COMPLEXITY OF BEZOUT'S THEOREM IV: PROBABILITY OF SUCCESS; EXTENSIONS

Michael Shub[1] and Steve Smale[1]

Abstract. We estimate the probability that a given number of projective Newton steps applied to a linear homotopy of a system of n homogeneous polynomial equations in $n+1$ complex variables of fixed degrees will find all the roots of the system. We also extend the framework of our analysis to cover the classical implicit function theorem and revisit the condition number in this context. Further complexity theory is developed.

1. Introduction.

1A. Bezout's Theorem Revisited.

Let $f : \mathbb{C}^{n+1} \to \mathbb{C}^n$ be a system of homogeneous polynomials

$$f = (f_1, \dots, f_n),\ \deg f_i = d_i,\ i = 1, \dots, n.$$

The linear space of such f is denoted by $\mathcal{H}_{(d)}$ where $d = (d_1, \dots, d_n)$. Consider the algorithm proposed in Bez I[2] to solve $f(\zeta) = 0$ approximately. This goes by following solutions of f_t, $0 \le t \le 1$ where $f_t = tf + (1-t)g$. Here $g \in \mathcal{H}_{(d)}$ is a certain universal system whose zeros are supposed to be known. Each step of this path-following algorithm is a version of Newton's method called projective Newton. Our complexity result here puts a bound on the number of these steps, a bound depending only on d and the probability of success σ. Note that some $f \in \mathcal{H}_{(d)}$ have a continuum of solutions so that the introduction of σ is natural.

Theorem A. *A number of projective Newton steps sufficient to find all the approximate zeros of $f \in \mathcal{H}_{(d)}$ with probability σ of success is*

$$\frac{cD^3\mathcal{D}n^2(n+1)(N-1)(N-2)}{1-\sigma}$$

[1] Supported in part by NSF grants.

[2] Bez I, Bez II, Bez III refer respectively to the three Shub–Smale [1993] references

where $D = \max_i(d_i)$, $\mathcal{D} = \prod_{i=1}^n d_i$ *and* N *is the dimension of* $\mathcal{H}_{(d)}$.

Remarks. (1) An *approximate zero* is defined in Bez I (also see below) without recourse to an arbitrary ε. Newton's method starting at an approximate zero converges quadratically, immediately to an associated actual zero.

(2) Projective Newton steps are defined in **Shub** [1993], and in Bez I, one can see the detailed full algorithm.

(3) See Bez I, II, III for background. The present paper uses some of these results. In particular, Theorem A uses the main theorem of Bez I and Theorem C of Bez II.

(4) For the case $n = 1$, the number of steps is $\frac{cd^6}{1-\sigma}$.

(5) **Renegar** [1987a] is an important predecessor to this paper. His result specialized to the case $n = 1$ has a factor $\frac{d^{26}}{(1-\sigma)^4}$. In **Smale** [1981], there is a similar result ($n = 1$) with a $\frac{d^9}{(1-\sigma)^7}$, but this paper has no theory for systems. There are much better results dealing with $n = 1$, and no probability σ. For example Shub–Smale [1985], [1986], Kim–Sutherland [1991], Renegar [1987b], Neff [1993] and especially Pan [1987].

(6) The constant c in Theorem A and (4) can be estimated from the explicit constants of Bez I and is not very large.

(7) The proof of Theorem A is given in section 2 below.

(8) Let us elaborate on σ. The space $\mathcal{H}_{(d)}$ is given a unitarily invariant Hermitian inner product which induces a Riemannian metric and probability measure on the associated projective space $\mathcal{P}(\mathcal{H}_{(d)})$. Then given σ, $0 < \sigma < 1$, there is a set in $\mathcal{P}(\mathcal{H}_{(d)})$ of measure σ such that for f in that set, the bound of Theorem A holds.

(9) What is g of Theorem A? Our theory asserts the existence of such a g, but it is not known how to find it. In fact that is the main problem of Bez III (even for $n = 1$). Besides the references in Bez III, see also Conway–Sloan [1988], especially section 2.3 and the references there.

(10) For finding one root of $f \in \mathcal{H}_{(d)}$ it seems likely that one can eliminate the factor $\mathcal{D}$ in the bound of Theorem A. One might use Theorem B of Bez II.

(11) The number of steps in Theorem A must be interpreted as the number of parallel steps; the algorithm moves along $\mathcal{D}$ paths at a time, one for each root, and each path takes the number of steps displayed in Theorem A.

1B. The Condition Number Revisited.

Consider the implicit function theorem. Assume that $F : \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}^m$ is a C^1 map (F could be defined locally or over $\mathbb{C}$). Suppose that $F(a_0, y_0) = 0$ and that the matrix $\frac{\partial F}{\partial y}(a_0, y_0) : \mathbb{R}^m \to \mathbb{R}^m$ is non-singular. Then there exists a C^1 map $G : \mathcal{U}(a_0) \to \mathbb{R}^m$ such that $F(a, G(a)) \equiv 0$ for all a in the open set $\mathcal{U}(a_0) \subset \mathbb{R}^k$.

We call the matrix $DG(a_0) : \mathbb{R}^m \to \mathbb{R}^m$ the *condition matrix* at (a_0, y_0). The *condition number* (essentially as in Wilkinson [1963], and Wosniakowski [1977]) is then defined by $\mu = \mu(a_0, y_0) = \|DG(a_0)\|$, operator norm. Thinking of a as input and y as output, μ is a bound on the infinitesimal output error in terms of the infinitesimal input error.

Remark. The map G is given only implicitly, yet the condition matrix

$$DG(a_0) = \frac{\partial F}{\partial y}(a_0, y_0)^{-1} \frac{\partial F}{\partial a}(a_0, y_0)$$

is given explicitly, and so is its norm, the condition number.

Example 1. Let $\mathcal{P}_d = \{(a_0, \ldots, a_d) = a \in \mathbb{R}^{n+1}\}$ represent the space of polynomials of degree d and define $F : \mathcal{P}_d \times \mathbb{R} \to \mathbb{R}$ by $F(a, y) = \sum_0^d a_i y^i$. Then $\mu(f, \zeta)$ bounds the infinitesimal change in the solution of $f(\zeta) = 0$ as a function of an infinitesimal change in the coefficients (see e.g. Wilkinson [1963], Demmel [1987], [1988], Bez I, II, III).

Example 2. Generalize Example 1 to systems of polynomials $f : \mathbb{R}^n \to \mathbb{R}^n$ (or $f : \mathbb{C}^n \to \mathbb{C}^n$).

Example 3. Let $\mathcal{T}$ be a linear subspace of $\mathcal{P}_d$ over $\mathbb{C}$ and $F : \mathcal{T} \times \mathbb{C} \to \mathbb{C}$ be $F(f, \zeta) = f(\zeta)$. Defining the condition number of these sparse systems is a great convenience.

Example 4. The special case of Example 3, $\mathcal{T} = \{f \in \mathcal{P} \mid f(x) = x^d - a\}$ defines the condition number for the d^{th} root.

As in Example 3, if $\mathcal{P}_d$ is replaced by a linear subspace $\mathcal{T} \subset \mathcal{P}_d$, the "sparse case", only infinitesimal changes in $\mathcal{T}$ are taken into account in the condition number, say $\mu_{\mathcal{T}}(f, \zeta)$. Then $\mu_{\mathcal{T}}(f, \zeta) \leq \mu(f, \zeta)$ and for certain $\mathcal{T}$, $\mu_{\mathcal{T}}$ may be much smaller than $\mu(f, \zeta)$.

We extend the framework of the implicit function theorem. Let F be as above. Let $V = \{(a, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid F(a, y) = 0\}$ and $\pi_1 : V \to \mathbb{R}^k$ be the restriction of the projection, and suppose that zero is a regular value of F.

Proposition. *$D\pi_1$ is singular at $(a, y) \in V$ if and only if $\frac{\partial F}{\partial y}(a, y)$ is singular.*

Here $D\pi_1(a, y) : T_{a,y}(V) \to \mathbb{R}^k$ is the derivative defined on the tangent space. The proof is standard.

Suppose $D\pi_1$ is non-singular at $(a_0, y_0) \in V$ and $G : \mathcal{U}(a_0) \to \mathbb{R}^m$ is as above. If g is the branch of π_1^{-1} taking a_0 to y_0 then $\pi_2 g(a) = G(a)$, where $\pi_2 : V \to \mathbb{R}^m$ is the restriction of the projection. The branch g is defined on a neighborhood of a_0 with values in V, but may be "topologically continued" to become defined on a larger domain of $\mathbb{R}^k$.

The last situation permits a direct extension to Riemannian manifolds. Let X be a manifold of "inputs", Y a manifold of "outputs" and $V \subset X \times Y$ the manifold of solutions to some computational problem. (Algorithms attempt to invert, or approximate the inverse of, $\pi_1 : V \to X$.) We suppose $\dim V = \dim X$ as some kind of local uniqueness hypothesis. If $(a, y) \in V$ is a solution then the *condition matrix* $DG(a) : T_a(X) \to T_y(Y)$ is defined provided $D\pi_1(a, y)$ is non-singular. In this case $\mu(a, y) = \|DG(a)\|$, otherwise "$\mu(a, y) = \infty$" and (a, y) is ill conditioned. The set of ill-conditioned (a, y) in V is defined by the "equation" $\operatorname{Det} D\pi_1(a, y) = 0$ and denoted by Σ'. Let $\pi_1(\Sigma') = \Sigma \subset X$ be the set of ill-conditioned inputs.

This framework includes the condition number defined in Bez I, and clarifies it. This is Example 5. Moreover there is an aspect of universality in the preceeding treatment of condition number.

Example 5 (See Bez I). $X = \mathcal{P}(\mathcal{H}_{(d)})$, $Y = \mathcal{P}(\mathbb{C}^{n+1})$ and

$$V = \{(f, \zeta) \in \mathcal{P}(\mathcal{H}_{(d)}) \times \mathcal{P}(\mathbb{C}^{n+1}) \mid f(\zeta) = 0\}.$$

$DG(f) : T_f(\mathcal{P}(\mathcal{H}_{(d)})) \to T_\zeta(\mathcal{P}(\mathbb{C}^{n+1}))$. $\mu(f, \zeta) = \|DG(f)\|$.

It can be shown that

$$\mu(f, \zeta) = \|f\| \, \|Df(\zeta)|_{N_\zeta}^{-1} \Delta(\|\zeta\|^{d_i - 1})\|.$$

Here $Df(\zeta)|_{N_\zeta} : N_\zeta \to \mathbb{C}^n$ is the derivative restricted to $N_\zeta = \{v \in \mathbb{C}^{n+1} \mid \langle v, \zeta \rangle = 0\}$, and $\Delta(\|\zeta\|^{d_i-1})$ is the diagonal matrix $\operatorname{Diag}(\|\zeta\|^{d_1-1}, \ldots, \|\zeta\|^{d_n-1})$. Compare Bez I, II.

