NEWTON'S METHOD FOR OVERDETERMINED
SYSTEMS OF EQUATIONS

J. P. DEDIEU AND M. SHUB

ABSTRACT. Complexity theoretic aspects of continuation methods for the solution of square or underdetermined systems of polynomial equations have been studied by various authors. In this paper we consider overdetermined systems where there are more equations than unknowns. We study Newton's method for such a system.

I. INTRODUCTION

Complexity theoretic aspects of continuation methods for the solution of systems of polynomial equations have been studied by Renegar [11], Smale [20]–[22], Shub and Smale [15]–[19], and Dedieu and Shub [4]. These papers have considered square or underdetermined systems. In this paper we consider overdetermined systems, where there are more equations than unknowns. We study Newton's method for such a system and then apply it to the elements of a path in the space of problems to produce a path of solutions. This is the approach of Renegar, Smale, Shub-Smale and Dedieu-Shub referred to above.

A) Newton's method: The affine case. We study here Newton's method to find the zeros of an analytic function

$$f : \mathbb{E} \to \mathbb{F}$$

with $\mathbb{E}$ and $\mathbb{F}$ two real or complex Hilbert spaces. In fact the domain of f may be an open set in $\mathbb{E}$ but, with abuse of notation, we continue to write $f : \mathbb{E} \to \mathbb{F}$.

For a continuous linear operator $A : \mathbb{E} \to \mathbb{F}$ with closed image, the Moore-Penrose inverse of A is the composition of two maps

$$A^\dagger : \mathbb{F} \to \mathbb{E}, \ A^\dagger = i\pi,$$

where π is the orthogonal projection onto $im\ A$ and i, defined on $im\ A$, is the right inverse of A whose image is the orthogonal complement of $ker\ A$ in $\mathbb{E}$. When A is injective with closed image,[1] then $A^\dagger = (A^*A)^{-1}A^*$ with A^* the adjoint of A. When A is surjective, then $A^\dagger = A^*(AA^*)^{-1}$.

Newton's method is defined, when $Df(x)$ has closed image, by

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

[1]Recall from the closed graph theorem that an injective map with closed image is an isomorphism onto its image, i.e., $A^{-1} : Im\ A \to \mathbb{E}$ is continuous.

and Newton's sequence starting at x is given by $x_k = N_f^k(x)$. In the rest of this section we suppose that $Df(x)$ is injective and has a closed image or, at least, lies in the considered domain.

When Newton's sequence converges to $\zeta \in \mathbb{E}$, then ζ satisfies the three following properties (which are clearly equivalent):

1. $Df(\zeta)^\dagger f(\zeta) = 0$,
2. $f(\zeta) \in im\ Df(\zeta)^\perp$,
3. ζ is a stationary point of the gradient of $F(x) = \|f(x)\|^2$, i.e., $grad\|f(\zeta)\|^2 = 0$, but $f(\zeta)$ is not necessarily equal to zero.

The stationary points of $grad\,F(x)$ are related to the nonlinear least squares problem

$$\min_{x \in \mathbb{E}} \|f(x)\|^2,$$

so that Newton's method provides a way to compute its solutions. Notice that this iteration doesn't require the computation of the second derivative $D^2 f(x)$.

This method to solve the nonlinear least squares problem was originally introduced by Gauss in 1809 is called the Gauss-Newton method in the literature: see Seber-Wild [13] and Dennis-Schnabel [5].

The convergence properties of Newton's sequence have been studied in two different contexts: Kantorovich-style theorems (see for example Ostrowski [10], Ortega-Rheinboldt [9]) using data in a neighborhood of a root, and Smale's α-theory using data at one point (Smale [22], Royden [13], Wang [25], and Shub-Smale [15, 18, 19] for the case $n \geq m$ and $Df(x)$ onto; see also Blum-Cucker-Shub-Smale [2]). We follow here the second point of view. Smale's α-theory involves three invariants, which are

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

$$\beta(f, x) = \|Df(x)^\dagger f(x)\|,$$

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

The convergence properties of the sequence $N_f^{(k)}(x), k \geq 0$, can be described when $Df(x)$ is surjective, in terms of these invariants. They need modification in the injective case since, via $Df(x)^\dagger$, we lose the information about the component of $f(x)$ on $im\ Df(x)^\perp$. For this reason we introduce

$$\alpha_1(f, x) = \beta_1(f, x)\gamma_1(f, x),$$

$$\beta_1(f, x) = \|Df(x)^\dagger\|\|(f(x))\|,$$

$$\gamma_1(f, x) = \sup_{k \geq 2} \left(\|Df(x)^\dagger\| \frac{\|D^k f(x)\|}{k!} \right)^{\frac{1}{k-1}}.$$

The main and well-known properties of Newton's method in the case of surjective derivative are

1. fixed points correspond to zeros of f, and
2. convergence to fixed points is quadratic.

We have seen that in the case of injective derivatives Newton's method may have fixed points which are not zeros. Convergence to these fixed points may fail to be

quadratic, as the following simple example shows. We consider

$$f(x) = \begin{pmatrix} x \\ x^2 + a \end{pmatrix}, \ f : \mathbb{R} \to \mathbb{R}^2,$$

where $a \in \mathbb{R}$ is given. Here $x = 0$ is a stationary point of

$$F(x) = x^2 + (x^2 + a)^2 = x^4 + (2a + 1)x^2 + a^2.$$

When $a = 0$ then $x = 0$ is a zero of f; when $a \neq 0$ then $f(0) \neq 0$. Newton's iterate is given by

$$N_f(x) = x - \frac{2x^3 + (2a + 1)x}{1 + 4x^2},$$

so that $DN_f(0) = -2a$. The conclusion is clear: when $a = 0$ then $DN_f(0) = 0$ and Newton's sequence converges quadratically to $x = 0$. When $a \neq 0$ and $|a| < \frac{1}{2}$, then $N_f^k(x)$ converges linearly to $x = 0$; when $|a| > \frac{1}{2}$, then $x = 0$ is a repulsive point for Newton's iteration. At $a = \frac{1}{2}$, $DN_f(0) = -1$ and N_f goes through a period doubling bifurcation. There is now a period two attracting orbit for N_f near 0, so Newton's method fails to converge to a fixed point near 0. For a equal 6, N_f appears by computer experiment to have gone through a whole period doubling cascade. Another remarkable fact is the following. Since

$$D^2 F(x) = 12x^2 + 2(2a + 1),$$

$x = 0$, as a stationary point of F, is a strict local minimum when $|a| < \frac{1}{2}$. In Theorem 4 below we prove a generalization of this fact.

Our main results on Newton's method are of two types, gamma theorems and alpha theorems. Gamma theorems give an estimate of the size of a disc of convergence of Newton's method about a zero. Alpha theorems give a criteria for convergence of Newton's method at a point from the value of alpha at that point. Theorem 1 and 2 are gamma theorems and Theorem 3 is an alpha theorem. Let us denote $\psi(v) = 1 - 4v + 2v^2$. This function decreases from 1 to 0 on the interval $[0, 1 - \frac{\sqrt{2}}{2}]$.

