A Family of variable metric proximal methods

J. Frederic Bonnans, Jean Charles Gilbert, Claude Lemaréchal, Claudia Sagastizábal

INRIA
INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Family of Variable Metric Proximal Methods

Joseph Frédéric Bonnans , Jean Charles Gilbert , Claude Lemaréchal , Claudia Sagastizábal

N° 1851

Février 1993

PROGRAMME 5

Traitement du signal,

automatique

et productique

ISSN 0249-6399

INRIA
ROCQUENCOURT

A Family of Variable Metric Proximal Methods

Joseph Frédéric Bonnans , Jean Charles Gilbert , Claude Lemaréchal , Claudia Sagastizábal

Programme 5 — Traitement du signal, automatique et productique
Projet Promath

Rapport de recherche n ° 1851 — Février 1993 — 33 pages

Abstract: We consider conceptual optimization methods combining two ideas: the Moreau-Yosida regularization in convex analysis, and quasi-Newton approximations of smooth functions. We outline several approaches based on this combination, and establish their global convergence. Then we study theoretically the local convergence properties of one of these approaches, which uses quasi-Newton updates of the objective function itself. Also, we obtain a globally and superlinearly convergent BFGS proximal method. At each step of our study, we single out the assumptions that are useful to derive the result concerned.

Key-words: Bundle methods, convex optimization, global and superlinear convergence, mathematical programming, proximal point, quasi-Newton algorithms, variable metric.

(Résumé : tsvp)

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33 1) 39 63 55 11 – Télécopie : (33 1) 39 63 53 30

Une famille de méthodes de quasi-Newton proximales

Résumé : Nous considérons des méthodes conceptuelles d'optimisation combinant deux idées: la régularisation de Moreau-Yosida en analyse convexe et les approximations quasi-Newtoniennes des fonctions régulières. Nous développons quelques approches basées sur cette combinaison et établissons leur convergence globale. Nous étudions ensuite d'un point de vue théorique les propriétés de convergence locale d'une de ces approches, dans laquelles les mises-à-jour utilisent la fonction originale. Nous présentons aussi une méthode de BFGS proximale qui converge globalement et superlinéairement. A chaque étape de notre développement, nous précisons les hypothèses minimales nécessaires à l'obtention des résultats.

1. Introduction. We consider in this paper algorithms to solve

$$\min\{f(x) : x \in \mathbb{R}^N\}, \tag{1}$$

where f is always assumed closed proper convex (we follow the terminology of [30]: f takes its values in $\mathbb{R} \cup \{+\infty\}$ but is not identically $+\infty$; closedness means lower semi-continuity). Additional assumptions on f will also be made, when studying rates of convergence.

Our algorithms are based on the use of the *proximal mapping*: given $x \in \mathbb{R}^N$ and a symmetric positive definite $N \times N$ matrix M, f is perturbed to the strongly convex function:

$$\varphi_M(z) := f(z) + \frac{1}{2}\langle M(z-x), z-x\rangle\,; \tag{2}$$

$\langle u, v\rangle := u^{\mathrm{T}}v$ is the usual dot product in $\mathbb{R}^N$ and $|\cdot|$ the associated norm. Note that φ_M has a unique minimizer. The image of x under the proximal mapping is:

$$p_M(x) := \operatorname{argmin}\{\varphi_M(z) : z \in \mathbb{R}^N\}. \tag{3}$$

Throughout, we will find it convenient to use the notation

$$x^p := p_M(x).$$

A traditional way of solving problem (1) via the proximal mapping (3) is the *proximal point algorithm*: see [19] and [31]. This method generates a minimizing sequence $\{x_n\}$ by the recurrence formula

$$x_{n+1} := x_n^p = p_M(x_n), \tag{4}$$

with a possibly varying matrix M of the form $M = c_n I$, $c_n > 0$. In view of the optimality condition for (3)

$$x^p = x - M^{-1}g^p, \quad \text{for some } g^p \in \partial f(x^p),$$

the proximal point algorithm can be seen as a "preconditioned implicit gradient method" to minimize f. The method is implicit since the subgradient used in the formula is evaluated at x^p, not at x, and the preconditioning is realized by the matrix M.

Another motivation for this approach is the *Moreau-Yosida regularization* of f: see [21] and [31]. This is the function f^p whose value at $x \in \mathbb{R}^N$ is:

$$f^p(x) := \varphi_M(x^p) = \min\{f(z) + \frac{1}{2}\langle M(z-x), z-x\rangle : z \in \mathbb{R}^N\}. \tag{5}$$

Indeed, the minima of f coincide with those of f^p and this latter function is convex, finite everywhere, and fairly smooth: with no additional assumption, f^p has a Lipschitz continuous gradient given by the formula

$$\nabla f^p(x) = M(x - x^p) = g^p. \tag{6}$$

Then the proximal point algorithm written in the form

$$x_{n+1} := x_n^p = x_n - M^{-1}M(x_n - x_n^p) = x_n - M^{-1}\nabla f^p(x_n)$$

can also be viewed as a preconditioned "explicit" gradient method to minimize f^p.

Thus, the Moreau-Yosida regularization provides a link between classical and nonsmooth optimization: a natural and attractive idea is to minimize f^p by a variable metric method of the type

$$x_{n+1} := x_n - t_n M_n^{-1}\nabla f^p(x_n) = x_n - t_n M_n^{-1}M(x_n - x_n^p). \tag{7}$$

The stepsize $t_n > 0$ can be computed as usual, and the matrix M_n can be generated according to a quasi-Newton formula [8], $M_{n+1} := \mathrm{qN}(M_n, y_n, s_n)$, using

$$s_n := x_{n+1} - x_n, \qquad y_n := \nabla f^p(x_{n+1}) - \nabla f^p(x_n) = M(x_{n+1} - x_{n+1}^p - x_n + x_n^p) \tag{8}$$

(other choices for s_n and y_n are possible, see [18]).

For example, the BFGS formula can be used; because its convergence just requires Lipschitz continuity of the gradient ([25]), the resulting method will converge always globally, and superlinearly in the "good" cases when f^p has a Lipschitz continuous Hessian. Now come implementation issues: how can we compute x_n^p? and how will its computation –or rather its approximation– affect convergence properties? As pointed out in [9], [1], *bundle methods* are a possible proposal. Given $x = x_n$, they provide a way of constructing a sequence $\{p^k\}$ tending to $p_M(x)$ when $k \to \infty$; more importantly, they also provide an efficient stopping criterion to apply a recurrence formula such as (4), the proximal point being replaced by its approximation p^k. We refer to [15], [5] for an accurate account of bundle methods from this point of view.

Starting from these ideas, we distinguish three possibilities.

- **Algorithmic Pattern 1 (AP1):**

Step 0: The symmetric positive definite matrix M is fixed throughout, say $M = I$. Start with an initial x_1 and some matrix M_1. Set the iteration counter $n = 1$.

Step 1: Given x_n, generate a sequence $p_n^k \to p_M(x_n)$, for example by a bundling algorithm, until the associated stopping criterion is satisfied.

Step 2: Compute a stepsize $t_n > 0$ to obtain

$$x_{n+1} := x_n - t_n M_n^{-1}M(x_n - p_n^k).$$

Step 3: Update M_n by a quasi-Newton formula using (8). Increase n by 1 and loop to Step 1.

Unfortunately, bundle methods, which produce the estimate p_n^k in Step 1, rely heavily on the update formula $x_{n+1} = p_n^k$. The reason is that Step 1 is stopped when $f(p_n^k)$ is sufficiently smaller than $f(x_n)$; but this decrease does not seem to allow $f(x_{n+1}) < f(x_n)$ in Step 2. We refer to [20] for first steps into the analysis of the above strategy.

REMARK 1.1. Incidentally, a second question is the choice of M: after all, the best matrix for (1)-(3) should be $M = 0$, in which case no update of x_n would be needed. Among other things, M should somehow take into account the scaling of the problem. □

A way round this difficulty is to take in (5) a *varying* matrix M yielding $x_{n+1} = x_n^p$. This results in the following variant:

- **Algorithmic Pattern 2 (AP2):**

STEP 0: Start with some initial point x_1 and matrix M_1. Set $n = 1$.

STEP 1: Given x_n and M_n, generate a sequence $p_n^k \to p_{M_n}(x_n)$, for example by a bundling algorithm, until the associated stopping criterion is satisfied.

STEP 2: Take

$$x_{n+1} := p_n^k.$$

STEP 3: Update M_n by a quasi-Newton formula using (8). Increase n by 1 and loop to Step 1.

The need for an artificial M is thus eliminated (barring the initial M_1), and the spirit of bundle methods is preserved; but now, the difficulty is in the quasi-Newton field: we no longer have a fixed Moreau-Yosida regularization f^p, whose Hessian is going to be approximated by $\{M_n\}$: we rather have a varying function f^p which depends on M_n, giving birth to a sort of vicious circle.

REMARK 1.2. Exploratory experiments with this latter algorithm indicate that some eigenvalues of M_n may have a tendency to approach 0; in view of Remark 1.1 this is not bad (f^p becomes closer to the true objective f), but will certainly result in delicate analysis and numerical implementation. On the other hand, preliminary experiments also indicate that this pattern deserves study: the algorithm behaves quite well on a benchmark of test problems for nonsmooth optimization ([32]). □

In this paper, we concentrate on a third alternative, based on an idea of [28]:

- **Algorithmic Pattern 3 (AP3):** Take (AP2) but, instead of (8), let the quasi-Newton update use more simply

$$s_n = x_{n+1} - x_n, \qquad y_n = \nabla f(x_{n+1}) - \nabla f(x_n). \tag{9}$$

Then the algorithm is just that of (AP2) with the following last step:

STEP 3: Update M_n by a quasi-Newton formula using (9). Increase n by 1 and loop to Step 1.

Naturally this has little meaning in the framework of nonsmooth optimization: (9) requires differentiability from f. Furthermore, we will pay little attention to implementability issues, i.e., on the actual computation of each proximal point x_n^p. Our ambition here is limited to exploring preliminary results to combine methods for nonsmooth optimization and classical quasi-Newton methods.

The paper is organized as follows. In the next section we state an abstract algorithmic pattern which accommodates any of the above strategies (AP1-3), and we give conditions guaranteeing global convergence. This section does not rely on the actual computation of proximal points x_n^p, neither on specific formulæ generating the matrices M_n. We obtain in § 2 our global results without any additional assumption on f. In the following sections, when we consider the local analysis of specific quasi-Newton formulæ, we require ∇f to be locally Lipschitzian, we assume also that it admits directional derivatives at $\bar{x}$. In Section 3, we adapt to our case the criterion of [6] for superlinear convergence. Then we give superlinear convergence results for a wide class of quasi-Newton methods, including PSB and DFP, assuming that f has at $\bar{x}$ a Hessian, in a "strong" sense. Under the same assumptions, we concentrate in Section 4 on both global and superlinear convergence of a conceptual algorithm using the BFGS update. Finally Section 5 gives some concluding remarks.

