Local convergence analysis of the Gauss–Newton method under a majorant condition

O.P. Ferreira [a], M.L.N. Gonçalves [b,*], P.R. Oliveira [b]

[a] *IME/UFG, Campus II- Caixa Postal 131, CEP 74001-970 - Goiânia, GO, Brazil*
[b] *COPPE-Sistemas, Universidade Federal do Rio de Janeiro, 21945-970 Rio de Janeiro, RJ, Brazil*

ARTICLE INFO

ABSTRACT

The Gauss–Newton method for solving nonlinear least squares problems is studied in this paper. Under the hypothesis that the derivative of the function associated with the least square problem satisfies a majorant condition, a local convergence analysis is presented. This analysis allows us to obtain the optimal convergence radius and the biggest range for the uniqueness of stationary point, and to unify two previous and unrelated results.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

We consider the *nonlinear least squares* problem

$$\min_{x \in \Omega} \|F(x)\|^2, \tag{1}$$

where $\Omega \subseteq \mathbb{X}$ is an open set and $F : \Omega \to \mathbb{Y}$ is a continuously differentiable nonlinear function, with $\mathbb{X}$ and $\mathbb{Y}$ being real or complex Hilbert spaces. The interest in this problems arises in data fitting, when $\mathbb{X} = \mathbb{R}^n$ and $\mathbb{Y} = \mathbb{R}^m$, where m is the number of observations and n is the number of parameters; see [11,19].

Denote by $F'(x)$ the derivative of F at a point $x \in \Omega$. When the derivative $F'(x)$ is injective, the problem of finding a stationary point of problem (1), that is, a solution of the nonlinear equation

$$F'(x)^*F(x) = 0,$$

* Corresponding author.
E-mail addresses: orizon@mat.ufg.br (O.P. Ferreira), maxlng@cos.ufrj.br (M.L.N. Gonçalves), poliver@cos.ufrj.br (P.R. Oliveira).

where A^* denotes the adjoint of the operator A, is equivalent to finding a least squares solution of the overdetermined nonlinear equation

$$F(x) = 0. \tag{2}$$

This problem has been extensively studied by Dedieu and Kim [8] and Dedieu and Shub [10] for analytic functions, and in the Riemannian context by Adler et al. [1].

When $F'(x)$ is injective and has a closed image for all $x \in \Omega$, the Gauss–Newton method finds stationary points of the above problem. Formally, the Gauss–Newton method is described as follows. Given an initial point $x_0 \in \Omega$, define

$$x_{k+1} = x_k - [F'(x_k)^*F'(x_k)]^{-1}F'(x_k)^*F(x_k), \quad k = 0, 1, \ldots.$$

If the above method converges to $x_* \in \Omega$, then x_* is a stationary point of problem (1), but we cannot conclude that x_* is a solution of (1) or $F(x_*) = 0$. In order to ensure that a stationary point x_* is a solution of (1), we have to apply optimality conditions. It is worth pointing that, if $F'(x)$ is invertible for all $x \in \Omega$, then the Gauss–Newton method becomes the Newton method. Early works dealing with the convergence of the Newton and Gauss–Newton methods include [1–3,5–10,12–14,17,18,20,21,24].

It is well known that the Gauss–Newton method may fail or even fail to be well defined; see the example on p. 225 of [11] and the example in [10]. To ensure that the method is well defined and converges to a stationary point of (1), some conditions must be imposed. For instance, classical convergence analysis (see [11,19]) requires that F' satisfies the Lipschitz condition and that the initial iterate is 'close enough' to the solution, but it cannot make us clearly see how big the convergence radius of the ball is. For the analytic function, Dedieu and Shub [10] have given an estimate of the convergence radius and a criterion for convergence of the Gauss–Newton method.

Our aim in this paper is to present a new local convergence analysis for the Gauss–Newton method under a majorant condition as introduced by Kantorovich [16], and successfully used by Ferreira [12], Ferreira and Gonçalves [13] and Ferreira and Svaiter [14] for studying the Newton method. 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 as introduced in [24] and used by Chen, Li [6,7] and Li et al. [17,18] for studying the Gauss–Newton and Newton methods. The convergence analysis presented provides a clear relationship between the majorant function, which relaxes the Lipschitz continuity of the derivative, and the function associated with the nonlinear least square problem (see, for example, Lemmas 13–15). Besides, the results presented here have made the conditions and the proof of convergence simpler. They also allow us to obtain the biggest range for the uniqueness of the stationary point and the optimal convergence radius for the method as regards the majorant function. Moreover, two previous and unrelated results pertaining to the Gauss–Newton method are unified: namely, the result for analytic functions that appeared in Dedieu and Shub [10] and the classical one for functions with Lipschitz derivative (see, for example, [11] and [19]).

The organization of the paper is as follows. In Section 1.1, we list some notation and basic results used in our presentation. In Section 2, the main result is stated, and in Section 2.1, some properties involving the majorant function are established. In Section 2.2, we present the relationships between the majorant function and the nonlinear function F, and in Section 2.3, the optimal ball of convergence and the uniqueness of the stationary point are established. In Section 2.4, the main result is proved, and some applications of this result are given in Section 3.

1.1. Notation and auxiliary results

The following notation and results are used throughout our presentation. Let $\mathbb{X}$ and $\mathbb{Y}$ be Hilbert spaces. The open and closed balls at $a \in \mathbb{X}$ and radius $\delta > 0$ are denoted, respectively, by

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

The set $\Omega \subseteq \mathbb{X}$ is an open set, the function $F : \Omega \to \mathbb{Y}$ is continuously differentiable, and $F'(x)$ has a closed image in Ω.

Let $A : \mathbb{X} \to \mathbb{Y}$ be a continuous and injective linear operator with closed image. The Moore–Penrose inverse $A^\dagger : \mathbb{Y} \to \mathbb{X}$ of A is defined by

$$A^\dagger := (A^*A)^{-1}A^*,$$

where A^* denotes the adjoint of the linear operator A.

Lemma 1 (*Banach's Lemma*)**.** *Let $B : \mathbb{X} \to \mathbb{X}$ be a continuous linear operator and $I : \mathbb{X} \to \mathbb{X}$ the identity operator. If $\|B - I\| < 1$, then B is invertible and $\|B^{-1}\| \le 1/(1 - \|B - I\|)$.*

Proof. See the proof of Lemma 1, p. 189 of [22] with $A = I$ and $c = \|B - I\|$. □

Lemma 2. *Let $A, B : \mathbb{X} \to \mathbb{Y}$ be continuous linear operators with closed images. If A is injective, $E = B - A$ and $\|EA^\dagger\| < 1$, then B is injective.*

Proof. In fact, $B = A + E = (I + EA^\dagger)A$, from the condition $\|EA^\dagger\| < 1$, we have from Lemma 1 that $I + EA^\dagger$ is invertible. So, B is injective. □

The next lemma is proved in [23] (see also [25]) for an $m \times n$ matrix with $m \ge n$ and $rank(A) = rank(B) = n$; that proof holds in a more general context, as will be stated below.

