To appear in *Optimization Methods & Software*
Vol. 00, No. 00, Month 20XX, 1–16

RESEARCH ARTICLE

A quasi-Newton modified Linear-Programming-Newton Method

María de los Ángeles Martínez[a*] and Damián Fernández[ab]

[a]*CIEM, CONICET, Ciudad Universitaria, X5000HUA Córdoba, Argentina;* [b]*FaMAF, Universidad Nacional de Córdoba, Ciudad Universitaria, X5000HUA Córdoba, Argentina*

(*Received 00 Month 20XX; final version received 00 Month 20XX*)

We consider a method to solve constrained system of nonlinear equations based on a modification of the Linear-Programming-Newton method and replacing the first order information with a quasi-Newton secant update, providing a computationally simple method. The proposed strategy combines good properties of two methods: the least change secant update for unconstrained system of nonlinear equations with isolated solutions and the Linear-Programming-Newton for constrained nonlinear system of equations with possible nonisolated solutions. We analyze the local convergence of the proposed method under a standard error bound condition proving its linear convergence for nonisolated solutions. Numerical experiments were done in order to show the proposed convergence rate.

Keywords: constrained nonlinear system of equations; nonisolated solutions; quasi-Newton method; local convergence.

AMS Subject Classification: 90C30; 65K05

1. Introduction

The aim of this work is to present a quasi-Newton type method for the solution of constrained system of equations and analyze its local convergence properties. The proposed method will help us to solve the problem of finding z such that

$$F(z) = 0, \qquad z \in \Omega \tag{1}$$

where $\Omega \subseteq \mathbb{R}^n$ is a nonempty and closed set and $F : \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable function with F' locally Lipschitz continuous.

In order to solve the problem (1) we propose the following iterative procedure.

Algorithm 1
Step 0: choose $\kappa > 0$, $z^0 \in \Omega$, $M_0 \in \mathbb{R}^{m\times n}$ and set $k = 0$.
Step 1: define $z^{k+1} = z^k + d^k$ where (d^k, γ_k) is a solution of

$$\begin{array}{ll} \underset{d,\gamma}{\text{minimize}} & \gamma \\ \text{subject to} & \|F(z^k) + M_k d\| \leq \kappa\gamma, \\ & \|d\| \leq \gamma, \\ & z^k + d \in \Omega. \end{array}$$

*Corresponding author. Email: mmartinez@famaf.unc.edu.ar

Step 2: if $F(z^{k+1}) = 0$, stop. Else, compute $M_{k+1} \in \mathbb{R}^{m \times n}$.
Step 3: set $k = k + 1$ and go to Step 1.

This Algorithm is an adaptation of the Linear-Programming-Newton method proposed in [6], where quadratic convergence was proved for nonisolated solutions and for relaxed smoothness assumptions on F. In [6], M_k is a suitable substitute of the Jacobian of F at z^k. However, the calculation of this replacement of the Jacobian (or the exact one) may be error-prone and expensive. This kind of issue is well-known in the literature of unconstrained nonlinear system of equations and can be managed by using quasi-Newton methods [1, 4]. Thus we proposed an adaptation in the spirit of quasi-Newton methods, by replacing the computation of the exact Jacobian by a computationally less expensive approximation with a convenient updating rule.

On the other hand, for unconstrained system of equations quasi-Newton methods are used to link the advantage of the local behavior of the standard Newton method with a consistent globalization strategy. In [12] it was shown that a globally convergent method can be formulated for F continuously differentiable or F in certain class of piecewise continuously differentiable mappings. Also, subproblem [12, problem (5.9)] was introduced to have a better scaled Linear-Programming subproblem. We shall stress that the subproblem in Step 1 of Algorithm 1 is an adaptation of [12, problem (5.9)]. This adaptation consists in taking a fixed parameter κ instead of a variable parameter $\|F(z^k)\|$, change that was made to obtain a suitable feasible point of the subproblem in order to guarantee the fulfillment of a uniform error bound.

It is well-known that problem (1) can be reformulated as a nonlinear least squares minimization problem. But in this case minimization algorithms can achieve stationary points that do not necessarily solve problem (1). Moreover, for nonconvex objective functions, the local isolatedness of the stationary point is a standard hypothesis. However, we should mention the work [18] that studies gradient projection methods and where R-linear convergence was proved for noninsolated stationary points. In contrast, the proposed method does not require the calculation of the exact Jacobian and a Q-linear convergence is obtained for a noninsolated solution set.

In order to solve the nonlinear least squares minimization problem, it is known from the literature that the Levenberg-Marquardt method is one of the best options. Under mild assumptions it generates a sequence that converges quadratically to a possible noninsolated solution [8, 10, 22]. However, this method requires the computation of the exact Jacobian.

Another quasi-Newton strategy to solve problem (1) was proposed in [21]. This work deals with a nonlinear least squares minimization reformulation of problem (1) and attempt to solve it by using a trust-region method. Convergence of the proposed algorithm was shown assuming that points in $F^{-1}(0)$ are isolated and that F_i are continuously differentiable outside $F_i^{-1}(0)$ and semismooth on $F_i^{-1}(0)$. Also, we should mention methods developed for particular nonlinear systems given by a reformulation of a Karush-Kuhn-Tucker (KKT) system [7, 20] and a reformulation of a mixed complementarity problem [17].

In order to simplify the convergence analysis, we consider a least change secant quasi-Newton update matrix (see [5]). Also the lines of the convergence analysis follows [2]. We stress that the Broyden's update [3] and the Powell-symmetric-Broyden (PSB) update [19] are particular cases of this general scheme. In those cases were the approximation matrix must be symmetric positive definite, a slight modification can be done in order to incorporate the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update (see [9, 13, 14]). In contrast to standard quasi-Newton results, where local Q-superlinear convergence is showed, we can prove only local Q-linear convergence. This result is still meaningful since we can solve problems with nonisolated solutions and without convexity/monotonicity assumptions by solving simple subproblems and without computing the Jacobian.

