NEWTON'S METHOD FOR ANALYTIC SYSTEMS OF EQUATIONS WITH
CONSTANT RANK DERIVATIVES.

Jean-Pierre Dedieu
MIP
Université Paul Sabatier
31062 Toulouse Cedex 04, France.
dedieu@cict.fr

Myong-Hi Kim
Department of Mathematics
State University of New York
College at Old Westbury, NY 11568
myonghi@math.sunysb.edu

ABSTRACT. In this paper we study the convergence properties of Newton's sequence for analytic systems of equations with constant rank derivatives. Our main result is an alpha-theorem which insures the convergence of Newton's sequence to a least-square solution of this system.

Subject Classification: 65H10.
Key words: Newton's Method, System of Equations, Least-Square Solution.
1This paper was completed when Jean-Pierre Dedieu was visiting IBM T. J. Watson Research Center in March 1999.
1. Introduction.

Newton's method is a classical numerical method to solve a system of nonlinear equations

$$f : \mathbf{E} \rightarrow \mathbf{F}$$

with E and F two Euclidean spaces or more generally two Banach spaces. If $x \in \mathbf{E}$ is an approximation of a zero of this system then, Newton's method updates this approximation by linearizing the equation $f(y) = 0$ around $x$ so that

$$f(x) + Df(x)(y - x) = 0.$$  

When $Df(x)$ is an isomorphism we obtain the classical Newton's iterate

$$y = N_f(x) = x - Df(x)^{-1}f(x).$$

When E and F are two Euclidean spaces and when $Df(x)$ is not an isomorphism we choose its Moore-Penrose inverse $Df(x)^\dagger$ instead of its classical inverse:

$$y = N_f(x) = x - Df(x)^\dagger f(x).$$

We recall that the Moore-Penrose inverse of a linear operator

$$A : \mathbf{E} \rightarrow \mathbf{F}$$

is the composition of two maps: $A^\dagger = B \circ \Pi_{\text{Im } A}$ where $\Pi_{\text{Im } A}$ is the orthogonal projection in F onto $\text{Im } A$ and $B$ is the right inverse of $A$ whose image is the orthogonal complement of Ker $A$ in E i.e. the inverse of the restriction

$$A|_{(\text{Ker } A)^\perp} : (\text{Ker } A)^\perp \rightarrow \text{Im } A.$$

We have $A^\dagger = (A^*A)^{-1}A^*$ when $A$ is injective, $A^\dagger = A^*(AA^*)^{-1}$ when $A$ is surjective, where $A^*$ denotes the adjoint of $A$. Notice that $A^\dagger A = \Pi_{(\text{Ker } A)^\perp}$ and $AA^\dagger = \Pi_{\text{Im } A}$.

For underdetermined systems, when $Df(x)$ is surjective, $Df(x)^\dagger$ is injective in F and hence the zeros of $f(x)$ corresponds to the fixed points of the Newton operator

$$N_f(x) = x - Df(x)^\dagger f(x).$$

The case of overdetermined systems is completely different. This iteration has been introduced for the first time by Gauss in 1809 [6] and, for this reason, it is called Newton-Gauss iteration. When $Df(x)$ is injective, the fixed points of $N_f(x)$ do not necessarily correspond to the zeros of $f$ but to the least-square solutions of $f(x) = 0$, i.e. to the stationary points of $F(x) = \|f(x)\|^2$. In other words $N_f(x) = x$ if and only if $D(\|f(x)\|^2) = 0$.

In this paper, our aim is to study the properties of Newton’s iteration for analytic systems of equations with constant rank derivatives. This case generalizes both the underdetermined case (Rank $Df(x) = \text{Dim } \mathbf{F}$) and the overdetermined case of (Rank $Df(x) = \text{Dim } \mathbf{E}$). It has been considered for the first time by Ben-Israel [2].

We consider an analytic function $f : \mathbf{E} \rightarrow \mathbf{F}$ between two Euclidean spaces. We let $n = \text{Dim } \mathbf{E}$ and $m = \text{Dim } \mathbf{F}$. We also consider the case of a function $f$ defined in an open set $U \subset \mathbf{E}$ but by abuse of notation we continue to write $f : \mathbf{E} \rightarrow \mathbf{F}$.
As in the injective-overdetermined case, the fixed points of Newton’s operator do not necessarily correspond to the zeros of $f$ but to the least square solutions of this system:

**Proposition 1.** The following statements are equivalent:

1. $N_f(x) = x$,
2. $Df(x)^{\dagger}f(x) = 0$,
3. $Df(x)^*f(x) = 0$,
4. $f(x) \in \text{Im} \ Df(x)^{\perp}$,
5. $DF(x) = 0$ with $F(x) = \|f(x)\|^2$.

The proof is easy and left to the reader. □

There are two points of view to analyze the convergence properties for Newton’s method: Kantorovich like theorems and Smale’s alpha-theory. Let $x \in \mathbb{E}$ be given. Under which hypothesis does the sequence

$$x_{k+1} = N_f(x_k), \ x_0 = x,$$

converges to a zero $\xi$ of $f$?

Kantorovich gives an answer in terms of the behavior of $f$ in a neighborhood of $x$ with a weak regularity assumption, say $f$ is $C^2$. See Ostrowski [12] or Ortega-Rheinboldt [11].

Alpha-theory, which was introduced by Kim in [8], [9] for one variable polynomial equations and by Smale for general systems of equations in [18], gives an answer in terms of three invariants.

$$\alpha(f,x) = \beta(f,x)\gamma(f,x)$$

$$\beta(f,x) = \|Df(x)^{-1}f(x)\|$$

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

which only depend on the derivatives $D^k f(x)$ at the given starting point $x$. Here a stronger regularity assumption is made: $f$ is an analytic system of equations.

The main feature of Newton’s iteration is its quadratic convergence to the zeros of $f$. Alpha-theory gives the size of the basin of attraction around these zeros in terms of the invariant $\gamma(f,x)$. We have:

**Theorem 1.** (Smale) When $\xi$ is a zero of $f$ and $Df(\xi)$ is an isomorphism then, for any $x \in \mathbb{E}$ satisfying

$$\|x - \xi\|\gamma(f,\xi) \leq \frac{3 - \sqrt{7}}{2},$$

1. the sequence $x_{k+1} = N_f(x_k), \ x_0 = x$ is well defined,
2. for any $k \geq 0$,

$$\|x_k - \xi\| \leq \left(\frac{1}{2}\right)^{2k-1} \beta(f, x).$$
This theorem is extended by Shub and Smale in [14] to the case of underdetermined systems of equations with surjective derivatives. They introduce the following invariants,

\[
\alpha(f, x) = \beta(f, x) \gamma(f, x)
\]

\[
\beta(f, x) = \|Df(x)\| \|f(x)\|
\]

\[
\gamma(f, x) = \sup_{k \geq 2} \left\| \frac{D^k f(x)}{k!} \right\|^{\frac{1}{k-1}}
\]

when \(Df(x)\) is onto and \(\infty\) otherwise. They give the following:

