Convergence behavior of Gauss-Newton’s method and extensions of the Smale point estimate theory∗

Chong Li†, Nuchun Hu‡ and Jin-Hua Wang§

Abstract: The notions of Lipschitz conditions with $L$ average are introduced to the study of convergence analysis of Gauss-Newton’s method for singular systems of equations. Unified convergence criteria ensuring the convergence of Gauss-Newton’s method for one kind of singular systems of equations with constant rank derivatives are established and unified estimates of radii of convergence balls are also obtained. Applications to some special cases such as the Kantorovich type conditions, $\gamma$-conditions and the Smale point estimate theory are provided and some important known results are extended and/or improved.

Keyword: Gauss-Newton’s Method; Majorizing sequence; Convergence criterion; Convergence balls.

1 Introduction

Let $F : \Omega \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^l$ be a nonlinear operator with its Fréchet derivative denoted by $DF$. Finding solutions of the nonlinear operator equation

$$F(x) = 0$$

(1.1)

is a very general subject and has been studied extensively in both theoretical and applied areas of mathematics. In the case when $m = l$ and $DF(x)$ is invertible for each $x \in \Omega$, the most famous method to find an approximative solution is Newton’s method, which is defined by

$$x_{n+1} = x_n - DF(x_n)^{-1}F(x_n) \quad \text{for each } n = 0, 1, 2, \ldots,$$

(1.2)

where $x_0 \in \Omega$ is an initial point. Usually, the study about convergence issue of Newton’s method is mainly centered on two types: local and semi-local convergence analysis. The local convergence issue is, based on information around a solution, to seek estimates of the radii of convergence balls; while the semi-local one is, based on information around an initial point, to give criteria ensuring the convergence of Newton’s method.

Regarding the semi-local convergence of Newton’s method, one of the most important results is the well-known Kantorovich theorem (cf. [19]) which provides a simple and clear convergence criterion for operators with bounded second derivatives $D^2F$ (or the Lipschitz continuous first derivatives). Another

∗This project was supported in part by the National Natural Science Foundations of China (Grant No. 10671175; 10731069).
†Department of Mathematics, Zhejiang University, Hangzhou 310027, P R China (cli@zju.edu.cn)
‡Department of Mathematics, Zhejiang University, Hangzhou 310027, P R China (nuchun@zjnu.cn)
§Corresponding author, Department of Mathematics, Zhejiang University of Technology, Hangzhou 310032, P R China (wjh@zjut.edu.cn)
important one is the Smale $\alpha$-theory in [25], where the concepts of approximate zeroes were proposed and criteria to judge an initial point being an approximate zero were established for analytic operators, only depending on the information at the initial point; while, the best criterion was subsequently founded in [35] by Wang and Han.

Regarding the local convergence of Newton’s method, Traub and Wozniakowski in [28] and Wang in [32] independently gave the best estimate of the radii of convergence balls when the first derivatives are Lipschitz continuous around a solution. Another important one is due to the Smale $\gamma$-theory in [25], where the estimate of radii of convergence balls was given for analytic operators, only depending on the information at the solution.

Besides, there are a lot of works on the weakness and/or extension of the hypothesis made on the underlying operators; see for example, [1, 9, 10, 12, 13, 17, 33, 34, 37, 38] and references therein. In particular, Wang introduced in [33] and [34] the notions of Lipschitz conditions with $L$ average, where $L$ is a positive nondecreasing function on $[0, R)$ satisfying

$$\int_0^R L(u)du \geq 1.$$  

(1.3)

The center Lipschitz condition with $L$ average in the inscribe sphere makes us unify convergence criteria containing the Kantorovich theorem and the Smale $\alpha$-theory; while the radius Lipschitz conditions with $L$ average unify the estimates of the radii of convergence balls for operators with Lipschitz continuous first derivatives and analytic operators; see e.g. [33, 34].

Recent attentions are focused on the study of finding zeros of singular nonlinear systems by Gauss-Newton’s method, which is defined as follows. For a given initial point $x_0 \in \Omega$, define

$$x_{n+1} = x_n - DF(x_n)^\dagger F(x_n)$$  

for each $n = 0, 1, 2, \ldots$,  

(1.4)

where $DF(x_n)^\dagger$ denotes the Moore-Penrose inverse of the linear operator (or matrix) $DF(x_n)$. For example, Shub and Smale extended in [24] the Smale point estimate theory (i.e., $\alpha$-theory and $\gamma$-theory) to Gauss-Newton’s methods for underdetermined analytic systems with surjective derivatives. By introducing the following invariants for underdetermined analytic systems

$$\gamma(F, x) := \sup_{k \geq 2} \left\| \frac{D^k F(x)}{k!} \right\|^{\frac{1}{k}},$$

they proved in [24] that if, $DF(x_0)$ is surjective and

$$\alpha(F, x_0) := \gamma(F, x_0) ||D F(x_0)^\dagger F(x_0)|| < \alpha_0 := 0.130716944 \cdots,$$  

(1.5)

or 0 is a regular value of $F$ (i.e., $DF(x)$ is surjective for each $x \in F^{-1}(0)$) and

$$\gamma d(x_0, F^{-1}(0)) < 0.069778332 \cdots,$$  

(1.6)

then Gauss-Newton’s method (1.4) with initial point $x_0$ converges to a zero $x^*$ of $F$ and satisfies

$$\|x_{n+1} - x_n\| \leq \left( \frac{1}{2} \right)^{2^n - 1} \|x_1 - x_0\|$$  

for each $n = 0, 1, \ldots$,  

(1.7)

where $\gamma = \max_{x \in F^{-1}(0)} \gamma(F, x)$. This result was improved recently by He et al [16] in such a way that the criteria (1.5) and (1.6) are respectively replaced with

$$\alpha(F, x_0) \leq \frac{13 - 3\sqrt{17}}{4} = 0.157671 \cdots.$$  

(1.8)
and
\[ \gamma d(x_0, F^{-1}(0)) < u_0 = 0.0776121 \cdots. \] (1.9)

For overdetermined systems, Dedieu and Shub studied in [6] the local (linear) convergence properties of Gauss-Newton’s method for analytic systems with injective derivatives and provided estimates of the radii of convergence balls for Gauss-Newton’s method, which has been extended by Li et al [22] to the overdetermined systems with injective derivatives satisfying the Lipschitz condition with \( L \) average mentioned above. However, for the general singular systems with constant rank derivatives, convergence analysis of Gauss-Newton’s method becomes much more complicated and difficult; see for example [5, 40, 41], where local and semi-local convergence properties of Gauss-Newton’s method for systems of equations with constant rank derivatives are explored.

Our interests in the present paper are centered on one kind of special singular systems with constant rank derivatives, that is, the systems with their derivatives satisfy
\[
\|DF(y)^\dagger(I - DF(x)DF(x)^\dagger)F(x)\| \leq \kappa \|x - y\| \quad \text{for each } x, y \in \Omega,
\] (1.10)
where \( 0 \leq \kappa < 1 \). This kind of systems was studied in [7, 8, 15], and in particular, Häußler established in [15] the Kantorovich type convergence criterion. Recently, Hu et al provided in [18] a refinement of the study for this kind of systems, and as consequences, an improved convergence criterion and an estimate of the radii of convergence balls of Gauss-Newton’s method are obtained. This kind of singular systems clearly contains underdetermined systems with surjective derivatives as special cases. However, the Smale point estimate theory has not been found to be explored for this kind systems.

In the present paper, we introduce the notions of the Lipschitz conditions with \( L \) average mentioned above (but without assumption of (1.3)) to the study of convergence analysis of Gauss-Newton’s method for the singular systems satisfying (1.10). Unified convergence criteria, which include the Kantorovich type and the Smale type convergence criteria as special cases, are established in section 3. In particular, as an application to the underdetermined systems with surjective derivatives, Corollary 3.2 extends the corresponding result in [33, Theorem 3.1] even for the nonsingular systems, as shown by Example 3.1. Moreover, unified estimates for the radii of convergence balls of Gauss-Newton’s method are presented in section 4. Applications to the cases of the Kantorovich type condition, the \( \gamma \)-condition and the Smale point estimate theory as well as some more general analytic systems are provided in section 5. As detailed in section 5, when the results are applied to the underdetermined systems with surjective derivatives, some known results such as [16, Theorems 3.3 and 4.2, Corollaries 5.3 and 5.5], [24, Theorems 1.4 and 1.7] and [29, Theorem 2.3] are extended and/or improved accordingly. In particular, Criterion (1.9) is improved to the following one:
\[ \gamma d(x_0, F^{-1}(0)) < t_0 = 0.0858167 \cdots. \] (1.11)

2 Preliminaries

Let \( \mathbb{R} := \mathbb{R} \cup \{+\infty\} \) and \( \mathbb{R}_+ := [0, +\infty] \). Throughout the whole paper, we assume that \( L \) is a positive nondecreasing function on \( [0, R] \), where \( R \in \mathbb{R}_+ \). Let \( \beta > 0 \) and \( 0 \leq \lambda < 1 \). The majorizing function \( h_\lambda : [0, R] \to \mathbb{R} \) corresponding to \( (\lambda, L) \) is defined by
\[
h_\lambda(t) = \beta - (1 - \lambda)t + \int_0^t L(u)(t - u)du \quad \text{for each } 0 \leq t \leq R.
\] (2.1)
4 CONVERGENCE BEHAVIOR OF GAUSS-NEWTON’S METHOD

Note that in the case when $\lambda = 0$, (2.1) reduces to the one which is employed by Wang in [33]. Obviously,

$$h^\prime_\lambda(t) = -(1 - \lambda) + \int_0^t L(u)du \quad \text{for each } 0 \leq t < R$$  \hspace{1cm} (2.2)

and

$$h^\prime\prime_\lambda(t) = L(t) \quad \text{for a.e. } 0 \leq t < R.$$  \hspace{1cm} (2.3)

Define

$$r_\lambda := \sup \{ r \in (0, R) : \int_0^r L(u)du \leq 1 - \lambda \}$$  \hspace{1cm} (2.4)

and

$$b_\lambda := (1 - \lambda)r_\lambda - \int_0^{r_\lambda} L(u)(r_\lambda - u)du.$$  \hspace{1cm} (2.5)

Write $\Delta = \int_0^R L(u)du$. Then

$$r_\lambda = \begin{cases} R & \text{if } \Delta < 1 - \lambda, \\ t_\lambda & \text{if } \Delta \geq 1 - \lambda, \end{cases}$$  \hspace{1cm} (2.6)

where $t_\lambda \in [0, R]$ is such that $\int_0^{t_\lambda} L(u)du = 1 - \lambda$ (noting that such a point $t_\lambda \in [0, R]$ exists if $\Delta \geq 1 - \lambda$). Furthermore, it follows that

$$\begin{cases} b_\lambda \geq \int_0^{r_\lambda} L(u)udu & \text{if } \Delta < 1 - \lambda, \\ b_\lambda = \int_0^{r_\lambda} L(u)udu & \text{if } \Delta \geq 1 - \lambda. \end{cases}$$  \hspace{1cm} (2.7)

Let $\{t_{\lambda,n}\}$ denote the sequence generated by

$$t_{\lambda,0} = 0, \quad t_{\lambda,n+1} = t_{\lambda,n} - \frac{h_\lambda(t_{\lambda,n})}{h_0(t_{\lambda,n})} = \frac{h_\lambda(t_{\lambda,n})}{h_0'(t_{\lambda,n})}$$  \hspace{1cm} (2.8)

for each $n = 0, 1, \ldots$. In particular, in the case when $\lambda = 0$, the sequence $\{t_{\lambda,n}\}$ reduces to Newton’s sequence and is denoted simply by $\{t_n\}$. The following lemma describes some properties about the function $h_\lambda$ and the convergence property of the sequence $\{t_{\lambda,n}\}$, which are crucial for convergence analysis of Gauss-Newton’s method.

**Lemma 2.1.** Suppose that $\beta \leq b_\lambda$. Then the following assertions hold.

(i) The function $h_\lambda$ is strictly decreasing on $[0, r_\lambda]$ and has exact one zero $t^*_\lambda$ in $[0, r_\lambda]$ satisfying $\beta < t^*_\lambda$.

(ii) The sequence $\{t_{\lambda,n}\}$ defined by (2.8) is strictly increasing and converges to $t^*_\lambda$.

**Proof.** The assertion (i) follows directly from (2.2) and the definitions of $r_\lambda$ and $b_\lambda$. Below we prove the assertion (ii). To do this, note that $0 = t_{\lambda,0} < t_{\lambda,1} = \beta < t^*_\lambda$. Now assume that

$$t_{\lambda,n-1} < t_{\lambda,n} < t^*_\lambda$$  \hspace{1cm} (2.9)

for some $n \in \mathbb{N}$. By (2.3), $-h_0'$ is strictly decreasing on $[0, R]$. Hence

$$-h_0'(t_{\lambda,n}) > -h_0'(t^*_\lambda) \geq -h_0'(r_\lambda) = -h_0'(r_\lambda) + \lambda \geq 0,$$

where the first inequality holds because of (2.9) and the last one is because $-h_0'(r_\lambda) \geq 0$ by the definition of $r_\lambda$. Moreover, $h_\lambda(t_{\lambda,n}) > 0$ by the assertion (i). It follows that

$$t_{\lambda,n+1} = t_{\lambda,n} - \frac{h_\lambda(t_{\lambda,n})}{h_0'(t_{\lambda,n})} > t_{\lambda,n}.$$  \hspace{1cm} (2.10)
Define the function \( N_\lambda \) on \([0, t_\lambda^*]\) by
\[
N_\lambda(t) := t - \frac{h_\lambda(t)}{h_\lambda'(t)} \quad \text{for each } t \in [0, t_\lambda^*].
\] (2.11)

Note that
\[
h_\lambda'(t) < 0 \quad \text{for each } t \in [0, t_\lambda^*],
\] (2.12)
unless \( \lambda = 0 \) and \( t = t_\lambda^* = r_\lambda \), for which we adopt the convention that \( \frac{h_\lambda(t_\lambda^*)}{h_\lambda'(t_\lambda^*)} = \lim_{t \to t_\lambda^*} \frac{h_\lambda(t)}{h_\lambda'(t)} \), which is equal to 0 by L'Hospital's rule. Hence, the function \( N_\lambda \) is well-defined and continuous on \([0, t_\lambda^*]\). Moreover, thanks to (2.3), (2.12) and the assertion (i), we have that
\[
N_\lambda'(t) := 1 - \frac{h_\lambda'(t)h_\lambda''(t) - h_\lambda(t)h_\lambda'''(t)}{(h_\lambda'(t))^2} = \frac{-\lambda h_\lambda''(t) + h_\lambda(t)L(t)}{(h_\lambda'(t))^2} > 0 \quad \text{for a.e. } t \in [0, t_\lambda^*).
\]
This together with (2.9) and (2.10) implies that
\[
t_{\lambda,n} < t_{\lambda,n+1} = N_\lambda(t_{\lambda,n}) < N_\lambda(t_\lambda^*) = t_\lambda^*.
\] (2.13)
Therefore, by mathematical induction, (2.9) holds for all \( n \in \mathbb{N} \). Consequently, \( \{t_{\lambda,n}\} \) is increasing, bounded and so converges to a point \( t^* \in (0, t_\lambda^*], \) which is clearly a zero of \( h_\lambda \) in \([0, t_\lambda^*]\). Hence \( t^* = t_\lambda^* \) and the proof is complete.