2. Global Convergence. In this section we prove the global convergence of the algorithms described abstractly by the General Algorithmic Pattern (GAP) below. Let (x_n, M_n) be the current iterate with M_n symmetric positive definite. Then, according to (4) and (5), the corresponding proximal point will be:

$$x_n^p := p_{M_n}(x_n) = \operatorname{argmin}\{f(z) + \frac{1}{2}\langle M_n(z-x), z-x\rangle : z \in \mathbb{R}^N\}. \tag{10}$$

We set

$$W_n := M_n^{-1}.$$

LEMMA 2.1. *With the notation and assumptions of Section* 1, *the following holds:*
(*i*) *The proximal point* x_n^p *is well defined and given by*

$$x_n^p = x_n - W_n g_n^p, \tag{11}$$

with

$$g_n^p \in \partial f(x_n^p). \tag{12}$$

(*ii*)

$$f(x_n^p) \le f(x_n) - \langle W_n g_n^p, g_n^p\rangle. \tag{13}$$

(*iii*) x_n *minimizes* $f \iff x_n = x_n^p \iff g_n^p = 0$.
(*iv*) *For all* y *with* $f(y) \le f(x_n^p)$, *there holds*

$$\langle M_n(y - x_n^p), y - x_n^p\rangle \le \langle M_n(y - x_n), y - x_n\rangle. \tag{14}$$

Proof. The minimand in (10) is lower semi-continuous and strongly convex; moreover, for any $z \in \operatorname{dom}(\partial f)$ it has the subdifferential $\partial f(z) + M_n(z - x_n)$. Existence and uniqueness of its minimum (that is the proximal point) is therefore clear, as well as the optimality condition (11)–(12). To obtain (13), multiply (11) by g_n^p and use (12). The equivalences in (*iii*) follow easily from (*i*) and (*ii*). As for (*iv*), take y with $f(y) \le f(x_n^p)$. Using (12),

$$f(x_n^p) \ge f(y) \ge f(x_n^p) + \langle g_n^p, y - x_n^p\rangle,$$

so that, with (11),

$$\langle M_n(x_n - x_n^p), y - x_n^p \rangle \leq 0.$$

Then develop the relation $|M_n^{1/2}(x_n - y)|^2 = |M_n^{1/2}(x_n - x_n^p + x_n^p - y)|^2$ to obtain (14). □□

Thus, the decrease of f from x_n to x_n^p is at least $\langle g_n^p, W_n g_n^p \rangle$, a positive number unless x_n is optimal. The Moreau-Yosida regularization takes the value

$$f^p(x_n) = f(x_n^p) + \frac{1}{2}\langle x_n^p - x_n, M_n(x_n^p - x_n)\rangle \tag{15}$$

and, using (11) we set

$$\delta_n := f(x_n) - f^p(x_n) = f(x_n) - f(x_n^p) - \frac{1}{2}\langle g_n^p, W_n g_n^p\rangle\ . \tag{16}$$

Observe that, in view of (13),

$$\frac{1}{2}\langle g_n^p, W_n g_n^p\rangle \leq \delta_n \leq f(x_n) - f(x_n^p)\,. \tag{17}$$

We consider in this section a very general pattern, in which f is simply required to decrease at each iteration by at least a fixed fraction m of δ_n, interpreted as a "nominal decrease".

- **General Algorithmic Pattern (GAP):**

STEP 0: Start with some initial point x_1 and matrix M_1; choose some descent parameter $m \in]0,1[$; set $n = 1$.

STEP 1: With δ_n given by (16), compute x_{n+1} satisfying

$$f(x_{n+1}) \leq f(x_n) - m\delta_n \tag{18}$$

(Note: for this, Proposition 2.2 below is helpful).

STEP 2: Update M_n, increase n by 1 and loop to Step 1.

For a nominal decrease, the use of the value $f(x_n) - f(x_n^p)$ in (18) may seem more natural than our δ_n. A substantial advantage of (16), however, is that implementable methods are known to guarantee (18) without computing any proximal point. In fact, if f is replaced by some smaller function ψ in the proximal problem (10), we get a smaller optimal value, which can be used to overestimate the nominal decrease.

PROPOSITION 2.2. *With the notation above, let ψ be a closed convex function on $\mathbb{R}^N$ satisfying $\psi \leq f$ and set*

$$\pi := \operatorname{argmin}\{\psi(z) + \frac{1}{2}\langle M_n(z - x_n), z - x_n\rangle : z \in \mathbb{R}^N\}. \tag{19}$$

(i) If

$$f(\pi) \leq f(x_n) - m[f(x_n) - \psi(\pi)], \tag{20}$$

then (18) is satisfied by $x_{n+1} = \pi$.

(*ii*) *If x_n does not minimize f, there exists $\epsilon_n > 0$ such that $f(\pi) - \psi(\pi) \leq \epsilon_n$ implies*

$$f(\pi) - \psi(\pi) \leq (1-m)[f(x_n) - \psi(\pi)], \tag{21}$$

which is equivalent to (20).

Proof. In the inequality

$$\psi(\pi) - f(x_n) \leq \psi(\pi) + \frac{1}{2}\langle M_n(\pi - x_n), \pi - x_n\rangle - f(x_n),$$

over-estimate the right-hand side by replacing successively π by x_n^p and ψ by f. Using (11), (16) we obtain

$$\psi(\pi) - f(x_n) \leq f(x_n^p) + \frac{1}{2}\langle W_n g_n^p, g_n^p\rangle - f(x_n) = -\delta_n; \tag{22}$$

because $m > 0$, (i) is clearly proved.

Now the equivalence between (21) and (20) is straightforward. If x_n does not minimize f, then $x_n^p \neq x_n$ and $g_n^p \neq 0$. In view of (22), we see that (21) = (20) is satisfied whenever, for example,

$$f(\pi) - \psi(\pi) \leq (1-m)\delta_n =: \epsilon_n > 0.$$

□□

The idea underlying (20) is classical in line-searches and trust region algorithms, if we interpret ψ as a model for f, whose value at the trial iterate π is a target for $f(x_{n+1})$. Proposition 2.2 only says that our descent test (18) is passed whenever the model is accurate enough at π. Bundle methods, precisely, construct such a model which is piecewise affine, resulting in a quadratic program for the proximal problem (10); see for example [5].

In the convergence result below $\lambda_{\min}(W)$ denotes the smallest eigenvalue of a symmetric matrix W; (23) is natural to rule out perturbed functions φ_M of (2) departing too much from f.

THEOREM 2.3. *Assume that the closed convex function f has a nonempty bounded set of minima, and let $\{x_n\}$ be a sequence generated by* (GAP). *Then $\{x_n\}$ is bounded and, if*

$$\sum_{n=1}^{\infty} \lambda_{\min}(W_n) = \infty, \tag{23}$$

any accumulation point of $\{x_n\}$ minimizes f. The same properties hold for the sequence of proximal points $\{x_n^p\}$ and it also holds $\liminf |g_n^p| = 0$.

Proof. Our starting assumption implies that the level sets of f are bounded (see [30], Theorems 8.4 and 8.7 and [13], Proposition IV.3.2.5); the sequences $\{x_n\}$ and $\{x_n^p\}$ are therefore bounded by construction. In what follows, $\bar{f}$ will denote the minimal value of f.

Combining (17) and (18), we have

$$\frac{1}{2}\langle W_n g_n^p, g_n^p\rangle \leq \frac{1}{m}(f(x_n) - f(x_{n+1})), \tag{24}$$

which gives by summation

$$\sum_{n=1}^{\infty} \langle W_n g_n^p, g_n^p \rangle \leq \frac{2}{m}(f(x_1) - \bar{f}) < \infty$$

and therefore

$$\sum_{n=1}^{\infty} \lambda_{\min}(W_n)|g_n^p|^2 < \infty.$$

In view of (23), the sequence $\{|g_n^p|^2\}$ cannot be bounded away from 0: there exists a subset $N_1 \subset \mathbb{N}$ such that $\lim_{n\in N_1} g_n^p = 0$.

Extract from N_1 a further subset, say $N_2 \subset N_1$, such that $\{x_n^p\}_{n\in N_2}$ tends to some limit $\bar{x}$. Because of (12), the closedness of the subdifferential mapping implies that $0 \in \partial f(\bar{x}) : \bar{x}$ minimizes f and $f(\bar{x}) = \bar{f}$.

Now $\{f(x_n)\}$ is nonincreasing and has a limit f^*; also $\langle W_n g_n^p, g_n^p \rangle \to 0$ in view of (24). Pass to the limit in (18), written for $n \in N_2$: we obtain

$$f^* \leq f^* - m(f^* - \bar{f}),$$

which implies $f^* = \bar{f}$. Then any accumulation point of $\{x_n\}$ is also optimal. □□

3. Local Convergence. From now on, f is assumed differentiable (and therefore finite everywhere). We use the notation $g(x)$ for the gradient of f at x, as well as $g_n = g(x_n)$ and $g_n^p = g(x_n^p)$.

We specialize in this section the General Algorithm Pattern of Section 2 along the lines of (AP3) in Section 1: we suppose the proximal point x_n^p is computed exactly and the symmetric positive definite matrix M_n is updated at each iteration by a formula such that the *quasi-Newton equation* holds:

$$M_{n+1} s_n = y_n, \tag{25}$$

where

$$s_n := x_{n+1} - x_n, \quad y_n := g_{n+1} - g_n.$$

In these circumstances, the pure prox-form of (AP3) is clumsy, as observed in [20] and [28]. Indeed, take the "ideal" situation in which f is quadratic with a positive definite Hessian matrix A, and take $M_n = A$ in the algorithm. Then, x_n^p is the minimizer of

$$\langle g_n, x - x_n \rangle + \frac{1}{2} \langle 2A(x - x_n), x - x_n \rangle,$$

which is only half-way towards the real minimum of f. A natural cure would be to do a line-search along the direction $x_n^p - x_n$. This idea will be used in Section 4, but in the present local study, we assume that the "ideal" stepsize of 2 is taken.

In a word, we consider in this section the following algorithm:

- **Quasi-Newton proximal algorithm (qN-AP3):**

STEP 0: Start with some initial point x_1 and a positive definite matrix M_1. Set $n = 1$.

STEP 1: Compute $x_n^p := p_{M_n}(x_n)$.

STEP 2: Update:

$$x_{n+1} := x_n - 2M_n^{-1} g_n^p = x_n + 2(x_n^p - x_n). \tag{26}$$

STEP 3: Update M_n by a quasi-Newton formula satisfying (25). Increase n by 1 and loop to Step 1.

Keeping here the notation of the preceding sections, we set

$$e_n := x_n - \bar{x}, \quad e_n^p := x_n^p - \bar{x} \quad \text{and} \quad \sigma_n := |e_{n+1}| + |e_n|. \tag{27}$$

Recall that we have from Lemma 2.1, with $g_n^p := g(x_n^p)$,

$$g_n^p + M_n(x_n^p - x_n) = 0. \tag{28}$$

Finally, remark that (26) gives

$$x_n^p = \frac{1}{2}x_n + \frac{1}{2}x_{n+1} \quad \text{and} \quad e_n^p = \frac{1}{2}e_n + \frac{1}{2}e_{n+1}. \tag{29}$$

In this section, we study the local convergence properties of the sequence $\{x_n\}$ generated by Algorithm (qN-AP3). We always assume that the gradient of f has directional derivatives at $\bar{x}$, a minimum point of f; our smoothness assumptions are reviewed in Section 3.1. In Section 3.2, we prove the linear convergence of $\{x_n\}$, assuming that (x_1, M_1) is "good enough" and that a bounded deterioration property holds for $\{M_n\}$ as is done in [14] for *standard* quasi-Newton algorithms. We characterize the superlinear convergence in Section 3.3, giving the *prox-version* of the well-known characterization for superlinear convergence of [6]. Finally, under stronger smoothness assumptions, we obtain local and superlinear convergence results for a wide class of quasi-Newton formulæ, including the prox-versions of the PSB and DFP algorithms. For this we extend the approach of Grzegórski [12] to variational quasi-Newton methods with variable norms and to the "proximal" framework.

3.1. Smoothness assumptions. In this subsection we state the assumptions needed for the sequel. We start by recalling some classical notions. An operator H from $\mathbb{R}^N$ to $\mathbb{R}^N$ is *positively homogeneous* when $H(tv) = tHv$, for all $v \in \mathbb{R}^N$ and all $t \geq 0$. Such an operator is said *bounded* if

$$|H| := \sup_{|v|=1} |Hv|$$

is finite. It is equivalent to say that H is continuous at 0. Observe that we use the same notation for the Euclidean norm in $\mathbb{R}^N$ and for the induced operator norm.

For the local analysis, only the behaviour of f in some neighborhood of $\bar{x}$ is relevant. Actually, our assumptions throughout involve a *convex* neighborhood Ω of $\bar{x}$.

– First of all, we require the gradient to be *locally Lipschitzian* around $\bar{x}$: there is a constant L_g such that

$$\forall x, y \in \Omega, \quad |g(x) - g(y)| \leq L_g |x - y|. \tag{30}$$

– We postulate that g admits at $\bar{x}$ a *directional derivative* $g'(\bar{x}, d)$ for all $d \in \mathbb{R}^N$. To stress that we are only interested in g' at the fixed solution point $\bar{x}$, we will generally use the notation $\bar{H}$ for the mapping $d \mapsto g'(\bar{x}, d)$. In other words,

$$\bar{H}d := \bar{H}(d) = \lim_{t \downarrow 0} \frac{g(\bar{x} + td) - g(\bar{x})}{t}. \tag{31}$$

Observe that $\bar{H}$ is positively homogeneous by definition and bounded because of (30): $|\bar{H}| \leq L_g$.

