

Local convergence of Newton's method in Banach space from the viewpoint of the majorant principle

ORIZON P. FERREIRA[†]

*Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus II,
Caixa Postal 131, CEP 74001-970, Goiânia, GO, Brazil*

[Received on 25 July 2007; revised on 26 April 2008]

A local convergence analysis of Newton's method for solving nonlinear equations, based on Kantorovich's majorant principle, is presented in this paper. This analysis provides a clear relationship between the majorant function, which relaxes the Lipschitz continuity of the derivative, and the nonlinear operator under consideration. It also allows us to obtain the optimal convergence radius, the biggest range for the uniqueness of the solution, and to unify some previous and unrelated results.

Keywords: Newton's method; majorant principle; local convergence; Banach space.

1. Introduction

Newton's method and its variations are the most efficient methods known for solving nonlinear equations, searching for a local minimizer of a function and many other applications. One of the most important applications of Newton's method allows us to develop polynomial time algorithms in convex programming, see [Nesterov & Nemirovskii \(1994\)](#). Besides its practical applications, Newton's method is also a powerful theoretical tool, having a wide range of applications in pure mathematics (see [Blum *et al.*, 1997](#); [Krantz & Parks, 2002](#); [Moser, 1961](#); [Nash, 1956](#); [Wayne, 1996](#)). For a current historical perspective of Newton's method (see [Polyak, 2007](#)).

Newton's method may fail to converge and may even fail to generate an infinite sequence when a singular derivative point of a nonlinear operator, which defines the nonlinear equation, is reached. To ensure that the method is well defined that the generated sequence converges to a solution of the nonlinear equation and that the convergence rate is quadratic, some conditions must be imposed. For instance, the classical convergence analysis requires the initial iterate to be 'close enough' to a solution and the first derivative of the nonlinear operator to be nonsingular in this solution. Moreover, Lipschitz continuity of the first derivative is assumed (see, e.g. [Dennis & Schnabel, 1996](#); [Nesterov & Nemirovskii, 1994](#); [Ortega, 1990](#); [Ortega & Rheinboldt, 1970](#); [Ostrowski, 1996](#); [Rall, 1974](#); [Smale, 1986](#); [Traub & Wozniakowski, 1979](#)). One drawback of this analysis is that closeness to a solution, and consequently existence, must be known or given *a priori*. On the other hand, this analysis has the advantage of giving the optimal convergence radius with respect to the Lipschitz constant (see [Rall, 1974](#); [Traub & Wozniakowski, 1979](#); [Wang, 2000](#)).

One of the usual hypotheses used in convergence analysis of Newton's method is that the Lipschitz continuity of the derivative of the nonlinear operator in question or something like Lipschitz continuity is critical; that is, keeping control of the derivative is an important point in the convergence analysis of Newton's method. In the last few years, a couple of papers have dealt with the issue of convergence

[†]Email: orizon@mat.ufg.br

of Newton's method by relaxing the assumption of Lipschitz continuity of the derivative. Besides improving the convergence theory, these new modifications of the Lipschitz condition permit us to unify several unrelated previous results; works dealing with this subject include [Alvarez et al. \(2008\)](#), [Ferreira & Svaiter \(2007\)](#) and [Wang \(1999, 2000\)](#).

The aim of this paper is to present a new local convergence analysis for Newton's method based on *Kantorovich's majorant principle*, introduced by [Kantorovich \(1951\)](#), see also [Ferreira & Svaiter \(2007\)](#). In our analysis, the classical Lipschitz condition is relaxed using a majorant function. It is worth pointing out that this condition is equivalent to Wang's condition, introduced in [Wang \(2000\)](#). The analysis presented provides a clear relationship between the majorant function and the nonlinear operator under consideration. Also, as in [Wang \(2000\)](#), it allows us to obtain the biggest range for the uniqueness of the solution and the optimal convergence radius for the method with respect to the majorant function. Moreover, several unrelated previous results pertaining to Newton's method are unified.

The organization of the paper is as follows: In Section 1.1, some notation and basic results used in the paper are presented. In Section 2, the main result is stated and in Section 2.1 some properties of the majorant function are established. In Section 2.2, some relationships between the majorant function and the nonlinear operator are presented. In Section 2.3, the uniqueness of the solution and the optimal convergence radius are obtained. In Section 2.4, the main result is proved and some applications of this result are given in Section 3. Some final remarks are made in Section 4.

1.1 Notation and auxiliary results

The following notation and results are used throughout our presentation. Let X and Y be Banach spaces. The open and closed balls at x are denoted, respectively, by

$$B(x, \delta) = \{y \in X; \|x - y\| < \delta\} \quad \text{and} \quad B[x, \delta] = \{y \in X; \|x - y\| \leq \delta\}.$$

Let $\Omega \subseteq X$. The Fréchet derivative of $F : \Omega \rightarrow Y$ at $x \in \text{int}(\Omega)$ is the linear map $F'(x) : X \rightarrow Y$.

LEMMA 1.1 (Banach's lemma) Let $B : X \rightarrow X$ be bounded linear operator. If $I : X \rightarrow X$ is the identity operator and $\|B - I\| < 1$, then B is invertible and $\|B^{-1}\| \leq 1/(1 - \|B - I\|)$.

Proof. Take $A = I$ and $c = \|B - I\|$ in Lemma 1 of [Smale \(1986, p. 189\)](#). □

LEMMA 1.2 If $0 \leq t < 1$, then $\sum_{i=0}^{\infty} (i + 2)(i + 1)t^i = 2/(1 - t)^3$.

Proof. Take $k = 2$ in Lemma 3 of [Blum et al. \(1997, p. 161\)](#). □

PROPOSITION 1.3 Let $R > 0$. If $\varphi : [0, R) \rightarrow \mathbb{R}$ is continuously differentiable and convex, then

- (1) $(\varphi(t) - \varphi(\tau t))/t \leq \varphi'(t)(1 - \tau)$ for all $t \in (0, R)$ and $0 \leq \tau \leq 1$;
- (2) $(\varphi(u) - \varphi(\tau u))/u \leq (\varphi(v) - \varphi(\tau v))/v$ for all $u, v \in [0, R)$, $u < v$ and $0 \leq \tau \leq 1$.