Theorem 1. *Let x and $\zeta \in \mathbb{E}$ be such that $f(\zeta) = 0$, $Df(\zeta)$ is injective with closed image, and*

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

Then Newton's sequence $x_k = N_f^k(x)$ satisfies

$$\|x_k - \zeta\| \leq \left(\frac{1}{2}\right)^{2^k - 1} \|x - \zeta\|.$$

Theorem 2. *Let x and $\zeta \in \mathbb{E}$ satisfy $Df(\zeta)^\dagger f(\zeta) = 0$, $Df(\zeta)$ injective with closed image and*

$$v = \|x - \zeta\|\gamma_1(f, \zeta) < 1 \dot{-} \frac{\sqrt{2}}{2}.$$

If

$$\lambda = \frac{1}{\psi(v)}(v + \sqrt{2}(2 - v)\alpha_1(f, \zeta)) < 1,$$

then Newton's sequence satisfies

$$\|x_k - \zeta\| \le \lambda^k \|x - \zeta\|.$$

Remark 1. Since $v \to 0$ when $x \to \zeta$, the condition $\lambda < 1$ is satisfied when

$$\alpha_1(f, \zeta) < \frac{1}{2\sqrt{2}}$$

for any $x \in \mathbb{E}$ in a ball around ζ.

For any real $k \ge 1$ we now define

$$m(k, \lambda) = \frac{\lambda \psi(\lambda)^2 - 4\lambda^3 - 4k\lambda^2(\lambda - 1)}{\psi(\lambda)^2 + 4\lambda(\lambda - 1)},$$

$$M(k) = \max_{0 \le \lambda \le 1 - \frac{\sqrt{2}}{2}} m(k, \lambda),$$

$$m(k, \lambda_{opt}(k)) = M(k).$$

We have the following.

Theorem 3. *For any $x \in \mathbb{E}$, let us define $\lambda_{opt} = \lambda_{opt}(\|Df(x)\|\|Df(x)^\dagger\|)$. If $Df(x)$ is injective with closed image and $\alpha_1(f, x) \le M(\|Df(x)\|\|Df(x)^\dagger\|)$, then*

1) N_f *maps* $B(x, \frac{\lambda_{opt}}{\gamma_1(f, x)})$ *into itself,*

2) N_f *is a contraction in that ball with contraction constant* $1 - \frac{\alpha_1(f, x)}{\lambda_{opt}} < 1$,

3) *there is a unique* $\zeta \in \mathbb{E}$ *such that* $Df(\zeta)^\dagger f(\zeta) = 0$ *and*

$$\|\zeta - x\| < \frac{\lambda_{opt}}{\gamma_1(f, x)}.$$

Our last local result about Newton's method is the following:

Theorem 4. *Let $\zeta \in \mathbb{E}$ be such that $Df(\zeta)$ is injective, $Df(\zeta)^\dagger f(\zeta) = 0$ and $2\alpha_1(f, \zeta) < 1$. Then*

1) ζ *is an attractive fixed point for Newton's method,*

2) ζ *is a strict local minimum for*

$$F(x) = \|f(x)\|^2$$

We do not know if attracting fixed points for Newton's method are always local minima of $\|f\|^2$.

We may use Theorem 1 to give a complexity upper bound estimate for continuation methods. We state our result in greater generality.

First we recall that, given an analytic function $f : \mathbb{E} \to \mathbb{F}$ and points $x, \zeta \in \mathbb{E}$ with $f(\zeta) = 0$ and $\|N_f^k(x) - \zeta\| \le 2^{1-2^k} \|x - \zeta\|$ for all $k \ge 1$, then x is called an *approximate zero* of f and ζ its *associated zero*. Now suppose we are given a family of analytic functions $f_t : \mathbb{E} \to \mathbb{F}$ for $t \in [0, 1]$ and $\zeta_t \in \mathbb{E}$ depending differentiably on t such that $f_t(\zeta_t) = 0$ and $Df_t(\zeta_t)$ is injective with closed image for all $t \in [0, 1]$. Let L_ζ be the length of the curve ζ_t for $t \in [0, 1]$. Let $\gamma_1 = \sup_{t \in [0,1]} \gamma_1(f_t, \zeta_t)$.

We associate to a subdivision $0 = t_0 < t_1 < \cdots < t_p = 1$ a sequence x_i for $i = 0, \ldots, p$ by $x_0 = \zeta_0$ and $x_{i+1} = N_{f_{t_{i+1}}}(x_i)$.

Theorem 5. *There is a partition $0 = t_0 < t_i < \cdots < t_p = 1$ such that x_i is defined for all $0 \le i \le p$, x_p is an approximate zero for f_1 with associated zero ζ_1, and*

$$p = \lceil \frac{4}{3 - \sqrt{7}} \gamma_1 L_\zeta \rceil.$$

Now we state a version of Theorem 5 in terms of the path f_t. In order to do so we require that f_t be differentiable as a function of t. For simplicity we restrict our attention to $f_t \subset P_{(d)}$, the space of polynomial systems $f = (f_1, \ldots, f_m)$ where, for each $1 \leq i \leq m$, f_i is a polynomial in n variables of degree d_i and $(d) = (d_1, \ldots, d_m)$.

Since $f_t(\zeta_t) = 0$ we have $Df_t(\zeta_t)(\dot{\zeta}_t) + \dot{f}_t(\zeta_t) = 0$. Since $Df_t(\zeta_t)$ is injective it follows that $\dot{\zeta}_t = -Df_t^\dagger(\zeta_t)(\dot{f}(\zeta_t))$. For $f \in P_{(d)}$ and $\zeta \in \mathbb{C}^{n+1}$, let $K(f, \zeta)$ be the linear operator mapping $P_{(d)}$ to $\mathbb{C}^n$ defined by $K(f, \zeta)(g) = -Df(\zeta)^\dagger(g(\zeta))$ and $\mu(f, \zeta) = \|K(f, \zeta)\|$. Finally, let $\mu = \sup_t \mu(f_t, \zeta_t)$ and let L_f be the length of the path $f_t \subset P_{(d)}$.

Theorem 6. *There is a partition $0 = t_0 < t_1 < \cdots < t_p = 1$ such that x_i is defined for all $0 \leq i \leq p$, x_p is an approximate zero for f_1 with associated zero ζ_1, and*

$$p = \lceil \frac{4\gamma_1 \mu L_f}{3 - \sqrt{7}} \rceil.$$

Here $P_{(d)}$ and $\mathbb{C}^n$ have Hermitian products which make them Hilbert spaces, and μ, L_ζ, L_f are all defined with respect to the induced norms.

For estimates of μ see [3]. References [2], [3] and [18] have versions of Theorems 5 and 6 when Df is an isomorphism.

B) Newton's method: the multihomogeneous case. Let $\mathbb{E}_1, \ldots, \mathbb{E}_k$ be complex or real vector spaces and $\mathbb{F} = \mathbb{C}^m$ or $\mathbb{R}^m$. Let $\mathbb{E} = \mathbb{E}_1 \times \ldots \times \mathbb{E}_k$ and $((d)) = ((d_1), \ldots, (d_k)), (d_i) = (d_{1i}, \ldots, d_{ki})$ for $i = 1, \ldots, m$. Then $f : \mathbb{E} \to \mathbb{F}$ is multihomogeneous of degree $((d))$ if and only if the ith coordinate function satisfies

$$f_i(\lambda_1 x_1, \ldots, \lambda_k x_k) = \prod_{j=1}^{k} \lambda_j^{d_{ji}} f_i(x_1, \ldots, x_k)$$

for $(x_1, \ldots, x_k) \in \mathbb{E}$ and $(\lambda_1, \ldots, \lambda_k)$ a k-tuple of scalars, i.e., $(\lambda_1, \ldots, \lambda_k) \in \mathbb{G} = \mathbb{C}^k$ or $\mathbb{R}^k$ as the case may be.

