|^2 - o(\|\Delta x^k\|^2) \end{aligned} \tag{5.26}$$

and

$$\begin{aligned}\|\Delta y^k\| \geq & \ \|\Delta \bar{y}^k\| - \|\Delta y^k - \Delta \bar{y}^k\| \\ & \geq \alpha \|\Delta x^k\| - o(\|\Delta x^k\|). \end{aligned} \tag{5.27}$$

Then the third inequality in (5.17), (5.24), (5.26), (5.27), and the fact that $\{\delta_k^{1/2}/\delta_k^{1/3}\} \to 0$ imply that the update conditions (5.1) and (5.2) are satisfied for all sufficiently large k. ◻

Remark 5.1. A principal contribution of this paper is the update or reset tests (5.1) and (5.2) depending on ε_k and $\delta_k^{1/3}$. From the proof of Lemma 5.7 it can be seen that $\delta_k^{1/3} + \delta_{k+1}^{1/3}$ in (5.2) could be replaced by $\delta_k^{\gamma} + \delta_{k+1}^{\gamma}$, where $\gamma < 1/2$ if $\{\delta_k\}$ is chosen such that $\sum_{k=0}^{\infty} \delta_k^{\gamma} < \infty$.

THEOREM 5.8. *Suppose that all the assumptions in Lemma* 5.6 *hold. Then* $\{\|B_k\|\}$ *and* $\{\|B_k^{-1}\|\}$ *are bounded,* K *has finitely many elements,* (5.7) *holds, and*

$$\lim_{k\to\infty} \frac{\|(B_k - H_*)\Delta x^k\|}{\|\Delta x^k\|} = 0. \tag{5.28}$$

Proof. The first three results follow from Lemmas 5.6 and 5.7, the definition of K, and Theorem 5.5. So we can assume that there exists an integer $\bar{k}$ such that for any $k \geq \bar{k}$, conditions (5.1) and (5.2) are satisfied. As in the proof of Theorem 5.5, we know that (5.4) holds with $\alpha_1 = \alpha$ and $\alpha_2 = 1/\|M\|$. So, for any $k \geq \bar{k}$, we have

$$(\Delta x^k)^T \Delta y^k \geq \bar{\alpha}_1 \|\Delta x^k\|^2 \text{ and } (\Delta x^k)^T \Delta y^k \geq \bar{\alpha}_2 \|\Delta y^k\|^2,$$

where $\bar{\alpha}_1 = \alpha_1/(1 + c_3)$ and $\bar{\alpha}_2 = \alpha_2(1 - c_4)/(1 + c_3)$. As in the proof of Lemma 5.6, by letting $\bar{\varepsilon}_k = \varepsilon'_k + \frac{\sqrt{\|M\|^3\|M^{-1}\|}}{2\sqrt{1-c_4}}(\delta_k^{1/3} + \delta_{k+1}^{1/3})$, we obtain

$$\sum_{k=0}^{\infty} \bar{\varepsilon}_k < \infty$$

and for all $k \geq \bar{k}$

$$\frac{\|\Delta y^k - H_* \Delta x^k\|}{\|\Delta x^k\|} \leq \bar{\varepsilon}_k.$$

Then (5.28) follows from the proof of Theorem 3.2 in [3]. ◻

In order to obtain superlinear convergence for the BFGS bundle-type method, we need further assumptions on G_M. From now on we will assume that G_M is Fréchet differentiable at $\bar{x}$, which, together with assuming that F_M is strongly convex, implies that $\nabla G_M(\bar{x})$ is positive definite and, hence, invertible.

COROLLARY 5.9. *Suppose that* F_M *is strongly convex on* D *and* G_M *is Fréchet differentiable at* $\bar{x}$*. If there exists a constant* $L > 0$ *such that*

$$\frac{\|\Delta \bar{y}^k - \nabla G_M(\bar{x})\Delta x^k\|}{\|\Delta x^k\|} \leq L \max\{\|x^{k+1} - \bar{x}\|, \|x^k - \bar{x}\|\}, \tag{5.29}$$

then the sequence $\{x^k\}$ *generated by the BFGS bundle-type method satisfies*

$$\lim_{k\to\infty} \frac{\|(B_k - \nabla G_M(\bar{x}))\Delta x^k\|}{\|\Delta x^k\|} = 0. \tag{5.30}$$