Proof. See Theorem 4.1.1 and Remark 4.1.2 of [Hiriart-Urruty & Lemaréchal \(1993, p. 21\)](#). □

2. Local analysis for Newton's method

Our goal is to state and prove a local theorem for Newton's method. First, we will prove some results regarding the scalar majorant function, which relaxes the Lipschitz condition. Then we will show that Newton's method is well defined and converges. We will also prove the uniqueness of the solution in a suitable region, and the convergence rate and optimal ball of convergence will be established. The statement of the theorem is as follows.

THEOREM 2.1 Let X and Y be Banach spaces, $\Omega \subseteq X$ an open set and $F : \Omega \rightarrow Y$ a continuously differentiable function. Let $x_* \in \Omega$, $R > 0$ and $\kappa := \sup\{t \in [0, R) : B(x_*, t) \subset \Omega\}$. Suppose that $F(x_*) = 0$, $F'(x_*)$ is invertible and there exists an $f : [0, R) \rightarrow \mathbb{R}$ twice continuously differentiable such that

$$\|F'(x_*)^{-1}[F'(x) - F'(x_* + \tau(x - x_*))]\| \leq f'(\|x - x_*\|) - f'(\tau\|x - x_*\|) \quad (2.1)$$

for all $\tau \in [0, 1]$, $x \in B(x_*, \kappa)$ and

- (h1) $f(0) = 0$ and $f'(0) = -1$,
- (h2) f' is convex and strictly increasing.

Let $\nu := \sup\{t \in [0, R) : f'(t) < 0\}$, $\rho := \sup\{t \in (0, \nu) : f(t)/(tf'(t)) - 1 < 1\}$ and

$$r := \min\{\kappa, \rho\}.$$

Then the sequences with starting points $x_0 \in B(x_*, r) \setminus \{x_*\}$ and $t_0 = \|x_* - x_0\|$, respectively, namely

$$x_{k+1} = x_k - F'(x_k)^{-1}F(x_k), \quad t_{k+1} = |t_k - f(t_k)/f'(t_k)|, \quad k = 0, 1, \dots, \quad (2.2)$$

are well defined; $\{t_k\}$ is strictly decreasing, is contained in $(0, r)$ and converges to 0 and $\{x_k\}$ is contained in $B(x_*, r)$ and converges to the point x_* which is the unique zero of F in $B(x_*, \sigma)$, where $\sigma := \sup\{0 < t < \kappa : f(t) < 0\}$. Moreover, $\{t_{k+1}/t_k^2\}$ is strictly decreasing,

$$\|x_* - x_{k+1}\| \leq [t_{k+1}/t_k^2]\|x_k - x_*\|^2, \quad t_{k+1}/t_k^2 \leq f''(t_0)/(2|f'(t_0)|), \quad k = 0, 1, \dots \quad (2.3)$$

If, additionally, $f(\rho)/(f'(\rho)) - 1 = 1$ and $\rho < \kappa$, then $r = \rho$ is the best possible convergence radius.

REMARK 2.2 Combining both inequalities in (2.3), we conclude that $\{x_k\}$ converges Q -quadratically to x_* . Moreover, as $\{t_{k+1}/t_k^2\}$ is strictly decreasing, we have $t_{k+1}/t_k^2 < t_1/t_0^2$ for $k = 0, 1, \dots$. So the first inequality in (2.3) implies that $\|x_* - x_{k+1}\| \leq [t_1/t_0^2]\|x_k - x_*\|^2$ for $k = 0, 1, \dots$. As a consequence,

$$\|x_* - x_k\| \leq t_0(t_1/t_0)^{2^k-1}, \quad k = 0, 1, \dots$$

REMARK 2.3 Assumption (2.1) is crucial to our analysis. It is worth pointing out that, under appropriate regularity conditions on the nonlinear operator F , assumption (2.1) always holds in a suitable neighbourhood of x_* . For instance, if F is twice continuously differentiable, then the majorant function $f : [0, \kappa) \rightarrow \mathbb{R}$ defined by $f(t) = Kt^2/2 - t$, where $K = \sup\{\|F'(x_*)^{-1}F''(x)\| : x \in B(x_*, \kappa)\}$, satisfies assumption (2.1). Estimating the constant K is a very difficult problem. Therefore, the goal is to identify classes of nonlinear operators for which it is possible to obtain a majorant function. We will give some examples of such classes in Section 3.

Now, we will give some examples of majorant functions satisfying the conditions (h1) and (h2).

EXAMPLE 2.4 The following functions satisfy the conditions (h1) and (h2) (see also Section 3):

- (i) $f : \mathbb{R} \rightarrow \mathbb{R}$ such that $f(t) = e^t - 2t - 1$;
- (ii) $f : [0, 1) \rightarrow \mathbb{R}$ such that $f(t) = -\ln(1-t) - 2t$ and
- (iii) $f : [0, 1) \rightarrow \mathbb{R}$ such that $f(t) = (1-t)\ln(1-t)$.

From now on, we assume that the hypotheses of Theorem 2.1 hold.

2.1 The majorant function

In this section, we will prove that the constants κ associated to Ω , ν , ρ and σ associated to the majorant function f are positive. We will also prove the statements in Theorem 2.1 involving the sequence $\{t_k\}$. First, we will prove that κ , ν and σ are positive.

PROPOSITION 2.5 The constants κ , ν and σ are positive and $t - f(t)/f'(t) < 0$ for all $t \in (0, \nu)$.

Proof. Since Ω is open and $x_* \in \Omega$, we can immediately conclude that $\kappa > 0$. As $f'(0) = -1$, there exists a $\delta > 0$ such that $f'(t) < 0$ for all $t \in (0, \delta)$. So, $\nu > 0$. Now, because $f(0) = 0$ and $f'(0) = -1$, there exists a $\delta > 0$ such that $f(t) < 0$ for all $t \in (0, \delta)$. Hence $\sigma > 0$.