Some words about our notation. We use $\|\cdot\|$ for a norm on $\mathbb{R}^n$, $\mathbb{R}^m$ and for its induced norm on $\mathbb{R}^{m\times n}$. We define $\mathcal{B}_r(z) = \{w \mid \|z-w\| \le r\}$ and $\operatorname{dist}(w,Z) = \inf_{z\in Z}\|z-w\|$. In the sequel, we will use $\|\cdot\|_\star$ for a norm on $\mathbb{R}^{m\times n}$ associated to an inner product and $\mathcal{B}^\star_r(z)$ for the ball in this norm.

2. Local convergence

For a given $\kappa > 0$, $w \in \Omega$ and $M \in \mathbb{R}^{m\times n}$ we consider the following optimization problem,

$$\begin{array}{ll} \underset{d,\gamma}{\text{minimize}} & \gamma \\ \text{subject to} & \|F(w)+Md\| \le \kappa\gamma, \\ & \|d\| \le \gamma, \\ & w+d \in \Omega. \end{array} \tag{2}$$

Some simple but useful properties of this problem can be summarized as follows.

PROPOSITION 2.1 *Let $w \in \Omega$. Then,*

(i) the pair $\left(0, \frac{1}{\kappa}\|F(w)\|\right)$ is feasible for problem (2),
(ii) the optimization problem (2) has a solution with $\gamma \le \frac{1}{\kappa}\|F(w)\|$, and
(iii) the optimal value of problem (2) is zero if and only if w is a solution of (1).

Proof. If $d = 0$ then $\|F(w)+Md\| = \|F(w)\| = \kappa\frac{1}{\kappa}\|F(w)\|$, $\|d\| = 0 \le \frac{1}{\kappa}\|F(w)\|$ and $w+d = w \in \Omega$. So, (i) follows.

Now, if we add to problem (2) the restriction $\gamma \le \frac{1}{\kappa}\|F(w)\|$, the solution set does not change. Since this new problem has a nonempty and compact feasible set with continuous objective function, we can guarantee the existence of solution. Thus (ii) is valid.

The proof of (iii) is trivial. If $\gamma = 0$, then $\|F(w)+Md\| = 0$, $\|d\| = 0$ and $w+d \in \Omega$, so $\|F(w)\| = 0$ and $w \in \Omega$, i.e., w is a solution for (1). On the other hand, if w is a solution of (1) then $(d,\gamma) = (0,0)$ is a solution of (2).

■

In order to define a quasi-Newton algorithm, we shall provide a rule to generate a suitable matrix M. So, let us consider a closed convex set $\mathcal{X} \subseteq \mathbb{R}^{m\times n}$ such that

$$F'(w) \in \mathcal{X}, \quad \forall\, w \in \Omega.$$

Then for $z, w \in \mathbb{R}^n$ and $M \in \mathbb{R}^{m\times n}$ consider the following problem

$$\begin{array}{ll} \underset{N}{\text{minimize}} & \|N-M\|_\star^2 \\ \text{subject to} & N(z-w) = F(z)-F(w), \\ & N \in \mathcal{X}. \end{array} \tag{3}$$

Since this is a (strongly) convex optimization problem, we have that M_+ is the (unique) solution if and only if

$$\langle M_+ - M, N - M_+\rangle \ge 0, \quad \forall\, N \text{ s.t. } N(z-w) = F(z)-F(w),\ N \in \mathcal{X},$$

where $\langle \cdot, \cdot \rangle$ is the inner product associated to $\| \cdot \|_\star$. Hence, it can be seen that

$$\|M - M_+\|_\star^2 \leq \|M - N\|_\star^2 - \|M_+ - N\|_\star^2, \tag{4}$$

for all N such that $N(z - w) = F(z) - F(w)$ and $N \in \mathcal{X}$. Note that if $\| \cdot \|_\star$ is the Frobenius norm we recover the Boyden's update when $\mathcal{X} = \mathbb{R}^{m\times n}$ and the PSB update when $\mathcal{X}$ is the space of symmetric matrices.

In order to analyze local convergence properties of Algorithm (1), first we state some assumptions. We assume that problem (1) has a nonempty solution set

$$Z = \{z \in \Omega \mid F(z) = 0\}.$$

Let $z^* \in Z$ denote an arbitrary but fixed solution of (1). We will assume the existence of a local error bound around z^*.

Assumption 1 There exist $\varepsilon_0 > 0$, $\ell > 0$ such that

$$\operatorname{dist}(w, Z) \leq \ell \|F(w)\|, \quad \forall\, w \in \mathcal{B}_{\varepsilon_0}(z^*) \cap \Omega.$$

We shall stress that this error bound condition can be obtained with suitable assumptions depending on the structure of F, see for example [11] for complementarity systems and [16] for generalized Nash equilibrium problems.

By the smoothness assumptions on F, there exist $L_0 > 0$, $L_1 > 0$ such that

$$\|F(z) - F(w)\| \leq L_0 \|z - w\|, \quad \|F'(z) - F'(w)\|_\star \leq L_1 \|z - w\|,$$

for all $z, w \in \mathcal{B}_{\varepsilon_0}(z^*)$. Thus, shrinking ε_0 if necessary, it holds that

$$\|F(w)\| \leq L_0 \operatorname{dist}(w, Z), \quad \forall\, w \in \mathcal{B}_{\varepsilon_0}(z^*). \tag{5}$$

Also, let $\beta > 0$ satisfies $\|M\| \leq \beta \|M\|_\star$ for all $M \in \mathbb{R}^{m\times n}$.

PROPOSITION 2.2 *Given $\kappa > 0$, there exist $\varepsilon_1 > 0$ and $\eta_1 > 0$ such that for any $w \in \mathcal{B}_{\varepsilon_1}(z^*)$ and $M \in \mathcal{B}^\star_{\eta_1}(F'(z^*))$, if $(\bar{d}, \bar{\gamma})$ is a solution of problem (2) then*

$$\begin{aligned} \|F(w) + M\bar{d}\| &\leq \kappa \operatorname{dist}(w, Z), & (6)\\ \|\bar{d}\| &\leq \operatorname{dist}(w, Z). & (7) \end{aligned}$$

Proof. If $w \in Z$, by Proposition 2.1(iii) and the second inequality constraint in (2), we have $(\bar{d}, \bar{\gamma}) = (0, 0)$. Then (6) and (7) hold.