– We will often suppose that the directional derivative (31) exists in a *strong* sense at $\bar{x}$ ([22], [23], see also [4], where the word *strict* is used). This means

$$\lim_{\substack{(x,y) \to (\bar{x}, \bar{x}) \\ x \neq y}} \frac{g(x) - g(y) - \bar{H}(x - y)}{|x - y|} = 0. \tag{32}$$

– Our final results need the difference quotient in (32) to converge at a specific rate, namely: for some positive constant L and all $x, y \in \Omega$,

$$|g(x) - g(y) - \bar{H}(x - y)| \leq L(|x - \bar{x}| + |y - \bar{x}|)|x - y|. \tag{33}$$

Needless to say, (33) implies (32), which in turn implies (31).

It is interesting to relate our assumptions with some other notions of weakened differentiability already stated in the literature; see for example [23], [16], [14], [24]. We recall first that, under the Lipschitz property (30), the limit in (31) becomes uniform in d: (31) is then equivalent to

$$g(\bar{x} + h) = g(\bar{x}) + \bar{H}h + o(|h|), \quad \text{when } h \to 0, \tag{34}$$

that is, $\bar{H}$ is the B-derivative of g at $\bar{x}$, in the sense of [29].

Assumption (32) turns out to be rather strong, even though it is a purely punctual condition. In fact, it can be seen as in the proof of [23, Theorem 2], that it implies the linearity of $\bar{H}$; and this just means that $\bar{H}$ is the strong *Fréchet* derivative of g at $\bar{x}$. To grasp the essence of (32), consider the case when g has directional derivatives in a neighborhood of $\bar{x}$: (32) just expresses the continuity of the mapping $x \mapsto g'(x, \cdot)$ at $\bar{x}$; this comes from the following theorem, which is an equivalent formulation of Theorem 2 in [23].

THEOREM 3.1. *Let $g : \mathbb{R}^N \to \mathbb{R}^N$ be a mapping satisfying* (30) *and having directional derivatives $g'(x, \cdot)$ for all $x \in \Omega$. Then the three statements below are equivalent:*

(i) the directional derivative $\bar{H}$ of (31) *satisfies the stronger limit property* (32),

(ii) g is Fréchet differentiable at $\bar{x}$ in the strong sense,

(iii) the mapping $x \mapsto g'(x,\cdot)$ is continuous at $\bar{x}$; in other words,

$$\sup_{|d|=1} |g'(x,d) - g'(\bar{x},d)| \to 0\,, \quad \text{when } x \to \bar{x}\,.$$

For an interpretation of our last assumption (33), assume again the existence of $g'(x,\cdot)$ in a neighborhood of $\bar{x}$: (33) connotes something stronger than the above continuity property *(iii)*, namely the "radially Lipschitz" property stated in (35). This comes from the next result, an equivalent formulation of Lemma 2.2 in [14]. It is here that the convexity of Ω is important.

THEOREM 3.2. *The hypotheses are those of Theorem* 3.1. *In addition, assume there exists a constant L such that*

$$\sup_{|d|=1} |g'(x,d) - g'(\bar{x},d)| \le L|x-\bar{x}| \quad \text{for all } x \in \Omega\,. \tag{35}$$

Then, for all x and $y \in \Omega$:

$$|g(x) - g(y) - \bar{H}(x-y)| \le L \max\{|x-\bar{x}|, |y-\bar{x}|\}\,|x-y|\,,$$

so that (33) *holds.*

3.2. Linear convergence and bounded deterioration. In this subsection, we prove the linear convergence of Algorithm (qN-AP3) when the generated matrices M_n satisfy a "Bounded Deterioration" property (Theorem 3.4). Before doing this, it is useful and instructive to analyze one step of the algorithm (Lemma 3.3). Our results are obtained under an extra assumption: there exists a positive definite matrix $\bar{M}$ such that

$$|I - \bar{M}^{-1}\bar{H}| \le \bar{r} < 1. \tag{36}$$

Assumption (36) is just a way of expressing that the positively homogeneous operator $\bar{H}$ is not too far from the open set of positive definite matrices that is convenient for the convergence analysis. This assumption was also made by Ip and Kyparisis [14]. When $\bar{H}$ is linear, condition (36) implies the nonsingularity of $\bar{H}$. When $\bar{H}$ is only a continuous positively homogeneous operator, however, the surjectivity of $\bar{H}$ is guaranteed (see the proof of Lemma 2 in [23]) but not its injectivity.

LEMMA 3.3. *Suppose that* (30), (31) *and* (36) *hold. Then, for all $r > \bar{r}/(2-\bar{r})$, there exist positive constants $\hat{\epsilon}_1$, $\hat{\epsilon}_2$ and μ such that if one iterate (x_n, M_n) of Algorithm* (qN-AP3) *satisfies*

$$|x_n - \bar{x}| \le \hat{\epsilon}_1 \quad \text{and} \quad |M_n - \bar{M}| \le \hat{\epsilon}_2, \tag{37}$$

then M_n is positive definite with $|M_n^{-1}| \le \mu$ and the next iterate x_{n+1} satisfies

$$|x_{n+1} - \bar{x}| \le r|x_n - \bar{x}|. \tag{38}$$

Proof. Let $r > \bar{r}/(2-\bar{r})$; there exists $r' > \bar{r}$ such that $r = r'/(2-r')$: just take $r' := 2r/(1+r)$. Now, choose $\hat{\epsilon}_2 > 0$ so that

$$\hat{\epsilon}_2 < \frac{1}{|\bar{M}^{-1}|} \quad \text{and} \quad \hat{\epsilon}_2 |\bar{M}^{-1}|\,|\bar{H}| \left(\frac{1}{|\bar{M}^{-1}|} - \hat{\epsilon}_2\right)^{-1} \leq \frac{r'-\bar{r}}{2}, \tag{39}$$

and set

$$\mu := \left(\frac{1}{|\bar{M}^{-1}|} - \hat{\epsilon}_2\right)^{-1}. \tag{40}$$

By the first inequality of (39), μ is a positive constant. Now, because $g(\bar{x}) = 0$, we have in (34) $g(x) - \bar{H}(x-\bar{x}) = o(|x-\bar{x}|)$. Therefore, there exists $\epsilon_1 > 0$ such that

$$|x-\bar{x}| \leq \epsilon_1 \quad \Longrightarrow \quad |g(x) - \bar{H}(x-\bar{x})| \leq \frac{r'-\bar{r}}{2\mu}|x-\bar{x}|. \tag{41}$$

Then, define $\hat{\epsilon}_1 > 0$ by

$$\hat{\epsilon}_1 := \frac{\epsilon_1}{\mu^{1/2}(|\bar{M}| + \hat{\epsilon}_2)^{1/2}}. \tag{42}$$

Having determined the positive constants $\hat{\epsilon}_1$, $\hat{\epsilon}_2$ and μ, we now prove the conclusions of the lemma, assuming (37).

First, by (37) and (39), we have

$$|M_n - \bar{M}| \leq \hat{\epsilon}_2 < \frac{1}{|\bar{M}^{-1}|}.$$

Then, the identity $M_n = \bar{M}[I + \bar{M}^{-1}(M_n - \bar{M})]$ and the Banach perturbation lemma imply that M_n is nonsingular (in fact positive definite) and that $|M_n^{-1}| \leq \mu$, with μ defined in (40).

Next, observe that $x_{n+1} = x_n - 2M_n^{-1}g_n^p = x_n^p - M_n^{-1}g_n^p$. Thus an easy calculation gives

$$\begin{aligned} e_{n+1} &= e_n^p - M_n^{-1}g_n^p \\ &= (I - M_n^{-1}\bar{H})e_n^p - M_n^{-1}(g_n^p - \bar{H}e_n^p) \\ &= (I - \bar{M}^{-1}\bar{H})e_n^p + \bar{M}^{-1}(M_n - \bar{M})M_n^{-1}\bar{H}e_n^p - M_n^{-1}(g_n^p - \bar{H}e_n^p). \end{aligned} \tag{43}$$

We are going to bound the norm of the right hand side of (43) by a multiple of $|e_n^p|$. There is no difficulty with the first two terms. For the last term we shall use the implication (41) after having shown that $|e_n^p| \leq \epsilon_1$. To do this, observe that lemma 2.1 (iv) with $y = \bar{x}$ gives

$$\frac{1}{|M_n^{-1}|}\,|e_n^p|^2 \leq |M_n|\,|e_n|^2.$$

Hence, using (37), $|e_n| \leq \hat{\epsilon}_1$ and (42), we get

$$|e_n^p| \leq |M_n^{-1}|^{1/2}|M_n|^{1/2}|e_n| \leq \mu^{1/2}(|\bar{M}| + \hat{\epsilon}_2)^{1/2}\hat{\epsilon}_1 = \epsilon_1.$$

Now, using (36) and (41), (43) gives

$$|e_{n+1}| \leq \left(\bar{r} + |\bar{M}^{-1}|\,\hat{\epsilon}_2\,\mu\,|\bar{H}| + \mu\,\frac{r'-\bar{r}}{2\mu}\right)|e_n^p| \leq r'|e_n^p|,$$

where we used the second inequality of (39) and (40). Finally, by (29), $|e_n^p| \leq (|e_n| + |e_{n+1}|)/2$, and the last inequality becomes

$$\left(1 - \frac{r'}{2}\right)|e_{n+1}| \leq \frac{r'}{2}|e_n|.$$

The conclusion of the lemma follows from the definition of r':

$$|e_{n+1}| \leq \frac{r'}{2-r'}|e_n| = r|e_n|.$$

□□

Since $\bar{r}/(2-\bar{r}) < 1$, Lemma 3.3 allows us to take $r < 1$. Then an easy consequence of this result is: if the matrices M_n are maintained in a ball of radius $\hat{\epsilon}_2$ around $\bar{M}$ and if the first iterate x_1 is taken sufficiently close to $\bar{x}$, the sequence $\{x_n\}$ generated by Algorithm (qN-AP3) converges to $\bar{x}$ linearly with rate r. As we shall see, this property of the matrices M_n is satisfied when they are updated by a large class of formulæ, namely those satisfying the bounded deterioration assumption defined below. This assumption depends on a particular matrix norm $\|\cdot\|$ possibly different from $|\cdot|$. Note that, since all norms are equivalent in $\mathbb{R}^{N\times N}$, there exists a positive constant η such that

$$\frac{1}{\eta}\|\cdot\| \leq |\cdot| \leq \eta\|\cdot\|. \tag{44}$$

Bounded Deterioration Assumption (BDA). Let there exist a positive constant $C_{\rm BD}$, a symmetric positive definite matrix $\bar{M}$ and a neighborhood $\mathcal{U} = \Omega_x \times \Omega_M$ of $(\bar{x}, \bar{M})$, with Ω_M containing only nonsingular matrices, with the following property. If (x_n, M_n) is in $\mathcal{U}$, if (x_{n+1}, M_{n+1}) is generated by Algorithm (qN-AP3) from (x_n, M_n) and if x_{n+1} is also in Ω_x, then

$$\|M_{n+1} - \bar{M}\| \leq (1 + C_{\rm BD}\sigma_n)\|M_n - \bar{M}\| + C_{\rm BD}\sigma_n, \tag{45}$$

where the matrix norm $\|\cdot\|$ satisfies (44) and σ_n is defined by (27).

This assumption is weaker than the one usually obtainable in standard quasi-Newton methods (see [2]) in the sense that here inequality (45) is only assumed to be satisfied when x_n *and* x_{n+1} are close to $\bar{x}$. Usually no restriction of this type is supposed for (45) to be valid, but when variational quasi-Newton updates *with variable norms* are involved (see Section 3.4), only the above weak form of BDA can be obtained. As far as local convergence is concerned, however, our weaker form suffices: indeed, as shown in Lemma 3.3 (with $r < 1$), once (x_n, M_n) is close enough to $(\bar{x}, \bar{M})$, x_{n+1} is even closer to $\bar{x}$ than x_n.

Conditions for linear convergence are given in the next theorem. We denote by $B(z,\rho)$ the ball of radius $\rho > 0$ centered at z (in a normed space depending on the context).

THEOREM 3.4.