If $\dot{f} \in N_f \subset \mathcal{H}_{(d)}$, $\dot{\zeta} \in N_\zeta \subset \mathbb{C}^{n+1}$, then

$$\|\dot{f}\|_{T_f(\mathcal{P}(\mathcal{H}_{(d)}))} = \frac{\|\dot{f}\|_{\mathcal{H}_{(d)}}}{\|f\|_{\mathcal{H}_{(d)}}}$$

$$\|\dot{\zeta}\|_{T_\zeta(\mathcal{P}(\mathbb{C}^{n+1}))} = \frac{\|\dot{\zeta}\|_{\mathbb{C}^{n+1}}}{\|\zeta\|_{\mathbb{C}^{n+1}}}$$

hence μ may be thought of as a relative condition number as in Wozniakowski [1977]. In general, condition numbers for homogeneous problems will have natural relative condition number interpretations.

In Bez I, II, III we normalized μ using factors of $d_i^{1/2}$. That is

$$\mu_{\text{norm}}(f, \zeta) = \|f\| \, \|Df(\zeta)|_{N_\zeta}^{-1} \Delta(d_i^{1/2}) \Delta(\|\zeta\|^{d_i-1})\|.$$

Henceforth we will call that the *normalized condition number*. (In Bez I, this was called $\mu_{\text{proj}}(f, \zeta)$.) The normalization gave the condition number theorem, restated below, a shorter form.

Now one may describe μ in the case of sparse systems of homogeneous polynomials as well, as in Example 3. This permits the sharpening of various results of our previous papers in the sparse case, as will become evident in the following.

Remark. The condition matrix has an interpretation in economics as the matrix of comparative statics. For example it dictates how infinitesimal equilibrium allocations change as a function of infinitesimal endowment charges (see e.g. Smale [1981]).

In the situation of Example 5, we restate the condition number theorem proved in Bez I and Bez II.

Let $\Sigma' \subset V$ (see Example 5) be the ill-posed set, i.e. $(f,\zeta) \in \Sigma'$ if and only if ζ is a degenerate root of f (or the derivative $Df(\zeta) : \mathbb{C}^{n+1} \to \mathbb{C}^n$ has rank $< n$). The map $\pi_2 : V \to \mathcal{P}(\mathbb{C}^{n+1})$ has fiber V_ζ over $\zeta \in \mathcal{P}(\mathbb{C}^{n+1})$, given by

$$V_\zeta = \{f \in \mathcal{P}(\mathcal{H}_{(d)}) \mid f(\zeta) = 0\} = \pi_2^{-1}(\zeta).$$

Thus as a subspace of $\mathcal{P}(\mathcal{H}_{(d)})$, V_ζ has an induced metric d_ζ. The condition number theorem gives a formula for $\mu(f,\zeta)$ in terms of this fiber distance.

Condition Number Theorem (Bez I, II, III). *For* $(f,\zeta) \in V \subset \mathcal{P}(\mathcal{H}_{(d)}) \times \mathcal{P}(\mathbb{C}^{n+1})$

$$\mu(f,\zeta) = \frac{\|f\|}{\|\Delta(d_i^{1/2})f\| \sin d_\zeta((\Delta(d_i^{1/2})f,\zeta), \Sigma' \cap V_\zeta)}.$$

A corollary to the condition number theorem computes $\mu(f) = \max\limits_{\substack{\zeta \\ f(\zeta)=0}} \mu(f,\zeta)$ in terms of $\min d_\zeta$ in the obvious way

$$\mu(f) = \frac{\|f\|}{\|\Delta(d_i^{1/2})f\| \min_\zeta \sin d_\zeta((\Delta(d_i^{1/2})f,\zeta), \Sigma' \cap V_\zeta)}.$$

Next, we consider the condition number for the eigenvector, eigenvalue problem and show how it fits into the preceeding picture and how a coresponding condition number theorem holds. For background see Wilkinson [1984], Demmel [1987,1988].

Let $\mathcal{M}(n)$ be the space of all $n \times n$ complex matrices and V be the subvariety of $\mathcal{M}(n) \times \mathcal{P}(\mathbb{C}^n) \times \mathbb{C}$ defined by

$$V = \{(M, v, \lambda) \in \mathcal{M}(n) \times \mathcal{P}(\mathbb{C}^n) \times \mathbb{C} \mid Mv = \lambda v\}.$$

The tangent space to V at (M, v, λ) is defined by

$$(\dot{M}, \dot{v}, \dot{\lambda}) \in T_M(\mathcal{M}(n)) \times T_v(\mathcal{P}(\mathbb{C}^n)) \times \mathbb{C}$$

satisfying (just differentiate $Mv = \lambda v$)

$$(\dot{M} - \dot{\lambda} I)v + (M - \lambda I)\dot{v} = 0, \ \langle \dot{v}, v \rangle = 0.$$

If M is a regular value of the restriction $\pi_1 : V \to \mathcal{M}(n)$ of the projection, then $\dot{v}$ and $\dot{\lambda}$ are each linear functions of $\dot{M}$. These functions are the condition matrices, say

$$K_1(M, v, \lambda)\dot{M} = \dot{v}, \;\; K_2(M, v, \lambda)\dot{M} = \dot{\lambda}.$$

Multiplying (M, λ) by a sccalar c and leaving v fixed we see that

$$K_1(cM, v, c\lambda) = \frac{1}{c} K_1(M, v, \lambda)$$
$$K_2(cM, v, c\lambda) = K_2(M, v, \lambda).$$

The Hermitian structure on $\mathcal{P}(\mathbb{C}^n)$ is the usual one, so for

$$u_1, u_2 \in v^\perp, \;\; \langle u_1, u_2 \rangle = \frac{\langle u_1, u_2 \rangle_{\mathbb{C}^n}}{\langle v, v \rangle_{\mathbb{C}^n}}.$$

We take trace$(AB^*) = \langle A, B \rangle$ as a Hermitian structure on $\mathcal{M}(n)$. Here B^* is the Hermitian transpose of B. The induced norm is the Frobenius norm $\|A\|^2 = \sum_{i,j} |a_{ij}|^2$. The condition numbers are then defined from the condition matrices as usual, say

$$C_i(M, v, \lambda) = \|K_i(M, v, \lambda)\|, \;\; i = 1, 2.$$

Define the "ill-posed" variety Σ' by

$$\Sigma' = \{(M, v, \lambda) \in V \mid \operatorname{rank}(\lambda I - M)^k < n - 1, \text{ some integer } k > 0\}.$$

Consider

$$\begin{array}{ccc} & V & \\ \pi_1 \swarrow & & \searrow \pi_2 \\ \mathcal{M}(n) & & \mathcal{P}(\mathbb{C}^n) \times \mathbb{C} \end{array}$$

where π_1 and π_2 are the restrictions to V of the natural projections from the product space, and $\Sigma = \pi_1(\Sigma')$. On $V - \pi_1^{-1}(\Sigma')$, π_1 is an n-fold covering map. The fiber $V_{v,\lambda} = \pi_2^{-1}(v, \lambda)$, of π_2 is an affine subspace of $\mathcal{M}(n)$ of codimension n. Let $d_{v,\lambda}$ be induced metric on $V_{v,\lambda}$.

Second Condition Number Theorem. *For* $(M, v, \lambda) \in V$

(a) $$C_1(M, v, \lambda) = [d_{v,\lambda}((M, v, \lambda), \Sigma' \cap V_{v,\lambda})]^{-1}$$

(b) $$C_2(M, v, \lambda) \leq \left(\frac{\|M\|^2}{d_{v,\lambda}((M, v, \lambda), \Sigma' \cap V_{v,\lambda})} + 1 \right)^{1/2}.$$

As in the (first) condition number theorem one has an obvious corollary for $C_i(M)$ by taking the maximum over $(v, \lambda) \in \pi_1^{-1}(M)$.

The proof of the second condition number theorem is in section 3.

1C. Moore–Penrose, Newton and Complexity.

Newton's method can be generalized to search for zeros of maps $f : \mathbb{R}^n \to \mathbb{R}^m$, $n \geq m$, using the Moore–Penrose inverse of the derivative (as in Allgower–Georg [1993]).

We recall the definition of the Moore–Penrose inverse $A^\dagger$ of a surjective linear map $A : V \to W$ where V, W are finite dimensional vector spaces with inner products. $A^\dagger : W \to V$ is simply the inverse of A restricted to $(\ker A)^\perp$. It may also be described as the unique linear map $A^\dagger : W \to V$ satisfying $AA^\dagger = I$, and $A^\dagger A$ is the orthogonal projection onto $(\ker A)^\perp$. Note that $A^\dagger = A^*(AA^*)^{-1}$, A^* the adjoint.

Now we will define Newton's method for $f : \mathbb{R}^n \to \mathbb{R}^m$ (f could as well be defined on a domain of $\mathbb{R}^n$, or from $\mathbb{C}^n$ to $\mathbb{C}^m$). Let $N_f : \mathbb{R}^n \to \mathbb{R}^n$ be

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

and if x_0 is a given "starting point" in R^n, $x_i = N_f(x_{i-1})$. Note that N_f is well-defined at x provided $Df(x)$ is surjective. Moreover if $m = n$ N_f is the usual Newton method.

If $Df(x)$ is surjective then x is a fixed point of N_f if and only if $f(x) = 0$.

Proposition. *Suppose* 0 *is a regular value of* $f : \mathbb{R}^n \to \mathbb{R}^m$*. Let for* $\zeta \in f^{-1}(0)$, $W^s_\zeta = \{x \in \mathbb{R}^n \mid N^k_f(x) \to \zeta \text{ as } k \to \infty\}$*. By* N^k_f *we mean the* k^{th} *iterate of* N_f*. Then (a) the union over* $\zeta \in f^{-1}(0)$ *of* W^s_ζ *is a neighborhood of* $f^{-1}(0)$*, (b)* W^s_ζ *intersected with a small neighborhood of* $f^{-1}(0)$ *is a cell varying continuously in* ζ*, and (c)* $DN_f(\zeta)$ *restricted to* $\ker Df(\zeta)^\perp$ *is zero. The tangent space of* W^s_ζ *at* ζ *is the orthogonal complement to* $T_\zeta(f^{-1}(0)) = \ker Df(\zeta)$*.*

This extends the usual basin of attraction theory from the case $m = n$.

Proof of the Proposition. The existence and continuity of W^s_ζ are contained in Theorems 5.1 and 5.5 of Hirsch–Pugh–Shub [1977]. The fact that the union fills a neighborhood follows from a simple degree argument. We will make the size of the small neighborhood more precise in section 5, Proposition 2.