Lemma 3. *Let $A, B : \mathbb{X} \to \mathbb{Y}$ be continuous and injective linear operators with closed images. Assume that $E = B - A$ and $\|A^\dagger\|\|E\| < 1$; then*

$$\|B^\dagger\| \le \frac{\|A^\dagger\|}{1 - \|A^\dagger\|\|E\|}, \qquad \|B^\dagger - A^\dagger\| \le \frac{\sqrt{2}\|A^\dagger\|^2\|E\|}{1 - \|A^\dagger\|\|E\|}.$$

Proposition 4. *If $0 \le t < 1$, then $\sum_{i=0}^{\infty}(i + 2)(i + 1)t^i = 2/(1 - t)^3$.*

Proof. Take $k = 2$ in Lemma 3, p. 161 of [4]. □

Also, the following auxiliary results of elementary convex analysis will be needed.

Proposition 5. *Let $R > 0$. If $\varphi : [0, R) \to \mathbb{R}$ is convex, then*

$$D^+\varphi(0) = \lim_{u \to 0+} \frac{\varphi(u) - \varphi(0)}{u} = \inf_{0 < u} \frac{\varphi(u) - \varphi(0)}{u}.$$

Proof. See Theorem 4.1.1 on p. 21 of [15]. □

Proposition 6. *Let $\epsilon > 0$ and $\tau \in [0, 1]$. If $\varphi : [0, \epsilon) \to \mathbb{R}$ is convex, then $l : (0, \epsilon) \to \mathbb{R}$, defined by*

$$l(t) = \frac{\varphi(t) - \varphi(\tau t)}{t},$$

is increasing.

Proof. See Theorem 4.1.1 and Remark 4.1.2 on p. 21 of [15]. □

2. Local analysis for the Gauss–Newton method

Our goal is to state and prove a local theorem for the Gauss–Newton method. First, we will prove some results regarding the scalar majorant function, which relaxes the Lipschitz condition of the derivative of the function associated with the nonlinear least square problem. We will then show that the Gauss–Newton method is well defined and that it converges. We will also prove the uniqueness of the stationary point in a suitable region, and the convergence rate will be established. The statement of the theorem is as follows.

Theorem 7. *Let $\Omega \subseteq \mathbb{X}$ be an open set, and let $F : \Omega \to \mathbb{Y}$ be a continuously differentiable function. Let $x_* \in \Omega, R > 0$ and*

$$c := \|F(x_*)\|, \qquad \beta := \|F'(x_*)^\dagger\|, \qquad \kappa := \sup\{t \in [0, R) : B(x_*, t) \subset \Omega\}.$$

Suppose that $F'(x_*)^*F(x_*) = 0$, $F'(x_*)$ is injective and that there exists an $f : [0, R) \to \mathbb{R}$ that is continuously differentiable such that

$$\|F'(x) - F'(x_* + \tau(x - x_*))\| \le f'(\|x - x_*\|) - f'(\tau\|x - x_*\|), \tag{3}$$

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;
(h3) $\sqrt{2}\,c\,\beta^2 D^+ f'(0) < 1$.

   Let there be given positive constants $\nu := \sup\{t \in [0, R) : \beta[f'(t) + 1] < 1\}$,

$$\rho := \sup\left\{ t \in (0, \nu) : \frac{\beta[tf'(t) - f(t)] + \sqrt{2}c\beta^2[f'(t) + 1]}{t[1 - \beta(f'(t) + 1)]} < 1 \right\}, \qquad r := \min\{\kappa, \ \rho\}.$$

   Then, the Gauss–Newton method for solving (1), *with starting point $x_0 \in B(x_*, r)/\{x_*\}$,*

$$x_{k+1} = x_k - F'(x_k)^\dagger F(x_k), \quad k = 0, 1 \dots, \tag{4}$$

   is well defined, the generated sequence $\{x_k\}$ is contained in $B(x_*, r)$, converges to x_*, and

$$\|x_{k+1} - x_*\| \le \frac{\beta[f'(\|x_0 - x_*\|)\|x_0 - x_*\| - f(\|x_0 - x_*\|)]}{\|x_0 - x_*\|^2[1 - \beta(f'(\|x_0 - x_*\|) + 1)]}\|x_k - x_*\|^2$$

$$+ \frac{\sqrt{2}c\beta^2[f'(\|x_0 - x_*\|) + 1]}{\|x_0 - x_*\|[1 - \beta(f'(\|x_0 - x_*\|) + 1)]}\|x_k - x_*\|, \quad k = 0, 1, \dots. \tag{5}$$

   Moreover, if $[\beta(\rho f'(\rho) - f(\rho)) + \sqrt{2}c\beta^2(f'(\rho) + 1)]/[\rho(1 - \beta(f'(\rho) + 1))] = 1$ and $\rho < \kappa$, then $r = \rho$ is the best possible convergence radius.
   If, additionally,

(h4) $2\,c\,\beta_0\,D^+ f'(0) < 1$, *then x_* is the unique stationary point of $F(x)^*F(x)$ in $B(x_*, \sigma)$, where $0 < \sigma := \sup\{t \in (0, \kappa) : [\beta(f(t)/t + 1) + c\beta_0(f'(t) + 1)/t] < 1\}$, $\beta_0 := \|[F'(x_*)^*F'(x_*)]^{-1}\|$.*

Remark 1. Inequality (5) shows that, if $c = 0$ (the so-called zero-residual case), then the Gauss–Newton method is locally Q-quadratically convergent to x_*. This behaviour is quite similar to that of the Newton method (see [12,24]). If c is relatively small (the so-called small-residual case), inequality (5) implies that the Gauss–Newton method is locally Q-linearly convergent to x_*; see the example in [10]. However, if c is large (the so-called large-residual case), the Gauss–Newton method may not be locally convergent at all; see condition (h3) and the examples on p. 225 and p. 1101 of [11] and [10], respectively. Hence, we may conclude that the Gauss–Newton method performs better on zero-residual or small-residual problems than on large-residual problems, while the Newton method is equally effective in all these cases.

   For zero-residual problems, i.e., $c = 0$, Theorem 7 becomes the following.

