TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Operations Research and Control Systems
SWM ORCOS

Kantorovich-type Theorems for Generalized Equations

R. Cibulka, A. L. Dontchev, J. Preininger, T. Roubal and V. Veliov

Research Report 2015-16

November, 2015

Operations Research and Control Systems
Institute of Statistics and Mathematical Methods in Economics
Vienna University of Technology

Research Unit ORCOS
Wiedner Hauptstraße 8 / E105-4
1040 Vienna, Austria
E-mail: orcos@tuwien.ac.at

Kantorovich-type Theorems for Generalized Equations

R. Cibulka[1], **A. L. Dontchev**[2,3], **J. Preininger**[3], **T. Roubal**[4] **and V. Veliov**[3],

Abstract. We study convergence of the Newton method for solving generalized equations of the form $f(x) + F(x) \ni 0$, where f is a continuous but not necessarily smooth function and F is a set-valued mapping with closed graph, both acting in Banach spaces. We present a Kantorovich-type theorem concerning r-linear convergence for a general algorithmic strategy covering both nonsmooth and smooth cases. Under various conditions we obtain higher-order convergence. Examples and computational experiments illustrate the theoretical results.

Key Words. Newton's method, generalized equation, variational inequality, metric regularity, Kantorovich theorem, linear/superlinear/quadratic convergence.

AMS Subject Classification (2010) 49J53, 49J40, 65J15, 90C30.

[1]NTIS - New Technologies for the Information Society and Department of Mathematics, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 22, 306 14 Pilsen, Czech Republic, cibi@kma.zcu.cz. Supported by the project GA15-00735S.

[2]Mathematical Reviews, 416 Fourth Street, Ann Arbor, MI 48107-8604, USA, ald@ams.org. Supported by Austrian Science Foundation (FWF) Grant P26640-N25.

[3]Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Wiedner Hauptstrasse 8, A-1040 Vienna. Supported by Austrian Science Foundation (FWF) Grant P26640-N25.

[4]NTIS - New Technologies for the Information Society and Department of Mathematics, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 22, 306 14 Pilsen, Czech Republic, roubalt@students.zcu.cz. Supported by the project GA15-00735S.

1 Introduction

While there is some disagreement among historians who actually invented the Newton method, see [34] for an excellent reading about early history of the method, it is well documented in the literature that L. V. Kantorovich [22] was the first to obtain convergence of the method on assumptions involving the point where iterations begin. Specifically, Kantorovich considered the Newton method for solving the equation $f(x) = 0$ and proved convergence by imposing conditions on the derivative $Df(x_0)$ of the function f and the residual $\|f(x_0)\|$ at the starting point x_0. These conditions can be actually checked, in contrast to the conventional approach utilizing the assumption that the derivative $Df(\bar{x})$ at a (unknown) root $\bar{x}$ of the equation is invertible and then claim that if the iteration starts close enough to $\bar{x}$ then it generates a convergent to $\bar{x}$ sequence. For this reason Kantorovich's theorem is usually called a global convergence theorem[5] whereas conventional convergence theorems are regarded as local theorems.

The following version of Kantorovich's theorem is close to that in [27]; for a proof see [27] or [23].

Theorem 1.1 (Kantorovich)**.** *Let X and Y be Banach spaces. Consider a function $f : X \to Y$, a point $x_0 \in X$ and a real $a > 0$, and suppose that f is continuously Fréchet differentiable in an open neighborhood of the ball $\mathbb{B}_a(x_0)$ and its Fréchet derivative Df is Lipschitz continuous in $\mathbb{B}_a(x_0)$ with a constant $L > 0$. Assume that there exist positive reals κ and η such that*

$$\|Df(x_0)^{-1}\| \leq \kappa \quad \textit{and} \quad \|Df(x_0)^{-1}f(x_0)\| < \eta.$$

If $\alpha := \kappa L \eta a < \frac{1}{2}$ and $a \geq a_0 := \frac{1-\sqrt{1-2\alpha}}{\kappa L}$, then there exists a unique sequence $\{x_k\}$ satisfying the iteration

$$f(x_k) + Df(x_k)(x_{k+1} - x_k) = 0, \quad k = 0, 1, \ldots, \tag{1}$$

with a starting point x_0; this sequence converges to a unique zero $\bar{x}$ of f in $\mathbb{B}_{a_0}(x_0)$ and the convergence rate is r-quadratic; specifically

$$\|x_k - \bar{x}\| \leq \frac{\eta}{\alpha}(2\alpha)^{2^k}, \quad k = 0, 1, \ldots.$$

In his proof of convergence Kantorovich used a novel technique of *majorization* of the sequence of iterate increments by the increments of a sequence of scalars. Notice that the derivative Df is injective not only at x_0 but also at the solution $\bar{x}$; indeed, for any $y \in X$ with $\|y\| = 1$ we have

$$\|Df(\bar{x})y\| \geq \|Df(x_0)y\| - \|(Df(\bar{x}) - Df(x_0))y\| \geq \frac{1}{\kappa} - La_0 = \frac{\sqrt{1-2\alpha}}{\kappa} > 0.$$

In a related development, Kantorovich showed in [23, Chapter 18] that, under the same assumptions as in Theorem 1.1, to achieve linear convergence to a solution there is no need to calculate during iterations the derivative $Df(x_k)$ at the current point x_k— it is enough

[5] Some authors prefer to call such a result a semilocal convergence theorem.

to use at each iteration the value of the derivative $Df(x_0)$ at the starting point, i.e., the iteration (1) becomes

$$f(x_k) + Df(x_0)(x_{k+1} - x_k) = 0, \quad k = 0, 1, \ldots. \tag{2}$$

He called this method the *modified Newton process*. This method is also known as the *chord method*, see [24, Chapter 5].

The work of Kantorovich has been extended in a number of ways by, in particular, utilizing various extensions of the majorization technique, such as the method of nondiscrete induction, see e.g. [29]. We will not go into discussing these works here but rather focus on a version of Kantorovich's theorem due to R. G. Bartle [6], which has been largely forgotten if not ignored in the literature. A version of Bartle's theorem, without referring to [6], was given recently in [9, Theorem 5].

Specifically, Bartle [6] considered the equation $f(x) = 0$, for a function f acting between Banach spaces X and Y, which is solved by the iteration

$$f(x_k) + Df(z_k)(x_{k+1} - x_k) = 0, \quad k = 0, 1, \ldots, \tag{3}$$

where z_k are, to quote [6], "arbitrarily selected points ... sufficiently close to the solution desired." For $z_k = x_k$ one obtains the usual Newton method, and for $z_k = x_0$ the modified Newton/chord method, but z_k may be chosen in other ways. For example as x_0 for the first s iterations and then the derivative could be calculated again every s iterations, obtaining in this way a *hybrid* version of the method. If computing the derivatives, in particular in the case they are obtained numerically, involves time consuming procedures, it is quite plausible to expect that for large scale problems the chord method or a hybrid version of it would possibly be faster than the usual method. We present here the following somewhat modified statement of Bartle's theorem which fits our purposes:

Theorem 1.2 (Bartle [6]). *Assume that the function $f : X \to Y$ is continuously Fréchet differentiable in an open set O. Let $x_0 \in O$ and let there exist positive reals a and κ such that for any three points $x_1, x_2, x_3 \in I\!B_a(x_0) \subset O$ we have*

$$\|Df(x_1)^{-1}\| < \kappa \quad \text{and} \quad \|f(x_1) - f(x_2) - Df(x_3)(x_1 - x_2)\| \leq \frac{1}{2\kappa}\|x_1 - x_2\|, \tag{4}$$

and also

$$\|f(x_0)\| < \frac{a}{2\kappa}. \tag{5}$$

Then for every sequence $\{z_k\}$ with $z_k \in I\!B_a(x_0)$ there exists a unique sequence $\{x_k\}$ satisfying the iteration (3) with initial point x_0; this sequence converges to a root $\bar{x}$ of f which is unique in $I\!B_a(x_0)$ and the convergence rate is r-linear; specifically

$$\|x_k - \bar{x}\| \leq 2^{-k} a, \quad k = 0, 1, \ldots.$$

In a path-breaking paper Qi and Sun [30] extended the Newton method to a nonsmooth equation by employing Clarke's generalized Jacobian $\bar{\partial} f$ of a function $f : \mathbb{R}^n \to \mathbb{R}^n$ instead of the derivative Df and proved convergence for a class of nonsmooth functions. Specifically, consider the following iteration: given x_k choose any matrix A_k from $\bar{\partial} f(x_k)$ and then find the next iterate by solving the linear equation

$$f(x_k) + A_k(x_{k+1} - x_k) = 0, \quad k = 0, 1, \ldots. \tag{6}$$

The following convergence theorem was proved in [30, Theorem 3.2]:

Theorem 1.3. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous around a root $\bar{x}$ at which all matrices in $\bar{\partial} f(\bar{x})$ are nonsingular. Also assume that for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $x \in I\!B_\delta(\bar{x})$ and for every $A \in \bar{\partial} f(x)$ one has*

$$\|f(x) - f(\bar{x}) - A(x - \bar{x})\| \leq \varepsilon \|x - \bar{x}\|. \tag{7}$$

Then there exists a neighborhood U of $\bar{x}$ such that for every starting point $x_0 \in U$ there exists a sequence satisfying the iteration (6) *and every such sequence is superlinearly convergent to $\bar{x}$.*

A function f which is Lipschitz continuous around a point $\bar{x}$ and satisfies (7) is said to be *semismooth*[6] at $\bar{x}$. Accordingly, the method (6) is a *semismooth Newton method* for solving equations. For more advanced versions of Theorem 1.3, see e.g. [15, Theorem 7.5.3], [21, Theorem 2.42] and [14, Theorem 6F.1].

In the same paper Qi and Sun proved what they called a "global" theorem [30, Theorem 3.3], which is more in the spirit of Kantorovich's theorem; we will state and prove an improved version of this theorem in the next section.

In this paper we derive Kantorovich-type theorems for a generalized equation: find a point $x \in X$ such that

$$f(x) + F(x) \ni 0, \tag{8}$$

where throughout $f : X \to Y$ is a continuous function and $F : X \rightrightarrows Y$ is a set-valued mapping with closed graph. Many problems can be formulated as (8), for example, equations, variational inequalities, constraint systems, as well as optimality conditions in mathematical programming and optimal control.

Newton-type methods for solving nonsmooth equations and variational inequalities have been studied since the 70s. In the last two decades a number of new developments have appeared some of which have been collected in several books [15, 18, 19, 25, 33]. A broad presentation of convergence results for both smooth and nonsmooth problem with particular emphasis on applying Newton-type method to optimization can be found in the recent book [21]. A Kantorovich-type theorem for generalized equations under metric regularity is proven in [13, Theorem 2] using the majorization technique, see also the recent papers [2] and [32]. Related results for particular nonsmooth generalized equations are given in [16] and [28]. In [8] applications of the modified Newton method for solving optimization problems appearing in nonlinear model predictive control are reported.

