A globally trust-region LP-Newton method for nonsmooth functions under the Hölder metric subregularity *

Letícia Becher † Damián Fernández ‡ Alberto Ramos §

April 30, 2021

Abstract

We describe and analyse a globally convergent algorithm to find a possible nonisolated zero of a piecewise smooth mapping over a polyhedral set, such formulation includes Karush-Kuhn-Tucker (KKT) systems, variational inequalities problems, and generalized Nash equilibrium problems. Our algorithm is based on a modification of the fast locally convergent *Linear Programming* (LP)-Newton method with a trust-region strategy for globalization that makes use of the natural merit function. The transition between the global to local convergence occurs naturally under mild assumption. Our local convergence analysis of the method is performed under a Hölder metric subregularity condition of the mapping defining the possibly nonsmooth equation and the Hölder continuity of its gradient mapping of the selection mapping. We present numerical results that show the feasibility of the approach.

Keywords: Nonlinear equations, Newton method, global convergence, superlinear convergence, LP-Newton method

1 Introduction

Nonlinear system of equations are prominent in natural sciences, engineering, social science and many other areas. They appear frequently in the modelling of many real life applications, as fluid mechanics, quantum field theory, chemistry,

*This work was supported by CNPq (Grant 438185/2018-8, 307270/2019-0), CAPES, CONICET (PIP 11220150100500CO) and SeCyT-UNC (PID 33620180100326CB).

†Department of Mathematics, Federal University of Paraná, Curitiba, PR, Brazil. (leticiabecher2017@gmail.com).

‡CIEM, CONICET; FAMAF, Universidad Nacional de Córdoba, Argentina. (dfernandez@famaf.unc.edu.ar).

§Department of Mathematics, Federal University of Paraná, Curitiba, PR, Brazil. (albertoramos@ufpr.br).

economics, finance to cite few of them. Thus, we propose a method for solving constrained nonlinear system of the form

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

where $\Omega \in \mathbb{R}^n$ is a given polyhedral set and $F : \mathbb{R}^n \to \mathbb{R}^m$ is a continuous mapping. The set of solutions of (1) is denoted by $Z := \{x \in \Omega : F(x) = 0\}$, which is assumed to be nonempty. One of the most widely used algorithm for solving nonlinear systems of equations when $\Omega = \mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}^n$ is a smooth mapping is the Newton's method which is described in almost all text books on numerical analysis and continuous optimization [9, 26]. Other well-known algorithms are Gauss-Newton methods, inexact Newton methods, quasi-Newton methods, Levenberg-Marquardt (LM) methods which are very well studied in several books [25, 5, 10]. Such methods have local fast convergence properties under a nonsingularity assumption at the solution (which implies the locally uniqueness of the solution) and smoothness of the data.

In order to deal with the lack of smoothness and non-uniqueness of the solutions, several approaches have been considered in the literature. For the differentiability issues, we can mention the big breakthrough given by the development of semismooth Newton methods, see [29] and references therein. Observe that nondifferentiable system of equations arise naturally in the study of nonsmooth reformulations of the KKT system, complementarity systems [11], variational inequalities [17], etc.

Recently, there are several approaches to relax the nonsingularity assumption and still maintain a fast rate of convergence. A common way is to consider the *error bound condition* (EBC) near a solution $\bar{x}$, which state the existence of $\omega > 0$ such that $\operatorname{dist}(x, Z) \leq \omega \|F(x)\|$ holds for every x sufficiently close to $\bar{x}$. The EBC (or Lipschitzian error bound or metric subregularity) has been an important and useful property for the study of stability issues [27], complexity theory results [8], fast convergence rate, analysis of critical multipliers for nonlinear systems and optimization problems, e.g. [18, 13] and references therein. Indeed, it can be argued that the local quadratic convergence rate of the Newton's method was not due to the nonsingularity of the Jacobian but to the existence of an EBC near to the solution. Despite the advantages of the EBC, there are important problems where such condition fails, see for instance [1, 22] and references therein. A weaker condition than the EBC is the *Hölderian error bound condition*, that holds at $\bar{x} \in Z$, if exist $q \in (0, 1]$ and $\omega > 0$ such that $\operatorname{dist}(x, Z) \leq \omega \|F(x)\|^q$ holds for every x sufficiently close to $\bar{x}$. A simple example is $F(x) := x^6$ with $\bar{x} := 0$, in this case, $\bar{x} = 0$ is the unique solution, $q = 1/6$ and $\omega = 1$.

A very interesting method for solving (1) is the *Linear Programming* (LP)-Newton method, where at each iteration a linear optimization problem must be solved [12]. The LP-Newton method for constrained equations maintain several properties of the classical Newton's method, as the local superlinear convergence but it is able to deal with possibility of nonisolated solutions and nonsmooth mappings. Later, a globally convergent algorithm based on the LP-Newton subproblems and with a linesearch procedure, has recently been developed [14]. For other variants of the LP-Newton method, we cite [24, 23].

In this manuscript, we introduce and analyse a globally convergent iterative method for solving constrained nonlinear system of the form (1) where the mapping F is a piecewise smooth function, the set the solutions has not necessarily an isolated point and possibly the EBC may not hold at any solution. Similar to the LP-Newton, in each subproblem the new method solves a linear optimization problem, and fast convergence results can be obtained under Hölder-type conditions (Theorems 4.4 and 4.6). To obtain global convergence results, a trust-region-like procedure is proposed which guarantee convergence to stationary points in the sense of Definition 2.1. The new method neither use line-search techniques as [14, 15] nor force sequences as some globally inexact Newton's methods [7].

We organize the rest of this paper as follows: in Section 2, we survey some basic results and preliminary considerations. Section 3 presents our main algorithm with the global convergence analysis. The globalization is based on the trust-region procedure for the natural merit function, which is the norm of the mapping defining the system of equations. The transition from global to local convergence analysis is studied in Section 4 with the local convergence of the method under several assumptions including the Hölderian error bound condition of the mapping defining the system of equations. Our numerical results are in Section 5. Finally, in Section 6, we give some conclusions, remarks and observations.

2 Basic definitions and preliminary considerations

Our notation is standard in optimization and variational analysis [28]. The Euclidean n-dimensional space is denoted by $\mathbb{R}^n$. We use $\|\cdot\|$ to denote the norm of the space, which is not necessarily the same in all the spaces. We stress that different norms serve for different purpose, for example: $\|\cdot\| := \|\cdot\|_1$ in $\mathbb{R}^n$ is used to induce sparsity and $\|\cdot\| := \|\cdot\|_\infty$ in $\mathbb{R}^m$ for component-wise minimum error. We denote by $\mathbb{B}$ the closed unit ball and $\mathbb{B}(x,\varepsilon) := x + \varepsilon\mathbb{B}$ is the closed ball centered at x with radius $\varepsilon > 0$. The dual norm $\|\cdot\|_*$ is defined by $\|\nu\|_* := \max\limits_{x\in\mathbb{B}} \langle \nu, x\rangle$ where $\langle\cdot\,,\cdot\rangle$ is the Euclidean inner product. For instance, if $\|\cdot\|$ is the ℓ_1 norm, then $\|\cdot\|_*$ is the ℓ_∞ norm. For a given set X, we use $\mathrm{conv}(X)$, $\mathrm{cl}(X)$, $\mathrm{int}(X)$ to represent the convex hull, the closure and the interior of X, respectively. The distance from z to X is $\mathrm{dist}(z,X) := \inf\{\|z-x\| \mid x \in X\}$.

A sequence of nonnegative scalars $\{\alpha_k\}$ converges Q-superlinearly to zero with order $q > 1$ if there is $C > 0$ such that $\alpha_{k+1} \le C\alpha_k^q$ for k large enough. When $q = 2$, then $\{\alpha_k\}$ converges Q-quadratically to zero. Also, $\{\alpha_k\}$ converges R-superlinearly to zero with order q if it is bounded by a sequence converging Q-superlinearly to zero with order q.

For any function $f : \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ and any point $\bar{x}$ with $f(\bar{x}) < \infty$, $v \in \mathbb{R}^n$ is a *subgradient* of f at $\bar{x}$, denoted by $v \in \partial f(\bar{x})$, if there exists a sequence v^k

converging to v along with a sequence $x^k \to \bar{x}$ with $f(x^k) \to f(\bar{x})$ such that

$$f(x) \geq f(x^k) + \langle v^k, x - x^k \rangle + o(\|x - x^k\|).$$

Along this work we will assume that $F : \mathbb{R}^n \to \mathbb{R}^m$ is a *piecewise smooth mapping*, i.e., F is a continuous mapping and there exists a finite family of continuously differentiable mappings $F^1, \ldots, F^p : \mathbb{R}^n \to \mathbb{R}^m$, called *selection mappings*, such that

$$F(x) \in \{F^1(x), \ldots, F^p(x)\}.$$

By [11, Lemma 4.6.1], F is locally Lipschitz continuous. For this class of mappings (1) subsummes various important problem classes, as complementarity problems (which include KKT systems and generalized Nash equilibrium problems) and variational inequalities, see for instance [11, 19, 6]. The Jacobian of the selection mapping F^j at x will be denote by $G_j(x)$, and the index set of selection mappings active at $x \in \mathbb{R}^n$ is

$$\mathcal{A}(x) := \{j \in \{1, \ldots, p\} \mid F^j(x) = F(x)\}.$$

Now, we proceed with a new and local method for solving (1) that we call LP(η, θ)-Newton method. At each iteration, it solves a minimization problem which is a modification of the subproblem of the LP-Newton method proposed in [12] but, to deal with the possibly lack of derivative of F and the failure of the EBC, we add some parameters η and θ. On the other hand, the parameter ϱ in Algorithm 1 is related with the radius of a trust-region procedure, described in Algorithm 2, used as globalization strategy. The LP-Newton method of [12] is a particular of LP(η, θ)-Newton method by taking $\varrho = 1$, $\eta = 2$, $\theta = 1$ and $\|\cdot\| = \|\cdot\|_\infty$.

Algorithm 1 (Pure) LP(η, θ)-Newton method.

Take $\varrho, \eta, \theta > 0$. Initialize with $k = 0$ and $x^0 \in \Omega$.
1. If $F(x^k) = 0$, stop.
2. Choose $\mathcal{A}_k \subseteq \mathcal{A}(x^k)$ and compute (d^k, γ_k) as a solution of the subproblem

$$\begin{array}{ll} \underset{(d,\gamma)}{\text{minimize}} & \gamma \\ \text{subject to} & \|F(x^k) + G_j(x^k)d\| \leq \gamma \|F(x^k)\|^\eta, \quad \forall j \in \mathcal{A}_k, \\ & \|d\| \leq \gamma \varrho \|F(x^k)\|^\theta, \\ & x^k + d \in \Omega. \end{array} \tag{2}$$

3. Set $x^{k+1} := x^k + d^k$ and $k = k + 1$. Go to Step 1.

Observe that LP(η, θ)-Newton method is always well defined since the constraints are always feasible. When the involved norms are the ℓ_1, ℓ_∞ or more generally if $\|\cdot\|$ is the Minkowski gauge function of a symmetric polyhedral set with nonempty interior (see [28, Exercise 11.20]), the sub-problems (2) are linear optimization problems. To obtain superlinear convergence near to a solution $\bar{x}$ satisfying the Hölderian error bound condition, the parameters (η, θ) must be in some region described in Theorem 4.4 and Theorem 4.6 which depends on the Hölder exponent $q \in (0, 1]$.

To globalize the method, we propose Algorithm 2 which is described in details in Section 3. At each iteration k, Algorithm 2 looks for a sufficient decrease of the merit function $\|F(x)\|$. Such decreasing condition is the following:

$$\|F(x^k + d^k)\| \leq \|F(x^k)\| - \sigma \Delta_k, \tag{3}$$

where $\sigma \in (0, 1)$ and $\Delta_k := \|F(x^k)\| - \gamma_k \|F(x^k)\|^\eta$. The number Δ_k is a nonnegative scalar measuring the predicted reduction (Corollary 3.3). The globalization mechanism of Algorithm 2 is actually a trust-region method. This assertion will become clear in Section 3. Note that expression (3) is equivalent to $\sigma \leq (\|F(x^k)\| - \|F(x^k + d^k)\|)/\Delta_k$, which can be interpreted as an acceptable ratio between the achieved reduction and the predicted reduction (see (17)). It is worthy to note, that the transition from global to fast local convergence occurs naturally under weak assumptions. Finally, we observe that for our global convergence analysis, we only require that the parameters η, θ must be positive.

To study the global convergence of Algorithm 2, we need to analyse the natural way of globalizing the constrained nonlinear system of equations (1) which consists of solving (4), and to consider the first-order optimality condition associated to it.

$$\begin{array}{ll} \text{minimize} & \|F(x)\| \\ \text{subject to} & x \in \Omega. \end{array} \tag{4}$$

Clearly, the merit function in (4) is locally Lipschitz continuous since $f(x) := g(F(x))$ with $g(y) = \|y\|$. Thus, by [28, Theorem 8.15], the associated first-order necessary optimality condition at $\bar{x} \in \Omega$ is

$$0 \in \partial f(\bar{x}) + N_\Omega(\bar{x}).$$

Furthermore, from [28, Theorem 10.49], we can upper estimate $\partial f(\bar{x})$ by

$$\partial f(x) \subset \bigcup \{\partial(\nu F)(x) \mid \nu \in \partial g(F(x))\}, \tag{5}$$

where $(\nu F)(x) := \langle \nu, F(x) \rangle$ and $\partial g(y) = \{\nu \in \mathbb{R}^m \mid \|\nu\|_* \leq 1, \langle \nu, y \rangle = \|y\|\}$, see [28, Corollary 8.25]. We stress that for a nonsmooth mapping F the inclusion in (5) may be strict. Thus, for our analysis, we will consider an optimization problem where stationary points involve the right-hand side set of (5). It can be seen that, $\bar{x}$ is a local solution of (4) if and only if $(\bar{x}, \|F(\bar{x})\|)$ is a local solution of

$$\begin{array}{ll} \underset{(x,\gamma)}{\text{minimize}} & \gamma \\ \text{subject to} & \|F(x)\| \leq \gamma, \\ & x \in \Omega. \end{array} \tag{6}$$

First, consider the next lemma whose proof follows straightforward arguments.

Lemma 2.1. *If $K := \{(y,\tau) \in \mathbb{R}^m \times \mathbb{R} \mid \|y\| \leq \tau\}$. Then, K is a closed convex cone with (negative) polar set $K^* = \{(\vartheta, t) \in \mathbb{R}^m \times \mathbb{R} \mid \|\vartheta\|_* \leq -t\}$.*

Thus, necessary optimality conditions of (6) can be stated as follows.