Suppose that (30), (31) *and* (36) *hold and that the update of the matrices in Algorithm* (qN-AP3) *satisfies* (BDA) *with the same matrix* $\bar{M}$ *as in* (36). *Then, for all* $r \in]\bar{r}/(2-\bar{r}), 1[$, *there exist positive constants* ϵ_1 *and* ϵ_2, *such that*

$$|x_1 - \bar{x}| \leq \epsilon_1 \quad \textit{and} \quad |M_1 - \bar{M}| \leq \epsilon_2 \tag{46}$$

imply the following statements:

(*i*) *Algorithm* (qN-AP3) *is well defined in the sense that, for all* $n \geq 1$, M_n *is positive definite and* x_n^p *and* x_n *lie in* Ω_x.

(*ii*) *The sequences* $\{M_n\}$ *and* $\{M_n^{-1}\}$ *are bounded and the sequence* $\{\|M_n - \bar{M}\|\}$ *converges.*

(*iii*) *The sequence* $\{x_n\}$ *converges linearly to* $\bar{x}$ *at rate* r:

$$|x_{n+1} - \bar{x}| \leq r|x_n - \bar{x}|, \quad \forall n \geq 1. \tag{47}$$

Proof. Take $r \in]\bar{r}/(2-\bar{r}), 1[\neq \phi$ and let $\hat{\epsilon}_1 > 0$ and $\hat{\epsilon}_2 > 0$ be given by Lemma 3.3. Then choose $\epsilon_2 > 0$ such that

$$B(\bar{M}, 2\eta\epsilon_2) \subset \Omega_M \quad \text{and} \quad 2\eta^2\epsilon_2 \leq \hat{\epsilon}_2; \tag{48}$$

here η is defined in (44), Ω_M is introduced in (BDA) and $B(\cdot,\cdot)$ refers to $\|\cdot\|$. Next, choose $\epsilon_1 > 0$ such that

$$B(\bar{x}, \epsilon_1) \subset \Omega_x, \quad \epsilon_1 \leq \hat{\epsilon}_1 \quad \text{and} \quad 2C_{\text{BD}}\epsilon_1(2\eta\epsilon_2 + 1)\frac{1}{1-r} \leq \eta\epsilon_2, \tag{49}$$

where Ω_x and C_{BD} were introduced in (BDA).

The positive constants ϵ_1 and ϵ_2 being determined, suppose that (46) holds and let us prove by induction that for all $n \geq 1$:

$$\|M_n - \bar{M}\| \leq 2\eta\epsilon_2, \tag{50}$$

$$|x_{n+1} - \bar{x}| \leq r|x_n - \bar{x}|. \tag{51}$$

First, by (44),

$$\|M_1 - \bar{M}\| \leq \eta|M_1 - \bar{M}| \leq \eta\epsilon_2. \tag{52}$$

Therefore, (50) is satisfied for $n = 1$. As $\epsilon_1 \leq \hat{\epsilon}_1$ and $\epsilon_2 \leq \hat{\epsilon}_2$ (by (48) and $\eta \geq 1$), we have

$$|x_1 - \bar{x}| \leq \hat{\epsilon}_1 \quad \text{and} \quad |M_1 - \bar{M}| \leq \hat{\epsilon}_2.$$

By Lemma 3.3, this implies that the next iterate x_2 is well defined and that (51) holds for $n = 1$.

Now, assume for induction that (50) and (51) are satisfied for $n = 1, \ldots, m-1$. By (51), (46) and (49), the points $x_1, \ldots, x_m$ are in $B(\bar{x}, \epsilon_1) \subset \Omega_x$, and by (50) and (48), the matrices $M_1, \ldots, M_{m-1}$ are in $B(\bar{M}, 2\eta\epsilon_2) \subset \Omega_M$. Therefore, we can use (BDA) for $n = 1, \ldots, m-1$ to obtain:

$$\begin{aligned} \|M_{n+1} - \bar{M}\| - \|M_n - \bar{M}\| &\leq C_{\mathrm{BD}}\sigma_n(\|M_n - \bar{M}\| + 1) \\ &\leq 2C_{\mathrm{BD}}|x_n - \bar{x}|(\|M_n - \bar{M}\| + 1) \\ &\leq 2C_{\mathrm{BD}}r^{n-1}\epsilon_1(2\eta\epsilon_2 + 1), \end{aligned}$$

where we have used (46), (50) and (51). Adding up from 1 to $m-1$ and using (52) and (49), we get

$$\|M_m - \bar{M}\| \leq \|M_1 - \bar{M}\| + 2C_{\mathrm{BD}}\epsilon_1(2\eta\epsilon_2 + 1)\frac{1}{1-r} \leq 2\eta\epsilon_2.$$

This proves (50) for $n = m$. To get (51) for $n = m$, we use as before Lemma 3.3 after having observed that $|x_n - \bar{x}| \leq \hat{\epsilon}_1$ (by (51) and (46)) and $|M_n - \bar{M}| \leq \hat{\epsilon}_2$ (by (50) and (48)). This completes our induction argument.

The boundedness of $\{M_n^{-1}\}$ is given by Lemma 3.3.

Finally, the proof of the convergence of $\{\|M_n - \bar{M}\|\}$ follows a classical scheme. The sequence $\{M_n\}$ being bounded, the sequence $\{\|M_n - \bar{M}\|\}$ has limit points. Then, we proceed by contradiction, supposing that there are two limit points: $l_1 < l_2$. As the series $\sum_{n=1}^{\infty}\sigma_n$ converges, there is an index n_0 such that

$$\sum_{n=n_0}^{\infty}\sigma_n \leq \frac{l_2 - l_1}{3}C_{\mathrm{BD}}^{-1}(2\eta\epsilon_2 + 1)^{-1}.$$

We can also choose an index $n_1 \geq n_0$ such that $\|M_{n_1} - \bar{M}\| \leq l_1 + \frac{l_2 - l_1}{3}$. Then, using (BDA) and (50), we can write, for all $n \geq n_1$

$$\begin{aligned} \|M_n - \bar{M}\| &\leq \|M_{n_1} - \bar{M}\| + C_{\mathrm{BD}}\sum_{i=n_1}^{n-1}\Big(\sigma_i(\|M_i - \bar{M}\| + 1)\Big) \\ &\leq \|M_{n_1} - \bar{M}\| + C_{\mathrm{BD}}(2\eta\epsilon_2 + 1)\sum_{i=n_0}^{\infty}\sigma_i \\ &\leq \|M_{n_1} - \bar{M}\| + \frac{l_2 - l_1}{3} \\ &\leq l_2 - \frac{l_2 - l_1}{3}, \end{aligned}$$

contradicting the fact that l_2 is another limit point. □□

Let us point out that the "implicit" form of (qN-AP3) allows a better rate of convergence than the one obtained in [14] for "standard" quasi-Newton formulæ, namely $r \in]\bar{r}, 1[$.

3.3. Characterization of superlinear convergence. In this subsection we characterize the q-superlinear convergence of a sequence $\{x_n\}$ generated by Algorithm (qN-AP3) by comparing the effect of M_n and $\bar{H}$ on $(-s_n)$. In [14], $\bar{M}$ in (36) is used as an intermediate matrix in the comparison. A similar result can be obtained here but, instead of assuming (36), we prefer to impose more regularity on $\bar{H}$. When $\bar{H}$ is linear, both results are similar to the well-known characterization of Dennis and Moré [6].

LEMMA 3.5. *Let $H : \mathbb{R}^N \mapsto \mathbb{R}^N$ be positively homogeneous, continuous and injective. Then the following properties hold:*

(i) there exists a constant C_H such that $|Hu| \geq C_H|u|$ for all $u \in \mathbb{R}^N$,

(ii) for any two bounded sequences $\{u_n\}$ and $\{v_n\}$ in $\mathbb{R}^N$,

$$Hu_n - Hv_n \to 0 \quad \Longrightarrow \quad u_n - v_n \to 0\,,$$

(iii) if $u_n \to 0$ then

$$H(u_n + o(|u_n|)) = Hu_n + o(|u_n|)\,. \tag{53}$$

Proof.

(i) Let $C_H := \min_{|u|=1} |Hu| \geq 0$; by continuity there exists u_0 of norm 1 such that $|Hu_0| = C_H$. Then the injectivity of H implies $C_H > 0$; the conclusion follows from positive homogeneity.

(ii) Having an arbitrary cluster point w of $\{u_n - v_n\}$, extract a subsequence such that $u_n \to u$, $v_n \to v$ and $u_n - v_n \to w = u - v$. By continuity, $Hu_n \to Hu$, $Hv_n \to Hv$ and by assumption, $Hu = Hv$. Since H is injective, $u = v$, $w = 0$; the result follows.

(iii) If H is continuous, it is uniformly continuous on the ball $B(0,2) \subset \mathbb{R}^N$. When $u_n \neq 0$, $u_n/|u_n| + o(1) \in B(0,2)$ for large n. Hence, by uniform continuity,

$$H\left(\frac{u_n}{|u_n|} + o(1)\right) = H\frac{u_n}{|u_n|} + o(1).$$

Thanks to positive homogeneity, we have proved (53). □□

THEOREM 3.6. *Let $\{x_n\}$ be a sequence generated by the recursion formula (26) converging to a solution point $\bar{x}$. Suppose that (30), (31) hold and that $\bar{H}$ is continuous and injective. Then*

$$x_n \to \bar{x} \quad q\text{-superlinearly} \quad \Longleftrightarrow \quad (M_n - \bar{H})(-s_n) = o(|s_n|). \tag{54}$$

Proof. First, remembering that $s_n = -2M_n^{-1}g_n^p$, we have, due to (34)

$$M_n s_n = -2g_n^p = -2\bar{H}e_n^p + o(|e_n^p|).$$

Hence,

$$(M_n - \bar{H})(-s_n) = 2\bar{H}e_n^p - \bar{H}(-s_n) + o(|e_n^p|). \tag{55}$$

Let us prove the "$\Longrightarrow$" part. As $e_{n+1} = o(|e_n|)$, we have

$$2e_n^p = e_{n+1} + e_n = e_n + o(|e_n|)$$
$$-s_n = -e_{n+1} + e_n = e_n + o(|e_n|).$$

The last estimate also implies that $e_n = O(|s_n|)$. Combining these estimates with (55) and using (53) with $H = \bar{H}$, we get

$$(M_n - \bar{H})(-s_n) = o(|e_n|) = o(|s_n|).$$

Consider now the "$\Longleftarrow$" part. From (55),

$$\bar{H}(-s_n) + o(|s_n|) = \bar{H}(2e_n^p) + o(|e_n^p|). \tag{56}$$

Taking norms and applying Lemma 3.5 (i), we get

$$C_H|s_n| \leq |\bar{H}(-s_n)| \leq |\bar{H}(2e_n^p)| + o(|e_n^p|) + o(|s_n|)\,.$$

Using the boundedness of $\bar{H}$ we conclude

$$|s_n| = O(|e_n^p|). \tag{57}$$

On the other hand, after division of (56) by $|e_n^p|$:

$$\frac{o(|e_n^p|)}{|e_n^p|} + \frac{o(|s_n|)}{|e_n^p|} = \frac{\bar{H}(2e_n^p)}{|e_n^p|} - \frac{\bar{H}(-s_n)}{|e_n^p|}.$$

Thanks to (57), the left-hand side tends to 0. We are in a position to apply Lemma 3.5 (ii) with $u_n = 2e_n^p/|e_n^p|$ and $v_n = -s_n/|e_n^p|$. Thus

$$\frac{2e_n^p + s_n}{|e_n^p|} = \frac{2e_{n+1}}{|e_n^p|} \to 0\,,$$

which can be written $e_{n+1} = o(|e_n^p|) = o(|e_{n+1} + e_n|) = o(|e_{n+1}|) + o(|e_n|)$. This implies $e_{n+1} = o(|e_n|)$ and the q-superlinear convergence of $\{x_n\}$. □□

With this result, the relation corresponding to the classical characterization of [6] can be recovered. Note, incidentally, that the above proof still works for nonsmooth equations (instead of minimization) where g is not a gradient. When assuming more regularity on f, we can also establish a very useful characterization:

COROLLARY 3.7. *Let $\{x_n\}$ be a sequence generated by Algorithm* (qN-AP3) *converging to a solution point $\bar{x}$. Suppose that* (30)-(32) *hold and that $\bar{H}$ is invertible. Then*

$$x_n \to \bar{x} \;\; q\text{-superlinearly} \quad \Longleftrightarrow \quad (M_{n+1} - M_n)s_n = o(|s_n|). \tag{58}$$

Proof. Due to the quasi-Newton equation (25), the second statement in (58) is equivalent to $y_n - M_n s_n = o(|s_n|)$. Apply (32) with $x = x_{n+1}$, $y = x_n$: we have $y_n = \bar{H}s_n + o(|s_n|)$; since (32) also implies the linearity of $\bar{H}$, the conclusion follows from Theorem 3.6. □□

3.4. Superlinear convergence of variational quasi-Newton algorithms. In this subsection, we go more concretely into the specification of the matrices M_n for Algorithm (qN-AP3). We propose an update scheme and show (Lemma 3.8) that it satisfies (BDA) in Section 3.2. Then, the linear convergence follows from Theorem 3.4 (Theorem 3.9). We also show (Theorem 3.10) that the scheme can provide the q-superlinear convergence of the generated sequences.