We adopt the notations used in the book [14]. The set of all natural numbers is denoted by $I\!N$ and $I\!N_0 = I\!N \cup \{0\}$; the n-dimensional Euclidean space is $\mathbb{R}^n$. Throughout X and Y are Banach spaces both norms of which are denoted by $\|\cdot\|$. The closed ball centered at x with

[6] Sometimes one adds to (7) the condition that f is directionally differentiable in every direction.

radius r is denoted as $\mathbb{B}_r(x)$; the unit ball is $\mathbb{B}$. The distance from a point x to a set A is $\operatorname{dist}(x, A) = \inf_{y\in A} \|x-y\|$. A generally set-valued mapping $F : X \rightrightarrows Y$ is associated with its graph $\operatorname{gph} F = \{ (x, y) \in X \times Y \,|\, y \in F(x)\}$ and its domain $\operatorname{dom} F = \{ x \in X \,|\, F(x) \neq \emptyset\}$. The inverse of F is $y \mapsto F^{-1}(y) = \{ x \in X \,|\, y \in F(x)\}$. By $\mathcal{L}(X, Y)$ we denote a space of linear bounded operators acting from X into Y equipped with the standard operator norm.

Recall that a set-valued mapping $\Phi : X \rightrightarrows Y$ is said to be *metrically regular* at x_0 for y_0 if $y_0 \in \Phi(x_0)$ and there exist neighborhoods U of x_0 and V of y_0 and a positive constant κ such that the set $\operatorname{gph}\Phi \cap (U \times V)$ is closed and

$$\operatorname{dist}\big(x, \Phi^{-1}(y)\big) \le \kappa \operatorname{dist}\big(y, \Phi(x)\big) \quad \text{for all} \quad (x, y) \in U \times V. \tag{9}$$

The infimum over all $\kappa > 0$ in (9) is the regularity modulus of Φ at x_0 for y_0 denoted by $\operatorname{reg}(\Phi; x_0|y_0)$. If in addition the mapping $\sigma :\ V \ni y \mapsto \Phi^{-1}(y) \cap U$ is not multivalued on V, then Φ is said to be strongly metrically regular and then σ is a Lipschitz continuous function on V. More about metric regularity and the related theory can be found in [14].

2 Main theorem

In preparation to our main result presented in Theorem 2.2 we give a strengthened version of [30, Theorem 3.3] for the iteration (6) applied to an equation in Banach spaces.

Theorem 2.1. *Let $f : X \to Y$ be a continuous function and let the numbers $a > 0$, $\kappa > 0$, $\delta \ge 0$ be such that*

$$\kappa\delta < 1 \qquad \text{and} \qquad \|f(x_0)\| < (1-\kappa\delta)\frac{a}{\kappa}. \tag{10}$$

Consider the iteration (6) with a starting point x_0 and a sequence $\{A_k\}$ of linear and bounded mappings such that for every $k \in \mathbb{N}_0$ we have

$$\|A_k^{-1}\| \le \kappa \quad \text{and} \quad \|f(x)-f(x')-A_k(x-x')\| \le \delta\|x-x'\| \quad \text{for every } x, x' \in \mathbb{B}_a(x_0). \tag{11}$$

Then there exists a unique sequence satisfying the iteration (6) with initial point x_0. This sequence remains in $\operatorname{int}\mathbb{B}_a(x_0)$ and converges to a root $\bar{x} \in \operatorname{int}\mathbb{B}_a(x_0)$ of f which is unique in $\mathbb{B}_a(x_0)$; moreover, the convergence rate is r-linear:

$$\|x_k - \bar{x}\| < (\kappa\delta)^k a.$$

Proof. Let $\alpha := \kappa\delta$. We will show, by induction, that there is a sequence $\{x_k\}$ with elements in $\operatorname{int}\mathbb{B}_a(x_0)$ satisfying (6) with the starting point x_0 such that

$$\|x_{j+1} - x_j\| \le \alpha^j\kappa\|f(x_0)\| < a\alpha^j(1-\alpha), \quad j = 0, 1, \ldots. \tag{12}$$

Let $k := 0$. Since A_0 is invertible, there is a unique $x_1 \in X$ such that $A_0(x_1 - x_0) = -f(x_0)$. Therefore,

$$\|x_1 - x_0\| = \|A_0^{-1}A_0(x_1 - x_0)\| = \|A_0^{-1}f(x_0)\| \le \kappa\|f(x_0)\| < a(1-\alpha).$$

Hence $x_1 \in \operatorname{int}\mathbb{B}_a(x_0)$. Suppose that, for some $k \in \mathbb{N}$, we have already found points $x_0, x_1, \ldots, x_k \in \operatorname{int}\mathbb{B}_a(x_0)$ satisfying (12) for each $j = 0, 1, \ldots, k-1$. Since A_k is invertible, there is a unique $x_{k+1} \in X$ such that $A_k(x_{k+1} - x_k) = -f(x_k)$. Then (12) with $j := k - 1$ implies

$$\begin{aligned} \|x_{k+1} - x_k\| &= \|A_k^{-1} A_k(x_{k+1} - x_k)\| = \|A_k^{-1} f(x_k)\| \leq \kappa \|f(x_k)\| \\ &= \kappa \|f(x_k) - f(x_{k-1}) - A_{k-1}(x_k - x_{k-1})\| \\ &\leq \kappa\delta \|x_k - x_{k-1}\| \leq \alpha^k \kappa \|f(x_0)\| < a\alpha^k(1 - \alpha). \end{aligned}$$

From (12), we have

$$\|x_{k+1} - x_0\| \leq \sum_{j=0}^{k} \|x_{j+1} - x_j\| \leq \sum_{j=0}^{k} \alpha^j \kappa \|f(x_0)\| < a \sum_{j=0}^{\infty} \alpha^j (1 - \alpha) = a, \tag{13}$$

that is, $x_{k+1} \in \operatorname{int} I\!B_a(x_0)$. The induction step is complete.

For any natural k and p we have

$$\|x_{k+p+1} - x_k\| \leq \sum_{j=k}^{k+p} \|x_{j+1} - x_j\| \leq \sum_{j=k}^{k+p} \alpha^j \kappa \|f(x_0)\| < \frac{\alpha^k}{1 - \alpha} \kappa \|f(x_0)\| < a\alpha^k. \tag{14}$$

Hence $\{x_k\}$ is a Cauchy sequence; let it converge to $\bar{x} \in X$. Passing to the limit with $p \to \infty$ in (14) we obtain

$$\|\bar{x} - x_k\| \leq \frac{\alpha^k}{1 - \alpha} \kappa \|f(x_0)\| < a\alpha^k \quad \text{for each} \quad k \in I\!N_0.$$

In particular, $\bar{x} \in \operatorname{int} I\!B_a(x_0)$. Using (6) and (11), we get

$$0 \leq \|f(\bar{x})\| = \lim_{k \to \infty} \|f(x_k)\| = \lim_{k \to \infty} \|f(x_k) - f(x_{k-1}) - A_{k-1}(x_k - x_{k-1})\| \leq \lim_{k \to \infty} \delta \|x_k - x_{k-1}\| = 0.$$

Hence, $f(\bar{x}) = 0$. Suppose that there is $\bar{y} \in I\!B_a(x_0)$ with $\bar{y} \neq \bar{x}$ and $f(\bar{y}) = 0$. Then

$$\begin{aligned} \|\bar{y} - \bar{x}\| &\leq \kappa \|A_0(\bar{y} - \bar{x})\| = \kappa \|f(\bar{y}) - f(\bar{x}) - A_0(\bar{y} - \bar{x})\| \\ &\leq \kappa\delta \|\bar{y} - \bar{x}\| < \|\bar{y} - \bar{x}\|, \end{aligned}$$

which is a contradiction. Hence $\bar{x}$ is a unique root of f in $I\!B_a(x_0)$. □

Our main result which follows is an extension of Theorem 2.1 for generalized equations (8). We adopt the following model of an iterative procedure for solving (8). Given $k \in I\!N_0$, based on the current and prior iterates x_n ($n \leq k$) one generates a "feasible" element $A_k \in \mathcal{L}(X, Y)$ and then the next iterate x_{k+1} is chosen according to the following Newton-type iteration:

$$f(x_k) + A_k(x_{k+1} - x_k) + F(x_{k+1}) \ni 0. \tag{15}$$

In order to formalize the choice of A_k we consider a sequence of mappings $\mathcal{A}_k : X^k \to \mathcal{L}(X, Y)$, where $X^k = X \times \ldots \times X$ is the product of k copies of X. Thus, A_k does not need to be chosen in advance and may depend on the already obtained iterates. In particular, one may take $A_k = \mathcal{A}_0(x_0)$, that is, use the same operator for all iterations, as in the standard chord method. Another possibility is to use $A_k = Df(x_k)$ in the case of a differentiable f or $A_k \in \bar{\partial} f(x_k)$, the Clarke generalized Jacobian if applicable. Intermediate choices are also possible, for example to use the same operator A in m successive steps and then to update it at the current point: $\mathcal{A}_k(x_0, \ldots, x_k) = \mathcal{A}_{m[k/m]}(x_{m[k/m]})$, where $[s]$ is the integer part of s.

Theorem 2.2. *Let the scalars $a > 0$, $b > 0$, $\kappa > 0$, $\delta \geq 0$ and the points $x_0 \in X$, $y_0 \in f(x_0) + F(x_0)$ be such that*

(A1) $\kappa\delta < 1$ and $\|y_0\| < (1 - \kappa\delta)\min\{\frac{a}{\kappa}, b\}$.

Moreover, assume there exists a function $\omega : [0, a] \to [0, \delta]$ such that for every $k \in \mathbb{N}_0$ and every $x_1, \ldots, x_k \in \mathbb{B}_a(x_0)$ the linear and bounded operator $A_k := A_k(x_0, \ldots, x_k)$ appearing in the iteration (15) *has the following properties:*

(A2) the mapping

$$x \mapsto G_{A_k}(x) := f(x_0) + A_k(x - x_0) + F(x) \tag{16}$$

is metrically regular at x_0 for y_0 with constant κ and neighborhoods $\mathbb{B}_a(x_0)$ and $\mathbb{B}_b(y_0)$;

(A3) $\|f(x) - f(x_k) - A_k(x - x_k)\| \leq \omega(\|x - x_k\|)\,\|x - x_k\|$ for every $x \in \mathbb{B}_a(x_0)$.

Then for every $\alpha \in (\kappa\delta, 1)$ there exists a sequence $\{x_k\}$ generated by the iteration (15) *with starting point x_0 which remains in $\operatorname{int} \mathbb{B}_a(x_0)$ and converges to a solution $\bar{x} \in \operatorname{int} \mathbb{B}_a(x_0)$ of* (8)*; moreover, the convergence rate is r-linear; specifically*

$$\|x_k - \bar{x}\| < \alpha^k a \quad \text{and} \quad \operatorname{dist}(0, f(x_k) + F(x_k)) \leq \alpha^k \|y_0\| \quad \text{for every } k \in \mathbb{N}_0. \tag{17}$$

If $\lim_{\xi \to 0} \omega(\xi) = 0$, then the sequence $\{x_k\}$ is convergent r-superlinearly, that is, there exist sequences of positive numbers $\{\varepsilon_k\}$ and $\{\eta_k\}$ such that $\|x_k - \bar{x}\| \leq \varepsilon_k$ and $\varepsilon_{k+1} \leq \eta_k \varepsilon_k$ for all sufficiently large $k \in \mathbb{N}$ and $\eta_k \to 0$.

