Optimization: A Journal of Mathematical Programming and Operations Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gopt20

Convergence of the Gauss-Newton method for a special class of systems of equations under a majorant condition

M.L.N. Gonçalves [a] & P.R. Oliveira [b]

[a] IME/UFG, Campus II - Caixa Postal 131, Goiânia, Brazil

[b] COPPE-Sistemas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

Version of record first published: 25 Mar 2013.

To cite this article: M.L.N. Gonçalves & P.R. Oliveira (2013): Convergence of the Gauss-Newton method for a special class of systems of equations under a majorant condition, Optimization: A Journal of Mathematical Programming and Operations Research, DOI:10.1080/02331934.2013.778854

To link to this article: http://dx.doi.org/10.1080/02331934.2013.778854

Optimization, 2013
http://dx.doi.org/10.1080/02331934.2013.778854

Convergence of the Gauss–Newton method for a special class of systems of equations under a majorant condition

M.L.N. Gonçalves[a]* and P.R. Oliveira[b]

[a] *IME/UFG, Campus II – Caixa Postal 131, Goiânia, Brazil;* [b] *COPPE-Sistemas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil*

(*Received 12 July 2012; final version received 11 February 2013*)

In this paper, we study the Gauss–Newton method for a special class of systems of non-linear equation. On the hypothesis that the derivative of the function under consideration satisfies a majorant condition, semi-local convergence analysis is presented. In this analysis, the conditions and proof of convergence are simplified by using a simple majorant condition to define regions where the Gauss–Newton sequence is 'well behaved'. Moreover, special cases of the general theory are presented as applications.

Keywords: Gauss–Newton method; majorant condition; non-linear systems of equations; semi-local convergence

1. Introduction

Consider the *systems of non-linear equations*

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

where $F : \Omega \to \mathbb{R}^m$ is a continuously differentiable function and $\Omega \subseteq \mathbb{R}^n$ is an open set.

When $F'(x)$ is invertible, the Newton method and its variant (see [1–4]) are the most efficient methods known for solving (1). However, when $F'(x)$ is not necessarily invertible, a generalized Newton method, called the Gauss–Newton method (see [5–7]), defined by

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

where $F'(x_k)^{\dagger}$ denotes the Moore–Penrose inverse of the linear operator $F'(x_k)$, finds least squares solutions of (1) which may or may not be solutions of (1). These least squares solutions are related to the non-linear least squares problem

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

that is, they are stationary points of $G(x) = \|F(x)\|^2$. It is worth noting that, if $F'(x)$ is surjective, then least squares solutions of systems of non-linear equations are also solutions of systems of non-linear equations.

*Corresponding author. Email: maxlng@mat.ufg.br

We shall consider the same special class of systems of non-linear equations studied in [8–10], i.e. systems of non-linear equations where the function F under consideration satisfies

$$\left\| F'(y)^{\dagger}(I_{\mathbb{R}^m} - F'(x)F'(x)^{\dagger})F(x) \right\| \leq \kappa \|x - y\|, \qquad \forall\, x, y \in \Omega \tag{2}$$

for some $0 \leq \kappa < 1$ and $I_{\mathbb{R}^m}$ denotes the identity operator on $\mathbb{R}^m$. This special class of non-linear systems of equation contains underdertermined systems with surjective derivatives, because when $F'(x)$ is surjective we can prove that $k = 0$ in (2).

In recent years, papers have addressed the issue of convergence of the Newton method, including the Gauss–Newton method, by relaxing the assumption of Lipschitz continuity of the derivative (see [1–7,9–15] and references therein). These new assumptions also allow us to unify previously unrelated convergence results, namely results for analytical functions (α-theory or γ-theory) and the classical results for functions with Lipschitz derivative. The main new conditions that relax the condition of Lipschitz continuity of the derivative include the majorant condition, which we will use, and Wang's condition, introduced in [13] and used for example in [10,14,15] to study the Gauss–Newton method. In fact, on the hypothesis of this paper, it can be shown that these conditions are equivalent. In a way, however, the formulation as a majorant condition is better than Wang's condition, as it provides a clear relationship between the majorant function and the non-linear function under consideration. Besides, the majorant condition provides a simpler proof of convergence.

Following the ideas of the semi-local convergence analysis in [4,6], we will present a new semi-local convergence analysis of the Gauss–Newton method for solving (1), where F satisfies (2), under a majorant condition. The convergence analysis presented here communicates the conditions and proof quite simply. This is possible thanks to our majorant condition and to a demonstration technique introduced in [4] which, instead of looking only to the sequence generated, identifies regions where, for the problem under consideration, the Gauss–Newton sequence is well behaved as compared with a method applied to an auxiliary function associated with the majorant function. Moreover, two unrelated previous results relating to the Gauss–Newton method are unified, namely, results for analytical functions under an α-condition and the classical result for functions with Lipschitz derivative. Besides, convergence results for underdetermined systems with surjective derivatives will be also given.

The paper is organized as follows. Section 1.1 lists some notations and basic results used in the presentation. Section 2 states and proves the main results. Finally, special cases of the general theory are presented as applications in Section 3.

1.1. *Notation and auxiliary results*

The following notations and results are used throughout this presentation. Let $\mathbb{R}^n$ have a norm $\|.\|$. The open and closed balls at $a \in \mathbb{R}^n$ and radius $\delta > 0$ are denoted, respectively by

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

Let $\Omega \subseteq \mathbb{R}^n$ be an open set. The Fréchet derivative of $F : \Omega \to \mathbb{R}^m$ is the linear map $F'(x) : \mathbb{R}^n \to \mathbb{R}^m$. If φ is a real-valued function and u_0 be in the domain of φ, we use $D^-\varphi(u_0)$ to denote left-sided derivative of φ at u_0.

Given a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ (or an $m \times n$ matrix), the Moore–Penrose inverse of A is the linear operator $A^\dagger : \mathbb{R}^m \to \mathbb{R}^n$ (or an $n \times m$ matrix) which satisfies:

$$AA^\dagger A = A, \quad A^\dagger AA^\dagger = A^\dagger, \quad (AA^\dagger)^* = AA^\dagger, \quad (A^\dagger A)^* = A^\dagger A,$$

where A^* denotes the adjoint of A. The Kernel and image of A are denoted by $Ker(A)$ and $im(A)$, respectively. It is easily seen from the definition of the Moore–Penrose inverse that

$$A^\dagger A = \Pi_{Ker(A)^\perp}, \qquad AA^\dagger = \Pi_{im(A)}, \tag{3}$$

where Π_E denotes the projection of $\mathbb{R}^n$ onto subspace E.

We use $I_{\mathbb{R}^m}$ to denote the identity operator on $\mathbb{R}^m$. If A is surjective, then

$$A^\dagger = A^*(AA^*)^{-1}, \qquad AA^\dagger = I_{\mathbb{R}^m}. \tag{4}$$

Lemma 1 (Banach's Lemma) *Let $B : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous linear operator. If $\|B - I_{\mathbb{R}^n}\| < 1$, then B is invertible and $\|B^{-1}\| \le 1/\left(1 - \|B - I_{\mathbb{R}^n}\|\right)$.*

Proof See the proof of Lemma 1, p. 189 of Smale [16] with $A = I_{\mathbb{R}^n}$ and $c = \|B - I_x\|$. □

The next lemma is proved on p. 43 of [17] (see also [18]). It is on the perturbation of the Moore–Penrose inverse of A.

