# Convergence of Quasi-Newton Methods for Solving Constrained Generalized Equations[*]

R. Andreani[†]     R. M. Carvalho[‡]     L. D. Secchin[§]     G. N. Silva[¶]

May 19, 2021 (revised April 5, 2022)

## Abstract

In this paper, we focus on quasi-Newton methods to solve constrained generalized equations. As is well-known, this problem was firstly studied by Robinson and Josephy in the 70's. Since then, it has been extensively studied by many other researchers, specially Dontchev and Rockafellar. Here, we propose two Broyden-type quasi-Newton approaches to dealing with constrained generalized equations, one that requires the exact resolution of the subproblems, and other that allows inexactness, which is closer to numerical reality. In both cases, projections onto the feasible set are also inexact. The local convergence of general quasi-Newton approaches is established under a bounded deterioration of the update matrix and Lipschitz continuity hypotheses. In particular, we prove that a general scheme converges linearly to the solution under suitable assumptions. Furthermore, when a Broyden-type update rule is used, the convergence is superlinearly. Some numerical examples illustrate the applicability of the proposed methods.

## 1   Introduction

All of our study and contributions are focused on the problem known as the *Constrained Generalized Equation*. Basically, it consists of finding $x \in \mathbb{X}$ such that

$$x \in C, \qquad 0 \in f(x) + F(x), \tag{1}$$

[†]Department of Applied Mathematics, University of Campinas, Rua Sérgio Buarque de Holanda, 651, 13083-859, Campinas, SP, Brazil. E-mail: `andreani@unicamp.br`

[‡]Federal Institute of Piauí, Picos, PI, Brazil. E-mail: `rui.marques@ifpi.edu.br`

[§]Department of Applied Mathematics, Federal University of Espírito Santo, Rodovia BR 101, Km 60, 29932-540, São Mateus, ES, Brazil. E-mail: `leonardo.secchin@ufes.br`

[¶]Center of Exact Sciences and Technologies, Federal University of the West of Bahia, 47808-021, Barreiras, BA, Brazil. E-mail: `gilson.silva@ufob.edu.br`

where $f : \Omega \to \mathbb{Y}$ is a continuously differentiable function, $\mathbb{X}$ and $\mathbb{Y}$ are Banach spaces, $\Omega \subseteq \mathbb{X}$ is an open set, $C \subset \Omega$ is a nonempty closed convex set, and $F : \Omega \rightrightarrows \mathbb{Y}$ is a multifunction with a closed nonempty graph.

Generalized equations were first proposed by Robinson [36]. In that work, the author deals with the unconstrained problem, which aims to find $x \in \mathbb{X}$ such that

$$0 \in f(x) + F(x), \tag{2}$$

where $f$ and $F$ are basically the same as in problem (1). This problem differs from (1) by the absence of the constraint $x \in C$. Josephy [30] shows that several problems can be rewritten as in (2), namely, the general nonlinear optimization, the variational inequality and the equilibrium problems. In the last ten years, many researchers have devoted their efforts to studying the application of Newton's method and its variants to solve (2), see for instance [30, 6, 7, 10, 19, 24, 23, 25, 26, 27, 35, 37, 38]. In particular, we highlight the important contributions of Dontchev [6, 7, 24], Adly [2, 4], Bonnans [10], Ferreira [25, 27, 14] and their collaborators.

The problem addressed in this paper appeared in a recent work by Oliveira et al [14], where Newton's method for solving (1) was considered. The presence of the constrained set $C$ allows us to write, in addition to the problems already mentioned previously, others in the form (1). For instance, the *Constrained Variational Inequality Problem* (CVIP)

$$\text{find} \quad x \in U \cap V \quad \text{such that} \quad \langle f(x), y - x \rangle \geq 0 \quad \text{for all} \quad y \in U, \tag{3}$$

$U, V \subset \mathbb{X}$ closed convex sets, can be stated as

$$\text{find} \quad x \in U \cap V \quad \text{such that} \quad 0 \in f(x) + N_U(x),$$

where $N_U$ is the normal cone associated to $U$. Problem (3) has been extensively studied over the past ten years, see for instance [11, 29]. Another important equivalence to constrained generalized equations is the *Split Variational Inequality Problem* (SVIP), stated as follows: let $U, V \subset \mathbb{Y}$ be nonempty, closed convex sets, and $A : \mathbb{X} \to \mathbb{Y}$ be a linear operator. Let $f : \mathbb{X} \to \mathbb{X}$ and $g : \mathbb{Y} \to \mathbb{Y}$ be functions. Then SVIP consists of

$$\text{find} \quad x_* \in U \quad \text{such that} \quad \langle f(x_*), x - x_* \rangle \geq 0, \quad \text{for all} \quad x \in U,$$

and such that $y_* = A x_* \in V$ satisfies

$$\langle g(y_*), y - y_* \rangle \geq 0 \quad \text{for all} \quad y \in V.$$

Taking $D := U \times V$ and $V := \{w = (x, y) \in \mathbb{X} \times \mathbb{Y} \mid Ax = y\}$, SVIP is equivalent to the following CVIP [11, Lemma 5.1]:

$$\text{find} \quad w_* \in D \cap V \quad \text{such that} \quad \langle h(w_*), w - w_* \rangle \geq 0 \quad \text{for all} \quad w \in D,$$

where $w = (x, y)$ and $h(x, y) := (f(x), g(y))$. In turn, this CVIP is equivalent to the following constrained generalized equation:

$$\text{find} \quad w_* \in D \cap V \quad \text{such that} \quad 0 \in h(w_*) + N_D(w_*).$$

It is known that SVIP includes several optimization problems, for instance, *Split Minimization Problem* and *Common Solutions to Variational Inequalities Problem*. For more details about these problems, see [11, 29, 1, 12, 34].

Artacho et al [6] studied a quasi-Newton method for the unconstrained problem (2). The authors considered the following iterative scheme:

$$f(x_k) + B_k(x_{k+1} - x_k) + F(x_{k+1}) \ni 0, \qquad k = 0, 1, \cdots, \tag{4}$$

where $\{B_k\}$ is a sequence of bounded linear mappings between Banach spaces $X$ and $Y$ satisfying the classical Broyden update rule. They proved that if the multifunction $f + F$ is metrically regular at $x_*$ for 0 and the derivative mapping $f'$ is Lipschitz continuous, then the sequence $\{x_k\}$ generated by (4) is linearly convergent to $x_*$, see [6, Theorem 4.3]. More generally, Adly and Huynh [3] introduced quasi-Newton schemes like (4) for solving (2), allowing $f$ possibly not differentiable. In this case, the authors assume the regularity metric condition with respect to a kind of semismooth regularization of $f + F$. They proved that if $B_k$ satisfies a suitable modified Broyden update, the sequence $\{x_k\}$ generated by (4) is linearly convergent to a solution $x_*$ of (2) [3, Theorem 4.3]. Similar approach was employed in [8, 33].

In this paper, we propose two quasi-Newton schemes to solve the constrained generalized equation (1). The first one is based on the following idea: given $x_0 \in C$ and the initial $B_0$, near to $f'(x_*)$, we compute, in each iteration, an intermediate point $y_k$ such that

$$f(x_k) + B_k(y_k - x_k) + F(y_k) \ni 0, \qquad k = 0, 1, \cdots. \tag{5}$$

Since $y_k$ can be infeasible, that is, $y_k \notin C$, we project it onto $C$ by an inexact projection procedure, obtaining a new iterate $x_{k+1}$ almost feasible. Thus we prove that under suitable assumptions the main sequence $\{x_k\}$ converges linearly to a solution of (1). The second method follows an analogous idea. The difference is that (5) is allowed to be solved inexactly. Specifically, $y_k$ only must be in an open ball centred at $x_k$ that intersects the set $\{y \mid f(x_k) + B_k(y - x_k) + F(y_k)\}$. This strategy is more suitable to be implemented, since solving (5) exactly can be practically impossible even in simple problems. Naturally, allowing inexact solutions leads to a more complicated convergence theory, which is addressed in Section 4.

This paper is divided into two parts. In the first one, we use the quasi-Newton approach (5) to find a solution of problem (1) with $B_k$ satisfying the classical Broyden update rule. By assuming the regularity metric of the multifunction $f + F$ at $x_*$ to 0, we show that the sequence $\{x_k\}$ generated by (5) is linearly convergent to $x_*$. Firstly, we suppose

3

that $B_k$ satisfies a bounded deterioration condition to obtain a general convergence result. As a particular case, we show that the Broyden update satisfies this bounded deterioration. It is worth mentioning that the used bounded deterioration condition on $B_k$ was previously considered in [6, 3]. Furthermore, we use the fact that Lipschitz properties of the multifunction $F^{-1}$ are inherited by the multifunction $(f + F)^{-1}$, as demonstrated by Dontchev and Hager in [21]. In the second part, we address the problem (1) using an inexact approach and similar ideas proposed by Dontchev and Rockafellar in [24]. The proposed inexact quasi-Newton method is described by

$$(f(x_k) + B_k(y_k - x_k) + F(y_k)) \cap R_k(x_k, y_k) \neq \emptyset, \qquad k = 0, 1, \cdots, \qquad (6)$$

where $R_k : \mathbb{X} \rightrightarrows \mathbb{Y}$ is a sequence of multifunctions with closed graphs representing the inexactness. It is not difficult to see that if $F \equiv \{0\}$ and $R_k(x_k, y_k) = \mathbb{B}_{\eta_k \|f(x)\|}(0)$, $\eta_k > 0$, ($\mathbb{B}_{\eta_k \|f(x)\|}(0)$ is the open ball centered at 0 and radius $\eta_k \|f(x)\|$), then the iterative scheme (6) reduces to

$$\|f(x_k) + B_k(y_k - x_k)\| \leq \eta_k \|f(x_k)\|, \qquad k = 0, 1, \cdots,$$

which can be seen as an inexact quasi-Newton method for solving $f(x) = 0$, $x \in C$. Then, assuming the multifunction $f + F$ metrically regular at $x_*$ for 0, $R_k$ partially Aubin continuous, and $d(0, R_k(u, x_*))$ fulfilling a suitable boundedness property, we show that the sequence $\{x_k\}$ generated by (6) is linearly convergent to $x_*$, with $B_k$ satisfying the Broyden update. It should be mentioned that inexact quasi-Newton methods for solving the unconstrained problem (2) are considered in [13].

This work is organized as follows. In Section 2, we present the notations and basic necessary concepts. In Section 3, we present the first quasi-Newton algorithm, where subproblems are solved exactly, and its local convergence analysis. Section 4 is devoted to the inexact quasi-Newton algorithm and its convergence. In Section 5, we discuss the important particular case of Broyden-type methods. Numerical experiments are presented in Section 7, illustrating the theory. Finally, Section 8 brings our conclusions.

## 2   Preliminaries

In this section, we briefly present the basic concepts that we will use throughout the work. A detailed presentation can be found in [23].

Firstly we establish some notations. Unless otherwise stated, $\mathbb{X}$ and $\mathbb{Y}$ are Banach spaces. A generic norm will be denoted by $\|.\|$. The sets

$$\mathbb{B}_\delta(x) := \{y \in \mathbb{X} \mid \|x - y\| < \delta\}, \qquad \mathbb{B}_\delta[x] := \{y \in \mathbb{X} \mid \|x - y\| \leq \delta\}$$

will denote the *open* and *closed balls* of radius $\delta > 0$, centered at $x$, respectively. The set $\mathbb{R}^+$ is the set of all non-negative real numbers. The *vector space of all continuous*

*linear mappings* $A : \mathbb{X} \to \mathbb{Y}$ will be denoted by $\mathcal{L}(\mathbb{X}, \mathbb{Y})$, and the *norm* of $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$ is defined by $\|A\| := \sup \{\|Ax\| \mid \|x\| \leq 1\}$. Let $\Omega \subseteq \mathbb{X}$ be an open set and $f : \Omega \to \mathbb{Y}$ be Fréchet differentiable at all $x \in \Omega$ (the Fréchet derivative of $f$ at $x$ is the continuous linear mapping $f'(x) : \mathbb{X} \to \mathbb{Y}$). Given a multifunction $F : \mathbb{X} \rightrightarrows \mathbb{Y}$, its *graph* is the set $\operatorname{gph} F := \{(x, u) \in \mathbb{X} \times \mathbb{Y} \mid u \in F(x)\}$. The multifunction $F^{-1} : \mathbb{Y} \rightrightarrows \mathbb{X}$ defined by $F^{-1}(u) := \{x \in \mathbb{Y} \mid u \in F(x)\}$ denotes the *inverse* of $F$. Given $C, D \subset \mathbb{X}$,

$$d(x, D) := \inf_{y \in D} \|x - y\|, \qquad e(C, D) := \sup_{x \in C} d(x, D) \tag{7}$$

are the *distance from $x$ to $D$* and the *excess of $C$ beyond $D$*, respectively. The outer distance from a point $x \in \mathbb{X}$ to a subset $C \subset \mathbb{X}$, denoted by $d^+(x, C)$, is defined as $d^+(x, C) := \sup\{\|x - y\| \mid y \in C\}$. The following conventions are adopted: $d(x, \varnothing) = +\infty$, $e(\varnothing, D) = 0$ when $D \neq \varnothing$, and $e(\varnothing, \varnothing) = +\infty$.