Corollary 8. *Let $\Omega \subseteq \mathbb{X}$ be an open set, and let $F : \Omega \to \mathbb{Y}$ be a continuously differentiable function. Let $x_* \in \Omega, R > 0$ and*

$$\beta := \|F'(x_*)^\dagger\|, \qquad \kappa := \sup\{t \in [0, R) : B(x_*, t) \subset \Omega\}.$$

Suppose that $F(x_*) = 0$, $F'(x_*)$ is injective and that there exists an $f : [0, R) \to \mathbb{R}$ that is continuously differentiable such that

$$\|F'(x) - F'(x_* + \tau(x - x_*))\| \le f'(\|x - x_*\|) - f'(\tau\|x - x_*\|),$$

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 there be given positive constants $\nu =: \sup\{t \in [0, R) : \beta[f'(t) + 1] < 1\}$,

$$\rho := \sup\{t \in (0, \nu) : [\beta(tf'(t) - f(t))]/[t(1 - \beta(f'(t) + 1))] < 1\}, \qquad r := \min\{\kappa, \rho\}.$$

Then, the Gauss–Newton method for solving (1), *with initial point* $x_0 \in B(x_*, r)/\{x_*\}$,

$$x_{k+1} = x_k - F'(x_k)^\dagger F(x_k), \quad k = 0, 1 \ldots,$$

is well defined, and the generated sequence $\{x_k\}$ *is contained in* $B(x_*, r)$ *and converges to* x_*, *which is the unique zero of* F *in* $B(x_*, \sigma)$, *where* $0 < \sigma := \sup\{0 < t < \kappa : \beta[f(t)/t + 1] < 1\}$. *Moreover, it holds that*

$$\|x_{k+1} - x_*\| \le \frac{\beta[f'(\|x_0 - x_*\|)\|x_0 - x_*\| - f(\|x_0 - x_*\|)]}{\|x_0 - x_*\|^2[1 - \beta(f'(\|x_0 - x_*\|) + 1)]}\|x_k - x_*\|^2, \quad k = 0, 1, \ldots.$$

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

Remark 2. When $F'(x_*)$ is invertible, Corollary 8 is similar to the result on the Newton method for solving nonlinear equations $F(x) = 0$, obtained by Ferreira in Theorem 2.1 of [12].

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

In order to prove Theorem 7 we need some results. From here on, we assume that all the assumptions of Theorem 7 hold.

2.1. The majorant function

Our first goal is to show that the constant κ associated with Ω and the constants ν, ρ and σ associated with the majorant function f are positive. Also, we will prove some results related to the function f.

We begin by noting that $\kappa > 0$, because Ω is an open set and $x_* \in \Omega$.

Proposition 9. *The constant* ν *is positive, and it holds that*

$$\beta[f'(t) + 1] < 1, \quad t \in (0, \nu).$$

Proof. As f' is continuous in $(0, R)$ and $f'(0) = -1$, it is easy to conclude that

$$\lim_{t \to 0} \beta[f'(t) + 1] = 0.$$

Thus, there exists a $\delta > 0$ such that $\beta(f'(t) + 1) < 1$ for all $t \in (0, \delta)$. Hence, $\nu > 0$.
Using (h2) and the definition of ν, the last part of the proposition follows. $\square$

Proposition 10. *The following functions are increasing:*

(i) $[0, R) \ni t \mapsto 1/[1 - \beta(f'(t) + 1)]$;
(ii) $(0, R) \ni t \mapsto [tf'(t) - f(t)]/t^2$;
(iii) $(0, R) \ni t \mapsto [f'(t) + 1]/t$;
(iv) $(0, R) \ni t \mapsto f(t)/t$.

As a consequence, the following functions are increasing:

$$(0, R) \ni t \mapsto \frac{tf'(t) - f(t)}{t^2[1 - \beta(f'(t) + 1)]}, \qquad (0, R) \ni t \mapsto \frac{f'(t) + 1}{t[1 - \beta(f'(t) + 1)]}.$$

Proof. Item (i) is immediate, as f' is strictly increasing in $[0, R)$.

To prove item (ii), note that after some simple algebraic manipulations we have

$$\frac{tf'(t) - f(t)}{t^2} = \int_0^1 \frac{f'(t) - f'(\tau t)}{t}\, d\tau.$$

So, applying Proposition 6 with $f' = \varphi$ and $\epsilon = R$, the statement follows.

To establish item (iii), use (h2), $f'(0) = -1$ and Proposition 6 with $f' = \varphi, \epsilon = R$ and $\tau = 0$.

Assumption (h2) implies that f is convex. As $f(0) = 0$, we have $f(t)/t = [f(t) - f(0)]/[t - 0]$. Hence, item (iv) follows by applying Proposition 6 with $f = \varphi$ and $\tau = 0$.

To prove that the functions in the last part are increasing, combine item (i) with item (ii) for the first function and item (i) with item (iii) for the second function. $\quad\square$

Proposition 11. *The constant ρ is positive, and it holds that*

$$\frac{\beta[tf'(t) - f(t)] + \sqrt{2}c\beta^2[f'(t) + 1]}{t[1 - \beta(f'(t) + 1)]} < 1, \quad \forall t \in (0, \rho).$$

Proof. First, using (h1) and some algebraic manipulation we obtain

$$\frac{\beta[tf'(t) - f(t)] + \sqrt{2}c\beta^2[f'(t) + 1]}{t[1 - \beta(f'(t) + 1)]} = \frac{\beta\left[f'(t) - \frac{f(t)-f(0)}{t-0}\right] + \sqrt{2}\,c\,\beta^2\,\frac{f'(t)-f'(0)}{t-0}}{1 - \beta(f'(t) + 1)}.$$

Combining the last equation with the assumption that f' is convex, we obtain from Proposition 5 that

$$\lim_{t \to 0} \frac{\beta[tf'(t) - f(t)] + \sqrt{2}c\beta^2[f'(t) + 1]}{t[1 - \beta(f'(t) + 1)]} = \sqrt{2}c\beta^2 D^+ f'(0).$$

Now, using (h3), i.e., $\sqrt{2}c\beta^2 D^+ f'(0) < 1$, we conclude that there exists a $\delta > 0$ such that

$$\frac{\beta[tf'(t) - f(t)] + \sqrt{2}c\beta^2[f'(t) + 1]}{t[1 - \beta(f'(t) + 1)]} < 1, \quad t \in (0, \delta).$$

Hence, $\delta \leq \rho$, which proves the first statement.

To conclude the proof, we use the definition of ρ, the above inequality and the last part of Proposition 10. $\quad\square$

Proposition 12. *The constant σ is positive, and it holds that*

$$\beta(f(t)/t + 1) + c\beta_0(f'(t) + 1)/t < 1, \quad t \in (0, \sigma).$$

Proof. To prove that $\sigma > 0$ we need assumption (h4). First, note that condition (h1) implies that

$$\beta\left[\frac{f(t)}{t} + 1\right] + c\beta_0\frac{f'(t) + 1}{t} = \beta\left[\frac{f(t) - f(0)}{t - 0} - f'(0)\right] + c\beta_0\frac{f'(t) - f'(0)}{t - 0}.$$

Therefore, using the last equality together with the assumption that f' is convex and (h4), we have $\lim_{t\to 0}[\beta(f(t)/t + 1) + c\beta_0(f'(t) + 1)/t] = c\beta_0 D^+ f'(0) < 1/2$. Thus, there exists a $\delta > 0$ such that

$$\beta\left[\frac{f(t)}{t} + 1\right] + c\beta_0\frac{f'(t) + 1}{t} < 1, \quad t \in (0, \delta).$$

Hence, $\delta \leq \sigma$, which proves the first statement.

To conclude the proof, we use the definition of σ, the above inequality and items (iii) and (iv) in Proposition 10. $\quad\square$

2.2. Relationship of the majorant function with the nonlinear function

In this section, we will present the main relationships between the majorant function f and the function F associated with the nonlinear least square problem.

