Journal of Complexity 28 (2012) 346–363

Contents lists available at SciVerse ScienceDirect

Journal of Complexity

journal homepage: www.elsevier.com/locate/jco

A robust Kantorovich's theorem on the inexact Newton method with relative residual error tolerance

O.P. Ferreira [a,*], B.F. Svaiter [b]

[a] IME/UFG, Campus II - Caixa Postal 131, CEP 74001-970 - Goiânia, GO, Brazil

[b] IMPA, Estrada Dona Castorina, 110, Jardim Botânico, CEP 22460-320 - Rio de Janeiro, RJ, Brazil

A R T I C L E I N F O

Article history:
Received 19 November 2010
Accepted 26 January 2012
Available online 21 February 2012

Keywords:
Kantorovich's theorem
Inexact Newton method
Banach space

A B S T R A C T

We prove that under semi-local assumptions, the inexact Newton method with a *fixed* relative residual error tolerance converges Q-linearly to a zero of the nonlinear operator under consideration. Using this result we show that the Newton method for minimizing a self-concordant function or to find a zero of an analytic function can be implemented with a fixed relative residual error tolerance.

In the absence of errors, our analysis retrieve the classical Kantorovich Theorem on the Newton method.

0. Introduction

The Newton method and its variations, including the inexact Newton methods, are the most efficient methods known for solving nonlinear equations

$$F(x) = 0,$$

where $\mathbb{X}$ and $\mathbb{Y}$ are Banach spaces, $C \subseteq \mathbb{X}$ and $F : C \to \mathbb{Y}$ is continuous and continuously differentiable on $int(C)$. Although the Newton method is quite efficient, it is also computationally expensive, because in each iteration a linear system involving the Jacobian of F must be solved. The solution of this linear system accounts mostly for the computational burden of these algorithms and sometimes, computation of the Jacobian is also expensive. A number of successful strategies were developed to circumvent these difficulties, and we will recall some of them. The Jacobian of F may be computed by finite-difference differentiation, or may be interactively approximated as in secant-methods (*e.g.* BFGS). The linear system may be solved approximately using iterative methods for large scale problems, as SOR, splitting (*e.g.* Gauss–Seidel) or conjugate-gradient methods. Moreover, modern implementations of the conjugate gradients, coupled with preconditioning, allows for the approximated

* Corresponding author.
E-mail addresses: orizon@mat.ufg.br (O.P. Ferreira), benar@impa.br (B.F. Svaiter).

0885-064X/$ – see front matter
doi:10.1016/j.jco.2012.02.002

solution of large scale linear systems, not feasibly solvable by using Gaussian elimination/matrix factorization. The linear system may be solved by a LU/Cholesky factorization, which can be used for solving some next iteration linear-systems. Note that in all these cases, approximated solutions of the linear system are used, instead of the theoretical Newton step. Even when the Jacobian is available and the linear system is solved by Gaussian elimination or matrix factorization, these methods provide approximated solutions with small residuals, due to round-off errors in floating-point arithmetics.

Kantorovich's Theorem on the Newton method uses semi-local assumption on F to guarantee existence of a solution of the above equation, uniqueness of this solution in a prescribed region and also convergence of the Newton method to such a solution, see [8,9]. Semi-local convergence theorems for the Newton method has been instrumental in the modern complexity analysis of the solution of polynomial (or analytical) equations [2,19], linear and quadratic programming problems and linear semi-definite programming problems [15,16]. These convergence results have also been used in the design and convergence analysis of algorithms for these problems. In all these applications, homotopy methods are combined with the Newton method, which helps the algorithm to keep track of the solution of a parametrized perturbed version of the original problem.

In view of the above mention convenience/necessity of implementing the Newton method with inexact Newton steps, it is natural to investigate robustness of Kantorovich's and Kantorovich's-like theorems under errors in the computation of the Newton step. Whenever the Jacobian of F is available, the residual of the linear system for the approximate Newton step is ready available, and can be used for error-tolerance criterion. It would be most desirable to have an *a priori* prescribed residual error tolerance in the iterative solutions of linear system for computing the Newton steps, because this would prevent over-solving and/or under-solving the linear system in question. Indeed, in all homotopy methods, the (parametrized) nonlinear equation in never solved up to machine precision, with the possible exception of the last equation. Instead, a single Newton step is used to maintain the iterate in the good convergence neighborhood for the (current) homotopic problem. So, it is interesting to verify if this could also be accomplished using an inexact Newton step with a fixed relative residual error tolerance. Although the local convergence analysis of the Newton method with relative errors in the residue [3,4,14] or in the steep [25] are well understood, the convergence analysis of the method under general semi-local assumptions assuming *only* bounded relative residual errors is a new contribution of this paper. Previous works on this subject include [13,17]. The advantage of working with an error tolerance on the residual rests in the fact that the exact Newton step need not to be know for evaluating this error, which makes this criterion attractive for practical applications.

Recently, Kantorovich's theorem on the Newton method was extended to Riemannian manifolds using a new technique which simplifies the analyses and proof of this theorem, see [5]. The basic idea is to combine the modern formulation of Kantorovich's Theorem by means of majorant functions [26] with the definitions of good regions for the inexact Newton method [5] (see also [6]). In these regions, the majorant function bounds the nonlinear function which root is to be found, and the behavior of the inexact Newton iteration in these regions is estimated using iterations associated to the majorant function. Moreover, as a whole, the union of all these regions is invariant under inexact Newton's iteration. This technique was successfully employed for proving generalized versions of Kantorovich's theorem in Riemannian Manifolds and also in the analysis of the classical version of Kantorovich's theorem in Banach spaces, see [1,6,10–12,22–24]. In the present work, we will use the technique introduced in [5] to present a robust version of the Kantorovich's theorem on the inexact Newton method with residual relative error. It is worth to point out that, for null error tolerance the analysis presented merge in the usual semi-local convergence analysis on the Newton method, see [6].

This paper is organized as follows. In Section 1, some definitions and auxiliary results are presented. In Section 2 the main result is stated and some properties of the majorant function are established. The main relationships between the majorant function and the nonlinear operator used in the paper are presented in Section 3. In Section 4 a family of regions where the behavior of the inexact Newton iteration is estimated using the majorant function is introduced. We also show that the union of all these regions is invariant under the inexact Newton iteration with a fixed relative residual error tolerance. In Section 5 the main result is proved. In Section 6 we show that the Newton method for minimizing a self-concordant function under the usual semi-local assumption for these functions, can be implemented with a fixed residual error tolerance. Moreover, we show that the Newton method for finding a zero of an analytic function, under the usual semi-local assumption of the α-theory can be also be implemented with a fixed relative residual error tolerance.

1. Basics definitions and auxiliary results

Let $\mathbb{X}$ be a Banach space. The open and closed ball at x are denoted, respectively by

$$B(x,r)=\{y\in\mathbb{X};\ \|x-y\|<r\},\qquad B[x,r]=\{y\in\mathbb{X};\ \|x-y\|\leqslant r\}.$$

The following auxiliary results of elementary convex analysis will be needed.

Proposition 1.1. *Let $I\subset\mathbb{R}$ be an interval, and $\varphi: I\to\mathbb{R}$ be convex.*

1. *For any $u_0\in \operatorname{int}(I)$, the application*

$$u\mapsto \frac{\varphi(u_0)-\varphi(u)}{u_0-u},\quad u\in I,\ u\neq u_0,$$

is increasing and there exist (in $\mathbb{R}$)

$$D^-\varphi(u_0)=\lim_{u\to u_{0-}}\frac{\varphi(u_0)-\varphi(u)}{u_0-u}=\sup_{u<u_0}\frac{\varphi(u_0)-\varphi(u)}{u_0-u}.$$

2. *If $u,v,w\in I$, $u<w$, and $u\le v\le w$ then*

$$\varphi(v)-\varphi(u)\le[\varphi(w)-\varphi(u)]\frac{v-u}{w-u}.$$

Proof. See [7]. □

Proposition 1.2. *If $h:[a,b)\to\mathbb{R}$ is convex, differentiable at a, $h'(a)<0$ and*

$$\lim_{t\to b_-}h(t)=0,$$

then

$$a-\frac{h(a)}{h'(a)}\le b,$$

with equality if and only if h is affine in $[a,b)$.