The analysis relies on a Hilbert matrix norm $|\cdot|_n$ (e.g., a weighted Frobenius norm); typically $|\cdot|_n$ depends on x_n and x_{n+1}. With σ_n defined in (27), the norm $|\cdot|_n$ is said *locally comparable* to a fixed norm $\|\cdot\|$ if

$$\begin{aligned} &\exists \sigma_{\rm LC} > 0,\ \exists C_{\rm LC} > 0,\ \forall \sigma_n \le \sigma_{\rm LC},\ \forall M \in \mathbb{R}^{N\times N},\ \text{we have} \\ &\Big|\, |M|_n - \|M\| \,\Big| \le C_{\rm LC}\, \|M\|\, \sigma_n. \end{aligned} \tag{59}$$

Our approach follows that of [12]. Let $\mathcal{K}$ be a closed convex set of symmetric matrices intersecting the set $\{M \in \mathbb{R}^{N\times N} : Ms_n = y_n\}$, when σ_n is small. By a *variational quasi-Newton formula*, we mean a method associating to the current matrix M_n the (symmetric) update $M_{n+1} := \mathrm{qN}(M_n, y_n, s_n)$, unique solution of

$$\min_M \left\{ |M - M_n|_n^2 : M \in \mathcal{K},\ Ms_n = y_n \right\}. \tag{60}$$

We state here a "technical hypothesis" expressing that a fixed matrix $\bar{M}$ is close enough to the feasible set of (60):

$$\begin{aligned} &\exists \bar{M} \in \mathbb{R}^{N\times N} \text{ symmetric positive definite},\ \exists \sigma_{\rm TEX} > 0,\ \exists C_{\rm TEX} > 0, \\ &\forall \sigma_n \le \sigma_{\rm TEX},\ \exists \hat{M}_n \in \mathbb{R}^{N\times N}, \text{ such that} \\ &\hat{M}_n \in \mathcal{K},\ \hat{M}_n s_n = y_n,\ |\hat{M}_n - \bar{M}|_n \le C_{\rm TEX}\, \sigma_n. \end{aligned} \tag{61}$$

Before giving the convergence theorems, let us check that (BDA) is satisfied for the scheme above.

LEMMA 3.8. *Suppose that Algorithm* (qN-AP3) *updates the matrices M_n according to the scheme* (60) *and that conditions* (59) *and* (61) *are satisfied. Then Assumption* (BDA) *holds with $\bar{M}$ given by* (61).

Proof. Let

$$\sigma := \min(\sigma_{\rm LC}, \sigma_{\rm TEX}, \frac{1}{3C_{\rm LC}}).$$

Since M_{n+1} is the orthogonal projection of M_n onto a closed convex set containing $\hat{M}_n$, we have

$$|M_n - M_{n+1}|_n^2 + |M_{n+1} - \hat{M}_n|_n^2 \le |M_n - \hat{M}_n|_n^2. \tag{62}$$

In particular,

$$|M_{n+1} - \hat{M}_n|_n \le |M_n - \hat{M}_n|_n. \tag{63}$$

Let us show that (BDA) holds with $C_{\text{BD}} := 3\max(C_{\text{LC}}, C_{\text{TEX}})$ and $\bar{M}$ given by (61), when $\sigma_n \leq \sigma$. We have, using (63) and (61),

$$\begin{aligned} |M_{n+1} - \bar{M}|_n &\leq |M_{n+1} - \hat{M}_n|_n + |\hat{M}_n - \bar{M}|_n \\ &\leq |M_n - \hat{M}_n|_n + C_{\text{TEX}}\sigma_n \\ &\leq |M_n - \bar{M}|_n + 2C_{\text{TEX}}\sigma_n. \end{aligned}$$

Then, using (59), we get

$$\begin{aligned} (1 - C_{\text{LC}}\sigma_n)\|M_{n+1} - \bar{M}\| &\leq (1 + C_{\text{LC}}\sigma_n)\|M_n - \bar{M}\| + 2C_{\text{TEX}}\sigma_n, \\ \|M_{n+1} - \bar{M}\| &\leq \left(\frac{1 + C_{\text{LC}}\sigma_n}{1 - C_{\text{LC}}\sigma_n}\right)\|M_n - \bar{M}\| + \left(\frac{2C_{\text{TEX}}}{1 - C_{\text{LC}}\sigma_n}\right)\sigma_n. \end{aligned}$$

Since $\sigma_n \leq \sigma \leq 1/(3C_{\text{LC}})$:

$$\frac{1 + C_{\text{LC}}\sigma_n}{1 - C_{\text{LC}}\sigma_n} \leq 1 + 3C_{\text{LC}}\sigma_n \quad \text{and} \quad \frac{2C_{\text{TEX}}}{1 - C_{\text{LC}}\sigma_n} \leq 3C_{\text{TEX}}.$$

Hence

$$\|M_{n+1} - \bar{M}\| \leq (1 + 3C_{\text{LC}}\sigma_n)\|M_n - \bar{M}\| + 3C_{\text{TEX}}\sigma_n,$$

which is just a bounded deterioration property of the type (BDA). □□

Then, we can show linear convergence under the assumptions of Theorem 3.9 and superlinear convergence when (32) holds (Theorem 3.10).

THEOREM 3.9. *Suppose that* (30), (31) *and* (36) *hold. Suppose also that Algorithm* (qN-AP3) *updates the matrices* M_n *according to the scheme* (60) *and that conditions* (59) *and* (61) *hold, the latter with the same matrix* $\bar{M}$ *as in* (36). *Then, if* (x_1, M_1) *is close enough to* $(\bar{x}, \bar{M})$, *Algorithm* (qN-AP3) *is well defined and generates a sequence* $\{x_n\}$ *converging q-linearly to* $\bar{x}$ *and a sequence of symmetric positive definite matrices* $\{M_n\}$ *such that*

$$(M_{n+1} - M_n) \to 0. \tag{64}$$

Proof. According to Lemma 3.8, (BDA) is satisfied with the same matrix $\bar{M}$ as in (36). Then, Theorem 3.4 gives the first part of the result (the linear convergence of the sequence $\{x_n\}$), as well as

$$\|M_n - \bar{M}\| \to \delta. \tag{65}$$

It remains to prove (64).

Due to the linear convergence of $\{x_n\}$ to $\bar{x}$, we can suppose that $\sigma_n \leq \min(\sigma_{\text{LC}}, \sigma_{\text{TEX}})$ for all $n \geq 1$. As in the proof of Lemma 3.8, we have the inequality

$$|M_n - M_{n+1}|_n^2 + |M_{n+1} - \hat{M}_n|_n^2 \leq |M_n - \hat{M}_n|_n^2, \tag{66}$$

and we proceed to show that both $|M_{n+1} - \hat{M}_n|_n$ and $|M_n - \hat{M}_n|_n$ tend to δ. From (59), (65) implies

$$|M_n - \bar{M}|_n \to \delta \quad \text{and} \quad |M_{n+1} - \bar{M}|_n \to \delta. \tag{67}$$

Using (61), we get

$$\Big| |M_n - \hat{M}_n|_n - |M_n - \bar{M}|_n \Big| \leq |\hat{M}_n - \bar{M}|_n \to 0,$$
$$\Big| |M_{n+1} - \hat{M}_n|_n - |M_{n+1} - \bar{M}|_n \Big| \leq |\hat{M}_n - \bar{M}|_n \to 0.$$

From this and (67), we deduce

$$|M_n - \hat{M}_n|_n \to \delta \quad \text{and} \quad |M_{n+1} - \hat{M}_n|_n \to \delta.$$

Then, (66) implies

$$|M_{n+1} - M_n|_n \to 0$$

and by (59),

$$\|M_{n+1} - M_n\| \to 0.$$

□□

THEOREM 3.10. *Suppose that* (30)-(32) *hold and that* $\bar{H}$ *is positive definite. Suppose also that Algorithm* (qN-AP3) *updates the matrices* M_n *according to the scheme* (60) *and that conditions* (59) *and* (61) *hold, the latter with* $\bar{M} = \bar{H}$*. Then, if* (x_1, M_1) *is close enough to* $(\bar{x}, \bar{H})$*, Algorithm* (qN-AP3) *is well defined and generates a sequence* $\{x_n\}$ *converging q-superlinearly to* $\bar{x}$*.*

Proof. Assumption (32) implies that $\bar{H}$ is linear, hence (36) holds with $\bar{M} = \bar{H}$; we can apply then Theorem 3.9, which gives the q-linear convergence of the sequence $\{x_n\}$ and $(M_{n+1} - M_n) \to 0$. Now the q-superlinear convergence of $\{x_n\}$ follows from Corollary 3.7. □□

3.5. Application to some quasi-Newton methods. We now apply the theory of the previous subsection to some particular quasi-Newton update formulæ. The main issue is to check condition (61), and it is here that assumption (33) comes into play.

We first show that (61) holds for general quasi-Newton methods, provided f is sufficiently smooth. As in the previous subsection, $\mathcal{K}$ is a general closed convex set of symmetric matrices.

PROPOSITION 3.11. *Suppose that* f *is twice Fréchet differentiable in a neighborhood* $\mathcal{N}$ *of* $\bar{x}$*, with a Lipschitz continuous Hessian. If* $\nabla^2 f(x) \in \mathcal{K}$ *for all* $x \in \mathcal{N}$ *and* (59) *holds, then* (61) *is satisfied for* $\bar{M} = \nabla^2 f(\bar{x})$*.*

Proof. Let $\sigma_{\text{LC}} > 0$ be given by (59) and take $\sigma \in]0, \sigma_{\text{LC}}]$ such that $B(\bar{x}, \sigma) \subset \mathcal{N}$. When $\sigma_n < \sigma$, the segment $[x_n, x_{n+1}]$ is in $\mathcal{N}$, so that we can define

$$\hat{M}_n := \int_0^1 \nabla^2 f(x_n + \tau s_n) d\tau.$$

Clearly, $\hat{M}_n \in \mathcal{K}$ and $\hat{M}_n s_n = y_n$. Furthermore, with $\bar{M} = \nabla^2 f(\bar{x})$,

$$|\hat{M}_n - \bar{M}| \leq \int_0^1 |\nabla^2 f(x_n + \tau s_n) - \bar{M}| d\tau \leq \frac{L_H}{2} \sigma_n,$$

where L_H is a Lipschitz constant of the map $\mathcal{N} \ni x \mapsto \nabla^2 f(x)$. Combine this, (44) and (59) to obtain

$$|\hat{M}_n - \bar{M}|_n \leq (1 + C_{\text{LC}}\sigma_n)\|\hat{M}_n - \bar{M}\| \leq \eta(1 + C_{\text{LC}}\sigma_n)|\hat{M}_n - \bar{M}| \leq \eta(1 + C_{\text{LC}}\sigma_{\text{LC}})\frac{L_H}{2}\,\sigma_n.$$

We recognize (61). □□

We consider now the prox-versions of the PSB and DFP algorithms. Let $\mathcal{K}$ be the set of symmetric matrices and take the Frobenius norm $|\cdot|_F$ for $|\cdot|_n$ and $\|\cdot\|$. Then the solution of (60) is given by the PSB update formula (see [7]): $M_{n+1} = \text{PSB}(M_n, y_n, s_n)$, where

$$\text{PSB}(M, y, s) := M + \frac{(y - Ms)s^{\text{T}} + s(y - Ms)^{\text{T}}}{|s|^2} - \frac{\langle y - Ms, s\rangle}{|s|^4}ss^{\text{T}}.$$

Recall that $\langle u, v\rangle$ and $u^{\text{T}}v$ denote the same operation. We note here that more general scalar products can also be used, as described for example in [11] and in the appendix of [10]. Reproducing the present theory in this framework is then an easy exercise.