Lemma 13. *Let $x \in \Omega$. If $\|x - x_*\| < \min\{v, \kappa\}$, then $F'(x)^*F'(x)$ is invertible, and the following inequalities hold:*

$$\|F'(x)^\dagger\| \leq \frac{\beta}{1 - \beta[f'(\|x - x_*\|) + 1]}, \qquad \|F'(x)^\dagger - F'(x_*)^\dagger\| < \frac{\sqrt{2}\beta^2[f'(\|x - x_*\|) + 1]}{1 - \beta[f'(\|x - x_*\|) + 1]}.$$

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

Proof. Let $x \in \Omega$ such that $\|x - x_*\| < \min\{v, \kappa\}$. Since $\|x - x_*\| < v$, using the definition of β, inequality (3), and the last part of Proposition 9, we have

$$\|F'(x) - F'(x_*)\|\|F'(x_*)^\dagger\| \leq \beta[f'(\|x - x_*\|) - f'(0)] < 1.$$

For simplicity, the notation defines the following matrices:

$$A = F'(x_*), \qquad B = F'(x), \qquad E = F'(x) - F'(x_*). \tag{6}$$

The last definitions together with latter inequality imply that

$$\|EA^\dagger\| \leq \|E\|\|A^\dagger\| < 1,$$

which, using that $F'(x_*)$ is injective, implies in view of Lemma 2 that $F'(x)$ is injective. So, $F'(x)^*F'(x)$ is invertible, and by definition of r we obtain that $F'(x)^*F'(x)$ is invertible for all $x \in B(x_*, r)$.

We already know that $F'(x_*)$ and $F'(x)$ are injective. Hence, to conclude the lemma we use the definitions in (6) and then combine the above inequality and Lemma 3. $\square$

Now, it is convenient to study the linearization error of F at a point in Ω; for this, we define

$$E_F(x, y) := F(y) - [F(x) + F'(x)(y - x)], \quad y, x \in \Omega. \tag{7}$$

We will bound this error by the error in the linearization on the majorant function f,

$$e_f(t, u) := f(u) - [f(t) + f'(t)(u - t)], \quad t, u \in [0, R]. \tag{8}$$

Lemma 14. *If $\|x - x_*\| < \kappa$, then it holds that $\|E_F(x, x_*)\| \leq e_f(\|x - x_*\|, 0)$.*

Proof. Since $B(x_*, \kappa)$ is convex, we obtain that $x_* + \tau(x - x_*) \in B(x_*, \kappa)$, for $0 \leq \tau \leq 1$. Thus, as F is continuously differentiable in Ω, the definition of E_F and some simple manipulations yield

$$\|E_F(x, x_*)\| \leq \int_0^1 \|[F'(x) - F'(x_* + \tau(x - x_*))]\| \|x_* - x\| \, d\tau.$$

From the last inequality and assumption (3), we obtain

$$\|E_F(x, x_*)\| \leq \int_0^1 [f'(\|x - x_*\|) - f'(\tau\|x - x_*\|)]\|x - x_*\| \, d\tau.$$

Evaluating the above integral and using the definition of e_f, the statement follows. $\square$

Lemma 13 guarantees, in particular, that $F'(x)^*F'(x)$ is invertible in $B(x_*, r)$ and, consequently, the Gauss–Newton iteration map is well defined. Let us call G_F the Gauss–Newton iteration map for F in that region:

$$G_F : B(x_*, r) \to \mathbb{Y}$$
$$x \mapsto x - F'(x)^\dagger F(x). \tag{9}$$

One can apply a *single* Gauss–Newton iteration on any $x \in B(x_*, r)$ to obtain $G_F(x)$, which may not belong to $B(x_*, r)$, or even may not belong to the domain of F. So, this is enough to guarantee the

well definedness of only one iteration. To ensure that the Gauss–Newton iterations may be repeated indefinitely, we need the following result.

Lemma 15. *Let $x \in \Omega$. If $\|x - x_*\| < r$, then G_F is well defined, and it holds that*

$$\|G_F(x) - x_*\| \leq \frac{\beta[f'(\|x - x_*\|)\|x - x_*\| - f(\|x - x_*\|)]}{\|x - x_*\|^2[1 - \beta(f'(\|x - x_*\|) + 1)]} \|x - x_*\|^2$$

$$+ \frac{\sqrt{2}c\beta^2[f'(\|x - x_*\|) + 1]}{\|x - x_*\|[1 - \beta(f'(\|x - x_*\|) + 1)]} \|x - x_*\|.$$

In particular,

$$\|G_F(x) - x_*\| < \|x - x_*\|.$$

Proof. First, note that, as $\|x - x_*\| < r$, it follows from Lemma 13 that $F'(x)^*F'(x)$ is invertible; then $G_F(x)$ is well defined. Since $F'(x_*)^T F(x_*) = 0$, some algebraic manipulation and (9) yield

$$G_F(x) - x_* = F'(x)^\dagger[F'(x)(x - x_*) - F(x) + F(x_*)] + F'(x_*)^\dagger F(x_*) - F'(x)^\dagger F(x_*).$$

From the last equation, the properties of the norm, and (7), we obtain

$$\|G_F(x) - x_*\| \leq \|F'(x)^\dagger\| \|E_F(x, x_*)\| + \|F'(x_*)^\dagger - F'(x)^\dagger\| \|F(x_*)\|.$$

Since $c = \|F(x_*)\|$, combining the last inequality with Lemmas 13 and 14, we have

$$\|G_F(x) - x_*\| \leq \frac{\beta e_f(\|x - x_*\|, 0)}{1 - \beta(f'(\|x - x_*\|) + 1)} + \frac{\sqrt{2}c\beta^2(f'(\|x - x_*\|) + 1)}{1 - \beta(f'(\|x - x_*\|) + 1)}.$$

Now, using (8) and (h1), we conclude from the last inequality that

$$\|G_F(x) - x_*\| \leq \frac{\beta[f'(\|x - x_*\|)\|x - x_*\| - f(\|x - x_*\|)]}{1 - \beta(f'(\|x - x_*\|) + 1)} + \frac{\sqrt{2}c\beta^2[f'(\|x - x_*\|) + 1]}{1 - \beta(f'(\|x - x_*\|) + 1)},$$

which is equivalent to the first inequality of the lemma.