If there exists a constant $L > 0$ such that $\omega(\xi) \leq \min\{\delta, L\xi\}$ for each $\xi \in [0, a]$, then the convergence of $\{x_k\}$ is r-quadratic: specifically, there exists a sequence of positive numbers $\{\varepsilon_k\}$ such that for any $C > \frac{\alpha L}{\delta}$ we have $\varepsilon_{k+1} < C\varepsilon_k^2$ for all sufficiently large $k \in \mathbb{N}$.

If the mapping G_{A_k} defined in (16) *is not only metrically regular but also strongly metrically regular with the same constant and neighborhoods, then there is no other sequence $\{x_k\}$ satisfying the iteration (15) starting from x_0 which stays in $\mathbb{B}_a(x_0)$.*

Proof. Choose an $\alpha \in (\kappa\delta, 1)$ and then κ' such that

$$\frac{\alpha}{\delta} \geq \kappa' > \kappa \quad \text{and} \quad \|y_0\| < (1 - \alpha)\min\left\{\frac{a}{\kappa'}, b\right\}. \tag{18}$$

Such a choice of κ' is possible for $\alpha > \kappa\delta$ sufficiently close to $\kappa\delta$. We shall prove the claim for an arbitrary value of α for which (18) holds with an appropriately chosen $\kappa' > \kappa$. This is not a restriction, since then (17) will hold for any larger value of α.

We will show that there exists a sequence $\{x_k\}$ with the following properties, for each $k \in \mathbb{N}$:

(a) $\|x_k - x_0\| \leq \frac{1-\alpha^k}{1-\alpha}\kappa'\|y_0\| < (1 - \alpha^k)a$;

(b) $\|x_k - x_{k-1}\| \leq \alpha^{k-1}\gamma_0 \ldots \gamma_{k-1}\kappa'\|y_0\| < \alpha^{k-1}(1 - \alpha)a$,
where $\gamma_0 := 1$, $\gamma_i := \omega(\|x_i - x_{i-1}\|)/\delta$ for $i = 1, \ldots, k - 1$;

(c) $0 \in f(x_{k-1}) + A_{k-1}(x_k - x_{k-1}) + F(x_k)$,
where $A_{k-1} := A_{k-1}(x_0, \ldots, x_{k-1})$.

We use induction, starting with $k = 1$. Since $0 \in I\!\!B_b(y_0)$ and $y_0 \in G_{A_0}(x_0)$, using (A2) for G_{A_0} we have that

$$\mathrm{dist}\left(x_0, G_{A_0}^{-1}(0)\right) \leq \kappa\, \mathrm{dist}\left(0, G_{A_0}(x_0)\right) \leq \kappa \|y_0\|.$$

If $y_0 = 0$, then we take $x_1 = x_0$. If not, we have that

$$\mathrm{dist}\left(x_0, G_{A_0}^{-1}(0)\right) < \kappa' \|y_0\|$$

and then there exists a point $x_1 \in G_{A_0}^{-1}(0)$ such that

$$\|x_1 - x_0\| < \kappa' \|y_0\| < (1-\alpha)a.$$

Clearly, (a)–(c) are satisfied for $k := 1$ and γ_1 is well-defined.

Assume that for some $k \in I\!\!N$ the point x_k has already been defined in such a way that conditions (a)–(c) hold. We shall define x_{k+1} so that (a)–(c) remain satisfied for k replaced with $k+1$.

First, observe that (a) implies $x_k \in I\!\!B_a(x_0)$. Denote $r_k := f(x_0) - f(x_k) - A_k(x_0 - x_k)$. In view of (a), the fact that $\omega(\|x_0 - x_k\|) \leq \delta$ and (A3) with $x = x_0$, we have

$$\begin{aligned} \|r_k - y_0\| &\leq \|y_0\| + \|f(x_0) - f(x_k) - A_k(x_0 - x_k)\| \\ &\leq \|y_0\| + \delta \|x_0 - x_k\| \leq \|y_0\| + \frac{1-\alpha^k}{1-\alpha}\kappa'\delta\|y_0\| \\ &\leq \|y_0\| + \frac{1-\alpha^k}{1-\alpha}\alpha\|y_0\| = \frac{1-\alpha^{k+1}}{1-\alpha}\|y_0\| < b. \end{aligned}$$

If $r_k \in G_{A_k}(x_k)$ then we take $x_{k+1} = x_k$. If not, by (A2),

$$\mathrm{dist}\left(x_k, G_{A_k}^{-1}(r_k)\right) \leq \kappa\, \mathrm{dist}\left(r_k, G_{A_k}(x_k)\right) < \kappa'\, \mathrm{dist}\left(r_k, G_{A_k}(x_k)\right).$$

Then there exists a point $x_{k+1} \in G_{A_k}^{-1}(r_k)$ such that

$$\|x_{k+1} - x_k\| < \kappa'\, \mathrm{dist}\left(r_k, G_{A_k}(x_k)\right).$$

Due to (c), we get

$$G_{A_k}(x_k) = f(x_0) + A_k(x_k - x_0) + F(x_k) \ni f(x_0) + A_k(x_k - x_0) - f(x_{k-1}) - A_{k-1}(x_k - x_{k-1}).$$

Using (A3) with $x = x_k$ and then (b) and (18) we have

$$\begin{aligned} \|x_{k+1} - x_k\| &\leq \kappa' \|r_k - [f(x_0) - f(x_{k-1}) + A_k(x_k - x_0) - A_{k-1}(x_k - x_{k-1})]\| \\ &= \kappa' \|f(x_k) - f(x_{k-1}) - A_{k-1}(x_k - x_{k-1})\| \\ &\leq \kappa' \omega(\|x_k - x_{k-1}\|)\|x_k - x_{k-1}\| \;=\; \kappa'\delta\gamma_k\|x_k - x_{k-1}\| \qquad (19) \\ &\leq \alpha^k \gamma_0 \ldots \gamma_k \kappa' \|y_0\| < \alpha^k (1-\alpha) a. \qquad (20) \end{aligned}$$

Hence, condition (b) is satisfied for $k+1$ and γ_{k+1} is well-defined. By the choice of x_{k+1} we have

$$r_k \in G_{A_k}(x_{k+1}) = f(x_0) + A_k(x_{k+1} - x_0) + F(x_{k+1}),$$

hence, after rearranging, condition (c) holds for $k+1$. To finish the induction step, use (a) to obtain

$$\|x_{k+1}-x_0\| \leq \|x_{k+1}-x_k\|+\|x_k-x_0\| \leq \alpha^k\kappa'\|y_0\|+\frac{1-\alpha^k}{1-\alpha}\kappa'\|y_0\| = \frac{1-\alpha^{k+1}}{1-\alpha}\kappa'\|y_0\|.$$

Now we shall prove that the sequence $\{x_k\}$ identified in the preceding lines is convergent. By (b) (with γ_i replaced with 1), applied for $k := m, n \in \mathbb{N}$ with $m < n$, we have

$$\|x_n - x_m\| \leq \alpha^m \frac{1-\alpha^{n-m}}{1-\alpha}\kappa'\|y_0\|,$$

hence $\{x_k\}$ is a Cauchy sequence. Let $\bar{x} = \lim_{k\to\infty} x_k$. Then by (a),

$$\|\bar{x}-x_0\| \leq \frac{\kappa'}{1-\alpha}\|y_0\| < a,$$

that is, $\bar{x} \in \operatorname{int} I\!B_a(x_0)$. Using (b), for any $k \in I\!N_0$, and the second inequality in (18), we have

$$\begin{aligned}
\|x_k-\bar{x}\| &= \lim_{m\to\infty}\|x_k - x_{k+m}\| \leq \lim_{m\to\infty}\sum_{i=k}^{k-1+m}\|x_i - x_{i+1}\| \\
&\leq \lim_{m\to\infty}\sum_{i=k}^{k-1+m}\alpha^i\gamma_1\ldots\gamma_i\kappa'\|y_0\| \leq \alpha^k\gamma_1\ldots\gamma_k\lim_{m\to\infty}\sum_{i=k}^{k-1+m}\alpha^{i-k}\kappa'\|y_0\| \\
&\leq \alpha^k\gamma_1\ldots\gamma_k\frac{\kappa'\|y_0\|}{1-\alpha} \leq \alpha^k\gamma_1\ldots\gamma_k a =: \varepsilon_k. \qquad (21)
\end{aligned}$$

By the definition of ε_k we get

$$\varepsilon_{k+1} = \alpha\gamma_{k+1}\varepsilon_k.$$

Since $\gamma_{k+1} \leq 1$ we obtain linear convergence in (17). If $\lim_{\xi\to 0}\omega(\xi) = 0$, then $\gamma_k \to 0$ and we have r-superlinear convergence.

Finally, if there exists a constant L such that $\omega(\xi) \leq \min\{\delta, L\xi\}$ for each $\xi \in [0,a]$, then for each $k \in I\!N$ condition (b) implies that $\xi := \|x_{k+1}-x_k\| < a$; hence

$$\gamma_{k+1} \leq \min\{1, L\|x_{k+1}-x_k\|/\delta\} \leq \|x_{k+1}-x_k\|L/\delta \leq (\varepsilon_{k+1}+\varepsilon_k)L/\delta.$$

Fix any $C > \alpha L/\delta$. Since the sequence $\{\varepsilon_k\}$ is strictly decreasing and converges to zero, we obtain

$$\varepsilon_{k+1} \leq \frac{\alpha L}{\delta}(\varepsilon_k+\varepsilon_{k+1})\varepsilon_k < C\varepsilon_k^2 \quad \text{for all sufficiently large } k \in I\!N.$$

This implies r-quadratic convergence.

To show that $\bar{x}$ solves (8), let $y_k := f(x_k) - f(x_{k-1}) - A_{k-1}(x_k - x_{k-1})$ for $k \in I\!N$. From (c) we have $y_k \in f(x_k) + F(x_k)$. Using (A3) with $x = x_k$ and then using (b) we obtain that

$$\|y_k\| = \|f(x_k)-f(x_{k-1})-A_{k-1}(x_k-x_{k-1})\| \leq \delta\|x_k-x_{k-1}\| \leq \delta\alpha^{k-1}\kappa'\|y_0\| \leq \alpha^k\|y_0\|. \quad (22)$$

Thus $(x_k, y_k) \to (\bar{x}, 0)$ as $k \to \infty$. Since f is continuous and F has closed graph, we obtain $0 \in f(\bar{x}) + F(\bar{x})$. The second inequality in (17) follows from (22).