In the context of generalized equations, it is common to consider some regularity conditions over $F$. Here, we will use the following notions:

**Definition 2.1.** *Let $\Omega \subseteq \mathbb{X}$ be an open and nonempty set. We say that the multifunction $G : \Omega \rightrightarrows \mathbb{Y}$ is*

- *metrically regular at $\bar{x} \in \Omega$ for $\bar{u} \in \mathbb{Y}$ with modulus $\lambda > 0$ when $\bar{u} \in G(\bar{x})$, and there exist $a > 0$ and $b > 0$ such that $\mathbb{B}_a[\bar{x}] \subset \Omega$ and*

$$d(x, G^{-1}(u)) \leq \lambda d(u, G(x)) \quad \text{for all} \quad x \in \mathbb{B}_a[\bar{x}], \ u \in \mathbb{B}_b[\bar{u}]; \tag{8}$$

- *strongly metrically subregular at $\bar{x} \in \Omega$ for $\bar{u} \in \mathbb{Y}$ with modulus $\lambda > 0$ when $\bar{u} \in G(\bar{x})$, and there exists $a > 0$ such that*

$$\|x - \bar{x}\| \leq \lambda d(\bar{u}, G(x)) \quad \text{for all} \quad x \in \mathbb{B}_a[\bar{x}].$$

It is easy to see that strong metric subregularity of $G$ at $\bar{x}$ for $\bar{u}$ implies that $\bar{x}$ is an isolated point in $G(\bar{u})$. Another intermediate concept is the metric subregularity, which consists in relaxing metric regularity by requiring (8) with $u = \bar{u}$ fixed.

**Remark 2.2.** *It is known that a multifunction $\Gamma : \mathbb{X} \rightrightarrows \mathbb{Y}$ is metrically regular at $\bar{x} \in \mathbb{X}$ for $\bar{y} \in \mathbb{Y}$ with modulus $\lambda > 0$ if and only if $\Gamma^{-1} : \mathbb{Y} \rightrightarrows \mathbb{X}$ has the Aubin property at $\bar{y}$ for $\bar{x}$ with the same constant $\lambda$, i.e.,*

$$e(\Gamma^{-1}(y) \cap \mathcal{X}, \Gamma^{-1}(y')) \leq \lambda \|y - y'\| \quad \text{for all} \quad y, y' \in \mathcal{Y},$$

*where $\mathcal{X}$ and $\mathcal{Y}$ are neighborhoods of $\bar{x}$ and $\bar{y}$, respectively. See [23, Theorem 5A.3, p. 255].*

The next result establishes a connection between the metric regularity of $f + F$ and the Aubin property of an associated map, which proof is analogous to that presented in [24].

**Proposition 2.3.** *Let $\zeta > 0$ and assume that the multifunction $f + F$ is metrically regular at $\bar{x}$ for $0$ with modulus $\lambda > 0$, where $\lambda\zeta < 1$. Let $u \in \mathbb{X}$, $B_u$ some approximation to $f'(u)$ and consider the multifunction*

$$G_u(x) = f(u) + B_u(x - u) + F(x), \tag{9}$$

*where the operator $B_u$ is such that $\|B_u - f'(\bar{x})\| \leq \zeta$. Then, for every $\kappa > \lambda/(1 - \lambda\zeta)$, there exist positive numbers $a$ and $b$ such that*

$$e\left(G_u^{-1}(y) \cap \mathbb{B}_a[\bar{x}], G_u^{-1}(y')\right) \leq \kappa\|y - y'\| \quad \text{for all} \quad u \in \mathbb{B}_a[\bar{x}], \quad y, y' \in \mathbb{B}_b[0]. \tag{10}$$

Another important result is a generalization of the contraction mapping principle for set-valued mappings, stated below. It will be useful to prove the convergence of the quasi-Newton method in the next section. Its proof can be found in [23, Theorem 5E.2, p. 313].

**Theorem 2.4.** *Let $\Phi : \mathbb{X} \rightrightarrows \mathbb{X}$ be a multifunction and $\bar{x} \in \mathbb{X}$. Suppose that there exist scalars $\rho > 0$ and $\mu \in (0, 1)$ such that the following conditions hold:*

*(i) the set $\mathrm{gph}\,\Phi \cap (\mathbb{B}_\rho[\bar{x}] \times \mathbb{B}_\rho[\bar{x}])$ is closed;*

*(ii) $d(\bar{x}, \Phi(\bar{x})) \leq \rho(1 - \mu)$;*

*(iii) $e\left(\Phi(p) \cap \mathbb{B}_\rho[\bar{x}], \Phi(q)\right) \leq \mu\|p - q\|$ for all $p, q \in \mathbb{B}_\rho[\bar{x}]$.*

*Then, there exists $y \in \mathbb{B}_\rho[\bar{x}]$ such that $y \in \Phi(y)$.*

In the sequel, we present the feasible inexact projection used in our proposed algorithms, as well as some of their properties of interest. This type of projection was used in [14] within a Newton method for constrained generalized equations over Euclidean spaces. See also [28].

**Definition 2.5.** *Let $\theta \geq 0$, $C \subset \mathbb{X}$ be a closed convex set and $x \in C$. The* feasible inexact projection mapping *relative to $x$ with error tolerance $\theta$, denoted by $P_C(\cdot, x, \theta) : \mathbb{X} \rightrightarrows C$, is the multifunction*

$$P_C(y, x, \theta) := \left\{z \in C \mid \langle y - z, u - z \rangle \leq \theta\|y - x\|^2, \ \forall\, u \in C\right\}. \tag{11}$$

*We say that $w \in P_C(y, x, \theta)$ is a* feasible inexact projection *of $y$ onto $C$ with respect to $x$ and with error tolerance $\theta$.*

**Remark 2.6.** *It follows from [9, Proposition 2.1.3, p. 201] that, for each $y \in \mathbb{X}$, the exact projection $P_C(y)$ is a vector in $P_C(y, x, \theta)$. Hence, $P_C(y, x, \theta)$ is nonempty for all $y \in \mathbb{X}$ and $x \in C$.*

**Remark 2.7.** $P_C(y, x, 0) := \{P_C(y)\}$ *for all* $y \in \mathbb{X}$ *and* $x \in C$. *From now on, we write* $P_C(y, x, 0) = P_C(y)$.

**Remark 2.8.** *[14, Lemma 1] If* $y, \bar{y} \in \mathbb{X}$, $x, \bar{x} \in C$, *and* $\theta \geq 0$, *then we have*

$$\|w - P_C(\bar{y}, \bar{x}, 0)\| \leq \|y - \bar{y}\| + \sqrt{2\theta}\|y - x\|$$

*for any* $w \in P_C(y, x, \theta)$.

Finally, we enunciate an useful version of the Aubin property suitable for multifunctions with two blocks of variables. We say that a multifunction $T : V \times W \rightrightarrows S$ is *partially Aubin continuous at* $(\bar{v}, \bar{w}) \in V \times W$ *with respect to* $w$ *uniformly in* $v$ *for* $\bar{s} \in S$ *with modulus* $\lambda > 0$ [23] (or simply, *partially Aubin continuous at* $(\bar{v}, \bar{w})$ *w.r.t.* $w$ *for* $\bar{s}$ *with modulus* $\lambda > 0$) if $\bar{s} \in T(\bar{v}, \bar{w})$ and there are neighborhoods $\mathcal{V}$ of $\bar{v}$, $\mathcal{W}$ of $\bar{w}$ and $\mathcal{S}$ of $\bar{s}$ such that

$$e(T(v, w) \cap \mathcal{S}, T(v, w')) \leq \lambda\|w - w'\| \quad \text{for all} \quad v \in \mathcal{V}, \quad w, w' \in \mathcal{W}.$$

# 3 The quasi-Newton method and its local convergence analysis

In this section, we propose the first quasi-Newton method for solving (1). Here, it is required that the auxiliary iterate $y_k$ be an exact solution of the correspondent unconstrained subproblem $0 \in f(x_k) + B_k(y - x_k) + F(y)$. As we already mentioned, $y_k$ may be infeasible. So, an inexact projection onto $C$ is employed to achieve feasibility at the limit, in the spirit of Definition 2.5. Algorithm 1 below formalizes this idea.

---

**Algorithm 1:** Quasi-Newton with Inexact Projections (QN-INEXP)

**Step 0.** Let $x_0 \in C$, $B_0$ and $\{\theta_k\} \subset [0, +\infty)$. Set $k \leftarrow 0$.

**Step 1.** If $0 \in f(x_k) + F(x_k)$ then **stop** returning $x_k$ as solution. Otherwise, compute $y_k \in \mathbb{X}$ such that

$$0 \in f(x_k) + B_k(y_k - x_k) + F(y_k).$$

**Step 2.** If $y_k \in C$, set $x_{k+1} = y_k$. Otherwise, take any $x_{k+1} \in C$ satisfying

$$x_{k+1} \in P_C(y_k, x_k, \theta_k).$$

**Step 3.** Compute $B_{k+1}$, set $k \leftarrow k + 1$ and go to **Step 1**.

---

**Remark 3.1.** *The projection in* **Step 3** *can be computed as an approximate feasible solution of the problem* $\min_{z \in C}\{\|z - y_k\|^2/2\}$ *satisfying* $\langle y_k - x_{k+1}, z - x_{k+1} \rangle \leq \theta_k \|y_k - x_k\|^2$ *for all* $z \in C$. *In [5], a Frank-Wolfe algorithm is design to compute such a projection. The choice of* $\theta_k$ *will be detailed in Section 7.*

Next, we state the local convergence of the QN-INEXP method. This is the main result of this section. The point $x_*$ will always refer to a solution of (1).

**Theorem 3.2.** *As in (1), let* $\Omega \subset \mathbb{X}$ *be an open set,* $f : \Omega \to \mathbb{Y}$ *be a Fréchet differentiable function,* $F : \Omega \rightrightarrows \mathbb{Y}$ *be a multifunction with closed graph and* $C \subset \Omega$ *be a nonempty closed convex set. Furthermore, let* $x_*$ *such that* $0 \in f(x_*) + F(x_*)$, $x_* \in C$. *Suppose the following conditions hold:*

*(i) $f + F$ is metrically regular at $x_*$ for $0$ with modulus $\lambda > 0$;*

*(ii) there exist $\epsilon > 0$ and a neighborhood $\mathcal{X}$ of $x_*$ such that*

$$\left\| f(u) - f(v) - f'(x_*)(u - v) \right\| \leq \epsilon \|u - v\| \quad \text{for all } u, v \in \mathcal{X}; \tag{12}$$

*(iii) $B_0$ is chosen so that*

$$\|B_0 - f'(x_*)\| < \frac{1}{2\lambda}; \tag{13}$$

*(iv) $\theta_k \geq 0$ for all $k \geq 0$ and $\hat{\theta} := \sup \theta_k < \frac{1}{2}$;*

*(v) there exists a constant $c > 0$ such that, for each $k \geq 0$, $B_{k+1}$ satisfies the bounded deterioration condition*

$$\|B_{k+1} - f'(x_*)\| \leq \|B_k - f'(x_*)\| + c(\|x_k - x_*\| + \|y_k - x_*\|). \tag{14}$$

*Then there exists a neighborhood $\mathcal{U}$ of $x_*$ such that, starting from any $x_0 \in C \cap \mathcal{U} \backslash \{x_*\}$, there is a sequence $\{x_k\} \subset C \cap \mathcal{U}$ generated by the* QN-INEXP *method that converges linearly to $x_*$.*

*Proof.* Let us consider the radii $a > 0$ and $b > 0$ associated with the metric regularity of $f + F$ (see Definition 2.1). Taking $\lambda' > \lambda$, we can assume without loss of generality that $a$ is small enough to $\mathbb{B}_a(x_*) \subset \mathcal{X}$,

$$\lambda \left( \delta + \frac{2ca}{1 - \bar{\gamma}} \right) < 1 \quad \text{and} \quad \frac{\lambda' \left( \epsilon + \delta + \frac{2ca}{1-\bar{\gamma}} \right)}{1 - \lambda \left( \delta + \frac{2ca}{1-\bar{\gamma}} \right)} < \frac{1 - \sqrt{2\hat{\theta}}}{1 + \sqrt{2\hat{\theta}}}, \tag{15}$$

where $\delta := \|B_0 - f'(x_*)\| < 1/(2\lambda)$ and

$$\bar{\gamma} = \left[ (1 + \sqrt{2\theta_0}) \frac{\lambda'(\epsilon + \delta)}{1 - \lambda\delta} + \sqrt{2\theta_0} \right].$$

8

We define the constant

$$\gamma := \max \left\{ \bar{\gamma}, (1 + \sqrt{2\hat{\theta}}) \frac{\lambda' \left( \epsilon + \delta + \frac{2ca}{1-\bar{\gamma}} \right)}{1 - \lambda \left( \delta + \frac{2ca}{1-\bar{\gamma}} \right)} + \sqrt{2\hat{\theta}} \right\}. \tag{16}$$

It is immediate from (15) that $0 < \gamma < 1$. In the sequel, we use induction to prove that, starting from any $x_0$ close enough to $x_*$, is possible to generate a sequence $\{x_k\}$ linearly convergent to $x_*$.