Define

$$\varepsilon_1 = \min\left\{\frac{\varepsilon_0}{2}, \frac{\kappa}{3\beta L_1}\right\}, \quad \eta_1 = \frac{\kappa}{2\beta}.$$

For $w \in \mathcal{B}_{\varepsilon_1}(z^*) \setminus Z$ let $\hat{w} \in Z$ be so that $\|w - \hat{w}\| = \operatorname{dist}(w, Z)$. Then, for $d = \hat{w} - w$ we have

$$\|d\| = \operatorname{dist}(w, Z) \leq \|w - z^*\| \leq \varepsilon_1.$$

On the other hand, since $\|\hat{w} - z^*\| \leq 2\|w - z^*\| \leq \varepsilon_0$ and $M \in \mathcal{B}_{\eta_1}(F'(z^*))$,

$$\begin{aligned} \|F(w) + Md\| &\leq \|F(w) + F'(w)d - F(\hat{w})\| + \|(M - F'(w))d\| \\ &\leq \frac{\beta L_1}{2}\|d\|^2 + \beta \left(\|F'(z^*) - F'(w)\|_\star + \|M - F'(z^*)\|_\star\right)\|d\| \\ &\leq \left(\frac{\beta L_1}{2}\varepsilon_1 + \beta L_1 \varepsilon_1 + \beta \eta_1\right)\|d\| \\ &\leq \kappa\|d\| = \kappa \operatorname{dist}(w, Z). \end{aligned}$$

Since $(d, \operatorname{dist}(w, Z))$ is feasible for problem (2), then $\bar{\gamma} \leq \operatorname{dist}(w, Z)$. Hence $\bar{d}$ satisfies (6) and (7).

■

LEMMA 2.3 *Suppose that Assumption 1 holds. Then for $\kappa > 0$ there exist $\varepsilon_2 > 0$ and $\eta_2 > 0$ such that if $w \in \mathcal{B}_{\varepsilon_2}(z^*)$, $M \in \mathcal{B}^\star_{\eta_2}(F'(z^*))$ and $(\bar{d}, \bar{\gamma})$ is a solution of problem (2) then*

$$\operatorname{dist}(w + \bar{d}, Z) \leq 2\kappa\ell \operatorname{dist}(w, Z). \tag{8}$$

Proof. If $w \in Z$ then $(\bar{d}, \bar{\gamma}) = (0, 0)$, as previously shown. Then (8) holds.

Since F is continuously differentiable, there exists $r > 0$ such that if $w, w + d \in \mathcal{B}_r(z^*)$ the following holds

$$\|F(w + d) - F(w) - F'(z^*)d\| \leq \frac{\kappa}{2}\|d\|.$$

Define

$$\varepsilon_2 = \min\left\{\frac{\varepsilon_0}{4}, \varepsilon_1, \frac{r}{2}\right\}, \quad \eta_2 = \min\left\{\eta_1, \frac{\kappa}{2\beta}\right\}, \tag{9}$$

where ε_1, η_1 are given by Proposition 2.2.

Let $w \in \mathcal{B}_{\varepsilon_2}(z^*) \setminus Z$ and $M \in \mathcal{B}^\star_{\eta_2}(F'(z^*))$. By (7) we have

$$\|w + \bar{d} - z^*\| \leq \|w - z^*\| + \operatorname{dist}(w, Z) \leq 2\|w - z^*\| \leq \min\left\{\frac{\varepsilon_0}{2}, r\right\}.$$

Then,

$$\begin{aligned} \|F(w + \bar{d})\| &\leq \|F(w + \bar{d}) - F(w) - F'(z^*)\bar{d}\| + \beta\|F'(z^*) - M\|_\star\|\bar{d}\| \\ &\quad + \|F(w) + M\bar{d}\| \\ &\leq \tfrac{\kappa}{2}\|\bar{d}\| + \beta\eta_2\|\bar{d}\| + \kappa \operatorname{dist}(w, Z) \\ &\leq 2\kappa \operatorname{dist}(w, Z). \end{aligned}$$

Hence, by Assumption 1,

$$\operatorname{dist}(w + \bar{d}, Z) \leq \ell\|F(w + \bar{d})\| \leq 2\kappa\ell \operatorname{dist}(w, Z).$$

■

Proposition 2.4 *Suppose that Assumption 1 holds and let $\varepsilon_2 > 0$, $\eta_2 > 0$ be given by Lemma 2.3 for some $\kappa > 0$. If $w \in \mathcal{B}_{\varepsilon_2}(z^*)$, $M \in \mathcal{B}^{\star}_{\eta_2}(F'(z^*))$, $(\bar{d}, \bar{\gamma})$ is a solution of problem (2) and M_+ is the solution of problem (3) for $z = w + \bar{d}$, then for $c = 2L_1\left(\frac{1}{2} + \max\{1, 2\kappa\ell\}\right)$ it holds that*

$$\|M_+ - F'(\hat{z})\|_\star \leq \|M - F'(\hat{w})\|_\star + c\,\mathrm{dist}(w, Z),$$

where $\hat{z}, \hat{w} \in Z$ satisfy $\|\hat{z} - z\| = \mathrm{dist}(z, Z)$ and $\|\hat{w} - w\| = \mathrm{dist}(w, Z)$.