In the case of strong metric regularity of G_A the way x_{k+1} is constructed from x_k implies automatically that x_{k+1} is unique in $I\!B_a(x_0)$. □

Remark 2.3. Suppose that there exist $\beta \in (0,1]$ and $L > 0$ such that $\omega(\xi) \le \min\{L\xi^\beta, \delta\}$ for each $\xi \in [0,a]$. Then $\{x_k\}$ converges to $\bar{x}$ with r-rate $1+\beta$: there exists a sequence of positive numbers $\{\varepsilon_k\}$ converging to zero and $C > 0$ such that $\varepsilon_{k+1} \le C\varepsilon_k^{1+\beta}$ for all $k \in \mathbb{N}$. Indeed, for each $k \in \mathbb{N}$, (b) implies that $\xi := \|x_{k+1} - x_k\| < a$, hence

$$\gamma_{k+1} \le \frac{L}{\delta}\|x_{k+1} - x_k\|^\beta \le \frac{L}{\delta}(\varepsilon_{k+1} + \varepsilon_k)^\beta = \frac{L}{\delta}(1 + \alpha\gamma_{k+1})^\beta \varepsilon_k^\beta \le \frac{L}{\delta}(1+\alpha)^\beta \varepsilon_k^\beta.$$

Hence, taking $C := \alpha L(1+\alpha)^\beta/\delta$ we get

$$\varepsilon_{k+1} = \alpha\gamma_{k+1}\varepsilon_k \le C\varepsilon_k^{1+\beta} \quad \text{for all } k \in \mathbb{N}.$$

Remark 2.4. Theorem 2.1 follows from the strong regularity part of Theorem 2.2. Indeed, for the case of the equation condition (A1) is the same as (10). The first inequality in (11) means that the mapping G_{A_k} with $F \equiv 0$ is strongly metrically regular uniformly in k, and the second inequality is the same as (A3).

The following corollary is a somewhat simplified version of Theorem 2.2 which may be more transparent for particular cases.

Corollary 2.5. *Let a, b, κ, δ be positive reals and a point $(x_0, y_0) \in \operatorname{gph}(f+F)$ be such that condition (A1) in Theorem 2.2 holds. Let $\{A_k\}$ be a sequence of bounded linear operators from X to Y such that for every $k \in \mathbb{N}_0$ the mapping G_{A_k} defined in* (16) *is metrically regular at x_0 for y_0 with constant κ and neighborhoods $\mathbb{B}_a(x_0)$ and $\mathbb{B}_b(y_0)$, and*

$$\|f(x) - f(x') - A_k(x - x')\| \le \delta\|x - x'\| \quad \textit{for any} \quad x, x' \in \mathbb{B}_a(x_0).$$

Then for every $\alpha \in (\kappa\delta, 1)$ there exists a sequence $\{x_k\}$ satisfying (15) *with starting point x_0 which is convergent to a solution $\bar{x} \in \operatorname{int} \mathbb{B}_a(x_0)$ of (8) with r-linear rate as in* (17).

3 Some special cases

Consider first the generalized equation (8) where the function f is continuously differentiable around the starting point x_0. Then we can take $A_k = Df(x_k)$ in the iteration (15) obtaining

$$f(x_k) + Df(x_k)(x_{k+1} - x_k) + F(x_{k+1}) \ni 0. \tag{23}$$

In the following theorem we obtain q-superlinear and q-quadratic convergence of the iteration (23) by concatenating the main Theorem 2.2 with conventional convergence results from [14], Theorems 6C.1 and 6D.2.

Theorem 3.1. *Consider the generalized equation* (8), *a point $(x_0, y_0) \in \operatorname{gph}(f+F)$ and positive reals κ, δ, a and b such that condition (A1) in Theorem 2.2 is satisfied. Suppose that the function f is continuously differentiable in an open set containing $\mathbb{B}_a(x_0)$, for every $z \in \mathbb{B}_a(x_0)$ the mapping*

$$x \mapsto G_z(x) := f(x_0) + Df(z)(x - x_0) + F(x)$$

is metrically regular at x_0 for y_0 with constant κ and neighborhoods $\mathbb{B}_a(x_0)$ and $\mathbb{B}_b(y_0)$, and also

$$\|f(x) - f(x') - Df(x)(x - x')\| \le \delta \|x - x'\| \quad \text{for all } x, x' \in \mathbb{B}_a(x_0).$$

Then there exists a sequence $\{x_k\}$ which satisfies the iteration (23) *with starting point x_0 and converges q-superlinearly to a solution $\bar{x}$ of* (8) *in* $\operatorname{int} \mathbb{B}_a(x_0)$. *If the derivative mapping Df is Lipschitz continuous in $\mathbb{B}_a(x_0)$, then the sequence $\{x_k\}$ converges q-quadratically to $\bar{x}$.*

Proof. Clearly, for any sequence $\{x_k\}$ in $\mathbb{B}_a(x_0)$ and for each $k \in \mathbb{N}_0$ the mapping $A_k := Df(x_k)$ satisfies (A2) and (A3) of Theorem 2.2 with $\omega(\xi) := \delta$, $\xi \ge 0$. From condition (A1) there exists $\alpha \in (\kappa\delta, 1)$ such that

$$\|y_0\| < (1 - \alpha)b. \tag{24}$$

Hence we can apply Theorem 2.2, which yields the existence of a sequence $\{x_k\}$ satisfying (23) and converging to a solution $\bar{x} \in \operatorname{int} \mathbb{B}_a(x_0)$ of (8); furthermore

$$\|\bar{x} - x_0\| \le \frac{\alpha}{\delta(1-\alpha)} \|y_0\|.$$

Hence, for $v_0 := f(\bar{x}) - f(x_0) - Df(\bar{x})(\bar{x} - x_0)$ we have

$$\begin{aligned} \|y_0 + v_0\| &= \|y_0 + f(\bar{x}) - f(x_0) - Df(\bar{x})(\bar{x} - x_0)\| \le \|y_0\| + \delta\|\bar{x} - x_0\| \\ &\le \|y_0\| + \frac{\alpha}{1-\alpha}\|y_0\| = \frac{\|y_0\|}{1-\alpha} < b, \end{aligned}$$

where we use (24). Clearly, the mapping

$$x \mapsto G'(x) := f(\bar{x}) + Df(\bar{x})(x - \bar{x}) + F(x) = v_0 + G_{\bar{x}}(x)$$

is metrically regular at x_0 for $y_0 + v_0$ with constant κ and neighborhoods $\mathbb{B}_a(x_0)$ and $\mathbb{B}_b(y_0 + v_0)$. Let $r, s > 0$ be so small that

$$\mathbb{B}_r(\bar{x}) \subset \mathbb{B}_a(x_0) \quad \text{and} \quad \mathbb{B}_s(0) \subset \mathbb{B}_b(y_0 + v_0).$$

Then since $0 \in G'(\bar{x})$, the mapping G' is metrically regular at $\bar{x}$ for 0 with constant κ and neighborhoods $\mathbb{B}_r(\bar{x})$ and $\mathbb{B}_s(0)$. Hence we can apply Theorems 6C.1, resp. 6D.2, in [14], according to which there exists a neighborhood O of $\bar{x}$ such that for any starting point in O there exists a sequence $\{x'_k\}$ which is q-superlinearly, resp. q-quadratically, convergent to $\bar{x}$. But for some k sufficiently large the iterate x_k of the initial sequence will be in O and hence it can be taken as a starting point of a sequence $\{x'_k\}$ which converges q-superlinearly, resp. q-quadratically, to $\bar{x}$. ◻

In the theorem coming next we utilize an auxiliary result which follows from Proof I, with some obvious adjustments, of the extended Lyusternik-Graves theorem given in [14, Theorem 5E.1].

Lemma 3.2. *Consider a mapping $F : X \rightrightarrows Y$, a point $(x_0, y_0) \in \operatorname{gph} F$ and a function $g : X \to Y$. Suppose that there are $a' > 0$, $b' > 0$, $\kappa' \geq 0$, and $\mu \geq 0$ such that F is metrically regular at x_0 for y_0 with constant κ' and neighborhoods $I\!B_{a'}(x_0)$ and $I\!B_{b'}(y_0)$, the function g is Lipschitz continuous on $I\!B_{a'}(x_0)$ with constant μ, and $\kappa'\mu < 1$. Then for any positive constants a and b such that*

$$\frac{1}{1-\kappa'\mu}\left[(1+\kappa'\mu)a + \kappa' b\right] + a < a', \quad b + \mu\left(\frac{1}{1-\kappa'\mu}\left[(1+\kappa'\mu)a + \kappa' b\right] + a\right) < b', \tag{25}$$

the mapping $g+F$ is metrically regular at x_0 for $y_0+g(x_0)$ with any constant $\kappa > \kappa'/(1-\kappa'\mu)$ and neighborhoods $I\!B_a(x_0)$ and $I\!B_b(y_0 + g(x_0))$.

Theorem 3.3. *Let the numbers $a > 0$, $b > 0$, $\kappa > 0$ and $\delta > 0$ and the points $x_0 \in X$, $y_0 \in f(x_0) + F(x_0)$ be such that (A1) is fulfilled. Let the numbers a', b', κ' be such that:*

$$0 < \kappa' < \frac{\kappa}{1+\kappa\delta}, \quad a' > 2a(1+\kappa\delta) + \kappa b, \quad b' > (2a\delta + b)(1+\kappa\delta). \tag{26}$$

Let f be Fréchet differentiable in an open set containing $I\!B_a(x_0)$, let $\mathcal{T} \subset \mathcal{L}(X, Y)$, and let $A_k : X^k \to \mathcal{T}$ be any sequence with $\sup_{A\in\mathcal{T}} \|A - A_0(x_0)\| \leq \delta$. Assume that

(A2') the mapping $x \mapsto G(x) := f(x_0) + A_0(x_0)(x - x_0) + F(x)$ is metrically regular with constant κ' and neighborhoods $I\!B_{a'}(x_0)$ and $I\!B_{b'}(y_0)$;

(A3') $\|A - Df(x)\| \leq \delta$ whenever $A \in \mathcal{T}$ and $x \in I\!B_a(x_0)$.

Then the first claim in Theorem 2.2 holds.

Proof. We shall prove that conditions (A2) and (A3) in Theorem 2.2 are satisfied.

To check (A2), pick any $A \in \mathcal{T}$ and let G_A be the mapping from Theorem 2.2 (with $A_k := A$). Define $g(x) := (A - A_0)(x - x_0)$, $x \in X$, so that $G_A = G + g$. Then g is Lipschitz continuous with constant δ and we can apply Lemma 3.2 with $\mu := \delta$, which implies (A2).

It remains to check (A3). Let $\omega(\xi) := \delta$ for each $\xi \geq 0$. Pick arbitrary points $x_0, x_1, \ldots, x_k$ in $I\!B_a(x_0)$ and set $A_k := A_k(x_0, \ldots, x_k)$. Finally, fix any $x \in I\!B_a(x_0)$. By the mean value theorem there is $z \in I\!B_a(x_0)$ such that $f(x) - f(x_k) - Df(z)(x - x_k) = 0$. Hence

$$\|f(x) - f(x_k) - A_k(x - x_k)\| = \|Df(z)(x - x_k) - A_k(x - x_k)\| \leq \delta\|x - x_k\|.$$

This proves (A3) and therefore the theorem. □

Next, we state and prove a theorem regarding convergence of the Newton's method applied to a generalized equation, which is close to the original statement of Kantorovich. The result is somewhat parallel to [13, Theorem 2] but on different assumptions.