Take $x_0 \in C \cap \mathbb{B}_{r_*}(x_*) \backslash \{x_*\}$, where

$$r_* := \min \left\{ a , \frac{b}{\epsilon + 2\delta} , \frac{b(1-\bar{\gamma})}{2[(1-\bar{\gamma})(\epsilon+\delta) + 2ca]} \right\}.$$

To construct the next iterate $x_1$, let us verify the conditions in Theorem 2.4. Defining the auxiliary multifunction

$$\Phi_{x_0}(z) := G_{x_*}^{-1} \left( f(x_*) - f(x_0) - B_0(z - x_0) + f'(x_*)(z - x_*) \right),$$

it follows that

$$d(x_*, \Phi_{x_0}(x_*)) = d \left( x_*, G_{x_*}^{-1} \left( f(x_*) - f(x_0) - B_0(x_* - x_0) \right) \right).$$

By using (12) and the definition of $r_*$, we get that

$$\begin{aligned} &\| f(x_*) - f(x_0) - B_0(x_* - x_0) \| \\ &\leq \| f(x_*) - f(x_0) - f'(x_*)(x_* - x_0) \| + \| (f'(x_*) - B_0)(x_* - x_0) \| \\ &\leq (\epsilon + \delta)\|x_* - x_0\| \leq b. \end{aligned} \tag{17}$$

Then, since $0 \in G_{x_*}(x_*)$ (see (9)), we obtain from (17) and Definition 2.1

$$\begin{aligned} d(x_*, \Phi_{x_0}(x_*)) &\leq \lambda' d(f(x_*) - f(x_0) - B_0(x_* - x_0), G_{x_*}(x_*)) \\ &\leq \lambda' \| f(x_*) - f(x_0) - B_0(x_* - x_0) \| \leq \lambda'(\epsilon + \delta)\|x_* - x_0\|. \end{aligned}$$

Thus, we have $d(x_*, \Phi_{x_0}(x_*)) \leq \rho(1 - \lambda\delta)$ for

$$\rho := \frac{\lambda'(\epsilon + \delta)\|x_* - x_0\|}{1 - \lambda\delta}.$$

Now, let $p, q \in \mathbb{B}_\rho[x_*]$. Taking into account (15) and $x_0 \in \mathbb{B}_{r_*}(x_*) \backslash \{x_*\}$, we can verify that $\rho < r_*$. Therefore, for $s = p$ or $s = q$ we obtain that

$$\begin{aligned} &\| f(x_*) - f(x_0) - B_0(s - x_0) + f'(x_*)(s - x_*) \| \\ &\qquad \leq \| f(x_*) - f(x_0) - f'(x_*)(x_* - x_0) \| + \| f'(x_*)(x_* - x_0) - f'(x_*)(x_* - s) - B_0(s - x_0) \| \\ &\qquad \leq (\epsilon + 2\delta)\|x_* - x_0\| \leq b, \end{aligned}$$

9

where the second inequality holds since $p, q \in \mathbb{B}_\rho[x_*]$ and, by (15), $\rho < \|x_0 - x_*\|$. As $e(\varnothing, \Phi_{x_0}(q)) = 0$, we can assume that $\Phi_{x_0}(p) \cap \mathbb{B}_a[x_*] \neq \varnothing$ for all $p \in \mathbb{B}_\rho[x_*]$. Let $z \in \Phi_{x_0}(p) \cap \mathbb{B}_a[x_*]$. From Definition 2.1 with $\bar{x} = x_*$, $\bar{u} = 0$, $x = z$, $u = f(x_*) - f(x_0) - B_0(q - x_0) + f'(x_*)(q - x_*)$ and $G = \Phi_{x_0}(q)$, we have

$$d(z, \Phi_{x_0}(q)) \leq \lambda d\left(f(x_*) - f(x_0) - B_0(q - x_0) + f'(x_*)(q - x_*), \ G_{x_*}(z)\right).$$

Since $z \in \Phi_{x_0}(p)$ implies $f(x_*) - f(x_0) - B_0(p - x_0) + f'(x_*)(p - x_*) \in G_{x_*}(z)$, the definition of distance given in (7) ensures that

$$d(f(x_*) - f(x_0) - B_0(q - x_0) + f'(x_*)(q - x_*), \ G_{x_*}(z)) \leq \|B_0 - f'(x_*)\|\|p - q\|.$$

So, $d(z, \Phi_{x_0}(q)) \leq \lambda \|B_0 - f'(x_*)\|\|p - q\|$, which implies $e(\Phi_{x_0}(p) \cap \mathbb{B}_a[x_*], \Phi_{x_0}(q)) \leq \lambda \|B_0 - f'(x_*)\|\|p - q\|$. Futhermore, since $\rho < r_* \leq a$, we have $e(\Phi_{x_0}(p) \cap \mathbb{B}_\rho[x_*], \Phi_{x_0}(q)) \leq e(\Phi_{x_0}(p) \cap \mathbb{B}_a[x_*], \Phi_{x_0}(q))$. Hence, using (13), we obtain

$$e\left(\Phi_{x_0}(p) \cap \mathbb{B}_\rho[x_*], \Phi_{x_0}(q)\right) \leq \lambda \left\|B_0 - f'(x_*)\right\| \|p - q\| \leq \lambda\delta\|p - q\|.$$

As $\lambda\delta < 1$ (see (13)), we can apply Theorem 2.4 with $\Phi = \Phi_{x_0}$, $\bar{x} = x_*$ and $\mu = \lambda\delta$ to conclude that there exists $y_0 \in \Phi_{x_0}(y_0)$ such that

$$\|y_0 - x_*\| \leq \frac{\lambda'(\epsilon + \delta)}{1 - \lambda\delta}\|x_0 - x_*\|.$$

At this point, we have constructed $y_0$. The next iterate $x_1$ is obtained according to **Step 2**, that is, $x_1 := y_0 \in C \cap \mathbb{B}_{r_*}[x_*]$ if $y_0 \in C$ and $x_1 \in P_C(y_0, x_0, \theta_0)$ otherwise. Also, it follows from Remark 2.8 with $w = x_1$, $y = y_0$, $x = x_0$, $\bar{y} = x_*$, $\bar{x} = x_*$ and $\theta = \theta_0$ that

$$\|x_1 - P_C(x_*, x_*, 0)\| \leq (1 + \sqrt{2\theta_0})\|x_* - y_0\| + \sqrt{2\theta_0}\|x_* - x_0\|$$
$$\leq \left[(1 + \sqrt{2\theta_0})\frac{\lambda'(\epsilon + \delta)}{1 - \lambda\delta} + \sqrt{2\theta_0}\right]\|x_* - x_0\| < \bar{\gamma}\|x_0 - x_*\|.$$

Hence, remembering that $P_C(x_*, x_*, 0) = x_*$ we conclude that $\|x_1 - x_*\| \leq \bar{\gamma}\|x_0 - x_*\|$. Since $\bar{\gamma} < 1$, this yields $x_1 \in C \cap \mathbb{B}_{r_*}[x_*]$.

Now, suppose that for $k > 0$ there are $y_0, \cdots, y_{k-1}, x_1, \cdots, x_k$ such that $y_j \in \mathbb{B}_{r_*}[x_*]$ for all $0 \leq j \leq k - 1$, $x_j \in C \cap \mathbb{B}_{r_*}[x_*]$ for all $1 \leq j \leq k$,

$$\|y_{j-1} - x_*\| \leq \bar{\gamma}\|x_{j-1} - x_*\| \qquad \text{and} \qquad \|x_j - x_*\| \leq \bar{\gamma}\|x_{j-1} - x_*\|, \quad \text{for} \quad j = 1, \cdots, k. \tag{18}$$

To complete the induction process we proceed analogously to the first step. As in (17), we need to show that $\|f(x_*) - f(x_k) - B_k(x_* - x_k)\| \leq b$. Indeed, since $x_k \in C \cap \mathbb{B}_{r_*}[x_*]$ we have

$$\|f(x_*) - f(x_k) - B_k(x_* - x_k)\| \leq \|f(x_*) - f(x_k) - f'(x_*)(x_* - x_k)\|$$
$$+ \|(f'(x_*) - B_k)(x_* - x_k)\| \leq \left(\epsilon + \|f'(x_*) - B_k\|\right)\|x_* - x_k\|. \tag{19}$$

10

On the other hand, from (14) we obtain that

$$\|f'(x_*) - B_k\| \leq \|f'(x_*) - B_{k-1}\| + c(\|x_{k-1} - x_*\| + \|y_{k-1} - x_*\|)$$
$$\leq \|f'(x_*) - B_{k-2}\| + c(\|x_{k-2} - x_*\| + \|y_{k-2} - x_*\|) + c(\|x_{k-1} - x_*\| + \|y_{k-1} - x_*\|)$$
$$\leq \cdots \leq \|f'(x_*) - B_0\| + c \left( \sum_{j=0}^{k-1} \|x_j - x_*\| + \sum_{j=1}^{k} \|y_{j-1} - x_*\| \right).$$

By (18), the last inequality becomes

$$\|f'(x_*) - B_k\| \leq \|f'(x_*) - B_0\| + c \left( \|x_0 - x_*\| + \sum_{j=1}^{k-1} \|x_j - x_*\| + \sum_{j=1}^{k} \|y_{j-1} - x_*\| \right)$$

$$\leq \delta + c\|x_0 - x_*\| + 2c \sum_{j=1}^{k} \bar{\gamma}^j \|x_0 - x_*\|$$

$$\leq \delta + 2c \sum_{j=0}^{k} \bar{\gamma}^j \|x_0 - x_*\| \leq \delta + \frac{2ca}{1 - \bar{\gamma}}. \tag{20}$$

Hence, from (19), (20) and the condition on $r_*$, we have

$$\|f(x_*) - f(x_k) - B_k(x_* - x_k)\| \leq \left( \epsilon + \delta + \frac{2ca}{1 - \bar{\gamma}} \right) r_* \leq b.$$

By Definition 2.1 and taking into account that $0 \in G_{x_*}(x_*)$, we obtain

$$d(x_*, \Phi_{x_k}(x_*)) \leq \lambda' \|f(x_*) - f(x_k) - f'(x_*)(x_* - x_k)\| + \lambda' \|f'(x_*) - B_k\| \|x_* - x_k\|$$

$$\leq \lambda' \left( \epsilon + \delta + \frac{2ca}{1 - \bar{\gamma}} \right) \|x_* - x_k\| = \bar{\rho} \left[ 1 - \lambda \left( \delta + \frac{2ca}{1 - \bar{\gamma}} \right) \right],$$

where

$$\bar{\rho} = \frac{\lambda' \left( \epsilon + \delta + \frac{2ca}{1 - \bar{\gamma}} \right)}{1 - \lambda \left( \delta + \frac{2ca}{1 - \bar{\gamma}} \right)} \|x_* - x_k\|.$$