It remains to show that $t - f(t)/f'(t) < 0$ for all $t \in (0, \nu)$. Since f' is strictly increasing, f is strictly convex. So

$$0 = f(0) > f(t) - tf'(t), \quad t \in (0, R).$$

If $t \in (0, \nu)$, then $f'(t) < 0$, which combined with last inequality yields the desired inequality. □

According to (h2) and the definition of ν , we have $f'(t) < 0$ for all $t \in [0, \nu)$. Therefore, the Newton iteration map for f is well defined in $[0, \nu)$. Let us call it n_f :

$$\begin{aligned} n_f : [0, \nu) &\rightarrow (-\infty, 0], \\ t &\mapsto t - f(t)/f'(t). \end{aligned} \tag{2.4}$$

PROPOSITION 2.6 The map $(0, \nu) \ni t \mapsto |n_f(t)|/t^2$ is strictly increasing and satisfies

$$|n_f(t)|/t^2 \leq f''(t)/(2|f'(t)|), \quad t \in (0, \nu).$$

Proof. Because $f(0) = 0$ and $f' < 0$ in $[0, \nu)$, we have, after simple manipulations, that

$$\frac{|n_f(t)|}{t^2} = \frac{1}{|f'(t)|} \int_0^1 \frac{f'(t) - f'(\tau t)}{t} d\tau, \quad t \in (0, \nu). \tag{2.5}$$

Using (h2) and statement (2) of Proposition 1.3, we conclude that the map $(0, \nu) \ni t \mapsto (f'(t) - f'(\tau t))/t$ is positive and strictly increasing for all $\tau \in (0, 1)$. Thus, the integral in (2.5) is positive and strictly increasing. Now, $f' < 0$ and strictly increasing in $[0, \nu)$, and hence the map $[0, \nu) \ni t \mapsto 1/|f'(t)|$ is strictly increasing. Therefore, both terms in the right-hand side of (2.5) are positive and increasing. So the first statement follows.

Part (1) of Proposition 1.3 implies that $(f'(t) - f'(\tau t))/t \leq f''(t)(1 - \tau)$, for all $0 \leq \tau \leq 1$, which substituting in (2.5) and performing the integral gives the desired inequality. □

PROPOSITION 2.7 The constant ρ is positive. As a consequence, $|n_f(t)| < t$ for all $t \in (0, \rho)$.

Proof. Using definition (2.4) and Proposition 2.5, simple algebraic manipulation gives

$$0 < f(t)/(tf'(t)) - 1 = |n_f(t)|/t, \quad t \in (0, \nu). \tag{2.6}$$

On the other hand, Proposition 2.6 implies that $|n_f(t)|/t^2$ is bounded near zero. So we obtain

$$\lim_{t \rightarrow 0} |n_f(t)|/t = \lim_{t \rightarrow 0} (|n_f(t)|/t^2)t = 0.$$

Combining (2.6) with the last equation, we conclude that there exists a $\delta > 0$ such that

$$0 < f(t)/(tf'(t)) - 1 < 1, \quad t \in (0, \delta).$$

Therefore, ρ is positive.

Proposition 2.6 implies, in particular, that the map $(0, \nu) \ni t \mapsto |n_f(t)|/t$ is strictly increasing. Thus, (2.6) and the definition of ρ imply that $|n_f(t)|/t = f(t)/(tf'(t)) - 1 < 1$, for all $t \in (0, \rho)$, as required. \square

Using (2.4), it is easy to see that the sequence $\{t_k\}$ is equivalently defined as

$$t_0 = \|x_* - x_0\|, \quad t_{k+1} = |n_f(t_k)|, \quad k = 0, 1, \dots \quad (2.7)$$

COROLLARY 2.8 The sequence $\{t_k\}$ is well defined, strictly decreasing and contained in $(0, \rho)$. Moreover, $\{t_{k+1}/t_k^2\}$ is strictly decreasing, $\{t_k\}$ converges to 0 and

$$t_{k+1}/t_k^2 \leq [f''(t_0)/(2|f'(t_0)|)], \quad k = 0, 1, \dots$$

Proof. Since $0 < t_0 = \|x_* - x_0\| < r \leq \rho$, using Proposition 2.7 and (2.7) it is simple to conclude that $\{t_k\}$ is well defined, strictly decreasing and contained in $(0, \rho)$. So we have proved the first part of the corollary.

Because $\{t_k\}$ is strictly decreasing, it follows from Proposition 2.6 that

$$t_{k+1}/t_k^2 = |n_f(t_k)|/t_k^2, \quad k = 0, 1, \dots,$$

is also strictly decreasing and, in particular,

$$t_{k+1}/t_k^2 \leq |n_f(t_0)|/t_0^2, \quad k = 0, 1, \dots \quad (2.8)$$

Again, as $\{t_k\}$ is strictly decreasing, we have $t_k < t_0$ for all k . Hence, using (2.8) we obtain

$$t_{k+1} \leq [|n_f(t_0)|/t_0]t_k, \quad k = 0, 1, \dots$$

In view of $t_0 = \|x_* - x_0\| < r \leq \rho$, Proposition 2.7 implies that $(|n_f(t_0)|/t_0) < 1$. So the above inequality implies that $\{t_k\}$ converges to 0. For the last statement, note that the second part of Proposition 2.6 implies that $|n_f(t_0)|/t_0^2 \leq f''(t_0)/(2|f'(t_0)|)$, which, combined with (2.8), yields the desired inequality. \square

2.2 Relationship between the majorant function and the nonlinear operator

In this section, we will present the main relationships between the majorant function f and the nonlinear operator F .

LEMMA 2.9 If $\|x - x_*\| < \min\{\kappa, \nu\}$, then $F'(x)$ is invertible and

$$\|F'(x)^{-1}F'(x_*)\| \leq 1/|f'(\|x - x_*\|)|.$$

In particular, F' is invertible in $B(x_*, r)$.

Proof. Since $\|x - x_*\| < \nu$, we have $f'(\|x - x_*\|) < 0$, which, together with (2.1) and (h1), implies that

$$\|F'(x_*)^{-1}F'(x) - I\| = \|F'(x_*)^{-1}[F'(x) - F'(x_*)]\| \leq f'(\|x - x_*\|) - f'(0) < -f'(0) = 1.$$

Thus, applying Lemma 1.1, we conclude that $F'(x_*)^{-1}F'(x)$ is invertible, which implies that $F'(x)$ is also invertible. Moreover,

$$\|F'(x)^{-1}F'(x_*)\| \leq \frac{1}{1 - \|F'(x_*)^{-1}F'(x) - I\|} \leq \frac{1}{1 - (f'(\|x - x_*\|) - f'(0))} = \frac{1}{|f'(\|x - x_*\|)|}.$$

We see that the last result follows by noting that $r \leq \min\{\kappa, \nu\}$. □

Newton iteration at a point happens to be a zero of the linearization of F at such a point. So we study the linearization error at a point in Ω :

$$E_F(x, y) := F(y) - [F(x) + F'(x)(y - x)], \quad y, x \in \Omega. \tag{2.9}$$

We will bound this error by the error in the linearization of the majorant function f :

$$e_f(t, u) := f(u) - [f(t) + f'(t)(u - t)], \quad t, u \in [0, R]. \tag{2.10}$$

LEMMA 2.10 If $\|x_* - x\| < \kappa$, then $\|F'(x_*)^{-1}E_F(x, x_*)\| \leq e_f(\|x - x_*\|, 0)$.