We assume throughout that f is analytic. The domain of f may be an open subset of $\mathbb{E}$, but with abuse of notation we continue to write $f : \mathbb{E} \to \mathbb{F}$.

The multihomogeneous projective Newton iteration was introduced by Dedieu and Shub [4] in the case of underdetermined systems. We will use here the results of that paper. The iteration is defined on $\mathbb{E}$ but is invariant under the natural identifications which define the product of the projective spaces $\mathbb{P}(\mathbb{E}_1) \times \ldots \times \mathbb{P}(\mathbb{E}_k)$. Indeed this is much of our motivation in defining Newton's iteration as we do, but it is important to keep in mind that implementations of the method reside in $\mathbb{E}$ itself!

For the rest of this section we will assume that $\mathbb{E}, \mathbb{F}$ and $\mathbb{G}$ are complex and finite dimensional vector spaces and that $\mathbb{E}_i$ has an Hermitian product $\langle, \rangle_i$. For the case where $\mathbb{E}, \mathbb{F}$ and $\mathbb{G}$ are real we would replace the Hermitian product by an inner product. Also, we denote

$$\mathbb{E}^* = (\mathbb{E}_1 \backslash \{0\}) \times \ldots \times (\mathbb{E}_k \backslash \{0\}).$$

If $\lambda = (\lambda_1, \ldots, \lambda_k) \in \mathbb{G}$ we define

$$\times \lambda : \mathbb{E} \to \mathbb{E}$$

by

$$\times \lambda x = (\lambda_1 x_1, \ldots, \lambda_k x_k).$$

For $x \in \mathbb{E}^*$, $x = (x_1, \ldots, x_k)$, we let $x_i^\perp$ be the Hermitian complement of x_i in $\mathbb{E}_i$,

$$x^\perp = \prod_{i=1}^{k} x_i^\perp \subset \mathbb{E} \text{ and } V_x = (x^\perp)^\perp \subset \mathbb{E}.$$

Notice that V_x is also the subspace of $\mathbb{E}$ spanned by the vectors $(0, \ldots, x_i, \ldots, 0)$, $i = 1, \ldots, k$. The dimension of V_x is k, since $x \in \mathbb{E}^*$.

We now define an Hermitian structure on $\mathbb{E}$, and hence on $x^\perp$, depending on x, by

$$\langle v, w \rangle_x = \sum_{i=1}^{k} \frac{\langle v_i, w_i \rangle_i}{\langle x_i, x_i \rangle_i}$$

for $x \in \mathbb{E}^*$, $v, w \in \mathbb{E}$. If $\lambda \in \mathbb{G}^*$, then $\times \lambda$ maps $x^\perp$ onto $(\times \lambda x)^\perp$ and

$$(*) \qquad \langle \times \lambda v, \times \lambda w \rangle_{(\times \lambda x)} = \langle v, w \rangle_x.$$

Condition $(*)$ says that $\times \lambda$ is an isometry from $x^\perp$ to $(\times \lambda x)^\perp$ as well as from $\mathbb{E}$ to $\mathbb{E}$ with their given Hermitian products.

We are now ready to define the multihomogeneous projective Newton iteration for f. We denote this map as $N_f : \prod_i \mathbb{E}_i \hookleftarrow$,

$$N_f(x) = x - (Df(x)|_{x^\perp})^\dagger f(x).$$

As for the affine versions of Newton's method, define

$$\gamma_1(f, x) = \max(1, \sup_{k \geq 2} \left\| (Df(x)|_{x^\perp})^\dagger \right\|_x \left\| \frac{D^k f(x)}{k!} \right\|_x^{\frac{1}{k-1}}),$$

$$\beta_1(f, x) = \left\| (Df(x)|_{x^\perp})^\dagger \right\|_x \| f(x) \|,$$

$$\alpha_1(f, x) = \beta_1(f, x) \gamma_1(f, x).$$

In the definition of $\gamma_1(f, x)$, $\| \ \|_x$ is the operator norm with respect to $\langle \ , \ \rangle_x$. These invariants satisfy the following:

$$\star_1(f, \times \lambda x) = \star_1(f, x)$$

for any $x \in \mathbb{E}^*$, $\lambda \in \mathbb{G}^*$. We recall that for $1 \leq i \leq k$ the Riemannian distance in $\mathbb{P}(\mathbb{E}_i)$ is given by

$$d_R(x_i, y_i) = \arccos \frac{|\langle x_i, y_i \rangle_i|}{\|x_i\|_i \|y_i\|_i},$$

and in $\mathbb{P}(\mathbb{E}_1) \times \ldots \times \mathbb{P}(\mathbb{E}_k)$ by

$$d_R(x, y) = \left(\sum_{i=1}^{k} d_R(x_i, y_i)^2 \right)^{1/2},$$

where $x = (x_1, \ldots, x_k)$ and $y = (y_1, \ldots, y_k) \in \mathbb{E}^*$. In fact we will use here the distances defined in $\mathbb{P}(\mathbb{E}_i)$ by

$$d_P(x_i, y_i) = \sin d_R(x_i, y_i),$$

and in $\mathbb{P}(\mathbb{E}_1) \times \ldots \times \mathbb{P}(\mathbb{E}_k)$ by

$$d_P(x, y) = \left(\sum_{i=1}^{k} d_P(x_i, y_i)^2 \right)^{1/2}$$

Our main theorem in this section is the following:

Theorem 7. *There is a universal constant $\gamma_u > 0$, approximately equal to .15872, with the following properties: Let $\zeta \in \mathbb{E}^*$ be a zero of f, with $Df(\zeta)|_{\zeta^\perp}$ injective and $x \in \mathbb{E}^*$, such that*

$$d_P(x, \zeta)\gamma_1(f, \zeta) \leq \gamma_u.$$

Then the multihomogeneous Newton sequence $x_0 = x$, $x_{k+1} = N_f(x_k)$, converges to ζ and, for each $k \geq 1$,

$$d_P(\zeta, x_k) \leq \left(\frac{1}{2}\right)^{2^k - 1} d_P(\zeta, x).$$

Theorems 5 and 6 now have their multihomogeneous analogues, which follow from Theorem 7 instead of Theorem 1. We don't bother to state them.

II. The proofs of Theorems 1–6

Our proofs of Theorems 1–4 proceed by a series of lemmas. The proofs of Theorems 1–3 are analogues of the proofs of the alpha and gamma theorems in [22] and [2].

We frequently use the notation π_G to denote orthogonal projection on G.