We conclude this section with some properties related to Moore-Penrose inverse, which are known in any textbooks on matrix analysis; see for example [3, 27, 30]. Let \( A : \mathbb{R}^m \to \mathbb{R}^l \) be a linear operator (or an \( l \times m \) matrix). Recall that an operator (or an \( m \times l \) matrix) \( A^\dagger : \mathbb{R}^l \to \mathbb{R}^m \) is the Moore-Penrose inverse of \( A \) if it satisfies the following four equations
\[
AA^\dagger A = A, \quad A^\dagger AA^\dagger = A^\dagger, \quad (AA^\dagger)^* = AA^\dagger, \quad (A^\dagger A)^* = A^\dagger A,
\]
where \( A^\dagger \) denotes the adjoint of \( A \). Let \( \ker A \) and \( \im A \) denote the kernel and image of \( A \), respectively. For a subspace \( E \) of \( \mathbb{R}^m \), we use \( \Pi_E \) to denote the projection onto \( E \). Then it is clear that
\[
A^\dagger A = \Pi_{(\ker A)^\perp} \quad \text{and} \quad AA^\dagger = \Pi_{\im A}.
\] (2.14)
In particular, in the case when \( A \) is surjective, or equivalently, when \( A \) is full row rank, \( AA^\dagger = I_l \). The following proposition gives a perturbation bound of the Moore-Penrose inverse, which will be useful.

**Proposition 2.1.** Let \( A \) and \( B \) be \( l \times m \) matrices. Assume that
\[
1 \leq \rank(A) \leq \rank(B) \quad \text{and} \quad \|B^\dagger\|\|A - B\| < 1.
\]
Then
\[
\rank(A) = \rank(B) \quad \text{and} \quad \|A^\dagger\| \leq \frac{\|B^\dagger\|}{1 - \|B^\dagger\|\|A - B\|}.
\]

### 3 Convergence criterion of Gauss-Newton’s method

Let \( B(x, r) \) and \( \overline{B}(x, r) \) stand respectively for the open and closed ball in \( \mathbb{R}^m \) with center \( x \) and radius \( r > 0 \). Let \( F : \Omega \subseteq \mathbb{R}^m \to \mathbb{R}^l \) be a nonlinear operator with continuous Fréchet derivative which is denoted by \( DF \). For the remainder of the whole paper we always assume that
\[
\|DF(y)^\dagger(I - DF(x)DF(x)^\dagger)F(x)\| \leq \kappa\|x - y\| \quad \text{for any } x, y \in \Omega
\] (3.1)
with $0 \leq \kappa < 1$. Let $L$ be a positive nondecreasing function on $[0, R)$ as in the previous section. The notion of the $L$-average Lipschitz condition was introduced by Li and Ng in [20], which is a modification of the one that was first introduced by Wang in [33] where the terminology of “the center Lipschitz condition in the inscribed sphere with $L$ average” was used. Throughout this section, let $x_0 \in \Omega$ be such that $DF(x_0) \neq 0$ and

$$
\text{rank}(DF(x)) \leq \text{rank}(DF(x_0)) \quad \text{for each } x \in \Omega.
$$

**Lemma 3.1.** Let $r > 0$ be such that $B(x_0, r) \subseteq \Omega$. Then $DF$ is said to satisfy the $L$-average Lipschitz condition on $B(x_0, r)$ if, for any $x, x' \in B(x_0, r)$ with $\|x - x_0\| + \|x' - x\| < r$,

$$
\|DF(x_0)F(DF'(x') - DF(x))\| \leq \int_{\|x - x_0\|}^{\|x - x_0\| + \|x' - x\|} L(u)du.
$$

**Definition 3.2.** Let $r > 0$ be such that $B(x_0, r) \subseteq \Omega$. Then $DF$ is said to satisfy the modified $L$-average Lipschitz condition on $B(x_0, r)$ if, for any $x, x' \in B(x_0, r)$ with $\|x - x_0\| + \|x' - x\| < r$,

$$
\|DF(x_0)\| \|DF(x') - DF(x)\| \leq \int_{\|x - x_0\|}^{\|x - x_0\| + \|x' - x\|} L(u)du.
$$

Before verifying the main theorem, we need two lemmas. The first one is known in [33, P.170]; while the second one is a consequence of Proposition 2.1. Recall that $r_0$ is defined by (2.4) for $\lambda = 0$ and recall from (2.2) that

$$
h_0(t) = -1 + \int_0^t L(u)du \quad \text{for each } 0 \leq t < R.
$$

**Lemma 3.1.** Let $0 \leq c < R$. Define

$$
\chi(t) = \frac{1}{t} \int_0^t L(c + u)(t - u)du \quad \text{for each } 0 \leq t < R - c.
$$

Then $\chi$ is increasing on $[0, R - c)$.

**Lemma 3.2.** Suppose that $0 < r \leq r_0$ satisfies $B(x_0, r) \subseteq \Omega$ and that $DF$ satisfies the modified $L$-average Lipschitz condition on $B(x_0, r)$. Then, for each $x \in B(x_0, r)$, $\text{rank}(DF(x)) = \text{rank}(DF(x_0))$ and

$$
\|DF(x)\| \leq h'_0(\|x - x_0\|)^{-1}\|DF(x_0)\|.
$$

Let

$$
\beta := \|DF(x_0)F(DF(x_0))\| \quad \text{and} \quad \lambda_0 = \kappa \left(1 - \int_0^\beta L(u)du\right).
$$

Recall that $b_\lambda$ is given by (2.5) and $t^*_\lambda$ is the unique zero of the function $h_\lambda$ in $[0, r_\lambda]$. Recall also that $\{t_{\lambda, n}\}$ is the sequence generated by (2.8).

**Theorem 3.1.** Let $\lambda \geq \lambda_0$. Suppose that

$$
\beta \leq b_\lambda \quad \text{and} \quad B(x_0, t^*_\lambda) \subseteq \Omega
$$

(3.8)
and that $DF$ satisfies the modified L-average Lipschitz condition on $B(x_0,t_\lambda^*)$. Let $\{x_n\}$ be the sequence generated by Gauss-Newton’s method (1.4) with initial point $x_0$. Then $\{x_n\}$ converges to a zero $x^*$ of $DF(\cdot)^\dagger F(\cdot)$ in $B(x_0,t_\lambda^*)$ and for each $n \geq 0$, the following estimates hold:

$$\|x_n - x^*\| \leq t_\lambda^* - t_{\lambda,n}$$

(3.9)

and

$$\|x_{n+1} - x_n\| \leq t_{\lambda,n+1} - t_{\lambda,n}.$$  

(3.10)

Proof. We first use mathematical induction to prove that

$$\|x_n - x_{n-1}\| \leq t_{\lambda,n} - t_{\lambda,n-1}$$

(3.11)

holds for each $n = 1, 2, \ldots$. Note first that (3.11) holds for $n = 1$ because

$$\|x_1 - x_0\| = \|DF(x_0)^\dagger F(x_0)\| = \beta = t_{\lambda,1} - t_{\lambda,0}.$$  

Assume now that (3.11) holds for all $n \leq k$. Write for each $s \in [0, 1]$,

$$x_k^s = x_{k-1} + s(x_k - x_{k-1}) \quad \text{and} \quad t_{\lambda,k}^s = t_{\lambda,k-1} + s(t_{\lambda,k} - t_{\lambda,k-1}).$$

Then, for each $s \in [0, 1]$,

$$\|x_k^s - x_0\| \leq \|x_k^s - x_{k-1}\| + \sum_{i=1}^{k-1} \|x_i - x_{i-1}\| \leq t_{\lambda,k}^s < t_\lambda^* \leq r_\lambda \leq r_0.$$  

(3.12)

In particular,

$$\|x_{k-1} - x_0\| \leq t_{\lambda,k-1} \quad \text{and} \quad \|x_k - x_0\| \leq t_{\lambda,k}.$$  

(3.13)

Hence, Lemma 3.2 is applicable and

$$\|DF(x_k)^\dagger\| \leq -h'_0(\|x_k - x_0\|)^{-1}\|DF(x_0)^\dagger\| \leq -h'_0(t_{\lambda,k})^{-1}\|DF(x_0)^\dagger\|,$$  

(3.14)

where the last inequality holds because $-h'_0(t)^{-1}$ is increasing monotonically. Furthermore, by inductional assumption, $\|x_k - x_{k-1}\| \leq t_{\lambda,k} - t_{\lambda,k-1}$. It follows from Lemma 3.1 and (2.3) that

$$\int_0^1 \frac{\|x_{k-1} - x_0\| + s\|x_k - x_{k-1}\|}{\|x_k - x_{k-1}\|} L(u)du \|x_k - x_{k-1}\| \leq t_{\lambda,k} - t_{\lambda,k-1}.$$  

(3.15)

Note that $h'_\lambda = h'_0 + \lambda$ and $-h'_\lambda(t_{\lambda,k-1}) - h'_0(t_{\lambda,k-1})(t_{\lambda,k} - t_{\lambda,k-1}) = 0$. Therefore

$$\int_0^1 \frac{\|x_{k-1} - x_0\| + s\|x_k - x_{k-1}\|}{\|x_k - x_{k-1}\|} L(u)du \|x_k - x_{k-1}\| \leq h_\lambda(t_{\lambda,k}) - \lambda(t_{\lambda,k} - t_{\lambda,k-1}).$$  

(3.16)

On the other hand, since $x_k - x_{k-1} = -DF(x_{k-1})^\dagger F(x_{k-1})$, it follows that

$$DF(x_k)^\dagger F(x_k) = \int_0^1 DF(x_k)^\dagger (DF(x_{k-1} + s(x_k - x_{k-1})) - DF(x_{k-1}))(x_k - x_{k-1})ds + DF(x_k)^\dagger (I - DF(x_{k-1})DF(x_{k-1})^\dagger)(x_k - x_{k-1}).$$  

(3.17)
Hence, by (3.1) and (3.14) together with the modified $L$-average Lipschitz condition (3.4), one has that
\[
\|x_{k+1} - x_k\| = \|DF(x_k)^\dagger F(x_k)\| \\
\leq \int_0^1 \|DF(x_k)^\dagger [DF(x_k) + s(x_k - x_{k-1}) - DF(x_{k-1})]\| x_k - x_{k-1}\| ds \\
+ \kappa \|x_k - x_{k-1}\| \\
\leq -h_0'(t_{\lambda,k})^{-1} \int_0^1 \|x_k - x_{k-1} + s(x_k - x_{k-1})\| L(u)du \|x_k - x_{k-1}\| ds + \kappa \|x_k - x_{k-1}\|.
\] (3.18)

Combining this with (3.16), one gets that
\[
\|x_{k+1} - x_k\| \leq -h_0'(t_{\lambda,k})^{-1} h_\lambda(t_{\lambda,k}) + (\kappa + \lambda h_0'(t_{\lambda,k})^{-1})(t_{\lambda,k} - t_{\lambda,k-1}).
\] (3.19)

Noting $\beta = t_{\lambda,1} \leq t_{\lambda,k}$ and in view of the definition of $\lambda_0$ in (3.7), we have that
\[
\lambda \geq \kappa \left(1 - \int_0^\beta L(u)du\right) \geq -h_0'(t_{\lambda,k})\kappa,
\]
that is, $\kappa + \lambda h_0'(t_{\lambda,k})^{-1} \leq 0$. Consequently, (3.19) entails
\[
\|x_{k+1} - x_k\| \leq -h_0'(t_{\lambda,k})^{-1} h_\lambda(t_{\lambda,k}) = t_{\lambda,k+1} - t_{\lambda,k}.
\]

This shows that (3.11) holds for $n = k + 1$ and so for each $n = 1, 2, \ldots$. Thus Lemma 2.1 is applicable to concluding that \(\{x_n\}\) converges to some point $x^* \in \overline{B(x_0, t_{\lambda,k}^*)}$ and (3.9) holds. Since
\[
\|DF(x^*)^\dagger F(x_n)\| \leq \|DF(x^*)^\dagger (I_{R^D} - DF(x_n)DF(x_n)^\dagger) F(x_n)\| \\
+ \|DF(x^*)^\dagger - DF(x_n)DF(x_n)^\dagger\| \|DF(x_n)\| \|x_{n+1} - x_n\|,
\]
one sees that $x^*$ is a zero of $DF(\cdot)^\dagger F(\cdot)$ and the proof is complete.

Clearly, the criterion (3.8) in Theorem 3.1 depends upon the choice of $\lambda$ and the smaller $\lambda$ is, the better the criterion is. Therefore the best choice of $\lambda$ is $\lambda \left(1 - \int_0^\beta L(u)du\right)$; but in this case, the criterion (3.8) would be implicit. Corollary 3.1 below provides a simple choice of $\lambda$ (i.e., take $\lambda = \kappa$) such that the criterion (3.8) is explicit.