To end the proof, first note that the right-hand side of the first inequality of the lemma is equivalent to

$$\left[\frac{\beta[f'(\|x - x_*\|)\|x - x_*\| - f(\|x - x_*\|)]}{\|x - x_*\|[1 - \beta(f'(\|x - x_*\|) + 1)]} + \frac{\sqrt{2}c\beta^2[f'(\|x - x_*\|) + 1]}{\|x - x_*\|[1 - \beta(f'(\|x - x_*\|) + 1)]} \right] \|x - x_*\|.$$

On the other hand, as $x \in B(x_*, r)/\{x_*\}$, i.e., $0 < \|x - x_*\| < r \leq \rho$, we apply Proposition 11 with $t = \|x - x_*\|$ to conclude that the quantity in the brackets above is less than one. So, the last inequality of the lemma follows. $\square$

2.3. Optimal ball of convergence and uniqueness

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

Lemma 16. *If $(\beta(\rho f'(\rho) - f(\rho)) + \sqrt{2}c\beta^2(f'(\rho) + 1))/\rho(1 - \beta(f'(\rho) + 1)) = 1$ and $\rho < \kappa$, then $r = \rho$ is the best possible convergence ratio.*

Proof. Define the function $h : (-\kappa, \kappa) \to \mathbb{R}$ by

$$h(t) = \begin{cases} -t/\beta + t - f(-t), & t \in (-\kappa, 0], \\ -t/\beta + t + f(t), & t \in [0, \kappa). \end{cases} \tag{10}$$

It is straightforward to show that $h(0) = 0$, $h'(0) = -1/\beta$, $h'(t) = -1/\beta + 1 + f'(|t|)$, and that

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

So, $F = h$ satisfies all the assumptions of Theorem 7 with $c = |h(0)| = 0$. Thus, as $\rho < \kappa$, it suffices to show that the Gauss–Newton method applied to solve (1), with $F = h$ and starting point $x_0 = \rho$, does not converge. Since $c = 0$, our assumption becomes

$$\beta(\rho f'(\rho) - f(\rho))/\rho(1 - \beta(f'(\rho) + 1)) = 1. \tag{11}$$

Hence the definition of h in (10) together with the last equality yields

$$x_1 = \rho - \frac{h'(\rho)^T h(\rho)}{h'(\rho)^T h'(\rho)} = \rho - \frac{-\rho/\beta + \rho + f(\rho)}{-1/\beta + 1 + f'(\rho)} = -\rho\left(\frac{\beta(\rho f'(\rho) - f(\rho))}{\rho(1 - \beta(f'(\rho) + 1))}\right) = -\rho.$$

Again, the definition of h in (10) and assumption (11) give

$$x_2 = -\rho - \frac{h'(-\rho)^T h(-\rho)}{h'(-\rho)^T h'(-\rho)} = -\rho - \frac{\rho/\beta - \rho - f(\rho)}{-1/\beta + 1 + f'(\rho)} = \rho\left(\frac{\beta(\rho f'(\rho) - f(\rho))}{\rho(1 - \beta(f'(\rho) + 1))}\right) = \rho.$$

Therefore, the Gauss–Newton method for solving (1), with $F = h$ and starting point $x_0 = \rho$, produces the cycle

$$x_0 = \rho, \qquad x_1 = -\rho, \qquad x_2 = \rho, \ldots;$$

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

Lemma 17. *If, additionally, (h4) holds, then x_* is the unique stationary point of $F(x)^*F(x)$ in $B(x_*, \sigma)$.*

Proof. Assume that $y \in B(x_*, \sigma)$, $y \neq x_*$ is also a stationary point of $F(x)^*F(x)$. Hence,

$$y - x_* = y - x_* - F'(y)^\dagger F(y).$$

Using $F'(x_*)^*F(x_*) = 0$, after some algebraic manipulation, the above equality becomes

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

Combining the last equation with properties of the norm and definitions of c, β and β_0, we obtain

$$\|y - x_*\| \leq \beta \int_0^1 \|F'(x_*) - F'(x_* + u(y - x_*))\| \|y - x_*\| du + c\beta_0 \|F'(x_*)^* - F'(y)^*\|.$$

Using (3) with $x = x_* + u(y - x_*)$ and $\tau = 0$ in the first term of the right-hand side, and $x = y$ and $\tau = 0$ in the second term of the right-hand side in the last inequality, we have

$$\|y - x_*\| \leq \beta \int_0^1 [f'(u\|y - x_*\|) - f'(0)]\|y - x_*\| du + c\beta_0[f'(\|y - x_*\|) - f'(0)].$$

Evaluating the above integral and using (h1), the latter inequality becomes

$$\|y - x_*\| \leq \left(\beta\left[\frac{f(\|y - x_*\|)}{\|y - x_*\|} + 1\right] + c\beta_0\left[\frac{f'(\|y - x_*\|) + 1}{\|y - x_*\|}\right]\right)\|y - x_*\|.$$

Since $0 < \|y - x_*\| < \sigma$, using Proposition 12 with $t = \|y - x_*\|$, we have $\|y - x_*\| < \|y - x_*\|$, which is a contradiction. Therefore, $y = x_*$. □

Remark 4. Note that in the above lemma we have used the fact that condition (3) holds only for $\tau = 0$.

2.4. Proof of Theorem 7

First of all, note that (4) together (9) implies that the sequence $\{x_k\}$ satisfies

$$x_{k+1} = G_F(x_k), \quad k = 0, 1, \ldots. \tag{12}$$

Proof. Since $x_0 \in B(x_*, r)/\{x_*\}$, i.e., $0 < \|x_k - x_*\| < r$, by combination of Lemma 13, the last inequality in Lemma 15 and an induction argument it is easy to see that $\{x_k\}$ is well defined and remains in $B(x_*, r)$.

Now, our goal is to show that $\{x_k\}$ converges to x_*. As $\{x_k\}$ is well defined and contained in $B(x_*, r)$, combining (12) with Lemma 15, we have

$$\|x_{k+1} - x_*\| \leq \frac{\beta[f'(\|x_k - x_*\|)\|x_k - x_*\| - f(\|x_k - x_*\|)]}{\|x_k - x_*\|^2[1 - \beta(f'(\|x_k - x_*\|) + 1)]} \|x_k - x_*\|^2$$
$$+ \frac{\sqrt{2}c\beta^2[f'(\|x_k - x_*\|) + 1]}{\|x_k - x_*\|[1 - \beta(f'(\|x_k - x_*\|) + 1)]} \|x_k - x_*\|,$$

for all $k = 0, 1, \ldots$. Using again (12) and the second part of Lemma 15, it is easy to conclude that

$$\|x_k - x_*\| < \|x_0 - x_*\|, \quad k = 1, 2 \ldots. \tag{13}$$

Hence, by combining the last two inequalities with the last part of Proposition 10, we obtain that

$$\|x_{k+1} - x_*\| \leq \frac{\beta[f'(\|x_0 - x_*\|)\|x_0 - x_*\| - f(\|x_0 - x_*\|)]}{\|x_0 - x_*\|^2[1 - \beta(f'(\|x_0 - x_*\|) + 1)]} \|x_k - x_*\|^2$$
$$+ \frac{\sqrt{2}c\beta^2[f'(\|x_0 - x_*\|) + 1]}{\|x_0 - x_*\|[1 - \beta(f'(\|x_0 - x_*\|) + 1)]} \|x_k - x_*\|,$$

for all $k = 0, 1, \ldots$, which is inequality (5). Now, using (13) and the last inequality, we have

$$\|x_{k+1} - x_*\| \leq \left[\frac{\beta[f'(\|x_0 - x_*\|)\|x_0 - x_*\| - f(\|x_0 - x_*\|)] + \sqrt{2}c\beta^2[f'(\|x_0 - x_*\|) + 1]}{\|x_0 - x_*\|[1 - \beta(f'(\|x_0 - x_*\|) + 1)]} \right]$$
$$\times \|x_k - x_*\|,$$

for all $k = 0, 1, \ldots$. Applying Proposition 11 with $t = \|x_0 - x_*\|$, it is straightforward to conclude from the latter inequality that $\{\|x_k - x_*\|\}$ converges to zero. So, $\{x_k\}$ converges to x_*. The optimal convergence radius was proved in Lemma 16 and the last statement of the theorem was proved in Lemma 17. $\square$

3. Special cases

In this section, we present two special cases of Theorem 7. They include the classical convergence theorem on the Gauss–Newton method under the Lipschitz condition and Smale's theorem on the Gauss–Newton method for analytic functions.