Proof. Since $B(x_*, \kappa)$ is convex, $x_* + (1 - u)(x - x_*) \in B(x_*, \kappa)$ for all $0 \leq u \leq 1$. Thus, since F is continuously differentiable in Ω , the definition of E_F and some simple manipulation yield

$$\|F'(x_*)^{-1}E_F(x, x_*)\| \leq \int_0^1 \|F'(x_*)^{-1}[F'(x) - F'(x_* + (1 - u)(x - x_*))]\| \|x_* - x\| du,$$

which, combined with assumption (2.1), gives

$$\|F'(x_*)^{-1}E_F(x, x_*)\| \leq \int_0^1 [f'(\|x - x_*\|) - f'((1 - u)\|x - x_*\|)] \|x - x_*\| du.$$

Performing the above integral and using the definition of e_f , the statement follows. □

Lemma 2.9 guarantees, in particular, that F' is invertible in $B(x_*, r)$ and consequently, the Newton iteration map is well defined. Let us call N_F , the Newton iteration map for F in that region:

$$\begin{aligned} N_F &: B(x_*, r) \rightarrow Y, \\ x &\mapsto x - F'(x)^{-1}F(x). \end{aligned} \tag{2.11}$$

One can apply a *single* Newton iteration on any $x \in B(x_*, r)$ to obtain $N_F(x)$ which may not belong to $B(x_*, r)$ or may not even belong to the domain of F . So this is enough to guarantee the well definedness of only one iteration. To ensure that Newtonian iterations may be repeated indefinitely, we need some additional results.

LEMMA 2.11 Take $0 < t < r$. If $\|x - x_*\| \leq t$, then $\|N_F(x) - x_*\| \leq [n_f(t)/t^2]\|x - x_*\|^2$.

Proof. Since $F(x_*) = 0$, the inequality is trivial for $x = x_*$. Now assume that $0 < \|x - x_*\| \leq t$. Lemma 2.9 implies that $F'(x)$ is invertible. Thus, because $F(x_*) = 0$, direct manipulation yields

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

Using the above equation and Lemmas 2.9 and 2.10, it is easy to conclude that

$$\|x_* - N_F(x)\| \leq \| -F'(x)^{-1}F'(x_*) \| \|F'(x_*)^{-1}E_F(x, x_*)\| \leq e_f(\|x - x_*\|, 0)/|f'(\|x - x_*\|)|. \quad (2.12)$$

On the other hand, taking into account that $f(0) = 0$, the definitions of e_f and n_f imply that

$$e_f(\|x - x_*\|, 0)/|f'(\|x - x_*\|)| = -n_f(\|x - x_*\|) = |n_f(\|x - x_*\|)|.$$

As $\|x - x_*\| \leq t$, Proposition 2.6 gives $|n_f(\|x - x_*\|)|/\|x - x_*\|^2 \leq |n_f(t)|/t^2$. So the last equality becomes

$$e_f(\|x - x_*\|, 0)/|f'(\|x - x_*\|)| \leq [|n_f(t)|/t^2]\|x - x_*\|^2.$$

Hence, the desired inequality follows by combining (2.12) and the latter equation. \square

COROLLARY 2.12 If $0 < t < r$, then $N_F(B[x_*, t]) \subset B[x_*, |n_f(t)|]$. Moreover, $N_F(B(x_*, r)) \subset B(x_*, r)$.

Proof. Take $x \in B[x_*, t]$. Since $\|x - x_*\|/t \leq 1$, Lemma 2.11 implies that $\|N_F(x) - x_*\| \leq |n_f(t)|$ and the first inclusion follows. Because $r \leq \rho$, Proposition 2.7 implies that $|n_f(t)| \leq t$. So the last inclusion is an immediate consequence of the first one. \square

2.3 Uniqueness and optimal convergence radius

In this section, we will obtain the uniqueness of the solution and the optimal convergence radius.

LEMMA 2.13 Take $t \in (0, \kappa)$. If $f(t) < 0$, i.e. 0 is the unique zero of f in $[0, t]$, then x_* is the unique zero of F in $B[x_*, t]$. As a consequence, x_* is the unique zero of F in $B(x_*, \sigma)$.

Proof. Assume that $y \in B[x_*, t]$ and $F(y) = 0$. Since $F(x_*) = 0$ and $F(y) = 0$, we have

$$y - x_* = - \int_0^1 F'(x_*)^{-1}[F'(x_* + u(y - x_*)) - F'(x_*)](y - x_*)du.$$

Using (2.1) with $x = x_* + u(y - x_*)$ and $\tau = 0$, it is easy to conclude from the last equality that

$$\|y - x_*\| \leq \int_0^1 [f'(u\|y - x_*\|) - f'(0)]\|y - x_*\|du = f(\|y - x_*\|) - f(0) - f'(0)\|y - x_*\|.$$

Taking into account that $f(0) = 0$ and $f'(0) = -1$, the latter inequality becomes

$$f(\|y - x_*\|) \geq 0.$$

Now, since f is strictly convex and $f(t) < 0$, we shall have $f < 0$ in $(0, t]$; i.e. 0 is the unique zero of f in $[0, t]$. Hence, the above inequality implies $\|y - x_*\| = 0$, i.e. $y = x_*$. So x_* is the unique zero of F in $B[x_*, t]$. The second part follows from the definition of σ . \square

REMARK 2.14 Note that in the above lemma, we have used the fact that condition (2.1) holds only for $\tau = 0$.

LEMMA 2.15 If $f(\rho)/(\rho f'(\rho)) - 1 = 1$ and $\rho < \kappa$, then $r = \rho$ is the optimal convergence radius.