**Corollary 3.1.** Suppose that
\[
\beta \leq b_\kappa \quad \text{and} \quad \overline{B(x_0, t_{\lambda,k}^*)} \subset \Omega
\] (3.20)
and that $DF$ satisfies the modified $L$-average Lipschitz condition on $\overline{B(x_0, t_{\lambda,k}^*)}$. Let \(\{x_n\}\) be the sequence generated by Gauss-Newton’s method (1.4) with initial point $x_0$. Then \(\{x_n\}\) converges to a zero $x^*$ of $DF(\cdot)^\dagger F(\cdot)$ in $\overline{B(x_0, t_{\lambda,k}^*)}$ and the estimates (3.9) and (3.10) hold with $\lambda = \kappa$.

Recall that $b_0, t_n, t_{\lambda,n}$ are defined in (2.5), (2.8) and in Lemma 2.1 for $\lambda = 0$. We have the following improved version of Theorem 3.1 for $\kappa = 0$.

**Theorem 3.2.** Suppose that
\[
\|DF(y)^\dagger (I_{R^D} - DF(x)DF(x)^\dagger) F(x)\| = 0 \quad \text{for any } x, y \in \Omega
\] (3.21)
Take \( \tilde{\Omega} := \{ x_0, t_0^* \} \subseteq \Omega \) and that \( DF \) satisfies the modified \( L \)-average Lipschitz condition on \( B(x_0, t_0^*) \). Let \( \{ x_n \} \) be the sequence generated by Gauss-Newton’s method (1.4) with initial point \( x_0 \). Then \( \{ x_n \} \) converges to a zero \( x^* \) of \( DF(\cdot)^T F(\cdot) \) in \( B(x_0, t_0^*) \) and the following inequalities hold for each \( n = 1, 2, \ldots \):

\[
\| x_n - x^* \| \leq t_0^* - t_n, \quad (3.23)
\]

\[
\| x_{n+1} - x_n \| \leq t_{n+1} - t_n, \quad (3.24)
\]

and

\[
\| x_{n+1} - x_n \| \leq \left( \frac{t_{n+1} - t_n}{t_n - t_{n-1}} \right) \| x_n - x_{n-1} \|. \quad (3.25)
\]

**Proof.** By Theorem 3.1, to complete the proof, it remains to verify that (3.25) is true. We will proceed by mathematical induction. Since \( \| x_1 - x_0 \| = \beta = t_1 - t_0 \), the case when \( n = 1 \) follows from (3.10) (with \( n = 1 \) and \( \lambda = 0 \)). Assume that (3.25) is true for \( n = k - 1 \). Below, we will show that (3.25) holds for \( n = k \). As \( \kappa = 0 \) and \( \lambda = 0 \), (3.15) and (3.18) yield that

\[
\| x_{k+1} - x_k \| \leq -h_0'(t_k)^{-1}\int_0^1 \| x_{k-1} - x_0 \| \| x_k - x_{k-1} \| L(u) \| x_k - x_{k-1} \| ds
\]

\[
\leq -h_0'(t_k)^{-1}h_0(t_k) \left( \frac{\| x_k - x_{k-1} \|}{t_k - t_{k-1}} \right)
\]

\[
= \left( \frac{t_{k+1} - t_k}{t_k - t_{k-1}} \right) \| x_k - x_{k-1} \|,
\]

where the last equality holds because \(-h_0'(t_k)^{-1}h_0(t_k) = t_{k+1} - t_k\). The proof is complete. \( \square \)

In the case when \( DF(x_0) \) is full row rank, the modified \( L \)-average Lipschitz condition in above corollary can be replaced with the \( L \)-average Lipschitz condition.

**Corollary 3.2.** Let \( x_0 \in \Omega \) be such that \( DF(x_0) \) is full row rank. Suppose that

\[
\beta \leq b_0, \quad B(x_0, t_0^*) \subseteq \Omega, \quad (3.26)
\]

and that \( DF \) satisfies the \( L \)-average Lipschitz condition on \( B(x_0, t_0^*) \). Let \( \{ x_n \} \) be the sequence generated by Gauss-Newton’s method (1.4) with initial point \( x_0 \). Then \( \{ x_n \} \) converges to a zero \( x^* \) of \( F(\cdot) \) in \( B(x_0, t_0^*) \). Moreover, the estimates (3.23)-(3.25) and the following inequality hold for each \( n = 1, 2, \ldots \):

\[
\| DF(x_0)^T F(x_n) \| \leq \left( \frac{t_{n+1} - t_n}{t_n - t_{n-1}} \right) \| DF(x_0)^T F(x_{n-1}) \|. \quad (3.27)
\]

**Proof.** Take \( \tilde{\Omega} := B(x_0, t_0^*) \) and consider the operator \( \tilde{F} : \tilde{\Omega} \rightarrow \mathbb{R}^l \) defined by \( \tilde{F}(x) = DF(x_0)^T F(x) \) for each \( x \in \tilde{\Omega} \). We shall apply Theorem 3.2 to \( \tilde{F} \). For this end, we claim that \( DF(x) \) is full row rank for each \( x \in \tilde{\Omega} \). In fact, since

\[
\| DF(x_0)^T (DF(x) - DF(x_0)) \| \leq \int_0^{\| x-x_0 \|} L(u) du < \int_0^{t_0^*} L(u) du \leq 1.
\]

By the Banach Lemma, \((I_{\mathbb{R}^m} - DF(x_0)^T (DF(x_0) - DF(x)))^{-1}\) exists and

\[
\| (I_{\mathbb{R}^m} - DF(x_0)^T (DF(x_0) - DF(x)))^{-1} \| \leq \frac{1}{1 - \int_0^{\| x-x_0 \|} L(u) du} = -h_0'(\| x - x_0 \|)^{-1}. \quad (3.28)
\]
Noting that $DF(x_0)$ is full row rank, we have that $DF(x_0)DF(x_0)^\dagger = I_{\mathbb{R}^k}$ and
\[
DF(x) = DF(x_0)\left(I_{\mathbb{R}^m} - DF(x_0)^\dagger (DF(x_0) - DF(x))\right).
\]
This implies that $DF(x)$ is full row rank because $I_{\mathbb{R}^m} - DF(x_0)^\dagger (DF(x) - DF(x_0))$ is invertible; hence the claim stands. Thus, in view of the definition of the Moore-Penrose inverse, one sees that
\[
(DF(x))_{\dagger} = (DF(x_0)^\dagger DF(x))^\dagger = DF(x)^\dagger DF(x_0)
\]
for each $x \in \tilde{\Omega}$.

This implies that (3.21) holds with $\tilde{F}, \tilde{\Omega}$ in place of $F, \Omega$, and that \{x_n\} coincides with the sequence generated by Gauss-Newton’s method (1.4) for $\tilde{F}$ with initial point $x_0$. Furthermore, since by (3.29)
\[
(DF(x_0))_{\dagger} = (DF(x_0)^\dagger DF(x_0))^\dagger = DF(x_0)^\dagger DF(x_0),
\]
it follows that
\[
\|DF(x_0)\| = \|DF(x_0)^\dagger DF(x_0)\| = \|DF(x_0)^\dagger F(x_0)\| = \|DF(x_0)^\dagger F(x_0)\|
\]
and
\[
\|DF(x_0)\| = \|DF(x_0)^\dagger DF(x_0)\| = \|\Pi_{\ker DF(x_0)^\dagger}\| = 1.
\]
Therefore, thanks to (3.3) and (3.26), the assumptions in Theorem 3.2 hold with $\tilde{F}$ in place of $F$. Consequently, Theorem 3.2 is applicable and so \{x_n\} converges to a zero $x^*$ of $DF(\cdot)^\dagger \tilde{F}(\cdot)$. Noting that $DF(\cdot)^\dagger \tilde{F}(\cdot) = DF(\cdot)^\dagger F(\cdot)$ and $F(\cdot) = DF(\cdot)(DF(\cdot)^\dagger F(\cdot))$, it follows that $x^*$ is a zero of $F(\cdot)$. Moreover, the estimates (3.9), (3.10) (with $\lambda = 0$) and (3.25) hold. To complete the proof, it remains to show that (3.27) are true for each $n = 1, 2, \ldots$. To do this, fix $k \in \{1, 2, \ldots\}$. Since $DF(x_0)DF(x_0)^\dagger = I_{\mathbb{R}^k}$, it follows that
\[
DF(x_k-1)^\dagger DF(x_0)(I_{\mathbb{R}^m} - DF(x_0)^\dagger (DF(x_0) - DF(x_{k-1}))) = DF(x_{k-1})^\dagger DF(x_{k-1}).
\]
This together with (2.14) implies that
\[
\begin{align*}
\|DF(x_{k-1})^\dagger DF(x_0)\| &= \|\Pi_{\ker DF(x_{k-1})\perp} (I_{\mathbb{R}^m} - DF(x_0)^\dagger (DF(x_0) - DF(x_{k-1})))^{-1}\| \\
&\leq \| (I_{\mathbb{R}^m} - DF(x_0)^\dagger (DF(x_0) - DF(x_{k-1})))^{-1}\| \\
&\leq h_0(t_{k-1})^{-1},
\end{align*}
\]
where the last inequality holds because of (3.28) and the fact that $\|x_{k-1} - x_0\| \leq t_{k-1} < t_0^*$ (thanks to (3.10) with $\lambda = 0$). Consequently,
\[
\|x_k - x_{k-1}\| = \|DF(x_{k-1})^\dagger F(x_{k-1})\| \leq -h_0(t_{k-1})^{-1}\|DF(x_{k-1})^\dagger F(x_{k-1})\| \tag{3.34}
\]
(noting $DF(x_0)DF(x_0)^\dagger = I_{\mathbb{R}^l}$). Since $DF(x_{k-1})DF(x_{k-1})^\dagger = I_{\mathbb{R}^l}$, it follows that
\[
DF(x_0)^\dagger F(x_k) = DF(x_0)^\dagger (F(x_k) - F(x_{k-1}) - DF(x_{k-1}) (x_k - x_{k-1})).
\]
This together with (3.3) and (3.15) (with $\lambda = 0$) yields that
\[
\begin{align*}
\|DF(x_0)^\dagger F(x_k)\| &\leq \int_0^1 \|DF(x_0)^\dagger (DF(x_k-1 + s(x_k - x_{k-1})) - DF(x_{k-1}))(x_k - x_{k-1})\| ds \\
&\leq \int_0^1 \left[ \|x_k - x_{k-1}\| + s\|x_k - x_{k-1}\| \right] L(u) \|x_k - x_{k-1}\| ds \\
&\leq h_0(t_k) \left( \frac{\|x_k - x_{k-1}\|}{t_k - t_{k-1}} \right).
\end{align*}
\]
Combining this with (3.34) and (3.10) (with $\lambda = 0$ and $n = k$), we get that

$$
\|DF(x_0)^TF(x_k)\| \leq -h'_0(t_{k-1})^{-1}h_0(t_k)\|DF(x_0)^TF(x_{k-1})\|
$$

$$
\leq -h'_0(t_{k-1})^{-1}h_0(t_k)\|DF(x_0)^TF(x_{k-1})\|
$$

$$
= \left( \frac{t_{k+1} - t_k}{t_k - t_{k-1}} \right) \|DF(x_0)^TF(x_{k-1})\|.
$$

This shows that (3.27) holds for each $k \in \{1, 2, \ldots\}$ and completes the proof.

**Remark 3.1.** Even in the case when $DF(x_0)$ is invertible, Corollary 3.2 extends the corresponding result in [33, Theorem 3.1], which requires the additional assumption that $\int_0^R L(u)du \geq 1$. Example 3.1 below presents such a case when Corollary 3.2 is applicable but not [33, Theorem 3.1] (note that $DF(x_0)$ is invertible).

**Example 3.1.** Let $m = l = 1$ and $\Omega = [-1, 1]$. Consider the operator $F : \Omega \to \mathbb{R}$ defined by

$$
F(x) = \frac{7}{9} - x + \int_0^x \frac{x - u}{3\sqrt{1 - u}} du \text{ for each } x \in \Omega.
$$

Let $x_0 = 0$ and let $L : [0, 1) \to \mathbb{R}$ be defined by $L(u) = 1/(3\sqrt{1-u})$ for each $0 \leq u < 1$ (and so $R = 1$). Then $DF(x_0) = -1$ and it is routine to verify that $DF$ satisfies the $L$-average Lipschitz condition on $B(x_0, 1)$. Furthermore, we have $t_0 = r_0 = 1$ and $b_0 = 7/9$. Hence, $\beta = \|DF(x_0)^TF(x_0)\| = \frac{\sqrt{3}}{18} \leq b_0$ and $B(x_0, t_0^\beta) \subseteq \Omega$. Thus Corollary 3.2 is applicable and the sequence generated by Gauss-Newton’s method (1.4) with initial point $x_0$ converges to a zero of $F$. Below we will show that [33, Theorem 3.1] is not applicable. In fact, otherwise, there exists a positive nondecreasing integrable function $\tilde{L}$ on $[0, \tilde{R})$, with $\tilde{R} \leq 1$ and $\int_0^\tilde{R} \tilde{L}(u)du \geq 1$, such that $\beta \leq \tilde{b}_0$ and $DF$ satisfies the $\tilde{L}$-average Lipschitz condition on $B(x_0, t_0^\beta)$, where $b_0$ and $t_0^\beta$ are the ones corresponding to $\tilde{L}$. Then $\tilde{L} \geq L$ and $\tilde{L} \neq L$ a.e. on $[0, t_0^\beta]$. Let $h_0$ and $h_0$ be the majorizing functions corresponding to $(0, L)$ and $(0, \tilde{L})$, respectively. Then $\hat{h} \geq h$ on $[0, t_0^\beta]$ and so $\hat{h}_0 \geq t_0^\beta = 1$. This implies that $\tilde{R} = \tilde{r}_0 = r_0 = t_0^\beta = 1$. Consequently,

$$
\tilde{b}_0 = 1 - \int_0^1 \tilde{L}(u)(1-u)du < 1 - \int_0^1 L(u)(1-u)du = \tilde{b}_0 = \beta.
$$

Hence [33, Theorem 3.1] is not applicable.

## 4 Convergence ball of Gauss-Newton’s method

Recall that $r_0$ is defined by (2.4) for $\lambda = 0$. Through the whole section, let $x^* \in \Omega$ be such that $F(x^*) = 0$ and $DF(x^*) \neq 0$. Furthermore, we shall assume that $B(x^*, r_0) \subseteq \Omega$ and

$$
\text{rank}(DF(x)) \leq \text{rank}(DF(x^*)) \text{ for any } x \in \Omega. \tag{4.1}
$$

The following lemma estimates the quantity $\|DF(x_0)^TF(x_0)\|$, which will be used in the proof of the main theorem of this section.