Proof. Let us define $N \in \mathbb{R}^{m \times n}$ such that

$$N = \int_0^1 F'(w + t\bar{d})dt.$$

It can be seen that N is feasible for problem (3). Then, by (4) we obtain

$$\|M_+ - N\|_\star \leq \|M - N\|_\star. \tag{10}$$

On the other hand, using (7) and (9) we have

$$\begin{array}{l} \|\hat{w} - z^*\| \leq \|\hat{w} - w\| + \|w - z^*\| \leq 2\|w - z^*\| \leq \frac{1}{2}\varepsilon_0, \\ \|\hat{z} - z^*\| \leq \|\hat{z} - z\| + \|z - z^*\| \leq 2\|z - z^*\| \leq 4\|w - z^*\| \leq \varepsilon_0. \end{array}$$

Hence $w, z, \hat{w}, \hat{z} \in \mathcal{B}_{\varepsilon_0}(z^*)$. Then

$$\begin{aligned} \|N - F'(\hat{z})\|_\star &\leq \int_0^1 \left\|F'(w + t\bar{d}) - F'(\hat{z})\right\|_\star dt \\ &\leq L_1 \int_0^1 \|w + t\bar{d} - \hat{z}\| dt \\ &\leq L_1 \int_0^1 \left(\|w + t\bar{d} - z\| + \|z - \hat{z}\|\right) dt \\ &\leq L_1 \int_0^1 (1 - t)\|\bar{d}\| dt + L_1 2\kappa\ell \mathrm{dist}(w, Z) \\ &\leq L_1\left(\tfrac{1}{2} + 2\kappa\ell\right) \mathrm{dist}(w, Z), \end{aligned}$$

where we use that $z = w + \bar{d}$, (7) and (8). In a similar form, we obtain

$$\|N - F'(\hat{w})\|_\star \leq L_1\left(\tfrac{1}{2} + 1\right) \mathrm{dist}(w, Z).$$

Then, using (10) and the inequalities above, we conclude that

$$\begin{aligned} \|M_+ - F'(\hat{z})\|_\star &\leq \|M_+ - N\|_\star + \|N - F'(\hat{z})\|_\star \\ &\leq \|M - F'(\hat{w})\|_\star + \|F'(\hat{w}) - N\|_\star + \|N - F'(\hat{z})\|_\star \\ &\leq \|M - F'(\hat{w})\|_\star + c\,\mathrm{dist}(w, Z), \end{aligned}$$

where $c = 2L_1\left(\frac{1}{2} + \max\{1, 2\kappa\ell\}\right)$.

■

THEOREM 2.5 *Suppose that Assumption 1 is satisfied. Let $\{z^k\}$, $\{d^k\}$ and $\{M_k\}$ be generated by Algorithm 1 choosing M_{k+1} in step 2 as the solution of problem (3) with $z = z^{k+1}$, $w = z^k$ and $M = M_k$. If $\kappa \le \frac{1}{4\ell}$, there exist $\varepsilon > 0$ and $\eta > 0$ such that if*

$$z^0 \in \mathcal{B}_\varepsilon(z^*), \quad M_0 \in \mathcal{B}^\star_\eta(F'(z^*)),$$

then

(i) the sequence $\{z^k\}$ converges to some $\bar{z} \in Z$, and $\{\operatorname{dist}(z^k, Z)\}$ converges to zero,
(ii) the sequence $\{\|M_k - F'(\bar{z})\|_\star\}$ converges.

Proof. Let ε_2, η_2 be given by Lemma 2.3. We will show by induction that for all $j \ge 0$

$$\|z^j - z^*\| \le \varepsilon_2, \tag{11}$$
$$\|M_j - F'(z^*)\|_\star \le \eta_2. \tag{12}$$

Define

$$\varepsilon = \min\left\{\frac{\varepsilon_2}{3}, \frac{\eta_2}{28L_1}\right\}, \quad \eta = \frac{\eta_2}{2}. \tag{13}$$

Let $z^0 \in \mathcal{B}_\varepsilon(z^*)$ and $M_0 \in \mathcal{B}^\star_\eta(F'(z^*))$. Then (11) and (12) hold for $j = 0$.

Now, suppose that (11) and (12) hold for all $j \le k$. Then, by Proposition 2.2 we have that

$$\begin{aligned} \|F(z^j) + M_j d^j\| &\le \kappa \operatorname{dist}(z^j, Z), && (14)\\ \|d^j\| &\le \operatorname{dist}(z^j, Z), && (15) \end{aligned}$$

for all $j \le k$. Hence,

$$\begin{aligned} \|z^{k+1} - z^*\| &\le \|z^{k+1} - z^0\| + \|z^0 - z^*\| \\ &\le \sum_{j=0}^{k} \|z^{j+1} - z^j\| + \|z^0 - z^*\| \\ &\le \sum_{j=0}^{k} \operatorname{dist}(z^j, Z) + \|z^0 - z^*\| \\ &\le \sum_{j=0}^{k} (2\kappa\ell)^j \operatorname{dist}(z^0, Z) + \|z^0 - z^*\| \\ &\le \left(\sum_{j=0}^{k} \frac{1}{2^j} + 1\right) \|z^0 - z^*\| \\ &\le 3\|z^0 - z^*\| \le 3\varepsilon \le \varepsilon_2, \end{aligned}$$

where we use (15), (8), the fact that $2\kappa\ell \le \frac{1}{2}$ and (13). On the other hand, by Proposition

2.4 and the fact that $2\kappa\ell < 1$, we obtain

$$\begin{aligned}
\|M_{k+1} - F'(\hat{z}^{k+1})\|_\star &\le \|M_k - F'(\hat{z}^k)\|_\star + 3L_1 \operatorname{dist}(z^k, Z) \\
&\le \|M_0 - F'(\hat{z}^0)\|_\star + 3L_1 \sum_{j=0}^{k} \operatorname{dist}(z^j, Z) \\
&\le \|M_0 - F'(\hat{z}^0)\|_\star + 3L_1 \sum_{j=0}^{k} \frac{1}{2^j} \operatorname{dist}(z^0, Z) \\
&\le \|M_0 - F'(\hat{z}^0)\|_\star + 6L_1 \operatorname{dist}(z^0, Z).
\end{aligned}$$

Thus,

$$\begin{aligned}
\|M_{k+1} - F'(z^*)\|_\star &\le \|M_{k+1} - F'(\hat{z}^{k+1})\|_\star + \|F'(\hat{z}^{k+1}) - F'(z^*)\|_\star \\
&\le \|M_0 - F'(\hat{z}^0)\|_\star + 6L_1\|z^0 - z^*\| + 2L_1\|z^{k+1} - z^*\| \\
&\le \|M_0 - F'(z^*)\|_\star + \|F'(z^*) - F'(\hat{z}^0)\|_\star + 12L_1\|z^0 - z^*\| \\
&\le \|M_0 - F'(z^*)\|_\star + 2L_1\|z^0 - z^*\| + 12L_1\varepsilon \\
&\le \eta + 14L_1\varepsilon \le \eta_2,
\end{aligned}$$

where we use that $\|\hat{z} - z^*\| \le \|\hat{z} - z\| + \|z - z^*\| \le 2\|z - z^*\|$. Hence (11) and (12) hold for $j = k+1$.