Proposition 2.2. *If $(\bar{x}, \bar{\gamma})$ is a local solution of problem (6) then $\bar{\gamma} = \|F(\bar{x})\|$ and there exists $\nu \in \mathbb{R}^m$ such that*

$$0 \in \partial(\nu F)(\bar{x}) + N_\Omega(\bar{x}) \quad \textit{with} \quad \|\nu\|_* \leq 1, \ \langle \nu, F(\bar{x})\rangle = \|F(\bar{x})\|.$$

Proof. Clearly, (6) is equivalent to

$$\text{minimize } \pi_2(x,\gamma) + \delta_K(\widehat{F}(x,\gamma)) \text{ over } (x,\gamma) \in \Omega \times \mathbb{R},$$

where $\pi_2(x,\gamma) = \gamma$, $\widehat{F}(x,\gamma) = (F(x),\gamma)$, $K = \{(y,\tau) \in \mathbb{R}^m \times \mathbb{R} \mid \|y\| \leq \tau\}$ and δ_K is the indicator function of K, i.e., $\delta_K(x) = 0$ if $x \in K$ and $\delta_K(x) = \infty$ otherwise.

To obtain the result, we will show that the following constraint qualification holds:

$$0 \in \partial(\zeta\widehat{F})(\bar{x},\bar{\gamma}) + N_{\Omega\times\mathbb{R}}(\bar{x},\bar{\gamma}), \quad \zeta \in N_K(\widehat{F}(\bar{x},\bar{\gamma})) \ \Rightarrow \ \zeta = 0. \tag{7}$$

Indeed, take $\zeta := (\nu, s) \in \mathbb{R}^m \times \mathbb{R}$, it can be seen that $\partial(\zeta\widehat{F})(\bar{x},\bar{\gamma}) = \{(v,s) \mid v \in \partial(\nu F)(\bar{x})\}$. From $N_{\Omega\times\mathbb{R}}(\bar{x},\bar{\gamma}) = N_\Omega(\bar{x}) \times \{0\}$ and the first inclusion in (7), we get $s = 0$. Since K is a cone, $\zeta \in N_K(\widehat{F}(\bar{x},\bar{\gamma}))$ if and only if $\langle \zeta, \widehat{F}(\bar{x},\bar{\gamma})\rangle = 0$ with $\zeta \in K^*$. Since $s = 0$ and $\zeta = (\nu, s) \in K^*$, lemma 2.1 implies that $\nu = 0$ and hence, $\zeta = 0$.

Now, the fulfilment of the constraint qualification (7) and [28, Exercise 10.52] guarantee the existence of a vector $\zeta = (\nu, s)$ such that

$$0 \in \partial(\pi_2 + \zeta\widehat{F})(\bar{x},\bar{\gamma}) + N_{\Omega\times\mathbb{R}}(\bar{x},\bar{\gamma}) \quad \text{and} \quad \zeta \in N_K(\widehat{F}(\bar{x},\bar{\gamma})). \tag{8}$$

From the first inclusion of (8) and by $\pi_2'(x,\gamma) = (0,1)$, we see that

$$0 \in \partial(\nu F)(\bar{x}) + N_\Omega(\bar{x}), \quad \text{and} \quad s = -1.$$

The second inclusion of (8) implies $\|\nu\|_* \leq 1$ and $0 = \langle \zeta, \widehat{F}(\bar{x},\bar{\gamma})\rangle = \langle \nu, F(\bar{x})\rangle - \bar{\gamma} \leq \|F(\bar{x})\| - \bar{\gamma}$. By the feasibility of $(\bar{x},\bar{\gamma})$, we get $\bar{\gamma} = \|F(\bar{x})\|$. □

Clearly, in the previous result, if the local solution $(\bar{x},\bar{\gamma})$ satisfies $F(\bar{x}) \neq 0$, then $\|\nu\|_* = 1$, since $\|\nu\|_* < 1$ implies $\langle \nu, F(\bar{x})\rangle < \|F(\bar{x})\|$.

In this work, instead of $\partial(\nu F)(\bar{x})$ we will use its convex hull, i.e., $\operatorname{conv}\big(\partial(\nu F)(\bar{x})\big)$. Since νF is locally Lipschitz continuous, by [28, Theorem 8.49], $\operatorname{conv}\big(\partial(\nu F)(\bar{x})\big)$ is the set of Clarke subgradients of νF at $\bar{x}$. Moreover, by [28, Theorem 9.61]

$$\operatorname{conv}\big(\partial(\nu F)(\bar{x})\big) = \operatorname{conv}\big\{G_j(\bar{x})^T\nu \mid j \in \tilde{\mathcal{A}}(\bar{x})\big\},$$

where $\tilde{\mathcal{A}}(x) := \{j \in \{1,\ldots,p\} \mid x \in \operatorname{cl}(\operatorname{int}(D_j))\}$ with $D_j = \{x \in \mathbb{R}^n \mid j \in \mathcal{A}(x)\}$. Note that when no spurious selection mappings are active in those points where F is nonsmooth, formally, when $D_j = \operatorname{cl}(\operatorname{int}(D_j))$ for all j, we have that $\tilde{\mathcal{A}}(x) = \mathcal{A}(x)$ for all x. Thus, to be consistent, we introduce the following.

Definition 2.1. We will say that $(\bar{x}, \bar{\gamma})$ is a Clarke stationary point of problem (6) if $\bar{\gamma} = \|F(\bar{x})\|$ and

$$0 \in \operatorname{conv}\big\{G_j(\bar{x})^T\nu \mid j \in \mathcal{A}(\bar{x}),\ \|\nu\|_* \leq 1,\ \langle \nu, F(\bar{x})\rangle = \|F(\bar{x})\|\big\} + N_\Omega(\bar{x}).$$

When it is clear in the context, we refer $\bar{x}$ as a Clarke stationary point.

With Definition 2.1 in mind, Proposition 2.2 says that if $\bar{x}$ is a local solution of minimize $\|F(x)\|$ over $x \in \Omega$, then $(\bar{x}, \|F(\bar{x})\|)$ is a Clarke stationary point.

3 Global convergence

To globalize the LP(η, θ)-Newton method, we present Algorithm 2. Different to other globalized methods for solving (1) based on line-search procedures, we use a trust-region approach. Here, the parameter ϱ_k is related with trust region radius and Δ_k to the acceptance of the trial point.

Algorithm 2 Globalized convergent algorithm

Take $\sigma \in (0,1)$, $\beta \in (0,1)$, $\eta > 0$ and $\theta > 0$. Initialize with $k = 0$, $\varrho_0 > 0$ and $x^0 \in \Omega$.

1. If $F(x^k) = 0$, then stop. Otherwise, choose $\mathcal{A}_k \subseteq \mathcal{A}(x^k)$.

2. Compute (d^k, γ_k) as a solution of the subproblem

$$\begin{array}{ll} \underset{(d,\gamma)}{\text{minimize}} & \gamma \\ \text{subject to} & \|F(x^k) + G_j(x^k)d\| \leq \gamma \|F(x^k)\|^\eta, \quad \forall j \in \mathcal{A}_k, \\ & \|d\| \leq \gamma \varrho_k \|F(x^k)\|^\theta, \\ & x^k + d \in \Omega. \end{array}$$

3. Compute the nonnegative number Δ_k where

$$\Delta_k := \|F(x^k)\| - \gamma_k \|F(x^k)\|^\eta. \tag{9}$$

   If $\Delta_k = 0$, then stop.

4. If $\|F(x^k + d^k)\| \leq \|F(x^k)\| - \sigma\Delta_k$ go to Step 5. Otherwise, set $\varrho_k = \beta\varrho_k$, $\mathcal{A}_k = \mathcal{A}_k \cup \{j\}$ for (possible empty) $j \in \mathcal{A}(x^k + d^k) \cap \mathcal{A}(x^k)$, and go to Step 2.

5. Set $x^{k+1} = x^k + d^k$, $\varrho_{k+1} = \varrho_0$ and $k = k + 1$. Update η and θ, if were necessary. Go to Step 1.

Some brief comments about Algorithm 2: (i) In Step 2, $\mathcal{A}_k$ is expected to be smaller than $\mathcal{A}(x^k)$, since in some applications the cardinality of the later may be large enough, which may increase the computational cost of solving the subproblems; (ii) as we said, the number Δ_k in Step 3 serves as a measure of optimality and hence as a natural stopping criterion. Indeed, $\Delta_k = 0$ if and only if x^k is Clarke stationary (see Corollary 3.3 for the proper statement); and (iii) in Step 4, we only update ϱ_k and $\mathcal{A}_k$ if the sufficient decrease condition (3) fails. The algorithm is well-defined, since (3) occurs after a finite number of iterations as we will state in Proposition 3.4. Note that we propose a modest rule for updating $\mathcal{A}_k$ since we want to avoid large $\mathcal{A}_k$.

We continue with the global convergence analysis. The next proposition provides necessary and sufficient conditions for optimality of minimization problems having the structure of the subproblems that appears in LP(η,θ)-Newton method.

Proposition 3.1. *Let $x \in \Omega$ be a feasible point and let $\zeta, \vartheta, \varrho > 0$. Consider $\mathcal{A}$ a non-empty finite set of indices. Take a non-zero vector $F \in \mathbb{R}^n$ and matrices $G_j \in \mathbb{R}^{m\times n}$ for every $j \in \mathcal{A}$.*

Then, (d^*, γ_*) is a solution of the minimization problem

$$\begin{array}{ll} \underset{(d,\gamma)}{\textit{minimize}} & \gamma \\ \textit{subject to} & \|F + G_j d\| \le \gamma\zeta, \ \forall\, j \in \mathcal{A}, \\ & \|d\| \le \varrho\gamma\vartheta, \\ & x + d \in \Omega. \end{array} \tag{10}$$

if and only if there exist $\{\nu^j\}_{j\in\mathcal{A}} \subset \mathbb{R}^m$ and $\mu \in \mathbb{R}^n$ such that

$$0 \in \tfrac{1}{\zeta}\sum_{j\in\mathcal{A}} G_j^T \nu^j + \tfrac{1}{\vartheta}\mu + N_\Omega(x + d^*), \tag{11}$$

$$\sum_{j\in\mathcal{A}} \|\nu^j\|_* + \varrho\|\mu\|_* = 1, \tag{12}$$

$$\gamma_* = \max\left\{ \tfrac{1}{\zeta}\max_{j\in\mathcal{A}} \|F + G_j d^*\|, \tfrac{1}{\varrho\vartheta}\|d^*\| \right\} = \tfrac{1}{\zeta}\sum_{j\in\mathcal{A}} \langle \nu^j, F + G_j d^*\rangle + \tfrac{1}{\vartheta}\langle \mu, d^*\rangle. \tag{13}$$

Furthermore, the next relations hold

$$\gamma_* = \frac{1}{\zeta}\|F + G_r d^*\|, \ \textit{for some } r \in \mathcal{A}; \tag{14a}$$

$$\|\nu^j\|_* \left(\tfrac{1}{\zeta}\|F + G_j d^*\| - \gamma_*\right) = 0, \ \textit{for all } j \in \mathcal{A}; \quad \|\mu\|_* \left(\tfrac{1}{\varrho\vartheta}\|d^*\| - \gamma_*\right) = 0; \tag{14b}$$

$$\langle \nu^j, F + G_j d^*\rangle = \|\nu^j\|_* \|F + G_j d^*\|, \ \textit{for all } j \in \mathcal{A}; \quad \langle \mu, d^*\rangle = \|\mu\|_* \|d^*\|. \tag{14c}$$

Proof. Assuming $\mathcal{A} := \{j_1, \ldots, j_\ell\}$ for some $\ell \in \mathbb{N}$, (10) is equivalent to

$$\begin{array}{ll} \underset{(d,\gamma)}{\text{minimize}} & \gamma \\ \text{subject to} & \|\tilde{F}(d)\| \le \gamma, \quad d \in \tilde{\Omega}, \end{array}$$

where $\tilde{\Omega} = \{d \in \mathbb{R}^n \mid x+d \in \Omega\}$, $\tilde{F}(d) = \left(\frac{1}{\zeta}(F + G_{j_1}d), \ldots, \frac{1}{\zeta}(F + G_{j_\ell}d), \frac{1}{\vartheta}d\right)$ and the norm defined by $\|z\| = \max\left\{\max_{j\in\mathcal{A}} \|y^j\|, \varrho^{-1}\|x\|\right\}$ for any $z = (y^{j_1}, \ldots, y^{j_\ell}, x)$.

Since all the function are convex, necessary optimality conditions are sufficient. By Proposition 2.2, $\gamma_* = \|\tilde{F}(d^*)\|$ and there exists $\lambda = (\nu^{j_1}, \ldots, \nu^{j_\ell}, \mu)$ such that

$$0 \in \tilde{F}'(d^*)^T\lambda + N_{\tilde{\Omega}}(d^*), \tag{15}$$

$$\|\lambda\|_* \leq 1, \quad \langle\lambda, \tilde{F}(d^*)\rangle = \|\tilde{F}(d^*)\|. \tag{16}$$

From the derivative of the affine function $\tilde{F}$ and since $N_{\tilde{\Omega}}(d^*) = N_\Omega(x+d^*)$, we obtain (11) from (15). Furthermore, $\|\tilde{F}(d^*)\| > 0$, otherwise, $F + G_jd^* = 0$ for all $j \in \mathcal{A}$ and $d^* = 0$, implying $F = 0$, which is a contradiction. Thus, $\|\lambda\|_* = 1$ from (16).

On the other hand, it is not hard to see that for the norm in $\mathbb{R}^{m|\mathcal{A}|} \times \mathbb{R}^n$ we have

$$\|\lambda\|_* = \sum_{j\in\mathcal{A}} \|\nu^j\|_* + \varrho\|\mu\|_*.$$

Hence, we get (12). Now, (13) follows from $\gamma_* = \|\tilde{F}(d^*)\| = \langle\lambda, \tilde{F}(d^*)\rangle$.

For the last assertions, since

$$\begin{aligned}
0 &= \langle\lambda, \tilde{F}(d^*)\rangle - \gamma_* \\
&\leq \tfrac{1}{\zeta}\sum_{j\in\mathcal{A}} \|\nu^j\|_*\|F + G_jd^*\| + \tfrac{1}{\vartheta}\|\mu\|_*\|d^*\| - \left(\sum_{j\in\mathcal{A}} \|\nu^j\|_* + \varrho\|\mu\|_*\right)\gamma_* \\
&= \sum_{j\in\mathcal{A}} \|\nu^j\|_*\left(\tfrac{1}{\zeta}\|F + G_jd^*\| - \gamma_*\right) + \varrho\|\mu\|_*\left(\tfrac{1}{\varrho\vartheta}\|d^*\| - \gamma_*\right) \leq 0,
\end{aligned}$$

each term in the above summation should be zero. Thus (14b) holds.