Proof. Since h is convex, $h(a)+h'(a)(t-a)\le h(t)$ for any $t\in[a,b)$. Taking the limit $t\to b_-$ we obtain

$$h(a)+h'(a)(b-a)\le 0.$$

The desired inequality now follows multiplying this inequality by the strictly positive number $-1/h'(a)$. If the above inequality holds as an equality, then

$$h'(a)=\frac{-h(a)}{b-a}.$$

Let $a\le s<t<b$. Using again the convexity of h we have

$$h(a)+h'(a)(s-a)\le h(s)\le h(a)\frac{t-s}{t-a}+h(t)\frac{s-a}{t-a}.$$

Taking again the limit $t\to b_-$ in the above equation and using the previous equation we conclude that $h(s)=h(a)(b-s)/(b-a)$, i.e., h is affine. If h is affine then the inequality of the proposition holds trivially as an equality. □

2. The inexact Newton method with relative error

Our goal is to prove the following version of Kantorovich's theorem on the inexact Newton method with relative error.

Theorem 2.1. *Let $\mathbb{X}$ and $\mathbb{Y}$ be Banach spaces, $R \in \mathbb{R}$, $C \subseteq \mathbb{X}$ and $F : C \to \mathbb{Y}$ a continuous function, continuously differentiable on int(C). Take $x_0 \in \operatorname{int}(C)$ with $F'(x_0)$ non-singular. Suppose that $f : [0, R) \to \mathbb{R}$ is continuously differentiable, $B(x_0, R) \subseteq C$,*

$$\|F'(x_0)^{-1}[F'(y) - F'(x)]\| \le f'(\|y - x\| + \|x - x_0\|) - f'(\|x - x_0\|), \tag{2.1}$$

for any $x, y \in B(x_0, R)$, $\|x - x_0\| + \|y - x\| < R$,

$$\|F'(x_0)^{-1}F(x_0)\| \le f(0), \tag{2.2}$$

(h1) $f(0) > 0, f'(0) = -1$*;*
(h2) *f' is strictly increasing and convex;*
(h3) *$f(t) < 0$ for some $t \in (0, R)$.*

Let

$$\beta := \sup_{t \in [0,R)} -f(t), \qquad t_* := \min f^{-1}(\{0\}), \qquad \bar{\tau} := \sup\{t \in [0, R) : f(t) < 0\}.$$

Take $0 \le \rho < \beta/2$ and define

$$\kappa_\rho := \sup_{\rho < t < R} \frac{-(f(t) + 2\rho)}{|f'(\rho)|(t - \rho)}, \qquad \lambda_\rho := \sup\{t \in [\rho, R) : \kappa_\rho + f'(t) < 0\},$$
$$\Theta_\rho := \frac{\kappa_\rho}{2 - \kappa_\rho}. \tag{2.3}$$

Then for any $\theta \in [0, \Theta_\rho]$ and $z_0 \in B(x_0, \rho)$, the sequence generated by the inexact Newton method for solving $F(x) = 0$ with starting point z_0 and residual relative error tolerance θ: for $k = 0, 1, \ldots,$

$$z_{k+1} = z_k + S_k, \qquad \|F'(z_0)^{-1}[F(z_k) + F'(z_k)S_k]\| \le \theta \|F'(z_0)^{-1}F(z_k)\|,$$

is well defined (for any particular choice of each S_k),

$$\|F'(z_0)^{-1}F(z_k)\| \le \left(\frac{1 + \theta^2}{2}\right)^k [f(0) + 2\rho], \quad k = 0, 1, \ldots, \tag{2.4}$$

the sequence $\{z_k\}$ is contained in $B(z_0, \lambda_\rho)$ and converges to a point $x_* \in B[x_0, t_*]$, which is the unique zero of F on $B(x_0, \bar{\tau})$. Moreover, if

(h4) $\lambda_\rho < R - \rho$*,*

then the sequence $\{z_k\}$ satisfies, for $k = 0, 1, \ldots,$

$$\|x_* - z_{k+1}\| \le \left[\frac{1 + \theta}{2} \frac{D^- f'(\lambda_\rho)}{|f'(\lambda_\rho)|} \|x_* - z_k\| + \theta \frac{f'(\lambda_\rho + \rho) + 2|f'(\rho)|}{|f'(\lambda_\rho + \rho)|}\right] \|x_* - z_k\|. \tag{2.5}$$

If, additionally, $0 \le \theta < \kappa_\rho/(4 + \kappa_\rho)$ then $\{z_k\}$ converges Q-linearly as follows

$$\|x_* - z_{k+1}\| \le \left[\frac{1 + \theta}{2} + \frac{2\theta}{\kappa_\rho}\right] \|x_* - z_k\|, \quad k = 0, 1, \ldots.$$

Remark 2.2. In Theorem 2.1 if $\theta = 0$ we obtain the exact Newton method and its convergence properties. Now, taking $\theta = \theta_k$ in each iteration and letting θ_k goes to zero as k goes to infinity, inequality (2.5) implies that the generated sequence converges to the solution with asymptotic superlinear rate.

From now on, we assume that the hypotheses of Theorem 2.1 hold. The scalar function f in the above theorem is called a *majorant function* for F at point x_0. Before proceeding, we will analyze some basic properties of the majorant function. Condition **(h2)** implies the strict convexity of f. Note that t_* is the smallest root of $f(t) = 0$ and, since f is convex, if this equation has two roots, then the second one is $\bar{\tau}$.

Define

$$\bar{t} := \sup\{t \in [0, R) : f'(t) < 0\}. \tag{2.6}$$

Proposition 2.3. *The following statements on the majorant function hold*

(i) $f'(t) < 0$ *for any* $t \in [0, \bar{t})$, *(and* $f'(t) \geq 0$ *for any* $t \in [0, R) \setminus [0, \bar{t})$*);*
(ii) $0 < t_* < \bar{t} \leq \bar{\tau} \leq R$*;*
(iii) $\beta = -\lim_{t \to \bar{t}_-} f(t),\ 0 < \beta < \bar{t}$.

Proof. Item (i) follows from the second part of **(h1)**, **(h2)** and the definition (2.6).

Using the first inequality in **(h1)**, **(h3)** and the continuity of f we conclude that t_* is well defined and

$$0 < t_* < R.$$

Condition **(h2)** implies in strict convexity of f, hence condition **(h3)** and the definition of t_* imply that there exists $t \in (t_*, R)$ such that

$$0 > f(t) > f(t_*) + f'(t_*)(t - t_*) = f'(t_*)(t - t_*),$$

which implies that $0 > f'(t_*)$. Therefore, using item (i) and the definition of $\bar{t}$ we have

$$t_* < \bar{t} \leq R.$$

Since t_* is the smallest root of $f(t) = 0$ and f is strictly decreasing in $[0, \bar{t})$ we conclude that $f < 0$ in $[t_*, \bar{t})$. So, the definition of $\bar{\tau}$ implies that

$$\bar{t} \leq \bar{\tau} \leq R,$$

and the proof of item (ii) is concluded.

Using **(h3)** and the definition of β we obtain that $0 < \beta$. Since f is convex, combining this with **(h1)** we have

$$f(t) \geq f(0) - t > -t, \quad 0 \leq t < R,$$

with strict inequality for $t \neq 0$. We know that f is strictly decreasing and $f < 0$ in $[t_*, \bar{t})$. Hence, letting t goes to $\bar{t}_-$ in last inequality and using the definition of β the item (iii) follows. □

We will first prove Theorem 2.1 for the case $\rho = 0$ and $z_0 = x_0$. In order to simplify the notation in the case $\rho = 0$, we will use κ, λ and θ instead of κ_0, λ_0 and θ_0 respectively:

$$\kappa := \sup_{0<t<R} \frac{-f(t)}{t}, \qquad \lambda := \sup\{t \in [0, R) : \kappa + f'(t) < 0\}, \qquad \Theta := \frac{\kappa}{2 - \kappa}. \tag{2.7}$$

Proposition 2.4. *For* κ, λ, θ *as in* (2.7) *it holds that*

$$0 < \kappa < 1, \qquad 0 < \Theta < 1, \qquad t_* < \lambda \leq \bar{t} \leq \bar{\tau}, \tag{2.8}$$

and

$$\begin{aligned} &f'(t) + \kappa < 0, \quad \forall t \in [0, \lambda), \\ &\inf_{0 \leq t < R} f(t) + \kappa t = \lim_{t \to \lambda_-} f(t) + \kappa t = 0, \end{aligned} \tag{2.9}$$

Proof. Since f is convex, combining this with **(h1)** we have

$$f(t) \geq f(0) - t > -t, \quad 0 \leq t < R,$$

with strict inequality for $t \neq 0$. For $t \neq 0$, last inequality is equivalent to

$$\frac{-f(t)}{t} \leq 1 - \frac{f(0)}{t} < 1 - \frac{f(0)}{R} < 1, \quad 0 < t < R,$$

and, using also **(h3)**, we conclude that

$$0 < \kappa < 1, \qquad 0 < \Theta < 1,$$

where the bounds on Θ follows from its definition and the bound on κ. Moreover, as f' is continuous, strictly increasing and $f'(0) = -1$, we obtain

$$0 < \lambda, \qquad f'(t) + \kappa < 0, \quad \forall t \in [0, \lambda),$$
$$\inf_{0 \leq t < R} f(t) + \kappa t = \lim_{t \to \lambda_-} f(t) + \kappa t = 0,$$

where the last equalities follows from the definition of κ.