Now, since (11) and (12) hold for all $k \ge 0$, by (7) and Lemma 2.3 we have

$$\begin{aligned}
\|z^{k+j} - z^j\| &\le \sum_{i=j}^{k+j-1} \|z^{i+1} - z^i\| \\
&\le \sum_{i=j}^{k+j-1} \operatorname{dist}(z^i, Z) \\
&\le \sum_{i=j}^{k+j-1} \frac{1}{2^{i-j}} \operatorname{dist}(z^j, Z) \\
&\le 2\operatorname{dist}(z^j, Z) \\
&\le 2\frac{1}{2^j}\operatorname{dist}(z^0, Z).
\end{aligned}$$

So the sequence $\{z^k\}$ is a Cauchy sequence and thus, by the closedness of Z, it converges to some $\bar{z} \in Z$. Since $\operatorname{dist}(z^k, Z) \le \|z^k - \bar{z}\|$, then (i) holds. Also, taking limit for $k \to \infty$ we obtain

$$\|z^j - \bar{z}\| \le 2\operatorname{dist}(z^j, Z), \tag{16}$$

for all $j \ge 0$.

In order to show the convergence of $\{\|M_k - F'(\bar{z})\|_\star\}$, note that by Proposition 2.4, for $k > j$

$$\begin{aligned}\|M_k - F'(\hat{z}^k)\|_\star &\leq \|M_j - F'(\hat{z}^j)\|_\star + 3L_1 \sum_{i=j}^{k-1} \text{dist}(z^i, Z) \\ &\leq \|M_j - F'(\hat{z}^j)\|_\star + 3L_1 \sum_{i=j}^{k-1} \frac{1}{2^{i-j}} \text{dist}(z^j, Z) \\ &\leq \|M_j - F'(\hat{z}^j)\|_\star + 6L_1 \text{dist}(z^j, Z).\end{aligned}$$

Since $\hat{z}^k \to \bar{z}$ and F' is continuous, then we have

$$\limsup_{k\to\infty} \|M_k - F'(\bar{z})\|_\star \leq \|M_j - F'(\hat{z}^j)\|_\star + 6L_1 \text{dist}(z^j, Z).$$

Now, taking limit for $j \to \infty$ in the right-hand side we obtain

$$\limsup_{k\to\infty} \|M_k - F'(\bar{z})\|_\star \leq \liminf_{j\to\infty} \|M_j - F'(\bar{z})\|_\star.$$

Hence (ii) holds. ■

Corollary 2.6 *Under the hypotheses of Theorem 2.5, it holds that* $\{\text{dist}(z^k, Z)\}$ *converges linearly to* 0 *and* $\{z^k\}$ *converges linearly to* $\bar{z}$.

Proof. Let us define for any k

$$N_k = \int_0^1 F'(z^k + td^k)dt.$$

Then, by (4) for $w = z^k$, $z = z^{k+1}$, $M = M_k$ and using that N_k is feasible for this problem, we obtain

$$\|M_k - M_{k+1}\|_\star^2 \leq \|M_k - N_k\|_\star^2 - \|M_{k+1} - N_k\|_\star^2$$

By definition of N_k and continuity of F' we have that $N_k \to F'(\bar{z})$, concluding by Theorem 2.5(ii) that

$$M_{k+1} - M_k \to 0.$$

Also, since $M_{k+1}d^k = F(z^{k+1}) - F(z^k)$, by Proposition 2.2 we have

$$\begin{aligned}\|F(z^{k+1})\| &= \|F(z^k) + M_{k+1}d^k\| \\ &\leq \kappa \text{dist}(z^k, Z) + \|(M_{k+1} - M_k)d^k\| \\ &\leq (\kappa + \|M_{k+1} - M_k\|)\text{dist}(z^k, Z).\end{aligned}$$

Then,

$$\begin{aligned} \frac{\text{dist}(z^{k+1}, Z)}{\text{dist}(z^k, Z)} &\leq \frac{\ell \|F(z^{k+1})\|}{\text{dist}(z^k, Z)} \\ &\leq \ell(\kappa + \|M_{k+1} - M_k\|) \to \ell\kappa \leq \frac{1}{4}. \end{aligned}$$

Thus, by (16) and the fact that $\text{dist}(z^k, Z) \leq \|z^k - \bar{z}\|$,

$$\frac{\|z^{k+1} - \bar{z}\|}{\|z^k - \bar{z}\|} \leq \frac{2\text{dist}(z^{k+1}, Z)}{\text{dist}(z^k, Z)} \to 2\ell\kappa \leq \frac{1}{2}.$$

■

3. Computational results

In the previous section we analyzed the convergence of our algorithm under certain hypotheses, in this section we are going to show the performance of the algorithm not only for problems which satisfy the hypotheses mentioned above but also for others which do not satisfy them.

The first example that we present satisfies all the required hypotheses, the second one is given by a nonlinear complementarity problem that does not satisfy Assumption 1, and they both were taken from [6]. Since problem (1) includes a broad range of problems and not only KKT reformulations, the rest of the examples are feasible point problems taken from the Hock-Schittkowski Collection [15] where we reformulated the feasible sets by considering it as in (1), despite possible violation of Assumption 1.

We wrote an Octave implementation of Algorithm 1, using the Simplex method implemented in the built-in function glpk for solving subproblem (2) and taking matrices defined by the Broyden's update with $M_0 = F'(z^0)$. The stopping criteria used were

(1) $\|F(z^k)\|_\infty <$ 1e-10 (residual error),
(2) $\|z^{k+1} - z^k\| <$ 1e-16 and,
(3) itmax = 1500 (maximum number of iterations).