In order to show (14a), i.e., $\frac{1}{\zeta}\|F+G_rd^*\| = \gamma_*$ for some $r \in \mathcal{A}$, suppose that $\frac{1}{\zeta}\|F + G_jd^*\| < \frac{1}{\varrho\vartheta}\|d^*\|$ for all $j \in \mathcal{A}$. Then, from (13) and (14b) we get that $\gamma_* = \frac{1}{\varrho\vartheta}\|d^*\|$ and $\|\nu^j\|_* = 0$ for all $j \in \mathcal{A}$. Thus, from (13) and (11),

$$0 < \gamma_* = \frac{1}{\vartheta}\langle\mu, d^*\rangle = \frac{1}{\vartheta}\langle-\mu, x - (x + d^*)\rangle \leq 0,$$

where in the last inequality we use $x \in \Omega$ and $-\mu \in N_\Omega(x+d^*)$. This contradiction guarantee $\|F + G_rd^*\|/\zeta \geq \|d^*\|/(\varrho\vartheta)$ for some $r \in \mathcal{A}$. Now, using (14b), we obtain

$$
\begin{aligned}
0 &= \langle \lambda, \tilde{F}(d^*) \rangle - \gamma_* \\
&= \tfrac{1}{\zeta} \sum_{j \in \mathcal{A}} \langle \nu^j, F + G_j d^* \rangle + \tfrac{1}{\vartheta} \langle \mu, d^* \rangle - \left(\sum_{j \in \mathcal{A}} \|\nu^j\|_* + \varrho \|\mu\|_* \right) \gamma_* \\
&= \tfrac{1}{\zeta} \sum_{j \in \mathcal{A}} \big(\langle \nu^j, F + G_j d^* \rangle - \|\nu^j\|_* \|F + G_j d^*\| \big) + \tfrac{1}{\vartheta} \left(\langle \mu, d^* \rangle - \|\mu\|_* \|d^*\| \right) \le 0.
\end{aligned}
$$

Thus, each term in the summation should be zero and (14c) holds. $\square$

We stress that applying Proposition 3.1 to subproblems in Algorithm 2, the number Δ_k can be written as

$$
\Delta_k = \|F(x^k)\| - \gamma_k \|F(x^k)\|^\eta = \|F(x^k)\| - \max_{j \in \mathcal{A}_k} \|F(x^k) + G_j(x^k) d^k\|. \tag{17}
$$

From the last equality, we may understood Δ_k as a measure of the predicted reduction of the model. The proper statement will be given in Corollary 3.3. From Proposition 3.1, if $\Delta = \|F\| - \gamma_* \zeta$ we have the following relations:

$$
\begin{aligned}
0 &\ge \left\langle -\tfrac{1}{\zeta} \textstyle\sum_{j \in \mathcal{A}} G_j^T \nu^j - \tfrac{1}{\vartheta} \mu, x - (x + d^*) \right\rangle \\
&= \left\langle \tfrac{1}{\zeta} \textstyle\sum_{j \in \mathcal{A}} G_j^T \nu^j + \tfrac{1}{\vartheta} \mu, d^* \right\rangle \qquad (18) \\
&= \gamma_* - \left\langle \tfrac{1}{\zeta} \textstyle\sum_{j \in \mathcal{A}} \nu^j, F \right\rangle \ge \tfrac{1}{\zeta} \big(\zeta \gamma_* - (1 - \varrho \|\mu\|_*) \|F\| \big) \\
&= \tfrac{1}{\zeta} \left(\varrho \|\mu\|_* \|F\| - \Delta \right), \qquad (19)
\end{aligned}
$$

where we use (11) for the first relation, (13) for the third relation and (12) for the fourth relation. Also, by (14b) and (14c), we have

$$
\begin{aligned}
-\|\nu^j\|_* \Delta &= \|\nu^j\|_* (\zeta \gamma_* - \|F\|) = \|\nu^j\|_* \|F + G_j d^*\| - \|\nu^j\|_* \|F\| \\
&= \langle \nu^j, F + G_j d^* \rangle - \|\nu^j\|_* \|F\| \le \langle \nu^j, G_j d^* \rangle.
\end{aligned} \tag{20}
$$

In order to simplify further results, we state the following auxiliary proposition.

Proposition 3.2. *Let d^*, $\{\nu^j\}_{j \in \mathcal{A}}$ and μ satisfy (11) and (12). Suppose that $\langle \nu^j, F \rangle = \|\nu^j\|_* \|F\|$, $\langle G_j^T \nu^j, d^* \rangle \ge 0$ for all $j \in \mathcal{A}$, and $\mu = 0$. Then*

$$
0 \in \operatorname{conv} \left\{ G_j^T \nu \mid j \in \mathcal{A},\ \|\nu\|_* \le 1,\ \langle \nu, F \rangle = \|F\| \right\} + N_\Omega(x).
$$

Proof. Since $\langle G_j^T \nu^j, d^* \rangle \ge 0$, $\forall j \in \mathcal{A}$ and $\mu = 0$, from (11) and for any $y \in \Omega$,

$$
0 \ge \left\langle -\tfrac{1}{\zeta} \textstyle\sum_{j \in \mathcal{A}} G_j^T \nu^j - \tfrac{1}{\vartheta} \mu, y - (x + d^*) \right\rangle \ge \tfrac{1}{\zeta} \left\langle - \textstyle\sum_{j \in \mathcal{A}} G_j^T \nu^j, y - x \right\rangle.
$$

Thus, setting $\mathcal{A}^* = \{j \in \mathcal{A} \mid \nu^j \neq 0\}$, we obtain

$$0 \in \sum_{j\in\mathcal{A}} G_j^T \nu^j + N_\Omega(x) = \sum_{j\in\mathcal{A}^*} \|\nu^j\|_* G_j^T \frac{\nu^j}{\|\nu^j\|_*} + N_\Omega(x),$$

where $\langle \nu^j/\|\nu^j\|_*, F\rangle = \|F\|$ for every $j \in \mathcal{A}^*$ and, by (12), $\sum_{j\in\mathcal{A}^*} \|\nu^j\|_* = 1$. $\square$

Clearly, the stopping criterion in Step 1 of Algorithm 2 corresponds to solutions of problem (6). Now, we can state that the stopping criterion in Step 3 of Algorithm 2 corresponds to Clarke stationary points of problem (6).

Corollary 3.3. *Under the hypotheses of Proposition 3.1, let (d^*, γ_*) be a solution of problem (10). Then, the scalar $\Delta(\zeta, \vartheta, \varrho) := \|F\| - \gamma_*\zeta$ is always nonnegative, and $\Delta(\zeta, \vartheta, \varrho) = 0$ if and only if*

$$0 \in \operatorname{conv}\left\{G_j^T \nu \mid j \in \mathcal{A},\ \|\nu\|_* \leq 1,\ \langle \nu, F\rangle = \|F\|\right\} + N_\Omega(x). \tag{21}$$

When, $(\zeta, \vartheta, \varrho)$ is clear to the context, we write Δ instead of $\Delta(\zeta, \vartheta, \varrho)$.

Proof. First, we will see that Δ is non-negative. Indeed, since $(0, \frac{1}{\zeta}\|F\|)$ is feasible for problem (10), then $\gamma_* \leq \frac{1}{\zeta}\|F\|$. Clearly, the last inequality implies that $\Delta \geq 0$.

Now, we will prove the equivalence. If $\Delta = 0$, then $\mu = 0$ by (19) and $\langle G_j^T \nu^j, d^*\rangle \geq 0$ by (20). Combining this with (18) we obtain $\langle G_j^T \nu^j, d^*\rangle = 0$ for all $j \in \mathcal{A}$. Hence, by (14b), (14c) and by $0 = \|F\| - \gamma_*\zeta$, for all $j \in \mathcal{A}$, we have

$$\|\nu^j\|_*\|F\| = \zeta\|\nu^j\|_*\gamma_* = \|\nu^j\|_*\|F + G_j d^*\| = \langle \nu^j, F + G_j d^*\rangle = \langle \nu^j, F\rangle.$$

Thus, from Proposition 3.2, (21) holds.

For the converse, suppose that there exist ν^i, α_{ij} with $i = 1, \ldots, k$ and $j \in \mathcal{A}$ satisfying $\|\nu^i\|_* \leq 1$, $\langle \nu^i, F\rangle = \|F\|$, $\alpha_{ij} \in [0, 1]$, $\sum_{i=1}^k \sum_{j\in\mathcal{A}} \alpha_{ij} = 1$ and

$$0 \in \sum_{i=1}^k \sum_{j\in\mathcal{A}} \alpha_{ij} G_j^T \nu^i + N_\Omega(x).$$

Now, since $x + d^* \in \Omega$, we see that $\left\langle -\sum_{i=1}^k \sum_{j\in\mathcal{A}} \alpha_{ij} G_j^T \nu^i, d^*\right\rangle \leq 0$. Then, using this inequality, we get

$$\begin{aligned}\|F\| &= \textstyle\sum_{i=1}^k \sum_{j\in\mathcal{A}} \alpha_{ij}\langle \nu^i, F\rangle \leq \sum_{i=1}^k \sum_{j\in\mathcal{A}} \alpha_{ij}\langle \nu^i, F + G_j d^*\rangle \\ &\leq \textstyle\sum_{i=1}^k \sum_{j\in\mathcal{A}} \alpha_{ij}\|\nu^i\|_*\|F + G_j d^*\| \leq \sum_{i=1}^k \sum_{j\in\mathcal{A}} \alpha_{ij}\zeta\gamma_* = \zeta\gamma_*,\end{aligned}$$

which implies $\Delta \leq 0$ and thus $\Delta = 0$. $\square$

We should stress that the previous result shows that if we solve (10) for some positive parameters $\zeta, \vartheta, \varrho$ and $\Delta(\zeta, \vartheta, \varrho) = 0$ then, $\Delta(\zeta, \vartheta, \varrho) = 0$ for any positive parameters $\zeta, \vartheta, \varrho$.

Let us show that the reduction in Step 4 is performed finitely many times, and thus, Algorithm 2 is well-defined.

Proposition 3.4. *Suppose that $x \in \Omega$, $F(x) \neq 0$, $\zeta, \vartheta, \varrho_0 > 0$, $\mathcal{A}_0 \subseteq \mathcal{A}(x)$ and $\sigma, \beta \in (0,1)$. Let $\Delta_t = \|F(x)\| - \gamma_t \zeta$, where (d^t, γ_t) is a solution of*

$$\begin{array}{ll} \underset{(d,\gamma)}{\textit{minimize}} & \gamma \\ \textit{subject to} & \|F(x) + G_j(x)d\| \leq \gamma\zeta, \ \forall\, j \in \mathcal{A}_t, \\ & \|d\| \leq \varrho_t \gamma \vartheta, \\ & x + d \in \Omega. \end{array} \tag{22}$$

Consider the recursive rule $\varrho_{t+1} = \beta \varrho_t$ and $\mathcal{A}_{t+1} = \mathcal{A}_t \cup \{j\}$ for (possible empty) $j \in \mathcal{A}(x + d^t) \cap \mathcal{A}(x)$. Then, there exists a iteration $t_* \geq 0$ such that either $\Delta_{t_*} = 0$ or $\|F(x + d^{t_*})\| \leq \|F(x)\| - \sigma \Delta_{t_*}$.

Proof. By contradiction, consider that for all t we have $\Delta_t > 0$ and $\|F(x+d^t)\| > \|F(x)\| - \sigma\Delta_t$. Let us denote by $\{\nu^{tj}\}$ and μ^t the Lagrange multipliers given by Proposition 3.1. Since $\Delta_t > 0$, by (13) we have $\|F(x)\|/\zeta > \gamma_t \geq \|d^t\|/(\varrho_t\vartheta)$. Thus, $\{d^t/\varrho_t\}$ is bounded, $\varrho_t \to 0$ and $d^t \to \bar{d} = 0$. Also, by (13),

$$\Delta_t = \|F(x)\| - \zeta\gamma_t \leq \|F(x)\| - \|F(x) + G_j(x)d^t\| \text{ for every } j \in \mathcal{A}_t. \tag{23}$$

On the other hand, by (19), we obtain

$$\Delta_t \geq \varrho_t \|\mu^t\|_* \|F(x)\|. \tag{24}$$

As $d^t \to \bar{d} = 0$, we see that $\mathcal{A}(x + d^t) \subset \mathcal{A}(x)$ for t large enough. Now, by the updating's rule of $\mathcal{A}_t$, we see that $\mathcal{A}_t = \mathcal{A}$ and $\mathcal{A}(x + d^t) \subseteq \mathcal{A}$ for all t large enough. Then, by (23) and (24), for $j \in \mathcal{A}(x + d^t)$

$$\begin{aligned} \|F^j(x + d^t)\| - \|F^j(x) + G_j(x)d^t\| &= \|F(x + d^t)\| - \|F(x) + G_j(x)d^t\| \\ &> \|F(x)\| - \|F(x) + G_j(x)d^t\| - \sigma\Delta_t \\ &\geq (1-\sigma)\Delta_t \geq (1-\sigma)\varrho_t\|\mu^t\|_*\|F(x)\|. \end{aligned}$$

By continuity of the derivative of F^j, we have $\|F^j(x+d^t)\| - \|F^j(x) + G_j(x)d^t\| = o(\|d^t\|)$. Dividing the above inequality by ϱ_t, using that $\{d^t/\varrho_t\}$ is bounded and taking limits, we obtain $\mu^t \to \bar{\mu} = 0$. Now, taking subsequence if necessary, we have $\nu^{tj} \to \bar{\nu}^j$ for $j \in \mathcal{A}$. Thus (11) holds replacing d^*, μ and ν^j by $\bar{d}$, $\bar{\mu}$ and $\bar{\nu}^j$, respectively. Also, since $\bar{d} = 0$, from (14c) we have $\|\bar{\nu}^j\|_* \|F(x)\| = \langle \bar{\nu}^j, F(x) \rangle$. Thus, applying Proposition 3.2 and Corollary 3.3, we should have $\Delta_t = 0$ for all $t \geq 0$. $\square$

From the previous results we can show that the proposed algorithm is globally convergent under the following assumption.