Theorem 3.4. *Let the positive scalars L, κ, a, b and the points $x_0 \in X$, $y_0 \in f(x_0) + F(x_0)$ be such that the function f is differentiable in an open neighborhood of the ball $I\!B_a(x_0)$ and its derivative Df is Lipschitz continuous on $I\!B_a(x_0)$ with Lipschitz constant L and also the mapping*

$$x \mapsto G(x) := f(x_0) + Df(x_0)(x - x_0) + F(x) \tag{27}$$

is metrically regular at x_0 for y_0 with constant κ and neighborhoods $\mathbb{B}_a(x_0)$ and $\mathbb{B}_b(y_0)$. Furthermore, let $\kappa' > \kappa$ and assume that for $\eta := \kappa'\|y_0\|$ we have

$$h := \kappa' L\eta < \frac{1}{2}, \quad \bar{t} := \frac{1}{\kappa' L}(1-\sqrt{1-2h}) \leq a \quad \text{and} \quad \|y_0\| + L\bar{t}^2 \leq b. \tag{28}$$

Then there is a sequence $\{x_k\}$ generated by the iteration (23) with initial point x_0 which stays in $\mathbb{B}_a(x_0)$ and converges to a solution $\bar{x}$ of the generalized equation (8); moreover, the rate of the convergence is

$$\|x_k - \bar{x}\| \leq \frac{2\sqrt{1-2h}\,\Theta^{2^k}}{\kappa' L(1-\Theta^{2^k})}, \qquad \text{for } k = 1, 2, \ldots, \tag{29}$$

where

$$\Theta := \frac{1-\sqrt{1-2h}}{1+\sqrt{1-2h}}.$$

If the mapping G is not only metrically regular but also strongly metrically regular with the same constant and neighborhoods, then there is no other sequence $\{x_k\}$ generated by the method (23) starting from x_0 which stays in $\mathbb{B}_a(x_0)$.

Proof. In the sequel we will utilize the following inequality for $u, v \in \mathbb{B}_a(x_0)$:

$$\begin{aligned} \|f(u) - f(v) - Df(v)(u-v)\| &= \left\|\int_0^1 [Df(v+s(u-v)) - Df(v)](u-v)\,\mathrm{d}s\right\| \\ &\leq L\|u-v\|^2 \int_0^1 s\,\mathrm{d}s = \frac{L}{2}\|u-v\|^2. \end{aligned}$$

We apply a modification of the majorization technique from [17]. Consider a sequence of reals t_k satisfying

$$t_0 = 0, \quad t_{k+1} = s(t_k), \quad k = 0, 1, \ldots,$$

where

$$s(t) = t - (p'(t))^{-1}p(t), \quad p(t) = \frac{\kappa' L}{2}t^2 - t + \eta.$$

It is known from [17] that the sequence $\{t_k\}$ is strictly increasing, convergent to $\bar{t}$, and also

$$t_{k+1} - t_k = \frac{\kappa' L(t_k - t_{k-1})^2}{2(1-\kappa' L t_k)}, \qquad k = 0, 1, \ldots. \tag{30}$$

Furthermore,

$$\bar{t} - t_k \leq \frac{2\sqrt{1-2h}\,\Theta^{2^k}}{\kappa' L(1-\Theta^{2^k})}, \qquad \text{for } k = 0, 1, \ldots. \tag{31}$$

We will show, by induction, that there is a sequence $\{x_k\}$ in $\mathbb{B}_a(x_0)$ fulfilling (23) with the starting point x_0 which satisfies

$$\|x_{k+1} - x_k\| \leq t_{k+1} - t_k, \quad k = 0, 1, \ldots. \tag{32}$$

This implies that $\{x_k\}$ is a Cauchy sequence, hence convergent to some $\bar{x}$, which, by passing to the limit in (23), is a solution of the problem at hand. Combining (31), (30) and (32) we obtain (29).

Let $k = 0$. If $y_0 = 0$ then we take $x_1 = x_0$. If not, since $0 \in I\!B_b(y_0)$ and $y_0 \in G(x_0)$, from the metric regularity of the mapping G in (27) we obtain

$$\operatorname{dist}(x_0, G^{-1}(0)) \le \kappa \|y_0\| < \kappa' \|y_0\|,$$

hence there exists $x_1 \in G^{-1}(0)$ such that

$$\|x_1 - x_0\| < \kappa' \|y_0\| = \eta = t_1 - t_0.$$

Suppose that for some $k \in I\!N$ we have already found points $x_0, x_1, \ldots, x_k$ in $I\!B_a(x_0)$ generated by (23) such that

$$\|x_j - x_{j-1}\| \le t_j - t_{j-1} \quad \text{for each} \quad j = 1, \ldots, k.$$

Without loss of generality, let $x_k \neq x_0$; otherwise there is nothing to prove. We have

$$\|x_k - x_0\| \le \sum_{j=1}^{k} \|x_j - x_{j-1}\| \le \sum_{j=1}^{k} (t_j - t_{j-1}) = t_k - t_0 = t_k < \bar{t} \le a.$$

Furthermore, for every $x \in I\!B_{\bar{t}-t_k}(x_k) \subset I\!B_{\bar{t}}(x_0)$, we obtain

$$\begin{aligned} &\|f(x_0) + Df(x_0)(x - x_0) - f(x_k) - Df(x_k)(x - x_k)\| \\ &\quad \le \|f(x) - f(x_0) - Df(x_0)(x - x_0)\| + \|f(x) - f(x_k) - Df(x_k)(x - x_k)\| \\ &\quad \le \frac{L}{2}\left(\|x - x_0\|^2 + \|x - x_k\|^2\right) < L\bar{t}^2 \le b - \|y_0\|, \end{aligned}$$

in particular, we have $f(x_0) + Df(x_0)(x - x_0) - f(x_k) - Df(x_k)(x - x_k) \in I\!B_b(y_0)$. Moreover,

$$r \;:=\; \frac{\frac{1}{2}\kappa' L \|x_k - x_{k-1}\|^2}{1 - \kappa' L \|x_k - x_0\|} \le \frac{\kappa' L (t_k - t_{k-1})^2}{2(1 - \kappa' L t_k)} = t_{k+1} - t_k.$$

Since $x_k \in I\!B_a(x_0)$ is generated by (23) from x_{k-1}, we get

$$f(x_0) + Df(x_0)(x_k - x_0) - f(x_{k-1}) - Df(x_{k-1})(x_k - x_{k-1}) \in G(x_k). \tag{33}$$

Now consider the set-valued mapping

$$X \ni x \mapsto \Phi_k(x) := G^{-1}(f(x_0) + Df(x_0)(x - x_0) - f(x_k) - Df(x_k)(x - x_k)) \subset X.$$

If $x_k = x_{k-1}$ then take $x_{k+1} = x_k$. Suppose that $x_k \neq x_{k-1}$. From (33) we obtain

$$\begin{aligned} \operatorname{dist}(x_k, \Phi_k(x_k)) &= \operatorname{dist}(x_k, G^{-1}\left(f(x_0) + Df(x_0)(x_k - x_0) - f(x_k)\right) \\ &\le \kappa \operatorname{dist}(f(x_0) + Df(x_0)(x_k - x_0) - f(x_k), G(x_k)) \\ &\le \kappa \|f(x_k) - f(x_{k-1}) - Df(x_{k-1})(x_k - x_{k-1})\| \\ &\le \frac{1}{2}\kappa L \|x_k - x_{k-1}\|^2 < \frac{\frac{1}{2}\kappa' L \|x_k - x_{k-1}\|^2}{1 - \kappa' L \|x_k - x_0\|}(1 - \kappa' L \|x_k - x_0\|) \\ &= r(1 - \kappa' L \|x_k - x_0\|). \end{aligned}$$

Let $u, v \in \mathbb{B}_{\bar{t}-t_k}(x_k)$ and let $z \in \Phi_k(u) \cap \mathbb{B}_{\bar{t}-t_k}(x_k)$. Then

$$f(x_0) + Df(x_0)(u - x_0) - f(x_k) - Df(x_k)(u - x_k) \in G(z).$$

Hence,

$$\begin{aligned}
\operatorname{dist}(z, \Phi_k(v)) &= \operatorname{dist}(z, G^{-1}(f(x_0) + Df(x_0)(v - x_0) - f(x_k) - Df(x_k)(v - x_k)) \\
&\le \kappa \operatorname{dist}(f(x_0) + Df(x_0)(v - x_0) - f(x_k) - Df(x_k)(v - x_k), G(z)) \\
&\le \kappa \| f(x_0) + Df(x_0)(v - x_0) - f(x_k) - Df(x_k)(v - x_k) \\
&\qquad -(f(x_0) + Df(x_0)(u - x_0) - f(x_k) - Df(x_k)(u - x_k)) \| \\
&\le \kappa \| Df(x_0) - Df(x_k) \| \| u - v \| \le (\kappa' L \| x_k - x_0 \|) \| u - v \|.
\end{aligned}$$

Since $\mathbb{B}_r(x_k) \subset \mathbb{B}_{\bar{t}-t_k}(x_k)$, by applying the contraction mapping theorem [14, Theorem 5E.2] we obtain that there exists a fixed point $x_{k+1} \in \mathbb{B}_r(x_k)$ of Φ_k. Hence

$$x_{k+1} \in G^{-1}\left(f(x_0) + Df(x_0)(x_{k+1} - x_0) - f(x_k) - Df(x_k)(x_{k+1} - x_k) \right),$$

that is, x_{k+1} is a Newton iterate from x_k according to (23). Furthermore,

$$\| x_{k+1} - x_k \| \le r \le t_{k+1} - t_k.$$

Then

$$\| x_{k+1} - x_0 \| \le \sum_{j=1}^{k+1} \| x_j - x_{j-1} \| \le \sum_{j=1}^{k+1} (t_j - t_{j-1}) = t_{k+1} - t_0 = t_{k+1} < \bar{t} \ \le \ a.$$

The induction step is complete and so is the proof. □

At the end of this section we add some comments on the results presented in this paper and give some examples. First, we would like to reiterate that, in contrast to the conventional approach to proving convergence of Newton's method where certain conditions *at a solution* are imposed, the Kantorovich theorem utilizes conditions for *a given neighborhood of the starting point* associated with some constants, the relations among which gives the existence of a solution and convergence towards it. In the framework of the main Theorem 2.2, among the constants taken into account are the radius a of the given neighborhood of the starting point x_0, the norm of the residual $\|y_0\|$ at the starting point, the constant of metric regularity κ, and the constant δ measuring the "quality" of the approximation of the "derivative" of the function f by the operators A_k. These constants are interconnected through relations that cannot be removed even in the particular cases of finite dimensional smooth problems, or nonsmooth problems where elements of the Clarke's generalized Jacobian play the role of approximations. In the smooth case the constant δ may be measured by the diameter of the set $\{\|Df(x)\| : x \in \mathbb{B}_a(x_0)\}$ or by La if Df is Lipschitz continuous with a Lipschitz constant L. In the nonsmooth case however, it is not sufficient to assume that the diameter of the generalized Jacobian around x_0 is less than δ. One may argue that for any small δ there exists a positive ε such that the generalized Jacobian has the "strict derivative property" displayed in [14, 6F.3] but in order this to work we need ε to match a. Note that if the residual $\|y_0\| = 0$ then we can always choose the constant a sufficiently small, but this may not be the case for the Kantorovich theorem. It would be quite interesting to know exactly "how far" the conventional and the Kantorovich theorems are from each other in particular for problems involving nonsmooth functions.