Most of times Algorithm 1 stopped for the first criterion. We marked with an asterisk those cases where it stopped for the second one. We present below tables where final residuals and numbers of iterations for different values of κ are shown.

Example 3.1 Consider the following system of two inequalities

$$\begin{aligned} z_1^2 + z_2^2 - 1 &\leq 0, \\ (z_1 - 1)^2 + z_2^2 - 1 &\leq 0. \end{aligned}$$

Taking $\Omega = \mathbb{R}^2 \times \mathbb{R}^2_+$, this can be written as

$$F(z) = \begin{pmatrix} z_1^2 + z_2^2 - 1 + z_3 \\ (z_1 - 1)^2 + z_2^2 - 1 + z_4 \end{pmatrix} = 0, \quad z \in \Omega,$$

with slack variables z_3 and z_4.

Since the system of inequalities satisfies the Mangasarian-Fromovitz constraint qualification at any feasible point, then it can be shown that F satisfies Assumption 1 at any $z^* \in Z$. Furthermore, F is differentiable and its derivate is locally Lipschitz continuous, so all the previous conditions are satisfied. Then, we illustrate our convergence result with this example.

In Table 1 numerical results are shown.

Table 1. Numerical results taking $z^0 = (2, 0, 0, 1)$.

κ	$\sharp$ Iterations	residual
0.1	14	2.7756e-17
1e-2	11	2.2204e-16
1e-3	10	4.1453e-11
1e-4	9	8.4163e-12
1e-5	8	3.0211e-11
1e-6	8	8.6729e-12

In the following graphics, we show the quotient between $\|z^{k+1} - \bar{z}\|$ and $\|z^k - \bar{z}\|$ (left) and, $\log(\|z^{k+1} - z^k\|)$ (right) for each iteration k and some values of κ, where $\bar{z}$ is the last iterate. Note that the convergence of the sequence was at least linear for small values of κ. This fact can be observed in Figures 1, 2 and 3.

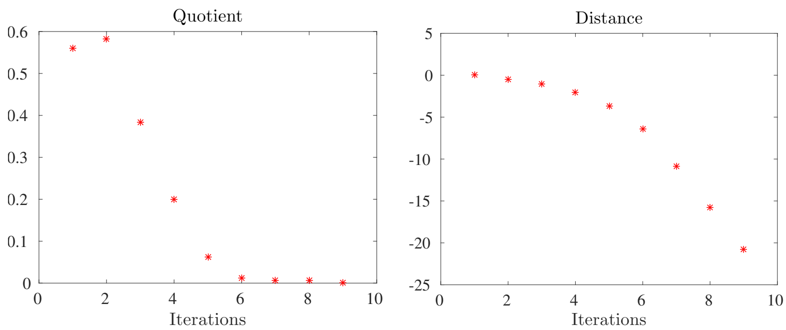

Figure 1. $\frac{\|z^{k+1}-\bar{z}\|}{\|z^k-\bar{z}\|}$ and $\log(\|z^{k+1} - z^k\|)$ for $\kappa = 10^{-4}$.

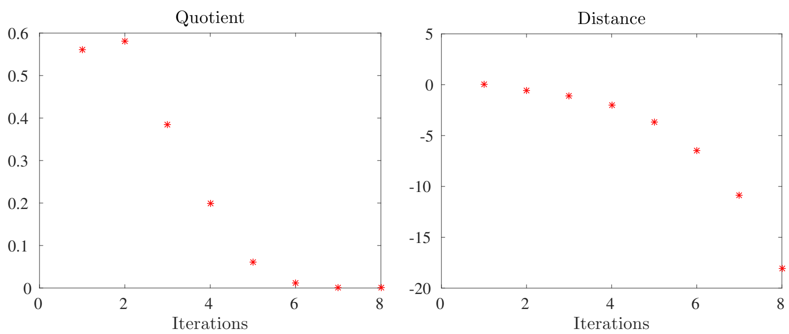

Figure 2. $\frac{\|z^{k+1}-\bar{z}\|}{\|z^k-\bar{z}\|}$ and $\log(\|z^{k+1} - z^k\|)$ for $\kappa = 10^{-5}$.

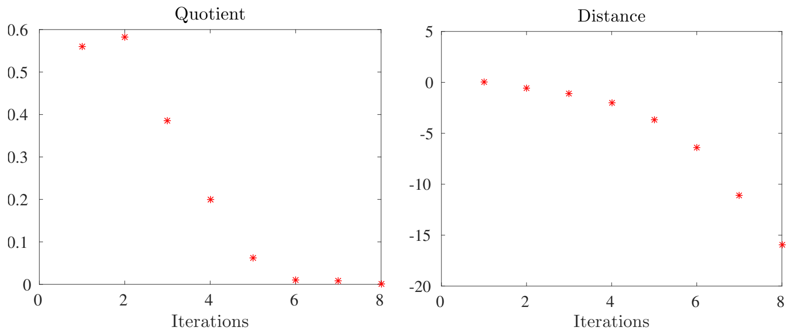

Figure 3. $\frac{\|z^{k+1}-\bar{z}\|}{\|z^k-\bar{z}\|}$ and $\log(\|z^{k+1}-z^k\|)$ for $\kappa = 10^{-6}$.

Example 3.2 As we mentioned at the beginning of the section, this problem was taken from [6]. Our intention is to show that although this problem does not satisfy Assumption 1, our algorithm can solved it generating a convergent sequence $\{z^k\}$.

We consider complementarity problem of the form

$$z \geq 0, \;\; T(z) \geq 0, \;\; z^T T(z) = 0,$$

where T is assumed to be differentiable with a locally Lipschitz continuous derivate.

For

$$T(z) = \begin{pmatrix} z_1 z_2 \\ z_1^2 + z_2 - 1 \end{pmatrix}$$

the problem can be formulated as

$$F(z) = \begin{pmatrix} z_1 z_2 - z_3 \\ z_1^2 + z_2 - 1 - z_4 \\ z_1 z_3 \\ z_2 z_4 \end{pmatrix}, \;\; z \in \Omega,$$

with slack variables z_3 and z_4 and $\Omega = \mathbb{R}^2_+ \times \mathbb{R}^2_+$.