**Theorem 2.** (Shub-Smale) Let \(f : \mathbb{R}^n \to \mathbb{R}^m\) have zero as a regular value and define

\[
\gamma = \max_{\xi \in f^{-1}(0)} \gamma(f, \xi)
\]

Then there is a universal constant \(C\) so that if \(d(x, f^{-1}(0)) < \frac{\xi}{\gamma}\) then

1. the sequence \(x_{k+1} = N_f(x_k), x_0 = x\), is well defined,
2. it converges to a zero of \(\xi\) of \(f\) and

\[
\|x_k - \xi\| \leq \left(\frac{1}{2}\right)^{2^{k-1}} \beta(f, x).
\]

The case of injective-overdetermined systems is slightly different. The main feature of Newton-Gauss iteration is a quadratic convergence to the zeros of \(f\) and a linear convergence to certain least-square solutions. Kantorovich like theorems are given in Ben-Israel [2], Dennis-Schnabel [5] and Seber-Wild [13]. Alpha-theory is studied by Dedieu-Shub in [4]. They introduce the following invariants,

\[
\alpha_1(f, x) = \beta_1(f, x) \gamma_1(f, x)
\]

\[
\beta_1(f, x) = \|Df(x)\| \|f(x)\|
\]

\[
\gamma_1(f, x) = \sup_{k \geq 2} \left\| \frac{D^k f(x)}{k!} \right\|^{\frac{1}{k-1}}
\]

which differ slightly from \(\alpha, \beta \) and \(\gamma\) introduced in the undetermined case. They prove the following theorems.

**Theorem 3.** (Dedieu-Shub) Let \(x \) and \(\xi \in \mathbb{E}\) be such that \(f(\xi) = 0\), \(Df(\xi)\) is injective and

\[
v = \|x - \xi\| \gamma_1(f, \xi) \leq \frac{3 - \sqrt{7}}{2}.
\]

Then Newton’s sequence \(x_k = N_f^{(k)}(x)\) satisfies

\[
\|x_k - \xi\| \leq \left(\frac{1}{2}\right)^{2^{k-1}} \|x - \xi\|.
\]
Theorem 4. (Dedieu-Shub) Let $x$ and $\xi \in \mathbf{E}$ satisfying $Df(\xi)^\dagger f(\xi) = 0$, $Df(\xi)$ injective and
\[ v = \|x - \xi\| \gamma_1(f, \xi) < 1 - \frac{\sqrt{2}}{2}. \]
If
\[ \lambda = \frac{v + \sqrt{2}(2 - v)\alpha_1(f, \xi)}{1 - 4v + 2v^2} < 1 \]
then Newton's sequence satisfies
\[ \|x_k - \xi\| \leq \lambda^k \|x - \xi\|. \]

Let us now come back to our problem: We recall that
\[ f : \mathbf{E} \to \mathbf{F} \]
is an analytic function with Rank $Df(x) \leq r$ for any $x \in \mathbf{E}$. We let
\[ V = f^{-1}(0) = \{ \xi \in \mathbf{E} : f(\xi) = 0 \} \]
and
\[ V_{ls} = \{ \xi \in \mathbf{E} : Df(\xi)^\dagger f(\xi) = 0 \}. \]
$V$ is the set of zeros of $f$ and $V_{ls}$ the set of least square solutions. See Proposition 1. The following proposition describes the smooth part of $V$:

Proposition 2. Let $\xi \in V$ with Rank $Df(\xi) = r$. Then
1. For any $x \in \mathbf{E}$ with $\|x - \xi\| \gamma_1(f, \xi) < 1 - \frac{\sqrt{2}}{2}$ one has Rank $Df(x) = r$,
2. $V \cap B_{(1 - \frac{\sqrt{2}}{2})/\gamma_1(f, \xi)(\xi)}$ is a submanifold in $\mathbf{E}$ with Dim = $n - r$.

Proof. The first assertion is proved in Lemma 1 below, the second assertion is a classical consequence of the first one, see Helgason [7], Chap. 1, Sect. 15.2. \qed

We do not have a similar result for $V_{ls}$: if $\xi \in V_{ls}$ with Rank $Df(\xi) = r$ is $V_{ls}$ a submanifold around $\xi$?

In order to state our next result we introduce some more notation. Let $\psi(u) = 1 - 4u + 2u^2$. It is decreasing from 1 to 0 when $0 \leq u \leq 1 - \frac{\sqrt{2}}{2}$. $\Pi_{E_1}$ denotes the orthogonal projection onto the subspace $E_1 \subset E$. For any linear operator $L : \mathbf{E} \to \mathbf{F}$,
\[ K(L) = \|L\| \|L^\dagger\| \]
denotes its condition number and $\|L\|$ the operator norm. We also use the following function
\[ A(v, K) = \frac{1}{\psi(v)} + \frac{2 - v}{(1 - v)^2} + \frac{1 + \sqrt{5}(1 - v)^2(2 - v)}{\psi(v)^2} \left( K + \frac{2v - v^2}{(1 - v)^2} \right) , \]
defined for $0 \leq v < 1 - \frac{\sqrt{2}}{2}$ and $K \geq 0$ and
\[ B(v, \alpha) = \frac{1 + \sqrt{5}(1 - v)^2(2 - v)}{\psi(v)^2} + \frac{\theta(\alpha)}{\alpha} , \]
with

$$\theta(\alpha) = \alpha \left( 2 + \frac{(1 + \sqrt{5})(1 + 2\alpha)}{(1 - 2\alpha)^2} \right)$$

defined for $0 \leq \alpha < 1 - \frac{\sqrt{5}}{2}$ and $0 \leq \alpha < \frac{1}{2}$. When $\xi_0$ is a zero of $f$ with Rank $Df(\xi_0) = r$, then for any $x_0 \in E$ in a neighborhood of $\xi_0$ Newton’s sequence starting at $x_0$ converges quadratically to a zero of $f$, but not necessarily equal to $\xi_0$. More precisely we prove here the following: let

$$\gamma_R = \max_{\xi \in B_R(\xi_0) \cap V} \gamma_1(f, \xi)$$

$$A_R = \max_{x \in B_R(\xi_0)} A(\|x - \xi\| \gamma_1(f, \xi), K(Df(\xi))).$$

**Theorem 5.** Let $\xi_0 \in E$, such that $f(\xi_0) = 0$ and Rank $Df(\xi_0) = r$. Let $R > 0$ satisfying the condition $R A_R \gamma_R \leq \frac{1}{2}$, with $\gamma_R$ and $A_R$ as above. Let $x_0 \in B_{\frac{1}{2} R}(\xi_0)$ such that $\xi_0 = \text{proj}_V x_0$ i.e. $\xi_0$ is the point in $V$ the closest to $x_0$. Then Newton’s sequence $x_k = N^{(k)}(x_0)$ is contained in $B_R(\xi_0)$ and

$$d(x_k, V) \leq \left( \frac{1}{2} \right)^{k-1} d(x_0, V).$$

As in the case of overdetermined systems with injective derivatives, the convergence of Newton’s sequence to the set of least square solutions fails to be quadratic. We have