We now give the complexity theory for using this method for finding zeros of analytic $f : \mathbb{R}^n \to \mathbb{R}^m$, generalizing Smale [1987], Bez I. Define for $x \in \mathbb{R}^n$,

$$\beta(f, x) = \|Df(x)^\dagger f(x)\| \text{ (or } \infty \text{ if } Df(x) \text{ is not surjective)}$$
$$\gamma(f, x) = \max_{k>1} \left\| Df(x)^\dagger \frac{D^k f(x)}{k!} \right\|^{\frac{1}{k-1}} \text{ (or } \infty \text{ if } Df(x) \text{ is not surjective)}$$

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

Theorem C1. *There is a universal constant α_0 approximately $\frac{1}{7}$ with this property. If $f, x = x_0$, are as above with $\alpha(f,x) < \alpha_0$, then all the Newton iterates $x_1, x_2, \dots$ are defined, converge to $\zeta \in \mathbb{R}^n$ with $f(\zeta) = 0$ and for all $k \geq 1$*

$$(*) \qquad \|x_{k+1} - x_k\| \leq \left(\frac{1}{2}\right)^{2^k - 1} \|x_1 - x_0\|.$$

A point $x_0 \in \mathbb{R}^n$ is called an *approximate zero* of f if $(*)$ is satisfied. Then ζ is called the associated zero.

The proof of Theorem C1 is in section 4.

Imagine an operation (an ideal operation) which produces from an approximate zero, the associated actual zero. This could be done in the Blum–Shub–Smale [1989] model of computation with a "6^{th} type of node" for example. We will assume in our complexity estimates below that such an operation exists. Justification is based partly on the gain in conceptual simplicity of the results and simplicity in the arguments. Moreover the results in Bez I give a mathematical justification. Using the robust α theory of Bez I (Theorem 3, I-2 and section II-3), one can bypass the use of this 6^{th} node, at a cost of more technical work. Using these arguments it seems likely that the use of the 6^{th} node here could be eliminated, obtaining estimates with slightly worse constants.

Let us see how one can use Theorem C1 to get global complexity results. Consider $f_t : \mathbb{R}^n \to \mathbb{R}^m$, $y_t \in \mathbb{R}^m$, a homotopy and path respectively for $0 \leq t \leq 1$ and let $\zeta_0 \in \mathbb{R}^n$ satisfy $f_0(\zeta_0) = y_0$. Define

$$A_{t,t'} = \max_{\substack{x \text{ subject to} \\ f_t(x) = y_t}} \alpha(f_{t'} - y_{t'}, x).$$

Observe $A_{t,t} \equiv 0$.

Hypothesis. Suppose $A_{t,t'} < \alpha_0$ whenever $|t - t'| \leq \Delta = \frac{1}{k}$, k a positive integer.

Corollary of Theorem C1. *Let f_t, y_t, ζ_0 be as above and satisfy the hypothesis. Then a number of steps (6^{th} node) sufficient to solve $f_1(\zeta_1) = y_1$ is k of the Hypothesis.*

The proof is immediate. Let $t_0 = 0$, $t_i = t_{i-1} + \Delta$, so $\alpha(f_{t_i}, \zeta_{t_{i-1}}) < \alpha_0$. Then the 6^{th} node yields ζ_{t_i} from $\zeta_{t_{i-1}}$ starting from ζ_0 with $f(\zeta_{t_i}) = y_{t_i}$.

One may use Moore–Penrose in place of projective Newton for finding roots of $f \in \mathcal{H}_{(d)}$. In projective Newton, at $z \in \mathbb{C}^{n+1}$ one restricts $Df(z) : \mathbb{C}^{n+1} \to \mathbb{C}^n$ to the orthogonal space $N_z = \{v \in \mathbb{C}^{n+1},\ \langle v, z \rangle = 0\}$. In Moore–Penrose this is simply replaced by the orthogonal space of $\ker Df(z)$. For $(f, \zeta) \in V - \Sigma'$, N_ζ and the orthogonal space of $\ker Df(\zeta)$ coincide, but this is not the case in general.

The main theorem of Bez I estimates the number of steps needed in following ζ_t from ζ_0 where $f_t(\zeta_t) = 0$, and f_t is a curve in $\mathcal{H}_{(d)}$. The same estimate can be proved for the Moore–Penrose version.

Theorem C2. *Let $F_t = (f_t, \zeta_t)$ be a homotopy path in $\mathcal{H}_{(d)} \times \mathbb{C}^{n+1}$ (so $f_t(\zeta_t) = 0$), and ζ_0 satisfy $f(\zeta_0) = 0$. Then $k = CLD^{3/2}\mu^2$ (the greatest integer in) Moore–Penrose steps are sufficient to produce $\zeta_{t_1}, \zeta_{t_2}, \ldots, \zeta_{t_k}$ with $t_k = 1$ and so $f_1(\zeta_1) = 0$.*

Here $\mu = \max_t \mu_{\text{norm}}(f_t, \zeta_t)$ is the normalized condition number. L is the length of the curve f_t in $\mathcal{P}(\mathcal{H}_{(d)})$. We use the 6^{th} node and our proof (see section 4) does use a couple of results from Bez I, but is much shorter with the concepts more clearly exposed than the proof of the Main Theorem in Bez I.

Consider the complexity of the problem of following the curve $F^{-1}(0)$ where $F : \mathbb{R}^{n+1} \to \mathbb{R}^n$ has zero as a regular value. Here the algorithm has Moore–Penrose as one ingredient of a predictor-corrector method just as in Allgower–Georg [1993].

Theorem C3. *The complexity (number of predictor corrector steps) sufficient to follow an arc A of $F^{-1}(0)$, $F : \mathbb{R}^{n+1} \to \mathbb{R}^n$ as above, is $C\gamma L$, where L is the length of A, C a constant (not more than 20) and $\gamma = \max_{x \in A} \gamma(F, x)$.*

Theorem C3 yields a way of dealing with the problem of producing a complexity theory for zero-finding of real polynomial systems. The difficulty here is that the set of ill-posed problems has codimension 1 so that paths will generally have to meet that set. Now one could use the parameter t of the homotopy for the extra variable. If one wants to follow a zero of $f_t : \mathbb{R}^n \to \mathbb{R}^n$, just define $F : \mathbb{R}^{n+1} \to \mathbb{R}^n$ by $F(t, x) = f_t(x)$ where $\mathbb{R}^{n+1} = \mathbb{R} \times \mathbb{R}^n$. It would be interesting to see this idea carried out to obtain explicit bounds.

Theorem C3 is proved in section 4.

The preceeding results extend to Riemannian manifolds provided with a computation of the exponential map.

The following result has a different version in Theorem D1 of section 1D.

Theorem C4. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ have zero as a regular value and define $\gamma = \max_{z \in f^{-1}(0)} \gamma(F, z)$. Then there is a universal constant C so that if the distance $d(z', F^{-1}(0)) < \frac{C}{\gamma}$, then z' is an approximate zero.*

For the proof see section 4.

1D. Complexity and Condition Number.

Approximate zeros were defined without reference to any actual zero. The corresponding zero was then derived. We now define x_0 as an *approximate zero of the second kind* (as in Smale [1987]) for $f : \mathbb{R}^n \to \mathbb{R}^n$ provided there is some $\zeta \in \mathbb{R}^n$, $f(\zeta) = 0$, and

$$\|x_k - \zeta\| \leq \left(\frac{1}{2}\right)^{2^{k-1}} \|x_0 - \zeta\|, \quad k = 1, 2, \ldots$$
$$x_k = x_{k-1} - Df(x_{k-1})^{-1} f(z_{k-1}).$$

Theorem D1 [Smale 1986]**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}^n$*,* $x, \zeta \in \mathbb{R}^n$*,* $f(\zeta) = 0$ *and*

$$\|x - \zeta\|\gamma(f, \zeta) < \frac{3 - \sqrt{7}}{2}.$$

Then x *is an approximate zero of the second kind.*

We will suppose that our "6^{th} node" has the power of producing the actual zero from an approximate zero of the second kind.

Now consider the setting of the implicit function theorem $F : \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R}^n$ where F may also be defined on some domain or over $\mathbb{C}$. Suppose that

$$t \to (a(t), \zeta(t)) \in \mathbb{R}^k \times \mathbb{R}^n, \text{ with } F(a(t), \zeta(t)) = 0,$$

is a curve, $0 \le t \le 1$, such that $\frac{\partial F}{\partial \zeta}(a(t), \zeta(t))$ is non-singular for all t. The idea is that $a(t)$ is given explicitly (the input of a problem) and $\zeta(t)$ is given implicitly. Suppose that $\zeta(0)$ is also given and that we want to find $\zeta(1)$. This is a general setting for path following algorithms.

For our algorithm and complexity result, it is convenient to write $\zeta_t = \zeta(t)$ and $F_t(x) = F(a(t), x)$, so that $F_t(\zeta_t) = 0$, $0 \le t \le 1$.

The algorithm (slightly idealized) depends on a partition $t_0 = 0$, $t_i < t_{i+1}$, $t_k = 1$ $i = 0, 1, \ldots, k$ and the condition is that ζ_{t_i} is an approximate zero of the second kind for $F_{t_{i+1}}$ and $\zeta_{t_{i+1}}$. Thus k "6^{th} node" operations are sufficient to produce ζ_1 and so k is the complexity.

We may estimate k thus by Theorem D1. Accordingly, the required condition to implement the above procedure is:

$$\|\zeta_{t_{i+1}} - \zeta_{t_i}\|\gamma(F_{t_{i+1}}, \zeta_{t_{i+1}}) \le \frac{3 - \sqrt{7}}{2}.$$

Let $\Delta_{i+1} = \|\zeta_{t_{i+1}} - \zeta_{t_i}\|$, $\gamma_i = \gamma(F_{t_{i+1}}, \zeta_{t_{i+1}})$. So

$$(*) \qquad k = \sum_1^k \frac{\Delta_i}{\Delta_i} = c \sum_1^k \Delta_i \gamma_i,$$

where $c = \frac{2}{3 - \sqrt{7}}$, is sufficient. This yields:

Theorem D2. *Suppose that* $F_t(\zeta_t)$ *is as described above,* $\gamma = \max_t \gamma(F_t, \zeta_t)$ *and* L *is the length of the curve* $t \to \zeta_t$*. Then given* ζ_0*, a number of steps ("*6^{th} *node") sufficient to reach* ζ*, is* $\frac{2}{3-\sqrt{7}} L\gamma$*.*

Use $(*)$ and that $\sum \Delta_i \gamma_i \le \gamma \sum \Delta_i \le \gamma L$. A number of Newton steps without using the 6^{th} node could probably be estimated at about three times the above, using the robust α theory of Bez I. Then one would obtain an approximate zero of f_1 instead of ζ_1.

Theorem D2, while dealing with an idealized algorithm, is nice because of its extreme simplicity in statement and proof. It displays the main complexity ingredients.

The condition $(*)$ is sharper. For example if $\gamma(F_t, \zeta_t)$ is monotone, the complexity is bounded by $\int_0^1 \|\zeta_t'\| \gamma_t dt$.