Next, we will present some elementary examples that illustrate the difference between the Newton method and the chord method with $A_k = A_0$ for all k, as well as the conditions for convergence appearing in the results presented.

Example 1. We start with the smooth one-dimensional example[7] to find a *nonnegative* root of $f(x) := (x-1)^2 - 4$; it is elementary to check that $\bar{x} = 3$ is the only solution. For every $x_0 > 1$ the usual Newton iteration is given by

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = \frac{x_k^2 + 3}{2(x_k - 1)}.$$

This iteration is convergent quadratically which agrees with the theory. The chord method,

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_0)} = \frac{2x_0 x_k - x_k^2 + 3}{2(x_0 - 1)},$$

converges linearly if there is a constant $c < 1$ and a natural number N such that

$$\frac{|x_{k+1} - 3|}{|x_k - 3|} = \frac{|2x_0 - x_k - 3|}{2|x_0 - 1|} \le c$$

for every $k \ge N$, but it may not be convergent for x_0 not close enough to 3. For example take $x_0 = 1 + \frac{2}{\sqrt{5}}$. Then the method oscillates between the points $1 + \frac{2}{\sqrt{5}}$ and $1 + \frac{6}{\sqrt{5}}$. The method converges q-superlinearly whenever

$$\lim_{k\to\infty} \frac{|x_{k+1} - 3|}{|x_k - 3|} = \lim_{k\to\infty} \frac{|2x_0 - x_k - 3|}{2|x_0 - 1|} = 0;$$

but this holds only for $x_0 = 3$. Hence, even in the case when there is convergence, it is not q-superlinear.

Let us check the assumptions of Theorem 2.2 with $\omega \equiv \delta$. Given x_0 and $a > 0$ we can calculate how large κ and δ have to be such that conditions (A2) and (A3) are fulfilled. Let us focus on the case $x_0 > 1$. For (A2) to hold we have to assume $a < x_0 - 1$. Then on $I\!B_a(x_0)$ we have that f' is positive and increasing. Hence (A2) and (A3) are satisfied for $\kappa = 1/f'(x_0 - a) = 1/(2(x_0 - a - 1)$ and $\delta = f'(x_0 + a) - f'(x_0 - a) = 4a$. For fixed x_0 let us find a such that (A1) holds as well, i.e.,

$$\|y_0\| < (1 - \kappa\delta)\frac{a}{\kappa} = 2a(x_0 - 3a - 1). \tag{34}$$

The right hand side is maximal for $a = \frac{x_0 - 1}{6}$. Expressing both sides of this inequality in terms of x_0, we obtain that if $x_0 \in (1 + 2\sqrt{6/7}, 1 + 2\sqrt{6/5})$ then we have convergence.

[7] Note that this problem can be written as a generalized equation.

The following example from [26], see also [25], example BE.1, shows lack of convergence of the nonsmooth Newton method if the function is not semismooth at the solution. But it is also an example which illustrates Corollary 2.5.

Example 2. Consider intervals $I(n) = [n^{-1}, (n-1)^{-1}] \subset \mathbb{R}$ and define $c(n) = \frac{1}{2}(n^{-1} + (n-1)^{-1})$ for $n \geq 2$. Let g_n be the linear function through the points $((n-1)^{-1}, (n-1)^{-1})$ and $(-c(n), 0)$, and h_n be the linear function through the points (n^{-1}, n^{-1}) and $(c(2n), 0)$. Then

$$g_n(x) = \frac{2n}{4n-1}x + \frac{2n-1}{(n-1)(4n-1)} \quad \text{and} \quad h_n(x) = \frac{4(2n-1)}{4n-3}x - \frac{4n-1}{n(4n-3)}.$$

Now define $f(x) = \min\{g_n(x), h_n(x)\}$ for $x \in I(n)$, $f(0) = 0$ and $f(x) = -f(-x)$ for $x < 0$. Then the equation $f(x) = 0$ has the single solution $\bar{x} = 0$ and we have that $\bar{\partial} f(0) = [\frac{1}{2}, 2]$. If we try to apply Corollary 2.5 for a neighborhood that contains $\bar{x} = 0$ we have to choose $\delta \geq \frac{3}{2}$ and $\kappa \geq 2$; but then $\kappa\delta > 1$. In this case for any starting point $x_0 \neq 0$ the Newton iteration does not converge, as shown in [26].

A similar example follows to which Corollary 2.5 can be applied.

Example 3. Define

$$g(x) := \begin{cases} 2 \text{ if } x \in \cup_{n\in\mathbb{Z}}[2^{2n-1}, 2^{2n}) \\ 3 \text{ if } x \in \cup_{n\in\mathbb{Z}}[2^{2n}, 2^{2n+1}) \end{cases}.$$

Let $f(x) := \int_0^x g(t)\mathrm{d}t$ for $x \geq 0$ and $f(x) = -f(-x)$ for $x < 0$. The function f is well defined on $\mathbb{R}$ with a unique root at $\bar{x} = 0$. For any starting point x_0 the assumptions for Corollary 2.5 are then fulfilled with $\kappa = \frac{1}{2}$ and $\delta = 1$ and each $a > 0$. Both the Newton and the chord method converge linearly.

4 Nonsmooth inequalities

Suppose that K is a nonempty subset of Y and let $F(x) := K$ for each $x \in X$. Then the generalized equation (8) reads as

$$f(x) + K \ni 0. \tag{35}$$

When $f : \mathbb{R}^n \to \mathbb{R}^m$ and $K := \mathbb{R}^m_+$ then the above inclusion corresponds to a system of m nonlinear (possibly nonsmooth) inequalities: find $x \in \mathbb{R}^n$ such that

$$f_1(x) \leq 0, \quad f_2(x) \leq 0, \quad \ldots, \quad f_m(x) \leq 0.$$

Kantorovich-type theorems for exact Newton's method for solving (35) with K being a closed convex cone and f being smooth can be found in [4, Chapter 2.6] and [31]. An inexact Newton's method is treated in a similar way in [16]. The paper [28] deals with a generalized equation of the form

$$g(x) + h(x) + K \ni 0, \tag{36}$$

where $g : X \to Y$ is a smooth function having a Lipschitz derivative on a neighborhood $O \subset X$ of a (starting) point $x_0 \in X$ and the function $h : X \to Y$ is Lipschitz continuous on O. The algorithm proposed therein reads as: given $x_k \in X$ find x_{k+1} satisfying

$$g(x_k) + h(x_k) + g'(x_k)(x_{k+1} - x_k) + K \ni 0. \tag{37}$$

Key assumptions are, similarly to [31, 4, 16], that $T := g'(x_0)(\cdot) + K$ maps X onto Y and

$$\|T^{-1}\|^{-} := \sup_{\|y\|\leq 1} \inf_{x \in T^{-1}(y)} \|x\| \leq b$$

for a sufficiently small number $b > 0$. Then Open Mapping Theorem [5, Theorem 2.2.1] (see also [14, Exercise 5C.4]) implies that T is metrically regular at zero for zero with any constant $\kappa > b$ and neighborhoods X and Y. Moreover, the Lipschitz constants of g' and h are assumed to be small compared to b. Clearly, (37) corresponds to our iteration scheme with $f := g + h$ and $A_k := g'(x_k)$, and, since A_k does not take into account the non-smooth part, it is expected to be slower in general (or not even applicable) as we will show on two toy examples below.

Consider a sequence $\{A_k\}$ in $\mathcal{L}(X, Y)$ and a starting point $x_0 \in X$. Given $k \in I\!N_0$, $x_k \in X$, and A_k, let

$$\Omega_k := \{u \in X \mid g(x_k) + h(x_k) + A_k(u - x_k) + K \ni 0\}.$$

The next iterate x_{k+1} generated by (15), which is sure to exist under the metric regularity assumption in Theorem 2.2, is any point lying in Ω_k such that

$$\|x_{k+1} - x_k\| \leq \kappa' \operatorname{dist}(-g(x_k) - h(x_k), K),$$

where $\kappa' > \kappa$ satisfies (18) and the right-hand side of the above inequality corresponds to a residual at the step k. To sum up, for the already computed x_k, the next iterate x_{k+1} can be found as a solution of the problem:

$$\text{minimize} \quad \varphi_k(x) \quad \text{subject to} \quad x \in \Omega_k,$$

where $\varphi_k : X \to [0, \infty)$ is a suitably chosen function. In [28], $\varphi_k = \|\cdot - x_k\|_2$ is used. In the following examples we solve the linearized problem in MATLAB using either function *fmincon* for $\varphi_k = \|\cdot - x_k\|_2^2$ or *quadprog* for $\varphi_k(x) := \frac{1}{2}x^T x - x_k^T x$. We will compare the following three versions of (15) for solving (36) with different choices of A_k at the step $k \in I\!N_0$ and current iterate x_k:

(C1) $A_k := g'(x_k)$;

(C2) $A_k \in \bar{\partial}(g + h)(x_k) = g'(x_k) + \bar{\partial}h(x_k)$;

(C3) $A_k := A_0$, where A_0 is a fixed element of $\bar{\partial}(g + h)(x_0) = g'(x_0) + \bar{\partial}h(x_0)$.

Example 4.1. Consider the system from [28]:

$$\begin{aligned} x^2 + y^2 - |x - 0.5| - 1 &\leq 0, \\ x^2 + (y - 1)^2 - |x - 0.5| - 1 &\leq 0, \\ (x - 1)^2 + (y - 1)^2 - 1 &= 0. \end{aligned} \tag{38}$$

Step k	fmincon			quadprog		
	(C1)	(C2)	(C3)	(C1)	(C2)	(C3)
0	5.0×10^{-2}	5.0×10^{-2}	5.0×10^{-2}	5.0×10^{-2}	5.0×10^{-2}	5.0×10^{-2}
1	2.4×10^{-2}	2.0×10^{-3}	2.0×10^{-3}	2.5×10^{-2}	2.0×10^{-3}	2.0×10^{-3}
2	1.2×10^{-2}	2.3×10^{-6}	2.3×10^{-6}	1.3×10^{-3}	2.3×10^{-6}	2.3×10^{-6}
4	3.1×10^{-3}	1.0×10^{-8}	1.0×10^{-8}	3.1×10^{-3}	6.5×10^{-9}	6.5×10^{-9}

Table 1: $\|(x_1^*, y_1^*) - (x_k, y_k)\|_\infty$ in Example 4.1 for $(x_0, y_0) = (0.55, 0.1)$.