Moreover, the sequences $\{\|B_k\|\}$ *and* $\{\|B_k^{-1}\|\}$ *are bounded.*

Proof. By using Theorems 5.5 and 5.8, (5.29), and Lemma 5.6 with $H_* = \nabla G_M(\bar{x})$ we obtain the results. ∎

Recall that a Lipschitz continuous function $H : \Re^n \to \Re^n$ is said to be directionally differentiable of degree 2 at x if

$$H(x + d) - H(x) - H'(x; d) = O(\|d\|^2),$$

where $H'(x; d)$ is the directional derivative of H at x in the direction d [22]. If $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are bounded, then (5.29) is satisfied if G_M is differentiable and directionally differentiable of degree 2 at $\bar{x}$. In fact, in this case, from Proposition 2.2 in [24], there exists a constant L_1 such that

$$\|\Delta \bar{y}^k - \nabla G_M(\bar{x})\Delta x^k\| \le L_1 \max\{\|x^{k+1} - \bar{x}\|^2, \|x^k - \bar{x}\|^2\}. \tag{5.31}$$

On the other hand, from (5.24), there exists a constant L_2 such that

$$\max\{\|x^{k+1} - \bar{x}\|, \|x^k - \bar{x}\|\} \le L_2\|\Delta x^k\|,$$

which, together with (5.31), implies that (5.29) holds with $L := L_1 L_2$.

If we do not wish to assume that $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are bounded, we may use Corollary 5.9 to obtain such boundedness by assuming that $G'_M(x; \cdot)$ is radially Lipschitz continuous at $\bar{x}$; i.e., the directional derivative of G_M exists on a neighborhood of $\bar{x}$ and there exists a constant $L > 0$ such that

$$\sup_{\|d\|=1} \|G'_M(x; d) - G'_M(\bar{x}; d)\| \le L\|x - \bar{x}\|$$

for all x in that neighborhood of $\bar{x}$. From Lemma 2.2 in Pang [20], this strong condition implies that (5.29) is satisfied. Also, from results in [20] and [24], this condition implies that G_M is strongly differentiable and directionally differentiable of degree 2 at $\bar{x}$.

LEMMA 5.10. *Suppose that all the assumptions in Corollary* 5.9 *hold. Then*

$$\|x^k + s^k - \bar{x}\| = o(\|x^k - \bar{x}\|). \tag{5.32}$$

Proof. Since Δx^k is a positive multiple of $s^k = -B_k^{-1}\tilde{G}_M(x^k)$, (5.30) implies that

$$\lim_{k\to\infty} \frac{\|(B_k - \nabla G_M(\bar{x}))s^k\|}{\|s^k\|} = 0$$

and

$$-\tilde{G}_M(x^k) - \nabla G_M(\bar{x})s^k = o(\|s^k\|).$$

So,

$$\begin{aligned}
\nabla G_M(\bar{x})s^k &= -\tilde{G}_M(x^k) + o(\|s^k\|) \\
&= -G_M(x^k) + o(\|G_M(x^k)\|) + o(\|s^k\|) \\
&= O(\|x^k - \bar{x}\|) + o(\|s^k\|),
\end{aligned}$$

which together with the invertibility of $\nabla G_M(\bar{x})$ implies that

$$\|s^k\| = O(\|x^k - \bar{x}\|). \tag{5.33}$$

On the other hand,

$$\begin{aligned}(B_k - \nabla G_M(\bar{x}))s^k &= -\tilde{G}_M(x^k) - \nabla G_M(\bar{x})s^k \\ &= -G_M(x^k) + o(\|G_M(x^k)\|) - \nabla G_M(\bar{x})s^k \\ &= -\nabla G_M(\bar{x})(x^k - \bar{x}) - \nabla G_M(\bar{x})s^k + o(\|x^k - \bar{x}\|) \\ &= -\nabla G_M(\bar{x})(x^k + s^k - \bar{x}) + o(\|x^k - \bar{x}\|). \qquad (5.34)\end{aligned}$$

From (5.34) and (5.33),

$$\begin{aligned}\frac{\|\nabla G_M(\bar{x})(x^k + s^k - \bar{x})\|}{\|x^k - \bar{x}\|} &= o(1) + \frac{\|(B_k - \nabla G_M(\bar{x}))s^k\|}{\|s^k\|}\frac{\|s^k\|}{\|x^k - \bar{x}\|} \\ &= o(1),\end{aligned}$$

which together with the invertibility of $\nabla G_M(\bar{x})$ implies (5.32). ∎

LEMMA 5.11. *Suppose that all the assumptions in Corollary* 5.9 *hold. Then*

$$F_M(x^k + s^k) \le F_M(x^k) + \sigma(s^k)^T \tilde{G}_M(x^k)$$

for all sufficiently large k.