**Theorem 6.** For $\xi_0 \in V_is$ with Rank $Df(\xi_0) = r$ and $0 < R < 1 - \frac{\sqrt{5}}{2}$, define

$$\Lambda = \max_{\xi \in B_R(\xi_0) \cap V_is} A(v, K(Df(\xi))) v + B(v, \alpha_1(f, \xi)) \alpha_1(f, \xi),$$

with $v = \|x - \xi\| \gamma_1(f, \xi)$, and

$$\alpha_1 = \max_{\xi \in B_R(\xi_0) \cap V_is} \alpha_1(f, \xi).$$

Let us suppose that $B_R(\xi_0) \cap V_is$ is a smooth submanifold in $E$, that $\Lambda < 1$ and $2\alpha_1 < 1$. Then, for any $x_0 \in E$ such that

$$x_0 - \xi_0 \in (T_{\xi_0} V_is)^\perp,$$

and $\|x_0 - \xi_0\| \leq \frac{1 - \Lambda}{2\Lambda} R,$

Newton’s sequence $x_k = N^{(k)}(x_0)$ is contained in $B_R(\xi_0)$ and

$$d(x_k, V) \leq \Lambda^k d(x_0, V).$$

Notice the following facts. The hypothesis in Theorem 6 is satisfied in a suitable neighborhood of $\xi_0 \in V_is$ when $V_is$ is smooth around $\xi_0$ and $\alpha_1(f, \xi_0)$ small enough i.e. when $\lim_{R \to 0} \Lambda < 1$.

The invariant $\alpha_1(f, \xi_0)$ is small when the residue function $F(\xi_0) = \|f(\xi_0)\|^2$ is itself small.

The nonconvergence of Newton’s sequence to least square solutions with large residues is a well known fact, see Dennis-Schnabel [5] and Dedieu-Shub [4].

When $\alpha_1(f, \xi_0)$ is small then $\xi_0$ is a strict local minimum for the residue function over $\xi_0 + (\ker Df(\xi_0))^\perp$. More precisely
Proposition 3. For any \( \xi \in V_{1n} \) with Rank \( DF(\xi) = r \) and \( \alpha_1(f, \xi) < \frac{1}{2} \) we have \( DF(\xi) = 0 \) and \( D^2F(\xi)(\hat{x}, \hat{x}) > 0 \) for any \( \hat{x} \in \ker DF(\xi), \hat{x} \neq 0 \).

In the following, under a simple assumption on \( f \) at \( x_0 \) we prove the existence of a least square solution \( \xi \) for \( f \) in a neighborhood of \( x_0 \) and the linear convergence of Newton’s sequence \( N^k_f(x_0) \) to \( \xi \).

Theorem 7. Suppose
\[
\alpha_1(f, x_0) K(Df(x_0)) \leq \frac{1}{48}.
\]
Then Newton’s sequence \( x_{k+1} = N_f(x_k) \) satisfies
\[
\|x_{k+1} - x_k\| \leq \left( \frac{1}{2} \right)^k \|x_1 - x_0\|.
\]
This sequence converges to a least square solution \( \xi \) of \( f \):
\[
DF(\xi)^\top f(\xi) = 0 \text{ and } \|\xi - x_0\| \leq 2\|x_1 - x_0\|.
\]

We close this section with some examples. Examples of “constant rank” systems of equations are given by distance geometry problems: an important tool in determining the three-dimensional structure of a molecule. Distance geometry problems are concerned with finding positions \( x_1, \ldots, x_n \) of \( n \) atoms in \( \mathbb{R}^3 \) such that
\[
\|x_i - x_j\| = \delta(i, j), \quad (i, j) \in S,
\]
where \( S \) is a subset of the atom pairs and \( \delta(i, j) \) is the given distance between atoms \( i \) and \( j \). When all these distances are given, this system has \( 3n \) unknowns and \( n(n-1)/2 \) equations. The dimension of the solution set, when it is nonempty, is at least 6 because these equations are invariant under translations and orthogonal transformations. Similar examples arise from the protein folding problem. For example the Lennard-Jones problem is to find the minimum energy structure of a cluster of \( n \) identical atoms using the Lennard-Jones potential energy:
\[
\min_{x_i \in \mathbb{R}^3} \sum_{1 \leq i \leq n} \sum_{i < j} p(\|x_i - x_j\|)
\]
with \( p(r) = r^{-12} - 2r^{-6} \). Typically \( n \) can take large values: 10 000 for example. This global optimization problem is still unsolved. We can see this problem as a nonlinear least square problem related to the system of equations
\[
(p(\|x_i - x_j\|) + 1)^{1/2} = 0, \quad i < j.
\]
Such a system enters in the category of “constant rank” systems. A good reference for such problems is the survey paper by A. Neumaier [10].
2. Proofs.

In this section we give the proofs of theorems 5, 6 and 7. We begin by a series of lemmas.

**Lemma 1.** Let \( x, y \in \mathbb{E} \) with Rank \( Df(y) \leq \text{Rank} \ Df(x) = r \) and \( u = \| x - y \| \gamma_1(f, x) < 1 - \frac{\sqrt{2}}{2} \). Then

1. \( Df(y) \) and \( \Pi_{\text{im} \ Df(x)} Df(y) \) have rank \( r \),
2. \( \Pi_{\text{Ker} \ Df(x)} + Df(x)^\dagger Df(y) \) is non-singular.
3. \( \| (\Pi_{\text{Ker} \ Df(x)} + Df(x)^\dagger Df(y))^{-1} \| \leq \frac{(1-u)^2}{\psi(u)} \).

**Proof.** \( Df(x)^\dagger(Df(x) - Df(y)) = -Df(x)^\dagger \sum_{k \geq 2} k \frac{D_k f(x)}{k!} (y - x)^{k-1} \) so that

\[
\| Df(x)^\dagger(Df(x) - Df(y)) \| \leq \frac{1}{(1-u)^2} - 1 < 1.
\]

By a classical linear algebra argument

\[
\text{id}_E - Df(x)^\dagger(Df(x) - Df(y)) = \Pi_{\text{Ker} \ Df(x)} + Df(x)^\dagger Df(y)
\]

is invertible and its inverse is bounded by

\[
\frac{1}{1 - \left(\frac{1}{(1-u)^2} - 1\right)} = \frac{(1-u)^2}{\psi(u)}.
\]

This proves 2 and 3. Moreover

\[
\Pi_{\text{im} \ Df(x)} Df(y) = Df(x)(\Pi_{\text{Ker} \ Df(x)} + Df(x)^\dagger Df(y)) = (\text{Rank} \ r) \circ (\text{nonsingular})
\]

has Rank \( r \). Thus \( \text{Rank} \ Df(y) \geq \text{Rank} \ \Pi_{\text{im} \ Df(x)} Df(y) = r \) and we are done. \( \square \)