Proof. Assume that $f(\rho)/(\rho f'(\rho)) - 1 = 1$ and $\rho < \kappa$. Define the function $h : (-\kappa, \kappa) \rightarrow \mathbb{R}$ by

$$h(t) = \begin{cases} -f(-t), & t \in (-\kappa, 0], \\ f(t), & t \in [0, \kappa). \end{cases} \tag{2.13}$$

It is straightforward to show that $h(0) = 0$, $h'(0) = -1$ and

$$|h'(0)^{-1}[h'(t) - h'(\tau t)]| \leq f'(|t|) - f'(\tau|t|), \quad \tau \in [0, 1], \quad t \in (-\kappa, \kappa).$$

So $F = h$, $X = Y = \mathbb{R}$, $\Omega = (-\kappa, \kappa)$ and $x_* = 0$ satisfy all the assumptions of Theorem 2.1. Thus, since $\rho < \kappa$, it suffices to show that the Newton method applied for solving $h(t) = 0$, with starting point $t_0 = -\rho$, does not converge. Given $f(\rho)/(\rho f'(\rho)) - 1 = 1$, the definition of h in (2.13) yields

$$t_1 = -\rho - h(-\rho)/h'(-\rho) = -\rho + f(\rho)/f'(\rho) = [f(\rho)/(\rho f'(\rho)) - 1]\rho = \rho.$$

Again, the definition of h in (2.13) and assumption $f(\rho)/(\rho f'(\rho)) - 1 = 1$ give

$$t_2 = \rho - h(\rho)/h'(\rho) = \rho - f(\rho)/f'(\rho) = -[f(\rho)/(\rho f'(\rho)) - 1]\rho = -\rho.$$

Therefore, Newton's method for solving $h(t) = 0$, with starting point $t_0 = -\rho$, produces the cycle

$$t_0 = -\rho, \quad t_1 = \rho, \quad t_2 = -\rho, \dots;$$

as a consequence, it does not converge. Therefore, the lemma is proved. □

2.4 Proof of Theorem 2.1

First, note that the first equation in (2.2) together with (2.11) implies that the sequence $\{x_k\}$ satisfies

$$x_{k+1} = N_F(x_k), \quad k = 0, 1, \dots, \tag{2.14}$$

which is indeed an equivalent definition of this sequence.

Proof. All the statements involving $\{t_k\}$ were proved in Corollary 2.8. As $x_0 \in B(x_*, r)$ and $r \leq \nu$, combining (2.14), inclusion $N_F(B(x_*, r)) \subset B(x_*, r)$ in Corollary 2.12 and Lemma 2.9, it is easy to conclude that $\{x_k\}$ is well defined and remains in $B(x_*, r)$.

We are going to prove that $\{x_k\}$ converges towards x_* . First, we will prove by induction that $\{t_k\}$ is a majorant sequence for $\{x_k\}$, i.e.

$$\|x_* - x_k\| \leq t_k, \quad k = 0, 1, \dots \tag{2.15}$$

Because $t_0 = \|x_* - x_0\|$, the above inequality holds for $k = 0$. Now, assume that $\|x_* - x_k\| \leq t_k$. First, note that from Corollary 2.8, we have $\{t_k\} \subset (0, r)$. Thus, using the first inclusion in Corollary 2.12, (2.7) and (2.14), we obtain that

$$\|x_* - x_{k+1}\| = \|x_* - N_F(x_k)\| \leq |n_f(t_k)| = t_{k+1},$$

and the proof by induction is complete. Since $\{t_k\}$ converges to 0, it follows from (2.15) that $\{x_k\}$ converges to x_* .

Now we will show the first inequality in (2.3). Note that from (2.15) and Lemma 2.11, we have

$$\|x_* - x_{k+1}\| = \|x_* - N_F(x_k)\| \leq [n_f(t_k)/t_k^2] \|x_* - x_k\|^2, \quad k = 0, 1, \dots$$

Therefore, the desired inequality follows from the definition of $\{t_k\}$ in (2.7) and the latter equation.

Uniqueness was proved in Lemma 2.13, and the last statement of the theorem was proved in Lemma 2.15. \square

3. Special cases

In this section, we will present three special cases of Theorem 2.1. They include the classical convergence theorem on Newton's method under the Lipschitz condition, Smale's theorem on Newton's method for analytical functions and the Nesterov–Nemirovskii theorem on Newton's method for self-concordant functions.

3.1 Convergence result under Lipschitz condition

In this section, we will present the classical convergence theorem for Newton's method under a Lipschitz condition; this has appeared in Rall (1974) and Traub & Wozniakowski (1979).

THEOREM 3.1 Let X, Y be Banach spaces, $\Omega \subseteq X$ an open set and $F : \Omega \rightarrow Y$ a continuously differentiable function. Let $x_* \in \Omega$ and $\kappa := \sup\{t \in [0, R) : B(x_*, t) \subset \Omega\}$. Suppose that $F(x_*) = 0$, $F'(x_*)$ is invertible and there exists a constant $K > 0$ such that

$$\|F'(x_*)^{-1}[F'(x) - F'(y)]\| \leq K\|x - y\|, \quad x, y \in \Omega. \quad (3.1)$$

Let $r = \min\{\kappa, 2/(3K)\}$. Then, the sequences with starting points $x_0 \in B(x_*, r) \setminus \{x_*\}$ and $t_0 = \|x_* - x_0\|$, respectively, namely

$$x_{k+1} = x_k - F'(x_k)^{-1}F(x_k), \quad t_{k+1} = ((K/2)t_k^2)/(1 - Kt_k), \quad k = 0, 1, \dots,$$

are well defined; $\{t_k\}$ is strictly decreasing, is contained in $(0, r)$ and converges to 0 and $\{x_k\}$ is contained in $B(x_*, r)$ and converges to the point x_* , which is the unique zero of F in $B(x_*, 2/K)$. Moreover, $\{t_{k+1}/t_k^2\}$ is strictly decreasing, $t_{k+1}/t_k^2 < 1/(2/K - 2t_0)$ for $k = 0, 1, \dots$ and

$$\|x_* - x_{k+1}\| \leq \frac{K}{2} \frac{1}{1 - Kt_k} \|x_k - x_*\|^2 \leq \frac{K}{2} \frac{1}{1 - K\|x_0 - x_*\|} \|x_k - x_*\|^2, \quad k = 0, 1, \dots$$

If, additionally, $2/(3K) < \kappa$, then $r = 2/(3K)$ is the best possible convergence radius.