Lemma 2 *Let $A, B : \mathbb{R}^n \to \mathbb{R}^m$ be continuous linear operators. Assume that*

$$1 \le rank(B) \le rank(A), \qquad \|A^\dagger\|\|A - B\| < 1.$$

Then

$$rank(A) = rank(B), \qquad \|B^\dagger\| \le \frac{\|A^\dagger\|}{1 - \|A^\dagger\|\|A - B\|}.$$

2. Semi-local analysis for the Gauss–Newton method

Our goal is to state and prove a semi-local theorem of the Gauss–Newton method for solving non-linear systems of equations, where the function under consideration satisfies (2). First, we will prove that this theorem holds for an auxiliary function associated with the majorant function. Then, we will prove well-definedness of the Gauss–Newton method and convergence. Convergence rates will also be established. The statement of the theorem is:

Theorem 3 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set and $F : \Omega \to \mathbb{R}^m$ a continuously differentiable function. Suppose that*

$$\left\| F'(y)^\dagger (I_{\mathbb{R}^m} - F'(x)F'(x)^\dagger)F(x) \right\| \le \kappa \|x - y\|, \qquad \forall\, x, y \in \Omega \tag{5}$$

for some $0 \le \kappa < 1$. Take $x_0 \in \Omega$ such that $\beta := \|F'(x_0)^\dagger F(x_0)\| > 0$, $F'(x_0) \neq 0$ and

$$rank(F'(x)) \le rank(F'(x_0)), \qquad \forall\, x \in \Omega. \tag{6}$$

Suppose that there exist $R > 0$ and a continuously differentiable function $f : [0, R) \to \mathbb{R}$ such that, $B(x_0, R) \subseteq \Omega$,

$$\|F'(x_0)^\dagger\|\|F'(y) - F'(x)\| \leq f'(\|y - x\| + \|x - x_0\|) - f'(\|x - x_0\|), \tag{7}$$

for any $x, y \in \Omega$, $\|x - x_0\| + \|y - x\| < R$, and moreover,

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

Take $\lambda \geq 0$ such that $\lambda \geq -\kappa f'(\beta)$ and consider the auxiliary function $h_\lambda : [0, R) \to \mathbb{R}$,

$$h_\lambda(t) := \beta + \lambda t + f(t). \tag{8}$$

If h_λ satisfies

(h3) $h_\lambda(t) = 0$ *for some $t \in (0, R)$,*

then $h_\lambda(t)$ has a smallest zero $t_\lambda^* \in (0, R)$, the sequences for solving $h_\lambda(t) = 0$ and $F(x) = 0$, with starting points $t_{\lambda,0} = 0$ and x_0, respectively,

$$t_{\lambda,k+1} = t_{\lambda,k} - h_0'(t_{\lambda,k})^{-1} h_\lambda(t_{\lambda,k}), \quad x_{k+1} = x_k - F'(x_k)^\dagger F(x_k), \quad k = 0, 1, \ldots, \tag{9}$$

are well defined, $\{t_{\lambda,k}\}$ is strictly increasing, is contained in $[0, t_\lambda^*)$, and converges to t_λ^*, $\{x_k\}$ is contained in $B(x_0, t_\lambda^*)$, converges to a point $x_* \in B[x_0, t_\lambda^*]$ such that $F'(x_*)^\dagger F(x_*) = 0$,

$$\|x_{k+1} - x_k\| \leq t_{\lambda,k+1} - t_{\lambda,k}, \qquad \|x_* - x_k\| \leq t_\lambda^* - t_{\lambda,k}, \qquad k = 0, 1, \ldots, \tag{10}$$

and

$$\|x_{k+1} - x_k\| \leq \frac{t_{\lambda,k+1} - t_{\lambda,k}}{(t_{\lambda,k} - t_{\lambda,k-1})^2} \|x_k - x_{k-1}\|^2, \quad k = 1, 2, \ldots. \tag{11}$$

Moreover, if $\lambda = 0$, the sequences $\{t_{\lambda,k}\}$ and $\{x_k\}$ converge Q-linearly and R-linearly (or, if $\lambda = 0$ and $h_0'(t_\lambda^*) < 0$, Q-quadratically and R-quadratically) to t_λ^* and x_*, respectively.

Remark 1 It is easily seen that the best choice of λ is the smallest possible. Hence, if $f'(\beta) \leq 0$ then $\lambda = -\kappa f'(\beta)$ is the best choice. Moreover, since $-f'(\beta) < -f'(0) = 1$ (**h2**), a possible choice for λ is κ, despite not being the best.

Remark 2 If $F'(x)$ is surjective, it follows from the second equation in (4) that $F'(x) F'(x)^\dagger = I_{\mathbb{R}^m}$. Thus, we can take $\lambda = 0$, because F satisfies (5) with $\kappa = 0$. Therefore, in this case, Theorem 3 extends the results obtained by Ferreira and Svaiter in Theorem 2 of [4].

From now on, we assume that the hypotheses of Theorem 3 hold.

2.1. *The auxiliary function and sequence* $\{t_{\lambda,k}\}$

In this section, we will study the auxiliary function, h_λ, which is associated with the majorant function, f, and prove all results regarding only the sequence $\{t_{\lambda,k}\}$. Remember that a function that satisfies (7), **(h1)** and **(h2)** is called a majorant function for the function F on $B(x_0, R)$. More details about the majorant condition can be found in [1–7].

PROPOSITION 4 *The following statements hold:*

(i) $h_\lambda(0) = \beta > 0$, $h'_\lambda(0) = \lambda - 1$;
(ii) h'_λ *is convex and strictly increasing.*

Proof It follows from the definition in (8) and assumptions **(h1)** and **(h2)**. □

PROPOSITION 5 *The function h_λ has a smallest root $t^*_\lambda \in (0, R)$, is strictly convex, and*

$$h_\lambda(t) > 0, \quad h'_0(t) < 0, \qquad t < t - h_\lambda(t)/h'_0(t) < t^*_\lambda, \qquad \forall\, t \in [0, t^*_\lambda). \tag{12}$$

Moreover, $h'_0(t^*_\lambda) \le 0$.

Proof As h_λ is continuous in $[0, R)$ and has a zero there (**(h3)**), it must have a smallest zero t^*_λ, which is greater than 0 because $h_\lambda(0) = \beta > 0$. Since, from item **(ii)** of Proposition 4, h'_λ is strictly increasing, then h_λ is strictly convex.

The first inequality in (12) follows from the assumption $h_\lambda(0) = \beta > 0$ and the definition of t^*_λ as the smallest root of h_λ. Since h_λ is strictly convex,

$$0 = h_\lambda(t^*_\lambda) > h_\lambda(t) + h'_\lambda(t)(t^*_\lambda - t), \qquad t \in [0, R),\ t \ne t^*_\lambda. \tag{13}$$

If $t \in [0, t^*_\lambda)$ then $h_\lambda(t) > 0$ and $t^*_\lambda - t > 0$, which combined with (13) yields $h'_\lambda(t) < 0$ for all $t \in [0, t^*_\lambda)$. Hence, using $\lambda \ge 0$ and $h'_\lambda(t) = \lambda + h'_0(t)$ for all $t \in [0, t^*_\lambda)$, the second inequality in (12) follows. The third inequality in (12) follows from the first and the second inequalities.

To prove the last inequality in (12), note that dividing the inequality in (13) by $-h'_\lambda(t)$ (which is strictly positive), together with some simple algebraic manipulations, gives

$$t - h_\lambda(t)/h'_\lambda(t) < t^*_\lambda, \quad \forall\, t \in [0, t^*_\lambda),$$

which, using the first inequality in (12) and $0 < -h'_\lambda(t) \le -h'_0(t)$ for all $t \in [0, t^*_\lambda)$, yields the desired inequality.

Since $h_\lambda > 0$ in $[0, t^*_\lambda)$ and $h_\lambda(t^*_\lambda) = 0$, we must have $h'_\lambda(t^*_\lambda) \le 0$. Thus, the last inequality of the proposition follows from the fact that $h'_\lambda(t^*_\lambda) = \lambda + h'_0(t^*_\lambda)$ and $\lambda \ge 0$. □

In view of the second inequality in (12), the following iteration map for h_λ is well defined in $[0, t^*_\lambda)$. Denoting this by n_λ:

$$\begin{aligned} n_\lambda : [0, t^*_\lambda) &\to \mathbb{R} \\ t &\mapsto t - h_\lambda(t)/h'_0(t). \end{aligned} \tag{14}$$

Note that in the case where $\lambda = 0$, the sequence n_λ reduces to a Newton sequence, which Ferreira and Svaiter used in [4] to obtain a semi-local convergence analysis of the Newton method under a majorant condition.