Numerical results are shown in Table 2.

Table 2. Numerical results taking $z^0 = (1, 1, 1, 1)$ and $z^0 = (2, 1, 1, 0)$.

κ	♯ Iterations	residual
0.1	1101	3.6278e-10 *
1e-2	119	3.9433e-12
1e-3	28	9.9180e-11
1e-4	41	2.9165e-11
1e-5	44	1.6547e-11
1e-6	19	0

κ	♯ Iterations	residual
0.1	18	2.2204e-16
1e-2	16	2.3921e-11
1e-3	16	7.4402e-12
1e-4	15	2.9631e-11
1e-5	14	2.4669e-11
1e-6	15	3.2511e-12

Next examples were taken from the Hock-Schittkowski Collection.

Example 3.3 The feasible set of the problem HS19 can be written as (1) taking

$$F(z) = \begin{pmatrix} -(z_1 - 5)^2 - (z_2 - 5)^2 + 100 + z_3 \\ (z_1 - 6)^2 + (z_2 - 5)^2 - 82.81 + z_4 \end{pmatrix},$$

with $\Omega = [13, 100] \times [0, 100] \times \mathbb{R}^2_+$.

Numerical results are shown in Table 3.

Table 3. Numerical results taking $z^0 = (20, 5, 0, 0)$.

κ	♯ Iterations	residual
0.1	15	2.6512e-13
1e-2	13	8.6597e-15
1e-3	10	2.8451e-15
1e-4	9	1.4140e-11
1e-5	8	4.5191e-12
1e-6	7	4.8402e-11

Example 3.4 The feasible set of the problem HS23 can be written as (1) taking

$$F(z) = \begin{pmatrix} -z_1 - z_2 + 1 + z_3 \\ -z_1^2 - z_2^2 + 1 + z_4 \\ -9z_1^2 - z_2^2 + 9 + z_5 \\ -z_1^2 + z_2 + z_6 \\ -z_2^2 + z_1 + z_7 \end{pmatrix},$$

with $\Omega = [-50, 50] \times [-50, 50] \times \mathbb{R}^5_+$.

Numerical results are shown in Table 4.

Table 4. Numerical results taking $z^0 = (2, 2, 0, 0)$.

κ	♯ Iterations	residual
0.1	48	6.1963e-11
1e-2	14	2.8282e-11
1e-3	13	5.7998e-13
1e-4	11	1.5838e-11
1e-5	12	6.5103e-13
1e-6	10	1.1697e-11

Example 3.5 The feasible set of the problem HS24 can be written as (1) taking

$$F(z) = \begin{pmatrix} -z_1/\sqrt{3} + z_2 + z_3 \\ -z_1 - \sqrt{3}z_2 + z_4 \\ z_1 + \sqrt{3}z_2 - 6 + z_5 \end{pmatrix},$$

with $\Omega = \mathbb{R}^5_+$.

Numerical results are shown in Table 5.

Example 3.6 The feasible set of the problem HS34 can be written as (1) taking

$$F(z) = \begin{pmatrix} -z_2 + \exp(z_1) + z_4 \\ -z_3 + \exp(z_2) + z_5 \end{pmatrix},$$

with $\Omega = [0, 100] \times [0, 100] \times [0, 10] \times \mathbb{R}^2_+$.

Numerical results are shown in Table 6.

Table 5. Numerical results taking $z^0 = (2, 1, 0, 0)$.

κ	♯ Iterations	residual
0.1	15	5.7732e-15
1e-2	8	1.3323e-15
1e-3	6	8.8818e-16
1e-4	4	3.1264e-13
1e-5	3	4.9195e-11
1e-6	2	1.3234e-12

Table 6. Numerical results taking $z^0 = (0, 0, 0, 0)$.

κ	♯ Iterations	residual
0.1	13	4.8242e-12
1e-2	10	3.4037e-11
1e-3	11	7.5634e-16
1e-4	11	5.5511e-17
1e-5	10	7.5562e-11
1e-6	10	2.0197e-13

Example 3.7 The feasible set of the problem HS56 can be written as (1) taking

$$F(z) = \begin{pmatrix} z_1 - 8\sin^2(z_4) \\ z_2 - 8\sin^2(z_5) \\ z_3 - 8\sin^2(z_6) \\ z_1 + 2z_2 + 2z_3 - 14\sin^2(z_7) \end{pmatrix},$$

with $\Omega = \mathbb{R}^7$.

Numerical results are shown in Table 7.

Table 7. Numerical results taking $z^0 = (1, 0, 0, 0, 0, 0, 0)$.

κ	♯ Iterations	residual
0.1	7	9.4079e-08 *
1e-2	4	9.6098e-09 *
1e-3	3	0
1e-4	2	9.9980e-09 *
1e-5	2	0
1e-6	2	2.1176e-22

Example 3.8 The feasible set of the problem HS60 can be written as (1) taking

$$F(z) = \; z_1(1 + z_2^2) + x_3^4 - 4 - 3\sqrt{2},$$

with $\Omega = [-10, 10] \times [-10, 10] \times [-10, 10]$.

Numerical results are shown in Table 8.

Table 8. Numerical results taking $z^0 = (1, 1, 1)$.

κ	♯ Iterations	residual
0.1	11	3.5527e-15
1e-2	10	1.9540e-14
1e-3	8	4.4409e-14
1e-4	8	2.8422e-14
1e-5	8	6.5725e-14
1e-6	9	2.6645e-14

Example 3.9 The feasible set of the problem HS74 can be written as (1) taking

$$F(z) = \begin{pmatrix} -z_4 + z_3 - 0.55 + x_5 \\ -z_3 + z_4 - 0.55 + x_6 \\ 1000\sin(-z_3 - 0.25) + 1000\sin(-z_4 - 0.25) + 894.8 - z_1 \\ 1000\sin(z_3 - 0.25) + 1000\sin(z_3 - z_4 - 0.25) + 894.8 - z_2 \\ 1000\sin(z_4 - 0.25) + 1000\sin(z_4 - z_3 - 0.25) + 1294.8 \end{pmatrix},$$

with $\Omega = [0, 1200] \times [0, 1200] \times [-0.55, 0.55] \times [-0.55, 0.55] \times \mathbb{R}^2_+$.