Note that $f'(t) < -\kappa < 0$ for all $t \in [0, \lambda)$. Since f' is strictly negative in $[0, \lambda)$, we conclude that $t_* < \lambda \leq \bar{t} \leq \bar{\tau}$ and the proof is concluded. □

3. Basic results

In this section we will obtain bounds on $\|F'^{-1}\|$ and on the linearization error on F using the majorant function f. This bounds will be used in the next section for analyzing the inexact Newton iterations. It is worth mentioning that in this section the inequality on **(h1)** and (2.2) will be used only for proving its last result and **(h3)** will not be used.

A Newton iteration at x requires non-singularity of $F'(x)$, which will be verified using the majorant function f.

Proposition 3.1. *If* $\|x - x_0\| \leq t < \bar{t}$ *then* $F'(x)$ *is non-singular and*

$$\|F'(x)^{-1}F'(x_0)\| \leq \frac{1}{-f'(t)}.$$

Proof. The definition (2.6) shows that $f'(t) < 0$. Direct manipulation, (2.1), **(h1)** and **(h2)** give us

$$\begin{aligned}\|F'(x_0)^{-1}F'(x) - I\| &= \|F'(x_0)^{-1}[F'(x) - F'(x_0)]\| \leq f'(\|x - x_0\|) - f'(0) \\ &= f'(t) + 1 < 1.\end{aligned}$$

Using Banach's Lemma and the last inequality above we conclude that $F'(x_0)^{-1}F'(x)$ is non-singular and

$$\|F'(x)^{-1}F'(x_0)\| = \|(F'(x_0)^{-1}F'(x))^{-1}\| \leq \frac{1}{1 - (f'(t) + 1)},$$

which is the desired inequality. □

The linearization errors on F and f are, respectively

$$E_F(y, x) := F(y) - [F(x) + F'(x)(y - x)], \quad x \in B(x_0, R),\ y \in C \tag{3.1}$$
$$e_f(v, t) := f(v) - [f(t) + f'(t)(v - t)], \quad t, v \in [0, R). \tag{3.2}$$

The linearization error of the majorant function bounds the linearization error of F.

Lemma 3.2. *If* $x, y \in \mathbb{X}$ *and* $\|x - x_0\| + \|y - x\| < R$ *then*

$$\|F'(x_0)^{-1}E_F(y, x)\| \leq e_f(\|x - x_0\| + \|y - x\|, \|x - x_0\|),$$

Proof. Since

$$x + u(y - x) \in B(x_0, R), \quad 0 \leq u \leq 1,$$

and F is continuously differentiable in $B(x_0, R)$, direct use of (3.1) gives

$$E_F(y, x) = \int_0^1 [F'(x + u(y - x)) - F'(x)](y - x)du.$$

Combining the above equality with (2.1) we have

$$\begin{aligned}\|F'(x_0)^{-1}E_F(y, x)\| &\leq \int_0^1 \|F'(x_0)^{-1}[F'(x + u(y - x)) - F'(x)]\| \, \|y - x\| \, du \\ &\leq \int_0^1 [f'(\|x - x_0\| + u\|y - x\|) - f'(\|x - x_0\|)]\|y - x\| \, du\end{aligned}$$

which after performing the integration and using the definition in (3.2) yields the desired inequality. □

Convexity of f and f' guarantee that $e_f(t + s, t)$ is increasing in s and t.

Lemma 3.3. *If* $0 \leq b \leq t$, $0 \leq a \leq s$ *and* $t + s < R$ *then*

$$e_f(a + b, b) \leq e_f(t + s, t),$$
$$e_f(a + b, b) \leq \frac{1}{2}\frac{f'(t + s) - f'(t)}{s} \, a^2, \quad s \neq 0.$$

Proof. First note that

$$e_f(a + b, b) = \int_0^a [f'(b + r) - f'(b)]dr.$$

Since f' is convex, for any $\tau_0 > 0$, the function $\tau \mapsto f'(\tau + \tau_0) - f'(\tau)$ is non-decreasing. So,

$$e_f(a + b, b) \leq \int_0^a [f'(t + r) - f'(t)]dr \ \leq \int_0^s [f'(t + r) - f'(t)]dr \tag{3.3}$$

where the second inequality follows from the convexity of f, which implies positivity of the integrand. To end the proof of first inequality, note that the last term on the above inequality is $e_f(t + s, t)$.

For proving second inequality, apply Proposition 1.1 item 2 with $u = t$, $v = t + r$, $w = t + s$ and $\varphi = f'$ in first inequality in (3.3) to conclude that

$$e_f(a + b, b) \leq \int_0^a [f'(t + s) - f'(t)] \, \frac{r}{s} \, dr,$$

which performing the integral gives the desired inequality. □

Now we are ready to bound the linearization error E_F using the linearization error on the majorant function.

Corollary 3.4. *If* $x, y \in \mathbb{X}$, $\|x - x_0\| \leq t$, $\|y - x\| \leq s$ *and* $s + t < R$ *then*

$$\|F'(x_0)^{-1}E_F(y, x)\| \leq e_f(t + s, t),$$
$$\|F'(x_0)^{-1}E_F(y, x)\| \leq \frac{1}{2}\frac{f'(s + t) - f'(t)}{s}\|y - x\|^2, \quad s \neq 0.$$

Proof. The results follows by direct combination of the Lemmas 3.2 and 3.3 by taking $b = \|x - x_0\|$ and $a = \|y - x\|$. □

The first inequality in the next corollary will be useful for obtaining asymptotic bounds on the sequence generated by the inexact Newton method with relative error tolerance, while the second inequality will be used to show that this method is robust with respect to the initial iterate.

Corollary 3.5. *For any $y \in B(x_0, R)$,*

$$-f(\|y - x_0\|) \leq \|F'(x_0)^{-1}F(y)\| \leq f(\|y - x_0\|) + 2\|y - x_0\|.$$

Proof. Using Lemma 3.2 with $x = x_0$, the definition of E_F and triangle inequality we have

$$\begin{aligned} e_f(\|y - x_0\|, 0) &\geq \|F'(x_0)^{-1}E_F(y, x_0)\| \\ &\geq \|F'(x_0)^{-1}F(x_0) + y - x_0\| - \|F'(x_0)^{-1}F(y)\| \\ &\geq \|y - x_0\| - \|F'(x_0)^{-1}F(x_0)\| - \|F'(x_0)^{-1}F(y)\|. \end{aligned}$$

Combining this inequality with the definition of e_f and using the assumption **(h1)** and (2.2) we obtain

$$f(\|y - x_0\|) - f(0) + \|y - x_0\| \geq \|y - x_0\| - f(0) - \|F'(x_0)^{-1}F(y)\|,$$

which is equivalent to the first inequality of the corollary.

To prove the second inequality, use again Lemma 3.2 the definition of E_F and triangle inequality to obtain

$$\begin{aligned} e_f(\|y - x_0\|, 0) &\geq \|F'(x_0)^{-1}E_F(y, x_0)\| \\ &\geq \|F'(x_0)^{-1}F(y)\| - \|F'(x_0)^{-1}F(x_0) + y - x_0\| \\ &\geq \|F'(x_0)^{-1}F(y)\| - \|F'(x_0)^{-1}F(x_0)\| - \|y - x_0\|. \end{aligned}$$

Using the above inequality, the definition of e_f, **(h1)** and (2.2) we have

$$f(\|y - x_0\|) - f(0) + \|y - x_0\| \geq \|F'(x_0)^{-1}F(y)\| - f(0) - \|y - x_0\|$$

which is equivalent to the second inequality of the corollary. □

Note that the first inequality on the above corollary proves that F has no zeros in the region $t_* < \|x - x_0\| < \bar{\tau}$.