In the main theorem of Bez I, the condition number $\mu(f, \zeta)$ turns up as the main ingredient. It is quite natural to ask why, since $\mu(f, \zeta)$ is an infinitesimal invariant reflecting other aspects of computation. We are now in a position to deal with that question.

Consider the environment of the previous theorem. One of the two complexity ingredients is L, the length of the curve ζ_t. This curve is only implicitly given, and hence L is also. It would be preferable to replace L by a more direct invariant of the input curve $a(t)$.

Here $F_t(\zeta_t) = F(a(t), \zeta(t)) = F(a(t), G(a(t)))$, and $G : U \to \mathbb{R}^n$ is defined on $a(t)$. Let L_a be the length of the curve $a(t)$. Recall that the condition number $\mu(a, \zeta)$, with $G(a) = \zeta$ at (a, ζ) is $\|DG(a)\|$ and so the condition number μ of $(a(t), \zeta(t))$ is $\max_t \mu(a(t), \zeta(t))$.

Proposition. $L \leq \mu L_a$.

Proof.

$$\begin{aligned} L &= \int_0^1 \|\zeta_t'\| dt = \int_0^1 \|G(a(t))'\| dt \\ &= \int_0^1 \|DG(a(t)) a'(t)\| dt \leq \int_0^1 \|DG(a(t))\| \, \|a'(t)\| dt \\ &\leq \mu \int_0^1 \|a'(t)\| dt = \mu L_a. \end{aligned}$$

The proposition and Theorem D2 yield the estimate:

$$(**) \qquad \text{complexity } k \leq \frac{2}{3 - \sqrt{7}} \gamma \mu L_a.$$

The last results give some way for taking advantage of sparsity in complexity estimates. In Bez I we found that for full systems, i.e. allowing all coefficients to be non-zero for systems of polynomials $f : \mathbb{C}^n \to \mathbb{C}^n$ of given degree $(d_1, \dots, d_n)$, the condition number μ squared was decisive. One factor of μ^2 came from an estimate on γ. The second factor of μ could be interpreted as the μ in $(**)$. But now in the theory just preceding, the sparse case leads to the condition number μ of $(**)$ which could be much smaller than the μ of Bez I.

We end section 1D with a couple of remarks.

(1) The condition matrix itself leads to an algorithm of predictor-corrector type, and its complexity may be estimated as above.

(2) Results here may be extended to maps of Riemannian manifolds using the exponential map extensively.

General Remark. The unitarily invariant norm on $\mathcal{H}_{(d)}$ used in Bez I, II, III, IV is described in detail by Weyl [1932] with his focus on explicit unitary invariance. Stein–Weiss [1971] use the same norm (and corresponding inner product), but don't discuss its unitary invariance. As we have mentioned earlier Eric Kostlan brought this approach to our attention.

2. Proof of Theorem A.

The proof uses the main results of Bez I and Bez II. First we follow Bez I to obtain: Theorem 1. We use some of the notation of Bez I. For example, we write $\mu_{\text{proj}}(f,\zeta)$ for $\mu_{\text{norm}}(f,\zeta)$, the normalized condition number.

Let $B_p(f,s)$ be the d_p ball of radius s around $f \in \mathcal{H}_{(d)} - \{0\}$, where d_p is the chordal metric

$$B_p(f,x) = \{g \in \mathcal{H}_{(d)} \mid g \not\equiv 0 \text{ and } d_p(f,g) < s\}.$$

Theorem 1. *Given $C_2 > 1$, $\exists C_1 > 1$ with the following property. If $g \in \mathcal{H}_{(d)}$, $g(x) = 0$ and $\mu_{\text{proj}}(g,x) < \infty$ then x may be continued to a zero $x(f)$ for all f in $B_p(g,s)$ where*

$$s = \frac{1}{C_1 D^{3/2} \mu^2_{\text{proj}}(g,x)}$$

and

$$\mu_{\text{proj}}(f, x(f)) \le C_2 \mu_{\text{proj}}(g,x).$$

Proof. We prove three preliminary propositions and a lemma.

Here we recall proposition 5(b) of §I-3 of Bez I.

Proposition 1. *Let $f, g \in \mathcal{H}_{(d)}$, then*

$$\mu_{\text{proj}}(f,x) \le \frac{\mu_{\text{proj}}(g,x)(1 + d_p(f,g))}{1 - D^{1/2} d_p(f,g) \mu_{\text{proj}}(g,x)}$$

as long as the denominator remains positive.

Lemma 1. *Let $f, g \in \mathcal{H}_{(d)}$. Let $g(x) = 0$. Then*
a) $\beta_0(f,x) \le \mu_{\text{proj}}(f,x) d_p(f,g)$
b) $\gamma_0(f,x) \le \frac{1}{2} \mu_{\text{proj}}(f,x) D^{3/2}$
c) $\alpha(f,x) \le \frac{1}{2} \mu_{\text{proj}}(f,x) d_p(f,g) D^{3/2}$.

Proof. From Propositions 2, 3 and Lemma 1 of §I-3 of Bez I

$$\beta_0(f,x) \le \mu_{\text{proj}}(f,x) \eta(f,x) \le \mu_{\text{proj}}(f,x) d_p(f,g)$$

which proves a), b) is Proposition 3 and c) results from multiplying a) and b).

Proposition 2. *There is a constant $K_1 > 0$ such that, if $g \in \mathcal{H}_{(d)}$ and $g(x) = 0$ then x may be continued to a zero $x(f)$ of f for all $f \in B_p(g, s)$ where $s = \frac{K_1}{\mu^2_{\text{proj}}(g,x) D^{3/2}}$. Moreover given constants $K_2, K_3, K_4, K_5 > 0$, K_1 may be chosen small enough such that*

a) $\mu_{\text{proj}}(f, x) \le (1 + K_2)\mu_{\text{proj}}(g, x)$

b) $\frac{\|x(f)-x\|}{\|x\|} \le \frac{2K_3}{\mu_{\text{proj}}(g,x) D^{3/2}}$

c) $\left(\frac{\|x(f)\|}{\|x\|}\right)^{D-1} \le 1 + K_4$

d) $\beta_0(f, x) \le \frac{K_3}{\mu_{\text{proj}}(g,x) D^{3/2}}$

e) $\gamma_0(f, x) \le \frac{(1+K_2)}{2} \mu_{\text{proj}}(g, x) D^{3/2}$

f) $\frac{\|x(f)-x\|}{\|x\|} \cdot \gamma_0(f, x) \le K_5$.

Proof. By Proposition 1

a)

$$\begin{aligned}\mu_{\text{proj}}(f, x) &\le \frac{\mu_{\text{proj}}(g, x)\left(1 + \frac{K_1}{\mu^2_{\text{proj}}(g,x) D^{3/2}}\right)}{\left(1 - \frac{K_1}{\mu_{\text{proj}}(g,x) D}\right)} \\ &\le \mu_{\text{proj}}(g, x)\left(\frac{1 + K_1}{1 - K_1}\right) \le \mu_{\text{proj}}(g, x)(1 + K_2)\end{aligned}$$

for K_1 small enough.

d)

$$\begin{aligned}\beta_0(f, x) &\le \mu_{\text{proj}}(f, x) d_p(f, g) \text{ by lemma 1a)} \\ &\le \mu_{\text{proj}}(g, x)\left(\frac{1 + K_1}{1 - K_1}\right) \cdot \frac{K_1}{\mu^2_{\text{proj}}(g, x) D^{3/2}} \\ &= \frac{\left(\frac{1+K_1}{1-K_1}\right) K_1}{\mu_{\text{proj}}(g, x) D^{3/2}} < \frac{K_3}{\mu_{\text{proj}}(g, x) D^{3/2}}\end{aligned}$$

for K_1 small enough.

e)

$$\begin{aligned}\gamma_0(f, x) &\le \frac{1}{2}\mu_{\text{proj}}(f, x) D^{3/2} \text{ by lemma 1b)} \\ &\le \frac{1}{2}\mu_{\text{proj}}(g, x)\left(\frac{1 + K_1}{1 - K_1}\right) \cdot D^{3/2} \le \frac{(1 + K_2)}{2}\mu_{\text{proj}}(g, x) D^{3/2}\end{aligned}$$

for K_1 small enough.

Thus $\alpha(f, x) = \beta_0(f, x)\gamma_0(f, x) \le \frac{1}{2}\left(\frac{1+K_1}{1-K_1}\right)^2 K_1$ so for K_1 small enough $\alpha(f, x) < \alpha_0$. $x(f)$ is then defined as the associated zero of $f \mid N_x$ for Newton's method with starting point x.

b) By d)

$$\beta(f,x) \le \frac{\left(\frac{1+K_1}{1-K_1}\right) K_1 \|x\|}{\mu_{\text{proj}}(g,x) D^{3/2}}$$

Thus by P.E., or Bez I,

$$\|x(f) - x\| \le \frac{2\left(\frac{1+K_1}{1-K_1}\right) K_1 \|x\|}{\mu_{\text{proj}}(g,x) D^{3/2}}$$

c)

$$\begin{aligned} \left(\frac{\|x(f)\|}{\|x\|}\right)^{D-1} &\le \left(\frac{\|x(f)-x\|}{\|x\|} + 1\right)^{D-1} \\ &\le \left(\frac{2\frac{1+K_1}{1-K_1} K_1}{D^{3/2}} + 1\right)^{D-1} \\ &\le e^{2\left(\frac{1+K_1}{1-K_1}\right) K_1} \end{aligned}$$

choose K_1 sufficiently small so that $e^{2\left(\frac{1+K_1}{1-K_1}\right) K_1} < 1 + K_4$.

f)

$$\begin{aligned} \frac{\|x(f)-x\|}{\|x\|}\gamma_0(f,x) &\le \frac{2K_3}{\mu_{\text{proj}}(g,x)D^{3/2}} \cdot \left(\frac{1+K_2}{2}\right) \mu_{\text{proj}}(g,x) D^{3/2} \\ &= K_3(1+K_2) \text{ by b) and e)} \end{aligned}$$

and we may choose K_3, K_2 sufficiently small so that $K_3(1+K_2) < K_5$.

Proposition 3. *Let $f \in \mathcal{H}_{(d)}$, $x \in \mathbb{C}^{n+1}$ and $y \in N_x = x + \operatorname{Null} x$. Then*

$$\mu_{\text{proj}}(f,y) \le \kappa \frac{(1-u)^2}{\psi(u)} \mu_{\text{proj}}(f,x) \left(\frac{\|y\|}{\|x\|}\right)^{D-1}$$

where $r_0 = \frac{\|y-x\|}{\|x\|}$, $u = r_0 \gamma_0(f \mid N_x, x)$, and

$$\kappa = \frac{(1+r_0^2)^{1/2}}{1 - r_0\left(\frac{(2-u)u}{(1-u)^2} + D\mu_{\text{proj}}(f,x) n(f,x)\right)\frac{(1-u)^2}{\psi(u)}}$$

as long as $0 \le u < 1 - \frac{\sqrt{2}}{2}$ and r_0 is small enough so that the denominator of κ remains positive.