Now, let $p, q \in \mathbb{B}_{\bar{\rho}}[x_*]$. From (16) we have $\bar{\rho} < r_*$. Thus, it is not difficult to see that $\|f(x_*) - f(x_k) - B_k(v - x_k) + f'(x_*)(v - x_*)\| < b$ for $v = p$ or $v = q$. Let $y \in \Phi_{x_k}(p) \cap \mathbb{B}_a[x_*]$. From Definition 2.1 with $\bar{x} = x_*$, $\bar{u} = 0$, $x_k = y$, $u = f(x_*) - f(x_k) - B_k(q - x_k) + f'(x_*)(q - x_*)$ and $G = \Phi_{x_k}(q)$, we obtain

$$d(y, \Phi_{x_k}(q)) \leq \lambda d \left( f(x_*) - f(x_k) - B_k(q - x_k) + f'(x_*)(q - x_*), \ G_{x_*}(y) \right).$$

11

Since $y \in \Phi_{x_k}(p)$ implies $f(x_*) - f(x_k) - B_k(p - x_k) + f'(x_*)(p - x_*) \in G_{x_*}(y)$, the definition of distance given in (7) gives

$$d(f(x_*) - f(x_k) - B_k(q - x_n) + f'(x_*)(q - x_*), \ G_{x_*}(y)) \leq \|B_k - f'(x_*)\| \|p - q\|.$$

Combining the two last inequalities we conclude that

$$d(y, \Phi_{x_k}(q)) \leq \lambda \|B_k - f'(x_*)\| \|p - q\|.$$

Taking the supremum with respect to $z \in \Phi_{x_k}(p) \cap \mathbb{B}_a[x_*]$ in the last inequality and using the definition of excess given in (7), we have

$$e\left(\Phi_{x_k}(p) \cap \mathbb{B}_a[x_*], \Phi_{x_k}(q)\right) \leq \lambda \|B_k - f'(x_*)\| \|p - q\|.$$

As $\bar{\rho} < r_* \leq a$, we have $e(\Phi_{x_k}(p) \cap \mathbb{B}_{\bar{\rho}}[x_*], \Phi_{x_k}(q)) \leq e(\Phi_{x_k}(p) \cap \mathbb{B}_a[x_*], \Phi_{x_k}(q))$. Hence, from the last inequality and the properties of the norm, we obtain

$$e\left(\Phi_{x_k}(p) \cap \mathbb{B}_\rho[x_*], \Phi_{x_k}(q)\right) \leq \lambda \left\|B_k - f'(x_*)\right\| \|p - q\| \leq \lambda \left(\delta + \frac{2ca}{1 - \bar{\gamma}}\right) \|p - q\|.$$

By (15) we have $\lambda \left(\delta + \frac{2ca}{1 - \bar{\gamma}}\right) < 1$, so we can apply Theorem 2.4 with $\Phi = \Phi_{x_k}$, $\bar{x} = x_*$ and $\mu = \lambda' \left(\delta + \frac{2ca}{1 - \bar{\gamma}}\right)$ to conclude that there exists $y_k \in \Phi_{x_k}(y_k)$ such that

$$\|y_k - x_*\| \leq \frac{\lambda' \left(\epsilon + \delta + \frac{2ca}{1 - \bar{\gamma}}\right)}{1 - \lambda \left(\delta + \frac{2ca}{1 - \bar{\gamma}}\right)} \|x_* - x_k\|.$$

As before, we take $x_{k+1} := y_k \in C \cap \mathbb{B}_{r_*}[x_*]$ if $y_k \in C$, and $x_{k+1} \in P_C(y_k, x_k, \theta_k)$ otherwise. From Remarks 2.7 and 2.8 with $w = x_{k+1}$, $y = y_k$, $x = x_k$, $\bar{y} = x_*$, $\bar{x} = x_*$, $\theta = \theta_k$, we get

$$\|x_* - x_{k+1}\| \leq \left[(1 + \sqrt{2\hat{\theta}}) \frac{\lambda' \left(\epsilon + \delta + \frac{2cra}{1 - \bar{\gamma}}\right)}{1 - \lambda \left(\delta + \frac{2ca}{1 - \bar{\gamma}}\right)} + \sqrt{2\hat{\theta}}\right] \|x_* - x_k\| < \gamma \|x_* - x_k\|.$$

Therefore, we conclude the induction process. $\qquad\square$

**Remark 3.3.** *In [3], Adly and Van Ngai considered a quasi-Newton method similar to (5). The main difference between their method and (5) is that $B_k$ is a multifunction from $\mathbb{X}$ to $\mathbb{Y}$. Also, in [3], the authors introduced a generalization of the semismooth to functions, see [3, Definition 2.3]. Another difference between these results consist in the assumption of regularity. But, applying Proposition 2.3 combined with [23, Theorem 3E.7] and [3, Remark 3.3], we conclude that these assumptions are equivalent. Hence, after some adjustments, assuming that $f$ is Fréchet differentiable and $C \subset \mathbb{X}$, we have that Theorem 3.2 extends [3, Theorem 3.2]. Evidently, the power of our work relies on the case $C \neq \mathbb{X}$. Nevertheless, Theorem 3.2 encompasses [6, Theorem 3.1] in the unconstrained case $C = \mathbb{X}$.*

12

**Remark 3.4.** *Although we do not known the solution $x_*$ a priori, the bounded deterioration condition (14) involving the derivative at $x_*$ acts as a theoretical expectation for the convergence. In the same way is the requirement (13) on the initial $B_0$. Specifically, the bounded deterioration as we stated is used in other related works, e.g., [6, Theorem 3.1]. It is worth mentioning that here we deal with general quasi-Newton schemes, in which case (14) plays an important role even for quasi-Newton methods for standard nonlinear programming [18]. On the other hand, the particular Broyden update rule considered in Section 5 satisfies the bounded deterioration (see Proposition 5.1), in accordance with standard nonlinear programming.*

## 4  The inexact quasi-Newton approach

In this section, we propose a version of QN-INEXP where subproblems need not to be solved exactly. Specifically, they become

$$(f(x_k) + B_k(y_k - x_k) + F(y_k)) \cap R_k(x_k, y_k) \neq \emptyset, \qquad k = 0, 1, \cdots, \tag{21}$$

where $\{B_k\}$ is a sequence of matrices and $R_k : \mathbb{X} \rightrightarrows \mathbb{Y}$ is a sequence of multifunctions with closed graphs representing the inexactness. To illustrate the flexibility of condition (21), we observe that when $F \equiv \{0\}$ and $R_k(x_k, x_{k+1}) := \mathbb{B}_{\eta_k \|f(x_k)\|}(0)$, $\eta_k > 0$, then we recover the inexact quasi-Newton method developed in [15] for nonlinear systems of equations. Also, if $R_k(x_k, x_{k+1}) = \{-r_k(x_k)\}$ where $\{r_k\}$ is a sequence of functions representing the inexactness, our method reduces to an instance of the inexact quasi-Newton method considered in [13]. Similarly to the previous section, we formally state our inexact quasi-Newton scheme with inexact projections in Algorithm 2.

---

**Algorithm 2:** Inexact Quasi-Newton with Inexact Projections (IQN-INEXP)

**Step 0.** Let $x_0 \in C$, $B_0$ and $\{\theta_k\} \subset [0, +\infty)$. Set $k \leftarrow 0$.

**Step 1.** If $0 \in f(x_k) + F(x_k)$ then **stop** returning $x_k$ as solution. Otherwise, compute $y_k \in \mathbb{X}$ such that

$$(f(x_k) + B_k(y_k - x_k) + F(y_k)) \cap R_k(x_k, y_k) \neq \emptyset.$$

**Step 2.** If $y_k \in C$, set $x_{k+1} = y_k$. Otherwise, take any $x_{k+1} \in C$ satisfying

$$x_{k+1} \in P_C(y_k, x_k, \theta_k).$$

**Step 3.** Compute $B_{k+1}$, set $k \leftarrow k + 1$ and go to **Step 1**.

---

As we made for QN-INEXP, we will establish the local convergence of IQN-INEXP. One

may think this can be done simply by adapting the proof of Theorem 3.2 by introducing inexactness, but this is not totally true. Here, differently from the previous theorem that uses the principle of contractions, the Coincidence Theorem will serve as support [24, Theorem 1].

**Theorem 4.1** (Coincidence Theorem). *Let $\mathbb{X}$ and $\mathbb{Y}$ be two metric spaces, $\rho : \mathbb{X} \times \mathbb{X} \to \mathbb{R}^+$ be a metric in $\mathbb{X}$ and consider the multifunctions $\Phi : \mathbb{X} \rightrightarrows \mathbb{Y}$ and $\Gamma : \mathbb{Y} \rightrightarrows \mathbb{X}$. Let $\bar{x} \in \mathbb{X}$ and $\bar{y} \in \mathbb{Y}$. Also, let $\eta$, $\kappa$ and $\mu$ be positive scalars such that $\mu\kappa < 1$. Suppose that one of the sets*

$$\operatorname{gph} \Phi \cap (\mathbb{B}_\eta[\bar{x}] \times \mathbb{B}_{\eta/\mu}[\bar{y}]) \quad or \quad \operatorname{gph} \Gamma \cap (\mathbb{B}_{\eta/\mu}[\bar{y}] \times \mathbb{B}_\eta[\bar{x}])$$

*is closed while the other is complete, or that both sets*

$$\operatorname{gph} (\Phi \circ \Gamma) \cap (\mathbb{B}_\eta[\bar{x}] \times \mathbb{B}_\eta[\bar{x}]) \quad and \quad \operatorname{gph} (\Gamma \circ \Phi) \cap (\mathbb{B}_{\eta/\mu}[\bar{y}] \times \mathbb{B}_{\eta/\mu}[\bar{y}])$$

*are complete. Also, suppose the following conditions hold:*

*(i) $d(\bar{y}, \Phi(\bar{x})) < \eta(1 - \kappa\mu)/(2\mu)$;*

*(ii) $d(\bar{x}, \Gamma(\bar{y})) < \eta(1 - \kappa\mu)/2$;*

*(iii) $e\left(\Phi(p) \cap \mathbb{B}_{\eta/\mu}[\bar{y}], \Phi(q)\right) \leq \mu\rho(p, q)$ for all $p, q \in \mathbb{B}_\eta[\bar{x}]$ such that $\rho(p, q) \leq \eta(1 - \kappa\mu)/\mu$;*

*(iv) $e\left(\Gamma(p) \cap \mathbb{B}_\eta[\bar{x}], \Gamma(q)\right) \leq \kappa\rho(p, q)$ for all $p, q \in \mathbb{B}_{\eta/\mu}[\bar{y}]$ such that $\mu(p, q) \leq \eta(1 - \kappa\mu)$.*

*Then there exist $\hat{x} \in \mathbb{B}_\eta[\bar{x}]$ and $\hat{y} \in \mathbb{B}_{\eta/\mu}[\bar{y}]$ such that $\hat{y} \in \Phi(\hat{x})$ and $\hat{x} \in \Gamma(\hat{y})$. If the mappings $\mathbb{B}_\eta[\bar{x}] \ni x \mapsto \Phi(x) \cap \mathbb{B}_{\eta/\mu}[\bar{y}]$ and $\mathbb{B}_{\eta/\mu}[\bar{y}] \ni y \mapsto \Gamma(y) \cap \mathbb{B}_\eta[\bar{x}]$ are single-valued, then the points $\hat{x}$ and $\hat{y}$ are unique in $\mathbb{B}_\eta[\bar{x}]$ and $\mathbb{B}_{\eta/\mu}[\bar{y}]$, respectively.*

Now, we apply Theorem 4.1 to obtain the desired convergence result for Algorithm 2.