The following linear algebra lemmas will be useful. Let \( A \) and \( B \) be \( m \times n \) real or complex matrices with non-zero singular values \( \sigma_1 \geq \cdots \geq \sigma_r > 0 \) and \( \tau_1 \geq \cdots \geq \tau_r > 0 \). Thus \( \text{Rank} \ A = \text{Rank} \ B = r \). Let us denote by \( \| A \| \) the usual spectral norm so that

\[
\| A \| = \sigma_1 \quad \text{and} \quad \| A^\dagger \| = \sigma_r^{-1}.
\]

We have (see Stewart-Sun [19], Chap. IV, Theorem 4-11):

**Lemma 2.** (Mirsky)

\[
\max | \sigma_i - \tau_i | \leq \| A - B \|
\]

We also need bounds for \( \| A^\dagger - B^\dagger \| \). The following lemma is valid in our context (see Stewart-Sun [19], Chap. III, Theorem 3.8):

**Lemma 3.** (Wedin)

\[
\| A^\dagger - B^\dagger \| \leq \frac{1 + \sqrt{5}}{2} \max(\| A^\dagger \|^2, \| B^\dagger \|^2) \| A - B \|.
\]
The constant \((1 + \sqrt{5})/2\) appearing in Lemma 3 may be improved according to the values of \(m, n\) and the ranks of \(A\) and \(B\). The precise statement is given in [19], Chapter III, Theorem 3.9. The case of Frobenius norm and arbitrary matrix norms are considered.

The following lemma generalizes a well-known result for square and non-singular matrices. It is probably well-known but we were not able to find it in the literature.

**Lemma 4.** Let \(A\) and \(B\) two \(m \times n\) matrices with \(\text{Rank } (A + B) \leq \text{Rank } A = r\) and \(\|A^\dagger\| \cdot \|B\| < 1\). Then

\[
\text{Rank } (A + B) = r \text{ and } \| (A + B)^\dagger \| \leq \frac{\|A^\dagger\|}{1 - \|A^\dagger\| \cdot \|B\|}.
\]

**Proof.** Let us denote by \(\sigma_1 \geq \cdots \geq \sigma_r > 0\) the non-zero singular values of \(A\) and by \(\rho_1 \geq \cdots \geq \rho_p \geq 0\) \((p = \min(m, n))\) the singular values of \(A + B\). By Lemma 2

\[
\frac{\sigma_r^{-1}}{\sigma_r} \leq \frac{\|A^\dagger\| \cdot \|B\|}{1 - \|A^\dagger\| \cdot \|B\|} < 1
\]

so that \(\rho_r > 0\) and consequently \(\text{Rank } (A + B) \geq r\). Since \(\text{Rank } (A + B) \leq r\) by the hypothesis, we have proved the equality. The nonzero singular values of \(A + B\) are

\[
\rho_1 \geq \cdots \geq \rho_r > 0.
\]

We have

\[
\| (A + B)^\dagger \| = \rho_r^{-1} = \frac{\sigma_r^{-1}}{1 - \frac{\sigma_r - \rho_r}{\sigma_r}} \leq \frac{\|A^\dagger\|}{1 - \|A^\dagger\| \cdot \|B\|}
\]

and we are done. \(\square\)

**Lemma 5.** Let \(x, y \in \mathbb{E}\) with \(\text{Rank } Df(y) \leq \text{Rank } Df(x) = r\) and \(u = \|x - y\| \gamma_1(f, x) < 1 - \frac{\sqrt{2}}{2}\). Then

1. \(\|Df(y) - Df(x)\| \leq \|Df(x)^\dagger\|^{-1} \frac{2u - u^2}{(1-u)^2}\),
2. \(\|Df(y)\| \leq \|Df(x)^\dagger\|^{-1} \left(K(Df(x)) + \frac{2u - u^2}{(1-u)^2}\right)\),
3. \(\|Df(y)^\dagger\| \leq \frac{(1-u)^2}{\psi(u)} \|Df(x)^\dagger\|\),
4. \(\|Df(x)^\dagger - Df(y)^\dagger\| \leq \frac{1 + \sqrt{2} (1-u)^2 (2u-u^2)}{2 \psi(u)} \|Df(x)^\dagger\|\).

**Proof.** \(Df(y) = Df(x) + \sum_{k \geq 2} k \frac{D^k f(x)}{k!} (y - x)^{k-1}\) so that

\[
\|Df(y) - Df(x)\| \leq \|Df(x)^\dagger\|^{-1} \left(\frac{1}{(1-u)^2} - 1\right)
\]

and this proves 1) and 2). Assertion 3) comes from Lemma 4 with \(A = Df(x)\) and \(B = Df(y) - Df(x)\). We have \(\text{Rank } (A + B) = r\) by Lemma 1

\[
\|A^\dagger\| \cdot \|B\| \leq \|Df(x)^\dagger\| \times \|Df(x)^\dagger\|^{-1} \frac{2u - u^2}{(1-u)^2} \leq 1
\]

by Lemma 5.1 and because \(u \leq 1 - \frac{\sqrt{2}}{2}\). Thus, by Lemma 4,

\[
\|Df(y)^\dagger\| \leq \frac{\|Df(x)^\dagger\|}{1 - \frac{2u - u^2}{(1-u)^2}} = \frac{(1-u)^2}{\psi(u)} \|Df(x)^\dagger\|.
\]
The last assertion is a consequence of Lemma 3, Lemma 5.1 and Lemma 5.3.

\[ \| Df(y)^\dagger - Df(x)^\dagger \| \leq \frac{1 + \sqrt{5} (1-u)^4}{2 \psi(u)^2} \| Df(x)^\dagger \|^2 \| Df(x)^\dagger \|^{-1} \frac{2u - u^2}{(1-u)^2}. \]

This achieves the proof of Lemma 5. \qed

**Lemma 6.** Let \( \xi \) and \( x \in \mathbf{E} \) with \( Df(\xi)^\dagger f(\xi) = 0 \), \( \text{Rank} \ Df(x) \leq \text{Rank} \ Df(\xi) = r \) and \( v = \| x - \xi \| \gamma_1(f, \xi) < 1 - \frac{\sqrt{2}}{2} \). Then

\[ \| Df(x)^\dagger f(\xi) \| \leq \frac{1 + \sqrt{5} (1-v)^2(2-v)}{2 \psi(v)^2} \| x - \xi \| \alpha_1(f, \xi). \]

**Proof.** It is a consequence of Lemma 5.4:

\[ \| Df(x)^\dagger f(\xi) \| = \| (Df(x)^\dagger - Df(\xi)^\dagger) f(\xi) \| \leq \| Df(x)^\dagger - Df(\xi)^\dagger \| \| f(\xi) \|
\]

\[ \leq \frac{1 + \sqrt{5} (1-v)^2(2-v)}{2 \psi(v)^2} v \| Df(\xi)^\dagger \| \| f(\xi) \|. \]