Proof. We can immediately prove that F, x_* and $f : [0, \kappa) \rightarrow \mathbb{R}$, defined by $f(t) = Kt^2/2 - t$, satisfy the inequality (2.1) and the conditions (h1) and (h2) in Theorem 2.1. In this case, it is easy to see that ρ and ν , as defined in Theorem 2.1, satisfy

$$\rho = 2/(3K) \leq \nu = 1/K,$$

and, as a consequence, $r := \min\{\kappa, 2/(3K)\}$. Moreover, $f(\rho)/(f'(\rho)) - 1 = 1$, $f(0) = f(2/K) = 0$ and $f(t) < 0$ for all $t \in (0, 2/K)$. Also, the sequence $\{t_k\}$ is equivalent to $t_{k+1} = |t_k - f(t_k)/f'(t_k)|$ for

$k = 0, 1, \dots$ and

$$t_{k+1}/t_k^2 = \frac{K}{2} \frac{1}{1 - Kt_k} = \frac{1}{2/K - 2t_k} < \frac{1}{2/K - 2t_0} = \frac{K}{2} \frac{1}{1 - K\|x_0 - x_*\|}, \quad k = 0, 1, \dots$$

Therefore, the result follows by invoking Theorem 2.1. □

REMARK 3.2 Note that since $\|x_* - x_0\| < 2/(3K)$, the last inequality in Theorem 3.1 implies that $\|x_* - x_{k+1}\| \leq [3K/2]\|x_k - x_*\|^2$ for $k = 0, 1, \dots$. Therefore, we conclude that

$$\|x_* - x_k\| \leq [2/(3K)]([3K/2]\|x_* - x_0\|)^{2^k}, \quad k = 0, 1, \dots$$

REMARK 3.3 If $F : \Omega \rightarrow Y$ satisfies the classical Lipschitz condition for the derivative, namely

$$\|F'(x) - F'(y)\| \leq L\|x - y\|, \quad x, y \in \Omega,$$

where $L > 0$, then it also satisfies the condition (3.1) with $K = L\|F'(x_*)^{-1}\|$. In this case, the best possible convergence radius for Newton's method is $r = 2/(3L\|F'(x_*)^{-1}\|)$.

3.2 Convergence result under Smale's condition

In this section, we will present a convergence theorem on Newton's method under Smale's condition. This is the corollary of Proposition 3 of Smale (1986, p. 195); see also Proposition 1 and Remark 1 of Blum *et al.* (1997, pp. 157, 158).

THEOREM 3.4 Let X and Y be Banach spaces, $\Omega \subseteq X$ an open set and $F : \Omega \rightarrow Y$ an analytic function. Let $x_* \in \Omega$ and $\kappa := \sup\{t \in [0, R) : B(x_*, t) \subset \Omega\}$. Suppose that $F(x_*) = 0$, $F'(x_*)$ is invertible and

$$\gamma := \sup_{n>1} \left\| \frac{F'(x_*)^{-1} F^{(n)}(x_*)}{n!} \right\|^{1/(n-1)} < +\infty. \tag{3.2}$$

Let $r = \min\{\kappa, (5 - \sqrt{17})/(4\gamma)\}$. Then, the sequences with starting points $x_0 \in B(x_*, r) \setminus \{x_*\}$ and $t_0 = \|x_* - x_0\|$, respectively, namely

$$x_{k+1} = x_k - F'(x_k)^{-1}F(x_k), \quad t_{k+1} = (\gamma t_k^2)/[2(1 - \gamma t_k)^2 - 1], \quad k = 0, 1, \dots,$$

are well defined; $\{t_k\}$ is strictly decreasing, is contained in $(0, r)$ and converges to 0 and $\{x_k\}$ is contained in $B(x_*, r)$ and converges to the point x_* , which is the unique zero of F in $B(x_*, 1/(2\gamma))$. Moreover, $\{t_{k+1}/t_k^2\}$ is strictly decreasing, $t_{k+1}/t_k^2 < \gamma/[2(1 - \gamma\|x_0 - x_*\|)^2 - 1]$ for $k = 0, 1, \dots$ and

$$\|x_{k+1} - x_*\| \leq \frac{\gamma}{2(1 - \gamma t_k)^2 - 1} \|x_k - x_*\|^2 \leq \frac{\gamma}{2(1 - \gamma\|x_0 - x_*\|)^2 - 1} \|x_k - x_*\|^2, \quad k = 0, 1, \dots$$

If, additionally, $(5 - \sqrt{17})/(4\gamma) < \kappa$, then $r = (5 - \sqrt{17})/(4\gamma)$ is the best possible convergence radius.

REMARK 3.5 Because $\|x_* - x_0\| < (5 - \sqrt{17})/(4\gamma)$, the last inequality in Theorem 3.4 implies that $\|x_* - x_{k+1}\| \leq [(4\gamma)/(5 - \sqrt{17})]\|x_k - x_*\|^2$ for $k = 0, 1, \dots$. Therefore, we have

$$\|x_* - x_k\| \leq [(5 - \sqrt{17})/(4\gamma)]([(4\gamma)/(5 - \sqrt{17})]\|x_0 - x_*\|)^{2^k}, \quad k = 0, 1, \dots$$

The next result gives a condition that is easier to check than condition (2.1) whenever the functions under consideration are twice continuously differentiable.

LEMMA 3.6 Let X and Y be Banach spaces, let $\Omega \subseteq X$ be an open set and let $F : \Omega \rightarrow Y$ be twice continuously differentiable. Let $x_* \in \Omega$, $R > 0$ and $\kappa := \sup\{t \in [0, R) : B(x_*, t) \subset \Omega\}$. If $F'(x_*)$ is invertible and there exists an $f : [0, R) \rightarrow \mathbb{R}$ twice continuously differentiable such that

$$\|F'(x_*)^{-1}F''(x)\| \leq f''(\|x - x_*\|), \quad (3.3)$$

for all $x \in B(x_*, \kappa)$, then F and f satisfy (2.1).