Step k	fmincon			quadprog		
	(C1)	(C2)	(C3)	(C1)	(C2)	(C3)
0	2.9×10^{-1}	2.9×10^{-1}	2.9×10^{-1}	2.9×10^{-1}	2.9×10^{-1}	2.9×10^{-1}
1	4.2×10^{-2}	4.2×10^{-2}	4.2×10^{-2}	4.2×10^{-2}	4.2×10^{-2}	4.2×10^{-2}
2	1.2×10^{-3}	1.2×10^{-3}	1.2×10^{-3}	1.2×10^{-3}	1.2×10^{-3}	1.2×10^{-3}
4	1.1×10^{-10}	5.2×10^{-10}	5.2×10^{-10}	7.9×10^{-13}	7.9×10^{-13}	5.2×10^{-13}
7	1.1×10^{-10}	5.2×10^{-10}	5.2×10^{-10}	1.6×10^{-16}	1.1×10^{-16}	1.1×10^{-16}

Table 2: $\|(x_2^*, y_2^*) - (x_k, y_k)\|_\infty$ in Example 4.1 for $(x_0, y_0) = (0, 0)$.

Observe that the exact solutions are given by $y = 1 \pm \sqrt{2x - x^2}$ if $0 \le x \le (11 - 6\sqrt{3})/26$ and $y = 1 - \sqrt{2x - x^2}$ when $(11 - 6\sqrt{3})/26 \le x \le 1/2$, in particular, the points $(x_1^*, y_1^*) := (0.5, 1 - \sqrt{3}/2)$ and $(x_2^*, y_2^*) = (1 - \sqrt{2}/2, 1 - \sqrt{2}/2)$ solve the problem. Then setting $g(x, y) := (x^2 + y^2 - 1, x^2 + (y-1)^2 - 1, (x-1)^2 + (y-1)^2 - 1)$, $h(x, y) := (-|x - 0.5|, -|x - 0.5|, 0)$, and $K := \mathbb{R}^2_+ \times \{0\}$ we arrive at (36). Denote

$$H(x, y) := \begin{pmatrix} 2x - \operatorname{sgn}(x - 0.5) & 2y \\ 2x - \operatorname{sgn}(x - 0.5) & 2(y-1) \\ 2(x-1) & 2(y-1) \end{pmatrix}, \quad \text{with} \quad \operatorname{sgn}(u) := \begin{cases} 1 & \text{if} \quad u > 0, \\ -1 & \text{otherwise.} \end{cases}$$

In (C2) we set $A_k := H(x_k, y_k)$ for each $k \in I\!N_0$ and in (C3) we put $A_0 := H(x_0, y_0)$.

From Table 1 we see that the convergence of (15) with the choice (C1) and the starting point $(0.55, 0.1)$ is much slower than (15) with the choice (C3). Both *quadprog* and *fmincon* are of almost the same efficiency.

From Table 2 we see that for the starting point $(0, 0)$ all the choices (C1)–(C3) provide similar accuracy but we get substantially better results when *quadprog* is used to solve the linearized problem.

Example 4.2. Consider the system

$$x^2 + y^2 - 1 \le 0 \quad \text{and} \quad -|x| - |y| + \sqrt{2} \le 0 \tag{39}$$

having four distinct solutions. Set $g(x, y) := (x^2 + y^2 - 1, 0)$, $h(x, y) := (0, -|x| - |y| + \sqrt{2})$, $K := \mathbb{R}^2_+$, and

$$H(x, y) = \begin{pmatrix} 2x & 2y \\ -\operatorname{sgn}(x) & -\operatorname{sgn}(y) \end{pmatrix}.$$

As before, in (C2) we set $A_k := H(x_k, y_k)$ for each $k \in I\!N_0$ and in (C3) we put $A_0 := H(x_0, y_0)$.

Step k	fmincon		quadprog	
	(C2)	(C3)	(C2)	(C3)
0	7.0×10^{-1}	7.0×10^{-1}	7.0×10^{-1}	7.0×10^{-1}
1	2.5×10^{-9}	2.5×10^{-9}	0	0
2	7.5×10^{-8}	7.5×10^{-8}	0	0
4	1.2×10^{-8}	1.2×10^{-8}	0	0
7	8.5×10^{-8}	8.5×10^{-8}	0	0
10	8.5×10^{-9}	3.7×10^{-9}	0	0

Table 3: $\|(-\sqrt{2}/2, -\sqrt{2}/2) - (x_k, y_k)\|_\infty$ in Example 4.2 for $(x_0, y_0) = (0, 0)$.

Step k	fmincon			quadprog		
	(C1)	(C2)	(C3)	(C1)	(C2)	(C3)
0	9.9×10^{2}	9.9×10^{2}	9.9×10^{2}	9.9×10^{2}	9.9×10^{2}	9.9×10^{2}
1	4.9×10^{2}	4.9×10^{2}	4.9×10^{2}	–	4.9×10^{2}	4.9×10^{2}
4	6.1×10^{1}	6.1×10^{1}	6.1×10^{1}	–	6.1×10^{1}	6.1×10^{1}
10	5.0×10^{-1}	6.0×10^{-1}	6.0×10^{-1}	–	5.8×10^{-1}	8.3×10^{-1}
21	7.0×10^{-1}	3.0×10^{-4}	1.5×10^{-1}	–	2.8×10^{-4}	1.4×10^{0}
40	7.0×10^{-1}	5.3×10^{-9}	1.5×10^{-1}	–	1.0×10^{-8}	1.4×10^{0}

Table 4: $\|(-\sqrt{2}/2, \sqrt{2}/2) - (x_k, y_k)\|_\infty$ in Example 4.2 for $(x_0, y_0) = (99, -999)$.

For the starting point $(0, 0)$ the method (15) with (C1) fails. The convergence for the remaining two choices (C2) and (C3) can be found in Table 3. Note that using *quadprog* we find a solution (up to a machine epsilon) after one step and the iteration using *fmincon* gives the precision 10^{-9} at most.

For the starting point $(99, -999)$ the method (15) with (C1) and (C3) does not converge - see Table 3. The only convergent scheme is (15) with (C2) (note that we start far away from the solution).

5 Numerical experiments for a model of economic equilibrium

In this section we present numerical results for a model of economic equilibrium presented in [12] and solved by using the Newton, the chord and the hybrid method with various parameter choices. A detailed description of the model is given in [12] so we shall not repeat it here.

The equilibrium problem considered is described by the variational inequality

$$0 \in g(p, m, x, \lambda, m^0, x^0) + N_C(p, m, x, \lambda), \tag{40}$$

where

$$g(p,m,x,\lambda,m^0,x^0) = \begin{pmatrix} \sum_{i=1}^r (x_i^0 - x_i) \\ \cdots \\ \lambda_i - \nabla_{m_i} u_i(m_i,x_i) \\ \cdots \\ \lambda_i p - \nabla_{x_i} u_i(m_i,x_i) \\ \cdots \\ m_i^0 - m_i + \langle p, x_i^0 - x_i \rangle \\ \cdots \end{pmatrix}$$

and N_C is the normal cone to the set

$$C = \mathbb{R}_+^n \times \mathbb{R}_+^r \times U_1 \times \cdots \times U_r \times \mathbb{R}_+^r.$$

Here r is the number of agents trading n goods, who start with initial vectors of goods x_i^0 and initial amount of money m_i^0. Further, x represents the vector of goods, p is the vector of prices, m is the vector of the amounts of money, U_i are closed subsets of $\mathbb{R}_+^n$. The functions u_i are utility functions and are given by

$$u_i(m_i,x_i) = \alpha_i \ln(m_i) + \chi_{\geq m_i^1}(m_i)\gamma_i(m_i - m_i^1)^2 + \sum_{j=1}^n \beta_{ij}\ln(x_{ij})$$

where $\gamma_i \in \mathbb{R}$, α_i, β_{ij} and m_i^1 are positive constants and $\chi_{\geq m_i^1}(m_i) = \begin{cases} 1 & m_i \geq m_i^1 \\ 0 & \text{otherwise} \end{cases}$, that is, when γ_i is different from zero then $\nabla_{m_i} u_i$, and hence g, are not differentiable.

The numerical implementation of Newton's method for this variational inequality has been done in Matlab. Each step of the method reduces to solving a linear complementarity problem (LCP). To solve these problems we used the Path-LCP solver available at [11]. For the linearization for the term involving χ we use the zero vector which is always an element of Clarke's generalized Jacobian of that function.

The computations are done for the following data (similar to [3]). We set the parameters as $n = r = 10$ (so in total we have 130 variables), $\alpha_i = \beta_{ij} = 1$ and $U_i = [0.94,\ 1.08]^n$ and use random initial endowments $m_i^0 \in [1, 1.3]$ and $x_{ij}^0 \in [0.94, 1.09]$.

First we consider at the smooth problem, that is, with $\gamma_i = 0$ for all $i = 1, 2, \ldots, 10$. We use the Newton method with starting points $p_j^s = m_i^s = x_{ij}^s = \lambda_i^s = 1$, where we update the Jacobian iteration every k steps. For $k = 1, 2, 3, 5, 100$ we get a solution with error $\varepsilon = 10^{-7}$ after $4, 5, 5, 6, 9$ iterations, respectively. Then, while the number of iterations needed increases the number of times to calculate a derivative decreases from 4 to 1. Table 5 shows the errors to the solution.

If we change the starting points to $p_j^s = m_i^s = x_{ij}^s = \lambda_i^s = 0.97$ the number of iterations needed increases to $4, 5, 7, 9, 32$. Again, the number of times we update the Jacobian decreases from 4 to 1. The errors are shown in Table 6. One can see that, as expected, the choice of the starting point becomes more important if the Jacobian is not updated after every iteration. This is even more evident if we change the starting values to $p_j^s = m_i^s = x_{ij}^s = \lambda_i^s = 0.96$, where the pure chord method without updating of the Jacobian does not converge, see Table 7.

Step	$k=1$	$k=2$	$k=3$	$k=5$	$k=100$
0	9.7×10^{-1}	9.7×10^{-1}	9.7×10^{-1}	9.7×10^{-1}	9.7×10^{-1}
1	2.0×10^{-1}	2.0×10^{-1}	2.0×10^{-1}	2.0×10^{-1}	2.0×10^{-1}
2	3.9×10^{-3}	3.5×10^{-2}	3.5×10^{-2}	3.5×10^{-2}	3.5×10^{-2}
3	1.5×10^{-6}	1.9×10^{-4}	3.3×10^{-3}	3.3×10^{-3}	3.3×10^{-3}
4	0	2.2×10^{-6}	2.0×10^{-6}	1.2×10^{-3}	1.2×10^{-3}
5	-	0	0	2.1×10^{-4}	2.1×10^{-4}
6	-	-	-	0	2.1×10^{-5}

Table 5: Absolute errors with starting values $p_j^s = m_i^s = x_{ij}^s = \lambda_i^s = 1$.