Proof. From the differentiability of G_M and the fact that $G_M(\bar{x}) = 0$, we have

$$F_M(x) = F_M(\bar{x}) + \frac{1}{2}(x - \bar{x})^T \nabla G_M(\bar{x})(x - \bar{x}) + o(\|x - \bar{x}\|^2).$$

From Lemma 5.10, $\|x^k + s^k - \bar{x}\| = o(\|x^k - \bar{x}\|)$, so

$$\|x^k - \bar{x}\| = \|s^k\| + o(\|s^k\|).$$

Therefore, from Lemma 5.10 and the boundedness of $\{B_k\}$,

$$\begin{aligned}&F_M(x^k + s^k) - F_M(x^k) - \sigma(s^k)^T \tilde{G}_M(x^k) \\ &= -\frac{1}{2}(x^k - \bar{x})^T \nabla G_M(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|^2) - \sigma(s^k)^T \tilde{G}_M(x^k) \\ &= -\frac{1}{2}(s^k)^T \nabla G_M(\bar{x})s^k + o(\|s^k\|^2) + \sigma(s^k)^T B_k s^k \\ &= -\sigma(s^k)^T(\nabla G_M(\bar{x}) - B_k)s^k + \left(\sigma - \frac{1}{2}\right)(s^k)^T \nabla G_M(\bar{x})s^k + o(\|s^k\|^2) \\ &= \left(\sigma - \frac{1}{2}\right)(s^k)^T \nabla G_M(\bar{x})s^k + o(\|s^k\|^2),\end{aligned}$$

which, together with the positive definiteness of $\nabla G_M(\bar{x})$ and the algorithm assumption that $\sigma < 1/2$, implies that for all sufficiently large k,

$$F_M(x^k + s^k) - F_M(x^k) - \sigma(s^k)^T \tilde{G}_M(x^k) < 0.$$

This completes the proof of this lemma. ∎

Now we have all the necessary material to give the superlinear convergence result.

THEOREM 5.12. *Suppose that F_M is strongly convex on D, G_M is Fréchet differentiable at $\bar{x}$, and there exists a constant $L > 0$ such that* (5.29) *holds. Then the sequence $\{x^k\}$ generated by the BFGS bundle-type method converges to $\bar{x}$ Q-superlinearly.*

Proof. From Lemmas 5.11 and 2.1 and line search criterion (3.2), for all sufficiently large k, we have

$$x^{k+1} = x^k + s^k.$$

Then the Q-superlinear convergence of $\{x^k\}$ follows from Lemma 5.10. □

6. Conclusions. This paper presents a globally and superlinearly convergent BFGS bundle-type method for the case where the Moreau–Yosida regularization function F_M and its gradient G_M are computed only approximately. It does not require the original objective to be differentiable at the solution. To accomplish this we employ a bundle method to implement $\varepsilon_k = \hat{F}_M(x^k) - \check{F}_M(x^k) = o(\|G_M(x^k)\|^2)$, which is an essential condition for superlinear convergence of an approximate Newton method applied to this type of problem [12]. Because of this requirement the subproblems may increase in difficulty as k increases. To try to alleviate this potential difficulty it may be beneficial to consider space decomposition as in [18] and to vary M in such a way that the subproblems are solved mainly in the subspace where the cutting-plane aspect of bundling is efficient. Also, if the variation in M and space decomposition are done properly, it may be possible to weaken the rate of convergence assumption to assuming that some regularization of f is strongly convex on a proper subset of $\Re^n$ when f is not differentiable at the solution.

In [7], Chen and Fukushima provide a globally and linearly convergent proximal quasi-Newton method and discuss local superlinear convergence conditions. Here we focus our attention on giving superlinear convergence conditions for a BFGS bundle-type method. It may be possible to generalize our results to an important subclass of the Broyden class of quasi-Newton methods by using the results in [4, 2] corresponding to some positive and negative values of the class parameter.