For this method, we have

PROPOSITION 3.12. *Suppose that* (30)-(33) *hold and that* $\bar{H}$ *is positive definite. Assume that Algorithm* (qN-AP3) *uses the PSB formula:* $M_{n+1} = \text{PSB}(M_n, y_n, s_n)$. *If* (x_1, M_1) *is close enough to* $(\bar{x}, \bar{H})$, *then the algorithm is well defined and* $x_n \to \bar{x}$ *q-superlinearly.*

Proof. Take

$$\hat{M}_n := \text{PSB}(\bar{H}, y_n, s_n)$$

and define $\delta_n := y_n - \bar{H}s_n$. Then

$$\hat{M}_n - \bar{H} = \frac{\delta_n s_n^{\text{T}} + s_n\delta_n^{\text{T}}}{|s_n|^2} - \frac{\langle \delta_n, s_n\rangle}{|s_n|^4}s_n s_n^{\text{T}}.$$

Taking $x = x_n$ and $y = x_{n+1}$ in (33), we obtain $\delta_n = O(|s_n|\,|\sigma_n|)$. Recall also that $|uv^{\text{T}}| = |u|\,|v|$. Therefore

$$|\hat{M}_n - \bar{H}| = O(|\sigma_n|).$$

On the other hand, since $\hat{M}_n \in \mathcal{K}$ and $\hat{M}_n s_n = y_n$, condition (61) holds with $|\cdot|_n = |\cdot|_F$ and $\bar{M} = \bar{H}$. We can now apply Theorem 3.10 to terminate the proof. □□

Consider now the DFP formula ([7]):

$$\text{DFP}(M, y, s) := M + \frac{(y - Ms)y^{\text{T}} + y(y - Ms)^{\text{T}}}{\langle y, s\rangle} - \frac{\langle y - Ms, s\rangle}{\langle y, s\rangle^2}yy^{\text{T}}.$$

This formula is well defined when $\langle y, s\rangle \neq 0$ and gives a symmetric positive definite matrix when M is itself symmetric positive definite and $\langle y, s\rangle > 0$. The updated matrix can be characterized as the solution of a variational problem. For this, let us introduce the weighted Frobenius norm associated to a symmetric positive definite matrix W:

$$M \mapsto |M|_{W,F} := |W^{-1/2}MW^{-1/2}|_F.$$

Then, when $\langle y_n, s_n\rangle$ is positive, $\mathrm{DFP}(M_n, y_n, s_n)$ is the solution of problem (60) in which $\mathcal{K}$ is the set of symmetric matrices and $|\cdot|_n$ is the norm $|\cdot|_{W_n,F}$ where W_n is any matrix satisfying $W_n s_n = y_n$ (see [7]). As we shall see in the proof of the next proposition, an appropriate choice of the matrix W_n will allow us to satisfy (59) and (61).

PROPOSITION 3.13. *Suppose that* (30)-(33) *hold and that* $\bar{H}$ *is positive definite. Assume that Algorithm* (qN-AP3) *uses the DFP formula:* $M_{n+1} = \mathrm{DFP}(M_n, y_n, s_n)$. *If* (x_1, M_1) *is close enough to* $(\bar{x}, \bar{H})$, *then the algorithm is well defined and* $x_n \to \bar{x}$ *q-superlinearly.*

Proof. Because $\bar{H}$ is positive definite, it is easy to see that when $\sigma_n := |x_n - \bar{x}| + |x_{n+1} - \bar{x}|$ is sufficiently small, we have

$$\langle y_n, s_n\rangle \geq \alpha |s_n|^2 \quad \text{and} \quad |y_n| \leq L|s_n|, \tag{68}$$

for some positive constants α and L. From now on, we suppose that σ_n is sufficiently small to have (68).

The matrix

$$\hat{M}_n := \mathrm{DFP}(\bar{H}, y_n, s_n)$$

is positive definite and verifies $\hat{M}_n s_n = y_n$. Then M_{n+1} is solution of (60) with $|\cdot|_n := |\cdot|_{\hat{M}_n,F}$.

Defining $\delta_n := y_n - \bar{H} s_n$, we have

$$\hat{M}_n - \bar{H} = \frac{\delta_n y_n^{\mathrm{T}} + y_n \delta_n^{\mathrm{T}}}{\langle y_n, s_n\rangle} - \frac{\langle \delta_n, s_n\rangle}{\langle y_n, s_n\rangle^2} y_n y_n^{\mathrm{T}}.$$

By (33), $\delta_n = O(|s_n|\,|\sigma_n|)$. Therefore, using (68),

$$|\hat{M}_n - \bar{H}| = O(|\sigma_n|). \tag{69}$$

It follows that $\hat{M}_n^{-1/2}$ is bounded for σ_n small enough, then

$$|\hat{M}_n - \bar{H}|_n = |\hat{M}_n^{-1/2}(\hat{M}_n - \bar{H})\hat{M}_n^{-1/2}|_F = O(|\hat{M}_n - \bar{H}|).$$

Since $\hat{M}_n$ is symmetric and $\hat{M}_n s_n = y_n$, condition (61) holds with $\bar{M} = \bar{H}$.

Let us now prove condition (59) with $\|\cdot\| = |\cdot|_{\bar{H},F}$. Observe that

$$|M|_{W,F} = \Big(\mathrm{tr}(W^{-1/2} M W^{-1} M W^{-1/2})\Big)^{1/2} = \Big(\mathrm{tr}(M W^{-1})^2\Big)^{1/2}.$$

Then, for $M \in \mathbb{R}^{N\times N}$ with $\|M\| = 1$,

$$\Big||M|_n - \|M\|\Big| = \frac{\Big||M|_n^2 - \|M\|^2\Big|}{|M|_n + \|M\|} \leq \Big|\mathrm{tr}(M\hat{M}_n^{-1})^2 - \mathrm{tr}(M\bar{H}^{-1})^2\Big|,$$

because $|M|_n + \|M\| \geq 1$. Now, $A \in \mathbb{R}^{N\times N} \mapsto \mathrm{tr}\, A$ is linear. Therefore, for some constant $C_1 > 0$,

$$\Big||M|_n - \|M\|\Big| \leq C_1 |(M\hat{M}_n^{-1})^2 - (M\bar{H}^{-1})^2|.$$

Using the relation $|B^2 - A^2| = |B(B-A) + (B-A)A| \leq (|A| + |B|)|B - A|$, we get

$$\Big| |M|_n - \|M\| \Big| \leq C_1 |M| \Big(|\hat{M}_n^{-1}| + |\bar{H}^{-1}| \Big) |\hat{M}_n^{-1} - \bar{H}^{-1}|.$$

Because the norms $|\cdot|$ and $\|\cdot\|$ are equivalent and $A \mapsto A^{-1}$ is infinitely differentiable on the set of nonsingular matrices, one has for σ_n sufficiently small

$$\Big| |M|_n - \|M\| \Big| = O(|\sigma_n|),$$

where we used (69). Now condition (59) holds by homogeneity in M.

The conclusion of the theorem follows from Theorem 3.10. □□

4. A BFGS-Proximal Method. In this section, we study the particularization of the algorithmic pattern (AP3), in which the proximal point x_n^p is computed exactly and the BFGS formula is used to update the matrices M_n. In this case, satisfactory global and q-superlinear convergence results can be obtained, in the sense that, given *any* initial pair (x_1, M_1), with M_1 symmetric and positive definite, the generated sequence $\{x_n\}$ converges superlinearly to a solution of problem (1). The precise results are given in Theorems 4.2 and 4.8 below.

To obtain these convergence results, f is always supposed differentiable (and therefore finite everywhere). Then we will again use the notation $g(x)$ for the gradient of f at x, as well as $g_n = g(x_n)$ and $g_n^p = g(x_n^p)$.

For given vectors s and y in $\mathbb{R}^N$, the BFGS update of an $N \times N$ symmetric matrix M is the matrix

$$\text{BFGS}(M, y, s) := M - \frac{Mss^{\mathrm{T}}M}{\langle Ms, s\rangle} + \frac{yy^{\mathrm{T}}}{\langle y, s\rangle} \tag{70}$$

(see [7] for instance). Observe that the trace of the matrix $M_+ = \text{BFGS}(M, y, s)$ is given by

$$\operatorname{tr} M_+ = \operatorname{tr} M - \frac{|Ms|^2}{\langle Ms, s\rangle} + \frac{|y|^2}{\langle y, s\rangle}. \tag{71}$$

When M is positive definite, the BFGS formula is well defined if $\langle y, s\rangle \neq 0$. However, the stronger condition

$$\langle y, s\rangle > 0$$

is generally required since this is a necessary and sufficient condition to have the updated matrix positive definite.

The algorithm considered in this section is stated as follows:

- **BFGS-proximal algorithm (BFGS-AP3):**

STEP 0: Choose an initial point $x_1 \in \mathbb{R}^N$ and an initial symmetric positive definite matrix M_1. Take m in $]0, 1[$. Set $n = 1$.

STEP 1: Given x_n and M_n, compute $x_n^p := p_{M_n}(x_n)$ and set $s_n^p := x_n^p - x_n$.

STEP 2: Compute the next iterate by:

$$x_{n+1} := x_n + t_n s_n^p.$$

The stepsize $t_n \geq 1$ is chosen to satisfy the general descent condition (18) and

$$\langle y_n, s_n \rangle > 0, \tag{72}$$

where $s_n = x_{n+1} - x_n$ and $y_n = g_{n+1} - g_n$. We also suppose that $t_n = 2$ is taken when the line-search conditions (18) and (72) allow it.

STEP 3: Update M_n by the BFGS formula:

$$M_{n+1} = \mathrm{BFGS}(M_n, y_n, s_n).$$

Increase n by 1 and loop to Step 1. □

In Step 2, the additional condition (72) is only required to guarantee the well posedness of the BFGS formula and the positive definiteness of the generated matrices. Note also that from Section 3 it is important to take $t_n = 2$ whenever possible for the sake of superlinear convergence. Step 2 is actually a line-search generating trial stepsizes $t \geq 1$ until (18) and (72) are simultaneously satisfied.

REMARK 4.1. Feasibility of this line-search is easy to establish. First of all, the requirement $t \geq 1$ is not classical but x_n^p, obtained for $t = 1$, satisfies the descent test (18) with a strict inequality. Then, by convexity of f, the stepsizes that satisfy (18) form a closed interval, say $\mathcal{I}_1$, containing 1 in its interior. As for (72), remark that the function

$$0 \leq t \mapsto \langle g(x_n + t s_n^p) - g_n, s_n^p \rangle =: d(t)$$

is nonnegative and nondecreasing and cannot be identically zero when f is bounded below in the direction s_n^p. This implies that the stepsizes satisfying (72) form an open interval $\mathcal{I}_2 =]t^a, +\infty[$, with finite t^a. We have to show that $\mathcal{I}_1$ and $\mathcal{I}_2$ intersect. There are 2 cases:

1. If $t^a < 1$, $\mathcal{I}_1 \cap \mathcal{I}_2$ contains some neighborhood of 1.
2. If $1 \leq t^a < +\infty$, the key is to observe that $f(x_n + t s_n^p)$ has the constant slope $\langle g_n^p, s_n^p \rangle = -\langle M_n s_n^p, s_n^p \rangle$ at any $t \in [0, t^a]$ (recall (11)). Hence

$$\langle M_n s_n^p, s_n^p \rangle = f(x_n) - f(x_n^p) \geq \delta_n.$$

Then, since $t^a \geq 1 > m$, we can write

$$f(x_n + t^a s_n^p) = f(x_n) - t^a \langle M_n s_n^p, s_n^p \rangle < f(x_n) - m\delta_n.$$

Thus, there is $\varepsilon > 0$ such that any stepsize in $]t^a, t^a + \varepsilon]$ satisfies (18) and (72).