Lemma 1. *When $Df(x)$ is injective and*

$$u = \|x - y\|\gamma_1(f, x) < 1 - \frac{\sqrt{2}}{2},$$

then:

1. *$Df(y)$ and $\pi_{im\ Df(x)}Df(y)$ are injective;*
2. *$Df(x)^\dagger Df(y)$ is nonsingular, and its inverse is equal to*

$$(\pi_{im\ Df(x)}Df(y))^\dagger\ Df(x);$$

3. *$\|(Df(x)^\dagger Df(y))^{-1}\| \leq \frac{(1-u)^2}{\psi(u)}$.*

Proof. We have

$$Df(x)^\dagger(Df(x) - Df(y)) = -Df(x)^\dagger \sum_{k \geq 2} k\frac{D^k f(x)}{k!}(y - x)^{k-1},$$

so that

$$\|Df(x)^\dagger(Df(x) - Df(y))\| \leq \sum_{k \geq 2} k\|Df(x)^\dagger\|\frac{\|D^k f(x)\|}{k!}\|y - x\|^{k-1}$$

$$\leq \sum_{k \geq 2} k\gamma_1(f, x)^{k-1}\|y - x\|^{k-1} = \frac{1}{(1 - u)^2} - 1 < 1,$$

since $u < 1 - \frac{\sqrt{2}}{2}$. By a classical argument $id_E - Df(x)^\dagger(Df(x) - Df(y)) = Df(x)^\dagger Df(y)$ is invertible, and its inverse is bounded in norm by

$$\|(Df(x)^\dagger Df(y))^{-1}\| \leq \frac{1}{1 - \left(\frac{1}{(1-u)^2} - 1\right)} = \frac{(1 - u)^2}{\psi(u)}.$$

Moreover we have

$$(\pi_{im\ Df(x)}Df(y))^{\dagger}Df(x)(Df(x)^{\dagger}Df(y))$$

$$= (\pi_{im\ Df(x)}Df(y))^{\dagger} \circ (\pi_{im\ Df(x)}Df(y)) = id_E$$

and this proves 2. $Df(y)$ is injective because $\pi_{im\ Df(x)}Df(y)$ is also injective. $\quad\square$

Below we will use the following lemmas. Let $A, B : \mathbb{E} \to \mathbb{F}$ be two injective linear bounded operators with closed range. Let us define

$$\mu_A = \inf_{\|x\|=1} \|Ax\|.$$

Lemma 2. *1)* $\|A^{\dagger}\| = \mu_A^{-1}$, *2)* $|\mu_A - \mu_B| \leq \|A - B\|$.

Proof. The proof is easy, and is left to the reader. $\quad\square$

Lemma 3 (Wedin Theorem).

$$\|A^{\dagger} - B^{\dagger}\| \leq \sqrt{2}\|A^{\dagger}\|\|B^{\dagger}\|\|A - B\|.$$

A proof of this lemma is given in Stewart-Sun [24] for $m \times n$ matrices with $m \geq n$ and rank $A = $ rank $B = n$. In fact this proof is valid in the more general context we deal with here.

Lemma 4. *If $Df(x)$ is injective and*

$$u = \|x - y\|\gamma_1(f, x) < 1 - \frac{\sqrt{2}}{2},$$

then

$$\|Df(y)^{\dagger}\| \leq \|Df(x)^{\dagger}\|\frac{(1 - u)^2}{\psi(u)}.$$

Proof. We have

$$Df(y) = Df(x) + \sum_{k \geq 2} k\frac{D^k f(x)}{k!}(y - x)^{k-1},$$

so that

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

By Lemma 1 $Df(x)$ and $Df(y)$ are both injective, so that, by Lemma 2, $|\mu_x - \mu_y| \leq \|Df(y) - Df(x)\|$. Moreover,

$$\mu_y^{-1} = \|Df(y)^{\dagger}\|, \qquad \mu_x^{-1} = \|Df(x)^{\dagger}\|,$$

$$\mu_y \geq \mu_x - \|Df(y) - Df(x)\| \geq \mu_x\left(2 - \frac{u}{(1 - u)^2}\right) = \mu_x\frac{\psi(u)}{(1 - u)^2},$$

because $u < 1 - \frac{\sqrt{2}}{2}$. Thus

$$\|Df(y)^{\dagger}\| = \mu_y^{-1} = \mu_x^{-1}(\mu_x\mu_y^{-1}) \leq \mu_x^{-1}\frac{(1 - u)^2}{\psi(u)} = \|Df(x)^{\dagger}\|\frac{(1 - u)^2}{\psi(u)}.$$

$\quad\square$

Lemma 5. *Let* $\zeta \in \mathbb{E}$ *with* $Df(\zeta)^{\dagger}f(\zeta) = 0$ *and* $Df(\zeta)$ *injective. For any* $x \in \mathbb{E}$ *such that*

$$v = \|\zeta - x\|\gamma_1(f, \zeta) < 1 - \frac{\sqrt{2}}{2}$$

we have

$$\|Df(x)^{\dagger}f(\zeta)\| \leq \sqrt{2}\frac{2v - v^2}{\psi(v)}\beta_1(f, \zeta).$$

Proof. By Lemma 3 we have

$$\|Df(x)^{\dagger}f(\zeta)\| = \|(Df(x)^{\dagger} - Df(\zeta)^{\dagger})f(\zeta)\|$$

$$\leq \sqrt{2}\|Df(x)^{\dagger}\|\|Df(\zeta)^{\dagger}\|\|Df(x) - Df(\zeta)\|\|f(\zeta)\|.$$

We now use Lemma 4 to bound $\|Df(x)^{\dagger}\|$. As in the proof of Lemma 3 we get

$$\|Df(x) - Df(\zeta)\| \leq \|Df(\zeta)^{\dagger}\|^{-1}\left(\frac{1}{(1-v)^2} - 1\right),$$

so that

$$\|Df(x)^{\dagger}f(\zeta)\| \leq \sqrt{2}\|Df(\zeta)^{\dagger}\|\frac{(1-v)^2}{\psi(v)}\|Df(\zeta)^{\dagger}\|$$

$$\times \|Df(\zeta)^{\dagger}\|^{-1}\left(\frac{1}{(1-v)^2} - 1\right)\|f(\zeta)\|,$$

and we are done. □

Lemma 6. *Under the hypothesis of Lemma 5 we have*

$$\|N_f(x) - \zeta\| \leq \|x - \zeta\|\frac{v}{\psi(v)} + \sqrt{2}\frac{2v - v^2}{\psi(v)}\beta_1(f, \zeta).$$

Proof. We have

$$N_f(x) - \zeta = x - \zeta - Df(x)^{\dagger}f(x)$$
$$= Df(x)^{\dagger}((Df(x)(x - \zeta) - f(x) + f(\zeta)) - Df(x)^{\dagger}f(\zeta)).$$

Thus

$$\|N_f(x) - \zeta\| \leq \|Df(x)^{\dagger}\|\|Df(x)(x - \zeta) + f(\zeta) - f(x)\| + \|Df(x)^{\dagger}f(\zeta)\|.$$

Moreover,

$$Df(x)(x - \zeta) - f(x) + f(\zeta) = \sum_{k \geq 1}(k - 1)\frac{D^k f(\zeta)}{k!}(z - \zeta)^k,$$

so that

$$\|Df(x)(x - \zeta) - f(x) + f(\zeta)\| \leq \|Df(\zeta)^{\dagger}\|^{-1}\|x - \zeta\|\frac{v}{(1-v)^2}.$$