PROPOSITION 6 *For each $t \in [0, t^*_\lambda)$ it holds that $\beta \le n_\lambda(t) < t^*_\lambda$.*

Proof Proposition 5 implies that h_λ is convex. Hence, using item **(i)** of Proposition 4, it is easy to see, by using convexity properties, that $(1 - \lambda)t - \beta \ge -h_\lambda(t)$, which combined with $\lambda \geq 0$ gives $t - \beta \geq -h_\lambda(t)$. Accordingly, the above definition implies that

$$n_\lambda(t) - \beta = t - \frac{h_\lambda(t)}{h_0'(t)} - \beta \geq -h_\lambda(t) - \frac{h_\lambda(t)}{h_0'(t)} = \frac{h_\lambda(t)}{-h_0'(t)}[h_0'(t) + 1], \qquad \forall\, t \in [0, t_\lambda^*).$$

Proposition 4 implies that $h_0'(0) = -1$ and h_0' is strictly increasing. Thus, we obtain $h_0'(t) + 1 \geq 0$, for all $t \in [0, t_\lambda^*)$. Therefore, combining the above inequality with the first two inequalities in (12), the first inequality of the proposition follows. To prove the last inequality of the proposition, combine (14) with the last inequality in (12).

PROPOSITION 7 *Iteration map n_λ maps $[0, t_\lambda^*)$ in $[0, t_\lambda^*)$, and it holds that*

$$t < n_\lambda(t), \qquad \forall\, t \in [0, t_\lambda^*).$$

Moreover, if $\lambda = 0$ or $\lambda = 0$ and $h_0'(t_\lambda^*) < 0$, we have the following inequalities, respectively,

$$t_\lambda^* - n_\lambda(t) \leqslant \frac{1}{2}(t_\lambda^* - t), \qquad t_\lambda^* - n_\lambda(t) \leq \frac{D^- h_0'(t_\lambda^*)}{-2h_0'(t_\lambda^*)}(t_\lambda^* - t)^2, \qquad \forall\, t \in [0, t_\lambda^*).$$

Proof The first two statements of the proposition follow trivially from the last inequalities in (12) and (14). Now, if $\lambda = 0$, then the sequence in (14) reduces to a Newton sequence. Hence, the second part of the proof follows the same pattern as the proof of Proposition 4 of [4] with $f = h_0$. □

The definition of $\{t_{\lambda,k}\}$ in Theorem 3 is equivalent to the following one

$$t_{\lambda,0} = 0, \quad t_{\lambda,k+1} = n_\lambda(t_{\lambda,k}), \qquad k = 0, 1, \ldots . \tag{15}$$

From which, using also Proposition 7, it is easy to prove that

COROLLARY 8 *The sequence $\{t_{\lambda,k}\}$ is well defined, is strictly increasing, is contained in $[0, t_\lambda^*)$, and converges to t_λ^*.*

Moreover, if $\lambda = 0$ or $\lambda = 0$ and $h_0'(t_\lambda^*) < 0$, the sequence $\{t_{\lambda,k}\}$ converges Q-linearly or Q-quadratically to t_λ^*, respectively, as follows

$$t_\lambda^* - t_{\lambda,k+1} \leq \frac{1}{2}(t_\lambda^* - t_{\lambda,k}), \qquad t_\lambda^* - t_{\lambda,k+1} \leq \frac{D^- h_0'(t_\lambda^*)}{-2h_0'(t_\lambda^*)}(t_\lambda^* - t_{\lambda,k})^2, \quad k = 0, 1, \ldots .$$

Hence, all statements involving only $\{t_{\lambda,k}\}$ in Theorem 3 are valid.

2.2. *Convergence*

In this section, we will prove well-definedness and convergence of the sequence $\{x_k\}$ specified in (9) in Theorem 3.

We start with two lemmas that highlight the relationships between the majorant function f and the non-linear function F.

PROPOSITION 9 *If $\|x - x_0\| \leq t < t_\lambda^*$, then $rank(F'(x)) = rank(F'(x_0)) \geq 1$ and*

$$\left\| F'(x)^\dagger \right\| \leq -\|F'(x_0)^\dagger\|/h_0'(t).$$

In particular, $rank(F'(x)) = rank(F'(x_0))$ *in* $B(x_0, t_\lambda^*)$.

Proof Take $x \in B[x_0, t]$, $0 \le t < t_\lambda^*$. Using the assumptions (7), **(h1)**, **(h2)**, $f'(t) = h_0'(t)$ and the second inequality in (12), we obtain

$$\|F'(x_0)^\dagger\|\|F'(x) - F'(x_0)\| \leqslant f'(\|x - x_0\|) - f'(0) \leqslant f'(t) + 1 = h_0'(t) + 1 < 1.$$

Combining the last inequality with (6) and Lemma 2, we conclude that $rank(F'(x)) = rank(F'(x_0)) \ge 1$ and

$$\|F'(x)^\dagger\| \leqslant \frac{\|F'(x_0)^\dagger\|}{1 - (f'(t) + 1)} = \frac{\|F'(x_0)^\dagger\|}{-f'(t)} = -\frac{\|F'(x_0)^\dagger\|}{h_0'(t)}.$$

□

It is convenient to study the linearization error of F at points in Ω. For that purpose, we define

$$E_F(x, y) := F(y) - \left[F(x) + F'(x)(y - x)\right], \qquad y,\, x \in \Omega. \tag{16}$$

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

$$e_f(t, u) := f(u) - \left[f(t) + f'(t)(u - t)\right], \qquad t,\, u \in [0, R). \tag{17}$$

LEMMA 10 *Take*

$$x, y \in B(x_0, R) \quad \text{and} \quad 0 \le t < v < R.$$

If $\|x - x_0\| \leqslant t$ *and* $\|y - x\| \leqslant v - t$, *then*

$$\|F'(x_0)^\dagger\|\|E_F(x, y)\| \leqslant e_f(t, v)\frac{\|y - x\|^2}{(v - t)^2}.$$

Proof The proof follows the same pattern as the proof of Lemma 7 of [4]. □

Proposition 9 guarantees, in particular, that $rank(F'(x)) \ge 1$ for all $x \in B(x_0, t_\lambda^*)$ 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:

$$\begin{aligned} G_F : B(x_0, t_\lambda^*) &\to \mathbb{R}^n \\ x &\mapsto x - F'(x)^\dagger F(x). \end{aligned} \tag{18}$$

One can apply a *single* Gauss–Newton iteration on any $x \in B(x_0, t_\lambda^*)$ to obtain $G_F(x)$, which may not belong to $B(x_0, t_\lambda^*)$ or even may not belong to the domain of F. Therefore, this is enough to guarantee well definedness of only one iteration. To ensure that Gauss–Newton iterations may be repeated indefinitely, we need some additional results.

First, we define some subsets of $B(x_0, t_\lambda^*)$, and we will prove that the desired inclusion holds for all points in these subsets.

$$K(t) := \left\{x \in \Omega \,:\, \|x - x_0\| \le t,\; \|F'(x)^\dagger F(x)\| \leqslant -\frac{h_\lambda(t)}{h_0'(t)}\right\}, \qquad t \in [0, t_\lambda^*), \tag{19}$$

$$K := \bigcup_{t \in [0, t_\lambda^*)} K(t). \tag{20}$$

In (19), $0 \leqslant t < t_\lambda^*$, therefore, $h_0'(t) \neq 0$ and $rank(F'(x)) \geq 1$ in $B[x_0, t] \subset B[x_0, t_\lambda^*)$ (Proposition 9). Hence, the definitions are consistent.