Exploiting these properties, the line-search can then be implemented by a simple bracketing algorithm as in [17]. Start from $t = 2$ and, at the current trial stepsize $t \geq 1$,

(i) perform the descent test; if it is not satisfied, t is too large, compute a smaller t;

(ii) if satisfied, test "$d(t) > 0$"; if yes, we are done; otherwise t is too small, compute a larger t.

4.1. Global convergence. Our global convergence result is a simple consequence of Theorem 2.3.

THEOREM 4.2. *Assume that the convex function f has a nonempty bounded set of minima and that its gradient mapping is locally Lipschitz continuous. Let $\{x_n\}$ be the sequence generated by Algorithm* (BFGS-AP3). *Then, all the accumulation points of $\{x_n\}$ and $\{x_n^p\}$ minimize f.*

Proof. In view of Theorem 2.3, we only have to prove (23). Let L be a Lipschitz constant for g on the set $\{x : f(x) \leq f(x_1)\}$ which, as already seen in the proof of Theorem 2.3, is compact. Applying for example [25] or Theorem X.4.2.2 of [13], we obtain

$$L \langle y_n, s_n \rangle \geq |y_n|^2,$$

and the trace relation (71) gives

$$\operatorname{tr} M_{n+1} \leq \operatorname{tr} M_n + L \leq \operatorname{tr} M_1 + nL \leq (n+1)C,$$

where $C := \max(\operatorname{tr} M_1, L)$.

As the largest eigenvalue is less than the trace, we get

$$\lambda_{\min}(M_n^{-1}) = \frac{1}{\lambda_{\max}(M_n)} \geq \frac{1}{\operatorname{tr} M_n} \geq \frac{1}{nC}.$$

Therefore, the convergence condition (23) holds and the result follows. □ ▯

4.2. The r-linear convergence. To prove superlinear convergence, it is known that a technically useful property is the r-linear convergence. This last property, interesting *per se*, can be established for (BFGS-AP3) under rather mild assumptions on f. We start with a result of general interest in convex analysis.

LEMMA 4.3. *Assume that the convex function f is differentiable. With $\bar{x}$ minimizing f, let $\alpha > 0$ and $x \in \mathbb{R}^N$ satisfy*

$$f(x) \geq f(\bar{x}) + \alpha |x - \bar{x}|^2. \tag{73}$$

Then

$$f(x) \leq f(\bar{x}) + (1/\alpha)|g(x)|^2. \tag{74}$$

Proof. Write the subgradient inequality at x and obtain with the Cauchy-Schwarz inequality

$$f(x) \leq f(\bar{x}) + |g(x)|\,|\bar{x} - x|,$$

so that with (73) and the nonnegativity of $f(x) - f(\bar{x})$,

$$f(x) - f(\bar{x}) \leq |g(x)| \sqrt{[f(x) - f(\bar{x})]/\alpha}.$$

The result follows. □▯

The next lemma is part of the theory of BFGS updates and can be stated independently of the present framework. We denote by θ_n the angle between $M_n s_n$ and s_n:

$$\cos\theta_n := \frac{\langle M_n s_n, s_n\rangle}{|M_n s_n||s_n|} = \frac{\langle M_n s_n^p, s_n^p\rangle}{|M_n s_n^p||s_n^p|},$$

and by $\lceil\cdot\rceil$ the roundup operator: $\lceil x\rceil = i$, when $i-1 < x \le i$ and $i \in \mathbb{N}$.

LEMMA 4.4. *Let $\{M_n\}$ be generated by the BFGS formula using pairs of vectors (y_n, s_n) satisfying*

$$\langle y_n, s_n\rangle \ge \alpha_1 |s_n|^2 \quad \textit{and} \quad \langle y_n, s_n\rangle \ge \alpha_2 |y_n|^2 \tag{75}$$

for all $n \ge 1$, where $\alpha_1 > 0$ and $\alpha_2 > 0$ are independent of n. Then for any $r \in]0,1[$, there exist positive constants γ_1 and γ_2, such that

$$\cos\theta_j \ge \gamma_1, \tag{76}$$

$$\frac{|M_j s_j|}{|s_j|} \le \gamma_2, \tag{77}$$

for at least $\lceil rn\rceil$ indices j in $\{1,\dots,n\}$.

Condition (76) on $\cos\theta_j$ was proved by [25], when the BFGS update is used for unconstrained problems with the Wolfe line-search. Byrd and Nocedal [3] showed that this result is true independently of any line-search: it can be stated, as above, only in terms of the updated matrices M_n and the vectors y_n and s_n. We found condition (77) also in [3].

We recall that the differentiable function f is said *strongly convex* on a domain $D \subset \mathbb{R}^N$, if it satisfies the equivalent properties for some $\alpha > 0$ (see [13] Theorem VI.6.1.2):

$$f(y) \ge f(x) + \langle g(x), y-x\rangle + \frac{\alpha}{2}|y-x|^2, \quad \text{for all } x, y \in D,$$

$$\langle g(y) - g(x), y-x\rangle \ge \alpha |y-x|^2, \quad \text{for all } x, y \in D.$$

THEOREM 4.5. *Assume that $\{x_n\}$ converges to a minimum point $\bar{x}$, in the neighborhood of which f is strongly convex and has a Lipschitz continuous gradient mapping. Then the convergence of $\{x_n\}$ is r-linear; this implies in particular that $\sum_{n\ge1} |x_n - \bar{x}| < \infty$.*

Proof. Since this is an asymptotic statement, we limit our attention to large enough n in all the proof below. The Lipschitz property of g ensures the second condition in (75) (see again [25]). The first one, as well as the growth condition (73), are ensured by strong convexity (i.e., strong monotonicity of the gradient mapping). Then our proof is based on an over-estimation of $f(x_n) - f(\bar{x})$ and begins by over-estimating $f^p(x_n) - f(\bar{x})$.

Inequality (17) gives $\langle M_n s_n^p, s_n^p\rangle/2 \le \delta_n = f(x_n) - f^p(x_n)$, so that

$$f^p(x_n) - f(\bar{x}) \le f(x_n) - f(\bar{x}) - \frac{1}{2}\langle M_n s_n^p, s_n^p\rangle. \tag{78}$$

To obtain an over-estimation of $f^p(x_n) - f(\bar{x})$, we under-estimate $\langle M_n s_n^p, s_n^p \rangle$, first in terms of $|g_n^p|^2$ and next in terms of $f^p(x_n) - f(\bar{x})$, using (74).

We start from

$$\langle M_n s_n^p, s_n^p \rangle = |M_n s_n^p||s_n^p| \cos\theta_n, \quad \text{for all } n \geq 1.$$

Fixing r in $]0,1[$, we denote by N_r^n the set of indices j in $\{1,\ldots,n\}$ for which (76) and (77) hold. Using successively (76), (77) and (74), and remembering from Lemma 2.1 that $g_n^p = -M_n s_n^p$, we write for all $j \in N_r^n$

$$\left\langle M_j s_j^p, s_j^p \right\rangle \geq \gamma_1 |M_j s_j^p||s_j^p| \geq \frac{\gamma_1}{\gamma_2} |M_j s_j^p|^2 = \frac{\gamma_1}{\gamma_2} |g_j^p|^2 \geq C_1 (f(x_j^p) - f(\bar{x})),$$

where $C_1 = \alpha\gamma_1/\gamma_2$. Adding $(C_1/2)\left\langle M_j s_j^p, s_j^p \right\rangle$ to the extreme sides and using (15) give

$$\left(1 + \frac{C_1}{2}\right) \left\langle M_j s_j^p, s_j^p \right\rangle \geq C_1 \left(f^p(x_j) - f(\bar{x})\right), \quad \text{for all } j \in N_r^n,$$

so that, as wished,

$$\frac{1}{2} \left\langle M_j s_j^p, s_j^p \right\rangle \geq C_2 \left(f^p(x_j) - f(\bar{x})\right), \quad \text{for all } j \in N_r^n,$$

where $C_2 = C_1/(2 + C_1)$. Combining this with (78) gives

$$f^p(x_j) - f(\bar{x}) \leq \left(\frac{1}{1 + C_2}\right) \left(f(x_j) - f(\bar{x})\right), \quad \text{for all } j \in N_r^n.$$

Now, using the line-search condition (18), we have for $j \in N_r^n$

$$\begin{aligned} f(x_{j+1}) - f(\bar{x}) &\leq (1 - m)\left(f(x_j) - f(\bar{x})\right) + m\left(f^p(x_j) - f(\bar{x})\right) \\ &\leq \left(1 - \frac{mC_2}{1 + C_2}\right)\left(f(x_j) - f(\bar{x})\right). \end{aligned}$$

Remark that we can write $1 - mC_2/(1 + C_2) =: \tau^{1/r}$ for some τ in $]0,1[$. Furthermore, as $|N_r^n| \geq rn$ (Lemma 4.4) and $f(x_{j+1}) - f(\bar{x}) \leq f(x_j) - f(\bar{x})$ for all j, we have

$$f(x_{n+1}) - f(\bar{x}) \leq \tau^{|N_r^n|/r}(f(x_1) - f(\bar{x})) \leq \tau^n (f(x_1) - f(\bar{x})), \quad \text{for all } n \geq 1.$$

Finally (73) allows us to deduce

$$|x_n - \bar{x}| \leq \left[\frac{f(x_n) - f(\bar{x})}{\alpha}\right]^{1/2} \leq \left[\frac{f(x_1) - f(\bar{x})}{\alpha}\right]^{1/2} (\sqrt{\tau})^{n-1}.$$

This implies that $\limsup_{n\to\infty} |x_n - \bar{x}|^{1/n} \leq \sqrt{\tau} < 1$, characterizing the r-linear convergence of x_n to $\bar{x}$. Finiteness of $\sum_{n\geq 1} |x_n - \bar{x}|$ follows. □□

4.3. Acceptability of the ideal stepsize. An important point for fast convergence is whether the stepsize $t_n = 2$ is accepted asymptotically by the line-search conditions (18) and (72). For this, and in particular for the descent condition (18), the candidate

$$x_n^+ := x_n + 2s_n^p \tag{79}$$

must be "superlinearly closer" to the minimum point $\bar{x}$ than x_n. This is the last condition involved in the next result.

THEOREM 4.6. *Assume that $\bar{x}$ is a minimum point of f at which* (30) *and* (31) *hold, and such that the directional-derivative operator $\bar{H}$ of g satisfies the following property:*

$$\exists \alpha > 0 \ \text{ such that } \ \langle \bar{H}z, z\rangle \geq \alpha |z|^2 \ \text{ for all } \ z \in \mathbb{R}^N. \tag{80}$$

If

$$|x_n^+ - \bar{x}| = o(|x_n - \bar{x}|), \tag{81}$$

then the point x_n^+ of (79) *is accepted by the line-search of Algorithm* (BFGS-AP3) *for n large enough.*

Proof. From (34), we have for z arbitrary in the neighborhood of $\bar{x}$:

$$g(z) = \bar{H}(z - \bar{x}) + o(|z - \bar{x}|), \tag{82}$$

so that in particular,

$$\langle g_n^p, s_n^p\rangle = \langle \bar{H}e_n^p, s_n^p\rangle + o(|e_n^p|\,|s_n^p|).$$

For n large enough, we write (82) with $z = \bar{x} + \tau(x_n - \bar{x})$; we multiply by $x_n - \bar{x}$ and we integrate from $\tau = 0$ to $\tau = 1$:

$$f(x_n) = f(\bar{x}) + \frac{1}{2}\langle \bar{H}e_n, e_n\rangle + o(|e_n|^2).$$

The same operation with x_n^p instead of x_n gives

$$f(x_n^p) = f(\bar{x}) + \frac{1}{2}\langle \bar{H}e_n^p, e_n^p\rangle + o(|e_n^p|^2).$$

These three relations give an estimate of $\delta_n = f(x_n) - f(x_n^p) + \frac{1}{2}\langle g_n^p, s_n^p\rangle$:

$$\delta_n = \frac{1}{2}\langle \bar{H}e_n, e_n\rangle - \frac{1}{2}\langle \bar{H}e_n^p, e_n\rangle + o(|e_n|^2),$$

where we have used (81): s_n^p and e_n^p have the order of magnitude of e_n. In the second term, use the relation

$$e_n = 2e_n^p - (x_n^+ - \bar{x}) = 2e_n^p + o(|e_n|)$$

to obtain

$$\delta_n = \frac{1}{2}\langle \bar{H}e_n, e_n\rangle - \langle \bar{H}e_n^p, e_n^p\rangle + o(|e_n|^2).$$

In summary, we have the following estimate for the right-hand side in (18):

$$\begin{aligned} f(x_n) - m\delta_n &= f(\bar{x}) + \frac{1-m}{2}\langle \bar{H}e_n, e_n\rangle + m\langle \bar{H}e_n^p, e_n^p\rangle + o(|e_n|^2) \\ &\geq f(\bar{x}) + \frac{1-m}{2}\alpha|e_n|^2 + o(|e_n|^2), \end{aligned}$$

because $m \in]0,1[$. On the other hand, (82) can again be used to obtain the estimate (we set $e_n^+ := x_n^+ - \bar{x}$)

$$f(x_n^+) = f(\bar{x}) + \frac{1}{2}\langle \bar{H}e_n^+, e_n^+\rangle + o(|e_n^+|^2) = f(\bar{x}) + o(|e_n|^2).$$

Because $(1-m)\alpha/2 > 0$, we conclude that our q-superlinear assumption ensures that (18) is eventually satisfied.