**Theorem 4.2.** *Let $\Omega \subset \mathbb{X}$ be an open set, $f : \Omega \to \mathbb{Y}$ be a Fréchet differentiable function, $F : \Omega \rightrightarrows \mathbb{Y}$ be a multifunction with closed graph and $C \subset \Omega$ be a nonempty closed convex set. Furthermore, let $x_*$ such that $0 \in f(x_*) + F(x_*)$, $x_* \in C$.*

*Suppose valid conditions (i) to (v) of Theorem 3.2. Also, suppose that the following additional conditions hold:*

*(vi) for each $k \geq 0$, the mapping $(u, x) \mapsto R_k(u, x)$ is partially Aubin continuous at $(x_*, x_*)$ w.r.t. $x$ for $0$ with modulus $\mu > 0$;*

*(vii) there are scalars $t \in \left(0, \frac{1 - \sqrt{2\hat{\theta}}}{1 + \sqrt{2\hat{\theta}}}\right)$, $0 < \gamma < t(1 - \lambda\mu)/2\mu$ and $\beta > 0$ such that*

$$d(0, R_k(u, x_*)) \leq \gamma\|u - x_*\| \quad for\ all \quad u \in \mathbb{B}_\beta(x_*), \quad k \geq 0. \tag{22}$$

14

*Then there exists a neighborhood $\mathcal{U}$ of $x_*$ such that, starting from any $x_0 \in C \cap \mathcal{U} \backslash \{x_*\}$, there is a sequence $\{x_k\} \subset C \cap \mathcal{U}$ generated by the* IQN-INEXP *method that converges linearly to $x_*$.*

*Proof.* From item (vi), there exist positive constants $a$ and $b$ satisfying

$$e\left(R_k(u, x) \cap \mathbb{B}_b[0], R_k(u, x')\right) \leq \mu \|x - x'\| \quad \text{for all} \quad u, x, x' \in \mathbb{B}_a[x_*]. \tag{23}$$

By (12), we can assume $a$ small enough so that

$$\|f(x) - f(x_*) - f'(x_*)(x - x_*)\| \leq \epsilon \|x - x_*\| \quad \text{for all} \quad x \in \mathbb{B}_a[x_*], \tag{24}$$

and

$$\left(\delta + \frac{2ca}{1 - \tilde{t}_0}\right) \lambda < 1, \tag{25}$$

where $c$ is the constant in (14) and $\tilde{t}_k := \left[(1 + \sqrt{2\theta_k})t + \sqrt{2\theta_k}\right]$, for all $k \geq 0$. Note that $\tilde{t}_k < 1$ from item (vii) and the fact that $\theta_k \leq \hat{\theta}$. Also, we choose a constant $\kappa$ such that $\kappa > \lambda$, $\kappa\mu < 1$, $\gamma < t(1 - \kappa\mu)/(2\mu)$,

$$\kappa > \varpi := \frac{\lambda}{1 - \lambda\left(\delta + \frac{2ca}{1 - \tilde{t}_0}\right)}$$

and

$$\epsilon + \delta + \frac{2ca}{1 - \tilde{t}_0} < \frac{t(1 - \kappa\mu)}{2\kappa}. \tag{26}$$

In order to apply Proposition 2.3, we define the following positive constant:

$$r_* := \min\left\{ a, \ \frac{b}{\left(\epsilon + \delta + \frac{2ca}{1 - \tilde{t}_0}\right)}, \ \beta, \ b\mu \right\}. \tag{27}$$

Choose any $x_0 \in \mathbb{B}_{r_*}[x_*] \backslash \{x_*\}$. Since $r_* \leq a$ and $(\epsilon + \delta)r_* \leq b$ we obtain

$$\|f(x_*) - f(x_0) - B_0(x_* - x_0)\|$$
$$\leq \|f(x_*) - f(x_0) - f'(x_*)(x_* - x_0)\| + \|(f'(x_*) - B_0)(x_* - x_0)\|$$
$$\leq (\epsilon + \delta)r_* \leq b.$$

Combining the above inequality with the fact that $0 \in f(x_*) + F(x_*)$ we have

$$-f(x_*) + f(x_0) + B_0(x_* - x_0) \in G_{x_0}(x_*) \cap \mathbb{B}_b[0]. \tag{28}$$

Then, since $\kappa > \varpi$ and $\lambda\delta < 1$, we can apply Proposition 2.3 with $u = x_0$, $\bar{x} = x_*$, $\zeta = \delta$, $y = -f(x_*) + f(x_0) + B_0(x_* - x_0)$, $y' = 0$ and $a = r_*$ to conclude that

$$e\left(G_{x_0}^{-1}(-f(x_*) + f(x_0) + B_0(x_* - x_0)) \cap \mathbb{B}_{r_*}[x_*], G_{x_0}^{-1}(0)\right)$$
$$\leq \kappa \| -f(x_*) + f(x_0) + B_0(x_* - x_0)\|.$$

Now, we use the definition of excess in (7) and (28) to obtain

$$d(x_*, G_{x_0}^{-1}(0)) \le \kappa \| - f(x_*) + f(x_0) + B_0(x_* - x_0) \|.$$

After simple manipulations and using (26), item (vii) and (24), the above inequality implies that

$$d(x_*, G_{x_0}^{-1}(0)) \le \kappa(\epsilon + \delta) \|x_* - x_0\| < \frac{t(1 - \kappa\mu)}{2} \|x_* - x_0\| < \frac{\eta(1 - \kappa\mu)}{2},$$

where the second inequality follows by (26) with $\eta := t\|x_0 - x_*\|$. On the other hand, as $\|x_0 - x_*\| \le r_* \le \beta$, $\kappa\mu < 1$ and $\gamma < t(1 - \kappa\mu)/(2\mu)$, we obtain from (22) that

$$d(0, R_0(x_0, x_*)) \le \gamma \|x_0 - x_*\| < \frac{t(1 - \kappa\mu)}{2\mu} \|x_0 - x_*\| = \frac{\eta(1 - \kappa\mu)}{2\mu}. \tag{29}$$

Hence, the conditions (i) and (ii) in Theorem 4.1 are satisfied with $\Phi(x) = R_0(x_0, x_*)$, $\Gamma = G_{x_0}^{-1}$, $\kappa$, $\mu$, $\bar{x} = x_*$, $\bar{y} = 0$ and $\eta = t\|x_0 - x_*\|$.

Now, since $x_0 \in \mathbb{B}_{r_*}[\bar{x}]$ and $t < 1$, we obtain that $\eta \le r_* \le a$ and $\eta/\mu \le b$. Thus, $\mathbb{B}_{\eta/\mu}[0] \subset \mathbb{B}_b[0]$ and $\mathbb{B}_\eta[x_*] \subset \mathbb{B}_a[x_*]$. Hence, for any $x, x' \in \mathbb{B}_a[x_*]$, condition (23) implies

$$e(R_0(x_0, x) \cap \mathbb{B}_{\eta/\mu}[0], R_0(x_0, x')) \le e(R_0(x_0, x) \cap B_b[0], R_0(x_0, x')) \le \mu \|x - x'\|,$$

that is, condition (iii) in Theorem 4.1 holds. Furthermore, from (10) we conclude that

$$e\left(G_{x_0}^{-1}(x) \cap \mathbb{B}_\eta[x_*], G_{x_0}^{-1}(x')\right) \le e\left(G_{x_0}^{-1}(x) \cap \mathbb{B}_a[x_*], G_{x_0}^{-1}(x')\right) \le \kappa \|x - x'\|$$

for all $x, x' \in \mathbb{B}_{\eta/\mu}[0] \subset \mathbb{B}_b[0]$. So, condition (iv) in Theorem 4.1 also holds. Therefore, we can apply Theorem 4.1 to ensure the existence of $y_0 \in \mathbb{B}_\eta[x_*]$ and $v_1 \in \mathbb{B}_{\eta/\mu}[0]$ such that $y_0 \in G_{x_0}^{-1}(v_1)$ and $v_1 \in R_0(x_0, y_0)$; that is,

$$v_1 \in G_{x_0}(y_0) \cap R_0(x_0, y_0) = (f(x_0) + B_0(y_0 - x_0) + F(y_0)) \cap R_0(x_0, y_0). \tag{30}$$

Moreover, Theorem 4.1 implies $\|y_0 - x_*\| \le t\|x_0 - x_*\|$. The inclusion (30) yields that $y_0$ satisfies (21) for $k = 0$. In particular, since $t < 1$, $y_0 \in \mathbb{B}_{r_*}[x_*]$. If $y_0 \in C \cap \mathbb{B}_\eta[x_*]$ then take $x_1 := y_0$. Otherwise, take $x_1 \in P_C(y_0, x_0, \theta_0)$. By using similar arguments as in Theorem 3.2 (see also Remark 2.8) we obtain

$$\|x_1 - x_*\| \le (1 + \sqrt{2\theta_0})\|x_* - y_0\| + \sqrt{2\theta_0}\|x_* - x_0\|$$
$$= \left[(1 + \sqrt{2\theta_0})t + \sqrt{2\theta_0}\right]\|x_* - x_0\| = \tilde{t}_0\|x_0 - x_*\|.$$

Since $\tilde{t}_0 < 1$, we have $x_1 \in C \cap \mathbb{B}_\eta[x_*] \subset C \cap \mathbb{B}_{r_*}[x_*]$. By induction, we suppose that there exist an integer $k > 1$ and points $x_1, x_2, \cdots, x_k \in C \cap \mathbb{B}_{r_*}[x_*]$ and $y_1, y_2, \cdots, y_k \in \mathbb{B}_{r_*}[x_*]$, satisfying

$$\|y_{j-1} - x_*\| \le t\|x_{j-1} - x_*\|, \quad \forall j = 1, 2, \cdots, k, \tag{31}$$
$$\|x_j - x_*\| \le \tilde{t}_0\|x_{j-1} - x_*\|, \quad \forall j = 1, 2, \cdots, k. \tag{32}$$

16

Without loss of generality, we assume that $y_{j-1}$, $x_{j-1}$ and $x_*$ are distinct from each other. Note that, as $x_k \in \mathbb{B}_{r_*}[x_*]$ and $r_* \leq \beta$, we can repeat the same argument from (29) with $x_0$ replaced by $x_k$, obtaining

$$d(0, R_k(x_k, x_*)) \leq \frac{\eta(1 - \kappa\mu)}{2\mu}.$$

To apply Proposition 2.3 in the induction step, firstly we need to show that $\|B_k - f'(x_*)\| \leq \zeta$ for some positive scalar $\zeta$ such that $\zeta\lambda < 1$. By combining (14), (31), (32) and using the fact that $x_0 \in \mathbb{B}_{r_*}[x_*]$, we have

$$\|f'(x_*) - B_k\| \leq \|f'(x_*) - B_{k-1}\| + c(\|x_{k-1} - x_*\| + \|y_{k-1} - x_*\|)$$

$$\leq \|f'(x_*) - B_0\| + c \sum_{j=0}^{k}(\|x_j - x_*\| + t\|x_j - x_*\|)$$

$$\leq \delta + (1+t)c\|x_0 - x_*\| + (1+t)c \sum_{j=1}^{k} \|x_j - x_*\|$$

$$\leq \delta + (1+t)c \sum_{j=0}^{k} \|x_j - x_*\| \leq \delta + 2c \sum_{j=0}^{\infty} \tilde{t}_0^j \|x_0 - x_*\| \leq \delta + \frac{2ca}{1 - \tilde{t}_0}.$$

That is, for all $k \geq 1$,

$$\|f'(\bar{x}) - B_k\| \leq \delta + \frac{2ca}{1 - \tilde{t}_0}. \tag{33}$$

The above inequality combined with (24) implies

$$\| - f(x_*) + f(x_k) + B_k(x_k - x_*)\|$$
$$\leq \| - f(x_*) + f(x_k) + f'(x_*)(x_* - x_k)\| + \|(f'(x_*) - B_k)(x_* - x_k)\|$$
$$\leq \epsilon\|x_* - x_k\| + \left(\delta + \frac{2ca}{1 - \tilde{t}_0}\right)\|x_* - x_k\| \leq \left(\epsilon + \delta + \frac{2ca}{1 - \tilde{t}_0}\right)r_* \leq b.$$

Thus, $-f(x_*) + f(x_k) + B_k(x_k - x_*) \in \mathbb{B}_b[0] \cap G_{x_k}(x_*)$. In particular, $x_* \in G_{x_k}^{-1}(-f(x_*) + f(x_k) + B_k(x_k - x_*))$. Then, using (25), (26), (33) and the last inequality we can apply Proposition 2.3 to obtain

$$d(x_*, G_{x_k}^{-1}(0)) \leq \kappa\left(\epsilon + \delta + \frac{2ca}{1 - \tilde{t}_0}\right)\|x_* - x_k\| \leq \frac{t(1 - \kappa\mu)}{2}\|x_* - x_k\| = \frac{\eta(1 - \kappa\mu)}{2}.$$

Thus the conditions (vi) and (vii) in Theorem 4.1 are satisfied with $u = x_k$. Now, since $x_k \in \mathbb{B}_{r_*}[x_*]$ and $\eta = t\|x_k - x_*\|$, condition (27) ensures that $\eta \leq r_* \leq a$ and $\eta/\mu \leq b$. So, (23) implies condition (iii) of Theorem 4.1. Furthermore, from (10) we conclude that condition (iv) in Theorem 4.1 holds for $\Gamma = G_{x_j}^{-1}$ with $j = k$. Thus, the assumptions of