Assumption 1. *$\|F(x)\| \le \|F^j(x)\|$ for all $j \in \{1,\dots,p\}$ and all $x \in \Omega$.*

This assumption was used in [14] and, as noticed in the mentioned work, it holds for a suitable reformulation of the mixed complementarity problem

$$a(z) = 0, \quad b(z) \ge 0, \quad c(z) \ge 0, \quad d(z) \ge 0, \quad \langle c(z), d(z)\rangle = 0,$$

where $a : \mathbb{R}^n \mapsto \mathbb{R}^l$, $b : \mathbb{R}^n \mapsto \mathbb{R}^s$, $c : \mathbb{R}^n \mapsto \mathbb{R}^m$ and $d : \mathbb{R}^n \mapsto \mathbb{R}^m$. The associated problem (1) is given by $\Omega = \mathbb{R}^n \times \mathbb{R}^s_+ \times \mathbb{R}^m_+ \times \mathbb{R}^m_+$ and

$$F(z,u,v,w) = \big(a(z),\, b(z) - u,\, c(z) - v,\, d(z) - w,\, \min\{v,w\}\big). \tag{25}$$

For this problem, Assumption 1 holds for any ℓ_r norm.

We now state the main result of this section.

Theorem 3.5. *If F satisfies Assumption 1, then for every $x^0 \in \Omega$, Algorithm 2 either*

(a) stops in some iterate x^ℓ such that $(x^\ell, \|F(x^\ell)\|)$ is a Clarke stationary point of problem (6), or

(b) it generates an infinite sequence $\{x^k\}$ such that every limit point $\bar{x}$ satisfies that $(\bar{x}, \|F(\bar{x})\|)$ is a Clarke stationary point of problem (6).

Proof. Let $\{x^k\}$ be the sequence of iterates given by the algorithm.

Part (a). Suppose that the sequence $\{x^k\}$ is finite. Let x^ℓ be the last element, then either $F(x^\ell) = 0$ or $\Delta_\ell = 0$. In the first case $(x^\ell, 0)$ is a solution of problem (6) and hence a Clarke stationary point by Proposition 2.2. In the second case $(x^\ell, \|F(x^\ell)\|)$ is a Clarke stationary point by Corollary 3.3.

Part (b). Assume that $\{x^k\}$ is a infinite sequence. Then, for all k, $F(x^k) \ne 0$, $\Delta_k > 0$ and $\|F(x^{k+1})\| \le \|F(x^k)\| - \sigma\Delta_k$. Let $\bar{x}$ be a limit point of $\{x^k\}$ and let $\mathcal{K}$ be the index set such that $\lim_{k\in\mathcal{K}} x^k = \bar{x}$. If $F(\bar{x}) = 0$, there is nothing to prove. Let $F(\bar{x}) \ne 0$. Thus $\{\|F(x^k)\|\}$ is a decreasing sequence such that $\lim_{k\to\infty} \|F(x^k)\| = \|F(\bar{x})\| > 0$ and $\lim_{k\to\infty} \Delta_k = 0$. Let $\zeta_k = \|F(x^k)\|^\eta$ and $\vartheta_k = \|F(x^k)\|^\theta$. Thus, (11)-(14c) hold replacing x, F, G_j, ζ, ϑ, ϱ, d^*, γ_*, $\mathcal{A}$, ν^j and μ by x^k, $F(x^k)$, $G_j(x^k)$, ζ_k, ϑ_k, ϱ_k, d^k, γ_k, $\mathcal{A}_k$, ν^{kj} and μ^k, respectively. From (19), we see $\Delta_k \ge \varrho_k\|\mu^k\|_*\|F(x^k)\|$ and taking limits, $\lim_{k\to\infty} \varrho_k\|\mu^k\|_* = 0$. Consider two cases depending if ϱ_k converges to zero or not.

(**b1**). If ϱ_k does not converges to zero. Take $\bar{\varrho} > 0$ such that $\varrho_k \ge \bar{\varrho} > 0$. In this case, $\mu^k \to \mu = 0$. Without loss of generality, assume that $\lim_{k\in\mathcal{K}} \nu^{kj} = \nu^j$, $\mathcal{A}_k = \mathcal{A}$ for all $k \in \mathcal{K}$ and $\lim_{k\in\mathcal{K}} d^k = d^*$. The result follows as in the "if part" proof of Corollary 3.3. Taking limits for $k \in \mathcal{K}$, from (20) and (18), we obtain $\langle G_j^T\nu^j, d^*\rangle = 0$ and hence $\langle \nu^{*j}, F(\bar{x})\rangle = \|\nu^{*j}\|_*\|F(\bar{x})\|$. Thus applying Proposition 3.2, we conclude that the point $(\bar{x}, \|F(\bar{x})\|)$ is a Clarke stationary point.

(**b2**). If $\varrho_k \to 0$, then, from Algorithm 2 and for $\tilde{\varrho}_k = \beta^{-1}\varrho_k$, we obtain that

$$\|F(x^k + \tilde{d}^k)\| > \|F(x^k)\| - \sigma\tilde{\Delta}_k \quad \text{with} \quad \tilde{\Delta}_k = \|F(x^k)\| - \zeta_k\tilde{\gamma}_k > 0,$$

where $(\tilde{d}^k, \tilde{\gamma}_k)$ is a solution of problem (2) for $\tilde{\varrho}_k$. Denote by $\tilde{\nu}^{kj}$ and $\tilde{\mu}^k$ the associated Lagrange multiplier for this problem. Following the algebraics in the proof of Proposition 3.4, $\{\tilde{d}^k/\tilde{\varrho}_k\}$ is bounded and for any $j \in \mathcal{A}_k$ we have

$$\begin{aligned}\|F(x^k + \tilde{d}^k)\| - \|F(x^k) + G_j(x^k)\tilde{d}^k\| &> \|F(x^k)\| - \|F(x^k) + G_j(x^k)\tilde{d}^k\| - \sigma\tilde{\Delta}_k \\ &\geq (1-\sigma)\tilde{\varrho}_k\|\tilde{\mu}^k\|_*\|F(x^k)\|.\end{aligned}$$

On the other hand, by Assumption 1, for any $j \in \mathcal{A}_k$

$$\|F(x^k+\tilde{d}^k)\| - \|F^j(x^k)+G_j(x^k)\tilde{d}^k\| \leq \|F^j(x^k+\tilde{d}^k)\| - \|F^j(x^k)+G_j(x^k)\tilde{d}^k\| = o(\|\tilde{d}^k\|)$$

where we use that F^j is continuously differentiable. Since $\|F(x^k)\| \geq \|F(\bar{x})\| > 0$ and $\{\tilde{d}^k/\tilde{\varrho}_k\}$ is bounded, we get $\lim_{k\to\infty} \tilde{\mu}^k = \bar{\mu} = 0$. Now, taking subsequence if necessary, assume $\mathcal{A}_k = \mathcal{A}$ for all sufficiently large $k \in \mathcal{K}$ and $\lim_{k\in\mathcal{K}} \tilde{\nu}^{kj} = \bar{\nu}^j$. Since $\lim_{k\in\mathcal{K}} \tilde{d}^k = d^* = 0$, from (14c) we obtain $\|\bar{\nu}^j\|_*\|F(\bar{x})\| = \langle \bar{\nu}^j, F(\bar{x})\rangle$ for all $j \in \mathcal{A}$. Thus, by Proposition 3.2, $(\bar{x}, \|F(\bar{x})\|)$ is a Clarke stationary point. □

As the reader has noted, our global convergence result depends on the behaviour of Δ_k independently of any positive parameters η and θ (see eq. (17)). However, a suitable choice can improve the rate of convergence as the next example shows.

Example 3.1. *Consider problem* (1) *with* $F(x) = x$ *and* $\Omega = \mathbb{R}$*. Taking* $x^0 > 0$*, Algorithm 2 generate sequences with*

$$d^k = -\frac{\varrho_k|x^k|^\theta}{\varrho_k|x^k|^\theta + |x^k|^\eta}x^k, \quad \gamma_k = \frac{|x^k + d^k|}{|x^k|^\eta}, \quad x^k + d^k = \frac{|x^k|^\eta}{\varrho_k|x^k|^\theta + |x^k|^\eta}x^k.$$

Hence, $\Delta_k = |x^k| - \gamma_k|x^k|^\eta = |x^k| - |x^k + d^k|$ *and the sufficient decrease condition in Step 4 is satisfied. Then,* $\varrho_k = \varrho_0$ *for all* k*. Obtaining* $x^k \to 0$ *and*

$$|x^{k+1}| = \frac{1}{\varrho_k + |x^k|^{\eta-\theta}}|x^k|^{1+\eta-\theta} \implies |x^{k+1}| \leq \frac{1}{\varrho_0}|x^k|^{1+\eta-\theta}.$$

The influence of the parameters η and θ in the method will be clear in Section 4, where a good choice of θ and η lead us to a fast rate of convergence of the algorithm under mild assumption. As we will see the parameters (η, θ) are a way to deal with the possible absence of derivative of F and the failure of the error bound condition near the solution set.

4 From global to local convergence

The purpose of this section is twofold: to show fast convergence results for Algorithm 1 under some assumptions that do not include local uniqueness of solutions or differentiability; and to see, under the same assumptions, that the transition of Algorithm 2 to Algorithm 1 occurs naturally since the sufficient decrease condition (3) in Algorithm 2 is accepted for any iterate sufficiently close to the solution set.

Let $\bar{x}$ be a solution of (1) and let $L > 0$ be the scalar that satisfies $\|F(x)\| \leq L\,\mathrm{dist}(x, Z)$ for all x close to $\bar{x}$. Clearly, the opposite inequality is the EBC, and our analysis will consider a wider class of functions.

Assumption 2. *There exist $\varepsilon_H > 0$, $\omega > 0$ and $q \in (0, 1]$ such that $\mathrm{dist}(x, Z) \leq \omega\|F(x)\|^q$ holds for every $x \in \mathbb{B}(\bar{x}, \varepsilon_H) \cap \Omega$.*

This assumption provides an upper bound for the distance from any point sufficiently close to $\bar{x}$ based on the value of $\|F(x)\|$. Such property is also known as *Hölder metric subregularity of order q*. When $q = 1$ we have the EBC, extensively used in local analysis and studied for instace in [21, 4].

Assumption 3. *There exist $\varepsilon_S > 0$, $K > 0$ and $\alpha > 0$ such that*

$$\|F^j(y) - F^j(x) - G_j(x)(y - x)\| \leq K\|y - x\|^{1+\alpha} \tag{26}$$

holds for all $j = 1, \ldots, p$, every $y \in \mathbb{B}(\bar{x}, \varepsilon_S) \cap \Omega$ and every $x \in \mathbb{B}(\bar{x}, \varepsilon_S) \cap \Omega \setminus Z$.

Clearly (26) holds if F is differentiable on $\mathbb{B}(\bar{x}, \varepsilon_S) \cap \Omega$ with derivative being Hölder continuous with exponent α. When F is piecewice affine, (26) holds for all α. The parameter α attempt to measure the degree of smoothness of the selection mappings.

Assumption 4. *There exists $\varepsilon_A > 0$ such that $\mathcal{A}(z) = \mathcal{A}(\bar{x})$, $\forall z \in \mathbb{B}(\bar{x}, \varepsilon_A) \cap Z$.*

It is easy to see that the set-valued mapping $\mathcal{A}$ is always upper semi-continuous. In this context, Assumption 4 implies continuity of $\mathcal{A}$ at $\bar{x}$ relative to Z. This property was also used in [12, Proposition 6] for convergence results. This condition holds for problems of the form (25) at points satisfying the strict complementarity condition.

Before continue, we stress out some differences between the results given in this section and the previous one. By Theorem 3.5, any limit point of the sequence $\{x^k\}$ generated by Algorithm 2 is Clarke stationary under Assumption 1 and the continuous differentiability of the selection mappings defined on $\mathbb{R}^n$. Meanwhile, in this section, under Assumptions 1 to 4 (which do not imply the differentiability of the selection mappings on all $\mathbb{R}^n$) Algorithm 1 converges to a solution for every initial point sufficiently close to the solution set. Furthermore, under the same assumptions, (3) always is accepted, and neither ϱ_k nor $\mathcal{A}_k$ are updated.

Now, to establish fast convergence rate for Algorithm 1, the parameters (η, θ) must be in a suitable set. Here, we consider two different regions defined as

$$\mathcal{R}_1(q, \alpha) = \left\{(\eta, \theta) \in \mathbb{R}^2_{++} \,\middle|\, q(1 + \alpha) - \frac{1}{1 + \alpha} > \eta - \theta > \frac{1}{q} - 1\right\}, \tag{27}$$

and

$$\mathcal{R}_2(q, \alpha) = \left\{(\eta, \theta) \in \mathbb{R}^2_{++} \,\middle|\, 1 < \theta(1 + \alpha) \leq q(1 + \alpha),\, 1 < \eta \leq q(1 + \alpha)\right\}. \tag{28}$$

Clearly, both sets depend on the parameter (q,α). They appear when we analyze the convergence rate of the sequence $\{x^k\}$ generated by Algorithm 1: $\mathcal{R}_1(q,\alpha)$ comes when we study the behavior of $\operatorname{dist}(x^k,Z)$, and $\mathcal{R}_2(q,\alpha)$ when we consider the behavior of $\|F(x^k)\|$, which are not necessarily comparable since the EBC may not hold. For a geometrical interpretation of $\mathcal{R}_1(q,\alpha)$ and $\mathcal{R}_2(q,\alpha)$, see figs. 1 and 2 in the next subsections.

Proposition 4.1. *Let $\bar{x}$ be a solution of* (1) *satisfying Assumptions 2 to 4. Then, there exists $\varepsilon_0>0$ such that if $x^k\in\mathbb{B}(\bar{x},\varepsilon_0)\cap\Omega$ with $F(x^k)\neq 0$, $\mathcal{A}_k\subseteq\mathcal{A}(x^k)$ and (d^k,γ_k) is a solution of problem* (2) *for $\varrho=\varrho_k$, it holds that*

$$\gamma_k \le \beta_0(\rho_k)\,\|F(x^k)\|^{q_0}, \tag{29}$$

$$\|F(x^k)+G_j(x^k)d^k\| \le \beta_1(\varrho_k)\operatorname{dist}(x^k,Z)^{q_1}, \quad \forall j\in\mathcal{A}_k, \tag{30}$$

$$\|d^k\| \le \beta_2(\varrho_k)\operatorname{dist}(x^k,Z)^{q_2}, \tag{31}$$

where q_0, q_1, q_2 are defined by

$$q_0=\min\{q(1+\alpha)-\eta, q-\theta\},\ q_1=\min\{1+\alpha,1+\eta-\theta\},\ q_2=\min\{1,1+\alpha-(\eta-\theta)/q\},$$

and β_0, β_1, β_2 by $\beta_0(\varrho_k)=\max\left\{K\omega^{1+\alpha},\omega/\varrho_k\right\}$, $\beta_1(\varrho_k)=\max\left\{K,L^{\eta-\theta}/\varrho_k\right\}$ and $\beta_2(\varrho_k)=\max\{1,\varrho_k K\omega^{(\eta-\theta)/q}\}$.