By Lemmas 4 and 5 we get

$$\|N_f(x) - \zeta\| \leq \frac{(1-v)^2}{\psi(v)}\|x - \zeta\|\frac{v}{(1-v)^2} + \sqrt{2}\frac{2v - v^2}{\psi(v)}\beta_1(f, v).$$

□

Proof of Theorem 1. We have

$$\|N_f(x) - \zeta\| \le \frac{v}{\psi(v)}\|x - \zeta\|.$$

When $v \le \frac{5-\sqrt{17}}{4}$ then $\frac{v}{\psi(v)} \le 1$, and by induction

$$\|N_f^k(x) - \zeta\| \le \left(\frac{v}{\psi(v)}\right)^{2^k - 1}\|x - \zeta\|.$$

When $v \le \frac{3-\sqrt{7}}{2}$ then $\frac{v}{\psi(v)} \le 1/2$, and we are done.   □

Proof of Theorem 2. We have

$$\|N_f(x) - \zeta\| \le \frac{v}{\psi(v)}\|x - \zeta\| + \sqrt{2}\frac{2v - v^2}{\psi(v)}\beta_1(f, \zeta),$$

or, equivalently,

$$\|N_f(x) - \zeta\| \le \left(\frac{v}{\psi(v)} + \sqrt{2}\frac{2 - v}{\psi(v)}\alpha_1(f, \zeta)\right)\|x - \zeta\|.$$

When $2\sqrt{2}\alpha_1(f, \zeta) < 1$ and v is small enough we have

$$\lambda = \frac{v}{\psi(v)} + \sqrt{2}\frac{2 - v}{\psi(v)}\alpha_1(f, \zeta) < 1,$$

so that

$$\|N_f(x) - \zeta\| \le \lambda\|x - \zeta\|.$$

An induction finishes the proof.   □

Lemma 7. *The derivative of the Moore-Penrose inverse of A, when A is an injective bounded linear operator with closed image, is given by*

$$DA^\dagger(E) = -A^\dagger E A^\dagger + (A^*A)^{-1}E^*\pi_{(im\ A)^\perp}.$$

Proof. Since $A^\dagger = (A^*A)^{-1}A^*$, the lemma follows by straightforward differentiation.   □

We deduce from this result an expression for the derivative of $N_f(x)$:

Lemma 8. *When $Df(x)$ is injective, then*

$$DN_f(x)\dot{x} = Df(x)^\dagger(D^2f(x)\dot{x})Df(x)^\dagger f(x)$$
$$- (Df(x)^*Df(x))^{-1}(D^2f(x)\dot{x})^*\pi_{im\ Df(x)^\perp}f(x).$$

The proof uses both the chain rule and Lemma 7.   □

Lemma 9. *When $Df(x)$ is injective, then*

$$\|DN_f(x)\| \le 4\alpha_1(f, x).$$

Proof. By Lemma 8 and the fact that for any linear map A, $\|A\| = \|A^*\|$,

$$\|DN_f(x)\| \le 2\|Df(x)^\dagger\|\|D^2f(x)\|\|Df(x)^\dagger\|\|f(x)\|$$
$$\le 4\gamma_1(f, x)\beta_1(f, x) = 4\alpha_1(f, x).$$

□

Lemma 10. *When $Df(x)$ is injective and $u = \|x - y\|\gamma_1(f,x) < 1 - \frac{\sqrt{2}}{2}$, we have*

1. $\beta_1(f,y) \leq \frac{(1-u)^2}{\psi(u)}(\beta_1(f,x) + \frac{u}{1-u}\|y - x\| + \|Df(x)\|\|Df(x)^\dagger\|\|y - x\|)$,

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

3. $\alpha_1(f,y) \leq \frac{1-u}{\psi(u)^2}(\alpha_1(f,x) + \frac{u^2}{1-u} + \|Df(x)\|\|Df(x)^\dagger\|u)$.

Proof. 3) is a consequence of 1) and 2). 1) goes as follows:

$$f(y) = f(x) + Df(x)(y - x) + \sum_{k \geq 2} \frac{D^k f(x)}{k!}(y - x)^k,$$

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 4. To prove 2) we start from

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

so that

$$\frac{D^k f(y)}{k!} \leq \sum_\ell \binom{k+\ell}{\ell} \frac{D^{k+\ell} f(x)}{(k+\ell)!}\|(y-x)\|^\ell$$

$$\leq \sum_\ell \binom{k+\ell}{\ell} \frac{\gamma_1^{k+\ell-1}}{(k+\ell)!}\|(y-x)\|^\ell\|Df(x)^\dagger\|^{-1}$$

$$= \frac{\gamma_1^{k-1}}{(1-u)^{k+1}}\|Df(x)^\dagger\|^{-1}.$$

By Lemma 4 we obtain

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

Thus

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

$\square$

Proof of Theorem 3. We have, by Lemmas 9 and 10,

$$\|DN_f(y)\| \leq 4\frac{1-u}{\psi(u)^2}(\alpha_1(f,x) + \frac{u^2}{1-u} + \|Df(x)\|\|Df(x)^\dagger\|u)$$

with $u = \|x - y\|\gamma_1(f,x) < 1 - \frac{\sqrt{2}}{2}$. By the mean value theorem this gives

$$\|N_f(y) - x\| \leq \|N_f(y) - N_f(x)\|\|x - y\| + \|N_f(x) - x\|$$

$$\leq 4\frac{1-u}{\psi(u)^2}(\alpha_1(f,x) + \frac{u^2}{1-u} + \kappa u)\|x - y\| + \beta_1$$

with $\kappa = \|Df(x)\|\|Df(x)^\dagger\|$. Suppose now that $\|x - y\|\gamma_1(f,x) \leq \lambda < 1 - \frac{\sqrt{2}}{2}$; then

$$\|N_f(y) - x\| \leq 4\frac{1-u}{\psi(u)^2}(\alpha_1 + \frac{u^2}{1-u} + \kappa u)\frac{\lambda}{\gamma_1} + \frac{\alpha_1}{\gamma_1}.$$

This quantity will be $\leq \frac{\lambda}{\gamma_1}$ when the following estimate is satisfied:

$$4\frac{1-u}{\psi(u)^2}(\alpha_1 + \frac{u^2}{1-u} + \kappa u)\lambda + \alpha_1 \leq \lambda,$$

given by

$$\alpha_1 \leq \frac{\lambda\psi(\lambda)^2 - 4\lambda^3 - 4\kappa\lambda^2(1-\lambda)}{\psi(\lambda)^2 + 4\lambda(1-\lambda)} = m(\kappa, \lambda).$$

The best possible λ is $\lambda_{opt}(\kappa)$ satisfying

$$m(\kappa, \lambda_{opt}) = \max_{0 \leq \lambda \leq 1 - \frac{\sqrt{2}}{2}} m(\kappa, \lambda).$$