LEMMA 11 *For each $t \in [0, t_\lambda^*)$, it holds that:*

(i) $K(t) \subset B(x_0, t_\lambda^*)$;
(ii) $\|G_F(G_F(x)) - G_F(x)\| \leq -\frac{h_\lambda(n_\lambda(t))}{h_0'(n_\lambda(t))}\left(\frac{\|G_F(x)-x\|}{n_\lambda(t)-t}\right)^2, \quad \forall x \in K(t)$;
(iii) $G_F\left(K(t)\right) \subset K\left(n_\lambda(t)\right)$.

As a consequence, $K \subset B(x_0, t_\lambda^*)$ and $G_F(K) \subset K$.

Proof Item **(i)** follows trivially from the definition of $K(t)$.

Take $t \in [0, t_\lambda^*)$, $x \in K(t)$. Using definition (19) and the first two statements in Proposition 7, we have

$$\|x - x_0\| \leq t, \qquad \|F'(x)^\dagger F(x)\| \leq -h_\lambda(t)/h_0'(t), \quad t < n_\lambda(t) < t_\lambda^*. \tag{21}$$

Therefore,

$$\begin{aligned}\|G_F(x) - x_0\| &\leqslant \|x - x_0\| + \|G_F(x) - x\| = \|x - x_0\| + \|F'(x)^\dagger F(x)\| \\ &\leqslant t - h_\lambda(t)/h_0'(t) = n_\lambda(t) < t_\lambda^*,\end{aligned}$$

and

$$G_F(x) \in B[x_0, n_\lambda(t)] \subset B(x_0, t_\lambda^*). \tag{22}$$

Since $G_F(x), n_\lambda(t)$ belong to the domains of F and f, respectively, using the definitions in (14) and (18), $h_\lambda(t) = \beta + \lambda t + f(t)$, linearization errors (16) and (17) and some algebraic manipulations, we obtain

$$\begin{aligned}h_\lambda(n_\lambda(t)) &= h_\lambda(n_\lambda(t)) - \left[h_\lambda(t) + h_0'(t)(n_\lambda(t) - t)\right] \\ &= e_f(t, n_\lambda(t)) - \lambda h_\lambda(t)/h_0'(t)\end{aligned} \tag{23}$$

and

$$\begin{aligned}F(G_F(x)) &= F(G_F(x)) - \left[F(x) + F'(x)(G_F(x) - x)\right] + (I_{\mathbb{R}^m} - F'(x)F'(x)^\dagger)F(x) \\ &= E_F(x, G_F(x)) + (I_{\mathbb{R}^m} - F'(x)F'(x)^\dagger)F(x).\end{aligned}$$

The last equation, together with simple algebraic manipulations, implies that

$$\begin{aligned}\|F'(G_F(x))^\dagger F(G_F(x))\| \leq\ & \|F'(G_F(x))^\dagger\|\|E_F(x, G_F(x))\| \\ &+ \|F'(G_F(x))^\dagger(I_{\mathbb{R}^m} - F'(x)F'(x)^\dagger)F(x)\|.\end{aligned}$$

As $\|G_F(x) - x_0\| \leq n_\lambda(t)$, it follows from Proposition 9 that $rank(F'(G_F(x))) \geq 1$ and

$$\|F'(G_F(x))^\dagger\| \leq -\|F'(x_0)^\dagger\|/h_0'(n_\lambda(t)).$$

From the two latter equations and (5) we have

$$\|F'(G_F(x))^\dagger F(G_F(x))\| \leq -\frac{\|F'(x_0)^\dagger\|}{h_0'(n_\lambda(t))}\|E(x, G_F(x))\| + \kappa\|G_F(x) - x\|.$$

On the other hand, using (21), Lemma (10) and (23) we have

$$\begin{aligned}\|F'(x_0)^\dagger\|\|E_F(x, G_F(x))\| &\leq e_f(t, n_\lambda(t))\left(\frac{\|G_F(x) - x\|}{n_\lambda(t) - t}\right)^2 \\ &\leq h_\lambda(n_\lambda(t))\left(\frac{\|G_F(x) - x\|}{n_\lambda(t) - t}\right)^2 + \lambda h_\lambda(t)/h_0'(t).\end{aligned}$$

Thus, the last two equations, together with the second equation in (21), imply

$$\begin{aligned}\|F'(G_F(x))^\dagger F(G_F(x))\| \leq \frac{-h_\lambda(n_\lambda(t))}{h_0'(n_\lambda(t))}&\left(\frac{\|G_F(x) - x\|}{n_\lambda(t) - t}\right)^2 \\ &+ (\kappa + \lambda(h_0'(n_\lambda(t)))^{-1})(-h_\lambda(t)/h_0'(t)).\end{aligned}$$

Taking $\lambda \geq -\kappa f'(\beta)$, the second inequality in (12) and (21), we obtain

$$\left(\kappa + \lambda(h_0'(n_\lambda(t)))^{-1}\right) \leq \kappa\left(1 - f'(\beta)(h_0'(n_\lambda(t)))^{-1}\right).$$

As $f'(t) = h_0'(t)$, using Proposition 6, **(h2)** and the second inequality in (12), we have

$$\kappa\left(1 - f'(\beta)(h_0'(n_\lambda(t))^{-1}\right) = \kappa\left(h_0'(\beta) - h_0'(n_\lambda(t))\right)(-h_0'(n_\lambda(t))^{-1} \leq 0.$$

Combining the three above inequalities with the first two inequalities in (12), we conclude

$$\|F'(G_F(x))^\dagger F(G_F(x))\| \leq \frac{-h_\lambda(n_\lambda(t))}{h_0'(n_\lambda(t))}\left(\frac{\|G_F(x) - x\|}{n_\lambda(t) - t}\right)^2.$$

Therefore, item **(ii)** follows from the last inequality and (18). Now, the last inequality combined with (14), (18) and the second inequality in (21) becomes

$$\|F'(G_F(x))^\dagger F(G_F(x))\| \leq \frac{-h_\lambda(n_\lambda(t))}{h_0'(n_\lambda(t))}.$$

This result, together with (22), shows that $G_F(x) \in K(n_\lambda(t))$, which proves item **(iii)**.

The next inclusion (the first in the second part), follows trivially from definitions (19) and (20). To check the last inclusion, take $x \in K$. Then $x \in K(t)$ for some $t \in [0, t_\lambda^*)$. Using item **(iii)** of the lemma, we conclude that $G_F(x) \in K(n_\lambda(t))$. To end the proof, note that $n_\lambda(t) \in [0, t_\lambda^*)$ and use the definition of K. □

Finally, we are ready to prove the main result of this section, which is an immediate consequence of the latter result. First note that the sequence $\{x_k\}$ (see (9)) satisfies

$$x_{k+1} = G_F(x_k), \qquad k = 0, 1, \dots, \tag{24}$$

which is indeed an equivalent definition of this sequence.

COROLLARY 12 *The sequence $\{x_k\}$ is well defined, is contained in $B(x_0, t_\lambda^*)$, converges to a point $x_* \in B[x_0, t_\lambda^*]$ such that $F'(x_*)^\dagger F(x_*) = 0$, and $\{x_k\}$ and $\{t_{\lambda,k}\}$ satisfy (10) and (11).*

Moreover, if $\lambda = 0$, the sequence $\{x_k\}$ converges R-linearly (or, if $\lambda = 0$ and $h_0'(t_\lambda^*) < 0$, R-quadratically) to x_*.

Proof Since $\|F'(x_0)^\dagger F(x_0)\| = \beta$, using item **(i)** of Proposition 4, we have

$$x_0 \in K(0) \subset K,$$

where the second inclusion follows trivially from (20). Using the above equation, the inclusions $G_F(K) \subset K$ (Lemma 11) and (24), we conclude that the sequence $\{x_k\}$ is well defined and lies in K. From the first inclusion in the second part of Lemma 11, we have trivially that $\{x_k\}$ is contained in $B(x_0, t_\lambda^*)$.