Lemma 3.6. *If $x \in \mathbb{X}$, $\|x - x_0\| \leq t < R$ then*

$$\|F'(x_0)^{-1}F'(x)\| \leq 2 + f'(t).$$

Proof. Simple algebraic manipulation together with assumption (2.1) give us

$$\|F'(x_0)^{-1}F'(x)\| \leq \|I + F'(x_0)^{-1}[F'(x) - F'(x_0)]\| \leq 1 + f'(\|x - x_0\|) - f'(0).$$

Hence, **(h1)**, **(h2)** and the last inequality imply the statement of the lemma. □

Lemma 3.7. *Take $\theta \geq 0$, $0 \leq t \leq \lambda$, $x_*, x, y \in \mathbb{X}$. If $\lambda < R$, $\|x - x_0\| \leq t$, $\|x_* - x\| \leq \lambda - t$, $F(x_*) = 0$ and*

$$\|F'(x_0)^{-1}[F(x) + F'(x)(y - x)]\| \leq \theta\|F'(x_0)^{-1}F(x)\|, \tag{3.4}$$

then

$$\|x_* - y\| \leq \left[\frac{1 + \theta}{2} + \frac{2\theta}{\kappa}\right]\|x_* - x\|, \tag{3.5}$$

$$\|x_* - y\| \leq \left[\frac{1 + \theta}{2}\frac{D^-f'(\lambda)}{|f'(\lambda)|}\|x_* - x\| + \theta\,\frac{2 + f'(\lambda)}{|f'(\lambda)|}\right]\|x_* - x\|. \tag{3.6}$$

Proof. Since $F(x_*) = 0$, direct algebraic manipulation and (3.1) yield

$$y - x_* = F'(x)^{-1}[E_F(x_*, x) + [F(x) + F'(x)(y - x)]].$$

Using (3.4), properties of the norm and some simple manipulations we conclude from last equality that

$$\|x_* - y\| \le \|F'(x)^{-1}F'(x_0)\|[\|F'(x_0)^{-1}E_F(x_*, x)\| + \theta\|F(x_0)^{-1}F(x)\|].$$

On the other hand, using again $F(x_*) = 0$ and the definition in (3.1) we have

$$-F'(x_0)^{-1}F(x) = F'(x_0)^{-1}[E_F(x_*, x) + F'(x)(x_* - x)],$$

which using the triangular inequality yields

$$\|F'(x_0)^{-1}F(x)\| \le \|F'(x_0)^{-1}E_F(x_*, x)\| + \|F'(x_0)^{-1}F'(x)\| \, \|x_* - x\|.$$

Combining two above inequalities with Proposition 3.1, Corollary 3.4 with $y = x_*$ and $s = \lambda - t$ and Lemma 3.6 we have

$$\|x_* - y\| \le \frac{1}{|f'(t)|}\left[\frac{1+\theta}{2}\frac{f'(\lambda) - f'(t)}{\lambda - t}\|x_* - x\| + \theta\,[2 + f'(t)]\right]\|x_* - x\|.$$

Since $\|x_* - x\| \le \lambda - t, f' < -\kappa < 0$ in $[0, \lambda)$ and f' is increasing the first inequality follows from last inequality.

Using Proposition 1.1 and taking in account that $f' < 0$ in $[0, \lambda)$ and increasing we obtain the second inequality from above inequality. □

4. The inexact Newton iteration with relative error

In the next lemma we study a single inexact Newton iteration with relative error θ.

Lemma 4.1. *Take $t, \varepsilon, \theta \ge 0$, and $x \in C$ such that*

$$\|x - x_0\| \le t < \bar{t}, \qquad \|F'(x_0)^{-1}F(x)\| \le f(t) + \varepsilon, \qquad t - (1+\theta)\frac{f(t)+\varepsilon}{f'(t)} < R. \tag{4.1}$$

If $y \in \mathbb{X}$ and

$$\|F'(x_0)^{-1}[F(x) + F'(x)(y - x)]\| \le \theta\|F'(x_0)^{-1}F(x)\|. \tag{4.2}$$

then

1. $\|y - x\| \le -(1+\theta)\frac{f(t)+\varepsilon}{f'(t)}$;
2. $\|y - x_0\| \le t - (1+\theta)\frac{f(t)+\varepsilon}{f'(t)} < R$;
3. $\|F'(x_0)^{-1}F(y)\| \le f(t - (1+\theta)\frac{f(t)+\varepsilon}{f'(t)}) + \varepsilon + 2\theta(f(t) + \varepsilon)$.

Proof. Using Proposition 3.1 and the first inequality in (4.1) we conclude that $F'(x)$ is non-singular and $\|F'(x)^{-1}F'(x_0)\| \le -1/f'(t)$. Therefore, using also the identity

$$y - x = F'(x)^{-1}F'(x_0)[F'(x_0)^{-1}[F(x) + F'(x)(y - x)] - F'(x_0)^{-1}F(x)],$$

triangular inequality and (4.2) we conclude that

$$\|y - x\| \le \frac{-1}{f'(t)}(1+\theta)\|F'(x_0)^{-1}F(x)\|.$$

To end the proof of item 1, use the above inequality and the second inequality on (4.1).

Item 2 follows from triangular inequality, item 1 and the first and the third inequalities in (4.1).

Using the definition of the error (3.1) we have

$$F(y) = E_F(y, x) + F'(x_0)[F'(x_0)^{-1}[F(x) + F'(x)(y - x)]].$$

Therefore, using the triangle inequality, (4.2) and the second inequality on (4.1) we have

$$\begin{aligned}\|F'(x_0)^{-1}F(y)\| &\le \|F'(x_0)^{-1}E_F(y,x)\| + \theta\|F'(x_0)^{-1}F(x)\| \\ &\le \|F'(x_0)^{-1}E_F(y,x)\| + \theta(f(t)+\varepsilon).\end{aligned}$$

Using (4.1), item 1, and Lemma 3.2 with $s = -(1+\theta)(f(t)+\varepsilon)/f'(t)$ we have

$$\begin{aligned}\|F'(x_0)^{-1}E_F(y,x)\| &\le e_f\left(t-(1+\theta)\frac{f(t)+\varepsilon}{f'(t)}, t\right) \\ &= f\left(t-(1+\theta)\frac{f(t)+\varepsilon}{f'(t)}\right) + \varepsilon + \theta(f(t)+\varepsilon).\end{aligned}$$

Direct combination of the two above equation yields the latter inequality in item 3. □

In view of Lemma 4.1 define, for $\theta \ge 0$, the auxiliary map $n_\theta : [0, \bar{t}\,) \times [0,\infty) \to \mathbb{R}\times\mathbb{R}$,

$$n_\theta(t,\varepsilon) := \left(t-(1+\theta)\frac{f(t)+\varepsilon}{f'(t)},\ \varepsilon + 2\theta(f(t)+\varepsilon)\right). \tag{4.3}$$

Let

$$\Omega := \{(t,\varepsilon)\in\mathbb{R}\times\mathbb{R} : 0\le t<\lambda,\ 0\le\varepsilon\le\kappa t, 0<f(t)+\varepsilon\}. \tag{4.4}$$

Lemma 4.2. *If $0\le\theta\le\Theta$, $(t,\varepsilon)\in\Omega$ and $(t_+,\varepsilon_+) = n_\theta(t,\varepsilon)$, that is,*

$$t_+ = t-(1+\theta)\frac{f(t)+\varepsilon}{f'(t)}, \qquad \varepsilon_+ = \varepsilon + 2\theta(f(t)+\varepsilon),$$

then $n_\theta(t,\varepsilon)\in\Omega$, $t<t_+$, $\varepsilon\le\varepsilon_+$ and

$$f(t_+)+\varepsilon_+ < \left(\frac{1+\theta^2}{2}\right)(f(t)+\varepsilon).$$

Proof. Since $0\le t<\lambda$, according to (2.7) we have $f'(t) < -\kappa < 0$. Therefore $t<t_+$ and $\varepsilon\le\varepsilon_+$.
As $\varepsilon\le\kappa t, f(t)+\varepsilon > 0$ and $-1\le f'(t) < f'(t)+\kappa < 0$,

$$\begin{aligned}-\frac{f(t)+\varepsilon}{f'(t)} &\le -\frac{f(t)+\kappa t}{f'(t)} \\ &= -\frac{f(t)+\kappa t}{f'(t)+\kappa}\left[1+\frac{\kappa}{f'(t)}\right] \le -\frac{f(t)+\kappa t}{f'(t)+\kappa}(1-\kappa).\end{aligned} \tag{4.5}$$

The function $h(s) := f(s)+\kappa s$ is differentiable at t, $h'(t)<0$, is strictly convex and

$$\lim_{s\to\lambda_-} h(t) = 0.$$

Therefore, using Proposition 1.2 we have $t - h(t)/h'(t) < \lambda$, which is equivalent to

$$-\frac{f(t)+\kappa t}{f'(t)+\kappa} < \lambda - t. \tag{4.6}$$

Combining the above inequality with (4.5) and the definition of t_+ we conclude that

$$t_+ < t + (1+\theta)(1-\kappa)(\lambda-t).$$

Using (2.7) and (2.8) we have $(1+\theta)(1-\kappa)\le 1-\theta<1$, which combined with the above inequality yields $t_+<\lambda$.