In that case N_f sends the ball centered at x with radius $\frac{\lambda}{\gamma_1(f,x)}$ into itself. Moreover, N_f is a contraction with contraction constant

$$\Lambda = 4\frac{1-\lambda}{\psi(\lambda)^2}(\alpha_1(f,x) + \frac{\lambda^2}{1-\lambda} + \|Df(x)\|\|Df(x)^\dagger\|\lambda) \leq 1 - \frac{\alpha_1}{\lambda_{opt}} < 1,$$

and we are done. $\qquad\square$

Proof of Theorem 4. With $F(x) = \|f\|^2$ we have $\frac{1}{2}DF(x) = Df(x)^*f(x)$ and

$$\frac{1}{2}D^2F(x) = Df(x)^*Df(x) + D^2f(x)^*f(x).$$

Moreover, when $Df(\zeta)^\dagger f(\zeta) = 0$, i.e., $f(\zeta) \in im\ Df(\zeta)^\perp$, the derivative of N_f is equal to

$$DN_f(\zeta) = -(Df(\zeta)^*Df(\zeta))^{-1}(D^2f(\zeta))^*f(\zeta).$$

Let us denote $\mu_\zeta = \inf_{\|x\|=1}\|Df(\zeta)x\|$, so that $\|Df(\zeta)^\dagger\| = \mu_\zeta^{-1}$. We want to prove that $\frac{1}{2}D^2F(\zeta)(x,x) > 0$ when $\|x\| = 1$. We have

$$\frac{1}{2}D^2F(\zeta)(x,x) = \|Df(\zeta)x\|^2 - \langle f(\zeta), D^2f(\zeta)(x,x)\rangle$$
$$\geq \mu_\zeta^2 - \|f(\zeta)\|\|D^2f(\zeta)(x,x)\|$$
$$= \mu_\zeta^2(1 - \|Df(\zeta)^\dagger\|^2\|f(\zeta)\|\|D^2f(\zeta)(x,x)\|)$$
$$\geq \mu_\zeta^2(1 - 2\alpha_1(f,\zeta)) > 0.$$

Moreover,

$$\|DN_f(\zeta)\| \leq \|(Df(\zeta)^*Df(\zeta))^{-1}\|\|D^2f(\zeta)\|\|f(\zeta)\|$$
$$= \mu_\zeta^{-2}\|D^2f(\zeta)\|\|f(\zeta)\| \leq 2\alpha_1(f,\zeta) < 1$$

so that ζ is an attractive point for Newton's method, and we are done. $\qquad\square$

Proof of Theorem 5. Choose $0 = t_0 < t_1 \cdots < t_p = 1$ such that $\|\zeta_{t_i} - \zeta_{t_{i+1}}\| < \frac{3-\sqrt{7}}{4}\frac{1}{\gamma_1}$. This is possible since $p = \lceil L_\zeta / \frac{3-\sqrt{7}}{4\gamma_1}\rceil$. Now we claim by induction that

A) $\|x_i - \zeta_{t_i}\| < \frac{3-\sqrt{7}}{4\gamma_1}$ and

B) $\|x_i - \zeta_{t_{i+1}}\| < \frac{3-\sqrt{7}}{2\gamma_1}$

For $i = 0$, $\|x_i - \zeta_{t_i}\| = 0$. From A) for i, B) follows for i since $\|\zeta_{t_i} - \zeta_{t_{i+1}}\| < \frac{3-\sqrt{7}}{4\gamma_1}$. Now A) follows for $i+1$ from B) for i by Theorem 1. Moreover, for each $i = 1, \ldots, p$, x_i is an approximate zero with associated zero ζ_i for f_{t_i} by Theorem 1. $\qquad\square$

Proof of Theorem 6. Theorem 6 follows from 5. We need only see that the length L_ζ of the curve ζ_t is less than or equal to the length L_f, of the curve f_t, times μ, which is proven in the next lemma.   □

Lemma 11. $L_\zeta \leq \mu L_f$.

Proof. We have

$$L_\zeta = \int_0^1 \|\dot{\zeta}_t\| dt \quad L_f = \int_0^1 \|\dot{f}_t\| dt$$

and $\|\dot{\zeta}_t\| \leq \mu(f_t, \zeta_t)\|\dot{f}_t\| \leq \mu\|\dot{f}_t\|$ by the definition of $\mu(f_t, \zeta_t)$ and μ.   □

III. The proof of Theorem 7

Proof of Theorem 7. It will be shown in Lemma 18 that

$$d_P(N_f(x), \zeta) \leq \frac{1}{2} d_P(x, \zeta)^2 \gamma_1(f, \zeta)$$

when $f(\zeta) = 0$, $Df(\zeta)|_{\zeta^\perp}$ is injective and

$$v = d_P(x, \zeta)\gamma_1(f, \zeta) \leq \gamma_u = .15872.$$

The proof follows easily by induction.   □

Remarks. 1. The quantities $N_f(x)$ and d_P appearing in Theorem 7 satisfy the following invariance properties : for any $\lambda \in \mathbb{G}^*$ and $x \in \mathbb{E}^*$, $N_f(\times \lambda x) = \times \lambda N_f(x)$ ([4], Proposition 1) and $d_P(\times \lambda x, \zeta) = d_P(x, \zeta)$. For this reason we can substitute for x the quantity $\times \lambda x$, so that $\times \lambda x - \zeta \in x^\perp$. This is accomplished with

$$\lambda_i = \frac{\langle \zeta_i, x_i \rangle_i}{\langle x_i, x_i \rangle_i}.$$

This quantity cannot be equal to zero because, in such a case, $d_P(\zeta_i, x_i) = 1$ and consequently $d_P(\zeta, x) \geq 1$, contrary to the hypothesis $v \leq .15872$. For this reason we suppose in the following that

$$x - \zeta \in x^\perp$$

In this case

$$d_P(x, \zeta) = \|x - \zeta\|_\zeta.$$

2. We also use here the concept of distance between two vector subspaces in $\mathbb{E}$. If V and W are two such subspaces, this distance is the maximum of the sine of a given $u \in V$ with its orthogonal projection in W. This distance is denoted here by $d_\zeta(V, W)$ because it is related to the Hermitian structure $\langle \, , \, \rangle_\zeta$. Various results concerning the distance between two vector subspaces are proved in [4], section 2.1.

Lemma 12. $d_\zeta(V_x, V_\zeta) = d_\zeta(V_\zeta, V_x) \leq v$.

Proof. See [4], Proposition 4.1 and Lemma 4.a.   □

Lemma 13. *When $f(\zeta) = 0$ and $Df(\zeta)|_{\zeta^\perp}$ is injective, then*
  1) $\ker Df(\zeta) = V_\zeta$,
  2) $im Df(\zeta) = im Df(\zeta)|_{\zeta^\perp}$.