Proof. Let $x \in B(x_*, \kappa)$ and $\tau \in [0, 1]$. It is simple to show that

$$\|F'(x_*)^{-1}[F'(x) - F'(x_* + \tau(x - x_*))]\| \leq \int_{\tau}^1 \|F'(x_*)^{-1}F''(x_* + t(x - x_*))\| \|x - x_*\| dt.$$

Now, since $\|x - x_*\| < \kappa$ and f satisfies (3.3), we obtain from the above inequality that

$$\|F'(x_*)^{-1}[F'(x) - F'(x_* + \tau(x - x_*))]\| \leq \int_{\tau}^1 f''(t\|x - x_*\|) \|x - x_*\| dt,$$

which, after evaluating the integrals, yields the desired result. \square

Proof of Theorem 3.4. Consider the real function $f : [0, 1/\gamma) \rightarrow \mathbb{R}$ defined by

$$f(t) = \frac{t}{1 - \gamma t} - 2t.$$

It is straightforward to show that f is analytic and

$$f(0) = 0, \quad f'(t) = 1/(1 - \gamma t)^2 - 2, \quad f'(0) = -1, \quad f''(t) = (2\gamma)/(1 - \gamma t)^3, \quad f^n(0) = n!\gamma^{n-1},$$

for $n \geq 2$. It follows from the last equalities that f satisfies (h1) and (h2). Now, using Lemma (3.6) and (3.2), it is easy to show that F and f satisfy (2.1); see, for example, Wang (1999). In this case, v , ρ and r , as defined in Theorem 2.1, satisfy

$$\rho = \frac{5 - \sqrt{17}}{4\gamma} < v = \frac{\sqrt{2} - 1}{\sqrt{2}\gamma} < \frac{1}{\gamma}, \quad r = \min \left\{ \kappa, \frac{5 - \sqrt{17}}{4\gamma} \right\}.$$

Moreover, $f(\rho)/(f'(\rho)) - 1 = 1$ and $f(0) = f(1/(2\gamma)) = 0$ and $f(t) < 0$ for $t \in (0, 1/(2\gamma))$. Also, $\{t_k\}$ is equivalent to $t_{k+1} = |t_k - f(t_k)/f'(t_k)|$ for $k = 0, 1, \dots$ and

$$t_{k+1}/t_k^2 = \frac{\gamma}{2(1 - \gamma t_k)^2 - 1} < \frac{\gamma}{2(1 - \gamma t_0)^2 - 1} = \frac{\gamma}{2(1 - \gamma \|x_0 - x_*\|)^2 - 1}, \quad k = 0, 1, \dots$$

Therefore, the result follows by applying Theorem 2.1. \square

3.3 Convergence result under the Nesterov–Nemirovskii condition

In this section, we will present a convergence theorem for Newton's method under the Nesterov–Nemirovskii condition.

Let $\Omega \subset \mathbb{R}^n$ be a convex set. A function $g : \Omega \rightarrow \mathbb{R}$ is called a -self-concordant with parameter $a > 0$ if $g \in C^3(\Omega)$, i.e. it is three times continuously differentiable in Ω , is convex and satisfies the following inequality:

$$|g'''(x)[h, h, h]| \leq 2a^{-1/2}(g''(x)[h, h])^{3/2}, \quad x \in \Omega, \quad h \in \mathbb{R}^n. \tag{3.4}$$

Take $x_* \in \Omega$ such that $g''(x_*)$ is invertible. Define the Hilbert space $X := (\mathbb{R}^n, \langle \cdot, \cdot \rangle_{x_*})$ as the Euclidean space \mathbb{R}^n with a new inner product, and the associated norm, defined by

$$\langle u, v \rangle_{x_*} := a^{-1} \langle g''(x_*)u, v \rangle, \quad \|u\|_{x_*} := \sqrt{\langle u, u \rangle_{x_*}}.$$

So the open and closed balls of radius $r > 0$ centred at x_* (Dikin's ellipsoid of radius $r > 0$ centred at x_*) in X are defined, respectively, as

$$W_r(x_*) := \{x \in \mathbb{R}^n : \|x - x_*\|_{x_*} < r\}, \quad W_r[x_*] := \{x \in \mathbb{R}^n : \|x - x_*\|_{x_*} \leq r\}.$$

THEOREM 3.7 Let $\Omega \subseteq \mathbb{R}^n$ be a convex set and $g : \Omega \rightarrow \mathbb{R}$ an a -self-concordant function. Let $x_* \in \Omega$ and suppose that $g'(x_*) = 0$ and $g''(x_*)$ is invertible. Let $X := (\mathbb{R}^n, \langle \cdot, \cdot \rangle_{x_*})$ be a Hilbert space, $\kappa := \sup\{u > 0 : W_u(x_*) \subset \Omega\}$ and

$$r := \min \left\{ \kappa, \frac{5 - \sqrt{17}}{4} \right\}.$$

Then, the sequences with starting points $x_0 \in B(x_*, r) \setminus \{x_*\}$ and $t_0 = \|x_* - x_0\|$, respectively, namely

$$x_{k+1} = x_k - g''(x_k)^{-1}g'(x_k), \quad t_{k+1} = t_k^2/[2(1 - t_k)^2 - 1], \quad k = 0, 1, \dots,$$

are well defined; $\{t_k\}$ is strictly decreasing, is contained in $(0, r)$ and converges to 0 and $\{x_k\}$ is contained in $B(x_*, r)$ and converges to the point x_* . Moreover, $\{t_{k+1}/t_k^2\}$ is strictly decreasing, $t_{k+1}/t_k^2 < 1/[2(1 - \|x_0 - x_*\|)^2 - 1]$ for $k = 0, 1, \dots$ and

$$\|x_{k+1} - x_*\| \leq \frac{1}{2(1 - t_k)^2 - 1} \|x_k - x_*\|^2 \leq \frac{1}{2(1 - \|x_0 - x_*\|)^2 - 1} \|x_k - x_*\|^2, \quad k = 0, 1, \dots$$

REMARK 3.8 Because $\|x_* - x_0\| < (5 - \sqrt{17})/4$, the last inequality in Theorem 3.4 implies that $\|x_* - x_{k+1}\| \leq [4/(5 - \sqrt{17})]\|x_k - x_*\|^2$ for $k = 0, 1, \dots$. Therefore, we have

$$\|x_* - x_k\| \leq [(5 - \sqrt{17})/4]([4/(5 - \sqrt{17})]\|x_0 - x_*\|)^{2^k}, \quad k = 0, 1, \dots$$