We will prove, by induction, that

$$x_k \in K(t_{\lambda,k}), \qquad k = 0, 1, \dots . \tag{25}$$

The above inclusion, for $k = 0$, is the first result in this proof. Assume now that $x_k \in K(t_{\lambda,k})$. Thus, using item **(iii)** of Lemma 11, (15) and (24), we conclude that $x_{k+1} \in K(t_{\lambda,k+1})$, which completes the induction proof of (25).

Now, using (25) and (19), we have

$$\|F'(x_k)^\dagger F(x_k)\| \le -h_\lambda(t_{\lambda,k})/h_0'(t_{\lambda,k}), \qquad k = 0, 1, \dots ,$$

which, using (9), becomes

$$\|x_{k+1} - x_k\| \le t_{\lambda,k+1} - t_{\lambda,k}, \qquad k = 0, 1, \dots .$$

So, the first inequality in (10) holds. As $\{t_{\lambda,k}\}$ converges to t_λ^*, the last inequality implies that

$$\sum_{k=k_0}^{\infty} \|x_{k+1} - x_k\| \le \sum_{k=k_0}^{\infty} t_{\lambda,k+1} - t_{\lambda,k} = t_\lambda^* - t_{\lambda,k_0} < +\infty,$$

for any $k_0 \in \mathbb{N}$. Hence, $\{x_k\}$ is a Cauchy sequence in $B(x_0, t_\lambda^*)$, and so converges to some $x_* \in B[x_0, t_\lambda^*]$. The last inequality also implies that the second inequality in 10 holds.

To prove that $F'(x_*)^\dagger F(x_*) = 0$, note that, with simple algebraic manipulation, (5) and (9), we obtain

$$\begin{aligned}\|F'(x_*)^\dagger F(x_k)\| &\leqslant \|F'(x_*)^\dagger \big(I - F'(x_k)F'(x_k)^\dagger\big) F(x_k)\| \\ &\quad + \|F'(x_*)^\dagger\| \|F'(x_k)F'(x_k)^\dagger F(x_k)\| \\ &\le \kappa \|x_k - x_*\| + \|F'(x_*)^\dagger\| \|F'(x_k)\| \|x_{k+1} - x_k\|.\end{aligned}$$

As F is continuously differentiable, we can take the limit in the last inequality to conclude that $F'(x_*)^\dagger F(x_*) = 0$.

Since $x_k \in K(t_{\lambda,k})$, for all $k = 0, 1, \dots$, the inequality in (11) follows by applying item **(ii)** of the Lemma 11 with $x = x_{k-1}$ and $t = t_{\lambda,k-1}$ and using the definitions in (15) and (24).

To end the proof, combine the second inequality in (10) with the last part of the Corollary 8. □

Therefore, it follows from Corollaries 8 and 12 that all statements in Theorem 3 are valid.

3. Special cases

In this section, we present some special cases of Theorem 3.

3.1. *Convergence result for $F'(x_0)$ surjective*

In this section, we present a theorem on the hypothesis that $F'(x_0)$ is surjective. In this case, we can use a majorant condition, which gives the propriety that $\{x_k\}$ is invariant under the function $\bar{F} \to A^\dagger F$, where $A : \mathbb{R}^n \to \mathbb{R}^m$ is any surjective linear operator.

THEOREM 13 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set and $F : \Omega \to \mathbb{R}^m$ a continuously differentiable function. Take $x_0 \in \Omega$ such that $\beta := \|F'(x_0)^\dagger F(x_0)\| > 0$ and $F'(x_0)$ is surjective. Suppose that there exist $R > 0$ and a continuously differentiable function $\bar{f} : [0, R) \to \mathbb{R}$ such that, $B(x_0, R) \subseteq \Omega$,*

$$\|F'(x_0)^\dagger (F'(y) - F'(x))\| \leq \bar{f}'(\|y - x\| + \|x - x_0\|) - \bar{f}'(\|x - x_0\|), \tag{26}$$

for any $x, y \in \Omega$, $\|x - x_0\| + \|y - x\| < R$, and moreover,

(h1) $\bar{f}(0) = 0$, $\bar{f}'(0) = -1$*;*
(h2) $\bar{f}'$ *is convex and strictly increasing.*

Consider the auxiliary function $h : [0, R) \to \mathbb{R}$,

$$h(t) := \beta + \bar{f}(t). \tag{27}$$

If h satisfies

(h3) $h(t) = 0$ *for some $t \in (0, R)$,*

then $h(t)$ has a smallest zero $t_* \in (0, R)$, the sequences for solving $h(t) = 0$ and $F(x) = 0$, with starting points $t_0 = 0$ and x_0, respectively,

$$t_{k+1} = t_k - h'(t_k)^{-1} h(t_k), \quad x_{k+1} = x_k - F'(x_k)^\dagger F(x_k), \quad k = 0, 1, \ldots, \tag{28}$$

are well defined, $\{t_k\}$ is strictly increasing, is contained in $[0, t_*)$, and converges Q-linearly to t_*, $\{x_k\}$ is contained in $B(x_0, t_*)$, and converges R-linearly to a point $x_* \in B[x_0, t_*]$ such that $F'(x_*)^\dagger F(x_*) = 0$,

$$\|x_{k+1} - x_k\| \leq t_{k+1} - t_k, \qquad \|x_* - x_k\| \leq t_* - t_k, \qquad k = 0, 1, \ldots, \tag{29}$$

$$\|x_{k+1} - x_k\| \leq \frac{t_{k+1} - t_k}{(t_k - t_{k-1})^2} \|x_k - x_{k-1}\|^2, \quad k = 1, 2, \ldots,$$

and

$$\|F'(x_0)^\dagger F(x_k)\| \leq \left(\frac{t_{k+1} - t_k}{t_k - t_{k-1}}\right) \|F'(x_0)^\dagger F(x_{k-1})\|, \quad k = 1, 2, \ldots. \tag{30}$$

If, additionally, $h'(t_*) < 0$, then the sequences $\{t_k\}$ and $\{x_k\}$ converge Q-quadratically and R-quadratically to t_* and x_*, respectively.

Proof Let $\bar{F} : \Omega \to \mathbb{R}^m$ be defined by

$$\bar{F}(x) = F'(x_0)^{\dagger} F(x), \qquad x \in \Omega. \tag{31}$$

On the hypothesis of the theorem, we will prove that $\bar{F}$ satisfies all assumptions of Theorem 3. Hence, with the exception of (30), the statements of the theorem follow from Theorem 3.

First of all, as $F'(x_0)$ is surjective, it follows from 4 that

$$F'(x_0)F'(x_0)^{\dagger} = I_{\mathbb{R}^m}. \tag{32}$$

Now, take $x \in B[x_0, t]$, $0 \le t \le t_*$. Using the assumptions (26), **(h1)** and **(h2)**, we obtain

$$\|F'(x_0)^{\dagger}[F'(x) - F'(x_0)]\| \leqslant \bar{f}'(\|x - x_0\|) - \bar{f}'(0) \leqslant \bar{f}'(t) + 1 < 1.$$

Using Lemma 1 and the above equation, we conclude that $\big(I_{\mathbb{R}^n} - F'(x_0)^{\dagger}(F'(x_0) - F'(x))\big)$ is non-singular and

$$\|\big(I_{\mathbb{R}^n} - F'(x_0)^{\dagger}(F'(x_0) - F'(x))\big)^{-1}\| \leqslant \frac{1}{1 - \big(\bar{f}'(t) + 1\big)} = -\frac{1}{\bar{f}'(t)} = -\frac{1}{h'(t)}. \tag{33}$$

Now, the equation in (32) implies that $F'(x) = F'(x_0)(I_{\mathbb{R}^n} - F'(x_0)^{\dagger}(F'(x_0) - F'(x)))$, which, using $F'(x_0)$ is surjective and $\big(I_{\mathbb{R}^n} - F'(x_0)^{\dagger}(F'(x_0) - F'(x))\big)$ is non-singular, yields $F'(x)$ is surjective for all $x \in B(x_0, t_*)$. Hence, using (31) and properties of the Moore–Penrose inverse, we have

$$(\bar{F}'(x))^{\dagger} = (F'(x_0)^{\dagger} F'(x))^{\dagger} = F'(x)^{\dagger} F'(x_0), \qquad \forall\, x \in \Omega.$$