Proof. By Euler's formula

$$Df_i(\zeta)(\times \lambda \zeta) = f_i(\zeta) \sum_{j=1}^{k} d_{ji} \lambda_j = 0,$$

so that $V_\zeta \subseteq ker Df(\zeta)$. Since $Df(\zeta)|_{\zeta^\perp}$ is injective we have $\zeta^\perp \cap ker Df(\zeta) = \{0\}$, so that $\zeta^{\perp\perp} = V_\zeta = ker Df(\zeta)$. □

Lemma 14. *If $v < 1$, $f(\zeta) = 0$ and $Df(\zeta)|_{\zeta^\perp}$ is injective, then $Df(\zeta)|_{x^\perp}$ is injective and*

$$im Df(\zeta) = im Df(\zeta)|_{\zeta^\perp} = im Df(\zeta)|_{x^\perp}.$$

Proof. Since $v < 1$, by Lemma 12 we have $d_\zeta(V_x, V_\zeta) = d_\zeta(V_\zeta, V_x) < 1$, so that, by [4], Proposition 4.3, $V_x \cap \zeta^\perp = \{0\}$ and the orthogonal projection $\pi : x^\perp \to \zeta^\perp$ is an isomorphism. Thus, for any $u \in x^\perp$ we have $Df(\zeta)u = Df(\zeta)\pi u$, so that $im Df(\zeta)|_{x^\perp} = im Df(\zeta)|_{\zeta^\perp}$, and the conclusion holds. □

Lemma 15. *If $v < 1$, $f(\zeta) = 0$ and $Df(\zeta)|_{\zeta^\perp}$ is injective, then:*

1) $\|Df(\zeta)|_{\zeta^\perp}^\dagger - Df(\zeta)|_{x^\perp}^\dagger\|_\zeta \leq \dfrac{v}{\sqrt{1-v^2}} \|Df(\zeta)|_{\zeta^\perp}^\dagger\|_\zeta.$

2) $\|Df(\zeta)|_{x^\perp}^\dagger\|_\zeta \leq \left(1 + \dfrac{v}{\sqrt{1-v^2}}\right) \|Df(\zeta)|_{\zeta^\perp}^\dagger\|_\zeta.$

Proof. By Lemma 14, for any $y \in im Df(\zeta)$ we can find a preimage $a \in \zeta^\perp$ and another preimage $b \in x^\perp$. In other words, $a = Df(\zeta)|_{\zeta^\perp}^\dagger y$, $b = Df(\zeta)|_{x^\perp}^\dagger y$ and $b - a \in ker Df(\zeta) = V_\zeta$. Since

$$d_\zeta(x^\perp, \zeta^\perp) = d_\zeta(V_\zeta, V_x)$$

(see [4], section 2.1), and $a = \pi_{V_\zeta} b$, $b \in x^\perp$, we get

$$\frac{\|a - b\|_\zeta}{\|a\|_\zeta} \leq \tan \arcsin d_\zeta(V_\zeta, V_x).$$

But $\tan \arcsin x = x/\sqrt{1-x^2}$ and, by Lemma 12, $d_\zeta(V_\zeta, V_x) \leq v < 1$, so that

$$\frac{\|a - b\|_\zeta}{\|a\|_\zeta} \leq \frac{d_\zeta(V_\zeta, V_x)}{\sqrt{1 - d_\zeta(V_\zeta, V_x)^2}} \leq \frac{v}{\sqrt{1-v^2}}.$$

We now notice that

$$\|Df(\zeta)|_{\zeta^\perp}^\dagger\|_\zeta = \max \frac{\|a\|_\zeta}{\|y\|}$$

and

$$\|Df(\zeta)|_{\zeta^\perp}^\dagger - Df(\zeta)|_{x^\perp}^\dagger\|_\zeta = \max \frac{\|a - b\|_\zeta}{\|y\|},$$

to complete the proof. □

Lemma 16. *If $v < .25$, $f(\zeta) = 0$ and $Df(\zeta)|_{\zeta^\perp}$ is injective, then*

1) $Df(x)|_{x^\perp}$ *is injective,*

2) $\|Df(x)|_{x^\perp}^\dagger\|_\zeta \leq \dfrac{1 + \dfrac{v}{\sqrt{1-v^2}}}{1 - \left(1 + \dfrac{v}{\sqrt{1-v^2}}\right)\left(\dfrac{1}{(1-v)^2} - 1\right)} \|Df(\zeta)|_{\zeta^\perp}^\dagger\|_\zeta.$

Proof. By Taylor's formula

$$Df(x) = Df(\zeta) + \sum_{k \geq 1} \frac{D^{k+1} f(\zeta)}{k!} (x - \zeta)^k,$$

so that

$$Df(\zeta)|_{x^\perp}^\dagger (Df(\zeta)|_{x^\perp} - Df(x)|_{x^\perp}) = -Df(\zeta)|_{x^\perp}^\dagger \sum_{k \geq 2} k \frac{D^k f(\zeta)}{k!} (x - \zeta)^{k-1}|_{x^\perp}.$$

Since $Df(\zeta)|_{x^\perp}$ is injective (Lemma 14), we have $Df(\zeta)|_{x^\perp}^\dagger Df(\zeta)|_{x^\perp} = id_{x^\perp}$. Thus

$$\|id_{x^\perp} - Df(\zeta)|_{x^\perp}^\dagger Df(x)|_{x^\perp}\|_\zeta \leq \sum_{k \geq 2} k \|Df(\zeta)|_{x^\perp}^\dagger\|_\zeta \frac{\|D^k f(\zeta)\|_\zeta}{k!} \|x - \zeta\|_\zeta^{k-1}$$

and by Lemma 15

$$\leq (1 + \frac{v}{\sqrt{1 - v^2}}) \sum_{k \geq 2} k \|Df(\zeta)|_{\zeta^\perp}^\dagger\|_\zeta \frac{\|D^k f(\zeta)\|_\zeta}{k!} \|x - \zeta\|_\zeta^{k-1}$$

$$\leq (1 + \frac{v}{\sqrt{1 - v^2}}) (\frac{1}{(1 - v)^2} - 1).$$

This last quantity is < 1 because $v < .25$. By a classical linear algebra argument, $Df(\zeta)|_{x^\perp}^\dagger Df(x)|_{x^\perp}$ is invertible and its inverse satisfies

$$(Df(\zeta)|_{x^\perp}^\dagger Df(x)|_{x^\perp})^{-1} \leq \frac{1}{1 - (1 + \frac{v}{\sqrt{1 - v^2}})(\frac{1}{(1 - v)^2} - 1)}.$$

Let us write

$$\pi_{im Df(\zeta)|_{x^\perp}} Df(x)|_{x^\perp} = Df(\zeta)|_{x^\perp} (Df(\zeta)|_{x^\perp}^\dagger Df(x)|_{x^\perp}),$$

the composition of an injective and an invertible map. We obtain that $\pi Df(x)|_{x^\perp}$ is injective, so that $Df(x)|_{x^\perp}$ is itself injective. Moreover,

$$Df(x)|_{x^\perp}^\dagger = (Df(\zeta)|_{x^\perp}^\dagger Df(x)|_{x^\perp})^{-1} (Df(\zeta)|_{x^\perp}^\dagger Df(x)|_{x^\perp}) Df(x)|_{x^\perp}^\dagger$$

$$= (Df(\zeta)|_{x^\perp}^\dagger Df(x)|_{x^\perp})^{-1} Df(\zeta)|_{x^\perp}^\dagger \pi_{im Df(x)|_{x^\perp}},$$

so that, by Lemma 15,

$$\|Df(x)|_{x^\perp}^\dagger\|_\zeta \leq \frac{1}{1 - (1 + \frac{v}{\sqrt{1 - v^2}})(\frac{1}{(1 - v)^2} - 1)} (1 + \frac{v}{\sqrt{1 - v^2}}) \|Df(\zeta)|_{\zeta^\perp}^\dagger\|_\zeta,$$

and we are done. $\square$

Lemma 17. *If $v < .25$, $f(\zeta) = 0$ and $Df(\zeta)|_{\zeta^\perp}$ is injective, then*

$$\|N_f(x) - \zeta\|_\zeta \leq \frac{(1 + \frac{v}{\sqrt{1 - v^2}}) \frac{1}{(1 - v)^2}}{1 - (1 + \frac{v}{\sqrt{1 - v^2}})(\frac{1}{(1 - v)^2} - 1)} v \|x - \zeta\|_\zeta.$$