LEMMA 3.9 Let $\Omega \subset \mathbb{R}^n$ be an open convex set and $g : \Omega \rightarrow \mathbb{R}$ an a -self-concordant function. Assume that $W_1(x_*) \subset \Omega$. Then

$$\|g''(x_*)^{-1}g'''(x)\|_{x_*} \leq \frac{2}{(1 - \|x - x_*\|_{x_*})^3}, \quad x \in W_1(x_*).$$

Proof. See Lemma 5.1 of Alvarez *et al.* (2008). □

Proof of Theorem 3.7. Consider the real function $f : [0, 1) \rightarrow \mathbb{R}$ defined by

$$f(t) = \frac{t}{1-t} - 2t.$$

It is straightforward to show that

$$f(0) = 0, \quad f'(t) = 1/(1-t)^2 - 2, \quad f'(0) = -1, \quad f''(t) = 2/(1-t)^3.$$

From the last equalities, it is easy to show that f satisfies (h1) and (h2). Now, combining Lemmas (3.6) and (3.9) and the latter equality, we conclude that $F = g'$ and f satisfy (2.1). In this case, ν , ρ and r as defined in Theorem 2.1 satisfy

$$\rho = \frac{5 - \sqrt{17}}{4} < \nu = \frac{\sqrt{2} - 1}{\sqrt{2}} < 1, \quad r = \min \left\{ \kappa, \frac{5 - \sqrt{17}}{4} \right\}.$$

Moreover, $f(\rho)/(f'(\rho)) - 1 = 1$ and $f(0) = f(1/2) = 0$ and $f(t) < 0$ for $t \in (0, 1/2)$. Also, $\{t_k\}$ is equivalent to $t_{k+1} = |t_k - f(t_k)/f'(t_k)|$ for $k = 0, 1, \dots$ and

$$t_{k+1}/t_k^2 = \frac{1}{2(1-t_k)^2 - 1} < \frac{1}{2(1-t_0)^2 - 1} = \frac{1}{2(1 - \|x_0 - x_*\|)^2 - 1}, \quad k = 0, 1, \dots$$

Therefore, the result follows by applying Theorem 2.1. □

4. Final remarks

Lemmas (3.6) and (3.9) imply that $f(t) = t/(1-t) - 2t$ is a majorant function to an a -self-concordant function $g : \Omega \rightarrow \mathbb{R}$. It is easy to see that f is not self-concordant. So we cannot apply Lemma 2.15, as before, to conclude that r in Theorem 3.7 is the best possible convergence radius.

Acknowledgements

The author would like to thank the two anonymous referees for their constructive comments, which have helped to substantially improve the presentation of the paper.

Funding

CNPq (302618/2005-8, 475647/2006-8); PRONEX Optimization (FAPERJ/CNPq); FUNAPE/UFG.

REFERENCES

- ALVAREZ, F., BOTLE, J. & MUNIER, J. (2008) A unifying local convergence result for Newton's method in Riemannian manifolds. *Found. Comput. Math.*, **8**, 197–226.
- BLUM, L., CUCKER, F., SHUB, M. & SMALE, S. (1997) *Complexity and Real Computation*. New York: Springer.
- DENNIS, J. E. JR. & SCHNABEL, R. B. (1996) *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Classics in Applied Mathematics, Vol. 16. Philadelphia: SIAM.
- FERREIRA, O. P. & SVAITER, B. F. (2007) Kantorovich's majorants principle for Newton's method. *Comput. Optim. Appl.*, doi:10.1007/s10589-007-9082 (to appear).

- HIRIART-URRUTY, J.-B. & LEMARÉCHAL, C. (1993) *Convex Analysis and Minimization Algorithms. Part I*. Berlin: Springer.
- KANTOROVICH, L. V. (1951) Principle of majorants and Newton's method. *Dokl. Akad. Nauk SSSR*, **76**, 104–144.
- KRANTZ, S. G. & PARKS, H. R. (2002) *The Implicit Function Theorem: History, Theory, and Applications*. Boston: Birkhäuser.
- MOSER, J. (1961) A new techniques for the construction of solutions of nonlinear differential equations. *Proc. Natl. Acad. Sci. USA*, **47**, 1824–1831.
- NASH, J. (1956) The embedding problem for Riemannian manifolds. *Ann. Math.*, **63**, 20–63.
- NESTEROV, Y. & NEMIROVSKII, A. (1994) *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM Studies in Applied Mathematics, vol. 13. Philadelphia: SIAM.
- ORTEGA, J. M. (1990) *Numerical Analysis: A Second Course*. Classics in Applied Mathematics, vol. 3. Philadelphia: SIAM.
- ORTEGA, J. M. & RHEIMBOLDT, W. C. (1970) *Interactive Solution of Nonlinear Equations in Several Variables*. New York: Academic Press.
- OSTROWSKI, A. M. (1996) *Solution of Equations and Systems of Equations*. New York: Academic Press.
- POLYAK, B. T. (2007) Newton's method and its use in optimization. *Eur. J. Oper. Res.*, **181**, 1086–1096.
- RALL, L. B. (1974) A note on the convergence of Newton's method. *SIAM J. Numer. Anal.*, **11**, 34–36.
- SMALE, S. (1986) Newton method estimates from data at one point. *The Merging of Disciplines: New Directions in Pure, Applied and Computational Mathematics* (R. Ewing, K. Gross & C. Martin eds). New York: Springer, pp. 185–196.
- TRAUB, J. F. & WOZNAKOWSKI, H. (1979) Convergence and complexity of Newton iteration for operator equation. *J. Assoc. Comput. Mach.*, **26**, 250–258.
- WANG, X. (1999) Convergence of Newton's method and inverse function theorem in Banach space. *Math. Comput.*, **68**, 169–186.
- WANG, X. (2000) Convergence of Newton's method and uniqueness of the solution of equation in Banach space. *IMA J. Numer. Anal.*, **20**, 123–134.
- WAYNE, C. E. (1996) *An Introduction to KAM Theory. Dynamical Systems and Probabilistic Methods in Partial Differential Equation*. Lectures in Applied Mathematics, vol. 31. Providence: American Mathematical Society.