The latter inequality and (32) imply that $\bar{F}'$ satisfies (5) with $\kappa = 0$, and that the second sequence in (28) coincides with the second sequence in (9). Moreover, using the last inequality, (31), (32) and (3), we obtain

$$\|\bar{F}'(x_0)^{\dagger}\bar{F}'(x_0)\| = \|(F'(x_0)^{\dagger}F'(x_0))^{\dagger}F'(x_0)^{\dagger}F'(x_0)\| = \|F'(x_0)^{\dagger}F'(x_0)\| \tag{34}$$

and

$$\|\bar{F}'(x_0)^{\dagger}\| = \|F'(x_0)^{\dagger}F'(x_0)\| = \|\Pi_{Ker(F'(x_0))^{\perp}}\| = 1. \tag{35}$$

Accordingly, (34) implies that $\|\bar{F}'(x_0)^{\dagger}\bar{F}'(x_0)\| > 0$, and (35) together with (26) and (31) implies that $\bar{F}'$ satisfies (7) with $f = \bar{f}$.

Therefore, with the exception (30), the results of the theorem follow from Theorem 3 with $F = \bar{F}$, $f = \bar{f}$, $h_\lambda = h$, $\lambda = 0$ and $t_\lambda^* = t_*$.

Our task is now to show that (30) holds.

Take $k \in \{1, 2, \dots\}$. Using the equation in (32), it follows by simple calculus that

$$F'(x_{k-1})^{\dagger}F'(x_0)\big(I_{\mathbb{R}^n} - F'(x_0)^{\dagger}(F'(x_0) - F'(x_{k-1}))\big) = F'(x_{k-1})^{\dagger}F'(x_{k-1}),$$

which, combined with (3), (33) and $\|x_{k-1} - x_0\| \le t_{k-1} \le t_*$, yields

$$\begin{aligned}\|F'(x_{k-1})^{\dagger}F'(x_0)\| &\le \|\Pi_{Ker(F'(x_{k-1}))^{\perp}}(I_{\mathbb{R}^n} - F'(x_0)^{\dagger}(F'(x_0) - F'(x_{k-1})))^{-1}\| \\ &\le \|(I_{\mathbb{R}^n} - F'(x_0)^{\dagger}(F'(x_0) - F'(x_{k-1})))^{-1}\| \\ &\le -(h'(t_{k-1}))^{-1}.\end{aligned}$$

Hence, using (28) and (32), we obtain

$$\|x_k - x_{k-1}\| = \|F'(x_{k-1})^{\dagger} F(x_{k-1})\| \le -(h'(t_{k-1}))^{-1} \|F'(x_0)^{\dagger} F(x_{k-1})\|. \qquad (36)$$

Since $F(x_{k-1})$ is also surjective, it follows from (4) that $F'(x_{k-1})F'(x_{k-1})^{\dagger} = I_{\mathbb{R}^m}$, which combined with Lemma 10 and (29) gives

$$\begin{aligned}\|F'(x_0)^{\dagger} F(x_k)\| &= \|F'(x_0)^{\dagger}(F(x_k) - F(x_{k-1}) - F'(x_{k-1})(x_k - x_{k-1})\| \\ &\le \|F'(x_0)^{\dagger}\| \|E_F(x_{k-1}, x_k)\| \\ &\le e_f(t_{k-1}, t_k) \frac{\|x_k - x_{k-1}\|}{(t_k - t_{k-1})} \\ &= h(t_k) \frac{\|x_k - x_{k-1}\|}{(t_k - t_{k-1})},\end{aligned}$$

where the latter equation is obtained by combining (17), (27) and (28). Given the last inequality, (36), and that $\{t_k\}$ and h' are strictly increasing, we have

$$\begin{aligned}\|F'(x_0)^{\dagger} F(x_k)\| &\le -\frac{h(t_k)}{h'(t_{k-1})} \frac{\|F'(x_0)^{\dagger} F(x_{k-1})\|}{(t_k - t_{k-1})} \\ &\le -\frac{h(t_k)}{h'(t_k)} \frac{\|F'(x_0)^{\dagger} F(x_{k-1})\|}{(t_k - t_{k-1})}.\end{aligned}$$

Therefore, the desired inequality is implied by the last inequality together with the definition of $\{t_k\}$ in (28). □

3.2. *Convergence result for Lipschitz condition*

In this section, we first present a theorem corresponding to Theorem 3, but under the Lipschitz condition instead of the general assumption (7). We also present a theorem corresponding to Theorem 13, but under the Lipschitz condition instead of assumption (26).

THEOREM 14 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set and $F : \Omega \to \mathbb{R}^m$ a continuously differentiable function. Suppose that*

$$\left\| F'(y)^{\dagger}(I_{\mathbb{R}^m} - F'(x)F'(x)^{\dagger})F(x) \right\| \le \kappa \|x - y\|, \qquad \forall\, x, y \in \Omega$$

for some $0 \le \kappa < 1$. Take $x_0 \in \Omega$ such that $\beta := \|F'(x_0)^{\dagger} F(x_0)\| > 0$, $F'(x_0) \ne 0$ and

$$rank(F'(x)) \le rank(F'(x_0)), \qquad \forall\, x \in \Omega.$$

Suppose that there exist $R > 0$ and $L > 0$, such that $B(x_0, R) \subseteq \Omega$,

$$\|F'(x_0)^{\dagger}\| \|F'(x) - F'(y)\| \le L\|x - y\|,$$

for any $x, y \in \Omega$, $\|x - x_0\| + \|y - x\| < R$. Take $\lambda = (1 - \beta L)\kappa$ and consider the auxiliary function $h_\lambda : [0, R) \to \mathbb{R}$,

$$h_\lambda(t) := \beta - (1 - \lambda)t + (Lt^2)/2.$$

If

$$\beta L \le \Delta := \frac{(1-\kappa)^2}{(\kappa^2-\kappa+1)+\sqrt{2\kappa^2-2\kappa+1}},$$

then $h_\lambda(t)$ *has a smallest zero* $t_\lambda^* = \big(1-\lambda-\sqrt{(1-\lambda)^2-2\beta L}\big)/L$, *the sequences for solving* $h_\lambda(t)=0$ *and* $F(x)=0$, *with starting points* $t_{\lambda,0}=0$ *and* x_0, *respectively,*

$$t_{\lambda,k+1}=t_{\lambda,k}-h_0'(t_{\lambda,k})^{-1}h_\lambda(t_{\lambda,k}),\quad x_{k+1}=x_k-F'(x_k)^\dagger F(x_k),\quad k=0,1,\ldots,$$

are well defined, $\{t_{\lambda,k}\}$ *is strictly increasing, is contained in* $[0,t_\lambda^*)$, *and converges to* t_λ^*, $\{x_k\}$ *is contained in* $B(x_0,t_\lambda^*)$, *converges to a point* $x_*\in B[x_0,t_\lambda^*]$ *such that* $F'(x_*)^\dagger F(x_*)=0$,

$$\|x_{k+1}-x_k\|\le t_{\lambda,k+1}-t_{\lambda,k},\qquad \|x_*-x_k\|\le t_\lambda^*-t_{\lambda,k},\qquad k=0,1,\ldots,$$

and

$$\|x_{k+1}-x_k\|\le \frac{t_{\lambda,k+1}-t_{\lambda,k}}{(t_{\lambda,k}-t_{\lambda,k-1})^2}\|x_k-x_{k-1}\|^2,\quad k=1,2,\ldots.$$

Moreover, if $\lambda=0$, *then the sequences* $\{t_{\lambda,k}\}$ *and* $\{x_k\}$ *converge Q-linearly and R-linearly (or, if* $\lambda=0$ *and* $h_0'(t_\lambda^*)<0$, *Q-quadratically and R-quadratically) to* t_λ^* *and* x_*, *respectively.*