\qed

**Lemma 7.** Under the hypothesis of Lemma 6, we have

\[ \| N_f(x) - \xi \| \leq \| \Pi_{\text{Ker} Df(x)}(x - \xi) \| + \frac{v \| x - \xi \|}{\psi(v)} + \frac{1 + \sqrt{5} (1-v)^2(2-v)}{2 \psi(v)^2} \| x - \xi \| \alpha_1(f, \xi). \]

**Proof.** We have

\[ N_f(x) - \xi = x - \xi - Df(x)^\dagger f(x) = \Pi_{\text{Ker} Df(x)}(x - \xi) + Df(x)^\dagger (Df(x)(x - \xi) - f(x) + f(\xi)) - Df(x)^\dagger f(\xi). \]

Using Taylor’s formula for both \( f(x) \) and \( Df(x) \) at \( \xi \) gives

\[ Df(x)(x - \xi) - f(x) + f(\xi) = \sum_{k \geq 1} (k - 1) \frac{Dk f(\xi)}{k!} (x - \xi)^k \]

so that

\[ \| Df(x)(x - \xi) - f(x) + f(\xi) \| \leq \| Df(\xi)^\dagger \|^2 \| x - \xi \| \sum_{k \geq 2} (k - 1) v^{k-1} \]

\[ = \| Df(\xi)^\dagger \|^2 \| x - \xi \| \frac{v}{(1-v)^2}. \]

By Lemma 5.3 we get

\[ \| Df(x)^\dagger (Df(x)(x - \xi) - f(x) + f(\xi)) \| \leq \frac{(1-v)^2}{\psi(v)} \| x - \xi \| \frac{v}{(1-v)^2} = \frac{v \| x - \xi \|}{\psi(v)}. \]

The conclusion comes from Lemma 6:

\[ \| N_f(x) - \xi \| \leq \| \Pi_{\text{Ker} Df(x)}(x - \xi) \| + \frac{v \| x - \xi \|}{\psi(v)} + \frac{1 + \sqrt{5} (1-v)^2(2-v)}{2 \psi(v)^2} \| x - \xi \| \alpha_1(f, \xi) \]

\qed
Lemma 8. Under the hypothesis of Lemma 6, we have
\[ \| \Pi_{\text{Ker} Df(x)}(x - \xi) \| \leq \| \Pi_{\text{Ker} Df(\xi)}(x - \xi) \| + v \| x - \xi \| \left( \frac{2 - v}{(1 - v)^2} + \frac{1 + \sqrt{5} (1 - v)^2 (2 - v)}{\psi(v)^2} \left( K(Df(\xi)) + \frac{2v - v^2}{(1 - v)^2} \right) \right). \]

Proof. \[
\Pi_{\text{Ker} Df(x)}(x - \xi) = \left( id_E - Df(x)^\dagger Df(x) \right) (x - \xi) \]
\[ = \Pi_{\text{Ker} Df(\xi)}(x - \xi) + Df(\xi)^\dagger (Df(\xi) - Df(x))(x - \xi) + (Df(\xi)^\dagger - Df(x)^\dagger) Df(x)(x - \xi) \]
\[ = a + b + c. \]

We give a bound for \( \| b \| \) via Lemma 5.1:
\[ \| b \| \leq \frac{2v - v^2}{(1 - v)^2} \| x - \xi \| \]
and a bound for \( \| c \| \) via Lemma 5.2 and 5.4:
\[ \| c \| \leq \frac{1 + \sqrt{5} (1 - v)^2 (2v - v^2)}{2 \psi(v)^2} \left( K(Df(\xi)) + \frac{2v - v^2}{(1 - v)^2} \right) \| x - \xi \|. \]

\[ \square \]

Lemma 9. Let \( \xi \) and \( x \in E \) with \( f(\xi) = 0 \), \( \text{Rank} Df(\xi) = r \) and \( v = \| x - \xi \| \gamma_1(f, \xi) \leq 1 - \frac{\sqrt{2}}{2} \). Then we have
\[ \| N_f(x) - \xi \| \leq \| \Pi_{\text{Ker} Df(\xi)}(x - \xi) \| + \| x - \xi \| v A(v, K(Df(\xi))) \]
with
\[ A(v, K) = \frac{1}{\psi(v)} + \frac{2 - v}{(1 - v)^2} + \frac{1 + \sqrt{5} (1 - v)^2 (2 - v)}{\psi(v)^2} \left( K + \frac{2v - v^2}{(1 - v)^2} \right) \]
and
\[ K(Df(\xi)) = \| Df(\xi) \| \| Df(\xi)^\dagger \|. \]

Proof. It is an easy consequence of Lemma 7 and Lemma 8 with \( f(\xi) = 0 \).

\[ \square \]

Proof of Theorem 5. Recall that \( \| x_0 - \xi_0 \| \leq \frac{\sqrt{2}}{2} R \). We first notice that, for any \( x \in B_R(\xi_0) \) we have
\[ \| x - \xi_0 \| \gamma(f, \xi_0) \leq R \gamma_{R, \xi_0} \leq \frac{1}{2A_{R, \xi_0}} < 1 - \frac{\sqrt{2}}{2}. \]
The last inequality is from the fact that \( A(v, K) \geq 3 \). Thus \( V \cap B_R(\xi_0) \) is a smooth submanifold in \( E \) (Proposition 2). Since \( \xi_0 \) is the projection of \( x_0 \) onto \( V \), and because \( V \cap B_R(\xi_0) \) is smooth, the orthogonality relation
\[ \Pi_{\text{Ker} Df(\xi_0)}(x_0 - \xi_0) = 0 \]
holds. By Lemma 9, we get
\[ \| N_f(x_0) - \xi_0 \| \leq \| x_0 - \xi_0 \|^2 \gamma_1(f, \xi_0) A(v_0, K_0) \leq \| x_0 - \xi_0 \| R \gamma_{R, \xi_0} A_{R, \xi_0} \leq \frac{1}{2} \| x_0 - \xi_0 \|, \]
so that \(x_1 = N_f(x_0)\) is in \(B_{\frac{3}{2}}(\xi_0)\) and consequently projects on \(V\) in a point \(\xi_1 \in B_{\mathcal{R}}(\xi_0)\) because

\[
\|\xi_1 - \xi_0\| \leq \|x_1 - \xi_1\| + \|x_1 - \xi_0\| \leq 2\|x_1 - \xi_0\| \leq R.
\]

Now we proceed by induction. Let \(x_{k+1} = N_f(x_k)\) and \(\xi_k\) be the projection of \(x_k\) onto \(V\). Then

\[
x_{k+1} - \xi_{k+1} \leq \|x_{k+1} - \xi_k\| \leq \|x_k - \xi_k\|^2 \gamma_1(f, \xi_k) A(v_k, K_k)
\]

\[
\leq \left( \frac{1}{2} \right)^{2^k - 1} \|x_0 - \xi_0\|^2 \gamma_1(f, \xi_k) A(v_k, K_k)
\]

\[
\leq \left( \frac{1}{2} \right)^{2^k - 1} \|x_0 - \xi_0\| \|x_0 - \xi_0\| \gamma_{R, \xi_0} A_{R, \xi_0}
\]

\[
\leq \frac{1}{2} \left( \frac{1}{2} \right)^{2^k - 2} \|x - \xi_0\| = \left( \frac{1}{2} \right)^{2^k - 1} \|x - \xi_0\|.
\]