Using the definition of ε_+, inequality $\varepsilon \le \kappa\, t$ and (2.7) we obtain

$$\begin{aligned}\varepsilon_+ &\le 2\theta(f(t)+\varepsilon)+\kappa\, t\\ &= \kappa(t+(1+\theta)(f(t)+\varepsilon)).\end{aligned}$$

Using again the inequalities $f(t)+\varepsilon > 0$ and $-1 \le f'(t) < 0$ we have

$$f(t)+\varepsilon \le -\frac{f(t)+\varepsilon}{f'(t)}.$$

Combining the two above inequalities with the definition of t_+ we obtain $\varepsilon_+ \le \kappa\, t_+$.

For proving the two last inequalities, first note that from the definition of the linearization error in (3.2) we have

$$\begin{aligned}f(t_+)+\varepsilon_+ &= f(t)+f'(t)(t_+-t)+e_f(t_+,t)+2\theta(f(t)+\varepsilon)+\varepsilon\\ &= \theta(f(t)+\varepsilon)+e_f(t_+,t)\\ &= \theta(f(t)+\varepsilon)+\int_t^{t_+}(f'(u)-f'(t))du.\end{aligned}$$

Since f' is strictly increasing we conclude that the integral is positive. So, last equality implies that $f(t_+)+\varepsilon_+ \ge \theta(f(t)+\varepsilon) > 0$. Taking $s \in [t_+,\lambda)$ and using the convexity of f' we have

$$\begin{aligned}\int_t^{t_+}(f'(u)-f'(t))du &\le \int_t^{t_+}(f'(s)-f'(t))\frac{u-t}{s-t}\,du\\ &= \frac{1}{2}\frac{(t_+-t)^2}{s-t}(f'(s)-f'(t)).\end{aligned}$$

Substituting last inequality into above equation we have

$$\begin{aligned}f(t_+)+\varepsilon_+ &\le \theta(f(t)+\varepsilon)+\frac{1}{2}\frac{(t_+-t)^2}{s-t}(f'(s)-f'(t))\\ &= \left(\theta+\frac{1}{2}\frac{(1+\theta)^2}{(s-t)}\frac{f(t)+\varepsilon}{-f'(t)}\frac{f'(s)-f'(t)}{-f'(t)}\right)(f(t)+\varepsilon).\end{aligned}$$

On the other hand, because $f'(s)+\kappa < 0$ and $-1 \le f'(t)$ it easy to conclude that

$$\frac{f'(s)-f'(t)}{-f'(t)} = \frac{f'(s)+\kappa-f'(t)-\kappa}{-f'(t)} \le 1-\kappa.$$

Combining last two above inequalities with (4.5), (4.6) and taking in account that $(1+\theta)(1-\kappa) \le 1-\theta$ we conclude that

$$\begin{aligned}f(t_+)+\varepsilon_+ &\le \left(\theta+\frac{1}{2}(1+\theta)^2(1-\kappa)^2\frac{\lambda-t}{s-t}\right)(f(t)+\varepsilon)\\ &= \left(\theta+\frac{1}{2}(1-\theta)^2\frac{\lambda-t}{s-t}\right)(f(t)+\varepsilon),\end{aligned}$$

and the result follows taking the limit $s \to \lambda_-$. □

The outcome of an inexact Newton iteration is any point satisfying some error tolerance. Hence, instead of a mapping for Newton iteration, we shall deal with a *family* of mappings, describing all possible inexact iterations.

Definition 4.3. For $0 \le \theta$, $\mathcal{N}_\theta$ is the family of maps $N_\theta : B(x_0,\bar t) \to \mathbb{X}$ such that

$$\|F'(x_0)^{-1}[F(x)+F'(x)(N_\theta(x)-x)]\| \le \theta\|F'(x_0)^{-1}F(x)\|, \tag{4.7}$$

for each $x \in B(x_0,\bar t\,)$.

If $x \in B(x_0, \bar{t})$, then $F'(x)$ is non-singular. Therefore, for $\theta = 0$, the family $\mathcal{N}_0$ has a single element, namely the exact Newton iteration map

$$N_0 : B(x_0, \bar{t}) \to \mathbb{X}, \qquad x \mapsto N_0(x) = x - F'(x)^{-1}F(x).$$

Trivially, if $0 \le \theta \le \theta'$ then $\mathcal{N}_0 \subset \mathcal{N}_\theta \subset \mathcal{N}_{\theta'}$. Hence, $\mathcal{N}_\theta$ is non-empty for all $\theta \ge 0$.

Remark 4.4. For any $\theta \in (0, 1)$ and $N_\theta \in \mathcal{N}_\theta$

$$N_\theta(x) = x \iff F(x) = 0, \quad x \in B(x_0, \bar{t}).$$

This means that the fixed point of the inexact Newton iteration N_θ are the same fixed points of the *exact* Newton iteration, namely, the zeros of F.

The main tool for the analysis of the inexact Newton method with a relative residual tolerance will be a family of sets described below and analyzed in the ensuing proposition, which is a combination of Lemmas 4.1 and 4.2. Define

$$K(t, \varepsilon) := \{x \in \mathbb{X} : \|x - x_0\| \le t, \|F'(x_0)^{-1}F(x)\| \le f(t) + \varepsilon\}, \tag{4.8}$$

and

$$K := \bigcup_{(t,\varepsilon)\in\Omega} K(t, \varepsilon). \tag{4.9}$$

Recall that n_θ, Ω and $\mathcal{N}_\theta$ were defined in (4.3), (4.4) and Definition 4.3 respectively.

Proposition 4.5. *Take $0 \le \theta \le \Theta$ and $N_\theta \in \mathcal{N}_\theta$. Then for any $(t, \varepsilon) \in \Omega$ and $x \in K(t, \varepsilon)$*

$$N_\theta(K(t, \varepsilon)) \subset K(n_\theta(t, \varepsilon)) \subset K, \qquad \|N_\theta(x) - x\| \le t_+ - t,$$

where t_+ is the first component of $n_\theta(t, \varepsilon)$. Moreover,

$$n_\theta(\Omega) \subset \Omega, \qquad N_\theta(K) \subset K. \tag{4.10}$$

Proof. Combine definitions (4.3), (4.4), Definition 4.3, (4.8), (4.9) with Lemmas 4.1 and 4.2. □

5. Convergence analysis

Theorem 5.1. *Take $0 \le \theta \le \Theta$ and $N_\theta \in \mathcal{N}_\theta$. For any $(t_0, \varepsilon_0) \in \Omega$ and $y_0 \in K(t_0, \varepsilon_0)$ the sequences*

$$y_{k+1} = N_\theta(y_k), \qquad (t_{k+1}, \varepsilon_{k+1}) = n_\theta(t_k, \varepsilon_k), \quad k = 0, 1, \ldots, \tag{5.1}$$

are well defined,

$$y_k \in K(t_k, \varepsilon_k), \qquad (t_k, \varepsilon_k) \in \Omega \quad k = 0, 1, \ldots, \tag{5.2}$$

the sequence $\{t_k\}$ is strictly increasing and converges to some $\tilde{t} \in (0, \lambda]$, the sequence $\{\varepsilon_k\}$ is non-decreasing and converges to some $\tilde{\varepsilon} \in [0, \kappa\lambda]$,

$$\begin{aligned}\|F'(x_0)^{-1}F(y_k)\| &\le f(t_k) + \varepsilon_k \\ &\le \left(\frac{1+\theta^2}{2}\right)^k (f(t_0) + \varepsilon_0), \quad k = 0, 1, \ldots\end{aligned} \tag{5.3}$$

the sequence $\{y_k\}$ is contained in $B(x_0, \lambda)$ and converges to a point $x_* \in B[x_0, t_*]$ which is the unique zero of F in $B(x_0, \bar{\tau})$ and

$$\|y_{k+1} - y_k\| \le t_{k+1} - t_k, \qquad \|x_* - y_k\| \le \tilde{t} - t_k, \quad k = 0, 1, \ldots. \tag{5.4}$$

Moreover, if

$((\mathbf{h4})')$ $\lambda < R$,

then the sequence $\{y_k\}$ *satisfies, for* $k = 0, 1, \ldots$

$$\|x_* - y_{k+1}\| \leq \left[\frac{1+\theta}{2}\frac{D^- f'(\lambda)}{|f'(\lambda)|}\|x_* - y_k\| + \theta\,\frac{2+f'(\lambda)}{|f'(\lambda)|}\right]\|x_* - y_k\|. \tag{5.5}$$

If, additionally, $0 \leq \theta < \kappa/(4+\kappa)$ *then* $\{y_k\}$ *converges* Q*-linearly as follows*

$$\|x_* - y_{k+1}\| \leq \left[\frac{1+\theta}{2} + \frac{2\theta}{\kappa}\right]\|x_* - y_k\|, \quad k = 0, 1, \ldots. \tag{5.6}$$

Proof. Well definition of the sequences $\{(t_k, \varepsilon_k)\}$ and $\{y_k\}$ as defined in (5.1) follows from the assumptions on θ, (t_0, ε_0), y_0 and the two last inclusions on Proposition 4.5. Moreover, since (5.2) holds for $k = 0$, using the first inclusion in Proposition 4.5 and induction on k, we conclude that (5.2) holds for all k. The first inequality in (5.4) now follows from Proposition 4.5, (5.1) and (5.2) while the first inequality in (5.3) follows from (5.2) and the definition of $K(t, \varepsilon)$ in (4.8).