Step	$k=1$	$k=2$	$k=3$	$k=5$	$k=100$
0	1.1×10^{0}	1.1×10^{0}	1.1×10^{0}	1.1×10^{0}	1.1×10^{0}
1	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}	1.0×10^{0}
2	1.3×10^{-1}	7.6×10^{-1}	7.6×10^{-1}	7.6×10^{-1}	7.6×10^{-1}
3	1.8×10^{-3}	3.5×10^{-2}	4.2×10^{-1}	4.2×10^{-1}	4.2×10^{-1}
4	0	9.1×10^{-4}	1.7×10^{-2}	2.7×10^{-1}	2.7×10^{-1}
5	-	0	1.4×10^{-3}	1.6×10^{-1}	1.6×10^{-1}
6	-	-	1.9×10^{-4}	2.2×10^{-3}	1.0×10^{-1}

Table 6: Absolute errors with starting values $p_j^s = m_i^s = x_{ij}^s = \lambda_i^s = 0.97$.

Step	$k=1$	$k=2$	$k=3$	$k=5$	$k=100$
0	1.2×10^{0}	1.2×10^{0}	1.2×10^{0}	1.2×10^{0}	1.2×10^{0}
1	1.7×10^{0}	1.7×10^{0}	1.7×10^{0}	1.7×10^{0}	1.7×10^{0}
2	4.3×10^{-1}	1.8×10^{0}	1.8×10^{0}	1.8×10^{0}	1.8×10^{0}
3	1.6×10^{-2}	2.5×10^{-1}	1.8×10^{0}	1.8×10^{0}	1.8×10^{0}
4	1.1×10^{-5}	2.3×10^{-2}	4.4×10^{-1}	1.8×10^{0}	1.8×10^{0}
5	0	2.1×10^{-5}	2.1×10^{-1}	1.8×10^{0}	1.8×10^{0}
6	-	0	1.5×10^{-1}	4.7×10^{-1}	1.9×10^{0}

Table 7: Absolute errors with starting values $p_j^s = m_i^s = x_{ij}^s = \lambda_i^s = 0.96$.

Step	$k=1$	$k=2$	$k=3$	$k=5$	$k=100$
0	2.1×10^0	2.1×10^0	2.1×10^0	2.1×10^0	2.1×10^0
1	4.5×10^{-1}	4.5×10^{-1}	4.5×10^{-1}	4.5×10^{-1}	4.5×10^{-1}
2	6.2×10^{-2}	8.2×10^{-2}	8.2×10^{-2}	8.2×10^{-2}	8.2×10^{-2}
3	1.5×10^{-4}	6.9×10^{-4}	2.7×10^{-2}	2.7×10^{-2}	2.7×10^{-2}
4	0	9.1×10^{-6}	5.3×10^{-5}	1.3×10^{-2}	1.3×10^{-2}
5	-	0	5.9×10^{-7}	3.7×10^{-3}	3.7×10^{-3}
6	-	-	0	3.3×10^{-6}	1.1×10^{-3}

Table 8: Absolute errors with parameters $m_i^1 = 0.8$ and $\gamma_i = 0.5$.

Step	$k=1$	$k=2$	$k=3$	$k=5$	$k=100$
0	4.1×10^0	4.1×10^0	4.1×10^0	4.1×10^0	4.1×10^0
1	1.5×10^0	1.5×10^0	1.5×10^0	1.5×10^0	1.5×10^0
2	1.2×10^0	2.8×10^{-1}	2.8×10^{-1}	2.8×10^{-1}	2.8×10^{-1}
3	1.3×10^{-2}	3.0×10^{-2}	2.7×10^{-1}	2.7×10^{-1}	2.7×10^{-1}
4	1.1×10^{-5}	5.3×10^{-3}	2.3×10^{-3}	1.4×10^{-1}	1.4×10^{-1}
5	0	0	4.2×10^{-5}	6.9×10^{-2}	6.9×10^{-2}
6	-	-	1.5×10^{-6}	3.8×10^{-4}	8.0×10^{-2}

Table 9: Absolute errors with parameters $m_i^1 = 0.8$ and $\gamma_i = 1$.

Consider now the nonsmooth problem for various values of γ_i and m_i^1. The starting point for the iteration is always $p_j^s = m_i^s = x_{ij}^s = \lambda_i^s = 1$. The results for $m_i^1 = 0.8$ and $\gamma_i = 0.5$ are given in Table 8.

If we increase γ_i to 1 the convergence speed in general decreases; the results are in Table 9.

For negative values of γ_i the model becomes quite unstable. For example if we set $\gamma_i = -0.7$ then for $k = 1$ the method converges after 23 iterations while for $k = 2$ we get a different solution after only 13 iterations and for $k = 3$ we get yet another different solution after 8 iterations. The absolute differences to the solution of the first Newton method are given in Table 10.

Step	$k=1$	$k=2$	$k=3$	$k=5$	$k=100$
0	1.2×10^0	1.2×10^0	1.2×10^0	1.2×10^0	1.2×10^0
1	8.4×10^{-1}	8.4×10^{-1}	8.4×10^{-1}	8.4×10^{-1}	8.4×10^{-1}
2	7.5×10^{-1}	8.0×10^{-1}	8.0×10^{-1}	8.0×10^{-1}	8.0×10^{-1}
3	1.2×10^0	7.6×10^{-1}	7.8×10^{-1}	7.8×10^{-1}	7.8×10^{-1}
4	8.6×10^{-1}	8.5×10^{-1}	8.1×10^{-1}	7.7×10^{-1}	7.7×10^{-1}
8	8.5×10^{-1}	9.1×10^{-1}	1.2×10^0	1.2×10^0	7.6×10^{-1}
13	5.8×10^{-1}	8.6×10^{-1}	1.2×10^0	1.2×10^0	8.2×10^{-1}
23	0	8.6×10^{-1}	1.2×10^0	1.2×10^0	1.2×10^{-1}

Table 10: Absolute errors with parameters $m_i^1 = 0.8$ and $\gamma_i = -0.7$.

References

[1] S. Adly, R. Cibulka, H. Van Ngai, Newton's method for solving inclusions using set-valued approximations, *SIAM J. Optim.* **25** (2015) 159–184.

[2] S. Adly, H. Van Ngai, Nguyen, Van Vu, Newton's method for solving generalized equations: Kantorovich's and Smale's approaches, J. Math. Anal. Appl. **439** (2016) 396–418.

[3] F. J. Aragón Artacho, A. Belyakov, A. L. Dontchev, M. Lopez, Local convergence of quasi-Newton methods under metric regularity, *Comput. Optim. Appl.* **58** (2014) 225–247.

[4] I. K. Argyros, *Convergence and applications of Newton-type iterations*, Springer, 2008.

[5] J.-P. Aubin, H. Frankowska, *Set-valued analysis,* Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, 1990.

[6] R. G. Bartle, Newton's method in Banach spaces, *Proc. Amer. Math. Soc.* **6** (1955) 827–831.

[7] S. C. Billups, Algorithms for complementarity problems and generalized equations, PhD thesis, Technical Report 95-14, Computer Sciences Department, University of Wisconsin, Madison, 1995.

[8] K. Butts, A. Dontchev, M. Huang, I. Kolmanovsky, A perturbed chord (Newton-Kantorovich) method for constrained nonlinear model predictive control, Proceedings of NOLCOS 2016, IFAC-PapersOnLine 49-18 (2016) 253–258.

[9] P. G. Ciarlet, C. Mardare, On the Newton-Kantorovich theorem. *Anal. Appl. (Singap.)* **10** (2012) 249–269.

[10] S. P. Dirkse, Robust solution of mixed complementarity problems, PhD thesis, Computer Science Departement, University of Wisconsin, Madison, 1994.

[11] S. P. Dirkse, M. C. Ferris, T. Munson, http://pages.cs.wisc.edu/~ferris/path.html

[12] A. L. Dontchev, R. T. Rockafellar, Parametric stability of solutions in models of economic equilibrium, *J. Convex Analysis* **19** (2012) 975–997.

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

[14] A. L. Dontchev, R. T. Rockafellar, *Implicit functions and solution mappings. A view from variational analysis,* 2nd Edition, Springer, 2014.

[15] F. Facchinei, J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems,* Springer, New York, 2003.

[16] O. P. Ferreira, G. N. Silva, Inexact Newton's method to nonlinear functions with values in a cone, arXiv preprint, arXiv:1510.01947, 2015.

[17] W. B. Gragg, R. A. Tapia, Optimal error bounds for the Newton-Kantorovich theorem, *SIAM J. Numer. Anal.* **11** (1974) 10–13.

[18] M. Hintermüller, Semismooth Newton Methods and Applications, Department of Mathematics, Humboldt-University of Berlin.

[19] K. Ito, K. Kunisch, *Lagrange multiplier approach to variational problems and applications,* SIAM, Philadelphia, PA, 2008.

[20] A. F. Izmailov, A. S. Kurennoy, M. V. Solodov, The Josephy-Newton method for semismooth generalized equations and semismooth SQP for optimization, *Set-Valued Var. Anal.* **21** (2013) 17–45.

[21] A. F. Izmailov, M. V. Solodov, *Newton-type methods for optimization and variational problems*, Springer, 2014.

[22] L. V. Kantorovich, On Newton's method for functional equations (Russian), *Doklady Akad. Nauk SSSR* (N.S.) **59** (1948) 1237–1240.

[23] L. V. Kantorovich, G. P. Akilov, *Functional analysis* (Russian), 2nd Edition, revised, Nauka, Moscow, 1977.

[24] C. T. Kelley, *Solving nonlinear equations with Newton's method,* Fundamentals of Algorithms, SIAM, Philadelphia, PA, 2003.

[25] D. Klatte, B. Kummer, *Nonsmooth equations in optimization. Regularity, calculus, methods and applications,* Kluwer, New York, 2002.

[26] B. Kummer, Newton's method for non-differentiable functions, in "Advances in Mathematical Optimization", ed J. Guddat et al., Ser. Math. Res. **45**, Akademie-Verlag, Berlin 1988, 114–125.

[27] J. M. Ortega, The Newton-Kantorovich theorem, *Amer. Math. Monthly* **75** (1968) 658-660.

[28] A. Pietrus, Non differentiable perturbed Newton's method for functions with values in a cone, *Investigación Oper.* **35** (2014) 58–67.

[29] F. A. Potra, V. Pták, *Nondiscrete induction and iterative processes,* Research Notes in Mathematics **103**, Pitman, Boston, MA, 1984.

[30] L. Qi, J. Sun, A nonsmooth version of Newton's method, *Math. Programming A* **58** (1993) 353–367.

[31] S. M. Robinson, Extension of Newton's method to nonlinear functions with values in a cone, *Numer. Math.* **19** (1972) 341–347.

[32] Silva, G. N., Kantorovich's theorem on Newton's method for solving generalized equations under the majorant condition, Appl. Math. Comput. **286** (2016) 178–188.

[33] M. Ulbrich, *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces*, SIAM, Philadelphia, PA, 2011.

[34] T. J. Ypma, Historical development of the Newton-Raphson method, *SIAM Rev.* **37** (1995) 531–551.