Acknowledgment. The authors would like to thank two referees for their helpful comments.

REFERENCES

[1] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, *A family of variable metric proximal methods*, Math. Programming, 68 (1995), pp. 15–47.

[2] R. H. Byrd, D. C. Liu, and J. Nocedal, *On the behavior of Broyden's class of quasi-Newton methods*, SIAM J. Optim., 2 (1992), pp. 533–557.

[3] R. H. Byrd and J. Nocedal, *A tool for the analysis of quasi-Newton methods with application to unconstrained minimization*, SIAM J. Numer. Anal., 26 (1989), pp. 727–739.

[4] R. H. Byrd, J. Nocedal, and Y. Yuan, *Global convergence of a class of quasi-Newton methods on convex problems*, SIAM J. Numer. Anal., 24 (1987), pp. 1171–1189.

[5] J. V. Burke and M. Qian, *A variable metric proximal point algorithm for monotone operators*, SIAM J. Optim., to appear.

[6] J. V. Burke and M. Qian, *On the superlinear convergence of the variable metric proximal point algorithm using Broyden and BFGS matrix secant updating*, Math. Programming, 1998, submitted.

[7] X. Chen and 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.

[8] R. Correa and C. Lemaréchal, *Convergence of some algorithms for convex optimization*, Math. Programming, 62 (1993), pp. 261–275.

[9] J. E. Dennis and J. J. Moré, *Quasi-Newton methods, motivation and theory*, SIAM Rev., 19 (1977), pp. 46–89.

[10] U. Felgenhauer, *On the stable global convergence of particular quasi-Newton methods*, Optimization, 26 (1992), pp. 97–113.

[11] M. Fukushima, *A descent algorithm for nonsmooth convex optimization*, Math. Programming, 30 (1984), pp. 163–175.

[12] M. Fukushima and L. Qi, *A globally and superlinearly convergent algorithm for nonsmooth convex minimization*, SIAM J. Optim., 6 (1996), pp. 1106–1120.

[13] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, vols. 1 and 2, Springer-Verlag, Berlin, Heidelberg, New York, 1993.

[14] C. -M. Ip and J. Kyparisis, *Local convergence of quasi-Newton methods for B-differentiable equations*, Math. Programming, 56 (1992), pp. 71–89.

[15] K. C. Kiwiel, *Proximal control in bundle methods for convex nondifferentiable minimization*, Math. Programming, 46 (1990), pp. 105–122.

[16] C. Lemaréchal and C. Sagastizábal, *An approach to variable metric methods*, in *Systems Modelling and Optimization*, in Lecture Notes in Control and Inform. Sci. 197, J. Henry and J. -P. Yvon, eds., Springer-Verlag, Berlin, 1994.

[17] C. Lemaréchal and C. Sagastizábal, *Practical aspects of the Moreau-Yosida regularization* I: *Theoretical preliminaries*, SIAM J. Optim., 7 (1997), pp. 367–385.

[18] R. Mifflin, *A quasi-second-order proximal bundle algorithm*, Math. Programming, 73 (1996), pp. 51–72.

[19] J. J. Moreau, *Proximite et dualite dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[20] J. -S. Pang, *Newton's method for B-differentiable equations*, Math. Oper. Res., 15 (1990), pp. 311–341.

[21] M. J. D. Powell, *Some global convergence properties of a variable metric algorithm for minimization without exact line searches*, in Non-Linear Programming, R. W. Cottle and C. E. Lemke, eds., SIAM-AMS Proceedings 9, AMS, Provindence, RI, 1976.

[22] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–244.

[23] L. Qi, *Second-order analysis of the Moreau-Yosida regularization of a convex function*, Applied Mathematics report 94/20, revised version, School of Mathematics, The University of New South Wales, Sydney, Australia, 1995.

[24] L. Qi, *On superlinear convergence of quasi-Newton methods for nonsmooth equations*, Oper. Res. Lett., 20 (1997), pp. 223–228.

[25] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[26] R. T. Rockafellar, *Maximal monotone relations and the second derivatives of nonsmooth functions*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 2 (1985), pp. 167–186.

[27] K. Yosida, *Functional Analysis*, Springer-Verlag, Berlin, 1964.