Proof.

$$\begin{aligned} \mu_{\text{proj}}(f,y) &= \|f\| \, \|(Df_y \mid \text{Null}_y)^{-1}(\Delta(d_i^{1/2}\|y\|^{d_i-1})\| \\ &\le \|f\| \, \|(Df_y \mid \text{Null}_y)^{-1}(Df_y \mid \text{Null}_x)\| \, \|(Df_y \mid \text{Null}_x)^{-1}(Df_x \mid \text{Null}_x)\| \\ &\times \|(Df_x \mid \text{Null}_x)^{-1}\Delta(d_i^{1/2}\|x\|^{d_i-1})\| \, \|\Delta\left(\frac{\|y\|}{\|x\|}\right)^{d_i-1}\| \\ &\le \kappa \frac{(1-u)^2}{\psi(u)} \mu_{\text{proj}}(f,x) \left(\frac{\|y\|}{\|x\|}\right)^{D-1} \end{aligned}$$

by Proposition 1 §III-2 of Bez I, Lemma 3(2) of §II-1 of Bez I (initially from P.E.), the definition of μ_{proj} and the fact that the norm of a diagonal matrix is the largest norm of its entries.

Proof of Theorem 1. Choose $K_2, K_3, K_4, K_5, K_6, K_7, K_8 > 0$ so that $(1+K_2)(1+K_4)(1+K_6)(1+K_7) < C_2$, $\frac{(1+K_3^2)^{1/2}}{1-K_3\left(\left(\frac{2-K_8}{(1-K_8)^2}\right)K_8+1\right)\frac{(1-K_8)^2}{\psi(K_8)}} < (1+K_7)$, $\frac{(1-K_8)^2}{\psi(K_8)} < 1+K_6$, and $0 \le K_8 \le K_5$.

Let $C_1 = \frac{1}{K_1}$ where K_1 of Proposition 2 is chosen with respect to K_2, K_3, K_4, K_5. Then in Proposition 3 with $y = x(f)$

$$\kappa < (1+K_7), \quad \frac{(1-u)^2}{\psi(u)} < 1+K_6, \quad \frac{\|x(f)\|^{D-1}}{\|x\|} \le 1+K_4$$

and $\mu_{\text{proj}}(f,x) \le (1+K_2)\mu_{\text{proj}}(g,x)$. Thus

$$\begin{aligned}\mu_{\text{proj}}(f,x(f)) &\le (1+K_7)(1+K_6)(1+K_2)\mu_{\text{proj}}(g,x)(1+K_4)\\ &< C_2\mu_{\text{proj}}(g,x).\end{aligned}$$

□

Corollary of Theorem 1. *Let L be a great circle in $\mathcal{H}_{(d)}$ and N_ρ the special neighborhood of Σ in $\mathcal{H}_{(d)}$ as in Bez I. Then there is a universal constant c so that if $L \cap N_\rho \ne \emptyset$, then* $\text{Vol}(L \cap N_{2\rho}) \ge \frac{c\rho^2}{D^{3/2}}$.

*The "*Vol*" is the measure of a subset of the circle and a great circle is just the intersection of a real 2-dimensional linear subspace with unit sphere.*

By Theorem 1, there is a constant C_1 such that $\mu(f) = \mu(f,y_i) \le 2\mu(g,x_i) \le 2\mu(g)$ for all f with $d_p(f,g) \le \frac{1}{C_1(\mu(g))^2 D^{3/2}}$. Therefore, if $\mu(g) \le \frac{1}{2\rho}$ then $\mu(f) \le \frac{1}{\rho}$ for all f such that

$$d_p(f,g) \le \frac{1}{C_1\left(\frac{1}{2\rho}\right)^2 D^{3/2}} = \frac{4\rho^2}{C_1 D^{3/2}}.$$

Now let $f \in L \cap N_\rho$, i.e. $\mu(f) > \frac{1}{\rho}$ and $g \in L \cap (N_{2\rho})^c$ (using c for the complement) so $\mu(g) \le \frac{1}{2\rho}$. Then $d_p(f,g) > \frac{4\rho^2}{C_1 D^{3/2}}$. Hence if $f \in L \cap N_\rho$, then the interval of d_p length $\frac{4\rho^2}{C_1 D^{3/2}}$ around f in L is contained in $N_{2\rho}$. Since $d_p = \sin d_R$ this interval has Riemannian length greater than $\frac{4\rho^2}{C_1 D^{3/2}}$ and so that its "volume" is greater than $\frac{c\rho^2}{D^{3/2}}$. □

Let S be the unit sphere in Euclidean space $\mathbb{E}$ of some dimension and $\mathcal{L}$ be the space of great circles of S. The orthogonal group $\mathcal{O}$ of $\mathbb{E}$ acts isometrically on the product $S \times \mathcal{L}$ by $(x,L) \to (Ox, OL)$ for $O \in \mathcal{O}$. The subspace

$$V = \{(x,L) \in S \times \mathcal{L} \mid x \in L\}$$

is invariant under this action, and $\mathcal{O}$ acts transitively on V. For $x \in S$, let $\mathcal{L}_x = \{L \in \mathcal{L} \mid x \in L\}$.

Proposition 4. *Let $\mathcal{U} \subset S$, $W \subset \mathcal{L}$ be open sets. Then*

(a) *there is an $x \in S$ such that*

$$\frac{\operatorname{Vol}(\mathcal{L}_x \cap W)}{\operatorname{Vol}\mathcal{L}_x} \leq \frac{\operatorname{Vol} W}{\operatorname{Vol}\mathcal{L}}.$$

(b) $\int_{L\in\mathcal{L}} \operatorname{Vol}(\mathcal{U} \cap L) = \frac{\operatorname{Vol}\mathcal{L}\operatorname{Vol}L_0\operatorname{Vol}\mathcal{U}}{\operatorname{Vol}S}$.

Here L_0 is just a standard great circle and $\operatorname{Vol} L_0 = 2\pi$.

For the proof of the proposition let $\pi_1 : V \to S$, $\pi_2 : V \to \mathcal{L}$ be the restrictions of the projections and $\mathcal{U}' = \pi_1^{-1}\mathcal{U}$, $W' = \pi_2^{-1}W$.

First we prove 4(b). Let $x_0 \in S$ and $\mathcal{L}_0 = \mathcal{L}_{x_0}$. Let NJ_1 and NJ_2 be the normal Jacobians of the maps π_1 and π_2 respectively (from the coarea formula as in Bez II). These are constants by the orthogonal invariance. We use $\mathcal{X}$ to denote the characteristic function.

Lemma.

$$\int_{\mathcal{L}}\int_{\pi_2^{-1}L} \mathcal{X}(\mathcal{U}' \cap \pi_2^{-1}L) = \left(\frac{NJ_2}{NJ_1}\right)\operatorname{Vol}(\mathcal{U})\operatorname{Vol}(\mathcal{L}_0).$$

Proof of the Lemma.

$$\int_{\mathcal{L}}\int_{\pi_2^{-1}L} \frac{1}{NJ_2}\mathcal{X}(\mathcal{U}' \cap \pi_2^{-1}L) = \int_V \mathcal{X}(\mathcal{U}') =$$
$$\int_S \frac{1}{NJ_1}\int_{V_x} \mathcal{X}(\mathcal{U}' \cap V_x) = \frac{1}{NJ}\operatorname{Vol}(\mathcal{U})\operatorname{Vol}(\mathcal{L}_0)$$

where V_x is the fiber over x of π_1, proving the lemma.

In the lemma consider the special case $\mathcal{U} = S$, $\mathcal{U}' = V$ to obtain that

$$\frac{NJ_2}{NJ_1} = \frac{\operatorname{Vol}(\mathcal{L})\operatorname{Vol}(L_0)}{\operatorname{Vol}(\mathcal{L}_0)\operatorname{Vol}(S)}.$$

Observe that

$$\int_{\pi_2^{-1}L} \mathcal{X}(\mathcal{U}' \cap \pi_2^{-1}L) = \operatorname{Vol}(\mathcal{U} \cap L).$$

Putting this, and our evaluation of $\frac{NJ_2}{NJ_1}$ into the lemma yields Proposition 4b.

Now for part (a) of the proposition. The coarea formula gives

$$\int_V \mathcal{X}(W') = \int_S \frac{1}{NJ_1}\int_{V_x} \mathcal{X}(W' \cap V_x)$$

and

$$\int_V \mathcal{X}(W') = \int_{\mathcal{L}} \frac{1}{NJ_2} \int_{\pi_2^{-1}L} \mathcal{X}(W' \cap L) = \frac{1}{NJ_2} \operatorname{Vol} W \operatorname{Vol} L_0.$$

Thus

$$\begin{aligned}\int_S \int_{V_x} \mathcal{X}(W' \cap V_x) &= \frac{NJ_1}{NJ_2} \operatorname{Vol} W \operatorname{Vol} L_0 \\ &= \frac{\operatorname{Vol} \mathcal{L}_0 \operatorname{Vol} S \operatorname{Vol} W}{\operatorname{Vol} \mathcal{L}}\end{aligned}$$

by the computation of $\frac{NJ_2}{NJ_1}$ above. Therefore

$$\begin{aligned}\frac{\operatorname{Vol} W}{\operatorname{Vol} \mathcal{L}} &\geq \frac{\min_{x \in S} \int_{V_x} \mathcal{X}(W' \cap V_x)}{\operatorname{Vol} \mathcal{L}_0} \\ &= \frac{\operatorname{Vol}(\mathcal{L}_x \cap W)}{\operatorname{Vol} \mathcal{L}_x}\end{aligned}$$

for some particular x. This proves Proposition 4a.