Here \(K_k = K(f, \xi_k), v_k = \|x_k - \xi_k\| \gamma_1(f, \xi_k)\). Further we have \(\xi_{k+1} \in B_{\mathcal{R}}(\xi_0)\) by noting that

\[
\|\xi_{k+1} - \xi_k\| \leq \|x_{k+1} - \xi_{k+1}\| + \|x_{k+1} - \xi_k\| \leq 2 \|x_{k+1} - \xi_k\|,
\]

\[
\|\xi_{k+1} - \xi_0\| \leq \sum_{j=0}^{k} \|\xi_{j+1} - \xi_j\| \leq 2 \sum_{j=0}^{k} \|x_{j+1} - \xi_j\|
\]

\[
\leq 2 \sum_{j=0}^{k} \left( \frac{1}{2} \right)^{2^j - 1} \|x_0 - \xi_0\| \leq \frac{2}{1 - 1/4} \|x_0 - \xi_0\|
\]

\[
\leq \frac{4}{3} \|x_0 - \xi_0\| \leq R,
\]

which completes the induction. \(\square\)

The following lemmas will be used to prove Proposition 3 and to compute the tangent space \(T_{\xi_0} V_{\xi}\) for \(\xi_0 \in V_{\xi}\) as required in Theorem 5. We begin with an identity given in Stewart-Sun [19] Chapter III, §3.4.

**Lemma 10.** Let \(A\) and \(B\) be \(m \times n\) matrices with \(\text{Rank } A = \text{Rank } B = r\). Then

\[
B^\dagger = A^\dagger - A^\dagger (B - A) A^\dagger + (A^\star A)^\dagger (B - A)^\star \Pi_{\text{Im } A} \Pi_{\text{Ker } A} (B - A)^\star (AA^\star)^\dagger + O(\|B - A\|^2).
\]

**Lemma 11.** When \(\text{Rank } Df(x) = r\), the derivative of \(Df(x)^\dagger f(x)\) is given by

\[
D(Df(x)^\dagger f(x))(\dot{x}) = \Pi_{\text{Ker } Df(x)^\dagger}(\dot{x}) - Df(x)^\dagger (D^2 f(x)(\dot{x})) Df(x)^\dagger f(x) + (Df(x)^\star Df(x))^\dagger (D^2 f(x)(\dot{x}))^\star \Pi_{\text{Im } Df(x)^\dagger} f(x)
\]

\[
+ (Df(x)^\dagger Df(x))^\dagger (D^2 f(x)(\dot{x}))^\star \Pi_{\text{Ker } Df(x)^\dagger} f(x) - \Pi_{\text{Ker } Df(x)} (D^2 f(x)(\dot{x}))^\star (Df(x) Df(x)^\dagger)^\dagger f(x).
\]

**Proof.** Note that

\[
D(Df(x)^\dagger f(x))(\dot{x}) = D(Df(x)^\dagger)(\dot{x}) f(x) + Df(x)^\dagger Df(x)(\dot{x}).
\]

Now use Lemma 10 with \(A = Df(x)\) and the chain rule to \(\dagger \circ Df\). Notice that \(Df(y)\) has rank \(r\) in a neighborhood of \(x\). \(\square\)
Lemma 12. When $Df(\xi)\parallel f(\xi) = 0$ and $\text{Rank } Df(\xi) = r$, we have

$$D(Df(\xi)\parallel f(\xi))\dot{x} = \Pi_{(\text{Ker } Df(\xi))} \dot{x} + (Df(\xi)^*)^\dagger(D^2f(\xi)\dot{x})^\star f(\xi).$$

When $V_{\xi}$ is smooth around $\xi$, its tangent space is the kernel in $E$ of this linear operator.

Proof. In Lemma 11, use the fact $f(\xi) \in \text{Im } Df(\xi)\parallel$; this gives us $\Pi_{(\text{Im } Df(\xi))}\parallel f(\xi) = f(\xi)$ which simplifies the third term, and that $(Df(\xi)Df(\xi)^\dagger) f(\xi) = 0$ which annihilates the last term in Lemma 11. This is because $\text{Ker } (AA^*) = \text{Ker } (A^*) = \text{Im } A^\perp$, for any matrix $A$. \qed

Lemma 13. When $Df(\xi)\parallel f(\xi) = 0$, $\text{Rank } f(\xi) = r$ and $\alpha_1(f, \xi) < \frac{1}{2}$, then

$$\|\Pi_{\text{Ker } Df(\xi)}(x - \xi)\| \leq \|\Pi_{r_{\xi}V_{\xi}}(x - \xi)\| + \theta(\alpha_1(f, \xi))\|x - \xi\|$$

with

$$\theta(\alpha) = \alpha \left(2 + \frac{1 + \sqrt{5}}{2(1 - 2\alpha)^2}\right), 0 \leq \alpha < \frac{1}{2}.$$ 

Proof. We first notice that $D(Df(\xi)\parallel f(\xi))\dot{x}$ is always in $\text{Ker } Df(\xi)\parallel$ so that the rank of this operator is $\leq r$. Let us write $A = \Pi_{(\text{Ker } Df(\xi))}\parallel$ and $B\dot{x} = (Df(\xi)^*Df(\xi))^\dagger(D^2f(\xi)\dot{x})^\star f(\xi)$. We have

$$D(Df(\xi)\parallel f(\xi)) = A + B, \Pi_{\text{Ker } Df(\xi)} = \Pi_{\text{Ker } A} \text{ and } \Pi_{r_{\xi}V_{\xi}} = \Pi_{\text{Ker } (A + B)}.$$  

We also can notice that $\|A\| = \|A^\dagger\| = 1$ and $\|B\| \leq 2\alpha_1(f, \xi) < 1$, by the definition of $\alpha_1$. By Lemma 4 we get $\text{Rank } (A + B) = r$ and

$$\|(A + B)^\dagger\| \leq \frac{1}{1 - \|B\|} \leq \frac{1}{1 - 2\alpha_1(f, \xi)}.$$ 

We have

$$\Pi_{\text{Ker } A} - \Pi_{\text{Ker } (A + B)} = (A + B)^\dagger(A + B) - A^\dagger A = ((A + B)^\dagger - A^\dagger)(A + B) + A^\dagger B,$$

so that, by Lemma 3

$$\|\Pi_{\text{Ker } A} - \Pi_{\text{Ker } (A + B)}\| \leq \frac{1 + \sqrt{5}}{2} \max \left(\|(A + B)^\dagger\|^2, \|A^\dagger\|^2\right)\|B\| \left(\|A\| + \|B\|\right) + \|A^\dagger\| \|B\|$$

$$\leq \frac{1 + \sqrt{5}}{2} \frac{2\alpha_1(f, \xi)}{(1 - 2\alpha_1(f, \xi))^2}(1 + 2\alpha_1(f, \xi)) + 2\alpha_1(f, \xi) = \theta(\alpha_1(f, \xi)).$$