Direct inspection of the definition of Ω in (4.4) shows that

$$\Omega \subset [0, \lambda) \times [0, \kappa\lambda).$$

Therefore, using (5.2) and the definition of $K(t, \varepsilon)$ we have

$$t_k \in [0, \lambda), \qquad \varepsilon_k \in [0, \kappa\lambda), \qquad y_k \in B(x_0, \lambda), \quad k = 0, 1, \ldots.$$

Using (4.4) and Lemma 4.2 we conclude that $\{t_k\}$ is strictly increasing, $\{\varepsilon_k\}$ is non-decreasing and the second equality in (5.3) holds for all k. Therefore, in view of the first two above inclusions, $\{t_k\}$ and $\{\varepsilon_k\}$ converge, respectively, to some $\tilde{t} \in (0, \lambda]$ and $\tilde{\varepsilon} \in [0, \kappa\lambda]$. Convergence to $\tilde{t}$, together with the first inequality in (5.4) and the inclusion $y_k \in B(x_0, \lambda)$ implies that y_k converges to some $x_* \in B[0, \lambda]$ and that the second inequality on (5.4) holds for all k.

Using the inclusion $y_k \in B(x_0, \lambda)$, the first inequality in Corollary 3.5 and (5.3) we have

$$-f(\|y_k - x_0\|) \leq \left(\frac{1+\theta^2}{2}\right)^k (f(t_0) + \varepsilon_0), \quad k = 0, 1, \ldots.$$

According to (2.9), $f' < -\kappa$ in $[0, \lambda)$. Therefore, since $f(t_*) = 0$ and $t_* < \lambda$,

$$f(t) \leq -\kappa(t - t_*), \quad t_* \leq t < \lambda.$$

Hence, *if* $\|y_k - x_0\| \geq t_*$, we can combine the two above inequalities, setting $t = \|y_k - x_0\|$ in the second, to obtain

$$\|y_k - x_0\| - t_* \leq \left(\frac{1+\theta^2}{2}\right)^k \frac{f(t_0) + \varepsilon_0}{\kappa}.$$

Note that the above inequality remain valid even if $\|y_k - x_0\| < t_*$. Therefore, taking the limit $k \to \infty$ in the above inequality we conclude that $\|x_* - x_0\| \leq t_*$. Moreover, now that we know that x_* is in the interior of the domain of F, we can also take the limit $k \to \infty$ in (5.3) to conclude that $F(x_*) = 0$.

The "classical" version of Kantorovich's theorem on the Newton method for a generic majorant function (see e.g. [6]) guarantee that under the assumptions of Theorem 2.1, F has a unique zero in $B(x_0, \bar{\tau})$. Hence x_* must be this zero of F.

To prove (5.5) and (5.6), first note that from first inclusion in (5.2) we have $\|y_k - x_0\| \leq t_k$, for all $k = 0, 1, \ldots$. Now, since $\tilde{t} \in (0, \lambda]$ we obtain from second inequality in (5.4) that $\|x_* - y_k\| \leq \lambda - t_k$, for all $k = 0, 1, \ldots$. Therefore, using $(\mathbf{h4})'$, $F(x_*) = 0$ and first equality in (5.1), the desire inequalities follows by applying Lemma 3.7. For concluding the proof, note that for $0 \leq \theta < \kappa/(4+\kappa)$ the quantity in the bracket in (5.6) is less than one, which implies that the sequence $\{y_k\}$ converges Q-linearly. □

Proposition 5.2. *If* $0 \le \rho < \beta/2$ *then*

$$\rho < \bar{t}/2 < \bar{t}, \qquad f'(\rho) < 0.$$

Proof. Assumption $\rho < \beta/2$ and Proposition 2.3 item (iii) proves the first two inequalities of the proposition. The last inequality follows from the first inequality and Proposition 2.3 item (i).

Proof of Theorem 2.1. First we will prove Theorem 2.1 with $\rho = 0$ and $z_0 = x_0$. Note that, from the definition in (2.7), we have

$$\kappa_0 = \kappa, \qquad \lambda_0 = \lambda, \qquad \Theta_0 = \Theta.$$

Since

$$(0, 0) \in \Omega, \qquad x_0 \in K(0, 0),$$

using Theorem 5.1 we conclude that Theorem 2.1 holds for $\rho = 0$.

For proving the general case, take

$$0 \le \rho < \beta/2, \qquad z_0 \in B[x_0, \rho]. \tag{5.7}$$

Using Proposition 5.2 and (2.6) we conclude that $\rho < \bar{t}/2$ and $f'(\rho) < 0$. Define

$$g : [0, R - \rho) \to \mathbb{R}, \qquad g(t) = \frac{-1}{f'(\rho)}[f(t + \rho) + 2\rho]. \tag{5.8}$$

We claim that g is a majorant function for F at point z_0. Trivially, $B(z_0, R - \rho) \subset C$, $g'(0) = -1$, $g(0) > 0$. Moreover g' is also convex and strictly increasing. To end the proof that g satisfies **(h1)–(h3)**, using Proposition 2.3 item (iii) and second inequality in (5.7) we have

$$\lim_{t \to \bar{t} - \rho} g(t) = \frac{-1}{f'(\rho)}(2\rho - \beta) < 0.$$

Using Proposition 3.1 we have

$$\|F'(z_0)^{-1}F'(x_0)\| \le \frac{-1}{f'(\rho)}. \tag{5.9}$$

Therefore, using also the second inequality of Corollary 3.5 we have

$$\begin{aligned}
\|F'(z_0)^{-1}F(z_0)\| &\le \|F'(z_0)^{-1}F'(x_0)\| \, \|F'(x_0)^{-1}F(z_0)\| \\
&\le \frac{-1}{f'(\rho)}[f(\|z_0 - x_0\|) + 2\|z_0 - x_0\|].
\end{aligned}$$

As $f' \ge -1$, the function $t \mapsto f(t) + 2t$ is (strictly) increasing. Combining this fact with the above inequality and (5.8) we conclude that

$$\|F'(z_0)^{-1}F'(z_0)\| \le g(0).$$

To end the proof that g is a majorant function for F at z_0, take $x, y \in \mathbb{X}$ such that

$$x, y \in B(z_0, R - \rho), \qquad \|x - z_0\| + \|y - x\| < R - \rho.$$

Hence $x, y \in B(x_0, R)$, $\|x - x_0\| + \|y - x\| < R$ and using (5.9) together with (2.1) we have

$$\begin{aligned}
\|F'(z_0)^{-1}[F'(y) - F'(x)]\| &\le \|F'(z_0)^{-1}F'(x_0)\| \, \|F'(x_0)^{-1}[F'(y) - F'(x)]\| \\
&\le \frac{-1}{f'(\rho)}[f'(\|y - x\| + \|x - x_0\|) - f'(\|x - x_0\|)].
\end{aligned}$$

Since f' is convex, the function $t \mapsto f'(s + t) - f'(s)$ is increasing for $s \ge 0$ and $\|x - x_0\| \le \|x - z_0\| + \|z_0 - x_0\| \le \|x - z_0\| + \rho$,

$$f'(\|y - x\| + \|x - x_0\|) - f'(\|x - x_0\|) \le f'(\|y - x\| + \|x - z_0\| + \rho) - f'(\|x - z_0\| + \rho).$$