We now apply Proposition 4 to the case where $\mathbb{E} = \mathcal{H}_{(d)}$, with $\mathcal{L}, S$ defined for this specialization. Then for $g \in S$ (i.e. $g \in \mathcal{H}_{(d)}$, $\|g\| = 1$), $\mathcal{L}_g$ is the space of all great circles in $\mathcal{H}_{(d)}$ containing g. Define

$$\mathcal{L}_{\rho,g} = \{L \in \mathcal{L}_g \mid L \cap N_\rho \neq \emptyset\}.$$

Theorem 2. *There is a constant $c > 0$ such that for each $d = (d_1, \ldots, d_n)$, there is a $g \in \mathcal{H}_{(d)}$ of norm 1 and*

$$\frac{\operatorname{Vol} \mathcal{L}_{\rho,g}}{\operatorname{Vol} \mathcal{L}_g} \leq c\rho^2 n^2 (n+1)(N-1)(N-2) D^{3/2} \mathcal{D}.$$

For the proof we use Proposition 4 taking $\mathcal{U} = N_{2\rho}$, and $W = \{L \in \mathcal{L} \mid L \cap N_\rho \neq \emptyset\}$. Using the Corollary of Theorem 1,

$$\operatorname{Vol} W \frac{c\rho^2}{D^{3/2}} \leq \int_{\mathcal{L}_\rho} \operatorname{Vol}(L \cap N_{2\rho}) \leq \int_{\mathcal{L}} \operatorname{Vol}(L \cap N_{2\rho})$$

which by Proposition 4b is

$$= \frac{\operatorname{Vol} \mathcal{L} \operatorname{Vol} L_0 \operatorname{Vol} N_{2\rho}}{\operatorname{Vol} S}.$$

By Proposition 4a there is $x = g$ so

$$\frac{\operatorname{Vol}(\mathcal{L}_x \cap W)}{\operatorname{Vol} \mathcal{L}_x} \leq \left(\frac{c\rho^2}{D^{3/2}}\right)^{-1} \frac{\operatorname{Vol} L_0 \operatorname{Vol} N_{2\rho}}{\operatorname{Vol} S}.$$

Now use Theorem C of Bez II. It yields (for $n > 1$):

$$\frac{\operatorname{Vol} N_{2\rho}}{\operatorname{Vol} S} \leq 4\rho^4 n^2(n+1)(N-1)(N-2)\mathcal{D}.$$

Joining this with the previous estimate yields Theorem 2 (note that $\mathcal{L}_x \cap W = \mathcal{L}_{\rho,g}$). The case of $n = 1$ is implied by Theorem D of Bez II.

Proof of Theorem A. Fix g as in Theorem 2. Give $f \in P(\mathcal{H}_{(d)})$, the homotopy $0 \leq t \leq 1$, $f_t = tf + (1-t)g$ has length less than or equal to one. Let $\rho(f, g) = \sup_\rho \{f_t\} \cap N_\rho = \emptyset$. Then by the main theorem of Bez I, $\frac{c_2 D^{3/2}}{\rho^2(f,g)}$ projective Newton steps are sufficient to find all the approximate zeros of f, with c_2 around 10. Now by Theorem 2 the probability that $\rho(f, g) > s$ is greater than or equal to the probability that the great circle through $f \cap N_s = \emptyset \geq 1 - cs^2 n^2(n+1)(N-1)(N-2)D^{3/2}\mathcal{D}$.

Thus the probability that $\frac{c_2 D^{3/2}}{s^2}$ steps suffice is greater than or equal to $1 - cs^2 n^2(n+1)(N-1)(N-2)D^{3/2}\mathcal{D}$ setting $t = cs^2 n^2(n+1)(N-1)(N-2)D^{3/2}\mathcal{D}$ we find that with probability $1 - t$, $\frac{c_3 n^2(n+1)(N-1)(N-2)D^3\mathcal{D}}{t}$ steps suffice.

3. Proof of the Second Condition Number Theorem.

Note that $V_{v,\lambda}$ is the set of M having v as a λ eigenvector and that if $M \in \Sigma' \cap V_{v,\lambda}$, then λ is a multiple eigenvalue.

The unitary group $\mathcal{U}(n)$ acts on $P(\mathbb{C}^n)$ in the natural way and acts on $\mathcal{M}(n)$ by sending $A \to UAU^{-1}$ for $U \in \mathcal{U}(n)$. Moreover if $(M, v, \lambda) \in V$, $Mv = \lambda v$ so, $UMU^{-1}(Uv) = \lambda(Uv)$. Thus V is invariant under the product action of $\mathcal{U}(n)$. Since $\pi_1 : V \to \mathcal{M}(n)$, $\pi_2 : V \to P(\mathbb{C}^n) \times \mathbb{C}$ both commute with the action of $\mathcal{U}(n)$, it follows that $K_i(M, v, \lambda)$, i 1, 2 are also invariant under the action of $\mathcal{U}(n)$. Thus $K_i(M, v, \lambda)$, $i = 1, 2$ only depends on the linear map $M \mid v^\perp$, where $v^\perp$ is the Hermitian complement of v in $\mathbb{C}^n$, and not on a particular basis.

Let $\hat{M} = \pi_{v^\perp} M \mid v^\perp$ and $M_1 = \pi_v M \mid v^\perp$ where $\pi_{v^\perp}$ and π_v are the orthogonal projections of $\mathbb{C}^n$ onto $v^\perp$ and v respectively. Since $\mathcal{U}(n)$ acts transitively on $P(\mathbb{C}^n)$ we may assume for proof that $v = (1, 0, \ldots, 0)$ and thus that $(M, v, \lambda) \in M_{v,\lambda}$ has the form $M = \begin{pmatrix} \lambda & M_1 \\ 0 & \hat{M} \end{pmatrix}$. In general we assume $\|v\| = 1$.

Lemma 1.
a) $K_1(M, v, \lambda)\dot{M} = (\lambda I_{n-1} - \hat{M})^{-1} \pi_{v^\perp} \dot{M}(v)$
b) $K_2(M, v, \lambda)\dot{M} = \frac{\langle y, \dot{M}v \rangle}{\langle y, v \rangle}$
where y satisfies $(M^* - \lambda I)y = 0$, M^* the adjoint of M.

Proof. The equations

$$(*) \qquad (\dot{\lambda} I - \dot{M})v + (\lambda I - M)\dot{v} = 0$$

(∗∗) $$\langle \dot v, v\rangle = 0$$

define $(\dot M, \dot v, \dot\lambda) \in T_{(M,v,\lambda)}V \subset T_M(\mathcal{M}(n)) \times T_v(P(\mathbb{C}^n)) \times \mathbb{C}$. Applying $\pi_{v^\perp}$ to (∗) gives

$$\pi_{v^\perp}\dot M(v) = \pi_{v^\perp}(\lambda I - M)\dot v,$$

and since $\dot v \in v^\perp$ by (∗∗) $\pi_{v^\perp}\dot M(v) = (\lambda I - \hat M)\dot v$ proving a).

For b) note that $\langle y, (\lambda I - M)u\rangle = 0$ for all $u \in \mathbb{C}^n$. Take the inner product of (∗) with y to obtain $\langle y, (\dot\lambda I - \dot M)v\rangle = 0$ so that

$$\dot\lambda = \frac{\langle y, \dot M v\rangle}{\langle y, v\rangle} = K_2(M, v, \lambda).$$

Lemma 2.

a) $\|K_1(M_1v, \lambda)\| = \|(\lambda I_{n-1} - \hat M)^{-1}\|$

b) $\|K_2(M_1v, \lambda)\| \le (1 + \|M\|^2\|(\lambda I_{n-1} - \hat M)^{-1}\|^2)^{1/2}$.

Proof. Assume $M = \begin{pmatrix} \lambda & M_1 \\ 0 & \hat M \end{pmatrix}$, $v = (1, 0, \ldots, 0)$ and write $\dot M = \begin{pmatrix} \dot\lambda & \dot M_1 \\ \dot M_2 & \dot{\hat M} \end{pmatrix}$. Then for a), $\dot v = 0$ for $\dot M$ of the form $\begin{pmatrix} \dot\lambda & \dot M_1 \\ 0 & \dot{\hat M} \end{pmatrix}$ since the eigenvector v is constant for these perturbations. Thus $K_1(M_1, v, \lambda)\dot M$ factors through projections on $\dot M_2$ and $\|\dot M_2\| = \|\pi_{v^\perp}\dot M_2(v)\|$.

For b) y may be taken as

(∗) $$y = v + (\lambda I_{n-1} - \hat M)^{*-1}(M^* - \lambda I)v,$$

for Image$(M^* - \lambda I) \subset \operatorname{Ker}(M - \lambda I)^\perp \subset v^\perp$. Applying $(M^* - \lambda I)$ to both sides of (∗) gives

$$\begin{aligned}(M^* - \lambda I)y &= (M^* - \lambda I)v + (M^* - \lambda I)(\lambda I_{n-1} - \hat M)^{*-1}(M^* - \lambda I)v \\ &= (M^* - \lambda I)v - (M^* - \lambda I)v = 0,\end{aligned}$$

using $(M^* - \lambda I)(\lambda I_{n-1} - \hat M)^{*-1} = -I_{n-1}$. Now from (∗)

$$\begin{aligned}\|y\|^2 &\le 1 + \|(\lambda I - \hat M)^{*-1}\|^2\|(M^* - \lambda I)v\|^2 \\ &\le 1 + \|(\lambda I - \hat M)^{*-1}\|^2\|M_1^* v\|^2 \\ &\le 1 + \|(\lambda I - \hat M)\|^2\|M\|^2\end{aligned}$$

using $\|A^*\| = \|A\|$. Also from (∗) it follows that $\langle y, v\rangle = \langle v, v\rangle = 1$. Thus

$$\|K_2(M, v, \lambda)\dot M\| = \left|\frac{\langle y, \dot M v\rangle}{\langle y, v\rangle}\right| \le \|y\|\,\|\dot M\|$$

using Lemma 1, Cauchy–Schwarz and the above. The estimate of $\|y\|^2$ finishes the proof of b).

For the proof of the theorem we also require

Proposition 1 (Eckart and Young [1936]). $\|A^{-1}\| = \frac{1}{d_F(A,S)}$ *where* $A \in \mathcal{M}(n)$, $S \subset \mathcal{M}(n)$ *is the set of singular matrices and* d_F *is the Frobenius distance.*

Proof of Theorem. Assume $M = \begin{pmatrix} \lambda & M_1 \\ 0 & \hat{M} \end{pmatrix}$, $v = (1, 0, \ldots, 0)$. Then it is simple to see that the closest matrix in $V_{v,\lambda} \cap \Sigma'$ is $N = \begin{pmatrix} \lambda & M_1 \\ 0 & \hat{N} \end{pmatrix}$ where $\hat{N}$ is the closest matrix in $\mathcal{M}(n-1)$ to $\hat{M}$ such that $\lambda I_{n-1} - \hat{N}$ is singular. That is,

$$\begin{aligned} d_{v,\lambda}((M, v, \lambda), \Sigma' \cap V_{v,\lambda}) &= d_F(\hat{M}, \hat{N}) \\ &= d_F(\lambda I_{n-1} - \hat{M}, S) \\ &= \|(\lambda I_{n-1} - \hat{M})^{-1}\|^{-1} \end{aligned}$$

by Proposition 1. Here S is the set of singular $(n-1) \times (n-1)$ matrices. Now Lemma 2 finishes the proof of the theorem.

4. The Proofs for Section 1C.

The proof of Theorem C1 is adapted from Smale [1986], or P.E.