The conclusion is now easy. \qed

Lemma 14. Let $\xi$ be given as in Lemma 13 and $x \in E$ with $v = \|x - \xi\|\gamma_1(f, \xi) < 1 - \frac{1}{\sqrt{2}}$. Then

$$\|N_f(x) - \xi\| \leq \|\Pi_{r_{\xi}V_{\xi}}(x - \xi)\| + A(v, K(Df(\xi)))v\|x - \xi\| + B(v, \alpha_1(f, \xi)\alpha_1(f, \xi))\|x - \xi\|$$

with

$$B(v, \alpha) = \frac{1}{2} \frac{(1 - v)^2(2 - v)}{\psi(v)^2} + \frac{\theta(\alpha)\alpha}{\alpha}.$$
Proof of Theorem 6. The proof of Theorem 6 is similar to the proof of Theorem 5 but uses Lemma 14 instead of Lemma 9. We define \( x_{k+1} = N_f(x_k) \) inductively and let \( \xi_k = \text{proj}_{V_i} x_k \). Inductively by Lemma 14,
\[
\|x_1 - \xi_1\| \leq \|x_1 - \xi_0\| \leq \Lambda \|x_0 - \xi_0\| \leq \frac{(1 - \Lambda)R}{2},
\]
recalling that \( \|x_0 - \xi_0\| \leq \frac{1 - \Lambda}{2 \Lambda} R \) and \( \Lambda < 1 \). Moreover because
\[
\|\xi_1 - \xi_0\| \leq \|\xi_1 - x_1\| + \|x_1 - \xi_0\| \leq \|\xi_1 - x_1\| + 2 \|x_0 - \xi_0\| \leq (1 - \Lambda)R < R
\]
so that \( \xi_1 \in B_R(\xi_0) \). Inductively by Lemma 14 with \( x = x_{k-1} \), we have
\[
\|x_k - \xi_k\| \leq \|x_k - \xi_{k-1}\| \leq \Lambda \|x_0 - \xi_0\|,
\]
Note that \( \|\xi_k - \xi_{k-1}\| \leq \|x_k - \xi_k\| + \|x_k - \xi_{k-1}\| \leq 2 \|x_k - \xi_{k-1}\| \). Moreover \( \xi_k \in B_R(\xi_0) \), because
\[
\|\xi_k - \xi_0\| \leq \sum_{j=1}^{k} \|\xi_j - \xi_{j-1}\| \leq \sum_{j=1}^{k} 2 \|x_j - \xi_{j-1}\| \leq 2 \sum_{j=1}^{k} \Lambda^j \|x_0 - \xi_0\| \leq 2 \Lambda \frac{\Lambda^k}{1 - \Lambda} \|x_0 - \xi_0\| \leq R,
\]
which completes the proof. \[\square\]

Proof of Proposition 3. We first notice that
\[
\|Df(\xi)\| = \mu^{-1} \text{ with } \mu = \min_{\dot{\xi} \in (\text{Ker } Df(\xi))^\perp} \|Df(\xi)\dot{\xi}\|.
\]
We also have
\[
\frac{1}{2} D^2 F(\xi, \dot{\xi}) = (D^2 f(\dot{\xi})f(\xi) + (Df(\xi))^2 f(\xi)) \dot{\xi}
\]
so that
\[
\frac{1}{2} D^2 F(\xi, \dot{\xi}) = f(\xi) f(\xi) \dot{\xi} + (Df(\xi))^2 f(\xi) \dot{\xi} \geq \mu^2 \|Df(\xi)\| \|\dot{\xi}\| \|f(\xi)\| \|D^2 f(\xi)\| \geq \mu^2 (1 - 2\alpha_1(f, \xi)) > 0.
\]

Lemma 15. Let \( x, y \in E \) and \( u = \|y - x\| \gamma_1(f, x) < 1 - \frac{\sqrt{2}}{2} \) as in Lemma 5. Then
\[
1. \beta_1(f, y) \leq \frac{(1 - u^2)}{\psi(u)} (\beta_1(f, x) + \frac{u}{2} \|y - x\| + K(Df(x))\|y - x\|),
\]
\[
2. \gamma_1(f, y) \leq \frac{\gamma_1(f, x)}{(1 - u)^2},
\]
\[
3. \alpha_1(f, y) \leq \frac{1 - u}{\psi(u)^2} \left( \alpha_1(f, x) + \frac{u^2}{2} + K(Df(x))u \right).
\]

Proof. 3) is a consequence of 1) and 2). 1) goes as follows: Recall that \( \gamma_1 = \text{sup} \left( \|Df(x)\| \|\frac{D^k f(x)}{k!}\| \right)^{1/k-1} \) and \( u = \|y - x\| \gamma_1(f, x) \). We have
\[
f(y) = f(x) + Df(x)(y - x) + \sum_{k \geq 2} \frac{D^k f(x)}{k!} (y - x)^k
\]
so that

$$\|f(y)\| \leq \|f(x)\| + \|Df(x)\| \|y - x\| + \|Df(x)^\dagger\|^{-1} \|y - x\| \frac{u}{1 - u}$$

and we conclude by Lemma 5.3. To prove 2) we start from

$$D^k f(y) = \sum_{\ell=0}^{\infty} \frac{(k + \ell)!}{\ell!} D^{k+\ell} f(x) (y - x)^\ell.$$

This gives

$$\frac{\|D^k f(y)\|}{k!} \leq \sum_{\ell=0}^{\infty} \left( \frac{k + \ell}{\ell} \right) \frac{D^{k+\ell} f(x)}{(k + \ell)!} \|y - x\|^\ell$$

noting that \((\frac{1}{1-u})^{(k)} = \frac{1}{(1-u)^{k+r}} = \sum_{\ell=0}^{\infty} \left( \frac{k + \ell}{\ell} \right) u^\ell\). By Lemma 5.3, we obtain

$$\|D f(y)^\dagger\| \frac{\|D^k f(y)\|}{k!} \leq \frac{(1-u)^2}{\psi(u)} \frac{\gamma_1^{k-1}}{(1-u)^{k+1}} = \frac{1}{\psi(u)} \frac{\gamma_1^{k-1}}{(1-u)^{k-1}}$$

thus

$$\gamma_1(f, y) \leq \frac{\gamma_1(f, x)}{(1-u)\psi(u)}.$$

In the following Lemmas we consider \(x_0, x \in E\) with \(\text{Rank } Df(x_0) = r\) and such that

$$u = \|x - x_0\|\gamma_1(f, x_0) \leq 2\alpha_1(f, x_0) < 1 - \frac{\sqrt{2}}{2}.$$  

We also introduce \(y = N_f(x)\). Our objective is to give an estimate for \(\|N_f(y) - N_f(x)\|\) in terms of \(\|y - x\|\). We begin a series of Lemmas. We often use the notations \(\alpha_0 = \alpha_1(f, x_0)\) and \(K_0 = K(Df(x_0))\).