Numerical results are shown in Table 9.

Table 9. Numerical results taking $z^0 = (800, 900, 0, 0, 0, 0)$.

κ	♯ Iterations	residual
0.1	342	5.9799e-11
1e-2	29	5.9711e-11
1e-3	16	1.7280e-11
1e-4	15	6.1164e-11
1e-5	13	2.6603e-11
1e-6	11	1.7963e-11

4. Final remarks

In this paper we have developed a quasi-Newton method for solving constrained systems of equations, based on the previous works [6, 12]. The proposed algorithm does not need any first order information, providing a computationally simple method that converges at least linearly. The rate of convergence is guaranteed even for nonisolated solutions.

Numerical examples show that the algorithm works well when all required hypotheses are satisfied, converging at least linearly as we expected. Nevertheless, if the hypotheses are not satisfied, our algorithm can still find a solution, as it is reflected in the second example. In general, for all the numerical experiments we found that the result improved notably as κ decreased. After a brief analysis of the results showed in Section 3, we conclude that a suitable value for the parameter κ may be less than 10^{-4}. Also, we observe that subproblem (2) is numerically stable (as suggested in [12]).

Funding

This work was partially supported by projects ANPCyT-PICT 2014 N°2534 and CONICET-PIP N°112-201101-00050. We would like to thank the anonymous referees for their valuable comments and thorough reading of the manuscript.

References

[1] P.T. Boggs, J.W. Tolle, and P. Wang, *On the local convergence of quasi-Newton methods for constrained optimization*, SIAM journal on control and optimization 20 (1982), pp. 161–171.
[2] J.F. Bonnans, *Local analysis of Newton-type methods for variational inequalities and nonlinear programming*, Applied mathematics & optimization 29 (1994), pp. 161–186.
[3] C.G. Broyden, *A class of methods for solving nonlinear simultaneous equations*, Mathematics of computation 19 (1965), pp. 577–593.
[4] C.G. Broyden, J.E. Dennis, and J.J. Moré, *On the local and superlinear convergence of quasi-Newton methods*, IMA Journal of Applied Mathematics 12 (1973), pp. 223–245.
[5] J.E. Dennis Jr and R.B. Schnabel, *Least change secant updates for quasi-Newton methods*, Siam Review 21 (1979), pp. 443–459.
[6] F. Facchinei, A. Fischer, and M. Herrich, *An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions*, Mathematical Programming 146 (2014), pp. 1–36.
[7] F. Facchinei, A. Fischer, C. Kanzow, and J. M. Peng, *A simply constrained optimization reformulation of KKT systems arising from variational inequalities*, Applied mathematics & optimization 40 (1999), pp. 19–37.
[8] J.y. Fan and Y.x. Yuan, *On the quadratic convergence of the levenberg-marquardt method without nonsingularity assumption*, Computing 74 (2005), pp. 23–39.
[9] D. Fernández, *A quasi-Newton strategy for the sSQP method for variational inequality and optimization problems*, Mathematical Programming 137 (2013), pp. 199–223.
[10] A. Fischer, *Local behavior of an iterative framework for generalized equations with nonisolated solutions*, Mathematical Programming 94 (2002), pp. 91–124.
[11] A. Fischer, M. Herrich, A.F. Izmailov, and M.V. Solodov, *Convergence conditions for newton-type methods applied to complementarity systems with nonisolated solutions*, Computational Optimization and Applications 63 (2016), pp. 425–459.
[12] A. Fischer, M. Herrich, A.F. Izmailov, and M.V. Solodov, *A globally convergent LP-Newton method*, SIAM Journal on Optimization 26 (2016), pp. 2012–2033.
[13] R. Fletcher, *A new variational result for quasi-Newton formulae*, SIAM Journal on Optimization 1 (1991), pp. 18–21.
[14] O. Güler, F. Gürtuna, and O. Shevchenko, *Duality in quasi-Newton methods and new variational characterizations of the DFP and BFGS updates*, Optimization Methods & Software 24 (2009), pp. 45–62.
[15] W. Hock and K. Schittkowski, *Test examples for nonlinear programming codes*, Journal of Optimization Theory and Applications 30 (1980), pp. 127–129.
[16] A.F. Izmailov and M.V. Solodov, *On error bounds and newton-type methods for generalized nash equilibrium problems*, Computational Optimization and Applications 59 (2014), pp. 201–218.
[17] C. Kanzow, *Strictly feasible equation-based methods for mixed complementarity problems*, Numerische Mathematik 89 (2001), pp. 135–160.
[18] Z.Q. Luo and P. Tseng, *Error bounds and convergence analysis of feasible descent methods: a general approach*, Annals of Operations Research 46 (1993), pp. 157–178.
[19] M.J. Powell, *A new algorithm for unconstrained optimization*, in *Nonlinear programming (Proc. Sympos., Univ. of Wisconsin, Madison, Wis., 1970)*, Academic Press, New York, 1970, pp. 31–65.
[20] H. Qi, L. Qi, and D. Sun, *Solving Karush–Kuhn–Tucker systems via the trust region and the conjugate gradient methods*, SIAM journal on optimization 14 (2003), pp. 439–463.
[21] M. Ulbrich, *Nonmonotone trust-region methods for bound-constrained semismooth equations with applications to nonlinear mixed complementarity problems*, SIAM Journal on Optimization 11 (2001), pp. 889–917.
[22] N. Yamashita and M. Fukushima, *On the Rate of Convergence of the Levenberg-Marquardt Method*, in *Topics in Numerical Analysis. Computing Supplementa*, G. Alefeld and X. Chen, eds., Vol. 15, Springer, Vienna, 2001, pp. 239–249.