Proof. Choose $\varepsilon_0\le\dfrac{1}{2}\min\{\varepsilon_H,\varepsilon_S,\varepsilon_A\}$ satisfying $\mathcal{A}(x)\subset\mathcal{A}(\bar{x})$, $\operatorname{dist}(x,Z)\le\min\{1,1/L\}$ and $\|F(x)\|\le L\operatorname{dist}(x,Z)$ for any $x\in\mathbb{B}(\bar{x},\varepsilon_0)\cap\Omega$. Take $x^k\in\mathbb{B}(\bar{x},\varepsilon_0)$ and let $z^k\in Z$ be such that $\|x^k-z^k\|=\operatorname{dist}(x^k,Z)$. Observe that $z^k\in\mathbb{B}(\bar{x},2\varepsilon_0)$ because

$$\|z^k-\bar{x}\|\le\operatorname{dist}(x^k,Z)+\|x^k-\bar{x}\|\le 2\|x^k-\bar{x}\|\le 2\varepsilon_0.$$

Using Assumptions 3 and 4, for every $j\in\mathcal{A}_k\subset\mathcal{A}(\bar{x})=\mathcal{A}(z^k)$ we get that

$$\begin{aligned}\|F(x^k)+G_j(x^k)(z^k-x^k)\| &= \|F^j(z^k)-F^j(x^k)-G_j(x^k)(z^k-x^k)\|\\ &\le K\|z^k-x^k\|^{1+\alpha}=K\operatorname{dist}(x^k,Z)^{1+\alpha}\\ &=\frac{K\operatorname{dist}(x^k,Z)^{1+\alpha}}{\|F(x^k)\|^{\eta}}\|F(x^k)\|^{\eta}.\end{aligned}$$

Clearly, we always have that

$$\|z^k-x^k\|=\frac{\operatorname{dist}(x^k,Z)}{\varrho_k\|F(x^k)\|^{\theta}}\varrho_k\|F(x^k)\|^{\theta}.$$

Thus, if (d^k,γ_k) is a solution of (2) then

$$\gamma_k\le\max\left\{\frac{K\operatorname{dist}(x^k,Z)^{1+\alpha}}{\|F(x^k)\|^{\eta}},\frac{\operatorname{dist}(x^k,Z)}{\varrho_k\|F(x^k)\|^{\theta}}\right\}.$$

Now, in one hand, since $\text{dist}(x^k, Z) \leq \omega \|F(x^k)\|^q$ and $\|F(x^k)\| \leq 1$, we have

$$\gamma_k \leq \max\left\{K\omega^{1+\alpha}, \frac{\omega}{\varrho_k}\right\} \|F(x^k)\|^{\min\{q(1+\alpha)-\eta, q-\theta\}}.$$

Obtaining (29).

On the other hand, we will bound $\gamma_k \|F(x^k)\|^\eta$ and $\gamma_k \varrho_k \|F(x^k)\|^\theta$. Indeed, by $\|F(x^k)\| \leq L\,\text{dist}(x^k, Z)$ and $\text{dist}(x^k, Z) \leq 1$, we get

$$\begin{aligned}\gamma_k \|F(x^k)\|^\eta &\leq \max\left\{K \text{dist}(x^k, Z)^{1+\alpha}, \frac{1}{\varrho_k} \text{dist}(x^k, Z)\|F(x^k)\|^{\eta-\theta}\right\} \\ &\leq \max\left\{K, L^{\eta-\theta}/\varrho_k\right\} \text{dist}(x^k, Z)^{\min\{1+\alpha, 1+\eta-\theta\}},\end{aligned}$$

and from $\|F(x^k)\|^{-1} \leq \omega^{1/q} \text{dist}(x^k, Z)^{-1/q}$, we obtain

$$\begin{aligned}\gamma_k \varrho_k \|F(x^k)\|^\theta &\leq \max\left\{\varrho_k K \frac{\text{dist}(x^k, Z)^{1+\alpha}}{\|F(x^k)\|^{\eta-\theta}}, \text{dist}(x^k, Z)\right\} \\ &\leq \max\{1, \varrho_k K \omega^{(\eta-\theta)/q}\} \,\text{dist}(x^k, Z)^{\min\{1, 1+\alpha-(\eta-\theta)/q\}}.\end{aligned}$$

Since (d^k, γ_k) is feasible for problem (2), then (30) and (31) hold. □

4.1 From global to local convergence in region $\mathcal{R}_1(q, \alpha)$

Let us describe our approach. First, Proposition 4.2 will show that the sufficient decrease condition (3) holds for every iterate near to $\bar{x}$, thus no reduction of the trust region is needed. And then, we will state the main result of this subsection in Theorem 4.4.

Now, some comments about region $\mathcal{R}_1(q, \alpha)$ defined in (27). It can be seen that $\mathcal{R}_1(q, \alpha) \neq \emptyset$ if $q > q_+(\alpha) := (\sqrt{\alpha^2 + 4(1+\alpha)^3} - \alpha)/2(1+\alpha)^2$, and in this case $q(1+\alpha) > 1$. A graphic representation is given in fig. 1. It is interesting to see that $\alpha \mapsto q_+(\alpha)$ is a non-increasing function with $\lim_{\alpha \to 0} q_+(\alpha) = 1$ and $\lim_{\alpha \to \infty} q_+(\alpha) = 0$. Thus, to ensure a rapid rate of convergence when the selection mappings are not sufficiently smooth, more regular must be the constrained set Z, that is, the Hölder exponent q must be close to 1. Also, if $\alpha = 1$, we have $q > q_+(1) = (\sqrt{33} - 1)/8$, the same lower bound considered in [1].

Proposition 4.2. *Let $\bar{x}$ be a solution of* (1) *satisfying Assumptions 1 to 4, with $(\eta, \theta) \in \mathcal{R}_1(q, \alpha)$. Then, there exists $\varepsilon_Q > 0$ such that for every $x^k \in \mathbb{B}(\bar{x}, \varepsilon_Q) \cap \Omega$ with $F(x^k) \neq 0$ and for any solution (d^k, γ_k) of problem* (2) *with $\mathcal{A}_k \subseteq \mathcal{A}(x^k)$ and $\varrho = \varrho_k$, it holds that $\Delta_k > 0$ and*

$$\|F(x^k + d^k)\| \leq \|F(x^k)\| - \sigma \Delta_k.$$

Proof. Let ε_0 be given by Proposition 4.1 and choose $\varepsilon_Q < \varepsilon_0$. By (31), d^k is close to zero if $x^k \in \mathbb{B}(x^*, \varepsilon_Q)$ for ε_Q small enough. Thus, for a small ε_Q we will have $x^k + d^k \in \mathbb{B}(x^*, \varepsilon_S)$. From Proposition 3.1, there exists $r \in \mathcal{A}_k$ such that $\gamma_k \|F(x^k)\|^\eta = \|F(x^k) + G_r(x^k) d^k\|$ and $\Delta_k = \|F(x^k)\| - \|F(x^k) + G_r(x^k) d^k\|$.

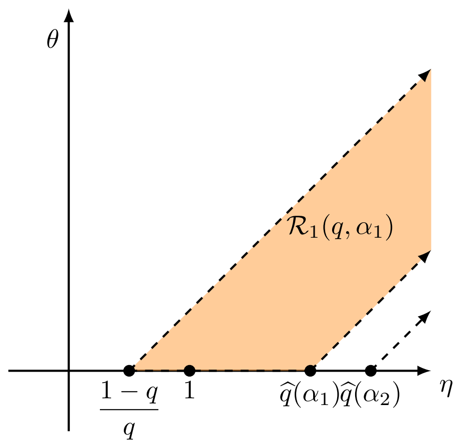

Figure 1: The region $\mathcal{R}_1(q,\alpha)$ is an open polyhedral with vertices $(\max{(1-q)/q}, 0)$ and $(\widehat{q}(\alpha), 0)$ where $\widehat{q}(\alpha) := q(1+\alpha)-(1+\alpha)^{-1}$. Furthermore, $\mathcal{R}_1(q,\alpha_1) \subset \mathcal{R}_1(q,\alpha_2)$ whenever $\alpha_1 \leq \alpha_2$. In the figure, $q > 1/2$.

Now, using Assumptions 1 and 3 and the triangle inequality, we get that

$$\begin{aligned}
\|F(x^k{+}d^k)\| - \|F(x^k)\| &\leq \|F^r(x^k+d^k)\| - \|F(x^k)\| \\
&\leq \|F^r(x^k+d^k) - F^r(x^k) - G_r(x^k)d^k\| + \|F(x^k) + G_r(x^k)d^k\| - \|F(x^k)\| \\
&\leq K\|d^k\|^{1+\alpha} - \Delta_k.
\end{aligned}$$

From the last inequality, we see

$$\|F(x^k+d^k)\| - \|F(x^k)\| + \sigma\Delta_k \leq K\|d^k\|^{1+\alpha} - (1-\sigma)\Delta_k. \tag{32}$$

On one hand, using (30) and Assumption 2, we have

$$\begin{aligned}
-\Delta_k &= \|F(x^k) + G_r(x^k)d^k\| - \|F(x^k)\| \leq \beta_1(\varrho_k)\,\mathrm{dist}(x^k,Z)^{q_1} - \omega^{1/q}\mathrm{dist}(x^k,Z)^{1/q} \\
&\leq -\mathrm{dist}(x^k,Z)^{1/q}\left(\omega^{1/q} - \beta_1(\varrho_k)\,\mathrm{dist}(x^k,Z)^{q_1-1/q}\right).
\end{aligned}$$

Thus, shrinking ε_Q we obtain that $-\Delta_k < 0$ since, as shown below, $q_1 - 1/q > 0$.

On the other hand, from (31) we have $\|d^k\|^{1+\alpha} \leq \beta_2(\varrho_k)^{1+\alpha}\mathrm{dist}(x^k,Z)^{q_2(1+\alpha)}$. Combining the last two inequalities with (32), we obtain

$$\begin{aligned}
&\frac{\|F(x^k+d^k)\| - \|F(x^k)\| + \sigma\Delta_k}{(1-\sigma)\mathrm{dist}(x^k,Z)^{1/q}} \\
&\leq -\left(\omega^{-1/q} - \beta_1(\varrho_k)\mathrm{dist}(x^k,Z)^{q_1-1/q} - \frac{K\beta_2(\varrho_k)^{1+\alpha}}{1-\sigma}\mathrm{dist}(x^k,Z)^{q_2(1+\alpha)-1/q}\right).
\end{aligned}$$

The right-hand side is negative for ε_Q small enough, since $q_1 - 1/q > 0$ and $q_2(1+\alpha) - 1/q > 0$. The last assertion comes from the definition of $\mathcal{R}_1$ (27) and the fact that $q_1 - 1/q > 0$ if and only if $1 < qq_1 = \min\{q(1+\alpha), q(1+\eta-\theta)\}$, if and only if $q(1+\alpha) > 1$, $\eta - \theta > \dfrac{1}{q} - 1$. And $q_2(1+\alpha) - 1/q > 0$ if and only if $1 < q(1+\alpha)q_2 = \min\{q(1+\alpha), q(1+\alpha)(1+\alpha-(\eta-\theta)/q)\}$, if and only if $q(1+\alpha) > 1$ and $q(1+\alpha) - \dfrac{1}{1+\alpha} > \eta - \theta$.

Thus, for a small ε_Q, the sufficient descent condition (3) is satisfied. $\square$

As a consequence of this result, (3) holds and ϱ_k is not updated in Step 4 of Algorithm 2. Thus, for any x^k close enough to $\bar{x}$, the parameter ϱ_k remains fixed.

Corollary 4.3. *Under the hypotheses of Proposition 4.2, and shrinking ε_Q, if $x^k \in \mathbb{B}(\bar{x}, \varepsilon_Q) \cap \Omega$ with $F(x^k) \neq 0$ and (d^k, γ_k) is a solution of problem* (2) *with $\mathcal{A}_k \subseteq \mathcal{A}(x^k)$, it holds that*

$$\operatorname{dist}(x^k + d^k, Z) \leq \beta_3(\varrho_k)\operatorname{dist}(x^k, Z)^{q_3}, \tag{33}$$

where $q_3 = \min\{q(1+\alpha), q(1+\alpha)(1+\alpha-(\eta-\theta)/q), q(1+\eta-\theta)\}$ and $\beta_3(\varrho_k) = 2^q\omega \max\{\beta_1(\varrho_k), K\beta_2(\varrho_k)^{1+\alpha}\}^q$.

Proof. Shrink ε_Q given by Proposition 4.2 such that, by (31), we get $x^k + d^k \in \mathbb{B}(\bar{x}, \varepsilon_S) \cap \mathbb{B}(\bar{x}, \varepsilon_H)$ and $\operatorname{dist}(x^k, Z) \leq 1$ if $x^k \in \mathbb{B}(\bar{x}, \varepsilon_Q)$.

Using Assumptions 1 and 3, (30) and (31), and taking $j \in \mathcal{A}_k$ we have that

$$\begin{aligned}
\|F(x^k + d^k)\| &\leq \|F^j(x^k + d^k)\| \\
&\leq \|F^j(x^k + d^k) - F^j(x^k) - G_j(x^k)d^k\| + \|F(x^k) + G_j(x^k)d^k\| \\
&\leq K\|d^k\|^{1+\alpha} + \beta_1(\varrho_k)\operatorname{dist}(x^k, Z)^{q_1} \\
&\leq 2\max\{K\beta_2(\varrho_k)^{1+\alpha}, \beta_1(\varrho_k)\}\operatorname{dist}(x^k, Z)^{\min\{(1+\alpha)q_2, q_1\}}.
\end{aligned}$$

Thus, by Assumption 2, we obtain

$$\begin{aligned}
\operatorname{dist}(x^k + d^k, Z) &\leq \omega\|F(x^k + d^k)\|^q \\
&\leq \omega\left(2\max\{K\beta_2(\varrho_k)^{1+\alpha}, \beta_1(\varrho_k)\}\right)^q \operatorname{dist}(x^k, Z)^{\min\{q(1+\alpha)q_2, qq_1\}}.
\end{aligned}$$

Thus (33) follows by setting $\beta_3(\varrho_k) := 2^q\omega\max\{\beta_1(\varrho_k), K\beta_2(\varrho_k)^{1+\alpha}\}^q$, and by the definition of q_1 and q_2

$$q_3 = \min\{q(1+\alpha)q_2, qq_1\} = \min\{q(1+\alpha), q(1+\alpha)(1+\alpha-(\eta-\theta)/q), q(1+\eta-\theta)\}.$$

Observe also that, $q(1+\alpha)q_2 > 1$ and $qq_1 > 1$. Obtaining $q_3 > 1$. $\square$

Here we present the main result of this subsection. The proof follows the ideas of [30, 1] used for the local convergence of the LM method but adapted to our LP-Newton scheme. Observe that Theorem 4.4 provides a region in the parameter space to ensure Q-superlinear convergence, and also says that iterates given by Algorithm 2 does not perform the globalization strategy when some point x^k is sufficiently close to a point $\bar{x}$ satisfying Assumptions 1 to 4.