It remains to take care of (72). From (82), setting $s_n^+ := x_n^+ - x_n = -e_n + o(|e_n|)$, we write

$$\langle g(x_n^+), s_n^+\rangle = \langle \bar{H}e_n^+, s_n^+\rangle + o(|e_n^+|\,|s_n^+|) = o(|e_n|^2),$$

$$\langle g(x_n), s_n^+\rangle = \langle \bar{H}e_n, s_n^+\rangle + o(|e_n|^2) = -\langle \bar{H}e_n, e_n\rangle + o(|e_n|^2).$$

We therefore obtain

$$\langle g(x_n^+) - g(x_n), s_n^+\rangle = \langle \bar{H}e_n, e_n\rangle + o(|e_n|^2) \geq \alpha|e_n|^2 + o(|e_n|^2),$$

and this again is eventually positive. □□

4.4. The q-superlinear convergence. Let us give one more general result from the theory of BFGS updates (see [3]).

LEMMA 4.7. *If $\{M_n\}$ is generated by the BFGS formula using pairs of vectors (y_n, s_n) such that*

$$\langle y_n, s_n\rangle > 0 \ \text{ for all } \ n \geq 1 \quad \text{and} \quad \sum_{n\geq 1}\frac{|y_n - Ms_n|}{|s_n|} < \infty,$$

where M is a fixed symmetric positive definite matrix, then

$$(M_n - M)s_n = o(|s_n|). \tag{83}$$

We have now all the necessary material to give our superlinear convergence result.

THEOREM 4.8. *Assume that the sequence $\{x_n\}$ generated by Algorithm* (BFGS-AP3) *converges to an optimal point $\bar{x}$, and that* (30), (33) *hold. Assume also that $\bar{H}$ is positive definite. Then, the convergence of x_n to $\bar{x}$ is q-superlinear.*

Proof. First of all, we establish the necessary local properties of the gradient mapping. Take x and y in the neighborhood of $\bar{x}$ and apply (34):

$$g(x) - g(y) = \bar{H}(x-y) + o(|x-y|).$$

This implies the Lipschitz continuity of g near $\bar{x}$. Multiply this last relation by $x - y$: because $\bar{H}$ is positive definite, g is (locally) strongly monotone, i.e., f is (locally) strongly convex. Thus, starting with Theorem 4.5 (all the assumptions required are satisfied): $\{x_n\}$ converges r-superlinearly to $\bar{x}$.

Now, since (33) holds, we have

$$\frac{|y_n - \bar{H}s_n|}{|s_n|} \leq L(|x_{n+1} - \bar{x}| + |x_n - \bar{x}|).$$

Therefore, by the r-linear convergence of $\{x_n\}$,

$$\sum_{n\geq 1} \frac{|y_n - \bar{H}s_n|}{|s_n|} < +\infty.$$

This and Lemma 4.7 give $(M_n - \bar{H})s_n = o(|s_n|)$.

Finally, the latter estimate and Theorem 3.6 imply that $x_n + 2s_n^p - \bar{x} = o(|e_n|)$. Then Theorem 4.6 shows that the stepsize $t_n = 2$ is accepted by the line-search. Hence $e_{n+1} = o(|e_n|)$ and the convergence is q-superlinear. □□

Let us conclude this section by a consequence of Theorems 4.2 and 4.8: if g is locally Lipschitzian, and if f has a minimum point $\bar{x}$ satisfying the assumptions of Theorem 4.8, then Algorithm (BFGS-AP3) is globally and q-superlinearly convergent.

5. Conclusion. The essential content of this paper is a theoretical investigation of algorithms for nonsmooth optimization combining quasi-Newton techniques with Moreau-Yosida regularizations. When doing so, we have privileged approaches lending themselves to implementations via bundle methods.

Ideally, this should be achieved by the algorithmic pattern AP2; see [18] for implementable proposals. However, the local properties of this algorithm turn out to be rather hard to analyze; as for AP1, studied by [20], some technicalities are needed when turning to implementation aspects. We have therefore adopted here AP3, which appears as a good compromise between theoretical simplicity and practical significance.

As stated in Sections 3 and 4, AP3 is quite comparable to a standard quasi-Newton algorithm. By analogy with differential equations, AP3 could be viewed as a trapezoidal integration scheme: two successive iterates are computed using the derivatives g and H at their mid-point x_n^p. As a by-product, the tools of the present work could therefore be applied to standard quasi-Newton algorithms (i.e., explicit integration schemes). Keeping this in mind, our local theory of §3 is then fairly comparable to that of [14]. In particular, it should be pointed out that the relevant smoothness assumptions are basically the same. Our role in this matter has been to extract from [14] the key properties of f, to be satisfied at the solution point $\bar{x}$ only. In other words, we used the conclusions of Theorems 3.1 and 3.2, instead of their premises.

On the other hand, such a local study with weakened assumptions is related to the resolution of nonsmooth equations, studied in [23], [27], [16], [26], [24], among others. There exist Newton formulæ which converge superlinearly under fairly general assumptions (semi-smoothness of g). Indeed, a Newton scheme uses directly the Hessian $\nabla^2 f(x_n)$, which gives

by definition reliable second-order information at x_n; the role of semi-smoothness is then to ensure that this information remains valid all the way to convergence. By contrast, we need here apparently restrictive assumptions such as (3.8); in a quasi-Newton context, however, they seem rather minimal. For the quasi-Newton equation (3.1) to be any good, the values $g(x_n)$ and $g(x_{n+1})$ must reflect the values $g(x)$ at neighboring x's; this is precisely the role of (3.8).

Acknowledgements. We are indebted to an anonymous referee, whose constructive remarks were very helpful for the final version of this paper.

REFERENCES

[1] A. Auslender, "Numerical methods for nondifferentiable convex optimization", *Mathematical Programming Study*, 30(1987) 102-126.

[2] C.G. Broyden, J.E. Dennis, and J.J. Moré, "On the local and superlinear convergence of quasi-Newton methods", *Journal of the Institute of Mathematics and its Applications*, 12(1973) 223-245.

[3] R.H. Byrd and J. Nocedal, "A tool for the analysis of quasi-Newton methods with application to unconstrained minimization", *SIAM Journal on Numerical Analysis*, 26(1989) 727-739.

[4] F.H. Clarke, *Optimization and nonsmooth analysis* (Wiley, 1983).

[5] R. Correa and C. Lemaréchal, "Convergence of some algorithms for convex minimization", *Mathematical Programming*, 62(1993) 261-275.

[6] J.E. Dennis and J.J. Moré, "A characterization of superlinear convergence and its application to quasi-Newton methods", *Mathematics of Computation*, 28(1974) 549-560.

[7] J.E. Dennis and J.J. Moré, "Quasi-Newton methods, motivation and theory", *SIAM Review*, 19(1977) 46-89.

[8] J.E. Dennis and R.B. Schnabel, A view of unconstrained optimization, in: G.L. Nemhauser, A.H.G. Rinnooy Kan, and M.J. Todd, eds., *Handbook in operations research and management science*, volume 1, pp. 1-72 (Elsevier Science Publishers B.V., North-Holland, 1989).

[9] M. Fukushima, "A descent algorithm for nonsmooth convex programming", *Mathematical Programming*, 30(1984) 163-175.

[10] J.Ch. Gilbert and C. Lemaréchal, "Some numerical experiments with variable-storage quasi-Newton algorithms", *Mathematical Programming*, 45(1989) 407-435.

[11] W.A. Gruver and E. Sachs, *Algorithmic methods in optimal control*, Research Notes in Mathematics #47 (Pitman, 1980).

[12] S.M. Grzegórski, "Orthogonal projections on convex sets for Newton-like methods", *SIAM Journal on Numerical Analysis*, 22(1985) 1208-1219.

[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", *Mathematical Programming*, 56(1992) 71-89.

[15] K.C. Kiwiel, "Proximity control in bundle methods for convex nondifferentiable minimization", *Mathematical Programming*, 46(1990) 105-122.

[16] B. Kummer, "Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis", in: W. Oettli and D. Pallaschke, eds., *Advances in optimization*, number 382 in Lecture Notes in Economics and Mathematical Systems, pp. 171-194 (Springer-Verlag, Berlin, Heidelberg, New York, 1991).

[17] C. Lemaréchal, "A view of line-searches", in: A. Auslender, W. Oettli, and J. Stoer, eds., *Optimization and optimal control*, number 30 in Lecture Notes in Control and Information Science, pp. 59-78 (Springer-Verlag, Berlin, Heidelberg, New York, 1981).

[18] C. Lemaréchal and C. Sagastizábal, "An approach to variable metric bundle methods", in: *Proceedings IFIP Conference Systems Modelling and Optimization*, Springer-Verlag, to appear. Also as Rapport de Recherche INRIA #2128, 1993.

[19] B. Martinet, "Régularisation d'inéquations variationnelles par approximations successives", *Revue Française d'Informatique et Recherche Opérationnelle*, R-3(1970) 154-179.

[20] R. Mifflin, "A quasi-second-order proximal bundle algorithm", Technical Report 93-3, University of Washington (Pullman, Washington, 1993).

[21] J.J. Moreau, "Proximité et dualité dans un espace hilbertien", *Bulletin de la Société Mathématique de France*, 93(1965) 273-299.

[22] J.M. Ortega and W.C. Rheinboldt, *Iterative solution of nonlinear equations in several variables* (Academic Press, New York, 1970).

[23] J.-S. Pang, "Newton's method for B-differentiable equations", *Mathematics of Operations Research*, 15(1990) 311-341.

[24] J.S. Pang and L.Q. Qi, "Nonsmooth equations: motivation and algorithms", *SIAM Journal on Optimization*, 3(1993) 443-465.

[25] M.J.D. Powell, "Some global convergence properties of a variable metric algorithm for minimization without exact line searches", in: R.W. Cottle and C.E. Lemke, eds., *Nonlinear Programming*, number 9 in SIAM-AMS Proceedings (American Mathematical Society, Providence, RI, 1976).

[26] L. Qi and J. Sun, "A nonsmooth version of Newton's method", *Mathematical Programming*, 58(1993) 353-367.

[27] L.Q. Qi, "Convergence analysis of some algorithms for solving nonsmooth equations", *Mathematics of Operations Research*, 18(1993) 227-244.

[28] M. Qian, "The variable metric proximal point algorithm: global and super-linear convergence", Manuscript GN-50, University of Washington, Department of Mathematics, (Seattle, WA 98195, 1992).

[29] S.M. Robinson, "Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity", *Mathematical Programming Study*, 30(1987) 45-66.

[30] R.T. Rockafellar, *Convex analysis*, (Princeton University Press, Princeton, New Jersey, 1970).

[31] R.T. Rockafellar, "Monotone operators and the proximal point algorithm", *SIAM Journal on Control and Optimization*, 14(1976) 877-898.

[32] C.A. Sagastizábal, "Quelques méthodes numériques d'optimisation. Application en gestion de stocks", PhD thesis, University of Paris I, Panthéon-Sorbonne (Paris, 1993).

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399