Combining the two above inequalities with the definition of g we obtain

$$\|F'(z_0)^{-1}[F'(y)-F'(x)]\| \le g'(\|y-x\|+\|x-z_0\|)-g'(\|x-z_0\|).$$

Note that for κ_ρ, λ_ρ and Θ_ρ as defined in (2.3), we have

$$\kappa_\rho = \sup_{0<t<R-\rho} \frac{-g(t)}{t}, \qquad \lambda_\rho = \sup\{t \in [0, R-\rho) : \kappa_\rho + g'(t) < 0\}, \qquad \Theta_\rho = \frac{\kappa_\rho}{2-\kappa_\rho},$$

which are the same as (2.3) with g instead of f. Therefore, applying Theorem 2.1 for F and the majorant function g at point z_0 and $\rho = 0$, we conclude that the sequence $\{z_k\}$ is well defined, remains in $B(z_0, \lambda_\rho)$, satisfies (2.4) and converges to some $z_* \in B[z_0, t_{*,\rho}]$ which is a zero of F, where $t_{*,\rho}$ is the smallest solution of $g(t) = 0$. Using (5.8) we conclude that $t_{*,\rho}$ is the smallest solution of

$$f(\rho+t)+2\rho = 0.$$

Hence, in view of Proposition 2.3 item (ii), we have $\rho + t_{*,\rho} < \bar{t} \le \bar{\tau}$ and $B[z_0, t_{*,\rho}] \subset B(x_0, \bar{\tau})$. Therefore, z_* is the unique zero of F in $B(x_0, \bar{\tau})$, which we already called x_*. Since

$$g'(t) = f'(t+\rho)/|f'(\rho)|, \qquad D^-g'(t) = D^-f'(t+\rho)/|f'(\rho)|, \quad t \in [0, R-\rho),$$

applying again Theorem 2.1 for F and the majorant function g at point z_0 and $\rho = 0$, we conclude that item **(h4)** also holds. □

6. Special cases

First we use Theorem 2.1 to analyze the convergence of the inexact Newton method with a relative residual error tolerance in the setting of Smale's α-theory. The first application of Kantorovich's Theorem, with exact Newton iterations, for Smale's α-theory were presented Wang and Han [21] (see also [20]). Smale's α-theory for inexact Newton steps with a quadratic error tolerance,

$$z_{k+1} = z_k + S_k, \qquad \|F'(z_0)^{-1}[F(z_k)+F'(z_k)S_k]\| \le \theta_n \|F'(z_0)^{-1}F(z_k)\|^2,$$
$$k = 0, \ldots, \quad \sup \theta_n < 1,$$

were analyzed by Shen and Li in [18]. Up to our knowledge, this is the first time an inexact Newton method with a relative error tolerance is analyzed in this framework.

Theorem 6.1. *Let $\mathbb{X}$ and $\mathbb{Y}$ be Banach spaces, $C \subseteq \mathbb{X}$ and $F : C \to \mathbb{Y}$ a continuous function and analytic $int(C)$. Take $x_0 \in \mathrm{int}(C)$ with $F'(x_0)$ non-singular. Define*

$$\gamma := \sup_{n>1} \left\| \frac{F'(x_0)^{-1}F^{(n)}(x_0)}{n!} \right\|^{1/(n-1)}.$$

Suppose that $B(x_0, 1/\gamma) \subset C$, $b > 0$ and that

$$\|F'(x_0)^{-1}F(x_0)\| \le b, \qquad b\gamma < 3-2\sqrt{2}, \qquad 0 \le \theta \le \frac{1-2\sqrt{\gamma b}-\gamma b}{1+2\sqrt{\gamma b}+\gamma b}.$$

Then, the sequence generated by the inexact Newton method for solving $F(x) = 0$ with starting point x_0 and residual relative error tolerance θ: for $k = 0, 1, \ldots,$

$$x_{k+1} = x_k + S_k, \qquad \|F'(x_0)^{-1}[F(x_k)+F'(x_k)S_k]\| \le \theta \|F'(x_0)^{-1}F(x_k)\|,$$

is well defined, the generated sequence $\{x_k\}$ converges to a point x_* which is a zero of F,

$$\|F'(x_0)^{-1}F(x_k)\| \le \left(\frac{1+\theta^2}{2}\right)^k b, \quad k = 0, 1, \ldots,$$

the sequence $\{x_k\}$ *is contained in* $B(x_0, \lambda)$, $x_* \in B[x_0, t_*]$ *and* x_* *is the unique zero of* F *in* $B(x_0, \bar{\tau})$, *where*

$$\lambda := \frac{b}{\sqrt{\gamma b} + \gamma b},$$

$$t_* = \frac{1 + \gamma b - \sqrt{1 - 6\gamma b + (\gamma b)^2}}{4}, \qquad \bar{\tau} = \frac{1 + \gamma b + \sqrt{1 - 6\gamma b + (\gamma b)^2}}{4}.$$

Moreover, the sequence $\{x_k\}$ *satisfies, for* $k = 0, 1, \ldots,$

$$\|x_* - x_{k+1}\| \le \left[\frac{1+\theta}{2}\frac{D^- f'(\lambda)}{|f'(\lambda)|}\|x_* - x_k\| + \theta\frac{f'(\lambda) + 2}{|f'(\lambda)|}\right]\|x_* - x_k\|.$$

If, additionally, $0 \le \theta < (1 - 2\sqrt{\gamma b} - \gamma b)/(5 - 2\sqrt{\gamma b} - \gamma b)$ *then* $\{x_k\}$ *converges Q-linearly as follows*

$$\|x_* - x_{k+1}\| \le \left[\frac{1+\theta}{2} + \frac{2\theta}{1 - 2\sqrt{\gamma b} - \gamma b}\right]\|x_* - x_k\|, \qquad k = 0, 1, \ldots.$$

Proof. Since the function $f : [0, 1/\gamma) \to \mathbb{R}$

$$f(t) = \frac{t}{1 - \gamma t} - 2t + b,$$

is a majorant function for F in x_0, [6]. Therefore, all results follow from Theorem 2.1, applied to this particular context. □

A semi-local convergence result for Newton method is instrumental in the complexity analysis of linear and quadratic minimization problems by means of self-concordant functions [15]. Kantorovich's theorem, with exact Newton iterations has already been used by Alvarez et al. [1] for the analysis of self-concordant minimization. Also in this setting, Theorem 2.1 provides a semi-local convergence result for Newton method with a relative error tolerance.

Theorem 6.2. *Let* $C \subseteq \mathbb{R}^n$ *be an open convex set and let* $g : C \to \mathbb{R}$ *be an a-self-concordant function with parameter* $a > 0$. *For* $x \in C$, *let*

$$\|v\|_x := \sqrt{v^T g''(x) v}, \quad v \in \mathbb{R}^n,$$

$$W_r(x) := \{z : \|z - x\|_x < r\}, \quad W_r[x] := \{z : \|z - x\|_x \le r\}.$$

Suppose that $x_0 \in C$, $g''(x_0)$ *is non-singular,* $b > 0$

$$\|g''(x_0)^{-1} g'(x_0)\|_{x_0} \le b < 3 - 2\sqrt{2}, \qquad 0 \le \theta \le \frac{1 - 2\sqrt{b} - b}{1 + 2\sqrt{b} + b}.$$

Then the sequence generated by the inexact Newton method for solving $g'(x) = 0$ *with starting point* x_0 *and residual relative error tolerance* θ: *for* $k = 0, 1, \ldots,$

$$x_{k+1} = x_k + S_k, \qquad \|g''(x_0)^{-1}[g'(x_k) + g''(x_k)S_k]\|_{x_0} \le \theta\|g''(x_0)^{-1} g'(x_k)\|_{x_0},$$

is well defined, converges to a point x_* *which is the (unique, global) minimizer of* g,

$$\|g''(x_0)^{-1} g'(x_k)\|_{x_0} \le \left(\frac{1+\theta^2}{2}\right)^k b, \quad k = 0, 1, \ldots,$$

the sequence $\{x_k\}$ *is contained in* $W_\lambda(x_0)$ *and* $x_* \in W_{t_*}(x_0)$, *where*

$$\lambda := \frac{b}{\sqrt{b} + b}, \qquad t_* = \frac{1 + b - \sqrt{1 - 6\,b + b^2}}{4}.$$

Moreover, the sequence $\{x_k\}$ satisfies, for $k = 0, 1, \ldots,$

$$\|x_* - x_{k+1}\| \leq \left[\frac{1+\theta}{2}\frac{D^-f'(\lambda)}{|f'(\lambda)|}\|x_* - x_k\| + \theta\frac{f'(\lambda)+2}{|f'(\lambda)|}\right]\|x_* - x_k\|.$$