3.1. Convergence result for the Lipschitz condition

In this section, we show a correspondent theorem to Theorem 7 under the Lipschitz condition (see [11,19]) instead of the general assumption (3).

Theorem 18. *Let $\Omega \subseteq \mathbb{X}$ be an open set, and let $F : \Omega \to \mathbb{Y}$ be a continuously differentiable function. Let $x_* \in \Omega$ and*

$$c := \|F(x_*)\|, \qquad \beta := \|F'(x_*)^\dagger\|, \qquad \kappa := \sup\{t \in [0, R) : B(x_*, t) \subset \Omega\}.$$

Suppose that $F'(x_*)^*F(x_*) = 0$, $F'(x_*)$ is injective, and that there exists a $K > 0$ such that

$$\sqrt{2}c\beta^2 K < 1, \qquad \|F'(x) - F'(y)\| \leq K\|x - y\|, \quad \forall x, y \in B(x_*, \kappa).$$

Let

$$r := \min\{\kappa, (2 - 2\sqrt{2}K\beta^2 c)/(3K\beta)\}.$$

Then, the Gauss–Newton method for solving (1), with initial point $x_0 \in B(x_*, r)/\{x_*\}$,

$$x_{k+1} = x_k - F'(x_k)^\dagger F(x_k), \quad k = 0, 1, \ldots,$$

is well defined, the sequence generated $\{x_k\}$ *is contained in* $B(x_*, r)$, *converges to* x_*, *and*

$$\|x_{k+1} - x_*\| \leq \frac{\beta K}{2(1 - \beta K \|x_0 - x_*\|)} \|x_k - x_*\|^2$$

$$+ \frac{\sqrt{2}c\beta^2 K}{1 - \beta K \|x_0 - x_*\|} \|x_k - x_*\|, \quad k = 0, 1, \ldots.$$

Moreover, if $(2 - 2\sqrt{2}K\beta^2 c)/(3K\beta) < \kappa$, *then* $r = (2 - 2\sqrt{2}K\beta^2 c)/(3K\beta)$ *is the best possible convergence radius.*

If, additionally, $2c\beta_0 K < 1$, *then* x_* *is the unique stationary point of* $F(x)^*F(x)$ *in* $B(x_*, (2 - 2c\beta_0 K)/(\beta K))$, *where* $\beta_0 := \|[F'(x_*)^*F'(x_*)]^{-1}\|$.

Proof. It is immediate to prove that F, x_* and $f : [0, \kappa) \to \mathbb{R}$ defined by $f(t) = Kt^2/2 - t$ satisfy inequality (3) and conditions (h1) and (h2). Since $\sqrt{2}c\beta^2 K < 1$ and $2c\beta_0 K < 1$, conditions (h3) and (h4) also hold. In this case, it is easy to see that the constants ν and ρ as defined in Theorem 7 satisfy

$$0 < \rho = (2 - 2\sqrt{2}K\beta^2 c)/(3K\beta) \leq \nu = 1/\beta K,$$

and, as a consequence, $0 < r = \min\{\kappa, \rho\}$. Moreover, it is straightforward to show that

$$[\beta(\rho f'(\rho) - f(\rho)) + \sqrt{2}c\beta^2(f'(\rho) + 1)]/[\rho(1 - \beta(f'(\rho) + 1))] = 1,$$

and $[\beta(f(t)/t + 1) + c\beta_0(f'(t) + 1)/t] < 1$ for all $t \in (0, (2 - 2c\beta_0 K)/(\beta K))$. Therefore, as F, r, f and x_* satisfy all the hypotheses of Theorem 7, taking $x_0 \in B(x_*, r) \setminus \{x_*\}$, the statements of the theorem follow from Theorem 7. $\square$

For zero-residual problems, i.e., $c = 0$, Theorem 18 becomes the following.

Corollary 19. *Let* $\Omega \subseteq \mathbb{X}$ *be an open set, and let* $F : \Omega \to \mathbb{Y}$ *be a continuously differentiable function. Let* $x_* \in \Omega$ *and*

$$\beta := \|F'(x_*)^\dagger\|, \qquad \kappa := \sup\{t \in [0, R) : B(x_*, t) \subset \Omega\}.$$

Suppose that $F(x_*) = 0$, $F'(x_*)$ *is injective and that there exists a* $K > 0$ *such that*

$$\|F'(x) - F'(y)\| \leq K\|x - y\|, \quad \forall x, y \in B(x_*, \kappa).$$

Let

$$r := \min\{\kappa, 2/(3K\beta)\}.$$

Then, the Gauss–Newton method for solving (1), *with initial point* $x_0 \in B(x_*, r)/\{x_*\}$,

$$x_{k+1} = x_k - F'(x_k)^\dagger F(x_k), \quad k = 0, 1, \ldots,$$

is well defined, and the sequence generated $\{x_k\}$ *is contained in* $B(x_*, r)$ *and converges to* x_*, *which is the unique zero of* F *in* $B(x_*, 2/(\beta K))$. *Moreover, it holds that*

$$\|x_{k+1} - x_*\| \leq \frac{\beta K}{2(1 - \beta K \|x_0 - x_*\|)} \|x_k - x_*\|^2, \quad k = 0, 1, \ldots.$$

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

Remark 5. When $F'(x_*)$ is invertible, Corollary 19 merges with the results on the Newton method for solving nonlinear equations $F(x) = 0$, obtained by Ferreira in Theorem 3.1 and Remark 3.3 of [12].

In the following numerical example from Dedieu and Shub [10], we apply the results of this section.