17

Theorem 4.1 are satisfied with $\eta = t\|x_k - x_*\|$, and hence there exists $y_k \in \mathbb{B}_\eta[x_*]$ and $v_{k+1} \in \mathbb{B}_{\eta/\mu}[0]$ such that $y_k \in G_{x_k}^{-1}(v_{k+1})$ and $v_{k+1} \in R_k(x_k, y_k)$; that is,

$$v_{k+1} \in G_{x_k}(y_k) \cap R_k(x_k, y_k) = (f(x_k) + B_k(y_k - x_k) + F(y_k)) \cap R_k(x_k, y_k). \qquad (34)$$

Moreover, Theorem 4.1 implies that $\|y_k - x_*\| \leq t\|x_k - x_*\|$. The inclusion (34) also ensures that $y_k$ satisfies (21) for all $k \geq 0$. Since $t < 1$ we have $y_k \in \mathbb{B}_{r_*}[x_*]$. Again, if $y_k \in C \cap \mathbb{B}_\eta[x_*]$ then take $x_{k+1} := y_k$ and $x_{k+1} \in P_{C \cap \mathbb{B}_c[x_*]}(y_k, x_k, \theta_k)$ otherwise. By using similar arguments employed in Theorem 3.2, we finally get

$$\|x_{k+1} - x_*\| \leq \left[(1 + \sqrt{2\theta_k})t + \sqrt{2\theta_k}\right]\|x_* - x_k\| = \tilde{t}_k\|x_k - x_*\|.$$

As $\tilde{t}_k < 1$, we obtain $x_{k+1} \in \mathbb{B}_{r_*}[x_*]$, concluding the proof. $\qquad \square$

**Remark 4.3.** *In [24, Theorem 4], a related result to Theorem 4.2 for the unconstrained case, $C = \mathbb{X}$, was provided. But the authors suppose that $B_k = f'(x_k)$ for all $k$. So, even for $C = \mathbb{X}$, Theorem 4.2 generalizes such previous result since we only require that $B_k$ approximate $f'(x_k)$.*

# 5 Linear Convergence of the Broyden-type quasi-Newton methods

In this section, we consider both QN-INEXP and IQN-INEXP methods when $B_{k+1}$ in **Step 3** is computed by the classical Broyden update scheme. Let us denote by $\langle \cdot, \cdot \rangle$ scalar products in $\mathbb{R}^n$. The Broyden update rule is defined as

$$B_{k+1} := B_k + \frac{(z_k - B_k s_k)\langle s_k, \cdot \rangle}{\|s_k\|^2}, \qquad (35)$$

where $z_k := f(y_k) - f(x_k)$ and $s_k := y_k - x_k$. Despite of (13), a practical usual choice for $B_0$ is $B_0 = f'(x_0)$.

The aim of this section is to show that the Broyden update rule satisfies the bounded deterioration property (14), a crucial condition in Theorems 3.2 and 4.2. Therefore, (35) is one practical choice that fulfils the general framework of our quasi-Newton methods.

**Proposition 5.1.** *Let $\mathbb{X}$ and $\mathbb{Y}$ be Hilbert spaces. Suppose that the Fréchet derivative mapping $f'$ is Lipschitz continuous with constant $L$ in a convex neighborhood $\mathcal{X}$ of a point $x_*$. Given $B_k \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$ and $x_k, y_k \in \mathcal{X}$, $y_k \neq x_k$, the operator $B_{k+1}$ defined as in (35) satisfies*

$$\|B_{k+1} - f'(x_*)\| \leq \|B_k - f'(x_*)\| + \frac{L}{2}(\|y_k - x_*\| + \|x_* - x_k\|).$$

*Proof.* The proof is straightforward from the proof of [6, Proposition 4.2] by taking $x_{k+1} := y_k$. $\qquad \square$

18

Using Proposition 5.1, the local convergence of the Broyden variants of QN-InexP and IQN-InexP follows directly from Theorems 3.2 and 4.2, respectively. Next we state such results.

**Corollary 5.2.** *Let $\Omega \subset \mathbb{X}$ be an open set, $f : \Omega \to \mathbb{Y}$ be a Fréchet differentiable function, $F : \Omega \rightrightarrows \mathbb{Y}$ be a multifunction with closed graph and $C \subset \Omega$ be a nonempty closed convex set. Furthermore, let $x_*$ such that $0 \in f(x_*) + F(x_*)$, $x_* \in C$.*

*Suppose $f + F$ is metrically regular at $x_*$ for $0$ with modulus $\lambda > 0$, and that $f$ has Fréchet derivative Lipschitz continuous locally around $x_*$. Consider the* QN-InexP *method with $\theta_k \geq 0$ for all $k$ and assume that*

$$\hat{\theta} := \sup \theta_k < \frac{1}{2}, \quad \delta := \|B_0 - f'(x_*)\| < \frac{1}{2\lambda}.$$

*Then, if we choose $B_{k+1}$ as in* (35), *there exists a neighborhood $\mathcal{U}$ of $x_*$ such that, starting from any $x_0 \in C \cap \mathcal{U} \backslash \{x_*\}$, there is a sequence $\{x_k\} \subset C \cap \mathcal{U}$ generated by the* QN-InexP *method that converges linearly to $x_*$.*

*If, in addition, items* (vi) *and* (vii) *of Theorem 4.2 are fulfilled, then the same conclusion is valid for the* IQN-InexP *method with $B_{k+1}$ as in* (35).

**Remark 5.3.** *As the classic Broyden update* (35) *satisfies* (14) *(Proposition 5.1), the second part of the above corollary generalizes [3, Theorem 3.2], since we are considering an inexact quasi-Newton method. Even if $C = \mathbb{X}$, the previous corollary extends [24, Theorem 4], since we are proving the convergence of the Broyden update method.*

# 6 Superlinear Convergence under the Dennis-Moré condition

We dedicate this section to show under what conditions we can obtain superlinear convergence of the inexact quasi-Newton method (21) by using the Broyden update (35). For this purpose, we will use the Hilbert-Schmidt norm of an operator $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$ between Hilbert spaces $\mathbb{X}$ and $\mathbb{Y}$, defined as

$$\|A\|_{HS} := \sqrt{\sum_{i \in I} \|Ae_i\|^2},$$

where $\{e_i, i \in I\}$ is an orthonormal basis of $\mathbb{X}$. We denote by $\mathcal{T}(\mathbb{X}, \mathbb{Y}) := \{A \in \mathcal{L}(\mathbb{X}, \mathbb{Y}) \mid \|A\|_{HS} < +\infty\}$ the set of Hilbert-Schmidt operators.

It is straightforward to show, as done in [6], that Proposition 5.1 is valid also for the Hilbert-Schmidt norm. To show the desired superlinear convergence we will use the Dennis-Moré theorem. This theorem gives a characterization for the superlinear convergence in quasi-Newton methods and these results can be adapted to generalized equations, see

[6, 17]. Briefly, if a quasi-Newton method generates a sequence $\{x_k\}$ which stays near $x_*$ and $x_{k+1} \neq x_k$ for all $k$, then $\{x_k\}$ converges superlinearly to $x_*$ if, and only if, it is convergent and

$$\lim_{k \to \infty} \frac{\|(B_k - f'(x_*))\widehat{s}_k\|}{\|\widehat{s}_k\|} = 0, \tag{36}$$

where $\widehat{s}_k$ is taken as $x_{k+1} - x_k$.

It is well known that the Broyden update (35) with $y_k = x_{k+1}$, applied to a smooth equation in finite dimensions with a nonsingular Jacobian at the reference solution $x_*$, satisfies (36); see e.g. [31, Theorem 7.2.4]. There are extensions of this claim for infinite dimensional Hilbert spaces [32, 39]. In [20], the Dennis-Moré condition (36) was applied to generalized equations in Banach spaces. In our context, we have the following statement:

**Proposition 6.1.** *Let $\Omega \subset \mathbb{X}$ be an open set, $f : \Omega \to \mathbb{Y}$ be a function which is Fréchet differentiable, $F : \Omega \rightrightarrows \mathbb{Y}$ be a multifunction with closed graph and $C \subset \Omega$ be a nonempty closed convex set. Furthermore, let $x_*$ be a solution of (1). Suppose that $f + F$ is strongly metrically subregular at $x^*$ for $0$ with modulus $\lambda > 0$. Let $\{x_k\}$ be a sequence generated by the* IQN-INEXP *method with $\theta_k \to 0$, and assume that*

$$d^+(0, R_k(x, x_*)) \leq \gamma_k \|x - x_*\|, \tag{37}$$

*for all $x$ in some neighborhood of $x_*$, $k \geq 0$ and $\gamma_k \to 0$. Then, if $\{x_k\}$ and $\{y_k\}$ both converge to $x_*$ and (36) holds with $\widehat{s}_k = s_k = y_k - x_k$, $y_k \neq x_k$, then $\{x_k\}$ converges superlinearly to $x_*$.*

*Proof.* The strong subregularity of $f + F$ at $x_*$ for $0$ with modulus $\lambda > 0$ implies that (see Definition 2.1)

$$\|y_k - x_*\| \leq \lambda d(0, f(y_k) + F(y_k)) \tag{38}$$

for all $k$ large enough, let us say, $k \geq k_0$. In turn, **Step 1** of IQN-INEXP (expression (21)) ensures, for each $k \geq 0$, the existence of $u_k \in R_k(x_k, y_k)$ such that

$$u_k - f(x_k) - f'(x_*)s_k - (B_k - f'(x_*))s_k + f(y_k) \in f(y_k) + F(y_k). \tag{39}$$

Thus, combining (38) and (39) we obtain, for all $k \geq k_0$,

$$\begin{aligned}
\|y_k - x_*\| &\leq \lambda \|u_k - f(x_k) - f'(x_*)s_k - (B_k - f'(x_*))s_k + f(y_k)\| \\
&\leq \lambda \|u_k\| + \lambda \|f(y_k) - f(x_k) - f'(x_*)s_k\| + \lambda \|(B_k - f'(x_*))s_k\|. \tag{40}
\end{aligned}$$

From the Fréchet differentiability of $f$, there is an index $k_1 \geq k_0$ and a sequence $\{\epsilon_k\}_{k \geq k_1}$ such that $\gamma_k \leq \epsilon_k \to 0$ and, for each $k \geq k_1$,

$$\|f(z) - f(x_*) - f'(x_*)(z - x_*)\| \leq \epsilon_k \|z - x_*\| \quad \text{for all } z \in \mathcal{X}_k, \tag{41}$$

20

where $\mathcal{X}_k$ is a neighborhood of $x_*$ containing $x_k$ and $y_k$ (remember that $y_k \to x_*$ by hypothesis). Also, from (36), we can increase if necessary the terms of the sequence $\{\epsilon_k\}$ to satisfy

$$\|(B_k - f'(x_*))s_k\| \le \epsilon_k \|s_k\| = \epsilon_k \|y_k - x_k\| \le \epsilon_k \|x_k - x_*\| + \epsilon_k \|y_k - x_*\| \quad (42)$$

for all $k \ge k_1$. Due to condition (37), $x_k \to x_*$ and $u_k \in R_k(x_k, x_*)$ for all $k$, we have

$$\|u_k\| = \|0 - u_k\| \le \sup\{\|0 - x\| \mid x \in R(x_k, x_*)\} = d^+(0, R_k(x_k, x_*)) \le \epsilon_k \|x_k - x_*\| \quad (43)$$

for all $k \ge k_1$. Using (41) and the fact that $x_k, y_k \in \mathcal{X}_k$, the second term of sum in (40) can be bounded as

$$
\begin{aligned}
\|f(y_k) - f(x_k) - f'(x_*)s_k\| &= \|f(y_k) - f(x_k) - f'(x_*)(y_k - x_k) + f(x_*) - f(x_*) + f'(x_*)(x_* - x_*)\| \\
&\le \|f(y_k) - f(x_*) - f'(x_*)(y_k - x_*)\| + \|f(x_k) - f(x_*) - f'(x_*)(x_k - x_*)\| \\
&\le \epsilon_k \|y_k - x_*\| + \epsilon_k \|x_k - x_*\|.
\end{aligned}
$$

Thus, we use the previous inequality, (42) and (43) in (40) we obtain

$$\|y_k - x_*\| \le 3\lambda \epsilon_k \|x_k - x_*\| + 2\lambda \epsilon_k \|y_k - x_*\|$$

for all $k \ge k_1$. As $\epsilon_k \to 0$, we can suppose without loss of generality that $k_1$ is such that $2\lambda \epsilon_k < 1$ for all $k \ge k_1$, and hence we conclude that

$$\|y_k - x_*\| \le \frac{3\lambda \epsilon_k}{1 - 2\lambda \epsilon_k} \|x_k - x_*\|$$

for all $k \ge k_1$. Now, from Remark 2.8 we have

$$\|x_{k+1} - x_*\| \le (1 + \sqrt{2\theta_k})\|y_k - x_*\| + \sqrt{2\theta_k}\|x_k - x_*\|$$

for all $k$. So, combining the previous two inequalities we arrive at

$$\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} \le r_k := (1 + \sqrt{2\theta_k})\frac{3\lambda \epsilon_k}{1 - 2\lambda \epsilon_k} + \sqrt{2\theta_k},$$

whenever $k \ge k_1$. Since $\epsilon_k \to 0$ and $\theta_k \to 0$, we have $r_k \to 0$ and thus $\{x_k\}$ converges to $x_*$ superlinearly. This concludes the proof. $\square$

The next result, which proof can be found in [7, Theorem 4.8], establishes a condition for the Broyden update to satisfy the Dennis-Moré condition (36) with $\widehat{s}_k = s_k := y_k - x_k$.