**Lemma 4.1.** Let $0 < r \leq r_0$. Suppose that $DF$ satisfies the modified $L$-average Lipschitz condition on $B(x^*, r)$. Then, for each $x_0 \in B(x^*, r)$, $\text{rank}(DF(x_0)) = \text{rank}(DF(x^*))$ and

$$
\|DF(x_0)^TF(x_0)\| \leq \frac{\|x_0 - x^*\| + \|x_0 - x^*\|}{1 - \int_0^\|x_0 - x^*\| L(u)(u - \|x_0 - x^*\|)du} \int_0^\|x_0 - x^*\| L(u)du. \tag{4.2}
$$
Consequently, it follows that $\Delta = 1 - 1 - \int_0^{\|x_0 - x^*\|} L(u)du$.

Then, the following assertions hold.

(i). Let $x_0 \in B(x^*, r)$. By Lemma 3.2, we have that $\text{rank}(DF(x_0)) = \text{rank}(DF(x^*))$ and

$$\|DF(x_0)^\dagger\| \leq \frac{\|DF(x^*)\|}{1 - \int_0^{\|x_0 - x^*\|} L(u)du}.$$  

(4.3)

Since

$$-DF(x_0)^\dagger F(x_0) = DF(x_0)^\dagger (F(x^*) - F(x_0) - DF(x_0)(x^* - x_0)) + \int_0^1 (DF(x_0) - DF(x^* + s(x_0 - x^*)))(x_0 - x^*)ds$$

$$+ \Pi_{(\ker DF(x_0))}(x^* - x_0),$$

it follows from (4.3) and the assumed modified $L$-average Lipschitz condition that

$$\|DF(x_0)^\dagger F(x_0)\| \leq \frac{1}{1 - \int_0^{\|x_0 - x^*\|} L(u)du} \int_0^1 \int_{\|x_0 - x^*\|} L(u)du\|x_0 - x^*\|ds + \|x_0 - x^*\|$$

$$= \frac{\|x_0 - x^*\| + \int_0^{\|x_0 - x^*\|} L(u)(u - \|x_0 - x^*\|)du}{1 - \int_0^{\|x_0 - x^*\|} L(u)du}.$$  

(4.4)

Hence (4.2) is proved. \qed

Recall that $r_\kappa$ and $b_\kappa$ are respectively defined by (2.4) and (2.5) for $\lambda = \kappa$.

**Lemma 4.2.** Suppose that $DF$ satisfies the modified $L$-average Lipschitz condition on $B(x^*, r_0)$. Let $x_0 \in B(x^*, r_0)$ and let $\bar{L} : [0, R - \|x_0 - x^*\|) \to \mathbb{R}$ be defined by

$$\bar{L}(u) = \frac{L(u + \|x_0 - x^*\|)}{1 - \int_0^{\|x_0 - x^*\|} L(u)du} \quad \text{for each } u \in [0, R - \|x_0 - x^*\|).$$

(4.5)

Then the following assertions hold.

(i) $r_\kappa \leq \bar{r}_\kappa + \|x_0 - x^*\| \leq r_0$, where $\bar{r}_\kappa$ is given by (2.4) with $\bar{L}$ and $\kappa$ in place of $L$ and $\lambda$.

(ii) $DF$ satisfies the modified $\bar{L}$-average Lipschitz condition on $B(x_0, r_0 - \|x_0 - x^*\|)$.

**Proof.** (i). It suffices to verify that

$$\int_0^{r_\kappa} L(u)du \leq \int_0^{\bar{r}_\kappa + \|x_0 - x^*\|} L(u)du \leq \int_0^{r_0} L(u)du.$$  

(4.6)

Note that the first inequality in (4.7) is trivial if $\bar{r}_\kappa = R - \|x_0 - x^*\|$. We assume that $\bar{r}_\kappa < R - \|x_0 - x^*\|$. Then, $\int_0^{R - \|x_0 - x^*\|} L(u)du > 1 - \kappa$ by the definition of $\bar{r}_\kappa$. Since

$$\Delta \leq 1 - \kappa \implies \int_0^{R - \|x_0 - x^*\|} L(u)du \leq 1 - \kappa,$$  

(4.7)

it follows that $\Delta > 1 - \kappa$. Hence, by the definitions of $r_\kappa$ and $\bar{r}_\kappa$,

$$\int_0^{r_\kappa} L(u)du = 1 - \kappa \quad \text{and} \quad \int_0^{\bar{r}_\kappa} L(u)du = 1 - \kappa.$$  

(4.8)

Consequently,

$$\int_0^{\|x_0 - x^*\|} L(u)du = \int_0^{r_\kappa} L(u + \|x_0 - x^*\|)du = (1 - \kappa) \left(1 - \int_0^{\|x_0 - x^*\|} L(u)du\right).$$

(4.9)
thanks to (4.6) and the definition of $\tilde{r}_{\kappa}$. Hence
\[
\int_{\|x_0-x^*\|}^{r_{\kappa}+\|x_0-x^*\|} L(u)\,du \geq (1-\kappa) - \int_{0}^{\|x_0-x^*\|} L(u)\,du = \int_{0}^{r_{\kappa}} L(u)\,du - \int_{0}^{\|x_0-x^*\|} L(u)\,du.
\]
This implies that the first inequality in (4.7) holds.

To show the second inequality in (4.7), we only need consider the case when $\Delta \geq 1$ because, otherwise, one has that $r_0 = R$ and hence $\tilde{r}_{\kappa} + \|x_0 - x^*\| \leq (R - \|x_0 - x^*\|) + \|x_0 - x^*\| = r_0$. Thus $\int_{0}^{r_{\kappa}} L(u)\,du = 1$. By (4.6) and the definition of $\tilde{r}_{\kappa}$, one has that
\[
\int_{\|x_0-x^*\|}^{r_{\kappa}+\|x_0-x^*\|} L(u)\,du = \int_{0}^{r_{\kappa}} L(u + \|x_0 - x^*\|)\,du \leq (1-\kappa) \left(1 - \int_{0}^{\|x_0-x^*\|} L(u)\,du\right)
\]
and so
\[
\int_{\|x_0-x^*\|}^{r_{\kappa}+\|x_0-x^*\|} L(u)\,du \leq 1 - \int_{0}^{\|x_0-x^*\|} L(u)\,du = \int_{0}^{r_{\kappa}} L(u)\,du - \int_{0}^{\|x_0-x^*\|} L(u)\,du.
\]
This implies that the second inequality in (4.7) holds and completes the proof of (i).

(ii) Let $x, x' \in B(x_0, r_0 - \|x_0 - x^*\|)$ satisfy $\|x - x_0\| + \|x' - x\| < r_0 - \|x_0 - x^*\|$. Then
\[
\|x - x^*\| + \|x' - x\| \leq \|x - x_0\| + \|x_0 - x^*\| + \|x' - x\| < r_0.
\]
Since $DF$ satisfies the modified $L$-average Lipschitz condition on $B(x^*, r_0)$, it follows that
\[
\|DF(x^*)\|^1 \|DF(x') - DF(x)\| \leq \int_{\|x-x^*\|}^{\|x-x^*\|+\|x'-x\|} L(u)\,du \leq \int_{\|x_0-x^*\|+\|x-x_0\|}^{\|x_0-x^*\|+\|x'-x\|} L(u)\,du,
\]
where the last inequality holds because $L$ is nondecreasing. Using (4.3) and (4.6), one gets that
\[
\|DF(x_0)^1 \| \|DF(x') - DF(x)\| \leq \int_{\|x-x_0\|}^{\|x-x_0\|+\|x'-x\|} \tilde{L}(u)\,du.
\]
This shows that $DF$ satisfies the modified $\tilde{L}$-average Lipschitz condition on $B(x_0, r_0 - \|x_0 - x^*\|)$ and the proof is complete.

Define the function $\phi_\kappa$ on $[0, r_{\kappa}]$ by
\[
\phi_\kappa(t) = b_{\kappa} - (2-\kappa) t + \kappa(r_{\kappa} - t) \int_{0}^{t} L(u)\,du + 2 \int_{0}^{t} L(u)(t-u)\,du \quad \text{for each } t \in [0, r_{\kappa}].
\]  

**Lemma 4.3.** $\phi_\kappa$ is a strictly decreasing continuous function on $[0, r_{\kappa}]$, and has exact one zero $\tilde{r}_{\kappa}$ in $[0, r_{\kappa}]$ and it satisfies $\frac{b_{\kappa}}{2-\kappa} < \tilde{r}_{\kappa} < r_{\kappa}$.

**Proof.** Note that $L$ is nondecreasing on $[0, r_{\kappa}]$. Then
\[
\phi_\kappa'(t) = -(2-\kappa) + \kappa(r_{\kappa} - t)L(t) + (2-\kappa) \int_{0}^{t} L(u)\,du \quad \text{for a.e. } t \in [0, r_{\kappa}].
\]
Furthermore, for any $0 \leq t_1 < t_2 \leq r_\kappa$,
\[
\phi'_\kappa(t_2) - \phi'_\kappa(t_1) = \kappa(r_\kappa - t_1)(L(t_2) - L(t_1)) + \kappa L(t_2)(t_1 - t_2) + (2 - \kappa) \int_{t_1}^{t_2} L(u)du \\
\geq \kappa(r_\kappa - t_1)(L(t_2) - L(t_1)) + 2(1 - \kappa) \int_{t_1}^{t_2} L(u)du \\
> 0
\]
Hence $\phi'_\kappa$ is strictly increasing on $[0, r_\kappa]$. We claim that
\[
\phi'_\kappa(r_\kappa - \cdot) = \lim_{t \to r_\kappa^-} \phi'_\kappa(t) = -(2 - \kappa) + (2 - \kappa) \int_{0}^{r_\kappa} L(u)du \leq 0. \tag{4.12}
\]
Granting this, $\phi_\kappa$ is a strictly decreasing continuous function on $[0, r_\kappa]$. To verify (4.12), we assume that $r_\kappa = R$ (as, otherwise, (4.12) is clear). Thus, one has that $\int_{0}^{R} L(u)du \leq 1 - \kappa$. This implies that $\liminf_{t \to R-} (R - t)L(t) = 0$ because, otherwise, there exist $R_0 > 0$ and $a > 0$ such that $(R - t)L(t) \geq a$ for each $t \in [R_0, R)$, which contradicts that $\int_{0}^{R} L(u)du < +\infty$. Consequently,
\[
\lim_{t \to r_\kappa^-} \phi'_\kappa(t) = \liminf_{t \to R-} \phi'_\kappa(t) = -(2 - \kappa) + \liminf_{t \to R-} (R - t)L(t) + (2 - \kappa) \int_{0}^{r_\kappa} L(u)du
\]
and (4.12) is seen to hold. Finally, since
\[
\phi_\kappa \left( \frac{b_\kappa}{2 - \kappa} \right) = \kappa \left( r_\kappa - \frac{b_\kappa}{2 - \kappa} \right) \int_{0}^{b_\kappa} L(u)du + 2 \int_{0}^{b_\kappa} L(u) \left( \frac{b_\kappa}{2 - \kappa} - u \right) du > 0
\]
and
\[
\phi_\kappa(r_\kappa) = r_\kappa \left( \int_{0}^{r_\kappa} L(u)du - 1 \right) - \int_{0}^{r_\kappa} L(u)du < 0.
\]
it follows that $\phi_\kappa$ has exact one zero $\hat{r}_\kappa$ in $[0, r_\kappa]$ and $\frac{b_\kappa}{2 - \kappa} < \hat{r}_\kappa < r_\kappa$.

The following theorem gives an estimate of the radius of the convergence ball of Gauss-Newton’s method. As explained at the beginning of section 3, we assume that (3.1) holds for some $0 \leq \kappa < 1$. Recall that $\hat{r}_\kappa$ is the unique zero of $\phi_\kappa$ in $[0, r_\kappa]$.

**Theorem 4.1.** Suppose that $DF$ satisfies the modified $L$-average Lipschitz condition on $B(x^*, r_0)$. Let $x_0 \in B(x^*, \hat{r}_\kappa)$ and let $\{x_n\}$ be the sequence generated by Gauss-Newton’s method (1.4) with initial point $x_0$. Then $\{x_n\}$ converges to a zero of $DF(\cdot)^\dag F(\cdot)$.

**Proof.** Let $\hat{r} = r_0 - \|x_0 - x^*\|$ and $\bar{L} : [0, \hat{r}] \to \mathbb{R}$ be defined by (4.6). Also let $\hat{r}_\kappa$, $\bar{b}_\kappa$ be given by (2.4) and (2.5) with $\kappa$, $\bar{L}$ in place of $\lambda$, $L$. Then, by Lemma 4.2, $\hat{r}_\kappa \leq \hat{r}$ and $DF$ satisfies the modified $L$-average Lipschitz condition on $B(x_0, \hat{r})$. Since $\|x_0 - x^*\| \leq \hat{r}_\kappa < r_\kappa \leq r_0$, it follows from Lemma 4.1 that $\text{rank}(DF(x_0)) = \text{rank}(DF(x^*))$ and (4.2) holds. Consequently, $\text{rank}(DF(x)) \leq \text{rank}(DF(x_0))$ for each $x \in \Omega$ by (4.1); hence (3.2) holds.