Theorem 4.4. *Let $\bar{x}$ be a solution of* (1) *satisfying Assumptions 1 to 4, with $(\eta,\theta) \in \mathcal{R}_1(q,\alpha)$. Then, there exists $\varepsilon_1 > 0$ such that every sequence $\{x^k\}$ generated by Algorithm 1, with $\varrho_k := \varrho_0$, $\mathcal{A}_k := \mathcal{A} \subset \mathcal{A}(\bar{x})$ and any initial point $x^0 \in \mathbb{B}(\bar{x},\varepsilon_1) \cap \Omega$, satisfies that*

1. *$\{\operatorname{dist}(x_k,Z)\}$ is Q-superlinearly convergent to 0 with order q_3;*
2. *$\{\|F(x^k)\|\}$ is R-superlinearly convergent to 0. If $qq_3 > 1$, the sequence $\{\|F(x^k)\|\}$ is Q-superlinearly convergent to 0 with order at least qq_3;*
3. *$\{x^k\}$ converges Q-superlinearly to a solution $\widehat{x}$ with order q_2q_3.*

Proof. We split the proof into three parts. First, we will show that $\{x^k\}$ always stay in a bounded set; second, that $\{x^k\}$ is a Cauchy sequence, and finally, the rate of convergence of the sequences.

Suppose that $\{x^k\}$ is a sequence generated by Algorithm 1. Let ε_0, q_1, q_2 be given by Proposition 4.1 and ε_Q, q_3 be given by Corollary 4.3. Define $\varepsilon_* := \min\{1,\varepsilon_0,\varepsilon_Q\}$.

a) *Exists an $\varepsilon_1 > 0$ such that if $\|x^0 - \bar{x}\| \le \varepsilon_1$ then $\{x^k\} \subset \mathbb{B}(\bar{x},\varepsilon_*)$.* By induction, we will show that $x^k \in \mathbb{B}(\bar{x},\varepsilon_*)$ for every $k \in \mathbb{N}$. Let $\beta_{20} = \beta_2(\varrho_0)$, $\beta_{30} = \beta_3(\varrho_0)$ and

$$\varepsilon_1 = \min\left\{\left(\frac{\varepsilon_*}{1+2\beta_{20}}\right)^{q_2^{-1}}, 2^{-q_2^{-1}}\beta_{30}^{-(q_3-1)^{-1}}\right\}.$$

Since $\varepsilon_* \le 1$ and $1 \le q_2^{-1}$, then $x^0 \in \mathbb{B}(\bar{x},\varepsilon_*)$.

Now, take $k \in \mathbb{N}$. Assume that $x^i \in \mathbb{B}(\bar{x},\varepsilon_*)$ for $i = 0,\dots,k$. By Proposition 4.2 and Corollary 4.3, we have $x^{i+1} = x^i + d^i$ and $\varrho_i = \varrho_0$. For $0 \le i \le l \le k$, we have

$$\begin{aligned}\operatorname{dist}(x^l,Z) &\le \beta_{30}\operatorname{dist}(x^{l-1},Z)^{q_3} \le \beta_{30}^{1+q_3}\operatorname{dist}(x^{l-2},Z)^{q_3^2} \le \beta_{30}^{\sum_{r=0}^{i-1} q_3^r}\operatorname{dist}(x^{l-i},Z)^{q_3^i}\\ &\le \beta_{30}^{(q_3^i-1)(q_3-1)^{-1}}\operatorname{dist}(x^{l-i},Z)^{q_3^i} = \beta_{30}^{-(q_3-1)^{-1}}\left(\beta_{30}^{(q_3-1)^{-1}}\operatorname{dist}(x^{l-i},Z)\right)^{q_3^i}.\end{aligned} \tag{34}$$

Let $r_i = \left(\beta_{30}^{(q_3-1)^{-1}}\operatorname{dist}(x^i,Z)\right)^{q_2}$. Using (34) repeatedly we obtain

$$r_l \le r_{l-i}^{q_3^i} \le r_0^{q_3^l}.$$

Since $\operatorname{dist}(x^0,Z) \le \varepsilon_1$, we get $r_0 \le 1/2$. From (31) and (34) with $l = i + j$, we obtain

$$\begin{aligned}\|x^{l+1} - x^j\| &\le \textstyle\sum_{i=j}^{l}\|x^{i+1}-x^i\| \le \beta_{20}\sum_{i=j}^{l}\operatorname{dist}(x^i,Z)^{q_2} = \beta_{20}\sum_{i=0}^{l-j}\operatorname{dist}(x^{i+j},Z)^{q_2}\\ &\le \beta_{20}\textstyle\sum_{i=0}^{l-j}\beta_{30}^{-q_2(q_3-1)^{-1}}\left[\left(\beta_{30}^{(q_3-1)^{-1}}\operatorname{dist}(x^j,Z)\right)^{q_2}\right]^{q_3^i}\\ &= \beta_{20}\textstyle\sum_{i=0}^{l-j}\beta_{30}^{-q_2(q_3-1)^{-1}} r_j^{q_3^i} \le \beta_{20}\beta_{30}^{-q_2(q_3-1)^{-1}} r_j \sum_{i=0}^{l-j} r_0^{\lfloor q_3^j(q_3^i-1)\rfloor}\\ &\le \beta_{20}\beta_{30}^{-q_2(q_3-1)^{-1}} r_j \textstyle\sum_{i=0}^{\infty} r_0^i \le 2\beta_{20}\beta_{30}^{-q_2(q_3-1)^{-1}} r_j,\end{aligned} \tag{35}$$

where we use $r_j \le r_0^{q_3^j}$ and $r_0 \le 1/2$. From this inequality with $l = k$ and $j = 0$ we obtain

$$\begin{aligned}\|x^{k+1} - \bar{x}\| &\le \|x^0 - \bar{x}\| + \|x^{k+1} - x^0\| \le \varepsilon_1 + 2\beta_{20}\beta_{30}^{-q_2(q_3-1)^{-1}} r_0 \\ &\le \varepsilon_1 + 2\beta_{20}\beta_{30}^{-q_2(q_3-1)^{-1}} \left(\beta_{30}^{(q_3-1)^{-1}}\varepsilon_1\right)^{q_2} = \varepsilon_1 + 2\beta_{20}\varepsilon_1^{q_2} \\ &\le (1 + 2\beta_{20})\varepsilon_1^{q_2} \le \varepsilon_*,\end{aligned}$$

where we use $q_2 \le 1$. Hence, $x^k \in \mathbb{B}(\bar{x}, \varepsilon_*)$ for all $k \ge 0$.

Now, by Proposition 4.2 and by Step 4 of Algorithm 2, $x^{k+1} = x^k + d^k \ \forall k$ and ϱ_k is not updated. This observation says that Algorithm 1 and Algorithm 2 generate the same iterates when the initial point x^0 is sufficient close a solution

b) $\{x^k\}$ *is a Cauchy sequence.* This follows from (35) and from $r_j \le r_0^{q_3^j} \to 0$ if $j \to \infty$. Thus, $\{x^k\}$ converges to some point $\widehat{x}$ and since $\operatorname{dist}(x^k, Z)$ converges to 0, we get $\widehat{x} \in Z$.

c) *The sequence* $\{x^k\}$ *converges to a solution with Q-order* q_2q_3. Since $\{x^k\}$ converges to a solution $\widehat{x}$, taking limits for $l \to \infty$ and $j = k+1$ in (35), we obtain

$$\begin{aligned}\|x^{k+1} - \widehat{x}\| &\le 2\beta_{20}\beta_{30}^{-q_2(q_3-1)^{-1}} r_{k+1} \le 2\beta_{20}\beta_{30}^{-q_2(q_3-1)^{-1}} r_k^{q_3} \\ &\le 2\beta_{20}\beta_{30}^{-q_2(q_3-1)^{-1}} \left(\beta_{30}^{(q_3-1)^{-1}} \operatorname{dist}(x^k, Z)\right)^{q_2q_3} \le 2\beta_{20}\beta_{30}^{q_2}\|x^k - \widehat{x}\|^{q_2q_3}.\end{aligned}$$

The last expression implies that $\{x^k\}$ converges Q-superlinearly to $\widehat{x}$ with order q_2q_3.

Finally, note that $\|F(x^k)\|$ converges R-superlinearly to zero since $\{\operatorname{dist}(x^k, Z)\}$ is Q-superlinearly convergent to zero and $\|F(x^k)\| \le L\operatorname{dist}(x^k, Z)$. To obtain Q-superlinear convergence with order at least $qq_3 > 1$, we only need to observe that

$$\|F(x^{k+1})\| \le L\operatorname{dist}(x^{k+1}, Z) \le L\beta_{30}\operatorname{dist}(x^k, Z)^{q_3} \le L\beta_{30}\omega^{q_3}\|F(x^k\|^{qq_3}, \quad (36)$$

where we have used Corollary 4.3 and Assumption 2. □

Remark 1. Under mild assumptions, Theorem 4.4 guarantees that the sequences $\{\operatorname{dist}\,(x^k, Z)\}$ and $\{x^k\}$ converge Q-superlinearly to 0 and to $\widehat{x}$ with order of convergence $q_3(\eta,\theta) = \min\{q(1+\alpha), q(1+\alpha)(1+\alpha-(\eta-\theta)/q), q(1+\eta-\theta)\}$ and $q_2(\eta,\theta)q_3(\eta,\theta)$ respectively (where $q_2(\eta,\theta) = \min\{1, 1+\alpha-(\eta-\theta)/q\} \le 1$). Clearly, the order of convergence of $\operatorname{dist}(x^k, Z)$ is not lower than the order of convergence of x^k. It is natural to ask for what choice of $(\eta,\theta) \in \mathcal{R}_1(q,\alpha)$ we obtain the best order of convergence. Thus, in this region, the best order of convergence $q_3(\eta,\theta)$ and $q_2(\eta,\theta)q_3(\eta,\theta)$ is attained at (η,θ) with $\eta - \theta = q\dfrac{(1+\alpha)^2-1}{1+\alpha+q}$ whose optimal values are $q(1+\alpha)\dfrac{1+q(1+\alpha)}{1+\alpha+q}$ and $q(1+\alpha)(\dfrac{1+q(1+\alpha)}{1+\alpha+q})^2$, respectively. When, $q = 1$ and $\alpha = 1$, both orders of convergence are 2, as in Newton's method.

4.2 From global to local convergence in region $\mathcal{R}_2(q,\alpha)$

From Proposition 4.2, any iterate very close to a solution and under a suitable choice of the parameters, the sufficient decrease condition (3) is accepted and ϱ_k is not updated. Unfortunately to use $\mathcal{R}_1(q,\alpha)$, the parameter (η,θ) must satisfied certain inequalities that depends on the knowledge of the possibly unknown Hölder constant q, see (27), which may restrict the applicability of the method. Thus, a natural question is if it is possible to guaranteed a sufficient decrease condition when the parameters are in a simple region. This is the purpose of this subsection by considering $\mathcal{R}_2(q,\alpha)$.

It is not difficult to see $\mathcal{R}_2(q,\alpha) \neq \emptyset$ if $q(1+\alpha) > 1$. A graphic representation is given in fig. 2. We observe that $\mathcal{R}_1(q,\alpha)$ and $\mathcal{R}_2(q,\alpha)$ are not comparable in the sense that $\mathcal{R}_1(q,\alpha)$ is not a subset of $\mathcal{R}_2(q,\alpha)$ and vice versa. Furthermore, as will be shown in Theorem 4.6, to ensure a fast convergence (with order $\delta = \min\{\eta,(1+\alpha)\theta\}$), the parameter η and θ must be taken away from 1 and $(1+\alpha)^{-1}$, respectively.

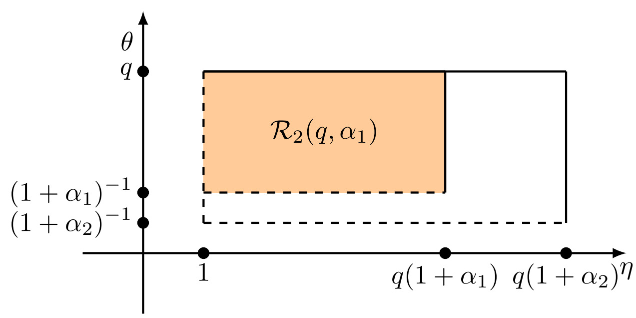

Figure 2: The region $\mathcal{R}_2(q,\alpha)$ clearly is a rectangle and $\mathcal{R}_2(q,\alpha_1) \subset \mathcal{R}_2(q,\alpha_2)$ whenever $\alpha_1 \leq \alpha_2$.

Now, we proceed by showing the fulfilment of sufficient decrease condition (3) when the parameters (η,θ) are in the region $\mathcal{R}_2(q,\alpha)$.

Proposition 4.5. *Let $\bar{x}$ be a solution of* (1) *satisfying Assumptions 1 to 4, with $(\eta,\theta) \in \mathcal{R}_2(q,\alpha)$. Then, there exists $\varepsilon_R > 0$ such that for every $x^k \in \mathbb{B}(\bar{x},\varepsilon_R)\cap\Omega$ with $F(x^k) \neq 0$ and for any solution (d^k,γ_k) of problem* (2) *with $\mathcal{A}_k \subseteq \mathcal{A}(x^k)$ and $\varrho = \varrho_k$, it holds that $\Delta_k > 0$ and*

$$\|F(x^k + d^k)\| \leq \|F(x^k)\| - \sigma\Delta_k.$$

Proof. Let ε_0 be given by Proposition 4.1. Since $\|F(x)\| \leq 1$ on $\mathbb{B}(\bar{x},\varepsilon_0)$ and $q_0 \geq 0$ by definition of $\mathcal{R}_2(q,\alpha)$ (28), then from (29) we have $\gamma_k \leq \beta_0(\varrho_k)$. Choose $\bar{\varepsilon}_R > 0$ such that $\bar{\varepsilon}_R < \varepsilon_0$, with $\beta_0(\varrho_k)\varrho_k\|F(x)\|^\theta \leq \bar{\varepsilon}_R/2$ and $\beta_0(\varrho_k)\|F(x)\|^{\eta-1} + K\beta_0(\varrho_k)^{1+\alpha}\varrho_k^{1+\alpha}\|F(x)\|^{\theta(1+\alpha)-1} \leq (1-\sigma)/2$ on $\mathbb{B}(\bar{x},\bar{\varepsilon}_R)$.