**Proposition 6.2.** *Consider a function $f : \mathbb{X} \to \mathbb{Y}$ and a point $x_* \in \mathbb{X}$ such that the derivative mapping $f'(x)$ is Lipschitz continuous around $x_*$ with respect to the Hilbert-Schmidt norm. Also, consider the Broyden update (35) in which $B_0 - f'(x_*)$ is a Hilbert-Schmidt operator. If the sequences $\{x_k\}$ and $\{y_k\}$ are linearly convergent to $x_*$ then they satisfy the Dennis-Moré condition (36).*

21

We finalize this section with the superlinear convergence of the QN-InexP method (Algorithm 1), which is a consequence of Propositions 6.1 and 6.2.

**Theorem 6.3.** *Consider the constrained generalized equation* (1) *with a solution* $x_*$ *and suppose that the derivative mapping* $f'(x)$ *is Lipschitz continuous around* $x_*$ *with respect to the Hilbert-Schmidt norm. Consider the* QN-InexP *method applied to* (1) *with the Broyden update* (35), *where* $B_0$ *is chosen to satisfy* (13) *and such that* $B_0 - f'(x_*)$ *is a Hilbert-Schmidt operator. If* $f + F$ *is strongly metrically subregular at* $x_*$ *for* 0, *then every sequence* $\{x_k\}$ *generated by* QN-InexP *that converges to* $x_*$ *is superlinearly convergent.*

## 7 Numerical experiments

In this section we consider the quasi-Newton scheme with matrices $B_k$ following (35). For simplicity, we suppose $\Omega = \mathbb{R}^n$. In order to write the inclusion $\bar{z} \in F(x)$ in a tractable way, we assume that the set $F(x)$ can be described by equality and inequality constraints, that is,

$$F(x) = \{z \in \mathbb{Y} \mid h(x, z) = 0, \ g(x, z) \leq 0\}, \tag{44}$$

where $h$ and $g$ are continuously differentiable functions. Thus

$$\bar{z} \in F(x) \ \Leftrightarrow \ h(x, \bar{z}) = 0, \ g(x, \bar{z}) \leq 0.$$

Therefore, determining a solution $y_k$ of the subproblem $0 \in f(x_k) + B_k(y - x_k) + F(y)$ is equivalent to find $y_k \in \mathbb{X}$ and $z_k \in \mathbb{Y}$ such that

$$z_k = -[f(x_k) + B_k(y_k - x_k)], \quad h(y_k, z_k) = 0, \quad g(y_k, z_k) \leq 0.$$

In turn, a sufficient condition for the above expressions to be satisfied is $(y_k, z_k)$ to be an optimal solution of

$$\min_{y,z} \frac{1}{2} \left\| z + \left[ f(x_k) + B_k(y - x_k) \right] \right\|_2^2 \tag{45a}$$

$$\text{subject to } h(y, z) = 0, \tag{45b}$$

$$g(y, z) \leq 0 \tag{45c}$$

with null objective value. This problem can be solved by standard nonlinear optimization methods. So, we have a practical way to test whether $\bar{z} \in F(x)$ or not, at least approximately since (45) is considered solved when an optimality accuracy is achieved. This is in agreement with the inexactness allowed in the IQN-InexP variant (Algorithm 2).

Taking into account (44), the stopping criterion $0 \in f(x_k) + F(x_k)$ is equivalent to

$$h(x_k, -f(x_k)) = 0, \quad g(x_k, -f(x_k)) \leq 0.$$

22

Thus, the following (approximate) criterion is natural to declare convergence:

$$e_k := \max\{\|h(x_k, -f(x_k))\|_\infty, \|g(x_k, -f(x_k))_+\|_\infty\} \leq 10^{-6}.$$

Note that $x_k \in C$ by construction in Algorithms 1 and 2, since the approximate projection in the sense of Definition 2.5 maintains feasibility.

In order to compute the iterate $x_{k+1} \in P_C(y_k, x_k, \theta_k)$ satisfying condition (11), we employ the Frank-Wolfe algorithm provided in [5], which is applied to the projection problem

$$\min_x \frac{1}{2}\|x - y_k\|_2^2 \quad \text{subject to} \quad x \in C. \tag{46}$$

We added large box-constraints $\ell \leq x \leq u$ to the above problem, since the Frank-Wolfe algorithm is designed to deal with compact convex feasible sets.

In the next sections, we present three illustrative numerical examples of our theory. Our implementation is made in Matlab© R2016b using double precision. All tests were run on GNU/Linux Ubuntu 20.04. To solve problem (45) we use the interior point method implemented in the `fmincon` routine with maximum number of iterations equals to 1000 and optimality tolerance $10^{-6}$. Regarding the projection step, the sequence $\{\theta_k\}$ was chosen as $\theta_0 = 10^{-2}$ and $\theta_k = \max\{0.9\,\theta_{k-1}, 10^{-8}\}$, $k \geq 1$. This choice is purely empiric, and tries to reflect the inexact computation allowed by theory (note that even $\theta_k \to 0$ is not required in Theorems 3.2 and 4.2). A more accurate choice may be supported by the inequality in Remark 2.8, but more research is necessary on this topic. The gradients of $h_i$ and $g_j$ are given in each case. Finally, since it is required that $x_0 \in C$, we project the given initial point onto $C$ if necessary by solving (46) once with Matlab's interior point method (the Frank-Wolfe algorithm of [5] also needs to be initialized within $C$).

## 7.1 Problem 1

Similarly to [6], let us consider $f : \mathbb{R}^2 \to \mathbb{R}^2$, $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ and $C$ given by

$$f(x_1, x_2) = (3x_1^3 - 2x_1^2, 0),$$

$$F(x_1, x_2) = \begin{cases} \{-x_1, x_1\} \times \{0\}, & x_1 \geq 0 \\ \emptyset \times \{0\}, & x_1 < 0 \end{cases},$$

$$C = \mathbb{R} \times [1/2, 1].$$

The solutions of the generalized equation $f(x) + F(x) \ni 0$ subject to $x \in C$ are $(0, x_2)$ and $(1, x_2)$, $x_2 \in [1/2, 1]$.

There are different ways to write $F(x)$ as (44). Each of them affects the numerical resolution of the problem by a previously selected algorithm (in our case, the interior point method of Matlab©). For this example, we choose $F(x) = \{z \in \mathbb{R}^2 \mid z_1^2 - x_1 = 0,\ z_2 = 0\}$.

Table 1 shows the execution starting from $x_0 = (0.7, 0)$. Columns "iter" and "conv. rate" stands for "iteration" and $\|x_k - x_*\|_\infty / \|x_{k-1} - x_*\|_\infty$, respectively, where $x_* = (0, 0.5)$

is the point to that algorithm approximates. From the third column, we can observe the linear convergence rate of $\{x_k\}$ to $x_*$ related in the theory (e.g. Theorem 3.2). We run the algorithm from several distinct initial points, and in all cases it converges to $(0, \min\{1, \max\{0.5, (x_0)_2\}\})$. That is, the sequence $\{(x_2)_k\}$ converges to the extreme of the interval $[1/2, 1]$ if the initial value is not in this interval, while otherwise it is constant. This behaviour is expected and reflects the initial projection onto $C$. The algorithm has a preference for solutions with $x_1 = 0$. This probably occurs by the fact that the term $3x_1^3$ of $f$ tends to be minimized in the objective function of (45). Evidently, the algorithm reaches a valid solution of the problem.

| iter | $e_k$ | conv. rate |
|------|-------|------------|
| 1 | 6.975990e$-$01 | |
| 2 | 3.075904e$-$01 | 4.554906e-01 |
| 3 | 4.694257e$-$02 | 1.472804e-01 |
| 4 | 1.808056e$-$04 | 3.850256e-03 |
| 5* | 2.367476e$-$10* | 1.309404e-06 |

Table 1: Resolution of Problem 1 starting from (0.7, 0.6).

## 7.2    Problem 2

The next problem is a modification of Problem 1 that imposes a stricter relationship between the variables $x_1$ and $x_2$:

$$f(x_1, x_2) = (3x_1^3 - 2x_1^2, -x_1 x_2),$$

$$F(x_1, x_2) = \left\{ \begin{array}{ll} \{-x_1, x_1\} \times \{1 - x_1\}, & x_1 \geq 0 \\ \emptyset \times \{0\}, & x_1 < 0 \end{array} \right.,$$

$$C = \mathbb{R} \times [0, 2].$$

The solutions are $(0, x_2)$, $x_2 \in [0, 2]$, and $(1, 0)$. Here, $F(x) = \{z \in \mathbb{R}^2 \mid z_1^2 - x_1 = 0, \ z_2 = 1 - x_1\}$ for $x_1 \geq 0$.

The algorithm is attracted to $(0, 1)$ for different initial point, including the solution $(0, 0)$. Evidently this depends on the implementation, and can be justified by the way that we initialize $z$ when solving problem (45). Table 2 shows the execution for $x_0 = (0.9, 0.5)$. As in Problem 1, we can observe a linear rate of convergence to $(0, 1)$.

| iter | $e_k$ | conv. rate |
|------|-------|------------|
| 1 | 5.785110e−01 | |
| 2 | 5.344404e−01 | 6.067537e−01 |
| 3 | 2.963137e−01 | 5.615067e−01 |
| 4 | 1.049330e−01 | 3.465348e−01 |
| 5 | 3.895669e−02 | 4.236108e−01 |
| 6 | 6.432412e−04 | 1.409197e−02 |
| 7 | 6.269612e−06 | 9.883571e−03 |
| 8* | 2.101206e−09* | 3.351656e−04 |

Table 2: Resolution of Problem 2 starting from (0.9, 0.5).

## 7.3 Problem 3

Here we consider the bilevel problem

$$\min_{x,y} \phi(x,y)$$

$$\text{subject to } (x,y) \in D$$

$$y \in \arg\min_{y} q(x,y)$$

$$\text{subject to } Yy \le c - Xx,$$

where $x \in R^n$, $y \in \mathbb{Y}$ and $c \in \mathbb{R}^p$. We assume that $D \ne \emptyset$ is closed and convex, and that $q(x,\cdot)$ is a continuously differentiable convex function for all $x$.

It is well known that computing a feasible point $(x,y)$ of the above problem is NP-hard in general, since $y$ must be a global minimizer of an optimization problem. This difficult is increased if the upper level constraints $(x,y) \in D$ contain lower level variables $y_i$'s. For details on bilevel problems, see [16]. One way to deal with bilevel problems is rewriting the lower level problem under their Karush-Kuhn-Tucker (KKT) conditions:

$$\nabla_y q(x,y) + Y^t \lambda = 0$$

$$\lambda \ge 0$$

$$(Xx + Yy - c)^t \lambda = 0$$

$$Yy \le c - Xx.$$

Since $q(x,\cdot)$ is convex, these conditions are sufficient to optimality of the lower level problem. Defining

$$F(x,y,\lambda) = \{0\}^2 \times N_{\mathbb{R}}(\lambda_1) \times \cdots \times N_{\mathbb{R}}(\lambda_p)$$

$$f(x,y,\lambda) = \left( \nabla_y q(x,y) - Y^t \lambda , \ Xx + Yy - c \right)$$

$$C = \left( D \cap \{(x,y) \mid Xx + Yy \le c\} \right) \times \mathbb{R}_+^p,$$

25

| iter | $e_k$ |
| --- | --- |
| 1 | 1.428588e-01 |
| 2 | 1.568136e-01 |
| 3 | 2.339251e-01 |
| 4 | 1.008833e-01 |
| 5 | 3.987558e-02 |
| $\vdots$ | $\vdots$ |
| 32 | 1.677399e-03 |
| 33 | 1.645132e-03 |
| 34 | 1.614623e-03 |
| 35* | 3.653530e-08* |

Table 3: Resolution of Problem 3.

a point $(x, y)$ is feasible for the bilevel problem if, and only if, there is a $\lambda$ such that

$$0 \in f(x, y, \lambda) + F(x, y, \lambda), \qquad (x, y, \lambda) \in C.$$

Note that the above generalized equation does not involve inequalities. They are totally encapsulated by the projection step.