Below we shall show that
\[
\beta = \|DF(x_0)^\dag F(x_0)\| \leq \bar{b}_\kappa. \tag{4.13}
\]
Granting this, $\hat{r} \geq \hat{r}_\kappa \geq \hat{t}_\kappa^*$ and Theorem 3.1 is applicable, where $\hat{t}_\kappa^*$ is the corresponding $t_\lambda^*$ for $\lambda = \kappa$ and $L = \bar{L}$. Therefore the conclusion follows.
For simplicity, we write $s_0 := \|x_0 - x^*\|$. Then, by (4.2),
\[
\beta \leq \frac{s_0 + \int_0^{s_0} L(u)(u-s_0)du}{1 - \int_0^{s_0} L(u)du}.
\] (4.14)

Since $r_\kappa \geq r_\kappa - \|x_0 - x^*\| = r_\kappa - s_0$ by Lemma 4.2 and the function $t \mapsto (1 - \kappa)t - \int_0^t L(u)(t-u)du$ is increasing on $[0, r_\kappa]$, it follows from the definition of $\hat{r}_\kappa$ that
\[
\hat{b}_\kappa \geq (1 - \kappa)(r_\kappa - s_0) - \int_0^{r_\kappa - s_0} L(u)(r_\kappa - s_0 - u)u = \frac{b_\kappa - (1 - \kappa)s_0 - (1 - \kappa)(r_\kappa - s_0) \int_0^{s_0} L(u)du + \int_0^{s_0} L(u)(r_\kappa - u)du}{1 - \int_0^{s_0} L(u)du}.
\] (4.15)

Thus, to verify (4.13), it suffices to verify that
\[
s_0 + \int_0^{s_0} L(u)(u-s_0)du \leq b_\kappa - (1 - \kappa)s_0 - (1 - \kappa)(r_\kappa - s_0) \int_0^{s_0} L(u)du + \int_0^{s_0} L(u)(r_\kappa - u)du,
\] (4.16)
or equivalently, $\phi_\kappa(s_0) \geq 0$, which is clear because $\phi_\kappa(s_0) \geq \phi_\kappa(\hat{r}_\kappa) = 0$ by Lemma 4.3 (as $s_0 \leq \hat{r}_\kappa$). The proof is complete.

By Lemma 4.3 and Theorem 4.1, the following corollary is straightforward.

**Corollary 4.1.** Suppose that $DF$ satisfies the modified $L$-average Lipschitz condition on $B(x^*, r_0)$. Let $x_0 \in B(x^*, \frac{b_\kappa}{2})$ and let $\{x_n\}$ be the sequence generated by Gauss-Newton’s method (1.4) with initial point $x_0$. Then $\{x_n\}$ converges to a zero of $DF(\cdot)^\dagger F(\cdot)$.

For the case when (3.21) holds, that is, $\kappa = 0$, $\phi_\kappa$ reduces to
\[
\phi_0(t) = b_0 - 2t + 2 \int_0^t L(u)(t-u)du \text{ for each } t \in [0, r_0]
\] (4.17)
and $\hat{r}_0$ is the corresponding zero of $\phi_0$. The following corollary is direct from Theorem 4.1.

**Corollary 4.2.** Suppose that (3.21) holds and that $DF$ satisfies the modified $L$-average Lipschitz condition on $B(x^*, r_0)$. Let $x_0 \in B(x^*, \frac{b_\kappa}{2})$ and let $\{x_n\}$ be the sequence generated by Gauss-Newton’s method (1.4) with initial point $x_0$. Then $\{x_n\}$ converges to a zero of $DF(\cdot)^\dagger F(\cdot)$.

Using Corollary 4.2, instead of Theorem 3.2, the argument for Corollary 3.2 works for the following corollary.

**Corollary 4.3.** Suppose that $DF(x^*)$ is full row rank and that $DF$ satisfies the $L$-average Lipschitz condition on $B(x^*, r_0)$. Let $x_0 \in B(x^*, \frac{b_\kappa}{2})$ and let $\{x_n\}$ be the sequence generated by Gauss-Newton’s method (1.4) with initial point $x_0$. Then $\{x_n\}$ converges to a zero of $F(\cdot)$.

## 5 Applications

This section is divided into four subsections: for the first two we consider the applications of our main results specializing respectively in the Kantorovich type condition and in the $\gamma$-condition studied by Wang and Han in [36] as well as the Smale type condition employed by Dedieu and Shub [6]. The third one is devoted to an extension of the Smale approximate zeros. The last one is devoted to a study of extensions of the Smale type condition for analytic operators used in [29] and of some examples studied by Wang in [34]. In particular, our results extend and improve some of the corresponding results in [16, 24, 29].

As in the previous sections, let $\beta = \|DF(x_0)^\dagger F(x_0)\|$ and $\kappa \in [0, 1)$ be such that (3.1) holds.
5.1 Kantorovich type condition

Throughout this subsection, let $L$ be a positive constant. Recall that an operator $T$ from $\Omega$ to a Banach space $X$ is said to be Lipschitz continuous on $\Omega_0 \subseteq \Omega$ with modulus $L$ if

$$\|T(x) - T(y)\| \leq L\|x - y\| \quad \text{for any } x, y \in \Omega.$$ 

Let $x_0 \in \Omega$ and $r > 0$ be such that $B(x_0, r) \subseteq \Omega$. It is clear that, if $\|DF(x_0)\|DF$ is Lipschitz continuous on $B(x_0, r)$ with modulus $L$, then $DF$ satisfies the modified $L$-average Lipschitz condition on $B(x_0, r)$.

For the case when $L$ is a constant function, by (2.4), (2.5) and (2.1), we have that

$$r_\lambda = \frac{1 - \lambda}{L}, \quad b_\lambda = \frac{(1 - \lambda)^2}{2L}$$

and

$$h_\lambda(t) = \beta - (1 - \lambda)t + \frac{L}{2}t^2 \quad \text{for each } t \geq 0.$$ (5.2)

Moreover, if $\beta \leq \frac{(1 - \lambda)^2}{2L}$, then the zero of $h_\lambda$ in $[0, r_\lambda]$ is given by

$$t_\lambda^* = \frac{1 - \lambda - \sqrt{(1 - \lambda)^2 - 2\beta L}}{L}.$$ (5.3)

Recall that $x_0 \in \Omega$ is such that $DF(x_0) \neq 0$ and (3.2) holds. Then we have the following theorem, which improves the corresponding result in [15].

**Theorem 5.1.** Let $\lambda = \lambda_0 := (1 - \beta L)\kappa$. Suppose that $\beta L \leq \Delta := \frac{(1 - \kappa)^2}{(\kappa^2 - \kappa + 1) + \sqrt{2\kappa^2 - 2\kappa + 1}}$, $B(x_0, \frac{\kappa}{L}) \subseteq \Omega$ (5.4) and that $\|DF(x_0)\|DF$ is Lipschitz continuous on $B(x_0, \frac{\kappa}{L})$ with modulus $L$. Let $\{x_n\}$ be the sequence generated by Gauss-Newton’s method (1.4) with initial point $x_0$. Then $\{x_n\}$ converges to a zero $x^*$ of $DF(\cdot)^T F(\cdot)$ in $B(x_0, \frac{\kappa}{L})$ and the following estimate holds:

$$\|x_n - x^*\| \leq t_\lambda^* - t_{\lambda,n} \quad \text{for each } n \geq 0,$$

where $t_{\lambda,n}$ is the sequence generated by (2.8) for the function $h_\lambda$ defined by (5.2).

**Proof.** By (5.1), we have $\beta \leq b_\lambda \iff \beta L \leq \Delta$. Therefore, (5.4) and (3.8) are the same and so the conclusion of Theorem 3.1 holds. This completes the proof. \qed

In the case when $\kappa = 0$, we have that $\lambda = 0$, $\Delta = \frac{1}{2}$ and the corresponding sequence $\{t_n\}$ coincides with the Newton sequence for $h_0$. It is well known (see for example [11, 23, 31]) that, if $\beta L \leq \frac{1}{2}$, $\{t_n\}$ has the closed form:

$$t_n = \frac{1 - q^{2n-1}}{1 - q^{2n}} t_0^*$$

for each $n \geq 0$, (5.5)

where

$$q = \frac{1 - \sqrt{1 - 2\beta L}}{1 + \sqrt{1 - 2\beta L}} \quad \text{and} \quad t_0^* = \frac{1 - \sqrt{1 - 2\beta L}}{L}.$$ (5.6)

Thus, applying Theorem 5.1 and Corollary 3.2, we immediately get the following corollary. Let $q$ and $t_0^*$ be defined by (5.6).
Corollary 5.1. Suppose that $\beta L \leq \frac{1}{2}$, $B(x_0, t^*_{0}) \subseteq \Omega$ and let $\{x_n\}$ be the sequence generated by Gauss-Newton’s method (1.4) with initial point $x_0$. Suppose that $\|DF(x_0)\|DF$ (resp. $DF(x_0)^{\dagger}DF$) is Lipschitz continuous on $B(x_0, t^*_{0})$ with modulus $L$ and that $\kappa = 0$ (resp. $DF(x_0)$ is full row rank). Then $\{x_n\}$ converges to a zero $x^*$ of $DF(\cdot)^{\dagger}F(\cdot)$ (resp. $F(\cdot)$) in $B(x_0, t^*_{0})$ and the following estimate holds:

$$\|x_n - x^*\| \leq \frac{q^{n-1}}{\sum_{i=0}^{\infty} q^i} t^*_{0} \quad \text{for each } n \geq 0. \quad (5.7)$$

Recall that $x^* \in \Omega$ is such that $F(x^*) = 0$, $DF(x^*) \neq 0$ and (4.1) holds. As before, we assume that $B(x^*, 1/L) \subseteq \Omega$. We have the following result, which was also considered in [18] by Hu, Shen and Li with another approach.

Theorem 5.2. Suppose that $\|DF(x^*)\|DF$ is Lipschitz continuous on $B(x^*, 1/L)$ with modulus $L$. Let $x_0 \in B \left( x^*, \frac{1-\sqrt{3\Delta+1}}{L} \right)$ and let $\{x_n\}$ be the sequence generated by Gauss-Newton’s method (1.4) with initial point $x_0$. Then $\{x_n\}$ converges to a zero of $DF(\cdot)^{\dagger}F(\cdot)$.

Proof. Let $\bar{L} := \frac{L}{1 - L\|x_0 - x^*\|}$ and $\bar{r} := \frac{1}{2} - \|x_0 - x^*\|$. Let $\bar{t}^*_\kappa$ and $\bar{r}_\kappa$ denote respectively the corresponding $\bar{t}^*_\kappa$ and $\bar{r}_\kappa$ given in (5.1) and (5.3) with $\bar{L}$ in place of $L$. Then, by Lemma 4.2, $\|DF(x_0)\|DF$ is Lipschitz continuous on $B(x_0, \bar{r})$ with modulus $\bar{L}$, and

$$\bar{t}^*_\kappa \leq \bar{t}^*_\kappa \leq \bar{r}_\kappa \leq r_0 - \|x_0 - x^*\| = \bar{r}$$

because $\lambda_0 \leq \kappa$ (where $\lambda_0 = (1 - \beta L)\kappa$). Hence $B(x^*, \bar{t}^*_\kappa) \subseteq B(x_0, \bar{r}) \subseteq \Omega$. Furthermore, by Lemma 4.1,

$$\beta \leq \frac{2 - L\|x^* - x_0\|}{2(1 - L\|x^* - x_0\|)} \|x^* - x_0\|.$$

It follows that

$$\beta \bar{L} \leq \frac{(2 - L\|x^* - x_0\|)L\|x^* - x_0\|}{2(1 - L\|x^* - x_0\|)^2} \leq \Delta, \quad (5.8)$$

since $L\|x_0 - x^*\| \leq 1 - \frac{1}{\sqrt{3\Delta+1}}$ and the function $t \mapsto \frac{(2-t)L}{2(1-t)^2}$ is increasing on $(0, 1)$. Thus, Theorem 5.1 is applicable and the conclusion holds. This completes the proof.

In particular, for the case when $\kappa = 0$ (so $\Delta = \frac{1}{2}$), we have the following corollary, which is a consequence of Theorem 5.2 (and its proof).

Corollary 5.2. Suppose that $\|DF(x^*)\|DF$ (resp. $DF(x^*)^{\dagger}DF$) is Lipschitz continuous on $B(x^*, 1/L)$ with modulus $L$ and that $\kappa = 0$ (resp. $DF(x^*)$ is full row rank). Let $x_0 \in B \left( x^*, \frac{2\sqrt{2}}{2\sqrt{2}+1} \right)$ and let $\{x_n\}$ be the sequence generated by Gauss-Newton’s method (1.4) with initial point $x_0$. Then $\{x_n\}$ converges to a zero of $DF(\cdot)^{\dagger}F(\cdot)$ (resp. $F(\cdot)$).

5.2 $\gamma$-condition

Throughout this subsection, we assume that $\gamma > 0$ and $F$ has continuous second derivative. The notion of the $\gamma$-condition for operators in Banach spaces was introduced in [36] by Wang and Han to study the Smale point estimate theory, which was recently extended in [21] to the setting of Riemannian manifolds.
Definition 5.1. Let $0 < r \leq \frac{1}{\gamma}$ be such that $B(x_0, r) \subseteq \Omega$. $F$ is said to satisfy the $\gamma$-condition (resp. the modified $\gamma$-condition) on $B(x_0, r)$ if (5.9) (resp. (5.10)) below holds.

\[ \|DF(x_0)^\dagger D^2 F(x)\| \leq \frac{2\gamma}{(1 - \gamma \|x - x_0\|)^3} \quad \text{for each } x \in B(x_0, r); \quad (5.9) \]

\[ \|DF(x_0)^\dagger \|D^2 F(x)\| \leq \frac{2\gamma}{(1 - \gamma \|x - x_0\|)^3} \quad \text{for each } x \in B(x_0, r). \quad (5.10) \]

Let $L$ be the function defined by

\[ L(u) = \frac{2\gamma}{(1 - \gamma u)^3} \quad \text{for each } u \text{ with } 0 \leq u < \frac{1}{\gamma}. \quad (5.11) \]

The following proposition can be easily proved by definitions.

Proposition 5.1. Let $0 < r \leq \frac{1}{\gamma}$ be such that $B(x_0, r) \subseteq \Omega$. Then $F$ satisfies the $\gamma$-condition (resp. the modified $\gamma$-condition) on $B(x_0, r)$ if and only if $DF$ satisfies the $L$-average Lipschitz condition (resp. the modified $L$-average Lipschitz condition) on $B(x_0, r)$.