Proof It can immediately be proved that F, x_0 and $f:[0,R)\to\mathbb{R}$ defined by $f(t)=Lt^2/2-t$, satisfy the inequality (7), and conditions **(h1)** and **(h2)**. Hence,

$$h_\lambda(t):=\beta-(1-\lambda)t+(Lt^2)/2=\beta+\lambda t+f(t).$$

Since,

$$\beta L\le\Delta=\frac{(1-\kappa)^2}{(\kappa^2-\kappa+1)+\sqrt{2\kappa^2-2\kappa+1}}=\frac{(1-\kappa)^2}{(1-\kappa)^2+\kappa+\sqrt{2\kappa^2-2\kappa+1}}\le 1, \tag{37}$$

we have $\lambda=(1-\beta L)\kappa\ge 0$ and $\lambda=-\kappa f'(\beta)$. Moreover, the first inequality in (37) implies that $(1-\lambda)^2-2\beta L\ge 0$, i.e. h_λ satisfies **(h3)** and $t_\lambda^*=\big(1-\lambda-\sqrt{(1-\lambda)^2-2\beta L}\big)/L$ is its smallest root.

Therefore, taking f, h_λ, λ and t_λ^* as defined above, all the statements of the theorem follow from Theorem 3. □

Under the Lipschitz condition, Theorem 13 becomes:

THEOREM 15 *Let* $\Omega\subseteq\mathbb{R}^n$ *be an open set and* $F:\Omega\to\mathbb{R}^m$ *a continuously differentiable function. Take* $x_0\in\Omega$ *such that* $\beta:=\|F'(x_0)^\dagger F(x_0)\|>0$ *and* $F'(x_0)$ *is surjective. Suppose that there exist* $R>0$ *and* $L>0$, *such that* $B(x_0,R)\subseteq\Omega$,

$$\|F'(x_0)^\dagger(F'(x)-F'(y))\|\le L\|x-y\|,$$

for any $x,y\in\Omega$, $\|x-x_0\|+\|y-x\|<R$. *Consider the auxiliary function* $h:[0,\ R)\to\mathbb{R}$,

$$h(t):=\beta-t+(Lt^2)/2.$$

If $\beta L\le 1/2$, *then* $h(t)$ *has a smallest zero* $t_*=\big(1-\sqrt{1-2\beta L}\big)/L$, *the sequences for solving* $h(t)=0$ *and* $F(x)=0$, *with starting points* $t_0=0$ *and* x_0, *respectively,*

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

are well defined, $\{t_k\}$ *is strictly increasing, is contained in* $[0, t_*)$*, and converges Q-linearly to* t_**,* $\{x_k\}$ *is contained in* $B(x_0, t_*)$*, and converges R-linearly to a point* $x_* \in B[x_0, t_*]$ *such that* $F'(x_*)^\dagger F(x_*) = 0$*,*

$$\|x_{k+1} - x_k\| \le t_{k+1} - t_k, \qquad \|x_* - x_k\| \le t_* - t_k, \qquad k = 0, 1, \dots,$$

$$\|x_{k+1} - x_k\| \le \frac{t_{k+1} - t_k}{(t_k - t_{k-1})^2}\|x_k - x_{k-1}\|^2, \quad k = 1, 2, \dots,$$

and

$$\|F'(x_0)^\dagger F(x_k)\| \le \left(\frac{t_{k+1} - t_k}{t_k - t_{k-1}}\right)\|F'(x_0)^\dagger F(x_{k-1})\|, \quad k = 1, 2, \dots.$$

If, additionally, $\beta L < 1/2$*, then the sequences* $\{t_k\}$ *and* $\{x_k\}$ *converge Q-quadratically and R-quadratically to* t_* *and* x_**, respectively.*

Proof The proof follows the same pattern as the proof of Theorem 14. □

3.3. *Convergence result under Smale's condition*

In this section, we first present a theorem corresponding to Theorem 3, but under Smale's α-condition (see [11,12,16]). We also present a theorem corresponding to Theorem 13, but under Smale's α-condition instead of assumption (26).

To simplify, we take $\lambda = \kappa$ in the next theorem. As seen in Remark 1, this is always a possible choice for λ.

THEOREM 16 *Let* $\Omega \subseteq \mathbb{R}^n$ *be an open set and* $F : \Omega \to \mathbb{R}^m$ *an analytic function. Suppose that*

$$\left\|F'(y)^\dagger(I_{\mathbb{R}^m} - F'(x)F'(x)^\dagger)F(x)\right\| \le \kappa\|x - y\|, \qquad \forall\, x, y \in \Omega$$

for some $0 \le \kappa < 1$*. Take* $x_0 \in \Omega$ *such that* $\beta := \|F'(x_0)^\dagger F(x_0)\| > 0$*,* $F'(x_0) \ne 0$ *and*

$$rank(F'(x)) \le rank(F'(x_0)), \qquad \forall\, x \in \Omega.$$

Suppose that

$$\gamma := \|F'(x_0)^\dagger\| \sup_{n>1}\left\|\frac{F^{(n)}(x_0)}{n!}\right\|^{1/(n-1)} < +\infty, \qquad B(x_0, 1/\gamma) \subseteq \Omega. \tag{38}$$

Consider the auxiliary function $h_\kappa : [0,\ 1/\gamma) \to \mathbb{R}$*,*

$$h_\kappa(t) := \beta - (2-\kappa)t + t/(1-\gamma t).$$

If

$$\alpha := \beta\gamma \le 3 - 2\sqrt{2},$$

then $h_\kappa(t)$ *has a smallest zero* $t_\kappa^* = \left(1-\kappa+\alpha-\sqrt{(1-\kappa+\alpha)^2 - 4(2-\kappa)\alpha}\right)/(2\gamma(2-\kappa))$*, the sequences for solving* $h_\kappa(t) = 0$ *and* $F(x) = 0$*, with starting points* $t_{\kappa,0} = 0$ *and* x_0*, respectively,*

$$t_{\kappa,k+1} = t_{\kappa,k} - h_0'(t_{\kappa,k})^{-1}h_\kappa(t_{\kappa,k}), \quad x_{k+1} = x_k - F'(x_k)^\dagger F(x_k), \quad k = 0, 1, \dots,$$

are well defined, $\{t_{\kappa,k}\}$ *is strictly increasing, is contained in* $[0, t_\kappa^*)$*, and converges to* t_κ^**,* $\{x_k\}$ *is contained in* $B(x_0, t_\kappa^*)$*, converges to a point* $x_* \in B[x_0, t_\kappa^*]$ *such that* $F'(x_*)^\dagger F(x_*) = 0$*,*

$$\|x_{k+1} - x_k\| \le t_{\kappa,k+1} - t_{\kappa,k}, \qquad \|x_* - x_k\| \le t_\kappa^* - t_{\kappa,k}, \qquad k = 0, 1, \dots ,$$

and

$$\|x_{k+1} - x_k\| \le \frac{t_{\kappa,k+1} - t_{\kappa,k}}{(t_{\kappa,k} - t_{\kappa,k-1})^2}\|x_k - x_{k-1}\|^2, \quad k = 1, 2, \dots .$$

Moreover, if $\kappa = 0$*, then the sequences* $\{t_{\kappa,k}\}$ *and* $\{x_k\}$ *converge Q-linearly and R-linearly (or, if* $\kappa = 0$ *and* $h_0'(t_\kappa^*) < 0$*, Q-quadratically and R-quadratically) to* t_κ^* *and* x_**, respectively.*

We need the following results to prove the above theorem.