Lemma. *Let* $f : \mathbb{R}^n \to \mathbb{R}^m$, $z, z' \in \mathbb{R}^n$ *and* $u = \|z' - z\|\gamma(f, z)$. *Let* $V_x = \ker Df(x)^\perp$ *and* $\pi_x : \mathbb{R}^n \to V_x$ *be the orthogonal projection. Suppose* $u < 1 - \frac{\sqrt{2}}{2}$. *Then*

(a)
$$\|Df(z)^\dagger Df(z') - \pi_z\| \le \left(\frac{1}{1-u}\right)^2 - 1 < 1$$

(b)
$$\|Df(z')|_{V_z}^{-1} Df(z)|_{V_z}\| \le \frac{(1-u)^2}{\psi(u)}$$

where $\psi(u) = 2u^2 - 4u + 1$

(c)
$$\|Df(z')^\dagger Df(z)\| \le \frac{(1-u)^2}{\psi(u)}.$$

Proof of the Lemma. For (a) expand $Df(z')$ by the Taylor series

$$Df(z') = Df(z) + \sum_{k=2} \frac{D^k f(z)}{(k-1)!}(z' - z)^{k-1}.$$

Apply $Df(z)^\dagger$, noting $Df(z)^\dagger Df(z) = \pi_z$ to obtain

$$\|Df(z)^\dagger Df(z') - \pi_z\| \le \sum_{k=1} k u^{k-1} - 1 = \left(\frac{1}{1-u}\right)^2 - 1$$

(compare P.E.).

Next note that

$$\|Df(z)|_{V_z}^{-1} Df(z')|_{V_z} - I_{V_z}\| \leq \|Df(z)^{\dagger} Df(z') - \pi_z\|,$$

so as in P.E., (b) follows. Here $I_{V_z} : V_z \to V_z$ is the identity. Finally (c) is a consequence of the fact that for any surjective linear $A : \mathbb{R}^{n+1} \to \mathbb{R}^n$ and hyperplane $V \subset \mathbb{R}^{n+1}$, $v \in \mathbb{R}^n$, $\|A|_{\ker A^{\perp}}^{-1} v\| \leq \|A|_V^{-1} v\|$.

Using the lemma, the proof of Theorem C1 follows just as in P.E.

We now prove Theorem C2. This uses Moore–Penrose (i.e., Newton's method extended by Moore–Penrose) to follow a curve (f_t, ζ_t) in $\mathcal{H}_{(d)} \times \mathbb{C}^{n+1}$ with $f_t(\zeta_t) = 0$, using $x'_{i+1} = N_{f_{t_{i+1}}}(x_i)$. Using the sixth node we may take $x_i = \zeta_{t_i}$, each $i = 0, 1, \ldots, k$ provided that $\alpha(f_t, \zeta_{t_i}) < \alpha_0$ for $t_i \leq t \leq t_{i+1}$.

By Bez I, especially the higher derivative estimate,

$$\alpha(f_t, \zeta_{t_i}) \leq \mu_{\text{norm}}(f_t, \zeta_{t_i})^2 \frac{\eta D^{3/2}}{2}$$

where $\eta \leq d_p(f_t, f_{t_i}) = d_p$, is as in Bez I and $d_p(f,g) = \sin d(f,g)$, $d(f,g)$ the Riemannian distance in $P(\mathcal{H}_{(d)})$.

Use Proposition 5, see I-3, Bez I to obtain

$$\begin{aligned} \mu_{\text{norm}}(f_t, \zeta_{t_i}) &\leq \frac{\mu_{\text{norm}}(f_{t_i}, \zeta_{t_i})(1 + d_p)}{1 - D^{1/2} d_p \mu_{\text{norm}}(f_{t_i}, \zeta_{t_i})} \\ &\leq \frac{\mu(1 + d_p)}{1 - D^{1/2} d_p \mu}, \ \mu = \max_t \mu_{\text{norm}}(f_t, \zeta_t). \end{aligned}$$

To apply Theorem C1, we thus need

$$\left(\frac{\mu(1 + d_p)}{1 - D^{1/2} d_p \mu} \right)^2 \frac{D^{3/2}}{2} d_p < \alpha_0.$$

There is a universal constant c which makes $d_p \leq \frac{c}{\mu^2 D^{3/2}}$ sufficient. Thus we obtain the complexity $k = \frac{1}{c} \mu^2 D^{3/2} L$. □

Proof of Theorem C3. Let zero be a regular value of $F : \mathbb{R}^{n+1} \to \mathbb{R}^n$, $u \in F^{-1}(0)$, $\dot{u} \in T_u(F^{-1}(0))$, and $\|\dot{u}\| = 1$. Suppose that

$$(*) \qquad \alpha(F, u + h\dot{u}) < \alpha_0.$$

Then the predictor-corrector algorithm with the 6^{th} node produces

$$u \to u + h\dot{u} \to u' \in F^{-1}(0)$$

where u' is the zero associated to the approximate zero $u + h\dot{u}$.

The estimate $(*)$ has two parts, the estimate for $\gamma(F, u + h\dot{u})$ and for $\beta(F, u + h\dot{u})$. One uses Lemma 2c of P.E., extended by Moore–Penrose, to see that

$$\gamma(F, u + h\dot{u}) \leq c\gamma, \text{ for } h\gamma \leq c_1$$

and for all $u \in A \subset F^{-1}(0)$.

The Taylor expansion of F about u yields

$$F(u + h\dot{u}) = \sum_{k=2} \frac{D^k F(u)(h\dot{u})^k}{k!}$$

since the first two terms are zero. Compose with $DF(u + h\dot{u})^\dagger$ to obtain

$$\beta(F, u + h\dot{u}) \leq \sum_{k=2} \frac{\|DF(u + h\dot{u})^\dagger DF(u) DF(u)^\dagger D^k F(u)\| h^k}{k!}.$$

Here we used $DF(u)DF(u)^\dagger = I$.

Lemma (c) from the beginning of this section applies to yield

$$\|DF(u + h\dot{u})^\dagger DF(u)\| \leq \frac{(1 - h\gamma)^2}{\psi(h\gamma)}, \; h\gamma \leq 1 - \frac{\sqrt{2}}{2}.$$

Thus we obtain $\beta(F, u + h\dot{u}) \leq ch^2\gamma$ and so $\alpha(F, u + h\dot{u}) \leq c(h\gamma)^2$ for $h\gamma \leq c'$. Thus $(*)$ is satisfied provided that $h\gamma$ is less than an easily estimated constant.

To finish the proof of Theorem C3 we need to relate h to the complexity. It is sufficient to show that $\|u' - u\| \geq ch$ where $u' \in F^{-1}(0)$ is obtained by the predictor-corrector step, and c, as usual, is a new universal constant. In fact it is sufficient to show that $\|(u + h\dot{u}) - u'\| \leq c_1 h$, c_1 a small universal constant. Since $u + h\dot{u}$ is an approximate zero, $\|u + h\dot{u} - u'\| \leq 2\beta(F, u + h\dot{u})$ and $\beta(F, u + hu) \leq c(h\gamma)h$. But $h\gamma$ can be assumed to be universally small as above. □

Proof of Theorem C4.

Lemma. *With the setting of the theorem, let* $u = \|z - z'\|\gamma(F, z)$ *where* $F(z) = 0$. *Then if* $\psi(u) > 0$, $\alpha(F, z') \leq \frac{u}{\psi(u)^2}$.

Proof of Lemma. Use Proposition 2 of II-1 of Bez I, extended to Moore–Penrose as before. Then

$$\alpha(F, z') \leq \frac{(1 - u)\alpha(F, z) + u}{\psi(u)^2}.$$

But $\alpha(F, z) = 0$ since $F(z) = 0$, proving the lemma.

There is a constant c such that if $u < c$, $\frac{u}{\psi(u)^2} < \alpha_0$. So if $z \in F^{-1}(0)$ and $\|z - z'\| < \frac{c}{\gamma}$, $\alpha(F, z') < \alpha_0$. □

Then we are finished by Theorem C1.

5. Estimates of DN_f.

Proposition 1. *Let f be an analytic map $f : \mathbb{R}^n \to \mathbb{R}^m$, $x \in \mathbb{R}^n$, $f(x) = 0$. Let $u = \|x-y\|\gamma(f,x)$. Then*

$$\|DN_f(y) - DN_f(x)\| \leq \frac{(1-u)^2}{\psi(u)}\left(\frac{1}{(1-u)^2} - 1\right) + \frac{2u}{\psi(u)^2}$$

as long as $\psi(u) > 0$, i.e. $u < 1 - \frac{\sqrt{2}}{2}$.

Proof.

$$\begin{aligned} DN_f(y) - DN_x(x) &= D(Df^\dagger \circ f)(y) - D(Df^\dagger \circ f)(x) \\ &= D(Df(y)^\dagger)f(y) - Df(y)^\dagger Df(y) \\ &\qquad - D(Df(y)^\dagger)f(x) + Df(x)^\dagger Df(x) \\ &= \underbrace{D(Df(y)^\dagger)f(y)}_{A} + \underbrace{Df(x)^\dagger Df(x) - Df(y)^\dagger Df(y)}_{B} \end{aligned}$$

we will prove in the lemmas below that

$$(*) \qquad \|A\| \leq \frac{2u}{\psi(u)^2}$$

and

$$(**) \qquad \|B\| \leq \frac{(1-u)^2}{\psi(u)}\left(\frac{1}{(1-u)^2} - 1\right).$$

Recall that $Df(x)^\dagger Df(x)$ is orthogonal projection on $\ker Df(x)^\perp$, and $Df(y)^\dagger Df(y)$ is orthogonal projection on $\ker Df(y)^\perp$. First we estimate the norm of $\sigma : \ker Df(x) \to \ker Df(x)^\perp$ such that $\ker Df(y) = \text{graph}(\sigma) = \{(v, \sigma(v)) \mid v \in \ker Df(x)\}$. Write $Df(x)$ as $(0, C)$

$$C : \ker Df^\perp \to H \qquad 0 : \ker Df \to H$$

the zero map and $Df(y) = (D_1, D_2)$ in these coordinates.

Lemma 1. $\|\sigma\| \leq \frac{(1-u)^2}{\psi(u)}\left(\frac{1}{(1-u)^2} - 1\right)$.

Proof. $\sigma = -D_2^{-1}D_1$ so $\|\sigma\| \leq \|D_2^{-1}C\|\,\|C^{-1}D_1\|$. Now $\|D_2^{-1}C\| \leq \frac{(1-u)^2}{\psi(u)}$ by Lemma b) §1c), and $\|C^{-1}D_1\| = \|Df(x)^\dagger \sum_2^\infty \frac{D^k f(x)(y-x)^{k-1}}{(k-1)!}\| \leq \frac{1}{(1-u)^2} - 1$.