Proof. Since we have supposed that $x - \zeta \in x^{\perp}$ and because $Df(x)|_{x^{\perp}}$ is injective, we have

$$N_f(x) - \zeta = Df(x)|_{x^{\perp}}^{\dagger}(Df(x)(x - \zeta) - f(x))$$

and by Taylor's formula for both Df and f at ζ

$$= Df(x)|_{x^{\perp}}^{\dagger} \sum_{k \geq 2}(k - 1)\frac{D^k f(\zeta)}{k!}(x - \zeta)^k,$$

so that

$$\|N_f(x) - \zeta\|_{\zeta} \leq \|Df(x)|_{x^{\perp}}^{\dagger}\|_{\zeta} \sum_{k \geq 2}(k - 1)\frac{\|D^k f(\zeta)\|_{\zeta}}{k!}\|x - \zeta\|_{\zeta}^k$$

and by Lemma 16

$$\leq \frac{1 + \dfrac{v}{\sqrt{1 - v^2}}}{1 - (1 + \dfrac{v}{\sqrt{1 - v^2}})(\dfrac{1}{(1 - v)^2} - 1)} \sum_{k \geq 2}(k - 1)\|x - \zeta\|_{\zeta} v^{k-1}.$$

This completes the proof. □

Lemma 18. *When $f(\zeta) = 0$, $Df(\zeta)|_{\zeta^{\perp}}$ is injective and $v = d_P(x, \zeta)\gamma_1(f, \zeta) \leq \gamma_u = .15872$, then*

$$d_P(N_f(x), \zeta) \leq \frac{1}{2}d_P(x, \zeta)^2 \gamma_1(f, \zeta).$$

Proof. This is an easy consequence of Lemma 17; .15872... is the smallest positive root of the equation

$$\frac{(1 + \dfrac{v}{\sqrt{1 - v^2}})\dfrac{1}{(1 - v)^2}}{1 - (1 + \dfrac{v}{\sqrt{1 - v^2}})(\dfrac{1}{(1 - v)^2} - 1)}v = \frac{1}{2}.$$

□

REFERENCES

[1] Allgower, E., and K. Georg, *Continuation and Path Following*, Acta Numerica, 1–64 (1993) MR **94k**:65076.

[2] Blum, L., F. Cucker, M. Shub, and S. Smale, *Complexity and Real Computation*, Springer, New York (1997). MR **99a**:68070

[3] Dedieu, J. P., *Condition Number Analysis for Sparse Polynomial Systems*, in: Foundations of Computational Mathematics, (F. Cucker and M. Shub, eds.), Springer, Heidelberg (1997).

[4] Dedieu, J. P. and M. Shub, *Multihomogeneous Newton Methods,* Math. Comp. (to appear).

[5] Dennis, J., and R. Schnabel, *Numerical Methods for Unconstrained Optimization and Non-linear Equations*, Prentice Hall (1983). MR **85j**:65001

[6] Lang, S., *Real Analysis*, Addison-Wesley, Reading, Mass. (1983). MR **87b**:00001

[7] Li, T. Y., *Numerical Solution of Multivariate Polynomial Systems by Homotopy Continuation Methods*, Acta Numerica, 6, 399–436 (1997). CMP 98:06

[8] Malajovich, G., *On Generalized Newton Algorithms*, Theoretical Computer Science **133**, 65–84 (1994). MR **95g**:65073

[9] Ortega, J., and W. Rheinboldt (editors), *Studies in Numerical Analysis. Vol. 2: Numerical Solutions of Nonlinear Problems*, SIAM, Philadelphia (1968). MR **42**:1302

[10] Ostrowski, A., *Solutions of Equations in Euclidean and Banach Spaces*, Academic Press, New York, (1973). MR **50**:11760

[11] Renegar, J., *On the Efficiency of Newton's Method in Approximating All Zeros of Systems of Complex Polynomials*, Math. of Oper. Research **12**, 121–148 (1987). MR **88j**:65112
[12] Royden, H., *Newton's Method*, Preprint, (1986).
[13] Seber, G., and C. Wild, *Nonlinear Regression*, John Wiley (1989). MR **90j**:62004
[14] Shub, M., *Some Remarks on Dynamical Systems and Numerical Analysis*, in: Dynamical Systems and Partial Differential Equations: Proceedings of the VII ELAM (L. Lara-Carrero and J. Lewowicz, eds.), Universidad Simon Bolivar, Editoricil Equinoccio, pp. 69–92 (1986). MR **88j**:58005
[15] Shub, M. and S. Smale, *Complexity of Bézout's Theorem I: Geometric Aspects*, J. Am. Math. Soc. 6, 459–501 (1993a). MR **93k**:65054
[16] Shub, M. and S. Smale, *Complexity of Bézout's Theorem II: Volumes and Probabilities*, in: Computational Algebraic Geometry, Progress in Mathematics (F. Eyssette and A. Galligo, eds.), vol. 109, Birkhäuser, 267–285 (1993b). MR **94m**:68086
[17] Shub, M. and S. Smale, *Complexity of Bézout's Theorem III: Condition Number and Packing*, J. of Complexity **9**, 4–14 (1993c). MR **94g**:65152
[18] Shub, M. and S. Smale, *Complexity of Bézout's Theorem IV: Probability of Success, Extensions*, SIAM J. Numer. Anal. **33**, 128-148 (1996). MR **97k**:65310
[19] Shub, M. and S. Smale, *Complexity of Bézout's Theorem V: Polynomial Time*, Theoretical Computer Science **133**, 141–164 (1994). MR **96d**:65091
[20] Smale, S., *On the Efficiency of Algorithms of Analysis*, Bull. A.M.S. **13**, 87-121 (1985). MR **86m**:65061
[21] Smale, S., *Algorithms for Solving Equations*, in: Proceedings of the International Congress of Mathematicians, A.M.S., pp. 172–195 (1986a). MR **89d**:65060
[22] Smale, S., *Newton's Method Estimates from Data at One Point*, in: The Merging of Disciplines: New Directions in Pure, Applied and Computational Mathematics (R. Ewing, K. Gross, and C. Martin, eds.), Springer (1986b), 185–196. MR **88e**:65076
[23] Smith, S., *Optimization Techniques on Riemannian Manifolds*, Fields Institute Communications **3**, 113–136 (1994). MR **95g**:58062
[24] Stewart, G. and J. Sun, *Matrix Perturbation Theory*, Academic Press (1990). MR **92a**:65017
[25] Wang, X, *Some Results Relevant to Smale's Reports,* in: Proceedings of the Smalefest (M. V. Hirsch, J. E. Marsden, and M. Shub, eds.), Springer, pp. 456–465 (1993). CMP 94:03

LAO, Université Paul Sabatier, 31062 Toulouse, Cedex 04, France
E-mail address: dedieu@cict.fr

Department of Mathematical Sciences, IBM Research Division, T.J. Watson Research Center, Yorktown Heights, NY 10598
E-mail address: mshub@us.ibm.com