LEMMA 17 *Let* $\Omega \subseteq \mathbb{R}^n$ *be an open set and* $F : \Omega \to \mathbb{R}^m$ *an analytic function. Suppose that* $x_0 \in \mathbb{R}^n$ *and* γ *is defined in (39). Then, for all* $x \in B(x_0, 1/\gamma)$ *it holds that*

$$\|F'(x_0)^\dagger\|\|F''(x)\| \leqslant (2\gamma)/(1 - \gamma\|x - x_0\|)^3.$$

Proof The proof follows the same pattern as the proof of Lemma 21 of [7]. □

LEMMA 18 *Let* $\Omega \subseteq \mathbb{R}^n$ *be an open set and* $F : \Omega \to \mathbb{R}^m$ *be twice continuously differentiable on* Ω*. If there exists* $f : [0, R) \to \mathbb{R}$ *that is twice continuously differentiable with derivative* f' *convex and satisfies*

$$\|F'(x_0)^\dagger\|\|F''(x)\| \leqslant f''(\|x - x_0\|),$$

for all $x \in \Omega$ *such that* $\|x - x_0\| < R$*, then* F *and* f *satisfy (7).*

Proof The proof follows the same pattern as the proof of Lemma 22 of [7]. □

Proof of Theorem 16 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

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

for $n \ge 2$. It follows from the latter equalities that f satisfies **(h1)** and **(h2)**. Moreover, as $f''(t) = (2\gamma)/(1 - \gamma t)^3$, combining Lemmas 17 and 18, we have F and f satisfy (7) with $R = 1/\gamma$. Hence,

$$h_\kappa(t) := \beta - (2 - \kappa)t + t/(1 - \gamma t) = \beta + \lambda t + f(t).$$

Since $\lambda = \kappa$, we have $0 \le \lambda < 1$ and $\lambda = -\kappa f'(0) \ge -\kappa f'(\beta)$, where the latter inequality follows from **(h2)**. Moreover, $\alpha = \beta\gamma \le 3 - 2\sqrt{2}$ implies that $\big((1-\kappa+\alpha)^2 - 4(2-\kappa)\alpha\big) \ge 0$, i.e. h_κ satisfies **(h3)** and $t_\kappa^* = \big(1-\kappa+\alpha-\sqrt{(1-\kappa+\alpha)^2 - 4(2-\kappa)\alpha}\big)/(2\gamma(2-\kappa))$ is its smallest root.

Therefore, taking f, $\lambda = \kappa$, $h_\lambda = h_\kappa$ and $t_\lambda^* = t_\kappa^*$ as defined above, all the statements of the theorem follow from Theorem 3. □

Under Smale's α-condition, Theorem 13 becomes:

THEOREM 19 *Let $\Omega \subseteq \mathbb{R}^n$ be an open set and $F : \Omega \to \mathbb{R}^m$ an analytic function. Take $x_0 \in \Omega$ such that $\beta := \|F'(x_0)^\dagger F(x_0)\| > 0$ and $F'(x_0)$ is surjective. Suppose that*

$$\gamma := \sup_{n>1} \left\| \frac{F'(x_0)^\dagger F^{(n)}(x_0)}{n!} \right\|^{1/(n-1)} < +\infty, \qquad B(x_0, 1/\gamma) \subseteq \Omega. \tag{39}$$

Consider the auxiliary function $h : [0,\ 1/\gamma) \to \mathbb{R}$,

$$h(t) := \beta - 2t + t/(1 - \gamma t).$$

If

$$\alpha := \beta\gamma \leq 3 - 2\sqrt{2},$$

then $h(t)$ has a smallest zero $t_* = \left(1 + \alpha - \sqrt{(1+\alpha)^2 - 8\alpha}\right)/(4\gamma)$, the sequences for solving $h(t) = 0$ and $F(x) = 0$, with starting points $t_0 = 0$ and x_0, respectively,

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

are well defined, $\{t_k\}$ is strictly increasing, is contained in $[0, t_*)$, and converges Q-linearly to t_*, $\{x_k\}$ is contained in $B(x_0, t_*)$ and converges R-linearly to a point $x_* \in B[x_0, t_*]$ such that $F'(x_*)^\dagger F(x_*) = 0$,

$$\|x_{k+1} - x_k\| \leq t_{k+1} - t_k, \qquad \|x_* - x_k\| \leq t_* - t_k, \qquad k = 0, 1, \dots,$$

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

and

$$\|F'(x_0)^\dagger F(x_k)\| \leq \left(\frac{t_{k+1} - t_k}{t_k - t_{k-1}} \right) \|F'(x_0)^\dagger F(x_{k-1})\|, \quad k = 1, 2, \dots.$$

If, additionally, $\alpha := \beta\gamma < 3 - 2\sqrt{2}$, then the sequences $\{t_k\}$ and $\{x_k\}$ converge Q-quadratically and R-quadratically to t_* and x_*, respectively.

Proof The proof follows the same pattern as the proof of Theorem 16. □

4. Final remarks

We presented a new semi-local convergence analysis of the Gauss–Newton method for solving 1, where F satisfies 2, under a majorant condition. It would also be interesting to present a local convergence analysis of the Gauss–Newton method, under a majorant condition, for the problem under consideration. As a consequence, we would get convergence results for analytical functions under an γ-condition. This local analysis will be performed in the future.

Acknowledgements

M.L.N. Gonçalves was supported in part by CAPES-Brazil. P.R. Oliveira author was supported in part by CNPq.

References

[1] Ferreira OP. Local convergence of Newton's method in Banach space from the viewpoint of the majorant principle. IMA J. Numer. Anal. 2009;29:746–759.

[2] Ferreira OP. Local convergence of Newton's method under majorant condition. J. Comput. Appl. Math. 2011;235:1515–1522.

[3] Ferreira OP, Gonçalves MLN. Local convergence analysis of inexact Newton-like methods under majorant condition. Comput. Optim. Appl. 2011;48:1–21.

[4] Ferreira OP, Svaiter BF. Kantorovich's majorants principle for Newton's method. Comput. Optim. Appl. 2009;42:213–229.

[5] Ferreira O, Gonçalves M, Oliveira P. Local convergence analysis of inexact Gauss–Newton like methods under majorant condition. J. Comput. Appl. Math. 2012;236:2487–2498.

[6] Ferreira OP, Gonçalves MLN, Oliveira PR. Convergence of the Gauss–Newton method for convex composite optimization under a majorant condition. prepint. Available from: http://arxiv.org/abs/1107.3796

[7] Ferreira OP, Gonçalves MLN, Oliveira PR. Local convergence analysis of the Gauss–Newton method under a majorant condition. J. Complexity. 2011;27:111–125.

[8] Häussler WM. A Kantorovich-type convergence analysis for the Gauss–Newton-method. Numer. Math. 1986;48:119–125.

[9] Hu N, Shen W, Li C. Kantorovich's type theorems for systems of equations with constant rank derivatives. J. Comput. Appl. Math. 2008;219:110–122.

[10] Li C, Hu N, Wang J. Convergence behavior of Gauss–Newton's method and extensions of the Smale point estimate theory. J. Complexity. 2010;26:268–295.

[11] Dedieu J-P, Kim M-H. Newton's method for analytic systems of equations with constant rank derivatives. J. Complexity. 2002;18:187–209.

[12] Dedieu JP, Shub M. Newton's method for overdetermined systems of equations. Math. Comp. 2000;69:1099–1115.

[13] Wang X. Convergence of Newton's method and uniqueness of the solution of equations in Banach space. IMA J. Numer. Anal. 2000;20:123–134.

[14] Xu X, Li C. Convergence of Newton's method for systems of equations with constant rank derivatives. J. Comput. Math. 2007;25:705–718.

[15] Xu X, Li C. Convergence criterion of Newton's method for singular systems with constant rank derivatives. J. Math. Anal. Appl. 2008;345:689–701.

[16] Smale S. 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). New York (NY): Springer; 1986. p. 185–196.

[17] Lawson CL, Hanson RJ. Solving least squares problems. Prentice-Hall Series in Automatic Computation. Englewood Cliffs (NJ): Prentice-Hall; 1974.

[18] Ben-Israel A, Greville TNE. Generalized inverses. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 15. 2nd ed, Theory and Applications. New York (NY): Springer-Verlag; 2003.