Lemma 2. *Let $E \subset H_1 \times H_2$ be given as the graph of $\sigma : H_1 \to H_2$. Then $\|\pi_E - \pi_{H_1}\| \leq \|\sigma\|$ where π_E and π_H are orthogonal projection on E and H respectively.*

Proof. Let $A : V \to H_1 \times H_2$. Then orthogonal projection (Image A) is given by $A(A^*A)^{-1}A^*$, A^* the adjoint of A. Thus

$$\pi_E - \pi_{H_1} = \begin{pmatrix} (I+\sigma^*\sigma)^{-1} - I & (I+\sigma^*\sigma)^{-1}\sigma \\ \sigma(I+\sigma^*\sigma)^{-1} & \sigma(I+\sigma^*\sigma)^{-1}\sigma^* \end{pmatrix}$$

and the norm of the matrix as an operator is easily seen to be less than or equal to $\|\sigma\|$.

Lemma 1 and Lemma 2 prove $(**)$. Using that $\|\pi_{E^\perp} - \pi_{H_1^\perp}\| = \|(I-\pi_E)-(I-\pi_{H_1})\| = \|\pi_E - \pi_{H_1}\|$. Now we turn to $(*)$

$$\begin{aligned} D(Df(y)^\dagger)f(y) &= -Df^\dagger(y)(D^2f(y)Df(y)^* + D^2f(y)^*)(Df(y)Df(y)^*)^{-1}f(y) \\ &\quad + D^2f(y)^*(Df(y)Df(y)^*)^{-1}f(y) \\ &= \underbrace{-Df^\dagger(y)(D^2f(y)Df^*(y))(Df(y)Df(y)^*)^{-1}f(y)}_{E} \\ &\quad + \underbrace{(I - Df^\dagger(y)Df(y))}_{G}\underbrace{(D^2f(y)^*(Df(y)Df(y)^*)^{-1}f(y))}_{H} \end{aligned}$$

G is a projection so $\|G\| = 1$. We will prove $\|E\| < \frac{u}{\psi(u)^2}$ and $\|H\| < \frac{u}{\psi(u)^2}$.

Lemma 3. $\|E\| \leq \alpha(f, y)$.

Proof.

$$\begin{aligned} \|E\| &\leq \|Df(y)^\dagger D^2f(y)\| \, \|Df^*(Df(y)Df(y)^*)^{-1}f(y)\| \\ &\leq \gamma(f,y)\beta(f,y) = \alpha(f,y). \end{aligned}$$

The notation $D^2f(y)^*$ has the following interpretation $D^2f(y)(u, \ )$ for fixed u is linear. $D^2f(y)^*$ means the adjoint of this linear map.

Lemma 4. $\|D^2f(y)^*(Df(y)Df(y)^*)^{-1}f(y)\| \leq \alpha(f, y)$.

Proof.

$$\begin{aligned} \|D^2f(y)^*(Df(y)Df(y)^*)^{-1}f(y)\| &\leq \|D^2f(y)^*(Df(y)Df(y)^*)^{-1}Df(y)\| \times \\ &\quad \|Df(y)^*(Df(y)Df(y)^*)^{-1}f(y)\| \\ &\quad \text{using that } Df(y)Df(y)^*(Df(y)Df(y)^*)^{-1} = Id \\ &\leq \gamma(f,y)\beta(f,y) = \alpha(f,y) \text{ using that } \|A^*\| = \|A\| \text{ for linear } A. \end{aligned}$$

∎

Now the proof of Theorem C4 gives $\alpha(f,y) \le \frac{u}{\psi(u)^2}$ proving $(*)$ and the proposition.

We make part of Proposition §1C more precise for analytic $f : \mathbb{E} \to \mathbb{F}$, where $\mathbb{E}, \mathbb{F}$, are Hilbert spaces. Suppose 0 is a regular value for f and $\zeta \in f^{-1}(0)$. Write $\mathbb{E} = \ker Df(\zeta) \oplus \ker Df(\zeta)^{\perp}$. Let $S_1 = \zeta + \{(x,y) \in \ker Df(\zeta) \oplus \ker Df(\xi)^{\perp} \mid \|x\| \ge \|y\|\}$. Let $S_2 = \zeta + \{(x,y) \in \ker Df(\zeta) \oplus \ker Df(\xi)^{\perp} \mid \|x\| \le \|y\|\}$. Let $\zeta = (\zeta_1, \zeta_2)$ with respect to $\ker Df(\zeta) \oplus \ker Df(\zeta)^{\perp}$, and let $B_r(x)$ denote the ball of radious r centered at x.

Proposition 2. *There is a universal constant $c > 0$ such that for analytic f and ζ as above and $\gamma = \gamma(f,\zeta)$*

a) $V \equiv f^{-1}(0) \cap B_{\frac{c}{\gamma}}(\zeta) = \bigcap_{n \ge 0} N_f^n(S_1 \cap B_{\frac{c}{\gamma}}(\zeta))$

b) *V is the graph of C^1 function $\sigma_v : B_{\frac{c}{\gamma}}(\zeta_1) \to \ker Df(\zeta)^{\perp}$ and* $\|D\sigma_v(\zeta_1 + x)\| \le 3\|x\|\gamma$

c)

$$W^S_{\zeta,\mathrm{loc}} = \{(x,y) \in B_{\frac{c}{\gamma}}(\zeta) \mid N_f^n(x,y) \in B_{\frac{c}{\gamma}}(\zeta) \ \forall n > 0$$

and $N_f^n(x,y) \to \zeta$ *as* $n \to \infty\} = \bigcap_{n \ge 0} N_f^{-n}(S_2 \cap B_{\frac{c}{\gamma}}(\zeta))$

d) *$W^S_{\zeta,\mathrm{loc}}$ is the graph of a c^1 function*

$$\sigma_w : B_{\frac{c}{\gamma}}(\zeta_2) \to \ker Df(\zeta),$$
$$D\sigma_w(\zeta_2) \equiv 0 \ \textit{and} \sup_{y \in B_{\frac{c}{\gamma}}(\zeta_2)} \|D\sigma_w(y)\| \le 1.$$

Proof. To see V as a graph over $B_{\frac{c}{\gamma}}(\zeta_1)$ restrict f to $(\zeta_1 + x) \times \ker Df(\zeta)^{\perp}$ and apply Lemma 1b §1C to deduce that $\alpha(f \mid (\zeta_1 + x) \times \ker Df(\zeta)^{\perp}, \zeta_1 + x) \le \frac{u}{\psi(u)^2}$. Now choose u small enough so that this quantity is less than α_0. Now Hirsch–Pugh–Shub [1977] Theorem 5.1, the use of a bump function and remarks on center manifolds finishes the rest of the proof. The estimate of $\|D\sigma_v(\zeta_1 + x)\|$ follows from Lemma 1.

References

Allgower, E. and Georg, K., *Continuation and path following*, Acta Numerica (1993), 1–64.

Blum, L., Shub, M. and Smale, S., *On a theory of computation and complexity over the real numbers: NP-completeness recursive functions and universal machines*, Bulletin of the Amer. Math. Soc. **21** (1989), 1–46.

Conway, J. and Sloane, N., *Sphere Packings, Lattices and Groups*, Springer, New York, 1988.

Demmel, J., *On condition numbers and the distance to the nearest ill-posed problem*, Numerische Math. **51** (1987), 251–289.

Demmel, J., *The probability that a numerical analysis problem is difficult*, Math. Comp. **50** (1988), 449–480.

Eckart, C. and Young, G., *The approximation of one matrix another of lower rank*, Psychometrika **1** (1936), 211–218.

Edelman, A., Kostlan, E. and Shub, M., *How many eigenvalues of a random matrix are real?*, preprint, Berkeley (1992).

Hirsch, M., Pugh, C. and Shub, M., *Invariant Manifolds*, Springer Lecture Notes # 583, Springer, New York, 1977.

Kim, M. and Sutherland, S., *Families of parallel root-finding algorithms*, preprint, SUNY Stonybrook, Inst. of Math. Sci., Stonybrook, New York (1991).

Neff, A., *Specified precision polynomial root isolation is in NC*, J. of Computer and Systems Sci., to appear (1993).

Pan, V., *Sequential and parallel complexity of approximate evaluation and polynomial zeros*, Comput. Math. Appl., Vol 14, No. 8 (1987), 591–622.

Renegar, J., *On the efficiency of Newton's Method in approximating allz eros of systems of complex polynomials*, Math. of Operations Research **12** (1987a).

Renegar, J., *On the worst case arithmetic complexity of approximating zeros of polynomials*, J. of Complexity **3** (1987b), 90–113.

Shub, M., *Some remarks on Bezout's Theorem and complexity theory*, to appear in "From Topology to Computation", Proceedings of the Smalefest, Hirsch, M., Marsden, J. and Shub, M. (Eds.) (1993).

Shub, M. and Smale, S., *Computational complexity: On the geometry of polynomials and a theory of cost*, I, Ann. Sci. École Norm. Sup. (4) **18** (1985), 107–142.

Shub, M. and Smale, S., *Computational complexity: On the geometry of polynomials and a theory of cost*, II, SIAM J. Comput. **15** (1986), 145–161.

Shub, M. and Smale, S., *Complexity of Bezout's Theorem I: Geometric aspects (referred to as Bez I)*, J. of the Amer. Math. Soc. **6** (1993), 459–501.

Shub, M. and Smale, S., *Complexity of Bezout's Theorem II: Volumes and probabilities (referred to as Bez II)*, Computational Algebraic Geometry (F. Eyssette and A. Galligo, Eds.), Progress in Mathematics, Vol 109, Birkhauser (1993), 267–285.

Shub, M. and Smale, S., *Complexity of Bezout's Theorem III: Condition number and packing (referred to as Bez III)*, J. of Complexity **9** (1993), 4–14.

Smale, S., *Global Analysis and Economics, Chapter 8*, in Handbook of Math. Economics, Vol 1, Arrow, K., Intrilligator, M. Eds., North Holland, New York, 1981.

Smale, S., *The fundamental theorem of algebra and complexity theory*, Bull. Amer. Math. Soc. (N.S.) **4** (1981), 1–36.

Smale, S., *Newton's Method estimates from data at one point*, in "The Merging of Disciplines: New Directions in Pure, Applied and Computation Math", Ewing, R., Gross, K. and Martin, C. (Eds.), Springer, New York (referred to as P.E.) (1986).

Smale, S., *Algorithms for solving equations*, Proceedings of the International Congress of Mathematics, Berkeley 1986, Amer. Math. Soc., Providence, R.I. (1987).

Stein, E. and Weiss, G., *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton Univ. Press, Princeton, N.J., 1971.

Weyl, H., *The Theory of Groups and Quantum Mechanics*, Dover, 1932.

Wilkinson, J., *Rounding Errors in Algebraic Processes*, Prentice Hall, Englewood Cliffs, N.J., 1963.

Wilkinson, J., *Sensitivity of Eigenvalues*, vol. 25, Utilities Mathematica, 1984, pp. 5–76.

Wozniakowski, H., *Numerical stability for solving non-linear equations*, Num. Math. **27** (1977), 373–390.