**Lemma 16.** Suppose that \(u = \|x - x_0\|\gamma_1(f, x_0) \leq 2\alpha_1(f, x_0) \leq \frac{1}{24}\). Then

1. \(\alpha_1(f, x) \leq 4.2\alpha_1(f, x_0) K(f(x_0))\),
2. \(K(f(x)) \leq 1.25K(f(x_0))\).

**Proof.** From Lemma 15.3 with \(x\) and \(x_0\) instead of \(y\) and \(x\), we have

$$\alpha_1(f, x) \leq \frac{1-u}{\psi(u)^2}(\alpha_0 + \frac{u^2}{1-u} + K_0 u) \leq \frac{1-u}{\psi(u)^2}(3K_0 \alpha_0 + \frac{2\alpha_0 u}{1-u})$$

$$\leq \frac{1-u}{\psi(u)^2}(3K_0 + \frac{2u}{1-u}) \leq (1.37)\alpha_0(3K_0 + 0.03) \leq 4.2\alpha_0 K_0,$$

for \(u \leq 2\alpha_0 \leq \frac{1}{24}\). A bound for \(K(Df(x))\) is given by Lemma 5.2 and 5.3.

$$K(Df(x)) \leq \frac{(1-u)^2}{\psi(u)}(K(Df(x_0)) + \frac{2u - u^2}{\psi(u)}) \leq (1.122)(K_0 + 0.11) \leq 1.25K_0,$$

for \(u \leq \frac{1}{24}\).  

\(\Box\)
Lemma 17. When $y = N_f(x)$ then
\[
N_f(y) - N_f(x) = Df(x)^\dagger(Df(x)(y - x) + f(x) - f(y)) + (Df(x)^\dagger - Df(y)^\dagger)f(x) + (Df(x)^\dagger - Df(y)^\dagger)(f(y) - f(x)).
\]

Proof. Just note that $y - x = Df(x)^\dagger Df(x)(y - x)$, because $N_f(x) - x \in \text{Im } Df(x)^\dagger$.

In Lemma 17, $N_f(y) - N_f(x)$ appears as the sum of the three quantities. We will use the notation
\[
\|N_f(y) - N_f(x)\| \leq A + B + C,
\]
for the norm of each of these expressions.

Lemma 18. Let $u_x = \|y - x\| \gamma_1(f, x)$.

1. $A \leq \|x - y\| \frac{u_x}{1 - u_x}$.

2. $B \leq \frac{1 + \sqrt{5}(1 - u_x)^2(2 - u_x)}{\psi(u_x)^2} \alpha_1(f, x) \|y - x\|$.

3. $C \leq \frac{1 + \sqrt{5}(1 - u_x)^2(2 - u_x)}{\psi(u_x)^2} u_x(K(Df(x)) + \frac{u_x}{1 - u_x}) \|x - y\|$.

Proof. By using the Taylor series of $f(y)$ around $x$ and the definition of $\gamma_1(f, x)$ we obtain
\[
A \leq \|Df(x)^\dagger(Df(x)(y - x) + f(x) - f(y))\|
\leq \|Df(x)^\dagger\| \sum_{k=2}^{\infty} \left\|\frac{D^k f(x)}{k!}\right\| \|y - x\|^k = \|y - x\| \frac{u_x}{1 - u_x}.
\]

\[\text{From Lemma 5.4, we have}\]
\[
\|Df(x)^\dagger - Df(y)^\dagger\| \leq \frac{1 + \sqrt{5}(1 - u_x)^2(2 - u_x)}{\psi(u_x)^2} u_x \|Df(x)^\dagger\|,
\]
so that
\[
B \leq \frac{1 + \sqrt{5}(1 - u_x)^2(2 - u_x)}{\psi(u_x)^2} \alpha_1(f, x) \|y - x\|.
\]

The Taylor expansion of $f(y)$ at $x$ gives
\[
\|f(y) - f(x)\| \leq \|Df(x)^\dagger\|^{-1}(K(Df(x)) + \frac{u_x}{1 - u_x}) \|x - y\|.
\]
This yields, using Lemma 5.4,
\[
C \leq \frac{1 + \sqrt{5}(1 - u_x)^2(2 - u_x)}{\psi(u_x)^2} u_x(K(Df(x)) + \frac{u_x}{1 - u_x}) \|x - y\|
\]

Proof of Theorem 7. Let us denote $y = N_f(x)$ and $u = \|x - x_0\| \gamma_1(f, x_0)$. Under the hypothesis $\|y - x\| \leq \|x_1 - x_0\|$ and $u \leq \frac{1}{2a}$ we will prove that
\[
\|N_f(y) - N_f(x)\| \leq \frac{1}{2} \|y - x\|.
\]
First notice that, using Lemma 15.2,

\[ u_x = \| y - x \| \gamma_1(f, x) \leq \| x_1 - x_0 \| \frac{\gamma_1(f, x_0)}{(1 - u)\psi(u)} \leq 1.25\alpha_0 \leq \frac{1}{38}, \]

for \( u \leq \frac{1}{24} \). Hence we have

\[ A \leq \| y - x \| \frac{u_x}{1 - u_x} \leq \| y - x \| (1.25)\alpha_0 \frac{1}{1 - u_x} \leq (1.25)(1.03)\alpha_0\| y - x \| \leq 1.3\alpha_0\| y - x \|. \]

It is convenient to have the following estimate:

\[ E_x \leq \frac{1 + \sqrt{5} (1 - u_x)^2(2 - u_x)}{2 \psi(u_x)^2} \leq 3.78 \]

for \( u_x \leq \frac{1}{38} \). For \( B \), by Lemma 16.1, we have

\[ B \leq E_x \alpha_x \| y - x \| \leq (3.78)(4.2\alpha_0 K_0)\| y - x \| \leq 15.9\alpha_0 K_0\| y - x \|. \]

Using Lemma 16.2, we have

\[ C \leq E_x u_x (K_x + \frac{u_x}{1 - u_x})\| y - x \| \leq E_x u_x (1.25 K_0 + 0.03)\| y - x \| \]
\[ \leq (3.78)(1.25)\alpha_0 (1.28) K_0\| y - x \| \leq 6.1\alpha_0 K_0\| y - x \|. \]

Hence we have

\[ \| N_f(y) - N_f(x) \| \leq A + B + C \leq \| y - x \| \leq 24\alpha_0 K_0\| y - x \| \leq \frac{1}{2}\| y - x \|, \]

because \( \alpha_0 K_0 \leq \frac{1}{38} \). Now it is easy to prove, by induction over \( k \), that

\[ \| x_{k+1} - x_k \| \leq \left( \frac{1}{2} \right)^k \| x_1 - x_0 \| \]

This completes the proof.
REFERENCES