First, we will show $x^k + d^k \in \mathbb{B}(\bar{x},\bar{\varepsilon}_R)$ if $x^k \in \mathbb{B}(\bar{x},\bar{\varepsilon}_R/2)$. By triangle inequality

$$\|x^k + d^k - \bar{x}\| \leq \|x^k - \bar{x}\| + \|d^k\| \leq \|x^k - x^*\| + \gamma_k\varrho_k\|F(x^k)\|^\theta \leq \bar{\varepsilon}_R.$$

Hence if $\bar{\varepsilon}_R < \min\{\varepsilon_A, \varepsilon_H, \varepsilon_S\}$, we get $x^k + d^k \in \mathbb{B}(\bar{x}, \varepsilon_S)$. Now, take $j \in \mathcal{A}_k$, by Assumption 1 and the triangle inequality,

$$\begin{aligned}\|F(x^k + d^k)\| \leq & \|F^j(x^k + d^k) - F^j(x^k) - G_j(x^k)d^k\| + \|F(x^k) + G_j(x^k)d^k\| \\ \leq & K\|d^k\|^{1+\alpha} + \gamma_k\|F(x^k)\|^\eta \\ \leq & K\gamma_k^{1+\alpha}\varrho_k^{1+\alpha}\|F(x^k)\|^{\theta(1+\alpha)} + \gamma_k\|F(x^k)\|^\eta,\end{aligned}$$

when in the second inequality we use Assumption 3. Since $\Delta_k = \|F(x^k)\| - \gamma_k\|F(x^k)\|^\eta \leq \|F(x^k)\|$, rearranging the previous expression, we obtain

$$\begin{aligned}&\|F(x^k{+}d^k)\| - \|F(x^k)\| + \sigma\Delta_k \\ &\qquad \leq -\|F(x^k)\|\left\{(1-\sigma) - \gamma_k\|F(x^k)\|^{\eta-1} - K\gamma_k^{1+\alpha}\varrho_k^{1+\alpha}\|F(x^k)\|^{\theta(1+\alpha)-1}\right\}.\end{aligned}$$

Since $\gamma_k\|F(x^k)\|^{\eta-1} + K\gamma_k^{1+\alpha}\varrho_k^{1+\alpha}\|F(x^k)\|^{\theta(1+\alpha)-1}$ is less than $(1-\sigma)/2$ by the choice of $\bar{\varepsilon}_R$, the last inequality implies $\|F(x^k + d^k)\| - \|F(x^k)\| + \sigma\Delta(x^k) < 0$. Finally, $\Delta_k = \|F(x^k)\|(1 - \gamma_k\|F(x^k)\|^{\eta-1})$ is positive since $\gamma_k\|F(x^k)\|^{\eta-1} < 1$. Thus, for $\varepsilon_R := \bar{\varepsilon}_R/2$, the sufficient descent condition (3) holds. □

Clearly, if the error bound condition holds and the derivative of F^j is locally Lipschitz continuous on $\mathbb{R}^n$, we get $(q, \alpha) = (1, 1)$ and thus $(\eta, \theta) = (2, 1) \in \mathcal{R}_2(q, \alpha)$, which are the parameters used in the LP-Newton method. Now, we present our main result of this subsection.

Theorem 4.6. *Let $\bar{x}$ be a solution of* (1) *satisfying Assumptions 1 to 4, with $(\eta, \theta) \in \mathcal{R}_2(q, \alpha)$. Then, there exists $\varepsilon_2 > 0$ such that every sequence $\{x^k\}$ generated by Algorithm 1, with $\varrho_k := \varrho_0$, $\mathcal{A}_k := \mathcal{A} \subset \mathcal{A}(\bar{x})$ and any initial point $x^0 \in \mathbb{B}(\bar{x}, \varepsilon_2) \cap \Omega$, satisfies that*

1. *$\{F(x_k)\}$ is Q-superlinearly convergent to 0 with order $\delta = \min\{\eta, (1+\alpha)\theta\}$;*
2. *$\{dist(x^k, Z)\}$ is R-superlinearly convergent to* 0 *with order δ;*
3. *$\{x^k\}$ converges Q-superlinearly to a solution $\widehat{x}$ with order qq_2.*

Proof. As in the proof of Theorem 4.4, first, we will prove that $x^k \in \mathbb{B}(\bar{x}, \varepsilon_*)$ $\forall k \in \mathbb{N}$, where $\varepsilon_* := \min\{\varepsilon_0, \varepsilon_H\}/2$. Set $\beta_{20} = \beta_2(\varrho_0)$ and $\beta_{30} = \beta_3(\varrho_0)$.

Take $k \in \mathbb{N}$ and by induction, assume $x^i \in \mathbb{B}(\bar{x}, \varepsilon_*)$, for $i = 1, \ldots, k$. Since $\varepsilon_* < \varepsilon_0$, by (29) we have $\gamma_k \leq \bar{\gamma} = \beta_0(\varrho_0)$. Take $i = 1, \ldots, k-1$. For any $j \in \mathcal{A}(x^i)$, use Assumptions 1 and 3 and boundedness of γ_i to get

$$\begin{aligned}\|F(x^{i+1})\| \quad & \leq \|F^j(x^i + d^i) - F^j(x^i) - G_j(x^i)d^i\| + \|F(x^i) + G_j(x^i)d^i\| \\ & \leq K\|d^i\|^{1+\alpha} + \gamma_i\|F(x^i)\|^\eta \\ & \leq K\gamma_i^{1+\alpha}\varrho_0^{1+\alpha}\|F(x^i)\|^{\theta(1+\alpha)} + \gamma_i\|F(x^i)\|^\eta \leq \beta_f\|F(x^i)\|^\delta\end{aligned} \tag{37}$$

where $\beta_f = 2\max\{K\bar{\gamma}^{1+\alpha}\varrho_0^{1+\alpha}, \bar{\gamma}\}$ and $\delta = \min\{\eta, (1+\alpha)\theta\}$. By (28), $\delta > 1$. Thus, using (37) several times, for every $i = 1, \ldots, k$, we obtain

$$\|F(x^i)\| \leq \beta_f \|F(x^{i-1})\|^{\delta} \leq \beta_f^{1+\delta}\|F(x^{i-2})\|^{\delta^2} \leq \cdots \leq \beta_f^{\frac{\delta^i-1}{\delta-1}}\|F(x^0)\|^{\delta^i}. \tag{38}$$

Certainly, (28) implies $q_2 > 0$. Now, set $a_1 := \omega^{q_2}\beta_f^{\frac{-qq_2}{\delta-1}}$ and $b_1 := \beta_f^{\frac{1}{\delta-1}}$. By Proposition 4.1 and using $\operatorname{dist}(x^i, Z) \leq \omega\|F(x^i)\|^q$, for every $i = 1, \ldots, k$ we get

$$\begin{aligned}\operatorname{dist}(x^i, Z)^{q_2} &\leq \omega^{q_2}\|F(x^i)\|^{qq_2} \leq \omega^{q_2}\beta_f^{\frac{-qq_2}{\delta-1}}\left(\left(\beta_f^{\frac{1}{\delta-1}}\|F(x^0)\|\right)^{qq_2}\right)^{\delta^i} \\ &\leq a_1\left(\left(b_1\|F(x^0)\|\right)^{qq_2}\right)^{\delta^i}.\end{aligned}$$

We will use (38) and the last relation to show that $x^{k+1} \in \mathbb{B}(\bar{x}, \varepsilon_*)$. First, using repeatedly Proposition 4.1 for $i = 1, \ldots, k$, we obtain

$$\begin{aligned}\|x^{k+1} - \bar{x}\| &\leq \|x^0 - \bar{x}\| + \textstyle\sum_{i=0}^{k}\|d^i\| \leq \|x^0 - \bar{x}\| + \sum_{i=0}^{k}\beta_{20}\operatorname{dist}(x^i, Z)^{q_2} \\ &\leq \|x^0 - \bar{x}\| + \textstyle\sum_{i=0}^{k}\beta_{20}a_1\left(\left(b_1\|F(x^0)\|\right)^{qq_2}\right)^{\delta^i} \\ &\leq \|x^0 - \bar{x}\| + \beta_{20}a_1(b_1\|F(x^0)\|)^{qq_2}\textstyle\sum_{i=1}^{\infty}\left(\left(b_1\|F(x^0)\|\right)^{qq_2}\right)^{\delta^i-1}.\end{aligned} \tag{39}$$

Since q_2 is positive, we can choose x^0 sufficiently near to $\bar{x}$ such that $\left(b_1\|F(x^0)\|\right)^{qq_2}$ is less than 1/2 and hence $\displaystyle\sum_{i=1}^{\infty}\left(\left(b_1\|F(x^0)\|\right)^{qq_2}\right)^{\delta^i-1}$ is finite (here, we use that $\delta^i > i$ for i large enough). Thus, choose $\varepsilon_2 > 0$ small enough such that for every x^0 with $\|x^0 - \bar{x}\| \leq \varepsilon_2$ the right-sides of (39) is less than $\varepsilon_*/2$. With this choice of ε_2, we have $x^{k+1} \in \mathbb{B}(\bar{x}, \varepsilon_*)$. This observation also implies that Algorithm 1 and Algorithm 2 generate the same iterates when the initial point x^0 is sufficient close a solution. From all above inequalities, we obtain:

a) The inequality $\|F(x^k)\| \leq \beta_f^{\frac{-1}{\delta-1}}\left(b_1\|F(x^0)\|\right)^{\delta^k}$ holds for every $k \in \mathbb{N}$. The choice of ε_2 implies $b_1\|F(x^0)\| \leq 1/2$ and, as consequence, $\|F(x^k)\| \to 0$ when $k \to \infty$. From $\|F(x^{k+1})\| \leq \beta_f\|F(x^k)\|^{\delta}$, $\forall k \in \mathbb{N}$, the sequence $\{\|F(x^k)\|\}$ converges to 0 with Q-order of convergence δ.

b) Since $\operatorname{dist}(x^k, Z) \leq \omega\|F(x^k)\|^q$, then $\operatorname{dist}(x^k, Z)$ is R-superlinearly convergent to 0, since $\|F(x^k)\|^q$ is Q-superlinearly convergent to 0 with order δ.

c) To show that $\{x^k\}$ is a Cauchy sequence, it will be sufficient to prove that $\{\|d^k\|\}$ is summable. By Proposition 4.1 and Assumption 2, we obtain

$$\|d^k\| \leq \beta_{20}\operatorname{dist}(x^k, Z)^{q_2} \leq \beta_{20}a_1\left(\left(b_1\|F(x^0)\|\right)^{qq_2}\right)^{\delta^k}, \tag{40}$$

where the last term is summable by the choice of x^0. Note that $qq_2 > 0$ if, and only if $q(1+\alpha) > \eta - \theta$.

Let $\widehat{x}$ be the limit of $\{x^k\}$. From a expression similar to (39), we get

$$
\begin{aligned}
\|x^{k+1}-\widehat{x}\| &\leq \textstyle\sum_{i=0}^{\infty}\|d^{k+i}\| \leq \beta_{20}\sum_{i=0}^{\infty}\operatorname{dist}(x^{k+i},Z)^{q_2} \\
&\leq \beta_{20}\textstyle\sum_{i=0}^{\infty} a_1\left(\left(b_1\|F(x^k)\|\right)^{qq_2}\right)^{\delta^i} \\
&\leq \beta_{20}a_1\left(b_1\|F(x^k)\|\right)^{qq_2}\textstyle\sum_{i=1}^{\infty}\left(\left(b_1\|F(x^k)\|\right)^{qq_2}\right)^{\delta^i-1} \\
&\leq \beta_{20}a_1\left(b_1L\operatorname{dist}(x^k,Z)\right)^{qq_2}\textstyle\sum_{i=1}^{\infty}\left(\left(b_1\|F(x^k)\|\right)^{qq_2}\right)^{\delta^i-1} \\
&\leq \beta_{20}a_1b_1^{qq_2}L^{qq_2}\|x^k-\widehat{x}\|^{qq_2}\textstyle\sum_{i=1}^{\infty}\left(\left(b_1\|F(x^k)\|\right)^{qq_2}\right)^{\delta^i-1}, \qquad (41)
\end{aligned}
$$

where in the third inequality we use a modification of (40). The sum in the last inequality is finite with an upper bound that does not depend of k. So, by (41), there exists $M>0$ such that $\|x^{k+1}-\widehat{x}\| \leq M\|x^k-\widehat{x}\|^{qq_2}$, $\forall k\in\mathbb{N}$. This conclude with the demonstration. □

Theorem 4.6 is pleasant in the sense that we achieve a rapid rate of convergence for $\{\|F(x^k)\|\}$ and $\{\operatorname{dist}(x^k,Z)\}$, even if we do not know the Hölder exponent q, just by choosing η and $\theta(1+\alpha)$ close to 1. In Theorem 4.4, to obtain Q-superlinear convergence, the values of q are restricted. In fact, $q>(\sqrt{33}-1)/8$ if $\alpha=1$. On the other hand, using Theorem 4.6 to guarantee the convergence of the sequence $\{x^k\}$, q must be greater than $1/2$ if $\alpha\leq 1$. In view of $(\sqrt{33}-1)/8\approx 0.56>1/2$, Theorem 4.6 ensures the convergence of $\{x^k\}$ for a larger range of values for q, but possibly losing some order of convergence. Finally, when $q=1$, $\alpha=1$, $\eta=2$ and $\theta=1$, we get $q_3=2$ and $\delta=2$, which imply quadratic Q-convergence of $\operatorname{dist}(x^k,Z)$ and $\{\|F(x^k)\|\}$ by Theorems 4.4 and 4.6.