For the remainder of this subsection, let $L$ be the function defined by (5.11). Then, by (2.4), (2.5), (2.1) (with $\lambda = \kappa$) and the elementary calculation, one has that,

\[ r_\kappa = \left(1 - \sqrt{\frac{1}{2} - \kappa}\right) \frac{1}{\gamma}, \quad b_\kappa = (3 - \kappa - 2\sqrt{2 - \kappa}) \frac{1}{\gamma} \quad (5.12) \]

and

\[ h_\kappa(t) = \beta - (1 - \kappa) t + \frac{\gamma t^2}{1 - \gamma t} \quad \text{for each } t \text{ with } 0 \leq t < \frac{1}{\gamma}. \quad (5.13) \]

Moreover, if $\beta \leq b_\kappa$, then the zero of $h_\kappa$ in $[0, r_\kappa]$ is given by

\[ t_\lambda^* = \frac{(1 - \kappa) + \alpha - \sqrt{(1 - \kappa + \alpha)^2 - 4(2 - \kappa)\alpha}}{2(2 - \kappa)\gamma}. \quad (5.14) \]

In particular, in the case when $\kappa = 0$, the sequence $\{t_n\}$ generated by (2.8) for the function $h_0$ defined by (5.13) coincides with the Newton sequence for $h_0$. It is well known (see for example [21, 33, 35]) that, if $\alpha := \beta \gamma \leq 3 - 2\sqrt{2}$, $\{t_n\}$ has the closed form:

\[ t_n = \frac{1 - \xi^{2n-1}}{1 - \xi^{2n-1} \eta} t_0^* \quad \text{for each } n = 0, 1, \ldots, \quad (5.15) \]

where

\[ \xi = \frac{1 - \alpha - \sqrt{(1 + \alpha)^2 - 8\alpha}}{1 - \alpha + \sqrt{(1 + \alpha)^2 - 8\alpha}}, \quad \eta = \frac{1 + \alpha - \sqrt{(1 + \alpha)^2 - 8\alpha}}{1 + \alpha + \sqrt{(1 + \alpha)^2 - 8\alpha}} \quad (5.16) \]

and

\[ t_0^* = \frac{1 + \alpha - \sqrt{(1 + \alpha)^2 - 8\alpha}}{4\gamma}. \quad (5.17) \]

Furthermore, by [33, 35] (see also [21]), if $\alpha < 3 - 2\sqrt{2}$, then

\[ t_{n+1} - t_n \leq \xi^{2n-1} \beta \quad \text{for each } n = 0, 1, \ldots, \quad (5.18) \]
Proof. Since

Recall that \( x_0 \in \Omega \) is such that \( DF(x_0) \neq 0 \) and (3.2) holds.

**Theorem 5.3.** Suppose that

\[
\alpha := \beta \gamma \leq 3 - \kappa - 2\sqrt{2} - \kappa, \quad \mathcal{B}(x_0, t^*_\kappa) \subseteq \Omega
\]

and \( F \) satisfies the modified \( \gamma \)-condition on \( \mathcal{B}(x_0, t^*_\kappa) \). Let \( \{x_n\} \) be the sequence generated by Gauss-Newton’s method (1.4) with initial point \( x_0 \). Then \( \{x_n\} \) converges to a zero \( x^* \) of \( DF(\cdot)^t F(\cdot) \) in \( \mathcal{B}(x_0, t^*_\kappa) \) and for each \( n \geq 0 \) the following estimates hold:

\[
\|x_n - x^*\| \leq t^*_{\kappa} - t_{\kappa,n}
\]

and

\[
\|x_{n+1} - x_n\| \leq t_{\kappa,n+1} - t_{\kappa,n},
\]

where \( t_{\kappa,n} \) is defined by (2.8) (with \( \lambda = \kappa \)) for the function \( h_\kappa \) defined by (5.13).

Proof. Since \( b_0 = (3 - \kappa - 2\sqrt{2} - \kappa) \frac{1}{3} \), the condition (3.20) in Corollary 3.1 is the same as (5.20). On the other hand, by Proposition 5.1, \( DF^t F \) satisfies the modified \( \frac{L}{\gamma} \)-average Lipschitz condition on \( \mathcal{B}(x_0, t^*_\kappa) \). Therefore Corollary 3.1 is applicable to completing the proof. \( \square \)

Thus, applying Theorem 5.3 and Corollary 3.2, we immediately get the following corollary. In particular, (ii) improves [16, Theorem 3.3]. Recall that \( \xi, \eta \) are defined by (5.16) and \( t^*_0 \) by (5.17), respectively.

**Corollary 5.3.** Suppose that

\[
\alpha \leq 3 - 2\sqrt{2}, \quad \mathcal{B}(x_0, t^*_0) \subseteq \Omega
\]

and let \( \{x_n\} \) be the sequence generated by Gauss-Newton’s method (1.4) with initial point \( x_0 \). Then the following assertions hold:

(i) if \( F \) satisfies the modified \( \gamma \)-condition on \( \mathcal{B}(x_0, t^*_0) \) and if \( \kappa = 0 \), then \( \{x_n\} \) converges to a zero \( x^* \) of \( DF(\cdot)^t F(\cdot) \) in \( \mathcal{B}(x_0, t^*_0) \), and the following estimates hold for the case when \( \alpha < 3 - 2\sqrt{2} \):

\[
\|x_n - x^*\| \leq \frac{\xi^{2^n-1}(1 - \eta)}{1 - \xi^{2^n-1}\eta} t^*_0 \leq \xi^{2^n-1} t^*_0 \quad \text{for each} \ n \geq 0
\]

and

\[
\|x_{n+1} - x_n\| \leq \xi^{2^n-1}\|x_n - x_{n-1}\| \quad \text{for each} \ n \geq 1.
\]

(ii) if \( F \) satisfies the \( \gamma \)-condition on \( \mathcal{B}(x_0, t^*_0) \) and if \( DF(x_0) \) is full row rank, then \( \{x_n\} \) converges to a zero \( x^* \) of \( F \) in \( \mathcal{B}(x_0, t^*_0) \), and the estimates (5.24), (5.25) and (5.26) below hold for the case when \( \alpha < 3 - 2\sqrt{2} \):

\[
\|DF(x_0)^t F(x_n)\| \leq \xi^{2^n-1}\|DF(x_0)^t F(x_{n-1})\| \quad \text{for each} \ n \geq 1.
\]

For the function \( L \) defined (5.11), the function \( \phi_\kappa \) defined by (4.10) reduces to

\[
\phi_\kappa(t) = \frac{1}{\gamma} \left( (3 - \kappa - 2\sqrt{2} - \kappa) + (2 - \kappa)\gamma t + \kappa \left( 1 - \sqrt{\frac{1}{2 - \kappa} - \gamma t} \right) \left( \frac{1}{1 - \gamma t} - 1 \right) + \frac{2(\gamma t)^2}{1 - \gamma t} \right)
\]
for each \( t \in \left[ 0, \left( 1 - \sqrt{\frac{1}{2\gamma}} \right) \right] \). By Lemma 4.3, \( \phi_\kappa \) has exact one zero \( \tilde{r}_\kappa \) in \( \left[ 0, \left( 1 - \sqrt{\frac{1}{2\gamma}} \right) \right] \) satisfying

\[
\frac{3-\kappa-2\sqrt{2-2}\gamma}{(2-\kappa)\gamma} < \tilde{r}_\kappa < \frac{1-\sqrt{\gamma(2-\kappa)}}{\gamma}.
\]

In particular, in the case when \( \kappa = 0 \),

\[
\tilde{r}_0 = \frac{5 - 2\sqrt{2} - \sqrt{12\sqrt{2} - 15}}{\gamma} = \frac{0.0959757\ldots}{\gamma}.
\]

Recall that \( x^* \in \Omega \) such that \( F(x^*) = 0 \), \( DF(x^*) \neq 0 \) and (4.1) holds. Noting that \( r_0 = \frac{2-\sqrt{2}}{2\gamma} \), we assume that \( \mathcal{B} \left( x^*, \frac{2-\sqrt{2}}{2\gamma} \right) \subseteq \Omega \) for the remainder of this and next subsections. The following theorem is immediate from Theorem 4.1 and Proposition 5.1.

**Theorem 5.4.** Suppose that \( F \) satisfies the modified \( \gamma \)-condition on \( \mathcal{B} \left( x^*, \frac{2-\sqrt{2}}{2\gamma} \right) \). Let \( x_0 \in \mathcal{B}(x^*, \tilde{r}_\kappa) \) and let \( \{x_n\} \) be the sequence generated by Gauss-Newton’s method (1.4) with initial point \( x_0 \). Then \( \{x_n\} \) converges to a zero of \( DF(\cdot)^1F(\cdot) \).

Also the following corollary is direct, where (ii) improves [16, Theorem 4.2], which gives the estimate \( \tilde{r}_0 = (0.080851\ldots)/\gamma \) for the radius of the convergence ball. Recall that \( \tilde{r}_0 = (0.0959757\ldots)/\gamma \) is given by (5.28).

**Corollary 5.4.** Suppose that \( F \) satisfies the modified \( \gamma \)-condition (resp. the \( \gamma \)-condition) on \( \mathcal{B} \left( x^*, \frac{2-\sqrt{2}}{2\gamma} \right) \) and that \( \kappa = 0 \) (resp. \( DF(x^*) \) is full row rank). Let \( x_0 \in \mathcal{B}(x^*, \tilde{r}_0) \) and let \( \{x_n\} \) be the sequence generated by Gauss-Newton’s method (1.4) with initial point \( x_0 \). Then \( \{x_n\} \) converges to a zero of \( DF(\cdot)^1F(\cdot) \) (resp. \( F(\cdot) \)).

One typical and important class of examples satisfying the \( \gamma \)-conditions is the one of analytic functions. Following the Smale idea in [25], Shub and Smale introduced in [24] the following invariant for analytic underdetermined systems with surjective \( DF(x) \):

\[
\gamma(F, x) := \sup_{k \geq 2} \left\| DF(x)^1 \frac{D^k F(x)}{k!} \right\|^{\frac{1}{k!}}.
\]

For the case when \( DF(x) \) is not surjective, due to loss of the information on \( (\text{im}DF(x_0))^\perp \), Dedieu and Shub introduce in [6] (see also Dedieu and Kim [5]) the following invariant for analytic systems with constant rank derivatives:

\[
\gamma_M(F, x) := \sup_{k \geq 2} \left( \left\| DF(x)^1 \right\| \left\| \frac{D^k F(x)}{k!} \right\| \right)^{\frac{1}{k!}}.
\]

The following proposition, the proof of which is standard and so is omitted here (cf. [33]), shows that an analytic operator satisfies the \( \gamma \)-condition and the modified \( \gamma \)-condition, and so the conclusions of Theorem 5.3-5.4 and Corollary 5.3-5.4 hold.

**Proposition 5.2.** Let \( \gamma = \gamma(F, x_0) \) (resp. \( \gamma = \gamma_M(F, x_0) \)) and \( 0 < r \leq \frac{1}{\gamma} \) be such that \( \mathcal{B}(x_0, r) \subseteq \Omega \). Then \( F \) satisfies the \( \gamma \)-condition (the modified \( \gamma \)-condition) on \( \mathcal{B}(x_0, r) \).

### 5.3 Extension of the Smale approximate zeros

We first recall the notion of the approximate zero of an analytic mapping \( F \) from the domain \( \Omega \) in a Banach space to another. The following unified definition is taken from [33]. Consider Newton’s iteration with initial point \( x_0 \):

\[
x_{n+1} = x_n - DF(x_n)^{-1}F(x_n) \quad \text{for each } n = 0, 1, 2, \ldots.
\]
Definition 5.2. Suppose $x_0 \in \Omega$ is such that Newton’s iteration (5.29) is well-defined for $F$ and satisfies

$$e(x_n) \leq \left(\frac{1}{2}\right)^{2n-1} e(x_{n-1}) \quad \text{for all } n = 1, 2, \ldots,$$

(5.30)

where $e(x_n)$ denotes some measurement of the approximation degree between $x_n$ and the zero point $x^*$. Then $x_0$ is said to be an approximate zero of $F$ in the sense of $e(x_n)$.

The notions of approximate zeroes in the sense of $\|x_{n+1} - x_n\|$ and in the sense of $\|x_n - x^*\|$ were introduced in [25], and a more reasonable definition for the second kind was presented in [26] (see also [2]), which was also studied by Wang in [39]. The notion of the approximate zero in the sense of $\|DF(x_0)^{-1}F(x_n)\|$ was defined in [4], which, as shown in [33], is equivalent to that in the sense of $\|x_{n+1} - x_n\|$.

We now extend the notion of approximate zeroes to the Gauss-Newton method. Throughout this subsection, we assume that $F$ is analytic on $\Omega$ and that (3.21) holds, that is, $\kappa = 0$.

Definition 5.3. Suppose that $x_0 \in \Omega$ is such that the sequence $\{x_n\}$ generated by Gauss-Newton’s method (1.4) converges to a zero $x^*$ of $DF(\cdot)^TF(\cdot)$ (resp. $F$) and satisfies (5.30). Then $x_0$ is said to be a GNM-approximate solution (resp. approximate zero) of $F$ in the sense of $e(x_n)$.

Let $t_0^*$ be defined by (5.17). Our first result concerning on the rule to judge $x_0$ to be a GNM-approximate solution or an approximate zero is as follows.

Theorem 5.5. Let $\gamma > 0$. Suppose that

$$\alpha = \beta \gamma \leq \frac{13 - 3\sqrt{17}}{4} \quad \text{and} \quad B(x_0, t_0^*) \subseteq \Omega.$$  

(5.31)

Then the following assertions hold:

(i) If $F$ satisfies the modified $\gamma$-condition on $B(x_0, t_0^*)$, then $x_0$ is a GNM-approximate solution of $DF(\cdot)^TF(\cdot)$ in the sense of $\|x_{n+1} - x_n\|$.

(ii) If $F$ satisfies the $\gamma$-condition on $B(x_0, t_0^*)$ and $DF(x_0)$ is full row rank, then $x_0$ is an approximate zero of $F$ in the sense of $\|x_{n+1} - x_n\|$ and $\|DF(x_0)^TF(x_n)\|$.