Example 1. Let $F : \mathbb{R} \to \mathbb{R}^2$ such that $F(x) = (x, x^2 + c)^T$, where $c \in \mathbb{R}$ is given. Hence,

$$\min_{x \in \mathbb{R}} \|F(x)\|^2 = x^4 + (2c + 1)x^2 + c^2. \tag{14}$$

It is easy to check that $F'(x) = (1, 2x)^T$, $\|F'(x) - F'(y)\| = 2|x - y|$, for all $x, y \in \mathbb{R}$, and that $x_* = 0$ is a stationary point. Thus, $K = 2$ and $\beta = 1$. Therefore, we conclude from Theorem 18 that, if $|c| < \sqrt{2}/4$, then the Gauss–Newton method, for solving (14), with initial point $x_0 \in B(0, (1 - 2\sqrt{2}|c|)/3)$, is well defined, the sequence generated $\{x_k\}$ is contained in $B(0, (1 - 2\sqrt{2}|c|)/3)$, converges to 0, and

$$|x_{k+1}| \leq \frac{|x_k| + 2\sqrt{2}|c|}{1 - 2|x_0|} |x_k|, \quad k = 0, 1, \ldots.$$

Moreover, if $c = 0$, the sequence $\{x_k\}$ converges Q-quadratically for $x_* = 0$.

3.2. Convergence result under Smale's condition

In this section, we present a correspondent theorem to Theorem 7 under Smale's condition, see [22], which first appeared in [10].

Theorem 20. *Let $\Omega \subseteq \mathbb{X}$ be an open set, and let $F : \Omega \to \mathbb{Y}$ be an analytic function. Let $x_* \in \Omega$ and*

$$c := \|F(x_*)\|, \qquad \beta := \|F'(x_*)^\dagger\|, \qquad \kappa := \sup\{t > 0 : B(x_*, t) \subset \Omega\}.$$

Suppose that $F'(x_*)^* F(x_*) = 0$, $F'(x_*)$ is injective, and

$$\gamma := \sup_{n>1} \left\| \frac{F^{(n)}(x_*)}{n!} \right\|^{1/(n-1)} < +\infty, \qquad 2\sqrt{2}c\beta^2\gamma < 1. \tag{15}$$

Let $a := (2 + 3\beta - \sqrt{2}c\beta^2\gamma)$, $b := 4(1 + \beta)(1 - 2\sqrt{2}c\beta^2\gamma)$ and

$$r := \min\{\kappa, (a - \sqrt{a^2 - b})/(2\gamma(1 + \beta))\}.$$

Then, the Gauss–Newton method for solving (1), with initial point $x_0 \in B(x_*, r)/\{x_*\}$,

$$x_{k+1} = x_k - F'(x_k)^\dagger F(x_k), \quad k = 0, 1, \ldots,$$

is well defined, the sequence generated $\{x_k\}$ is contained in $B(x_*, r)$, converges to x_*, and

$$\|x_{k+1} - x_*\| \leq \frac{\beta\gamma}{(1 - \gamma\|x_0 - x_*\|)^2 - \beta\gamma(2\|x_0 - x_*\| - \gamma\|x_0 - x_*\|^2)} \|x_k - x_*\|^2$$

$$+ \frac{\sqrt{2}c\beta^2\gamma(2 - \gamma\|x_0 - x_*\|)}{(1 - \gamma\|x_0 - x_*\|)^2 - \beta\gamma(2\|x_0 - x_*\| - \gamma\|x_0 - x_*\|^2)} \|x_k - x_*\|, \quad k = 0, 1, \ldots.$$

Moreover, if $(a - \sqrt{a^2 - b})/(2\gamma(1 + \beta)) < \kappa$, then $r = (a - \sqrt{a^2 - b})/(2\gamma(1 + \beta))$ is the best possible convergence radius.

If, additionally, $4c\beta_0\gamma < 1$, then x_* is the unique stationary point of $F(x)^*F(x)$ in $B(x_*, \sigma)$, where $\sigma := (\omega_1 - \sqrt{\omega_1^2 - \omega_2})/(2\gamma(1 + \beta))$, $\omega_1 := (2 + \beta - c\beta_0)$, $\omega_2 := 4(1 + \beta)(1 - 2c\beta_0\gamma)$, $\beta_0 := \|[F'(x_*)^*F'(x_*)]^{-1}\|$.

Remark 6. The results in the above theorem appeared in [10], which also proved that a stationary point is a strict local minimum for (1).

We need the following result to prove the above theorem.

Lemma 21. *Let $\Omega \subseteq \mathbb{X}$ be an open set, and let $F : \Omega \to \mathbb{Y}$ be an analytic function. Suppose that $x_* \in \Omega$ and $B(x_*, 1/\gamma) \subset \Omega$, where γ is defined in (15). Then, for all $x \in B(x_*, 1/\gamma)$, it holds that*

$$\|F''(x)\| \leqslant (2\gamma)/(1 - \gamma\|x - x_*\|)^3.$$

Proof. Let $x \in \Omega$. Since F is an analytic function, we have

$$F''(x) = \sum_{n=0}^{\infty} \frac{1}{n!} F^{(n+2)}(x_*)(x - x_*)^n.$$

Combining (15) and the above equation, we obtain, after some simple calculus, that

$$\|F''(x)\| \leqslant \gamma \sum_{n=0}^{\infty} (n+2)(n+1)(\gamma \|x - x_*\|)^n.$$

On the other hand, as $B(x_*, 1/\gamma) \subset \Omega$, we have $\gamma \|x - x_*\| < 1$. So, from Proposition 4, we conclude that

$$\frac{2}{(1 - \gamma \|x - x_*\|)^3} = \sum_{n=0}^{\infty} (n+2)(n+1)(\gamma \|x - x_*\|)^n.$$

Combining the two above equations, we obtain the desired result. □

The next result gives a condition that is easier to check than condition (3), whenever the functions under consideration are twice continuously differentiable.

Lemma 22. *Let $\Omega \subseteq \mathbb{X}$ be an open set, $x_* \in \Omega$, and let $F : \Omega \to \mathbb{Y}$ be twice continuously on Ω. If there exists an $f : [0, R) \to \mathbb{R}$ that is twice continuously differentiable such that*

$$\|F''(x)\| \leqslant f''(\|x - x_*\|), \tag{16}$$

for all $x \in \Omega$, such that $\|x - x_*\| < R$, then F and f satisfy (3).

Proof. Taking $\tau \in [0, 1]$ and $x \in \Omega$, such that $x_* + \tau(x - x_*) \in \Omega$ and $\|x - x_*\| < R$, we obtain that

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

Now, as $\|x - x_*\| < R$ and f satisfies (16), we obtain from the last inequality that

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

Evaluating the latter integral, the statement follows. □

Proof of Theorem 20. Consider the real function $f : [0, 1/\gamma) \to \mathbb{R}$ defined by

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

It is straightforward to show that f is analytic and that

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

for $n \geq 2$. It follows from the last equalities that f satisfies (h1) and (h2). Since $2\sqrt{2}c\beta^2\gamma < 1$ and $4c\beta_0\gamma < 1$, conditions (h3) and (h4) also hold. Now, as $f''(t) = (2\gamma)/(1 - \gamma t)^3$, combining Lemmas 21 and 22, we conclude that F and f satisfy (3) with $R = 1/\gamma$. In this case, it is easy to see that the constants ν and ρ as defined in Theorem 7 satisfy

$$0 < \rho = (a - \sqrt{a^2 - b})/(2\gamma(1 + \beta)) < \nu = ((1 + \beta) - \sqrt{\beta(1 + \beta)})/(\gamma(1 + \beta)) < 1\gamma,$$

and, as a consequence, $0 < r = \min\{\kappa, \rho\}$. Moreover, it is not hard to see that

$$[\beta(\rho f'(\rho) - f(\rho)) + \sqrt{2}c\beta^2(f'(\rho) + 1)]/[\rho(1 - \beta(f'(\rho) + 1))] = 1,$$

and $[\beta(f(t)/t + 1) + c\beta_0(f'(t) + 1)/t] < 1$ for all $t \in (0, \sigma)$. Therefore, as $F, r, f,$ and x_* satisfy all hypotheses of Theorem 7, taking $x_0 \in B(x_*, r) \setminus \{x_*\}$, the statements of the theorem follow from Theorem 7. □

For zero-residual problems, i.e., $c = 0$, Theorem 20 becomes the following.