If, additionally, $0 \leq \theta < (1 - 2\sqrt{b} - b)/(5 - 2\sqrt{b} - b)$ then $\{x_k\}$ converges Q-linearly as follows

$$\|x_* - x_{k+1}\| \leq \left[\frac{1+\theta}{2} + \frac{2\theta}{1 - 2\sqrt{b} - b}\right]\|x_* - x_k\|, \quad k = 0, 1, \ldots.$$

Proof. The scalar function $f : [0, 1) \to \mathbb{R}$ defined by

$$f(t) = \frac{t}{1-t} - 2t + b,$$

is a majorant function for g' in x_0, [6]. Therefore, the proof follows from Theorem 2.1, applied to this particular context. □

Theorem 6.3. *Let $\mathbb{X}$ and $\mathbb{Y}$ be a Banach spaces, $C \subseteq \mathbb{X}$ and $F : C \to \mathbb{Y}$ a continuous function, continuously differentiable on $int(C)$. Take $x_0 \in \mathrm{int}(C)$ with $F'(x_0)$ non-singular. Suppose that exist constants $L > 0$ and $b > 0$ such that $bL < 1/2$, $B(x_0, 1/L) \subset C$ and*

$$\|F'(x_0)^{-1}[F'(y) - F'(x)]\| \leq L\|x - y\|, \quad x, y \in B(x_0, 1/L),$$

$$\|F'(x_0)^{-1}F(x_0)\| \leq b, \qquad 0 \leq \theta \leq \frac{1 - \sqrt{2bL}}{1 + \sqrt{2bL}}.$$

Then, the sequence generated by the inexact Newton method for solving $F(x) = 0$ with starting point x_0 and residual relative error tolerance θ: for $k = 0, 1, \ldots,$

$$x_{k+1} = x_k + S_k, \qquad \|F'(x_0)^{-1}[F(x_k) + F'(x_k)S_k]\| \leq \theta\|F'(x_0)^{-1}F(x_k)\|,$$

is well defined,

$$\|F'(x_0)^{-1}F(x_k)\| \leq \left(\frac{1+\theta^2}{2}\right)^k b, \quad k = 0, 1, \ldots$$

the sequence $\{x_k\}$ is contained in $B(x_0, \lambda)$, converges to a point $x_* \in B[x_0, t_*]$ which is the unique zero of F in $B(x_0, 1/L)$ where

$$\lambda := \frac{\sqrt{2bL}}{L}, \qquad t_* = \frac{1 - \sqrt{1 - 2Lb}}{L}.$$

Moreover, the sequence $\{x_k\}$ satisfies, for $k = 0, 1, \ldots,$

$$\|x_* - z_{k+1}\| \leq \left[\frac{1+\theta}{2}\frac{L}{1 - \sqrt{2bL}}\|x_* - x_k\| + \theta\frac{1 + \sqrt{2bL}}{1 - \sqrt{2bL}}\right]\|x_* - x_k\|.$$

If, additionally, $0 \leq \theta < (1 - \sqrt{2bL})/(5 - \sqrt{2bL})$ then the sequence $\{x_k\}$ converges Q-linearly as follows

$$\|x_* - x_{k+1}\| \leq \left[\frac{1+\theta}{2} + \frac{2\theta}{1 - \sqrt{2bL}}\right]\|x_* - x_k\|, \quad k = 0, 1, \ldots.$$

Proof. Since the function $f : [0, 1/L) \to \mathbb{R}$,

$$f(t) := \frac{L}{2}t^2 - t + b,$$

is a majorant function for F at point x_0, all result follow from Theorem 2.1, applied to this particular context. □

Acknowledgments

The first author was partially supported by CAPES, CNPq Grant 303732/2011-3, PRONEX-Optimization (FAPERJ/CNPq) and IMPA. The second author was partially supported by CNPq grants 303583/2008-8, 474945/2010-7, FAPERJ grant E-26/110.295/2009 and by PRONEX-Optimization (FAPERJ/CNPq).

References

[1] F. Alvarez, J. Bolte, J. Munier, A unifying local convergence result for Newton's method in Riemannian manifolds, Found. Comput. Math. 8 (2) (2008) 197–226.
[2] L. Blum, F. Cucker, M. Shub, S. Smale, Complexity and Real Computation, Springer-Verlag, New York, 1998, With a foreword by Richard M. Karp.
[3] J. Chen, W. Li, Convergence behaviour of inexact Newton methods under weak Lipschitz condition, J. Comput. Appl. Math. 191 (1) (2006) 143–164.
[4] R.S. Dembo, S.C. Eisenstat, T. Steihaug, Inexact Newton methods, SIAM J. Numer. Anal. 19 (2) (1982) 400–408.
[5] O.P. Ferreira, B.F. Svaiter, Kantorovich's theorem on Newton's method in Riemannian manifolds, J. Complexity 18 (1) (2002) 304–329.
[6] O.P. Ferreira, B.F. Svaiter, Kantorovich's majorants principle for Newton's method, Comput. Optim. Appl. 42 (2) (2009) 213–229.
[7] J.-B. Hiriart-Urruty, C. Lemaréchal, Convex Analysis and Minimization Algorithms. I, in: Grundlehren der Mathematischen Wissenschaften Fundamental Principles of Mathematical Sciences, vol. 305, Springer-Verlag, Berlin, 1993, Fundamentals.
[8] L.V. Kantorovič, The principle of the majorant and Newton's method, Doklady Akad. Nauk SSSR (N.S.) 76 (1951) 17–20.
[9] L.V. Kantorovich, G.P. Akilov, Functional analysis in normed spaces, in: A.P. Robertson. (Ed.), in: International Series of Monographs in Pure and Applied Mathematics, vol. 46, The Macmillan Co, New York, 1964, Translated from the Russian by D.E. Brown.
[10] C. Li, J. Wang, Newton's method on Riemannian manifolds: Smale's point estimate theory under the γ-condition, IMA J. Numer. Anal. 26 (2) (2006) 228–251.
[11] C. Li, J. Wang, Newton's method for sections on Riemannian manifolds: generalized covariant α-theory, J. Complexity 24 (3) (2008) 423–451.
[12] C. Li, J.-H. Wang, J.-P. Dedieu, Smale's point estimate theory for Newton's method on Lie groups, J. Complexity 25 (2) (2009) 128–151.
[13] I. Moret, A Kantorovich-type theorem for inexact Newton methods, Numer. Funct. Anal. Optim. 10 (3–4) (1989) 351–365.
[14] B. Morini, Convergence behaviour of inexact Newton methods, Math. Comp. 68 (228) (1999) 1605–1613.
[15] Y. Nesterov, A. Nemirovskii, Interior-Point Polynomial Algorithms in Convex Programming, in: SIAM Studies in Applied Mathematics, vol. 13, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.
[16] F.A. Potra, The Kantorovich theorem and interior point methods, Math. Program. 102 (1, Ser. A) (2005) 47–70.
[17] W. Shen, C. Li, Kantorovich-type convergence criterion for inexact Newton methods, Appl. Numer. Math. 59 (7) (2009) 1599–1611.
[18] W. Shen, C. Li, Smale's α-theory for inexact Newton methods under the γ-condition, J. Math. Anal. Appl. 369 (1) (2010) 29–42.
[19] S. Smale, Newton's method estimates from data at one point, in: The Merging of Disciplines: New Directions in Pure, Applied, and Computational Mathematics, Laramie, Wyo., 1985, Springer, New York, 1986, pp. 185–196.
[20] X. Wang, Convergence of Newton's method and inverse function theorem in Banach space, Math. Comp. 68 (225) (1999) 169–186.
[21] X.H. Wang, D.F. Han, On dominating sequence method in the point estimate and Smale theorem, Sci. China Ser. A 33 (2) (1990) 135–144.
[22] J.-H. Wang, S. Huang, C. Li, Extended Newton's method for mappings on Riemannian manifolds with values in a cone, Taiwanese J. Math. 13 (2B) (2009) 633–656.
[23] J.-h. Wang, C. Li, Uniqueness of the singular points of vector fields on Riemannian manifolds under the γ-condition, J. Complexity 22 (4) (2006) 533–548.
[24] J.-H. Wang, C. Li, Kantorovich's theorem for newton's method on lie groups, J. Zhejiang Univ. Sci. A 8 (6) (2007) 978–986. cited By (since 1996) 0.
[25] T.J. Ypma, Local convergence of inexact Newton methods, SIAM J. Numer. Anal. 21 (3) (1984) 583–590.
[26] P.P. Zabrejko, D.F. Nguen, The majorant method in the theory of Newton–Kantorovich approximations and the Pták error estimates, Numer. Funct. Anal. Optim. 9 (5–6) (1987) 671–684.