Proof. We only prove the assertion (i) because the proof of the assertion (ii) is almost the same. To show (i), we apply Corollary 5.3 (as $\alpha \leq \frac{13 - 3\sqrt{17}}{4} < 3 - 2\sqrt{2}$ by (5.31)) to get that $\{x_n\}$ converges to a zero $x^*$ of $DF(\cdot)^TF(\cdot)$ in $B(x_0, t_0^*)$ and (5.25) holds with $\xi$ given by $\xi = \xi(\alpha)$, where $\xi(\cdot)$ is defined by

$$\xi(t) = \frac{1 - t - \sqrt{(1 + t)^2 - 8t}}{1 - t + \sqrt{(1 + t)^2 - 8t}} \quad \text{for each } t \in [0, \frac{13 - 3\sqrt{17}}{4}].$$

Since $\xi(\cdot)$ increases as $t$ does on $[0, \frac{13 - 3\sqrt{17}}{4}]$ and the value of $\xi(\cdot)$ at $\alpha = \frac{13 - 3\sqrt{17}}{4}$ is $\frac{1}{2}$, we have that $\xi \leq \frac{1}{2}$. This together with (5.25) implies that $x_0$ is a GNM-approximate solution of $DF(\cdot)^TF(\cdot)$ in the sense of $\|x_{n+1} - x_n\|$ and completes the proof.

By Proposition 5.2, the following corollary is direct. In particular the assertion (ii) improves [16, Corollary 5.3] and so [24, Theorem 1.4].
Corollary 5.5. Suppose that (5.31) holds. Then the following assertions hold:

(i) If \( \gamma = \gamma_M(F,x_0) \), then \( x_0 \) is a GNM-approximate solution of \( DF(\cdot)^TF(\cdot) \) in the sense of \( \|x_{n+1} - x_n\| \).

(ii) If \( \gamma = \gamma(F,x_0) \) and \( DF(x_0) \) is full row rank, then \( x_0 \) is an approximate zero of \( F \) in the sense of \( \|x_{n+1} - x_n\| \) and \( \|DF(x_0)^TF(x_n)\| \).

Recall that \( x^* \in \Omega \) such that \( F(x^*) = 0, DF(x^*) \neq 0 \) and (4.1) holds. Let \( \psi \) be the function defined by \( \psi(t) = 1 - 4t + 2t^2 \) for each \( t \in [0, \frac{2-\sqrt{2}}{2\gamma}] \). Let \( t_0 = 0.0858167 \cdots \) be the smallest positive root of the equation

\[
\frac{t - 2t^2}{\psi(t)^2} = \frac{13 - 3\sqrt{17}}{4}.
\]

Corollary 5.6. Let \( \gamma = \gamma_M(F,x^*) \) (resp. \( \gamma = \gamma(F,x^*) \) and \( DF(x^*) \) is full row rank), and let \( x_0 \in B(x^*, t_0/\gamma) \). Suppose that \( B(x^*, \frac{2-\sqrt{2}}{2\gamma}) \subseteq \Omega \). Then the assertions (i) (resp. (ii)) in Corollary 5.5 hold.

Proof. We only prove the assertion (i) in Corollary 5.5. For this purpose, we write \( \bar{u} = \gamma \|x_0 - x^*\| \) and \( \bar{\gamma} = \frac{1}{\sqrt{\alpha_0(1-\alpha_0)}} \). By Propositions 5.1 and 5.2, \( DF \) satisfies the modified \( L \)-average Lipschitz condition on \( B(x^*, \frac{2-\sqrt{2}}{2\gamma}) \) with \( L \) defined by (5.11). Thus Lemmas 4.1 and 4.2 are applicable. Hence

\[
\beta \leq \frac{(1 - \bar{u})(1 - 2\bar{u}) \bar{u}}{\gamma}
\]

thanks to (4.2) and (5.11). Consequently,

\[
\bar{\alpha} := \beta \bar{\gamma} \leq \frac{\bar{u} - 2\bar{u}^2}{\psi(\bar{u})^2} \leq \frac{t_0 - 2t_0^2}{\psi(t_0)^2} = \frac{13 - 3\sqrt{17}}{4},
\]

because \( \bar{u} \leq t_0 \) and the function \( t \mapsto \frac{t - 2t^2}{\psi(t)^2} \) is increasing on \( [0, \frac{2-\sqrt{2}}{2\gamma}] \).

Moreover, since

\[
\bar{t}_0^* := 1 + \bar{\alpha} - \sqrt{(1 + \bar{\alpha})^2 - 8\alpha} \leq \frac{2 - \sqrt{2}}{2\bar{\gamma}} \leq \frac{2 - \sqrt{2}}{2\gamma} - \|x_0 - x^*\|,
\]

it follows from Lemma 4.2 that \( DF \) satisfies the modified \( L \)-average Lipschitz condition on \( B(x_0, \bar{t}_0^*) \) with \( \bar{L} \) defined by (4.6). By (4.6), for each \( u \in [0, \bar{t}_0^*], \)

\[
\bar{L}(u) = \frac{(1 - \bar{u})^2}{\psi(\bar{u})} \cdot \frac{2\gamma}{(1 - \bar{u} - \gamma u)^\gamma} \leq \frac{2\gamma}{(1 - \bar{u} - \gamma u)^\gamma} \leq \frac{2\gamma}{(1 - \bar{\gamma} u)^\gamma},
\]

because \( 0 < \psi(t) < 1 \) for all \( t \in (0, \frac{2-\sqrt{2}}{2\gamma}) \). This together with Proposition 5.1 implies that \( DF \) satisfies \( \bar{\gamma} \)-condition on \( B(x_0, \bar{t}_0^*) \). Thus, Corollary 5.5 is applicable to concluding that \( x_0 \) is a GNM-approximate solution of \( DF(\cdot)^TF(\cdot) \) in the sense of \( \|x_{n+1} - x_n\| \) and the proof is complete.

Part (b) in the following definition is taken from \([24]\).

Definition 5.4. Let \( y \in \mathbb{R}^l \), \( y \) is called a

(a) quasi-regular value if \( DF(x^*) \neq 0 \) and (4.1) holds for each \( x^* \in F^{-1}(y) \);

(b) regular value if \( DF(x^*) \) is full row rank for each \( x^* \in F^{-1}(y) \).
The following corollary, which is a direct consequence of Corollary 5.6, extends and improves [16, Colollary 5.5] (where \( t_0 = 0.0776121 \cdots \)) and so [24, Theorem 1.7].

**Corollary 5.7.** Suppose that \( F \) has \( 0 \) as a quasi-regular value (resp. a regular value). Let \( \gamma = \sup_{x^* \in F^{-1}(0)} \gamma_M(F,x^*) \) (resp. \( \gamma = \sup_{x^* \in F^{-1}(0)} \gamma(F,x^*) \)). Suppose that
\[
d(x_0,F^{-1}(0)) < \frac{t_0}{\gamma} \quad \text{and} \quad \bigcup_{x^* \in F^{-1}(0)} B \left( x^*, \frac{2 - \sqrt{2}}{2\gamma} \right) \subseteq \Omega.
\]
Then the assertion (i) (resp. (ii)) in Corollary 5.5 holds.

**5.4 Further applications to analytic systems and examples**

Let \( \gamma_n > 0 \) for \( n = 2, 3, 4, \ldots \) and \( F \) be an analytic operator. Wang and Zhao introduced in [29] the following condition to study the Smale point estimate theory for Newton’s method (assuming \( D F(x_0) \) is invertible):
\[
\|D F(x_0)^\dagger D^n F(x_0)\| \leq \gamma_n \quad \text{for each} \quad n \geq 2.
\]
This condition was used in [14] again by J.M. Gutiérrez et al to analyze the convergence of Moser’s method. As before, in the case when \( D F(x_0) \) is not full row rank, we need the following modified condition:
\[
\|D F(x_0)^\dagger\| \|D^n F(x_0)\| \leq \gamma_n \quad \text{for each} \quad n \geq 2.
\]

Set
\[
R = \lim_{n \to \infty} \sqrt[n]{\frac{\gamma_n}{n}},
\]
where we adopt the conventions that \( 1^0 = +\infty \) and \( 1^{+\infty} = 0 \). Then we have the following proposition, the proof of which is easy and so is omitted here.

**Proposition 5.3.** Suppose that condition (5.32) (resp. (5.33)) holds. Let \( 0 < r \leq R \) be such that \( B(x_0,r) \subseteq \Omega \). Then \( D F \) satisfies the L-average Lipschitz condition (resp. the modified L-average Lipschitz condition) on \( B(x_0,r) \) with \( L \) defined by
\[
L(u) = \sum_{n=2}^{\infty} \frac{\gamma_n}{(n-2)!} u^{n-2} \quad \text{for each} \quad u \text{ with } 0 \leq u < R.
\]
Let \( L \) be the function defined by (5.35). Then
\[
\Delta = \int_0^R L(u) du = \sum_{n=2}^{\infty} \frac{\gamma_n}{(n-1)!} R^{n-1},
\]
and there exists \( t_\kappa \in [0, R] \) such that \( \sum_{n=2}^{\infty} \frac{\gamma_n}{(n-1)!} (t_\kappa)^{n-1} = 1 - \kappa \) if \( \Delta \geq 1 - \kappa \). By (2.6), (2.5) and (2.1)(with \( \lambda = \kappa \)), one has that
\[
r_\kappa = \left\{ \begin{array}{ll}
t_\kappa & \text{if } \sum_{n=2}^{\infty} \frac{\gamma_n}{(n-1)!} R^{n-1} \geq 1 - \kappa, \\
R & \text{if } \sum_{n=2}^{\infty} \frac{\gamma_n}{(n-1)!} R^{n-1} < 1 - \kappa,
\end{array} \right.
\]
\[
b_\kappa = (1 - \kappa) r_\kappa - \sum_{n=2}^{\infty} \frac{\gamma_n}{n!} (r_\kappa)^n.
\]
and the corresponding majorizing function is
\[ h_\kappa(t) = \beta - (1 - \kappa)t + \sum_{n=2}^{\infty} \frac{\gamma_n}{n!} t^n \quad \text{for each } t \in [0, R]. \tag{5.38} \]

In the case when \( DF(x_0) \) is invertible, Wang and Zhao studied in [29] the Smale point estimate theory for Newton’s method under the strong assumption that \( R = +\infty \). By Proposition 5.3, our theorems obtained in sections 3 and 4 are applicable to establishing the corresponding results. Here we don’t intend to restate every theorem again but, as an example, only the following theorem, which extends [29, Theorem 2.3].

**Theorem 5.6.** Let \( x_0 \in \Omega \) be such that \( DF(x_0) \) is full row rank. Suppose that (5.32) holds and
\[ \beta := \|DF(x_0)^T F(x_0)\| \leq r_0 - \sum_{n=2}^{\infty} \frac{\gamma_n}{n!} (r_0)^n, \quad B(x_0, t_0^*) \subseteq \Omega, \tag{5.39} \]
where \( t_0^* \) is the unique zero of \( h_0 \) in \([0, r_0]\). Let \( \{x_n\} \) be the sequence generated by Gauss-Newton’s method (1.4) with initial point \( x_0 \). Then \( \{x_n\} \) converges to a zero \( x^* \) of \( F(\cdot) \) in \( B(x_0, t_0^*) \), and the estimates (3.23)-(3.25) and (3.27) hold.

As we have seen, to apply our results of the present paper, it is the key to determine the values of parameters \( r_\kappa, b_\kappa \) and \( \hat{r}_\kappa \). Usually, it is very technical to determine the bounds \( \gamma_n \) for \( \|DF(x_0)^T D^n F(x_0)\| \) or \( \|DF(x_0)^T\|D^n F(x_0)\| \) such that the values of these parameters can be figured out. Below we consider some special and important examples of \( \{\gamma_n\} \), which were used in [33] to extend the Smale point estimate theory.

**Example 5.1.** Let \( \gamma, c \in (0, +\infty) \) and \( m \in (-1, 0) \cup (0, +\infty) \). Consider the sequences \( \{\gamma_n\} \) respectively defined as follows:

<table>
<thead>
<tr>
<th>Type</th>
<th>( \gamma_n )</th>
<th>( \gamma_r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential type</td>
<td>( c \gamma^{n-1} )</td>
<td>( (c+1-\kappa) \ln \frac{c+1-\kappa}{c} -(1-\kappa) )</td>
</tr>
<tr>
<td>Binomial type</td>
<td>( c \frac{m+n-1}{m+1} \gamma^{n-1} )</td>
<td>( 1 - \kappa + c \frac{m+n}{m+1} \left( 1 - \frac{c+1-\kappa}{c} \right) )</td>
</tr>
<tr>
<td>The first logarithmic type</td>
<td>( c(n-1)! \gamma^{n-1} )</td>
<td>( 1 - \kappa - c \ln \frac{c+1-\kappa}{c} )</td>
</tr>
<tr>
<td>The second logarithmic type</td>
<td>( c(n-2)! \gamma^{n-1} )</td>
<td>( 1 - \kappa - c \left( 1 - e^{-\frac{c+1-\kappa}{c}} \right) )</td>
</tr>
</tbody>
</table>

where \( \frac{m+n-1}{m+1} \gamma^{n-1} = (m+1)(m+2) \cdots (m+n-1) \gamma^{n-1} \). For the general case, the parameters \( r_\kappa \) and \( b_\kappa \) have explicit forms which are given in Table 1, but not the parameter \( \hat{r}_\kappa \). For the special case when \( c = 1, m = 1, -\frac{1}{2}, \frac{1}{2} \) and \( \kappa = 0, \frac{1}{2} \), the corresponding values of the parameters \( r_\kappa, b_\kappa \) are given in Table 2.
5.5 Concluding remark

We used in the present paper the notions of the Lipschitz conditions with $L$ average to analyze the convergence behavior of Gauss-Newton's method for the singular systems satisfying (1.10), but without assumption of (1.3) for the involved function $L$. As it has been seen, the lack of the assumption (1.3) makes the study more complicated and the consideration seems original, in particular for the case when $\kappa \neq 0$. Our main results obtained in the present paper give unified convergence criteria and unified estimates for the radii of convergence balls of Gauss-Newton’s method. Applications to the cases of the Kantorovich type condition, the $\gamma$-condition and the Smale point estimate theory as well as some more general analytic systems are provided. When these results are applied to the underdetermined systems with surjective derivatives, some known results are extended and/or improved as noted in the introduction section. Below, we provide two examples to illustrate the applicability of our results. The first one is concerned with the case when $\kappa \neq 0$ and the second one with the case when $\kappa = 0$ but each derivative $DF(x)$ is not of full row rank. Hence, the results in [16, 24, 29] are not applicable.