Corollary 23. *Let $\Omega \subseteq \mathbb{X}$ be an open set, and let $F : \Omega \to \mathbb{Y}$ be an analytic function. Let $x_* \in \Omega$, and*

$$\beta := \|F'(x_*)^\dagger\|, \qquad \kappa := \sup\{t > 0 : B(x_*, t) \subset \Omega\}.$$

Suppose that $F(x_*) = 0$, $F'(x_*)$ is injective and

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

Let

$$r := \min\{\kappa, (2 + 3\beta - \sqrt{\beta(8 + 9\beta)})/(2\gamma(1 + \beta))\}.$$

Then, the Gauss–Newton method for solving (1), *with initial point $x_0 \in B(x_*, r)/\{x_*\}$,*

$$x_{k+1} = x_k - F'(x_k)^\dagger F(x_k), \quad k = 0, 1, \ldots,$$

is well defined, is contained in $B(x_*, r)$, and converges to x_*, which is the unique zero of F in $B(x_*, 1/(\gamma(1+\beta)))$. Moreover, it holds that

$$\|x_{k+1} - x_*\| \leq \frac{\beta\gamma}{(1 - \gamma\|x_0 - x_*\|)^2 - \beta\gamma(2\|x_0 - x_*\| - \gamma\|x_0 - x_*\|^2)} \|x_k - x_*\|^2,$$
$$k = 0, 1, \ldots.$$

If, additionally, $(2+3\beta-\sqrt{\beta(8+9\beta)})/(2\gamma(1+\beta)) < \kappa$, then $r = (2+3\beta-\sqrt{\beta(8+9\beta)})/(2\gamma(1+\beta))$ is the best possible convergence radius.

Remark 7. When $F'(x_*)$ is invertible, Corollary 23 is similar to the results on the Newton method for solving nonlinear equations $F(x) = 0$, obtained by Ferreira in Theorem 3.4 of [12].

4. Final remark

In Theorem 3 of Dedieu and Shub [10], an α-condition is given for the convergence of the Gauss–Newton method. It would also be interesting to see whether that theorem holds if we replace the α-condition with the majorant condition.

Acknowledgments

References

[1] R.L. Adler, J.P. Dedieu, J.Y. Margulies, M. Martens, M. Shub, Newton's method on Riemannian manifolds and a geometric model for the human spine, IMA J. Numer. Anal. 22 (3) (2002) 359–390.
[2] 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.
[3] I.K. Argyros, S. Hilout, Improved generalized differentiability conditions for Newton-like methods, J. Complexity 22 (3) (2010) 316–333.
[4] L. Blum, F. Cucker, M. Shub, S. Smale, Complexity and Real Computation, Springer-Verlag, New York, 1998. With a foreword by Richard M. Karp.
[5] J. Chen, The convergence analysis of inexact Gauss–Newton methods for nonlinear problems, Comput. Optim. Appl. 40 (1) (2008) 97–118.
[6] J. Chen, W. Li, Convergence of Gauss–Newton's method and uniqueness of the solution, Appl. Math. Comput. 170 (1) (2005) 686–705.
[7] J. Chen, W. Li, Convergence behaviour of inexact Newton methods under weak Lipschitz condition, J. Comput. Appl. Math. 191 (1) (2006) 143–164.
[8] J.P. Dedieu, M.H. Kim, Newton's method for analytic systems of equations with constant rank derivatives, J. Complexity 18 (1) (2002) 187–209.

 [9] J.P. Dedieu, P. Priouret, G. Malajovich, Newton's method on Riemannian manifolds: covariant alpha theory, IMA J. Numer. Anal. 23 (3) (2003) 395–419.
[10] J.P. Dedieu, M. Shub, Newton's method for overdetermined systems of equations, Math. Comp. 69 (231) (2000) 1099–1115.
[11] J.E. Dennis Jr., R.B. Schnabel, Numerical methods for unconstrained optimization and nonlinear equations, in: Classics in Applied Mathematics, vol. 16, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996. Corrected reprint of the 1983 original.
[12] O.P. Ferreira, Local convergence of Newton's method in Banach space from the viewpoint of the majorant principle, IMA J. Numer. Anal. 29 (3) (2009) 746–759.
[13] O.P. Ferreira, M.L.N. Gonçalves, Local convergence analysis of inexact Newton-like methods under majorant condition. Comput. Optim. Appl., in press (doi:10.1007/s10589-009-9249-2).
[14] O.P. Ferreira, B.F. Svaiter, Kantorovich's majorants principle for Newton's method, Comput. Optim. Appl. 42 (2) (2009) 213–229.
[15] 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.
[16] L.V. Kantorovich, The principle of the majorant and Newton's method, Doklady Akad. Nauk SSSR (N.S.) 76 (1951) 17–20.
[17] C. Li, N. Hu, J. Wang, Convergence behavior of Gauss–Newton's method and extensions of the Smale point estimate theory, J. Complexity 26 (3) (2010) 268–295.
[18] C. Li, W.H. Zhang, X.Q. Jin, Convergence and uniqueness properties of Gauss–Newton's method, Comput. Math. Appl. 47 (6–7) (2004) 1057–1067.
[19] J. Nocedal, S.J. Wright, Numerical optimization, in: Springer Series in Operations Research, Springer-Verlag, New York, 1999.
[20] P.D. Proinov, General local convergence theory for a class of iterative processes and its applications to Newton's method, J. Complexity 25 (1) (2009) 38–62.
[21] P.D. Proinov, New general convergence theory for iterative processes and its applications to Newton–Kantorovich type theorems, J. Complexity 26 (1) (2010) 3–42.
[22] 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.
[23] G.W. Stewart, On the continuity of the generalized inverse, SIAM J. Appl. Math. 17 (1969) 33–45.
[24] X. Wang, Convergence of Newton's method and uniqueness of the solution of equations in Banach space, IMA J. Numer. Anal. 20 (1) (2000) 123–134.
[25] P.A. Wedin, Perturbation theory for pseudo-inverses, Nord. Tidskr. Informationsbehandling (BIT) 13 (1973) 217–232.