5 Numerical experiences

In this section, we present some numerical experiments for the globalized LP(η,θ)-Newton (Algorithm 2). Experiments were made using Octave. In order to obtain a linear programming subproblem, we choose norms associated to polyhedral unit balls. Therefore, subproblems were solve by using the GNU Linear Programming Kit (GLPK) with default parameters. Also, to avoid numerical instabilities as in [16], we rewrite problem (2) as

$$
\begin{array}{ll}
\underset{(d,\gamma)}{\text{minimize}} & \gamma \\
\text{subject to} & \|F(x^k)+G_j(x^k)d\| \leq \gamma\|F(x^k)\|^{\eta-\theta}, \quad \forall j\in\mathcal{A}_k, \\
& \|d\|\leq\gamma\varrho, \quad x^k+d\in\Omega.
\end{array}
$$

It can be seen that if γ_k is the optimum value for this problem, then the definition of Δ_k must be $\Delta_k=\|F(x^k)\|-\gamma_k\|F(x^k)\|^{\eta-\theta}$. For the next examples, we choose $\eta=q(1+\alpha)$ (the largest possible for $\mathcal{R}_2(q,\alpha)$) and, to obtain $\eta-\theta>1$ with $\theta>1/(1+\alpha)$ (the smallest possible for $\mathcal{R}_2(q,\alpha)$), we choose θ as the middle point between $\eta-1$ and $1/(1+\alpha)$.

Example 5.1. *This problem, from [12, Example 6], correspond to* (1) *with*

$$F(x) = \begin{bmatrix} x_4 + x_5 - x_6 - x_9 \\ x_4 + x_2 + x_3 - x_7 - x_9 \\ x_2 + x_3 - x_9 \\ x_1 + x_2 - x_8 \\ x_1 + x_{10} \\ -x_1 + x_{11} \\ 1 - x_2 + x_{12} \\ -x_4 + x_{13} \\ -x_1 - x_2 - x_3 + x_{14} \\ \min\{x_5, x_{10}\} \\ \min\{x_6, x_{11}\} \\ \min\{x_7, x_{12}\} \\ \min\{x_8, x_{13}\} \\ \min\{x_9, x_{14}\} \end{bmatrix}, \qquad \Omega = \mathbb{R}^4 \times \mathbb{R}^5_+ \times \mathbb{R}^5_+.$$

Its solution set is $Z = \{(0, t, -t, 0, s, s, 0, t, 0, 0, 0, t-1, 0, 0) \mid t \geq 1, s \geq 0\}$. *Since* F *is a reformulation of a mixed complementarity problem Assumption 1 holds. Moreover, Assumptions 2 and 3 hold since* F *is a piecewise affine function.*

We run Algorithm 2 for 100 *uniformly distributed random initial points in* $(-10, 10)^4 \times (0, 10)^5 \times (0, 10)^5$ *with parameters* $(\eta, \theta) = (2, 1)$ *and* $(\eta, \theta) = (3, 1.17)$.

Taking ℓ_∞ *norm in both spaces, domain and range, we obtain for* $(\eta, \theta) = (2, 1)$ *and* $(\eta, \theta) = (3, 1.167)$, *convergence to solution in* 97 *and* 95 *cases, to stationary point in* 3 *and* 4 *cases, and* GLPK *fails in* 0 *and* 1 *case. For sequences converging to solutions, the comparison is made in table 1.*

Table 1: Convergence to solutions for example 5.1 using $\ell_\infty - \ell_\infty$ norms

	$(\eta,\theta) = (2,1)$		
Iterations	7.639 ± 0.482		
CPU-Time	0.016 ± 0.003		
Pure LP-N	100%		
	Q-quad	Q-sup	Q-lin
$\|F(x^k)\|$	100%	0%	0%
$\text{dist}(x^k, Z)$	100%	0%	0%
$\|x^k - \bar{x}\|$	100%	0%	0%

	$(\eta,\theta) = (3,1.167)$		
Iterations	11.632 ± 2.222		
CPU-Time	0.025 ± 0.006		
Pure LP-N	100%		
	Q-quad	Q-sup	Q-lin
$\|F(x^k)\|$	100%	0%	0%
$\text{dist}(x^k, Z)$	100%	0%	0%
$\|x^k - \bar{x}\|$	100%	0%	0%

Considering ℓ_1 *norm in domain and* ℓ_∞ *in range with* $(\eta, \theta) = (2, 1)$ *and* $(\eta, \theta) = (3, 1.167)$, *we obtain convergence to solution in* 63 *and* 86 *cases, to stationary points in* 30 *and* 13 *cases, and* 7 *and* 1 *fails of* GLPK. *In table 2 is summarized the information when it converges to solutions.*

Table 2: Convergence to solutions for example 5.1 using $\ell_1 - \ell_\infty$ norms

	$(\eta,\theta)=(2,1)$		
Iterations	15.762 ± 1.563		
CPU-Time	3.881 ± 1.043		
Pure LP-N	98%		
	Q-quad	Q-sup	Q-lin
$\|F(x^k)\|$	73%	27%	0%
$\mathrm{dist}(x^k,Z)$	100%	0%	0%
$\|x^k-\bar{x}\|$	100%	0%	0%

	$(\eta,\theta)=(3,1.167)$		
Iterations	32.953 ± 9.151		
CPU-Time	6.168 ± 1.579		
Pure LP-N	97%		
	Q-quad	Q-sup	Q-lin
$\|F(x^k)\|$	98%	2%	0%
$\mathrm{dist}(x^k,Z)$	100%	0%	0%
$\|x^k-\bar{x}\|$	100%	0%	0%

Example 5.2. *Consider the problem in [12, Example 5], correspond to* (1) *with*

$$F(x)=\begin{bmatrix} x_1x_2-x_3 \\ x_1^2+x_2-1-x_4 \\ \min\{x_1,x_3\} \\ \min\{x_2,x_4\} \end{bmatrix}, \qquad \Omega=\mathbb{R}^2_+\times\mathbb{R}^2_+.$$

The solution set of this problem is given by $Z=\{(t,0,0,t^2-1)\mid t\geq 1\}\cup\{(0,1,0,0)\}$*. As noticed in [12] this problem satisfies Assumption 2, also it can be seen that Assumptions 1 and 3 hold. Running Algorithm 2 for 100 uniformly distributed initial points in* $(0,10)^2\times(0,10)^2$ *we obtaing the following results.*

Taking ℓ_∞ *norm in domain and range with* $(\eta,\theta)=(2,1)$ *and* $(\eta,\theta)=(2,0.75)$*, we obtain convergence to solution in all cases. A brief summary is in table 3.*

Table 3: Convergence to solutions for example 5.2 using $\ell_\infty - \ell_\infty$ norms

	$(\eta,\theta)=(2,1)$		
Iterations	12.85 ± 6.217		
CPU-Time	0.052 ± 0.054		
Pure LP-N	56%		
	Q-quad	Q-sup	Q-lin
$\|F(x^k)\|$	58%	40%	2%
$\mathrm{dist}(x^k,Z)$	39%	57%	4%
$\|x^k-\bar{x}\|$	52%	15%	33%

	$(\eta,\theta)=(2,0.75)$		
Iterations	12.77 ± 5.703		
CPU-Time	0.056 ± 0.061		
Pure LP-N	51%		
	Q-quad	Q-sup	Q-lin
$\|F(x^k)\|$	67%	22%	11%
$\mathrm{dist}(x^k,Z)$	51%	32%	17%
$\|x^k-\bar{x}\|$	56%	13%	31%

For experiments with ℓ_1 *norm in domain and* ℓ_∞ *in range, taking* $(\eta,\theta)=(2,1)$ *and* $(\eta,\theta)=(2,0.75)$*, we obtain convergence to solution in 80 and 74 cases, to stationary points in 9 and 10 cases, with 11 and 16 fails from* GLPK*. table 4 shows a detailed information about convergence to solutions.*

6 Conclusions and Remarks

In this work, we have presented a new globally convergent method for solving systems of nonlinear equations given by a piecewise smooth mapping with pos-

Table 4: Convergence to solutions for example 5.2 using $\ell_1 - \ell_\infty$ norms

	$(\eta,\theta)=(2,1)$		
Iterations	17.15 ± 5.842		
CPU-Time	0.105 ± 0.069		
Pure LP-N	19%		
	Q-quad	Q-sup	Q-lin
$\|F(x^k)\|$	44%	49%	7%
$\mathrm{dist}(x^k, Z)$	40%	54%	7%
$\|x^k - \bar{x}\|$	34%	23%	43%

	$(\eta,\theta)=(2,0.75)$		
Iterations	15.568 ± 5.667		
CPU-Time	0.053 ± 0.037		
Pure LP-N	27%		
	Q-quad	Q-sup	Q-lin
$\|F(x^k)\|$	66%	26%	8%
$\mathrm{dist}(x^k, Z)$	58%	35%	7%
$\|x^k - \bar{x}\|$	66%	7%	27%

sible non-isolated solutions. We have employed a globalisation technique based on a trust-region strategy. Our approach is quite general and it can be employed in other contexts. Evenmore, we obtained verifiable optimality condition for a general framework for minimizing the natural merit function given by the norm of the mapping defining the system of equations, see (6). We have also analysed its local convergence under a Hölderian error bound condition of the underlying mapping and Hölder continuity of the derivative of the selection mappings. Furthermore, we found different regions for the parameters (η, θ) where a fast rate of convergence occurs. Such regions can be used for choose or update the parameters in order to increase the order of convergence. It is worthy to mention that, for our hybrid method the transition between global and local convergence occurs naturally under mild assumptions.

There are several possible extensions to the present study. One approach is to analyze the local convergence under an additional assumption, as the Kurdyka-Lojasiewicz inequality for the LP(η, θ)-Newton method. Such inequality has been useful in the analysis of global behaviour of descent methods for nonsmooth functions, see for instance [3] and references therein. Another possible approach is to consider constrained nonsmooth system of equations where Ω is given by functional constraints. A similar question was analyzed in [20] by using the LM method for smooth mappings. Here, we expect numerical optimization methods for solving (4) with nice convergence properties under the constraint qualifications considered in [2].

References

[1] M. Ahookhosh, F. A. Artacho, R. M. T. Fleming, and P. T. Voung. Local convergence of the Levenberg-Marquardt method under Holder metric subregularity. *Advances in Computational Mathematics*, 45:2771–2806, 2019.

[2] R. Andreani, J.M. Martínez, A. Ramos, and P.J.S. Silva. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Mathematics of Operations Research*, 43(3):693–717, 2018.

[3] H. Attouch, J. Bolte, and B. F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting and regularized Gauss-Seidel methods. *Mathematical Programming*, 1:2:91–129, 2013.

[4] R. Behling, A. Fischer, G. Haeser, A. Ramos, and K. Schonefeld. On the constrained error bound condition and the projected Levenberg-Marquardt method. *Optimization*, 66 (8):1397–1411, 2017.

[5] D. P. Bertsekas. *Nonlinear programming*. Athenas Scientific, 1999.

[6] S. Billups and K. Murty. Complementarity problems. *Journal of Computational and Applied Mathematics*, 124:303, 2000.

[7] E. G. Birgin, N. Krejic, and J. M. Martinez. Globally convergent inexact quasi-Newton methods for solving nonlinear systems. *Numerical Algorithms*, 32:249–260, 2003.

[8] J. Bolte, T. P. Nguyen, J. Peypouquet, and B. W. Suter. From error bounds to the complexity of first-order descent methods for convex functions. *Mathematical Programming*, 165(2):471–507, 2017.

[9] R. L. Burden and J. D. Faires. *Numerical Analysis*. Thomson Brooks/Cole, 2005.

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

[11] F. Facchinei and J.S. Pang. *Finite Dimensional Variational Inequalities and Complementarity Problems*. Springer, 2003.

[12] F. Facchinnei, A. Fischer, and M. Herrich. An LP-Newton method: Nonsmooth equations, KKT systems, and nonisolated solutions. *Mathematical Programming*, 146:1–36, 2014.

[13] D. Fernandez and M. V. Solodov. Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. *SIAM Journal on Optimization*, 22:2:384–407, 2012.

[14] A. Fischer, M. Herrich, A. Izmailov, and M. Solodov. A globally convergent LP-Newton method. *SIAM Journal of Optimization*, 26:2012–2033, 2016.

[15] A. Fischer, M. Herrich, A. F. Izmailov, W. Scheck, and M. V. Solodov. A globally convergent LP-Newton method for piecewise smooth constrained equations: escaping nonstationary accumulation points. *Computational Optimization and Applications*, 69:325–349, 2018.

[16] A. Fischer, M. Herrich, A. F. Izmailov, and M. V. Solodov. A globally convergent LP-Newton method. *SIAM Journal on Optimization*, 26(4):2012–2033, 2016.

[17] K. Ito and K. Kunisch. *Lagrange Multiplier approach to Variational Problems and Applications.* SIAM Society for Industrial and Applied Mathematics, 2008.

[18] A. F. Izmailov, A. S. Kurennoy, and M. V. Solodov. Critical solutions of nonlinear equations: Stability issues. *Mathematical Programming*, 168:475–507, 2018.

[19] A. F. Izmailov and M. V. Solodov. *Newton-type methods for Optimization and Variational Problems.* Springer Series in Operation Research and Financial Engineering, 2014.

[20] A. F. Izmailov, M. V. Solodov, and E. I. Uskov. A globally convergent Levenberg-Marquardt method for equality-constrained optimization. *Computational Optimization and Applications*, 72:215–239, 2019.

[21] C. Kanzow, N. Yamashita, and M. Fukushima. Levenberg-Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints. *Journal of Computational and Applied Mathematics*, 172:375–397, 2004.

[22] G. Li, B. S. Mordukhovich, and T. S. Pham. New error bounds for polynomial systems with applications to holderian stability in optimization and spectral theory of tensors. *Mathematical Programming*, 153:333–362, 2015.

[23] M. A. Martínez and D. Fernández. A quasi-Newton modified LP-Newton method. *Optimization Methods and Software*, 2017.

[24] M. A. Martínez and D. Fernández. On the local and superlinear convergence of a secant modified linear-programming-Newton method. *Journal of Optimization Theory and Applications*, pages 1–18, 2018.

[25] J. Nocedal and S. J. Wright. *Numerical optimization.* Springer, 2006.

[26] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of Nonlinear equations in several variables.* Academic press, 1970.

[27] J. S. Pang. Error bounds in mathematical programming. *Mathematical Programming*, 79:299–332, 1997.

[28] R. T. Rockafellar and R. Wets. *Variational Analysis.* Series: Grundlehren der mathematischen Wissenschaften, Vol. 317, 2009.

[29] M. Ulbrich. *Semismooth Newton methods for Variational Inequalities and Constrained Optimization.* Society for Industrial and Applied Mathematics, 2011.

[30] N. Yamashita and M. Fukushima. *On the rate of convergence of the Levenberg-Marquardt method.* G. Alefeld, X. Chen (eds): Topics in Numerical Analysis, Vol. 15, 2001.