Example 5.2. Let $\mathbb{R}^2$ be endowed with the $l_1$-norm. Consider the operator $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by

$$F(x) := (\sin(\xi_1 - \xi_2), \cos(\xi_1 - \xi_2) - 1)^T$$

for each $x = (\xi_1, \xi_2)^T \in \mathbb{R}^2$.

Then $F$ is analytic on $\mathbb{R}^2$, and

$$DF(x) = \begin{pmatrix} \cos(\xi_1 - \xi_2) & -\cos(\xi_1 - \xi_2) \\ \sin(\xi_1 - \xi_2) & \sin(\xi_1 - \xi_2) \end{pmatrix}$$

for each $x = (\xi_1, \xi_2)^T \in \mathbb{R}^2$.

Let $x = (\xi_1, \xi_2)^T \in \mathbb{R}^2$. Then $\text{rank}(DF(x)) = 1$ and the Moore-Penrose inverse is

$$DF(x)^\dagger = \frac{1}{2} \begin{pmatrix} \cos(\xi_1 - \xi_2) & -\sin(\xi_1 - \xi_2) \\ \cos(\xi_1 - \xi_2) & \sin(\xi_1 - \xi_2) \end{pmatrix}.$$ 

Furthermore, by mathematical induction, we can easily get that

$$D^k F(x)u_1 u_2 \cdots u_k = \begin{pmatrix} \sin(\xi_1 - \xi_2 + \frac{k\pi}{2}) \\ \cos(\xi_1 - \xi_2 + \frac{k\pi}{2}) \end{pmatrix} \prod_{i=1}^{k} (u_i^1 - u_i^2)$$

for each $k = 1, 2, \ldots$. 


<table>
<thead>
<tr>
<th>$\gamma_n$</th>
<th>$\kappa$</th>
<th>$\gamma r_\kappa$</th>
<th>$\gamma b_\kappa$</th>
<th>$\gamma r_\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma^{n-1}$</td>
<td>0</td>
<td>0.69314...</td>
<td>0.38629...</td>
<td>0.21897...</td>
</tr>
<tr>
<td>$n! \gamma^{n-1}$</td>
<td>0.5</td>
<td>0.40546...</td>
<td>0.10819...</td>
<td>0.08697...</td>
</tr>
<tr>
<td>$(2n-1)!! \gamma^{n-1}$</td>
<td>0.5</td>
<td>0.29285...</td>
<td>0.17157...</td>
<td>0.09597...</td>
</tr>
<tr>
<td>$(2n+1)!! \gamma^{n-1}$</td>
<td>0.5</td>
<td>0.18350...</td>
<td>0.05051...</td>
<td>0.03995...</td>
</tr>
<tr>
<td>$(n - 1)! \gamma^{n-1}$</td>
<td>0.5</td>
<td>0.69314...</td>
<td>0.22023...</td>
<td>0.12270...</td>
</tr>
<tr>
<td>$(n - 2)! \gamma^{n-1}$</td>
<td>0.5</td>
<td>0.23658...</td>
<td>0.16558...</td>
<td>0.09408...</td>
</tr>
</tbody>
</table>
where \( u_i = (u_{1i}, u_{2i}) \in \mathbb{R}^2 \) for each \( i = 1, 2, \ldots, k \). Hence
\[
\|DF(x)^T\| = \max \{ |\cos(\xi_1 - \xi_2)|, |\sin(\xi_1 - \xi_2)| \}
\]
and
\[
\|D^k F(x)\| = |\cos(\xi_1 - \xi_2)| + |\sin(\xi_1 - \xi_2)|.
\]

Consequently,
\[
\gamma(F, x) = \sup_{k > 1} \left( \|DF(x)^T\| \left\| \frac{D^k F(x)}{\|x\|} \right\| \right)^{1/k}
\]
\[
= \frac{1}{2} \max \{ |\cos(\xi_1 - \xi_2)|, |\sin(\xi_1 - \xi_2)| \}.
\]

Let \( \Omega = \{(\xi_1, \xi_2)^T : -\frac{\pi}{2} < \xi_1 < \frac{\pi}{2}, i = 1, 2, \} \subseteq \mathbb{R}^2 \). Noting that \( 0 \leq 1 - \cos(\xi_1 - \xi_2) \leq \frac{7}{10} \) for any \((\xi_1, \xi_2)^T \in \Omega\), one has that
\[
\|DF(y)^T (I - DF(x) DF(x)^T) F(x)\| = |(1 - \cos(\xi_1 - \xi_2)) \sin(\xi_1 - \xi_2 - (\xi_1 - \xi_2))| \leq \frac{7}{10} \|x - y\|
\]
holds for any \( y = (\xi_1, \xi_2)^T \in \Omega \), that is, (3.1) holds with \( \kappa = \frac{7}{10} \). Now we take the initial point \( x_0 = (\frac{130}{32}, 0)^T \). Then
\[
\beta(F, x_0) = \|DF(x_0)^T F(x_0)\| = \sin \frac{1}{32}
\]
and, thanks to (5.40),
\[
\gamma(F, x_0) = \frac{1}{2} \cos \frac{1}{32} \left( \sin \frac{1}{32} + \cos \frac{1}{32} \right).
\]

It follows that
\[
\alpha := \beta(F, x_0) \gamma(F, x_0) = \frac{1}{4} \sin \frac{1}{16} \left( \sin \frac{1}{32} + \cos \frac{1}{32} \right) < \frac{23 - 2\sqrt{130}}{10} = 3 - 2\sqrt{2} - \kappa
\]
and by (5.14)
\[
t^n_{\kappa} < \frac{1 - \kappa + \alpha}{2(2 - \kappa)^2} = \frac{6 + 5 \sin \frac{1}{16} (\sin \frac{1}{32} + \cos \frac{1}{32})}{26 \cos \frac{1}{32} (\sin \frac{1}{32} + \cos \frac{1}{32})} < \frac{1}{2}.
\]

Therefore \( B(x_0, t^n_{\kappa}) \subseteq \Omega \). Then, by Proposition 5.2, Theorem 5.3 is applicable to concluding that the sequence \( \{x_n\} \) generated by Gauss-Newton’s method (1.4) with initial point \( x_0 \) converges to a zero of \( DF(x)^T F(x) \) in \( B(x_0, t^n_{\kappa}) \).

Furthermore, we consider the point \( x^* = (0, 0)^T \). Then \( x^* \in \Omega \) satisfies \( F(x^*) = 0 \) and \( \gamma := \gamma(F, x^*) = \frac{1}{2} \) by (5.40). Thus, noting that \( \kappa = \frac{7}{10} \), the function \( \phi_\kappa \) defined by (5.27) reduces to
\[
\phi_\kappa(t) = 2 \left( \frac{23 - 2\sqrt{130}}{10} \frac{13t}{20} + \frac{7}{10} \left( \frac{13 - \sqrt{130}}{13} - \frac{t}{2} \right) \left( \frac{1}{1 - \frac{t}{2}} - 1 \right) \right)
\]
for each \( t \in \left[ \frac{26 - 2\sqrt{130}}{13}, \frac{26 - 2\sqrt{130}}{13} \right] \). Then \( \phi_\kappa \) has exactly one zero \( \hat{t}_\kappa = 0.03711716 \cdots \in \left[ \frac{26 - 2\sqrt{130}}{13}, \frac{26 - 2\sqrt{130}}{13} \right] \). Since \( \frac{2\sqrt{2}}{\pi} = 2 - \sqrt{2} < \frac{1}{2} \), we have \( \overline{B}(x^*, \frac{2\sqrt{2}}{\pi}) \subseteq \Omega \). Thus, by Proposition 5.2, Theorem 5.4 is applicable to concluding that for any \( x_0 \in \overline{B}(x^*, \hat{t}_\kappa) \), the sequence \( \{x_n\} \) generated by Gauss-Newton’s method (1.4) with initial point \( x_0 \) converges to a zero of \( DF(x)^T F(x) \).
Example 5.3. Let $\mathbb{R}^2$ be endowed with the $l_1$-norm and $\Omega = \mathbb{R}^2$. Let $\tau \in \mathbb{R}$ and let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$F(x) := (\sin(\xi_1 + \xi_2), \sin(\xi_1 + \xi_2) - \tau)^T$$

for each $x = (\xi_1, \xi_2)^T \in \mathbb{R}^2$.

Then $F$ is analytic on $\mathbb{R}^2$, and

$$DF(x) = \cos(\xi_1 + \xi_2) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

for each $x = (\xi_1, \xi_2)^T \in \mathbb{R}^2$.

Let $x = (\xi_1, \xi_2)^T \in \mathbb{R}^2$ with $\xi_1 + \xi_2 \neq \frac{\pi}{2} + k\pi, k \in \mathbb{N}$. Then $\text{rank}(DF(x)) = 1$ and the Moore-Penrose inverse is

$$DF(x)^\dagger = \frac{1}{4\cos(\xi_1 + \xi_2)} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$  

Therefore,

$$\beta(F, x) = \|DF(x)^\dagger F(x)\| = \frac{|\sin(\xi_1 + \xi_2) - \frac{\pi}{2}|}{|\cos(\xi_1 + \xi_2)|}.$$  

(5.43)

Clearly,

$$DF(y)^\dagger \left( I - DF(x) DF(x)^\dagger \right) F(x) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

for each $y = (\zeta_1, \zeta_2)^T$.

Hence (3.21) holds on $\mathbb{R}^2$. Moreover, by mathematical induction, we obtain that

$$D^k F(x)u_1 u_2 \cdots u_k = \begin{pmatrix} \sin(\xi_1 + \xi_2 + \frac{k\pi}{2}) \\ \sin(\xi_1 + \xi_2 + \frac{k\pi}{2}) \end{pmatrix} \prod_{i=1}^k (u_i^1 + u_i^2),$$

where $u_i = (u_i^1, u_i^2) \in \mathbb{R}$ for each $i = 1, 2, \cdots, k$. Therefore

$$\|DF(x)^\dagger\| = \frac{1}{2|\cos(\xi_1 + \xi_2)|} \quad \text{and} \quad \|D^k F(x)\| = 2|\sin(\xi_1 + \xi_2 + \frac{k\pi}{2})|.$$  

Then

$$\gamma_M(F, x) = \sup_{k > 1} \left( \|DF(x)^\dagger\| \|D^k F(x)\| \right)^{\frac{1}{k+1}} = \max \left\{ \frac{\sqrt{6}}{6}, \sup_{k \geq 1} \left( \frac{|\tan(\xi_1 + \xi_2)|}{(2k)!} \right)^{\frac{1}{k+1}} \right\}.  \tag{5.44}$$

Let $\tau = 1$. Take the initial point $x_0 = (\frac{\pi}{6}, \frac{5\pi}{48})^T$. Then

$$\gamma = \gamma_M(F, x_0) = \frac{1}{2}\tan\frac{13\pi}{48} \quad \text{and} \quad \beta = \beta(F, x_0) = \sec\frac{13\pi}{48}\left(\sin\frac{13\pi}{48} - \frac{1}{2}\right);$$

hence

$$\alpha = \beta\gamma = \frac{1}{2}\tan\frac{13\pi}{48}\frac{1}{\cos\frac{13\pi}{48}}(\sin\frac{13\pi}{48} - \frac{1}{2}) < 3 - 2\sqrt{2}.$$  

Noting that the inclusion $\overline{B(x_0, t_0/\alpha)} \subseteq \Omega$ is trivial, we have by Proposition 5.2 that Corollary 5.3 (i) is applicable to concluding that the sequence $\{x_n\}$ generated by Gauss-Newton’s method (1.4) with initial point $x_0$ converges to a zero $x^*$ of $DF(\cdot)^\dagger F(\cdot)$ in $\overline{B(x_0, t_0)}$.

Furthermore, take $x_0 = (\frac{\pi}{6}, \frac{\pi}{24})$. By (5.44) and (5.43), we have

$$\gamma = \gamma_M(F, x_0) = \frac{\sqrt{6}}{6} \quad \text{and} \quad \beta = \beta(F, x_0) = \frac{5\pi}{24}\left(\sin\frac{5\pi}{24} - \frac{1}{2}\right).$$
It follows that
\[ \alpha = \beta = \gamma = \sqrt{\frac{6}{6} \sec \frac{5\pi}{24} \left( \sin \frac{5\pi}{24} - \frac{1}{2} \right)} = \frac{13 - 3\sqrt{17}}{4}. \]

Thus, Corollary 5.5 (i) is applicable to concluding that \( x_0 \) is a GNM-approximate solution of \( DF(\cdot)\dagger F(\cdot) \) in the sense of \( \|x_{n+1} - x_n\| \).

Finally, let \( \tau = 0 \) and consider \( x^* = (0, 0)^T \). Then \( F(x^*) = 0 \). By (5.44), \( \gamma = \gamma_M(F, x^*) = \sqrt{\frac{2}{6}} \) and so
\[ \hat{r}_0 = \frac{0.0959757 \cdots}{\gamma} = 0.2350915 \cdots. \]

Thus Corollary 5.4 is applicable to concluding that for any \( x_0 \in B(x^*, \hat{r}_0) \) with \( \hat{r}_0 = 0.2350915 \cdots \), the sequence \( \{x_n\} \) generated by Gauss-Newton's method (1.4) with initial point \( x_0 \) converges to a zero of \( DF(\cdot)\dagger F(\cdot) \).

Furthermore, it is easy to see that
\[ F^{-1}(0) = \{ (\xi_1, \xi_2) | \xi_1 + \xi_2 = k\pi, k = 0, 1, \cdots \} \]
and that \( F \) has 0 as a quasi-regular value. Write \( \gamma = \sup_{x^* \in F^{-1}(0)} \gamma_M(F, x^*). \) Then by (5.44) we get that \( \gamma := \sqrt{\frac{6}{6}}. \) Therefore Corollary 5.7 is applicable to concluding that if \( d(x_0, F^{-1}(0)) < \frac{0.0858167 \cdots}{\gamma} \) = 0.2102071 \cdots, then \( x_0 \) is a GNM-approximate solution of \( DF(\cdot)\dagger F(\cdot) \) in the sense of \( \|x_{n+1} - x_n\| \).

References