To illustrate the functionality of our algorithm, let us consider the particular instance

$$q(x, y) = \frac{1}{2}(\|x\|^2 + \|y\|^2)$$

$$X = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}, \quad Y = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad c = \begin{bmatrix} -1 \\ -1 \end{bmatrix},$$

$$D = \{(x, y) \in \mathbb{R}^4 \mid x_1 + x_2 + y_1 + y_2 \le 1\}$$

In general, our implementation converges to different points when we vary the initial point. This is in agreement with the fact that, for each $x$, the lower level problem possibly admits a different minimizer. Starting from $(x_0, y_0, \lambda_0) = (0, 0, 0, 0, 0, 0)$, the algorithm reaches

$$x_* \approx (0.4012 , -0.2009 ), \quad y_* \approx ( 2.8779 \cdot 10^{-8} , 3.6537 \cdot 10^{-8} ), \quad \lambda_* \approx ( 0.0000 , 0.0000 ),$$

with $e_* \approx 3.65 \cdot 10^{-8}$ (in this case, $e_k$ encompasses the KKT residue of the lower level problem). The execution is presented in Table 3.

We do not calculate the "rate of convergence" for Problem 3 since it requires to know *a priori* the solution set of the generalized equation, i.e., the feasible set of the bilevel problem. So, it is hard to decide numerically if the iterate $(x_k, y_k, \lambda_k)$ is close to the solution set, or even close to a particular solution. On the other hand, it is not reasonable use the last iterate $(x_*, y_*, \lambda_*)$ to make this computation because this point carries approximation errors, leading to a false estimative for $\|x_k - x_*\|_\infty / \|x_{k-1} - x_*\|_\infty$.

26

# 8  Conclusions

In this paper, we deal with constrained generalized equations. This is a very general class of problems, encompassing several other contexts, such as standard nonlinear optimization, variational inequalities and equilibrium problems. We presented two general quasi-Newton frameworks for solving constrained generalized equations that employ an inexact projection step. Firstly, we discuss a quasi-Newton scheme where subproblems are solved exactly. Its local convergence is provided under a bounded deterioration condition on the update quasi-Newton operator/matrix. Secondly, we extend the proposed method allowing subproblems to be solved inexactly. The resulting (inexact quasi-Newton) method is closer to the numerical practice, where inexactness is naturally present. We also proved that the inexact scheme converges locally under mild assumptions.

We analysed the particular case when the classical Broyden update rule is employed. For both exact and inexact quasi-Newton methods, we show that, in this case, the deterioration condition is satisfied directly. Illustrative numerical experiments with the Broyden variant were made to align theory with practice.

As future works, we intend to answer whether the condition of bounded deterioration can be replaced by something weaker or whether, with additional assumptions, we could obtain another update rule besides Broyden's. Another line of research is to study the variational inequality problem

$$0 \in f(x) + N_C(x),$$

by considering the inexact quasi-Newton method

$$(f(x_k) + (B_k + \bar{B})(x_{k+1} - x_k) + N_C(x_{k+1})) \cap \mathbb{B}_{\eta_k \psi(x_k)}(0) \neq \emptyset, \quad k = 0, 1, \cdots, \qquad (47)$$

with $B_k$ satisfying the Broyden update, $\bar{B}$ an $n \times n$ matrix, $\{\eta_k\}$ a sequence of positive numbers converging to zero and $\psi$ a Lipschitz function. As in [22], we intend to consider Newton-Kantorovich theorem on (47). Moreover, we propose to apply the quasi-Newton iteration

$$\nabla h(x_k) + B_k(x_{k+1} - x_k) + N_C(x_{k+1}) \ni 0 \qquad (48)$$

to the first-order necessary optimality condition $\nabla h(x) + N_C(x) \ni 0$ of the optimization problem

$$\text{minimize} \quad h(x) \quad \text{subject to} \quad x \in C,$$

where $h$ is a twice continuously differentiable function and $C$ is a closed and convex set. We propose to update $B_k$ defined in (48) firstly by the Broyden update (35) with $z_k = \nabla h(x_{k+1}) - \nabla h(x_k)$ and $s_k = x_{k+1} - x_k$, and secondly using the BFGS method, that is,

$$B_{k+1} := B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k},$$

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

It is worth noting that, as reported in [22], the scheme (48) can be used for solving control-constrained optimal control problems. So, we also propose to do numerical experiments to the quasi-Newton method (48) and to compare the results with the ones obtained in [22].

During the review process, an anonymous referee asks about the possibly of weakening the assumption that $f + F$ is metrically regular in our convergence results. This is a very interesting issue. In fact, as pointed out recently in [40], metrically regularity fails to hold in important situations. So, a topic for further research is to establish the convergence of Quasi-Newton schemes under, for instance, Hölder-type hypotheses. Also, besides Definition 2.5, other type of projections should be considered in the convergence analysis, especially those with numerical appeal.

## Acknowledgments

## References

[1] M. Abbas, M. AlShahrani, Q. H. Ansari, O. S. Iyiola, and Y. Shehu. Iterative methods for solving proximal split minimization problems. *Numerical Algorithms*, 78(1):193–215, 2018.

[2] S. Adly, R. Cibulka, and H. V. Ngai. Newton's method for solving inclusions using set-valued approximations. *SIAM Journal on Optimization*, 25(1):159–184, 2015.

[3] S. Adly and V. N. Huynh. Quasi-Newton methods for solving nonsmooth equations: generalized Dennis-Moré theorem and Broyden's update. *J. Convex Anal.*, 25(4):1075–1104, 2018.

[4] S. Adly, H. Van Ngai, and V. V. Nguyen. Newton's method for solving generalized equations: Kantorovich's and Smale's approaches. *Journal of Mathematical Analysis and Applications*, 439(1):396–418, 2016.

[5] A. A. Aguiar, O. P. Ferreira, and L. F. Prudente. Subgradient method with feasible inexact projections for constrained convex optimization problems. *Optimization*, pages 1–23, 2021.

[6] F. J. Aragón Artacho, A. Belyakov, A. L. Dontchev, and M. López. Local convergence of quasi-Newton methods under metric regularity. *Computational Optimization and Applications*, 58(1):225–247, 2014.

[7] F. J. A. Artacho, A. L. Dontchev, M. Gaydu, M. H. Geoffroy, and V. M. Veliov. Metric regularity of Newton's iteration. *SIAM Journal on Control and Optimization*, 49(2):339–362, 2011.

[8] D. Azé and C. C. Chou. On a Newton type iterative method for solving inclusions. *Mathematics of Operations Research*, 20(4):790–800, 1995.

[9] D. Bertsekas. *Nonlinear Programming*. Optimization and Computation Series. Athena Scientific, Belmont, MA, 2 edition, 1999.

[10] J. F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization*, 29(2):161–186, 1994.

[11] Y. Censor, A. Gibali, and S. Reich. Algorithms for the split variational inequality problem. *Numerical Algorithms*, 59(2):301–323, 2012.

[12] Y. Censor, A. Gibali, S. Reich, and S. Sabach. Common solutions to variational inequalities. *Set-Valued and Variational Analysis*, 20(2):229–247, 2012.

[13] R. Cibulka, A. Dontchev, and A. Kruger. Strong metric subregularity of mappings in variational analysis and optimization. *Journal of Mathematical Analysis and Applications*, 457(2):1247–1282, 2018. Special Issue on Convex Analysis and Optimization: New Trends in Theory and Applications.

[14] F. R. de Oliveira, O. P. Ferreira, and G. N. Silva. Newton's method with feasible inexact projections for solving constrained generalized equations. *Computational Optimization and Applications*, 72(1):159–177, 2019.

[15] R. S. Dembo, S. C. Eisenstat, and T. Steihaug. Inexact Newton methods. *SIAM Journal on Numerical Analysis*, 19(2):400–408, 1982.

[16] S. Dempe. *Foundations of Bilevel Programming*, volume 61 of *Nonconvex Optimization and Its Applications*. Springer US, Dordrecht, 1 edition, 2002.

[17] J. E. Dennis and J. J. Moré. A characterization of superlinear convergence and its application to quasi-Newton methods. *Mathematics of Computation*, 28(126):549–560, 1974.

[18] J. E. Dennis and J. J. Moré. Quasi-Newton methods, motivation and theory. *SIAM Review*, 19(1):46–89, 1977.

[19] A. Dontchev. Local analysis of a Newton-type method based on partial linearization. In *The mathematics of numerical analysis (Park City, UT, 1995)*, volume 32, pages 295–306. Amer. Math. Soc., Providence, RI, 1996.

[20] A. L. Dontchev. Generalizations of the Dennis–Moré theorem. *SIAM Journal on Optimization*, 22(3):821–830, 2012.

[21] A. L. Dontchev and W. W. Hager. An inverse mapping theorem for set-valued maps. *Proceedings of the American Mathematical Society*, 121(2):481–489, 1994.

[22] A. L. Dontchev, M. Huang, I. V. Kolmanovsky, and M. M. Nicotra. Inexact Newton-Kantorovich methods for constrained nonlinear model predictive control. *IEEE Transactions on Automatic Control*, 64(9):3602–3615, 2019.

[23] A. L. Dontchev and R. T. Rockafellar. *Implicit Functions and Solution Mappings: A View from Variational Analysis*. Springer Monographs in Mathematics. Springer, New York, NY, 1 edition, 2009.

[24] A. L. Dontchev and R. T. Rockafellar. Convergence of inexact Newton methods for generalized equations. *Mathematical Programming*, 139(1):115–137, 2013.

[25] O. Ferreira and G. Silva. Local convergence analysis of Newton's method for solving strongly regular generalized equations. *Journal of Mathematical Analysis and Applications*, 458(1):481–496, 2018.

[26] O. P. Ferreira. A robust semi-local convergence analysis of Newton's method for cone inclusion problems in Banach spaces under affine invariant majorant condition. *J. Comput. Appl. Math.*, 279:318–335, 2015.

[27] O. P. Ferreira and G. N. Silva. Kantorovich's theorem on Newton's method for solving strongly regular generalized equation. *SIAM Journal on Optimization*, 27(2):910–926, 2017.

[28] M. L. Gonçalves and J. G. Melo. A Newton conditional gradient method for constrained nonlinear systems. *Journal of Computational and Applied Mathematics*, 311:473–483, 2017.

[29] H. He, C. Ling, and H.-K. Xu. A relaxed projection method for split variational inequalities. *Journal of Optimization Theory and Applications*, 166(1):213–233, 2015.

[30] N. Josephy. *Newton's method for generalized equations and the pies energy model*. PhD thesis, University of Wisconsin–Madison, 1979.

[31] C. T. Kelley. *Iterative methods for linear and nonlinear equations*. SIAM, 1995.

[32] C. T. Kelley and E. W. Sachs. A new proof of superlinear convergence for Broyden's method in Hilbert space. *SIAM Journal on Optimization*, 1(1):146–150, 1991.

[33] D. Klatte and B. Kummer. Approximations and generalized Newton methods. *Mathematical Programming*, 168(1):673–716, 2018.

[34] A. Moudafi. Split monotone variational inclusions. *Journal of Optimization Theory and Applications*, 150(2):275–283, 2011.

[35] S. M. Robinson. Extension of Newton's method to nonlinear functions with values in a cone. *Numerische Mathematik*, 19(4):341–347, 1972.

[36] S. M. Robinson. *Generalized equations and their solutions, Part I: Basic theory*, pages 128–141. Springer Berlin Heidelberg, Berlin, Heidelberg, 1979.

[37] S. M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5(1):43–62, 1980.

[38] S. M. Robinson. *Generalized Equations*, pages 346–367. Springer Berlin Heidelberg, Berlin, Heidelberg, 1983.

[39] E. W. Sachs. Broyden's method in Hilbert space. *Mathematical Programming*, 35(1):71–82, 1986.

[40] L. L. Wei Ouyang. Hölder strong metric subregularity and its applications to convergence analysis of inexact Newton methods. *Journal of Industrial & Management Optimization*, 17(1):169–184, 2021.